

## 4

### Estudo de casos

#### 4.1

##### Descrição

Para avaliar o desempenho do modelo NFHB-T2 proposto neste trabalho, foram utilizadas diversas bases de dados *benchmark* e aplicações reais de previsão e aproximação de funções. A primeira diz respeito ao “*benchmark*” para previsão da série caótica de *Mackey-Glass*, com diferentes níveis de ruído. As bases restantes referem-se a aplicações em sistemas dinâmicos, e em séries cujo comportamento considera-se que segue um processo estocástico: o “Movimento Geométrico Browniano”. Foram feitas também análises de desempenho em aplicações reais em diferentes setores: financeiro, elétrico e médico.

Os testes foram realizados usando o modelo NFHB-T2 com duas opções de particionamento: particionamento fixo, onde o método do *Gradient Descent* ajusta apenas os parâmetros dos conseqüentes das regras fuzzy do tipo 2; e particionamento adaptativo, onde o método do *Gradient Descent* ajusta os parâmetros dos antecedentes e dos conseqüentes das regras fuzzy do tipo 2. As duas opções para o particionamento das células filhas com a mesma variável de entrada, descritas na seção 3.4, foram avaliadas. Além disso, usou-se a opção 1 como método de seleção da seqüência de variáveis de entrada para o modelo, descrita na seção 3.6; esta opção baseia-se na execução de mini-sistemas ANFIS. Em todos os casos, para o processo de treinamento foi utilizado um conjunto de validação para interromper o treinamento, de forma a evitar o problema do super-aprendizado (*overfitting*).

As seções 4.2 e 4.3 apresentam os resultados obtidos com o modelo NFHB-T2 nos casos de aproximação de funções e previsão de séries, respectivamente. Nestas seções as tabelas resumem os resultados obtidos após várias tentativas, com diferentes configurações de parâmetros do modelo NFHB-T2, tais como o uso de particionamento fixo ou adaptativo, taxa de decomposição e opção para o particionamento das células filhas com mesma variável de entrada.

Nestas tabelas o melhor desempenho considerando o percentual de acerto no conjunto de teste está sombreado. A seção 4.4 apresenta comparações entre o desempenho do modelo NFHB-T2 e o de outros modelos, como o Neuro-Fuzzy Hierarquico BSP Takagi-Sugeno que faz uso de conjuntos fuzzy do tipo 1 [SOUZ99] e Redes Neurais “Multi Layer Perceptron”.

## 4.2

### Aproximação de funções: *Velocidade da Onda do Pulso de Sangue*

No caso de aproximação de funções, utiliza-se uma base de dados da área médica, desejando-se encontrar a velocidade da onda do pulso de sangue (VOP) para fins de diagnóstico patológico; valores grandes da VOP indicam algum tipo de patologia. Esta velocidade, mostrada na Figura 4.1, é definida como a velocidade média do pulso de sangue quando este se desloca da artéria carótida à artéria femural.

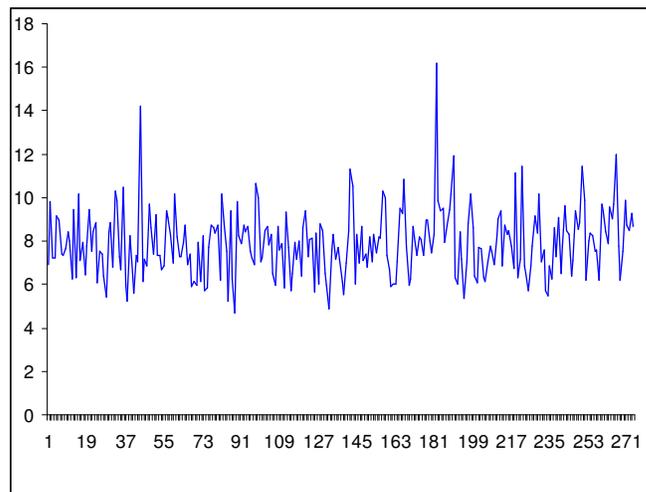


Figura 4.1 –Velocidade da Onda do Pulso de Sangue (VOP)

Os dados desta base são resultantes de um estudo realizado para definir diferentes valores normais esperados da VOP, para grupos de pessoas de diferentes idades. A VOP é, atualmente, medida através de um instrumento protótipo bastante caro. Deste modo, a possibilidade de inferir indiretamente seu valor, a partir de variáveis clínicas tradicionais, tais com idade, peso, altura,

pressão arterial, ritmo cardíaco e níveis de colesterol, é de grande importância clínica.

Uma série de 17 variáveis clínicas, além da VOP, foram medidas em um grupo de 392 pessoas, incluindo pacientes e voluntários. A Tabela 4.1 apresenta tais variáveis, que podem ser divididas em três grupos: antropométricas (1, 9, 10, 11, 12, 13, e 14), hemodinâmicas (2, 3, 4 e 15) e metabólicas (5, 6, 7, 8, 16 e 17).

Tabela 4.1 Variáveis Clínicas Medidas

Nº DA VARIÁVEL	NOME	TIPO DE VARIÁVEL
<b>1</b>	<b>Idade</b>	antropométricas
<b>2</b>	<b>Pressão Arterial Sistólica PAS</b>	hemodinâmicas
<b>3</b>	<b>Pressão Arterial Diastólica PAD</b>	hemodinâmicas
<b>4</b>	<b>Frequência Cardíaca</b>	hemodinâmicas
<b>5</b>	<b>Colesterol Total</b>	metabólicas
<b>6</b>	<b>Triglicerídeos TG</b>	metabólicas
<b>7</b>	<b>Colesterol HDL</b>	metabólicas
<b>8</b>	<b>Glicose</b>	metabólicas
<b>9</b>	<b>Peso</b>	antropométricas
<b>10</b>	<b>Altura</b>	antropométricas
<b>11</b>	<b>Cintura</b>	antropométricas
<b>12</b>	<b>Quadril</b>	antropométricas
13	Índice de massa corporal (BMI)	antropométricas
14	Relação Cintura / Quadril C/Q	antropométricas
15	Pressão de Pulso	hemodinâmicas
16	Colesterol LDL	metabólicas
17	Colesterol VLDL	metabólicas

Após uma análise do conjunto de variáveis, algumas foram excluídas por terem uma relação explícita com as outras. Este é o caso de: BMI – peso dividido pela altura elevada ao quadrado; C/Q – medida da cintura dividida pelo quadril; Pressão do pulso – pressão arterial sistólica menos a pressão arterial diastólica; Colesterol LDL e Colesterol VLDL – calculados a partir de outros níveis de colesterol e triglicerídeos. Portanto, só foram consideradas as 12 variáveis mostradas em negrito na Tabela 4.1. A seqüência de variáveis de entrada obtida através do método de seleção baseado em mini-sistemas ANFIS foi: 1, 10, 9, 3, 11, 5, 2, 12, 4, 6, 8, 7.

O conjunto possui ao todo 392 padrões com 12 atributos. Para a realização dos testes, o conjunto inicial de 392 padrões foi dividido, de forma aleatória, em três conjuntos: treinamento, validação e teste, em proporções de 70%, 20% e 10%,

respectivamente, resultando em 274 padrões para o conjunto de treinamento, 78 para validação e 40 para teste.

A Tabela 4.2 resume os resultados obtidos para a base de dados *Médicos*.

TABELA 4.2: Resultados obtidos com o modelo NFHB-T2 para a aproximação da VOP da base de dados Médicos.

Opção Particion. Células Filhas	Taxa Decomp. 0.001-0.05	Ajuste Anteced.	Num. Conseq.	MAPE Treino	MAPE Validação	MAPE Teste	U-Theil Teste
1	0.05	não	12	11.79	13.64	10.41	0.46
2	0.05	não	12	11.79	13.64	10.41	0.46
1	0.05	sim	90	10.57	14.00	12.66	0.55
2	0.05	sim	90	10.57	14.00	12.66	0.55
1	0.01	não	61	11.55	13.66	10.46	0.46
2	0.01	não	61	11.57	13.67	10.46	0.46
1	0.01	sim	247	9.91	13.31	11.37	0.52
2	0.01	sim	412	10.21	13.38	11.43	0.52

O erro MAPE (“Mean Absolute Percentage Error”) na Tabela 4.2 é calculado de acordo com a seguinte fórmula:

$$MAPE = \frac{\sum_{k=1}^N \left| \frac{a_k - y_k}{a_k} \right|}{N} \times 100 \% \quad \text{Equação 4.1}$$

onde:

$N$  = Número de padrões da base de dados

$a_k$  = Saída desejada para o  $k$ -ésimo padrão

$y_k$  = Saída obtida para o  $k$ -ésimo padrão

O coeficiente U de Theil na Tabela 4.2 mede o quanto os resultados estão melhores que uma previsão ingênua (i.e. “a melhor estimativa do próximo valor é o valor atual”). O coeficiente U de Theil é calculado pela expressão:

$$UTHEIL = \frac{\sqrt{\sum_{k=1}^N (a_k - y_k)^2}}{\sqrt{\sum_{k=1}^N (a_k - a_{k-1})^2}} \quad \text{Equação 4.2}$$

onde:

$N$  = Número de padrões da base de dados.

$a_k$  = Saída desejada para o  $k$ -ésimo padrão.

$y_k$  = Saída obtida para o  $k$ -ésimo padrão.

Assim, um coeficiente  $U$  de Theil menor do que 1 indica uma previsão melhor que a previsão ingênua. Portanto, quanto mais próximo de zero for este coeficiente, melhor a previsão.

Inspecionando-se a Tabela 4.2, observa-se um melhor desempenho de generalização para o modelo NFHB-T2 com taxa de decomposição de 0.05 e particionamento fixo. Neste caso, a opção para o particionamento das células filhas com a mesma variável de entrada descrita na seção 3.4 não altera os resultados. Uma justificativa é o fato de que a influência desta opção só é verificada quando as variáveis de entrada começam a se repetir na árvore, uma vez que o melhor resultado é obtido com uma árvore pequena (12 conseqüentes). Como a base de dados *Médicos* possui um número de variáveis de entrada grande (12 variáveis), a repetição não ocorre na árvore. Portanto pode-se dizer que nem todas as variáveis de entrada são aplicadas ao modelo NFHB-T2. Desta forma, pode não existir células com a mesma variável de entrada. Neste caso, o melhor desempenho foi obtido com o modelo NFHB-T2 de 5 níveis de hierarquia com variáveis de entrada 1, 10, 9, 3 e 11 para cada nível, respectivamente

A Figura 4.2 a seguir mostra as saídas do modelo NFHB-T2 para o conjunto de validação e teste da base de dados *Médicos*.

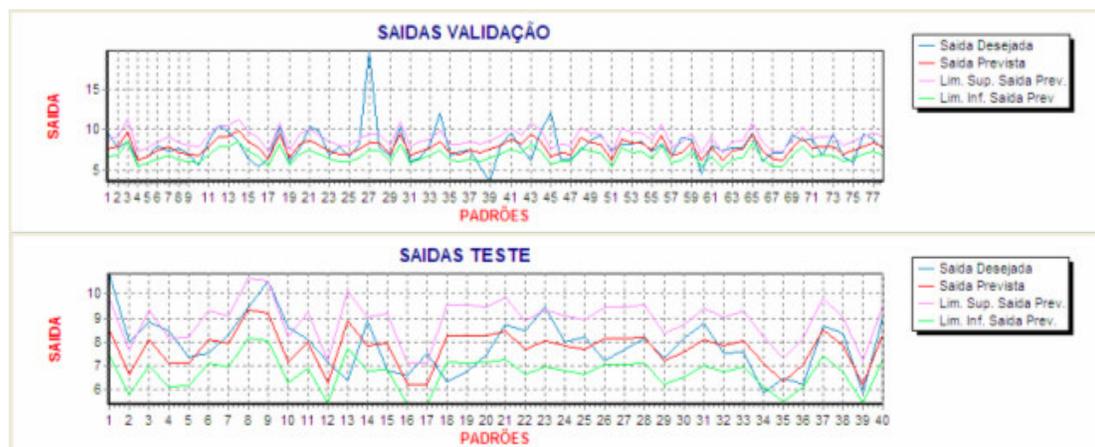


Figura 4.2 - Saídas do modelo NFHB-T2 para o conjunto de validação e teste para a base de dados *Médicos*.

Inspecionando-se a Figura 4.2 observa-se que os intervalos de confiança para as saídas defuzzificadas do modelo NFHB-T2 obtidos automaticamente através dos conjuntos fuzzy *intervalares* do tipo 1 de saída são coerentes, isto é, a maior parte das saídas desejadas encontra-se dentro destes intervalos de confiança.

### 4.3

#### Previsão de Séries Temporais

##### 4.3.1

##### Previsão da Série Caótica *Mackey-Glass*

A série caótica criada por Mackey e Glass é um dos *benchmarks* mais tradicionais desta área. Ela é dada pela seguinte equação:

$$x'_t = -0.1x_t + \frac{0.2x_{t-\tau}}{1 + x_{t-\tau}^{10}} \quad \text{Equação 4.3}$$

onde  $\tau=17$ .

Para este estudo de casos foram usados conjuntos de dados da série Mackey-Glass afetada por ruído aditivo uniforme de SNR (razão entre a variância do ruído e a variância dos dados) igual a 6.2% e a 18.6%. Estas séries são mostradas na Figura 4.3.

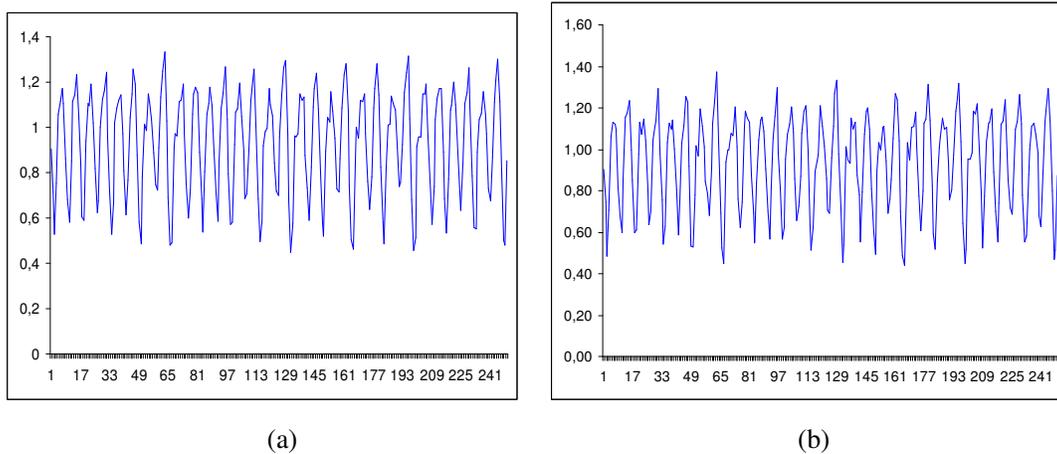


Figura 4.3 – (a) Série Mackey-Glass com ruído aditivo uniforme de SNR =6.2 %. (b) Série Mackey-Glass com ruído aditivo uniforme de SNR =18.6 %.

Estes dois conjuntos possuem, ao todo, 2182 padrões cada um. Para a realização dos testes, cada conjunto inicial de 2182 padrões foi dividido em três conjuntos: treinamento, validação e teste, em proporções de 70%, 20% e 10% respectivamente, resultando em 1527 padrões para o conjunto de treinamento, 437 para validação e 218 para teste.

Para a previsão destes dois conjuntos empregou-se a técnica do janelamento [SOUZ99], considerando-se janelas de tamanho igual a 4 (referentes aos valores anteriores). O tamanho da janela foi determinado através da análise de autocorrelação feita na série. A seqüência de variáveis de entrada obtida através do método de seleção baseado em mini-sistemas ANFIS foi:  $x_{t-2}, x_t, x_{t-3}, x_{t-1}$ .

A Tabela 4.3 resume os resultados da previsão um passo à frente obtidos para a base de dados *Mackey-Glass* com ruído de SNR igual a 6.2% e a 18.6%.

TABELA 4.3: Resultados obtidos com o modelo NFHB-T2 para previsão da base de dados *Mackey-Glass* com ruído de SNR=6.2% e SNR=18.6%.

SNR (%)	Opção Particion. Células Filhas	Taxa Decomp. 0.001-0.05	Ajuste Anteced.	Num. Conseq.	MAPE Treino	MAPE Validaç.	MAPE Teste	U-Theil Teste
6.2	1	0.05	não	32	2.722	2.313	3.094	0.161
6.2	2	0.05	não	32	2.839	2.466	3.233	0.166
6.2	1	0.05	sim	97	1.562	0.673	1.990	0.111
6.2	2	0.05	sim	106	1.770	1.023	2.289	0.127
6.2	1	0.01	não	160	2.276	1.818	2.556	0.133
6.2	2	0.01	não	161	2.295	1.842	2.574	0.134
6.2	1	0.01	sim	191	6.510	6.271	6.967	0.393
6.2	2	0.01	sim	412	4.749	4.084	4.952	0.295
18.6	1	0.05	não	29	5.321	3.339	6.581	0.333
18.6	2	0.05	não	29	4.877	2.553	6.091	0.308
18.6	1	0.05	sim	101	4.149	1.359	5.250	0.283
18.6	2	0.05	sim	87	4.340	1.478	5.269	0.287
18.6	1	0.01	não	151	5.427	3.929	6.324	0.341
18.6	2	0.01	não	159	4.735	2.556	5.787	0.312
18.6	1	0.01	sim	408	5.274	3.290	6.237	0.328
18.6	2	0.01	sim	234	7.637	5.632	8.543	0.437

Os erros MAPE e U-theil na Tabela 4.3 são calculados da mesma forma que no caso anterior, através das equações 4.1 e 4.2.

Inspecionando-se a Tabela 4.3, observa-se que os melhores desempenhos de generalização para a base de dados *Mackey-Glass* com ruídos de SNR igual a 6.2% e 18.6% foram obtidos com taxas de decomposição de 0.05, particionamentos adaptativos e opção 1 para os particionamentos das células

filhas com a mesma variável de entrada. Neste caso, os melhores desempenhos foram obtidos, respectivamente, com o modelo NFHB-T2 de 9 níveis de hierarquia com variáveis de entrada  $x_{t-2}, x_t, x_{t-3}, x_{t-1}, x_{t-2}, x_t, x_{t-3}, x_{t-1}, x_{t-2}$  para cada nível, respectivamente, e, com o modelo NFHB-T2 de 10 níveis de hierarquia com variáveis de entrada  $x_{t-2}, x_t, x_{t-3}, x_{t-1}, x_{t-2}, x_t, x_{t-3}, x_{t-1}, x_{t-2}, x_t$  para cada nível, respectivamente. Como já era esperado, observa-se que o erro *MAPE* cresce à medida que o SNR cresce.

As Figuras 4.4 e 4.5 a seguir mostram as saídas do modelo NFHB-T2 para o conjunto de validação e teste da base de dados *Mackey-Glass* para os dois níveis de ruído, respectivamente.

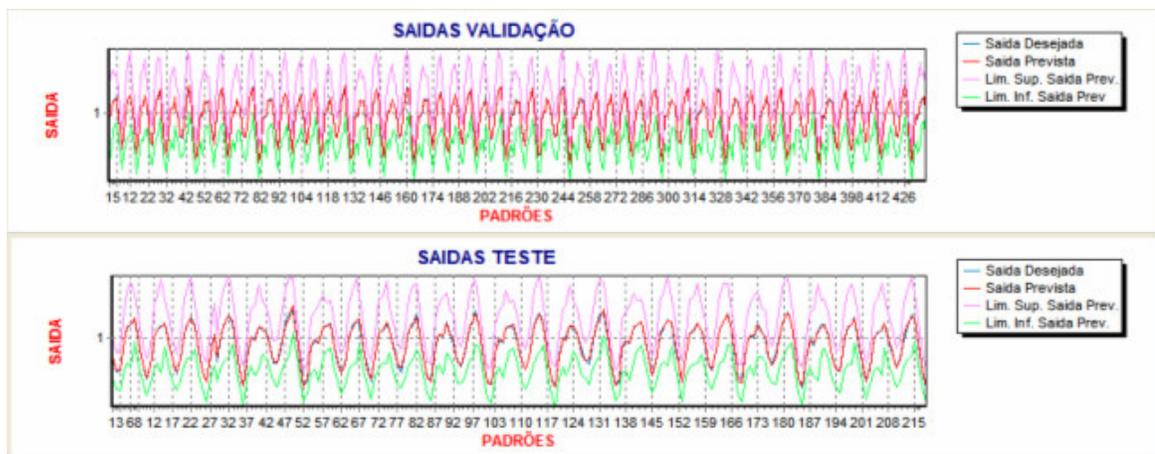


Figura 4.4 - Saídas do modelo NFHB-T2 para o conjunto de validação e teste da base de dados *Mackey-Glass* com ruído de SNR=6.2 %.

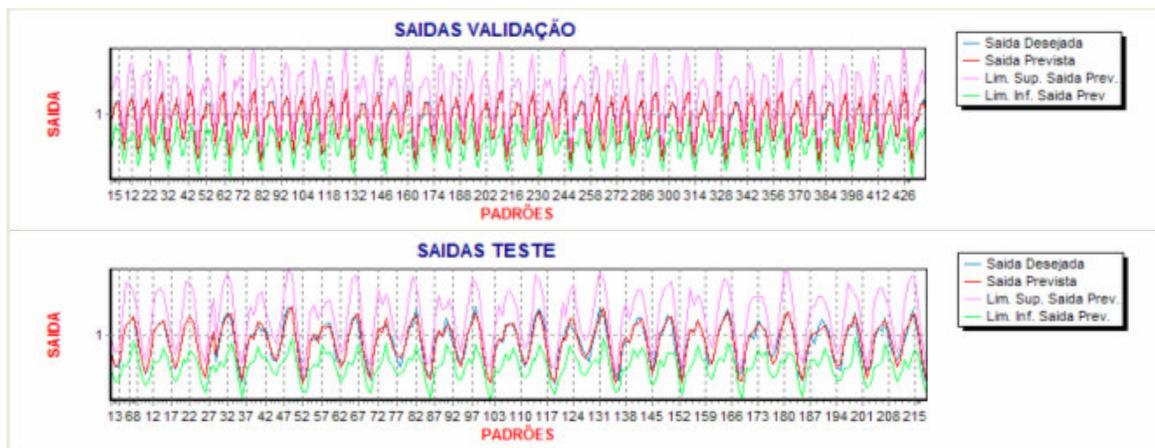


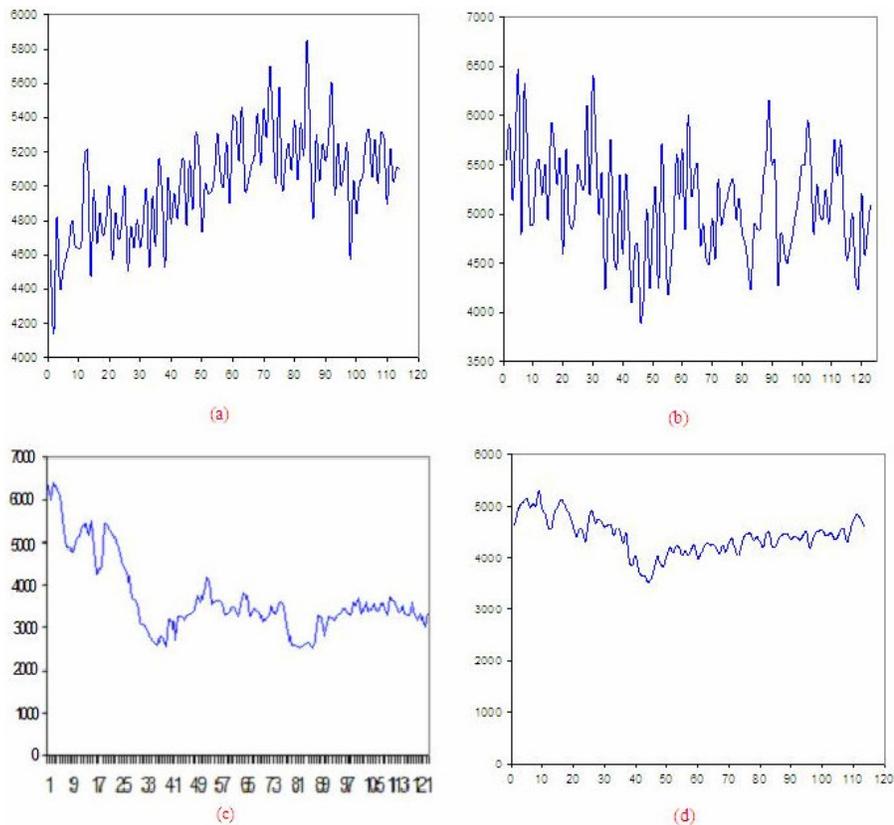
Figura 4.5 - Saídas do modelo NFHB-T2 para o conjunto de validação e teste da base de dados *Mackey-Glass* com ruído de SNR=18.6 %.

Da mesma forma que no caso da base de dados *Médicos*, os intervalos de confiança para as saídas defuzzificadas mostram coerência. Isto pode ser observado nas Figuras 4.4 e 4.5. Observa-se também que o modelo NFHB-T2 conseguiu reproduzir as séries Mackey-Glass com SNR igual a 6.2% e a 18.6 % muito bem.

### 4.3.2

#### Previsão de Séries Temporais de Negócios

Esta seção apresenta uma aplicação de previsão de séries temporais de negócios a partir de dados históricos mensais [NNFC06]. Para esta aplicação foram utilizadas seis séries de negócios diferentes, mostradas na Figura 4.6.



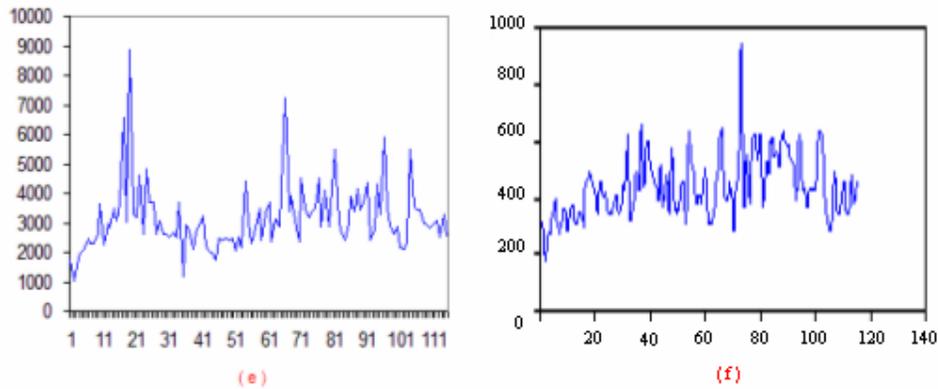


Figura 4.6 – (a) Série temporal de Negócios 1. (b) Série temporal de Negócios 2. (c) Série temporal de Negócios 3. (d) Série temporal de Negócios 4. (e) Série temporal de Negócios 5. (f) Série temporal de Negócios 6.

Os cinco conjuntos correspondentes às séries 1, 2, 3, 4, 5 e 6 possuem ao todo 114, 123, 121, 114, 111 e 115 padrões. Para a realização dos testes, cada conjunto inicial correspondente às cinco séries foi dividido em três conjuntos: treinamento, validação e teste, em proporções de 70%, 20% e 10% respectivamente.

Para a previsão das séries 1, 3 e 4 empregou-se de novo a técnica da janela considerando-se janelas de tamanho igual a 12 (referentes aos valores mensais passados) e informações adicionais (fora da janela) referentes ao ano e mês da previsão. Por outro lado, para a previsão da série 2 empregaram-se janelas de tamanho igual a 3 (referentes aos valores mensais passados); para as séries 5 e 6 empregaram-se janelas de tamanho igual a 12 (referentes aos valores mensais passados). O tamanho da janela foi determinado através da análise de autocorrelação efetuada na série.

A seqüência de variáveis de entrada obtida através do método de seleção baseado em mini-sistemas ANFIS foi:

Série 1:  $x_{t-11}, x_{t-3}, x_{t-8}, x_{t-2}, \text{mês}, \text{ano}, x_{t-9}, x_{t-6}, x_{t-10}, x_{t-4}, x_{t-1}, x_{t-7}, x_t$ .

Série 2:  $x_{t-2}, x_t, x_{t-1}$ .

Série 3:  $x_{t-7}, x_t, \text{ano}, x_{t-6}, x_{t-8}, x_{t-5}, \text{mês}, x_{t-3}, x_{t-1}, x_{t-4}, x_{t-2}, x_{t-9}, x_{t-11}, x_{t-10}$ .

Série 4:  $x_{t-4}, x_t, x_{t-2}, \text{mês}, x_{t-3}, x_{t-5}, x_{t-7}, x_{t-1}, x_{t-6}, \text{ano}, x_{t-8}, x_{t-11}, x_{t-10}, x_{t-9}$ .

Série 5:  $x_{t-10}, x_{t-1}, x_{t-7}, x_{t-4}, x_{t-6}, x_{t-3}, x_{t-2}, x_t, x_{t-11}, x_{t-8}, x_{t-5}, x_{t-9}$ .

Série 6:  $x_{t-11}, x_{t-6}, x_{t-2}, x_{t-4}, x_{t-10}, x_{t-3}, x_{t-1}, x_t, x_{t-5}, x_{t-7}, x_{t-9}, x_{t-8}$ .

Os resultados da previsão um passo à frente obtidos para estas cinco séries de *Negócios* estão resumidos nas Tabelas 4.4, 4.5, 4.6, 4.7, 4.8 e 4.9, respectivamente.

TABELA 4.4: Resultados obtidos com o modelo NFHB-T2 para previsão da série de Negócios 1.

<b>Opção Particion. Células Filhas</b>	<b>Taxa Decomp. 0.001-0.05</b>	<b>Ajuste Anteced.</b>	<b>Num. Conseq.</b>	<b>MAPE Treino</b>	<b>MAPE Validaç.</b>	<b>MAPE Teste</b>	<b>U-Theil Teste</b>
1	0.05	não	2	3.643	3.081	2.296	0.086
2	0.05	não	2	3.643	3.081	2.296	0.086
1	0.05	sim	52	4.088	3.064	1.767	0.072
2	0.05	sim	52	4.088	3.064	1.767	0.072
1	0.01	não	2	3.643	3.081	2.296	0.086
2	0.01	não	2	3.643	3.081	2.296	0.086
1	0.01	sim	141	3.520	3.385	2.183	0.084
2	0.01	sim	141	3.520	3.385	2.183	0.084

TABELA 4.5: Resultados obtidos com o modelo NFHB-T2 para previsão da série de Negócios 2.

<b>Opção Particion. Células Filhas</b>	<b>Taxa Decomp. 0.001-0.05</b>	<b>Ajuste Anteced.</b>	<b>Num. Conseq.</b>	<b>MAPE Treino</b>	<b>MAPE Validaç.</b>	<b>MAPE Teste</b>	<b>U-Theil Teste</b>
1	0.05	não	22	7.258	7.137	7.612	0.264
2	0.05	não	33	7.145	6.774	7.111	0.251
1	0.05	sim	79	9.695	5.527	10.08	0.367
2	0.05	sim	33	9.020	5.821	7.877	0.278
1	0.01	não	23	9.646	7.289	8.466	0.295
2	0.01	não	23	9.677	7.330	8.518	0.297
1	0.01	sim	298	8.368	5.145	8.714	0.288
2	0.01	sim	159	10.13	5.830	8.804	0.339

TABELA 4.6: Resultados obtidos com o modelo NFHB-T2 para previsão da série de Negócios 3

<b>Opção Particion. Células Filhas</b>	<b>Taxa Decomp. 0.001-0.05</b>	<b>Ajuste Anteced.</b>	<b>Num. Conseq.</b>	<b>MAPE Treino</b>	<b>MAPE Validaç.</b>	<b>MAPE Teste</b>	<b>U-Theil Teste</b>
1	0.05	sim	26	8.83	4.28	3.38	0.159
2	0.05	sim	26	8.83	4.28	3.38	0.159
1	0.01	sim	136	11.32	4.33	4.58	0.199
2	0.01	sim	136	11.32	4.33	4.58	0.199

TABELA 4.7: Resultados obtidos com o modelo NFHB-T2 para previsão da série de Negócios 4.

<b>Opção Particion. Células Filhas</b>	<b>Taxa Decomp. 0.001-0.05</b>	<b>Ajuste Anteced.</b>	<b>Num. Conseq.</b>	<b>MAPE Treino</b>	<b>MAPE Validaç.</b>	<b>MAPE Teste</b>	<b>U-Theil Teste</b>
1	0.05	não	28	2.732	1.637	2.161	0.090
2	0.05	não	28	2.732	1.637	2.161	0.090
1	0.05	sim	17	6.766	2.007	4.474	0.199
2	0.05	sim	17	6.766	2.007	4.474	0.199
1	0.01	não	43	8.931	3.196	6.660	0.264
2	0.01	não	43	8.931	3.196	6.660	0.264
1	0.01	sim	76	9.563	2.016	4.529	0.202
2	0.01	sim	76	9.563	2.016	4.529	0.202

TABELA 4.8: Resultados obtidos com o modelo NFHB-T2 para previsão da série de Negócios 5.

<b>Opção Particion. Células Filhas</b>	<b>Taxa Decomp. 0.001-0.05</b>	<b>Ajuste Anteced.</b>	<b>Num. Conseq.</b>	<b>MAPE Treino</b>	<b>MAPE Validaç.</b>	<b>MAPE Teste</b>	<b>U-Theil Teste</b>
1	0.05	não	10	27.39	23.40	13.58	0.401
2	0.05	não	10	27.39	23.40	13.58	0.401
1	0.05	sim	58	23.21	21.51	13.23	0.407
2	0.05	sim	58	23.21	21.51	13.23	0.407
1	0.01	não	36	27.36	23.38	13.67	0.402
2	0.01	não	36	27.36	23.38	13.67	0.402
1	0.01	sim	209	25.00	23.06	12.85	0.382
2	0.01	sim	209	25.00	23.06	12.85	0.382
1	0.001	sim	385	24.54	22.39	12.32	0.375

TABELA 4.9: Resultados obtidos com o modelo NFHB-T2 para previsão da série de Negócios 6.

<b>Opção Particion. Células Filhas</b>	<b>Taxa Decomp. 0.001-0.05</b>	<b>Ajuste Anteced.</b>	<b>Num. Conseq.</b>	<b>MAPE Treino</b>	<b>MAPE Validaç.</b>	<b>MAPE Teste</b>	<b>U-Theil Teste</b>
1	0.05	não	7	18.03	18.35	19.96	0.631
2	0.05	não	7	18.03	18.35	19.96	0.631
1	0.05	sim	36	15.53	16.63	16.98	0.587
2	0.05	sim	36	15.53	16.63	16.98	0.587
1	0.01	não	33	17.76	18.35	19.49	0.614
2	0.01	não	33	17.76	18.35	19.49	0.614
1	0.01	sim	257	10.18	14.39	12.26	0.394
2	0.01	sim	257	10.18	14.39	12.26	0.394

Inspecionando-se as tabelas, observa-se que os melhores desempenhos de generalização para as séries de negócios 1 e 3 foram obtidos com taxa de decomposição de 0.05 e particionamento adaptativo; já no caso da série de negócios 4, o melhor desempenho de generalização foi obtido com taxa de decomposição de 0.05 e particionamento fixo.

Em todos estes casos, a opção para o particionamento das células filhas com a mesma variável de entrada não altera os resultados, a justificativa para tal é similar ao caso da base de dados *Médicos*, com a diferença de que nessas séries utilizaram-se 14 variáveis de entrada, e os melhores resultados foram obtidos com árvores pequenas de 52, 26 e 28 conseqüentes respectivamente.

O melhor desempenho de generalização para a série de negócios 2 foi obtido com taxa de decomposição de 0.05, particionamento fixo e opção 2 para o particionamento das células filhas com a mesma variável de entrada, conforme mostrado na Tabela 4.5. Observa-se também nesta tabela que a opção para o particionamento das células filhas com a mesma variável de entrada altera os resultados. A justificativa para isto é que utilizaram-se somente 3 variáveis de entrada para a previsão desta série. Por outro lado, os melhores desempenhos de generalização para as séries de negócios 5 e 6 foram obtidos com taxas de decomposição de 0.001 e 0.01, respectivamente, ambas com particionamento adaptativo e opção 1 para o particionamento das células filhas com a mesma variável de entrada, conforme mostrado nas Tabelas 4.8 e 4.9, respectivamente.

Os modelos que apresentaram os melhores desempenhos foram:

Série 1: NFHB-T2 de 7 níveis de hierarquia com variáveis de entrada  $x_{t-11}$ ,  $x_{t-3}$ ,  $x_{t-8}$ ,  $x_{t-2}$ , *mês*, *ano*,  $x_{t-9}$  para cada nível, respectivamente;

Série 2: NFHB-T2 de 6 níveis de hierarquia com variáveis de entrada  $x_{t-2}$ ,  $x_t$ ,  $x_{t-1}$ ,  $x_{t-2}$ ,  $x_t$ ,  $x_{t-1}$  para cada nível, respectivamente;

Série 3: NFHB-T2 de 6 níveis de hierarquia com variáveis de entrada  $x_{t-7}$ ,  $x_t$ , *ano*,  $x_{t-6}$ ,  $x_{t-8}$ ,  $x_{t-5}$  para cada nível, respectivamente;

Série 4: NFHB-T2 de 5 níveis de hierarquia com variáveis de entrada  $x_{t-4}$ ,  $x_t$ ,  $x_{t-2}$ , *mês*,  $x_{t-3}$  para cada nível, respectivamente;

Série 5: NFHB-T2 de 13 níveis de hierarquia com variáveis de entrada  $x_{t-10}$ ,  $x_{t-1}$ ,  $x_{t-7}$ ,  $x_{t-4}$ ,  $x_{t-6}$ ,  $x_{t-3}$ ,  $x_{t-2}$ ,  $x_t$ ,  $x_{t-11}$ ,  $x_{t-8}$ ,  $x_{t-5}$ ,  $x_{t-9}$ ,  $x_{t-10}$  para cada nível respectivamente;

Série 6: modelo NFHB-T2 de 9 níveis de hierarquia com variáveis de entrada  $x_{t-11}$ ,  $x_{t-6}$ ,  $x_{t-2}$ ,  $x_{t-4}$ ,  $x_{t-10}$ ,  $x_{t-3}$ ,  $x_{t-1}$ ,  $x_t$ ,  $x_{t-5}$  para cada nível, respectivamente.

Das Tabelas 4.4, 4.5, 4.6, 4.7, 4.8 e 4.9 observa-se também que, para as mesmas configurações de parâmetros do modelo NFHB-T2 (particionamento fixo ou adaptativo e opção para o particionamento das células filhas com a mesma variável de entrada), quanto menor o valor da taxa de decomposição, maior é o número de conseqüentes.

As Figuras 4.7, 4.8, 4.9, 4.10, 4.11 e 4.12, a seguir, mostram as saídas do modelo NFHB-T2 para o conjunto de validação e teste da base de dados de *Negócios*.

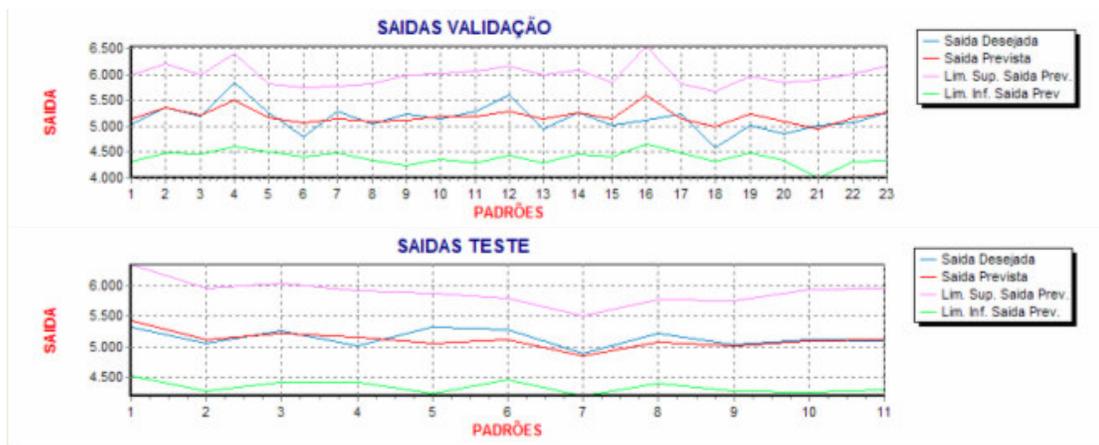


Figura 4.7 - Saídas do modelo NFHB-T2 para o conjunto de validação e teste da base de dados para a série temporal de Negócios 1.

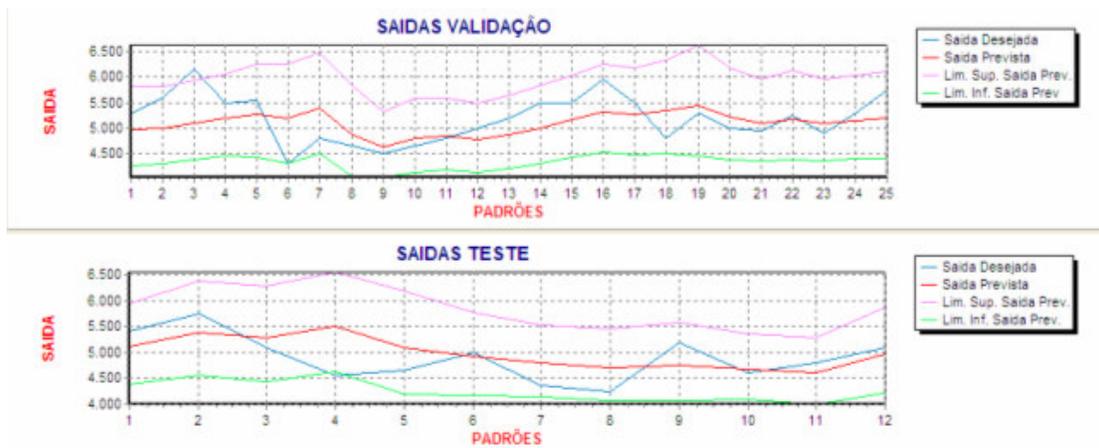


Figura 4.8 - Saídas do modelo NFHB-T2 para o conjunto de validação e teste da base de dados para a série temporal de Negócios 2.

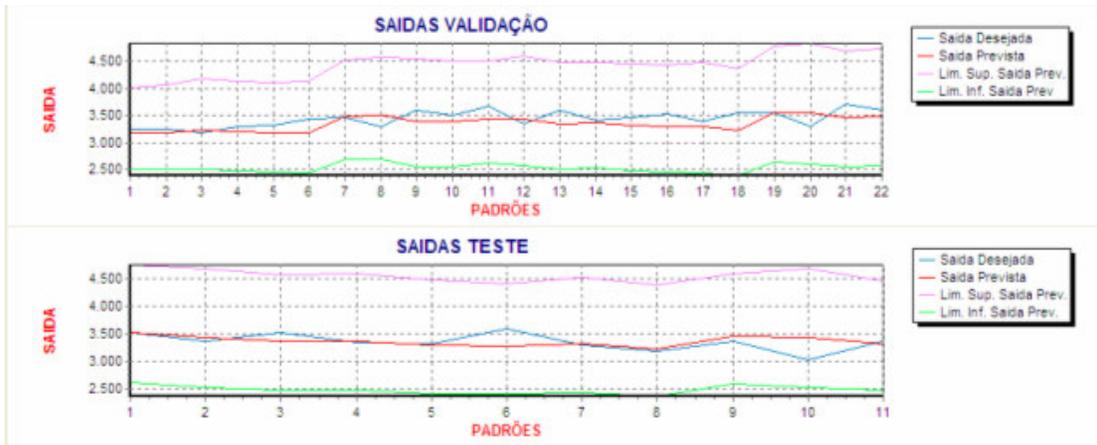


Figura 4.9 - Saídas do modelo NFHB-T2 para o conjunto de validação e teste da base de dados para a série temporal de Negócios 3.

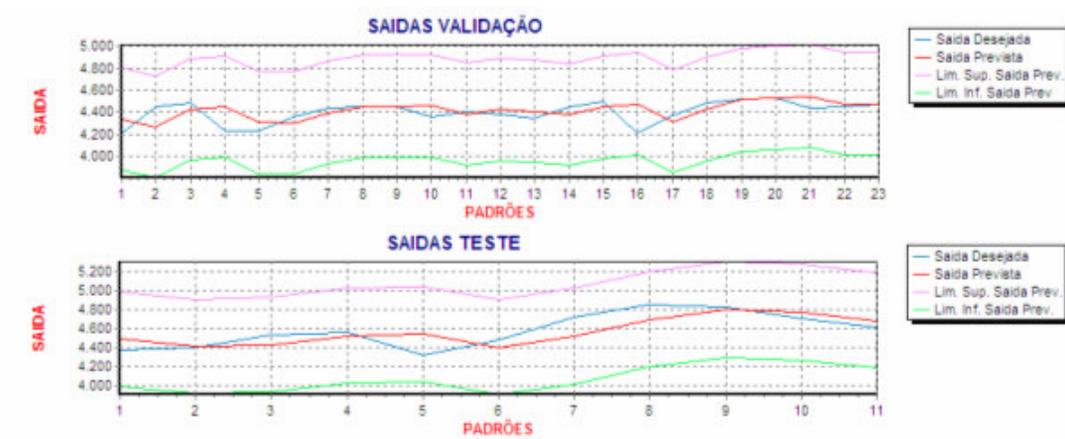


Figura 4.10 - Saídas do modelo NFHB-T2 para o conjunto de validação e teste da base de dados para a série temporal de Negócios 4.

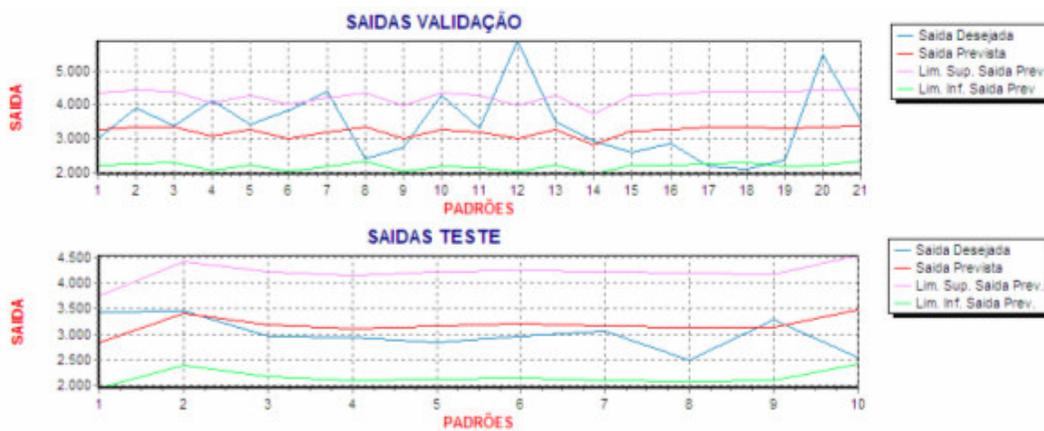


Figura 4.11 - Saídas do modelo NFHB-T2 para o conjunto de validação e teste da base de dados para a série temporal de Negócios 5.

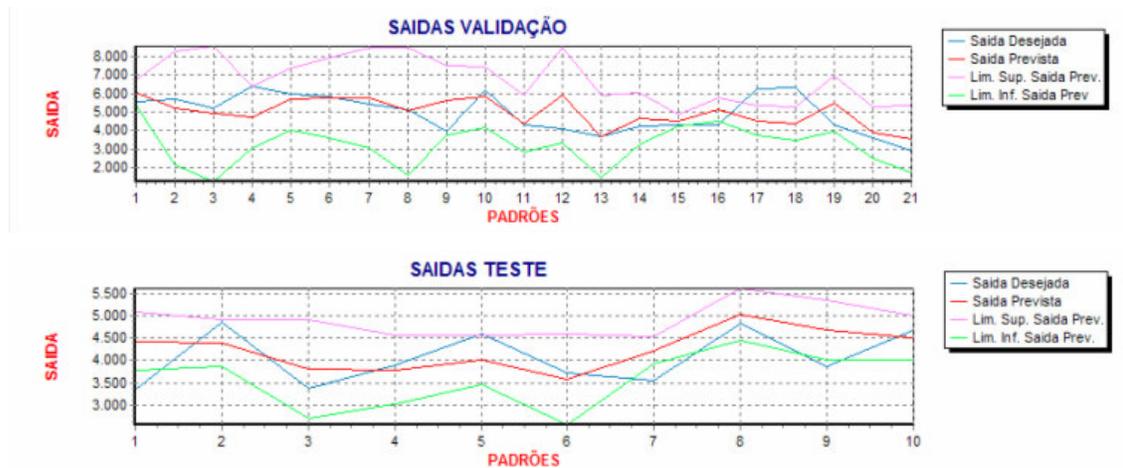


Figura 4.12 - Saídas do modelo NFHB-T2 para o conjunto de validação e teste da base de dados para a série temporal de Negócios 6.

Dos gráficos, observa-se novamente uma coerência nos intervalos de confiança obtidos para as saídas defuzzificadas do modelo NFHB-T2.

### 4.3.3

#### Previsão de Séries Temporais que seguem um Processo Estocástico: *Movimento Geométrico Browniano*

Esta seção apresenta a aplicação do modelo NFHB-T2 na previsão de séries temporais que seguem um processo estocástico: o Movimento Geométrico Browniano (MGB)<sup>1</sup>. Para esta aplicação utilizaram-se três séries com valores de tendência e volatilidade diferentes: a série 1 segue um MGB com tendência de 0.1 e volatilidade de 0.25; para a série 2 a tendência é de 0.1 e volatilidade é de 0.5; na série 3 tem-se uma tendência de 0.3 e uma volatilidade de 0.5. Estas três séries são mostradas na Figura 4.13.

<sup>1</sup> O Movimento Geométrico Browniano é um modelo amplamente usado em processos estocásticos em finanças. [DIXI94]

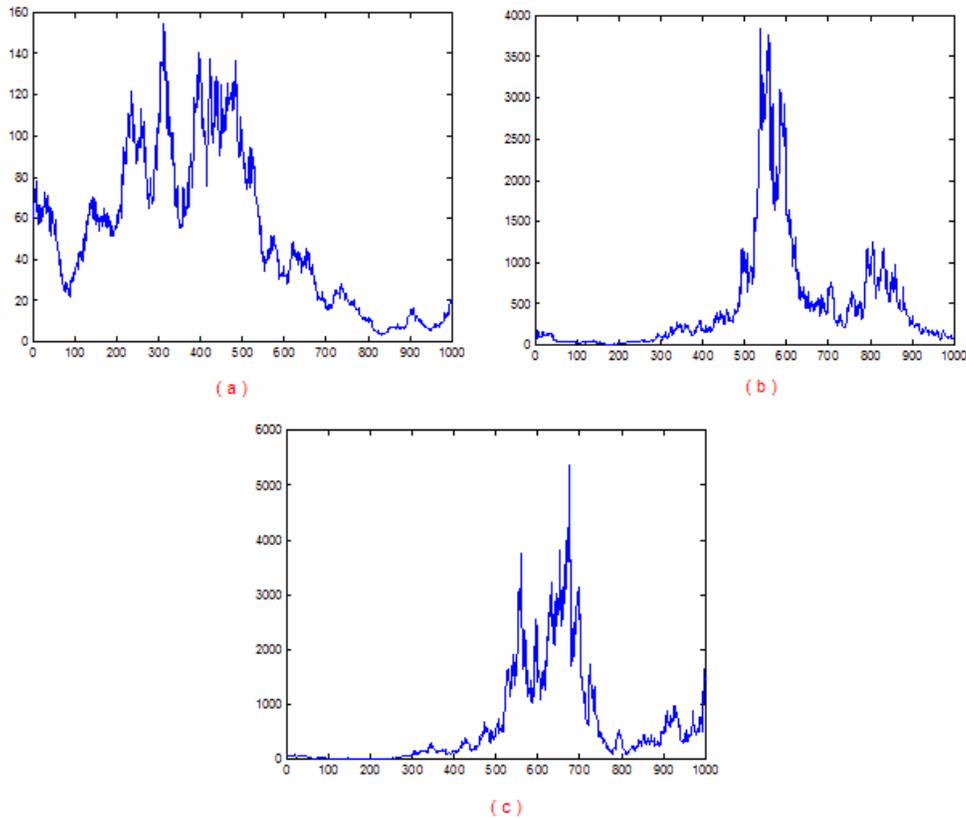


Figura 4.13 – (a) Série MGB 1. (b) Série MGB 2. (c) Série MGB 3.

Os três conjuntos de dados correspondentes às séries 1, 2 e 3 possuem ao todo 988 padrões. Para a realização dos testes, cada série foi dividida nos conjuntos de treinamento, validação e teste, nas mesmas proporções dos testes anteriores, resultando em 692 padrões para o conjunto de treinamento, 197 para validação e 99 para teste.

Da mesma forma que no caso das séries de *Negócios*, para a previsão das séries 1, 2 e 3 consideraram-se janelas de tamanho igual a 12 (referentes aos valores passados). O tamanho da janela foi determinado através da análise de autocorrelação efetuada na série.

A seqüência de variáveis de entrada obtida através dos mini-sistemas ANFIS foi:

Série 1:  $x_t, x_{t-1}, x_{t-2}, x_{t-3}, x_{t-5}, x_{t-4}, x_{t-7}, x_{t-6}, x_{t-8}, x_{t-9}, x_{t-11}, x_{t-10}$ .

Série 2:  $x_{t-6}, x_t, x_{t-1}, x_{t-2}, x_{t-5}, x_{t-9}, x_{t-8}, x_{t-7}, x_{t-11}, x_{t-10}, x_{t-3}, x_{t-4}$ .

Série 3:  $x_t, x_{t-2}, x_{t-1}, x_{t-10}, x_{t-7}, x_{t-6}, x_{t-11}, x_{t-9}, x_{t-5}, x_{t-4}, x_{t-8}, x_{t-3}$ .

Uma síntese dos resultados da previsão um passo à frente obtidos para estas três séries pode ser visto nas Tabelas 4.10, 4.11 e 4.12 a seguir.

TABELA 4.10: Resultados obtidos com o modelo NFHB-T2 para previsão da série que segue um MGB 1.

Opção Particion. Células Filhas	Taxa Decomp. 0.001-0.05	Ajuste Anteced.	Num. Conseq.	MAPE Treino	MAPE Validaç.	MAPE Teste	U-Theil Teste
1	0.05	não	14	6.180	6.511	6.968	0.514
2	0.05	não	14	6.180	6.511	6.968	0.514
1	0.05	sim	51	8.132	9.157	9.702	0.751
2	0.05	sim	51	8.132	9.157	9.702	0.751
1	0.01	não	44	6.172	6.531	6.903	0.511
2	0.01	não	44	6.172	6.531	6.903	0.511
1	0.01	sim	228	16.42	18.48	17.11	1.212
2	0.01	sim	248	12.25	16.82	16.62	1.183

TABELA 4.11: Resultados obtidos com o modelo NFHB-T2 para previsão da série que segue um MGB 2.

Opção Particion. Células Filhas	Taxa Decomp. 0.001-0.05	Ajuste Anteced.	Num. Conseq.	MAPE Treino	MAPE Validaç.	MAPE Teste	U-Theil Teste
1	0.05	não	7	42.61	16.91	26.78	0.856
2	0.05	não	7	42.61	16.91	26.78	0.856
1	0.05	sim	20	29.41	11.67	12.56	0.558
2	0.05	sim	20	29.41	11.67	12.56	0.558
1	0.01	não	24	38.28	13.02	15.05	0.592
2	0.01	não	24	38.28	13.02	15.05	0.592
1	0.01	sim	110	32.81	15.19	18.64	0.723
2	0.01	sim	110	32.81	15.19	18.64	0.723

TABELA 4.12: Resultados obtidos com o modelo NFHB-T2 para previsão da série de que segue um MGB 3.

Opção Particion. Células Filhas	Taxa Decomp. 0.001-0.05	Ajuste Anteced.	Num. Conseq.	MAPE Treino	MAPE Validaç.	MAPE Teste	U-Theil Teste
1	0.05	não	7	66.19	12.57	12.12	0.965
2	0.05	não	7	66.19	12.57	12.12	0.965
1	0.05	sim	62	63.04	13.23	12.68	0.991
2	0.05	sim	33	42.49	14.12	13.15	1.033
1	0.01	não	22	68.95	12.83	12.46	0.993
2	0.01	não	22	68.95	12.83	12.46	0.993

Observa-se que o modelo NFHB-T2 com taxa de decomposição de 0.01 e particionamento fixo obteve o melhor desempenho de generalização para a série 1; já no caso da série 2, o melhor desempenho de generalização foi obtido com o modelo NFHB-T2 com taxa de decomposição de 0.05 e particionamento adaptativo. O melhor desempenho de generalização para a série 3 foi obtido com o modelo NFHB-T2 com taxa de decomposição de 0.05 e particionamento fixo. Em todos estes casos, novamente a opção para o particionamento das células filhas com a mesma variável de entrada não altera os resultados; isto se deve aos motivos expostos na seção 4.3.2.

É importante ressaltar que os modelos que apresentaram os melhores desempenhos foram:

Série 1: NFHB-T2 de 7 níveis de hierarquia com variáveis de entrada  $x_t, x_{t-1}, x_{t-2},$

$x_{t-3}, x_{t-5}, x_{t-4}, x_{t-7}$  para cada nível, respectivamente;

Série 2: NFHB-T2 de 5 níveis de hierarquia com variáveis de entrada  $x_{t-6}, x_t, x_{t-1},$

$x_{t-2}, x_{t-5}$  para cada nível, respectivamente;

Série 3: NFHB-T2 de 4 níveis de hierarquia com variáveis de entrada  $x_t, x_{t-2}, x_{t-1},$

$x_{t-10}$  para cada nível, respectivamente;

As Figuras 4.14, 4.15 e 4.16 a seguir mostram as saídas do modelo NFHB-T2 para o conjunto de validação e teste da base de dados de *MGB*.

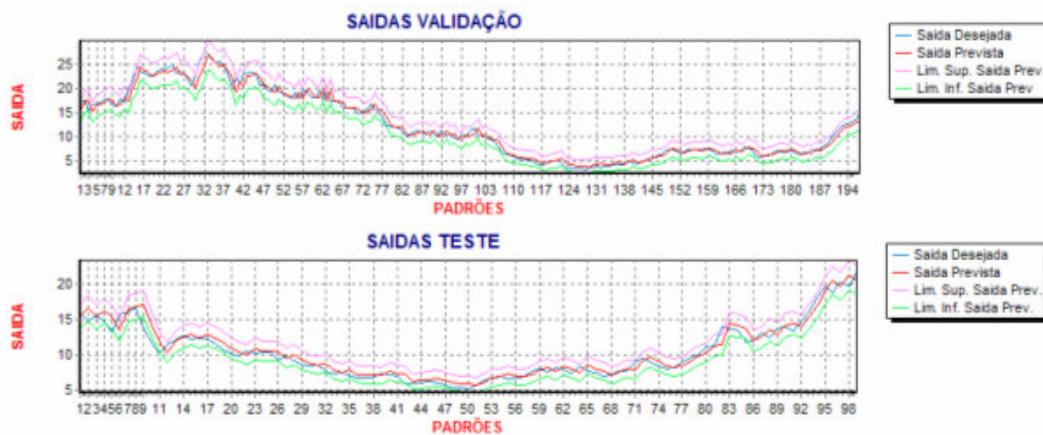


Figura 4.14 - Saídas do modelo NFHB-T2 para o conjunto de validação e teste da base de dados para a série que segue um MGB 1.

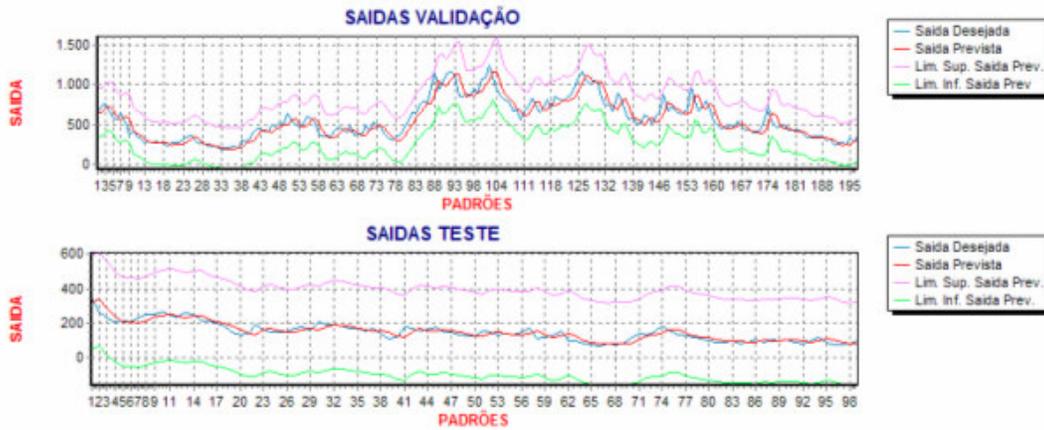


Figura 4.15 - Saídas do modelo NFHB-T2 para o conjunto de validação e teste da base de dados para a série que segue um MGB 2.

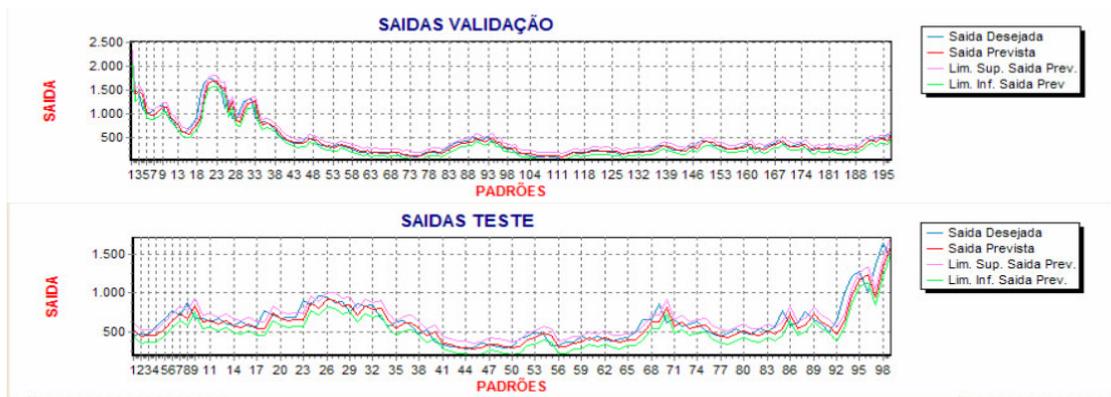


Figura 4.16 - Saídas do modelo NFHB-T2 para o conjunto de validação e teste da base de dados para a série que segue um MGB 3.

Da mesma forma que nos casos anteriores, das Figuras 4.14, 4.15 e 4.16 observa-se que a maioria das saídas desejadas encontra-se dentro dos intervalos de confiança obtidos para as saídas defuzzificadas do modelo NFHB-T2. Observa-se, também, que o modelo NFHB-T2 conseguiu reproduzir as séries muito bem na maioria dos casos.

#### 4.3.4

#### Previsão de Séries Temporais de Retornos

Esta seção apresenta uma aplicação de previsão de séries de retornos. Nesta aplicação, duas séries de retornos semanais diferentes – Brahma ON, Bemge ON – foram utilizadas (Figura 4.17).

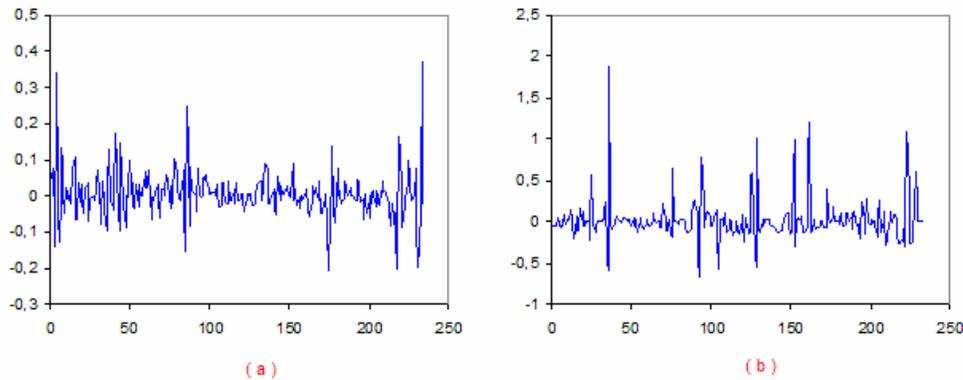


Figura 4.17 – (a) Série temporal de retornos Brahma ON. (b) Série temporal de retornos Bemge ON.

As séries Brahma ON e Bemge ON possuem ao todo 224 e 228 padrões, respectivamente. Estas séries foram divididas em conjuntos de treinamento, validação e teste, da mesma forma que nos experimentos anteriores.

Neste caso, considerou-se janelas de tamanho igual a 10 (referentes aos valores passados) para o caso da previsão da série Brahma ON, já no caso da série Bemge ON considerou-se janelas de tamanho igual a 6. O tamanho da janela foi determinado através da análise de autocorrelação efetuada na série. Para estas duas séries, a seqüência de variáveis de entrada obtida através do método de seleção baseado em mini-sistemas ANFIS foi:

Série 1:  $x_{t-1}, x_t, x_{t-8}, x_{t-7}, x_{t-3}, x_{t-5}, x_{t-6}, x_{t-4}, x_{t-2}, x_{t-9}$ .

Série 2:  $x_t, x_{t-2}, x_{t-1}, x_{t-3}, x_{t-4}, x_{t-5}$ .

As métricas de erro utilizadas foram MAD, NRMSE e MSE descritos abaixo. Não se utilizou o MAPE neste caso por existirem vários valores iguais a zero.

O erro MAD representa o desvio absoluto médio e é calculado segundo:

$$MAD = \frac{1}{N} \sum_{i=1}^n |a_k - y_k| \quad \text{Equação 4.4}$$

onde:

$N$  = Número de padrões da base de dados.

$a_k$  = Saída desejada para o  $k$ -ésimo padrão.

$y_k$  = Saída obtida para o  $k$ -ésimo padrão.

O erro NRMSE representa o erro quadrático médio normalizado:

$$NRMSE = \sqrt{\frac{\sum_{t=1}^n (a_k - y_k)^2}{\sum_{t=1}^n (a_k - Media)^2}}$$

Equação 4.5

onde:

*Media* = média das saídas desejadas para todos os padrões.

O erro MSE representa o erro quadrático médio:

$$MSE = \frac{1}{N} \sum_{t=1}^n (a_k - y_k)^2$$

Equação 4.6

Os resultados da previsão um passo à frente obtidos para estas duas séries de retornos estão mostrados nas Tabelas 4.13 e 4.14.

TABELA 4.13: Resultados obtidos com o modelo NFHB-T2 para previsão da série de retornos Brahma ON.

Opção Partic. Células Filhas	Taxa Decom. 0.001-0.05	Ajuste Antec.	N ° Consq	MAD Treino	MAD Valid	MAD Teste	NRMSE Treino	NRMSE Valid	NRMSE Teste	MSE Treino	MSE Valid	MSE Teste	U Theil Teste
1	0.05	não	16	0.1060	0.1020	0.1221	2.4150	2.2400	1.3670	0.0137	0.0124	0.0263	0.9639
2	0.05	não	16	0.1060	0.1020	0.1221	2.4150	2.2400	1.3670	0.0137	0.0124	0.0263	0.9639
1	0.01	não	98	0.0432	0.0421	0.1067	1.3060	1.2730	1.2720	0.0040	0.0040	0.0227	0.8963
2	0.01	não	98	0.0432	0.0421	0.1067	1.3060	1.2730	1.2720	0.0040	0.0040	0.0227	0.8963

TABELA 4.14: Resultados obtidos com o modelo NFHB-T2 para previsão da série de retornos Bemge ON.

Opção Partic. Células Filhas	Taxa Decom. 0.001-0.05	Ajuste Antec.	N ° Consq	MAD Treino	MAD Valid	MAD Teste	NRMSE Treino	NRMSE Valid	NRMSE Teste	MSE Treino	MSE Valid	MSE Teste	U Theil Teste
1	0.05	não	25	0.1632	0.1003	0.2508	1.1010	1.0660	1.2360	0.0872	0.0189	0.1429	0.8920
2	0.05	não	25	0.1632	0.1003	0.2508	1.1010	1.0660	1.2360	0.0872	0.0189	0.1429	0.8920
1	0.05	sim	71	0.1855	0.1106	0.2028	1.2990	1.2040	1.2070	0.1214	0.0241	0.1363	0.8712
2	0.05	sim	30	0.1905	0.1219	0.2354	1.3440	1.2880	1.3070	0.1299	0.0276	0.1599	0.9433
1	0.01	sim	300	0.2033	0.1207	0.2515	1.2710	1.1710	1.360	0.1162	0.0228	0.1730	0.9814
2	0.01	sim	75	0.1906	0.1269	0.2320	1.4100	1.3460	1.3130	0.1430	0.0301	0.1612	0.9472

O melhor desempenho de generalização para a série de retornos Brahma ON foi obtido com o modelo NFHB-T2 com taxa de decomposição de 0.01 e particionamento fixo. Neste caso, novamente a opção para o modo de particionamento das células filhas com a mesma variável de entrada não altera os resultados, pelos motivos expostos na seção 4.3.2. Para o caso da série de retornos Bemge ON, o melhor desempenho de generalização foi obtido com taxa de decomposição de 0.05, particionamento adaptativo e opção 1 para o particionamento das células filhas com a mesma variável de entrada. Isto pode ser observado nas Tabelas 4.13 e 4.14.

Os modelos que apresentaram os melhores desempenhos foram:

Série 1: NFHB-T2 de 8 níveis de hierarquia com variáveis de entrada  $x_{t-1}, x_t, x_{t-8}, x_{t-7}, x_{t-3}, x_{t-5}, x_{t-6}, x_{t-4}$  para cada nível, respectivamente;

Série 2: NFHB-T2 de 14 níveis de hierarquia com variáveis de entrada  $x_t, x_{t-2}, x_{t-1}, x_{t-3}, x_{t-4}, x_{t-5}, x_t, x_{t-2}, x_{t-1}, x_{t-3}, x_{t-4}, x_{t-5}, x_{t-2}, x_{t-1}$  para cada nível, respectivamente;

As Figuras 4.18 e 4.19 a seguir mostram as saídas do modelo NFHB-T2 para o conjunto de validação e teste da base de dados de *Retornos*.

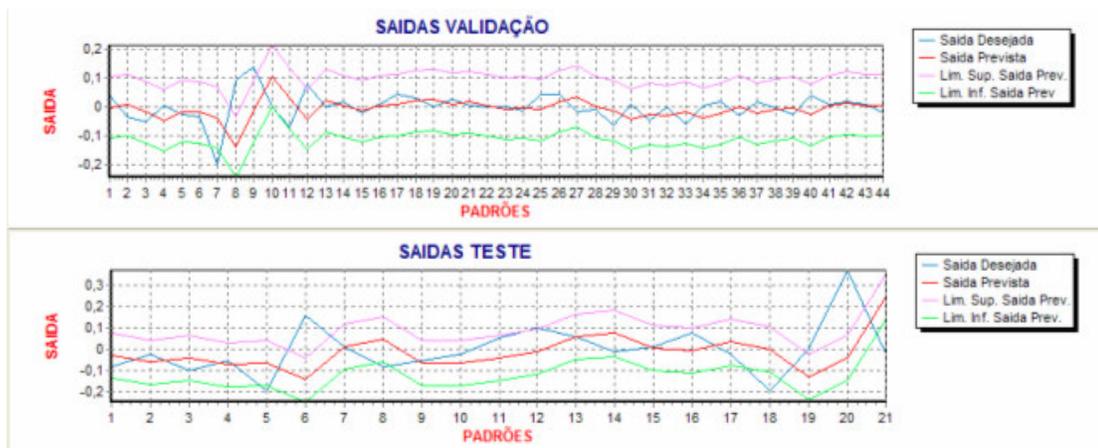


Figura 4.18 - Saídas do modelo NFHB-T2 para o conjunto de validação e teste da base de dados para a série de retornos Brahma ON

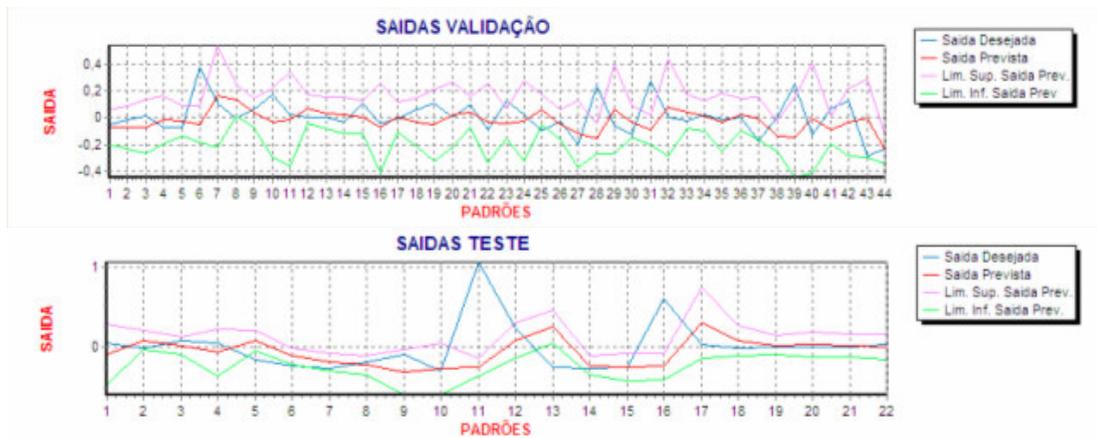


Figura 4.19 - Saídas do modelo NFHB-T2 para o conjunto de validação e teste da base de dados para a série de retornos Bemge ON

Como esperado, a maioria das saídas desejadas encontram-se dentro dos intervalos de confiança obtidos com o modelo NFHB-T2, conforme mostram as Figuras 4.18 e 4.19.

#### 4.3.5

##### Previsão de Séries Temporais em Sistemas Dinâmicos: *Santa Fe*

Esta seção apresenta uma aplicação de previsão de séries temporais em sistemas dinâmicos. A série considerada é a série *Santa Fe* [WEIG94].

O conjunto possui ao todo 340 padrões, sendo dividido em 273 padrões para o conjunto de treinamento, 46 para validação e 21 para teste.

Similar aos casos anteriores, para este série considerou-se janelas de tamanho igual a 4 (referentes aos valores passados). O tamanho da janela foi determinado através da análise de autocorrelação efetuada na série. O método de seleção da seqüência de variáveis de entrada obteve como resultados a seqüência:

$$x_{t-1}, x_t, x_{t-2}, x_{t-3}.$$

A Tabela 4.15 apresenta os resultados de previsão um passo à frente. Conforme a Tabela 4.15, o melhor desempenho de generalização foi obtido pelo modelo NFHB-T2 com taxa de decomposição de 0.05, particionamento adaptativo e opção 1 para o particionamento das células filhas com a mesma variável de entrada. Além disso, o melhor desempenho foi obtido com o modelo NFHB-T2 de 13 níveis de hierarquia com variáveis de entrada  $x_{t-1}, x_t, x_{t-2}, x_{t-3}, x_{t-1}, x_t, x_{t-2}, x_{t-3}, x_{t-1}, x_t, x_{t-2}, x_{t-3}, x_{t-1}$  para cada nível respectivamente.

TABELA 4.15: Resultados obtidos com o modelo NFHB-T2 para previsão da base de dados *Santa Fe*.

Opção Particion. Células Filhas	Taxa Decomp. 0.001-0.05	Ajuste Anteced.	Num. Conseq.	MAPE Treino	MAPE Validaç.	MAPE Teste	U-Theil Teste
1	0.05	não	26	24.37	21.59	4.80	0.277
2	0.05	não	26	24.38	21.58	4.80	0.277
1	0.05	sim	133	20.78	20.13	4.67	0.298
2	0.05	sim	127	19.36	18.89	6.15	0.345
1	0.01	não	105	22.89	21.30	5.10	0.291
2	0.01	não	105	22.93	21.47	5.13	0.294
1	0.01	sim	287	24.62	22.37	8.95	0.457
2	0.01	sim	146	35.76	23.06	10.93	0.671

A Figura 4.20 mostra as saídas do modelo NFHB-T2 para o conjunto de validação e teste da base de dados *Santa Fe*.

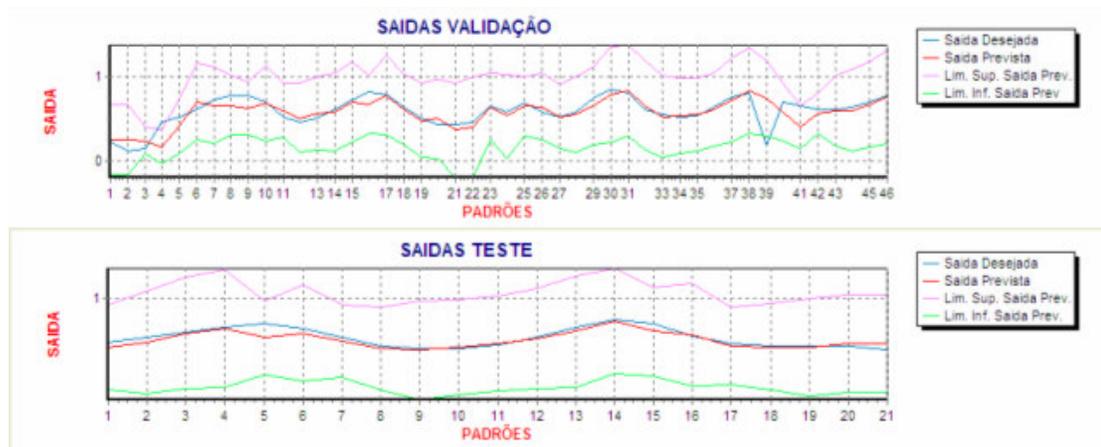


Figura 4.20 - Saídas do modelo NFHB-T2 para o conjunto de validação e teste da base de dados *Santa Fe*.

Da Figura 4.20 observa-se que o modelo NFHB-T2 conseguiu reproduzir bem a série. Observa-se também que a maioria das saídas desejadas encontra-se dentro dos intervalos de confiança obtidos para as saídas defuzzificadas do modelo NFHB-T2. Desta forma, assim como nos casos anteriores, estes intervalos de confiança demonstram coerência.

### 4.3.6

#### Previsão de Séries Temporais de Carga Elétrica: *Eletropaulo*

Esta seção apresenta uma aplicação de previsão de séries temporais de carga elétrica a partir de dados históricos diários. Para esta experiência, foram utilizados dados relativos aos valores diários de carga elétrica da concessionária *Eletropaulo*. Os dados obtidos correspondem ao período compreendido entre maio de 2002 e abril de 2005.

O conjunto possui ao todo 1089 padrões. Para a realização dos testes, o conjunto inicial de 1089 padrões foi dividido em 762 padrões para o conjunto de treinamento, 218 para validação e 109 para teste.

Conforme visto nos casos anteriores, para a previsão desta série empregaram-se janelas de tamanho igual a 7 (referentes aos valores diários passados). O tamanho da janela foi determinado através da análise de autocorrelação efetuada na série. O método baseado em mini-sistemas ANFIS selecionou a seguinte seqüência de variáveis de entrada:  $x_{t-6}, x_{t-4}, x_{t-3}, x_t, x_{t-5}, x_{t-2}, x_{t-1}$ .

Os resultados da previsão um passo à frente estão mostrados na Tabela 4.16.

TABELA 4.16: Resultados obtidos com o modelo NFHB-T2 para previsão da base de dados *Eletropaulo*.

Opção Particion. Células Filhas	Taxa Decomp. 0.001-0.05	Ajuste Anteced.	Num. Conseq.	MAPE Treino	MAPE Validaç.	MAPE Teste	U-Theil Teste
1	0.05	não	23	3.416	4.008	4.457	0.421
2	0.05	não	23	3.362	3.955	4.382	0.419
1	0.05	sim	83	2.689	3.244	3.435	0.349
2	0.05	sim	212	2.104	3.109	3.170	0.358
1	0.01	não	99	3.246	3.803	4.172	0.403
2	0.01	não	99	3.210	3.771	4.122	0.400
1	0.01	sim	346	3.770	4.012	4.575	0.432
2	0.01	sim	240	4.256	4.700	5.191	0.471

Conforme a Tabela 4.16, o modelo NFHB-T2 que apresentou melhor desempenho foi aquele com taxa de decomposição de 0.05, particionamento adaptativo e opção 2 para o particionamento das células filhas com a mesma variável de entrada. Neste caso, o melhor desempenho foi obtido com o modelo

NFHB-T2 de 13 níveis de hierarquia com variáveis de entrada  $x_{t-6}, x_{t-4}, x_{t-3}, x_t, x_{t-5}, x_{t-2}, x_{t-1}, x_{t-6}, x_{t-4}, x_{t-3}, x_t, x_{t-5}, x_{t-2}$  para cada nível, respectivamente.

As saídas do modelo NFHB-T2 para o conjunto de validação e teste da base de dados *Eletropaulo* são mostradas na Figura 4.21.

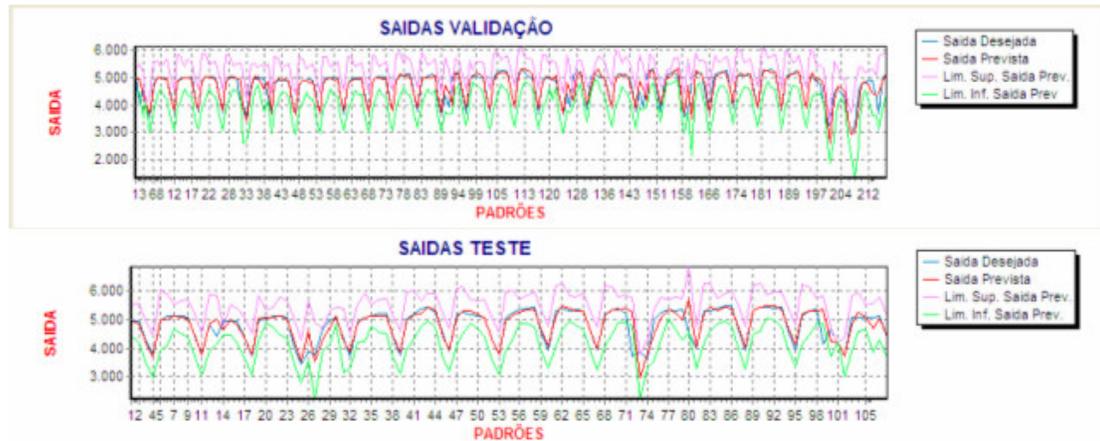


Figura 4.21 - Saídas do modelo NFHB-T2 para o conjunto de validação e teste da base de dados *Eletropaulo*.

Conforme a figura acima, novamente o modelo NFHB-T2 conseguiu reproduzir a série muito bem. Como esperado, a maioria das saídas desejadas encontram-se dentro dos intervalos de confiança obtidos pelo modelo NFHB-T2.

#### 4.4

#### Comparação do Modelo NFHB-T2 com outros modelos.

O desempenho do modelo NFHB-T2 na previsão dos conjuntos de dados *Mackey-Glass*, *Negócios*, *MGB*, *Santa Fe*, *Eletropaulo*, *Retornos* e na aproximação do conjunto de dados *Médicos*, expostos nas seções anteriores, foram comparados com o desempenho do modelo Neuro-Fuzzy Hierárquicos BSP Takagi-Sugeno que faz uso de conjuntos fuzzy tipo 1 [SOUZ99] e com Redes Neurais “Multi Layer Perceptron”. Foram utilizadas topologias de redes neurais com uma camada intermediária, treinadas com algoritmo de aprendizado supervisionado Levenberg-Marquardt [HAGA94], [HAGA96] – uma variação do Back-propagation de modo a obter maior velocidade de convergência, em modo batch de treinamento.

Para o caso das Redes Neurais MLP, cada topologia é definida pelo número de neurônios na camada escondida. Foram avaliadas diversas topologias, com este número variando entre 1 neurônio e 2 vezes o número de entradas + 5 neurônios. Ou seja, para redes com 10 entradas, foram avaliadas topologias desde 1 neurônio na camada escondida até 25 neurônios. Além disso, é importante ressaltar que o procedimento de treinamento usando o método de validação cruzada é repetido 10 vezes para cada topologia. Isso permite escolher, entre as 10 melhores redes, aquela que apresenta o melhor desempenho. É importante lembrar que, no caso dos modelos NFHB-T2 e NFHB com conjuntos fuzzy do tipo 1, não foi possível realizar o treinamento com este método de validação cruzada, em virtude da inicialização determinística dos seus parâmetros (vide seção 3.5).

Para fins de comparação, o modelo Neuro-Fuzzy Hierárquicos BSP com conjuntos fuzzy tipo 1 e as Redes Neurais MLP foram testados com os mesmos conjuntos de treinamento, validação e teste empregados nos experimentos descritos nas seções 4.2 e 4.3. Desta forma, empregaram-se as mesmas variáveis de entrada e, no caso de previsão, foi utilizado o mesmo tamanho de janela empregado na aproximação e previsão das bases de dados usadas para testar o modelo NFHB-T2.

As Tabelas 4.17 e 4.18 apresentam, de forma resumida, os resultados em termos de percentual de acerto médio nos conjuntos de teste obtidos na aproximação e na previsão um passo à frente das bases de dados respectivas. No caso das Redes Neurais MLP, são apresentados os valores mínimos deste percentual de acerto médio, as médias e os desvios padrão respectivos resultantes do processo de treinamento com o método de validação cruzada.

Inspecionando-se a Tabela 4.17 é possível observar que os resultados obtidos para o modelo NFHB-T2 no caso da aproximação de funções do banco de dados *Médicos* foram superiores aos melhores resultados encontrados pelos outros modelos. Uma possível justificativa é o fato que a base de dados *Médicos* possui dados reais, portanto com muita incerteza. É importante ressaltar que, em equipamentos de desempenho semelhante, o tempo de treinamento das RN MLP foi de cerca de quatro horas, em contraposição a um tempo de dezenas a centenas de segundos requerido pelos modelos NFHB-T2 e NFHB com conjuntos fuzzy do tipo 1 para realizar a mesma tarefa. Além disso, o modelo NFHB com conjuntos fuzzy do tipo 1 e as RN MLP têm limitações na modelagem de incertezas

presentes e não conseguem fornecer um intervalo de confiança para as suas saídas; estes modelos proporcionam unicamente saídas pontuais. O modelo NFHB-T2 tem a vantagem de modelar estas incertezas e de fornecer este intervalo de confiança, o que se constitui em uma informação importante para aplicações reais.

Tabela 4.17 Comparação do desempenho entre o Modelo NFHB-T2 e outros modelos para o caso de aproximação e previsão um passo à frente.

BASE DE DADOS	MODELO							
	NFHB-T2		NFHB Tipo 1		RN MLP			Desvio. Padrão. MAPE Teste
	MAPE Teste	U-Theil Teste	MAPE Teste	U-Theil Teste	MAPE Teste Min.	U-Theil Teste Min.	Média MAPE Teste	
APROXIMAÇÃO								
Médicos	10.41	0.460	11.87	0.528	9.99	0.435	11.85	0.951
PREVISÃO								
Mackey % SNR 6.2	1.99	0.111	1.90	0.112	1.87	0.110	1.91	0.018
Mackey % SNR 18.6	5.25	0.283	5.34	0.285	5.19	0.278	5.28	0.070
Negócios Série 1	1.77	0.072	1.62	0.068	0.67	0.027	0.90	0.033
Negócios Série 2	7.11	0.251	7.92	0.288	8.34	0.301	8.53	3.956
Negócios Série 3	3.38	0.159	4.66	0.169	4.41	0.194	6.51	2.495
Negócios Série 4	2.16	0.090	1.80	0.071	1.81	0.077	2.90	0.795
Negócios Série 5	12.32	0.375	15.24	0.481	15.50	0.491	20.94	4.791
Negócios Série 6	12.26	0.394	18.34	0.580	16.65	0.537	25.43	7.201
MGB Série 1	6.90	0.511	6.16	0.523	7.52	0.619	10.51	2.914
MGB Série 2	12.56	0.558	11.97	0.564	12.76	0.551	13.67	1.414
MGB Série 3	12.12	0.965	13.09	1.039	13.53	1.010	14.99	1.725
Santa Fe	4.67	0.298	5.04	0.313	5.12	0.275	5.91	0.874
Eletropaulo	3.17	0.358	4.49	0.372	3.26	0.307	4.14	0.280

Observa-se também na Tabela 4.17 que os resultados obtidos para o modelo NFHB-T2 no caso da previsão da série Mackey-Glass com ruído de SNR igual a 6.2% e a 18.6% foram muito próximos aos melhores resultados encontrados pelo modelo NFHB com conjuntos fuzzy do tipo 1 e pelas RN MLP. Neste caso, os melhores desempenhos nas previsões das séries Mackey-Glass com ruídos de SNR iguais a 6.2% e a 18.6% foram obtidos com o modelo NFHB com conjuntos fuzzy do tipo 1 e com o modelo NFHB-T2, respectivamente. Observe-

se que o tempo de treinamento das RN MLP foi de cerca de uma hora, muito superior aos poucos minutos requeridos pelos modelos NFHB-T2 e NFHB com conjuntos fuzzy do tipo 1. Há que considerar ainda as limitações já mencionadas do modelo NFHB com conjuntos fuzzy do tipo 1 e das RN MLP.

Tabela 4.18 Comparação do desempenho entre o Modelo NFHB-T2 e outros modelos para o caso de previsão um passo à frente da base de dados de *Retornos*.

BASE DE DADOS	MODELO																	
	NFHB-T2				NFHB Tipo 1				RN MLP									
	MAD Teste	NRMSE Teste	MSE Teste	U-theil Teste	MAD Teste	NRMSE Teste	MSE Teste	U-theil Teste	MAD Teste Min.	Média MAD Teste	Desvio. Padrão MAD Teste	NRMSE Teste Min.	Média NRMSE Teste	Desvio. Padrão NRMSE Teste	MSE Teste Min.	Média MSE Teste	Desvio. Padrão MSE Teste	U-theil Teste
Brahma	0.107	1.272	0.023	0.896	0.081	0.984	0.013	0.696	0.082	0.127	0.027	0.899	1.329	0.213	0.239	0.534	0.179	0.635
Bernge	0.203	1.207	0.136	0.871	0.199	0.999	0.090	0.722	0.203	0.231	0.024	0.948	1.068	0.094	1.862	2.382	0.438	0.685

No caso da previsão das séries de negócios 2, 3, 5 e 6, a Tabela 4.17 mostra que o modelo NFHB-T2 apresentou desempenho superior ao do modelo NFHB com conjuntos fuzzy do tipo 1 e ao das RN MLP. Observa-se também que no caso da previsão das séries de negócios 1 e 4, o modelo NFHB-T2 obteve resultados um pouco inferiores aos melhores resultados proporcionados pelo modelo NFHB com conjuntos fuzzy do tipo 1 e pela RN MLP. No caso da previsão da série de negócios 4, o melhor desempenho foi obtido com o modelo NFHB com conjuntos fuzzy do tipo 1; já no caso da previsão da série de negócios 1, o melhor desempenho foi o das RN MLP. Novamente, o tempo de treinamento das RN MLP foi muito maior quando comparado com o tempo de treinamento do modelo NFHB-T2 e ao do modelo NFHB com conjuntos fuzzy do tipo 1. Similarmente ao caso anterior, o modelo NFHB com conjuntos fuzzy do tipo 1 e as RN MLP têm limitações na modelagem de incertezas e não oferecem um intervalo de confiança para as suas saídas. Os resultados de previsão encontrados para este tipo de séries sugerem que o modelo NFHB-T2 apresenta um melhor desempenho na maioria dos casos das séries de negócios testadas. Isto se deve, possivelmente, ao fato de que tais séries possuem maior complexidade e muita incerteza.

A Tabela 4.17 mostra que os resultados obtidos para o modelo NFHB-T2 no caso da previsão das séries que seguem um MGB 1 e 2 foram um pouco inferiores quando comparados aos melhores resultados encontrados pelo modelo

NFHB com conjuntos fuzzy do tipo 1, e superiores quando comparados com os melhores resultados encontrados com as RN MLP. Nestes casos, os melhores desempenhos foram obtidos com o modelo NFHB com conjuntos fuzzy do tipo 1. Entretanto é importante destacar as limitações do modelo NFHB com conjuntos fuzzy do tipo 1 já mencionadas. Observa-se também que, no caso da previsão da série 3, o modelo NFHB-T2 apresentou desempenho superior quando se comparam os melhores resultados dos três modelos. O fato da série 3 possuir valores maiores de tendência e volatilidade faz com que seja mais complexa, de forma que o modelo NFHB-T2 é o mais adequado para tratá-la.

A partir da Tabela 4.17 é possível observar também que os resultados obtidos para o modelo NFHB-T2 no caso da previsão um passo à frente da série em sistemas dinâmicos *Santa Fe* e da série de carga elétrica *Eletropaulo* foram superiores em comparação com os melhores resultados encontrados pelos outros dois modelos. Uma possível justificativa é o fato que estas séries possuem dados reais, portanto com muita incerteza.

A Tabela 4.18 mostra que os resultados obtidos para o modelo NFHB-T2 no caso da previsão um passo à frente das séries de retornos foram inferiores aos melhores resultados encontrados pelo modelo NFHB com conjuntos fuzzy do tipo 1, e superiores aos melhores resultados encontrados pelas RN MLP. Neste caso, os melhores desempenhos foram obtidos com o modelo NFHB com conjuntos fuzzy do tipo 1. É importante ressaltar que as saídas deste foram muito próximas às suas médias, conforme mostrado nas Figuras 4.22 e 4.23. Isto sugere que este modelo poderia estar aprendendo a média destas séries e não efetuando uma previsão. O modelo NFHB-T2 consegue acompanhar as variações no tempo destas séries, apesar de apresentar maior erro na previsão. Isto sugere que o modelo NFHB-T2 consegue evitar a tendência de aprender a média – comum na maioria dos modelos quando prevêm séries de retornos. As Figuras 4.22 e 4.23 apresentam as saídas previstas pelo modelo NFHB-T2 e pelo modelo NFHB com conjuntos fuzzy do tipo 1, as saídas desejadas e as médias destas calculadas para o conjunto de teste da base de dados Brahma ON e Bemge ON, respectivamente. Para o cálculo destas médias consideraram-se inicialmente as saídas desejadas pertencentes ao conjunto de teste inteiro (100%); este conjunto de teste foi depois dividido em 50% e finalmente em 25%, calculando-se então a média de cada parte.

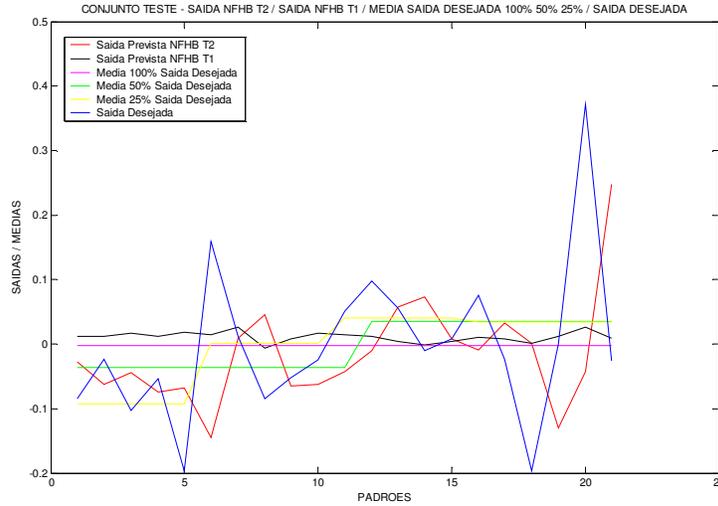


Figura 4.22 – Saídas do modelo NFHB-T2, do modelo NFHB com conjuntos fuzzy do tipo 1, saída desejada e média da saída desejada para o conjunto teste da base de dados Brahma ON.

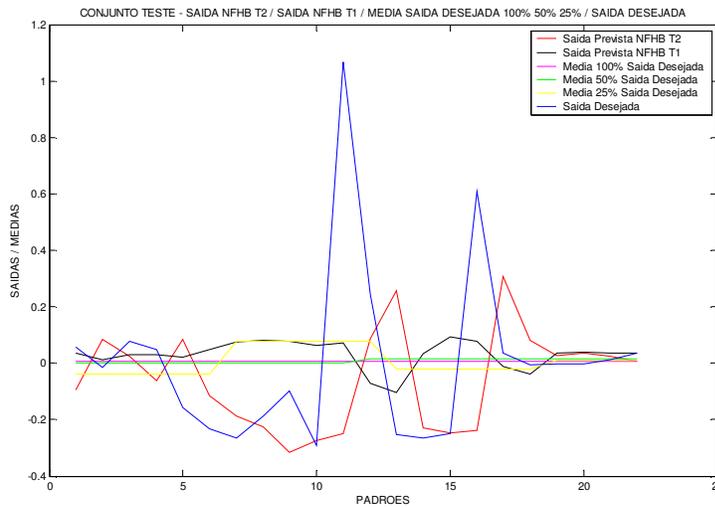


Figura 4.23 – Saídas do modelo NFHB-T2, do modelo NFHB com conjuntos fuzzy do tipo 1, saída desejada e média da saída desejada para o conjunto de teste da base de dados Bemge ON.

As Tabelas 4.19 e 4.20 apresentam, de forma resumida, os resultados em termos de percentual de acerto médio nos conjuntos de teste obtidos na previsão dez passos à frente das bases de dados *Santa Fe* e *Eletropaulo* e na previsão seis passos à frente das bases de dados de *Retornos*.

Tabela 4.19 Comparação do desempenho entre o Modelo NFHB-T2 e outros modelos para o caso de previsão dez passos à frente da base de dados de *Santa Fe* e *Eletropaulo*.

PASSOS À FRENTE					1	2	3	4	5	6	7	8	9	10	
CONFIGURAÇÃO					U-Theil	MAPE	MAPE								
		Taxa Descomp. 0.001-0.05	Ajuste Antec.	Nº Conseq.											
<b>NFHB do Tipo 1</b>	Santa Fe	0.05	sim	83	0.31	5.04	9.61	11.35	10.40	9.73	11.15	11.52	11.40	11.82	11.97
<b>NFHB-T2</b>	Santa Fe	0.05	sim	133	0.30	4.67	8.75	11.58	11.63	12.29	13.37	16.81	15.43	17.76	17.90
<b>NFHB do Tipo 1</b>	Eletropaulo	0.05	sim	118	0.37	4.49	6.44	7.09	7.35	7.34	7.22	7.50	7.22	7.30	7.23
<b>NFHB-T2</b>	Eletropaulo	0.05	sim	212	0.36	3.17	3.43	3.58	3.56	3.53	3.47	3.35	3.46	3.63	3.68

Tabela 4.20 Comparação do desempenho entre o Modelo NFHB-T2 e outros modelos para o caso de previsão seis passos à frente da base de dados *Retornos*.

BASE DE DADOS RETORNOS																							
PASSOS À FRENTE					1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6	
MODELO	SÉRIE	CONFIGURAÇÃO			U-Theil	MAD						NRMSE						MSE					
		Taxa Decomp. 0.001-0.05	Ajuste Antec.	Nº Conseq.																			
<b>NFHB Tipo 1</b>	Brahma	0.05	sim	2	0.696	0.081	0.083	0.082	0.085	0.084	0.085	0.984	0.983	0.999	0.998	0.997	0.995	0.013	0.013	0.014	0.015	0.015	0.015
<b>NFHB-T2</b>	Brahma	0.01	não	98	0.896	0.107	0.142	0.137	0.131	0.141	0.142	1.272	1.131	0.832	0.806	0.974	0.961	0.023	0.035	0.032	0.029	0.033	0.035
<b>NFHB Tipo 1</b>	Bemge	0.01	sim	9	0.722	0.199	0.194	0.209	0.215	0.224	0.235	0.999	0.991	1.015	1.004	1.004	1.004	0.090	0.091	0.100	0.103	0.108	0.114
<b>NFHB-T2</b>	Bemge	0.05	sim	71	0.871	0.203	0.249	0.254	0.270	0.281	0.284	1.207	0.887	0.824	0.790	0.723	0.950	0.136	0.183	0.199	0.213	0.221	0.235

A Tabela 4.19 mostra que os resultados obtidos para o modelo NFHB-T2 no caso da previsão de dois até dez passos à frente da série em sistemas dinâmicos *Santa Fe* foram inferiores aos melhores resultados encontrados pelo modelo NFHB com conjuntos fuzzy do tipo 1. Por outro lado, no caso da previsão de dois até dez passos à frente da série de carga elétrica *Eletropaulo*, os resultados obtidos para o modelo NFHB-T2 foram superiores.

A Tabela 4.20 mostra que os resultados obtidos para o modelo NFHB-T2 no caso da previsão de dois até seis passos à frente das séries de *Retornos* foram inferiores aos melhores resultados encontrados pelo modelo NFHB com conjuntos fuzzy do tipo 1.

A seguir, alguns dos conjuntos fuzzy *intervalares* do tipo 1 de saída do modelo NFHB-T2, considerados como intervalo de confiança, obtidos na previsão do conjunto de dados *MGB* e na aproximação do conjunto de dados *Médicos* expostos nas seções anteriores, foram comparados com os intervalos de confiança obtidos estatisticamente com as saídas previstas dos modelos NFHB-T2 e NFHB.

Os intervalos de confiança de 95% obtidos estatisticamente, considerando-se normalidade na distribuição dos dados, foram calculados de acordo com [MEDE99], [SPIE93]:

$$IC = x_{t+1} \pm 2\sigma_{\epsilon} \quad \text{Equação 4.7}$$

Onde:

$IC$  = Intervalo de confiança de 95%.

$x_{t+1}$  = Valor de previsão do modelo.

$\sigma_{\epsilon}$  = Desvio padrão do erro de previsão do modelo.

As Figuras 4.24, 4.25, 4.26, 4.27, 4.28, 4.29, 4.30, 4.31 e 4.32 mostram os intervalos de confiança obtidos automaticamente pelo modelo NFHB-T2 através do seu conjunto fuzzy *intervalar* do tipo 1 de saída, e, os intervalos de confiança obtidos estatisticamente com as saídas previstas dos modelos NFHB-T2 e NFHB para estas bases de dados. As Figuras 4.24, 4.27 e 4.30 correspondem às Figuras 4.2, 4.14, e 4.15 e foram repetidas para fins comparativos.

- Aproximação de funções: Velocidade da Onda do Pulso de Sangue (VOP)

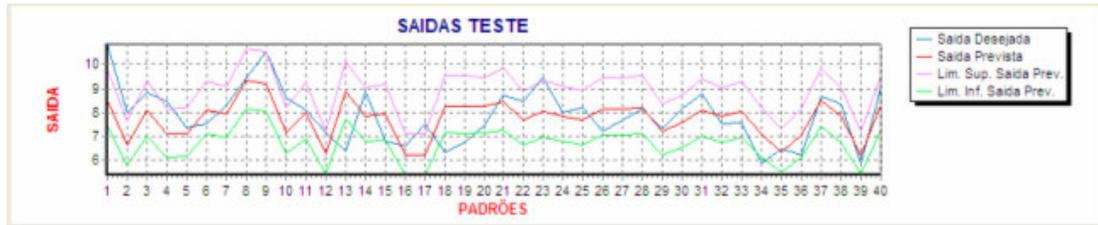


Figura 4.24 – Intervalo de confiança obtido automaticamente pelo modelo NFHB-T2 para o conjunto de teste da base de dados Médicos.

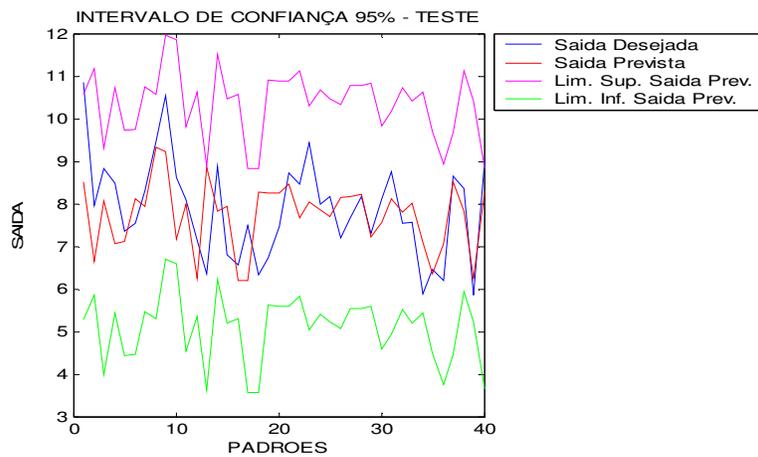


Figura 4.25 - Intervalo de confiança obtido estatisticamente com a saída prevista pelo modelo NFHB-T2 para o conjunto de teste da base de dados Médicos.

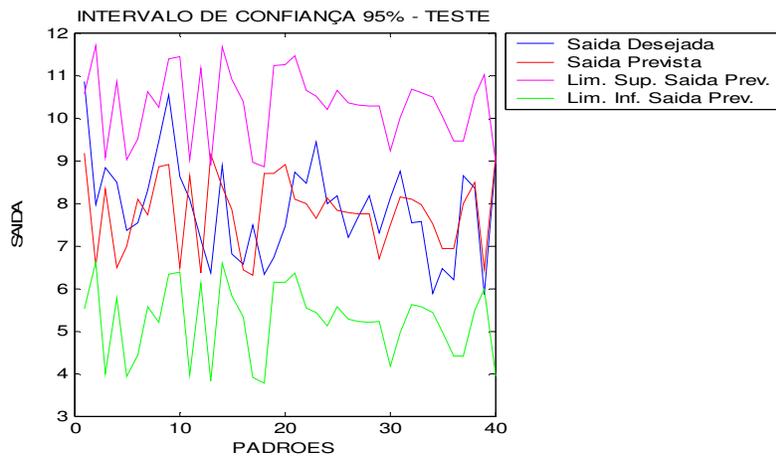


Figura 4.26 - Intervalo de confiança obtido estatisticamente com a saída prevista pelo modelo NFHB que usa conjuntos fuzzy do tipo 1 para o conjunto de teste da base de dados Médicos.

- Previsão de séries temporais: Série que segue um MGB – Série 1



Figura 4.27 – Intervalo de confiança obtido automaticamente pelo modelo NFHB-T2 para o conjunto de teste da base de dados MGB – Série 1.

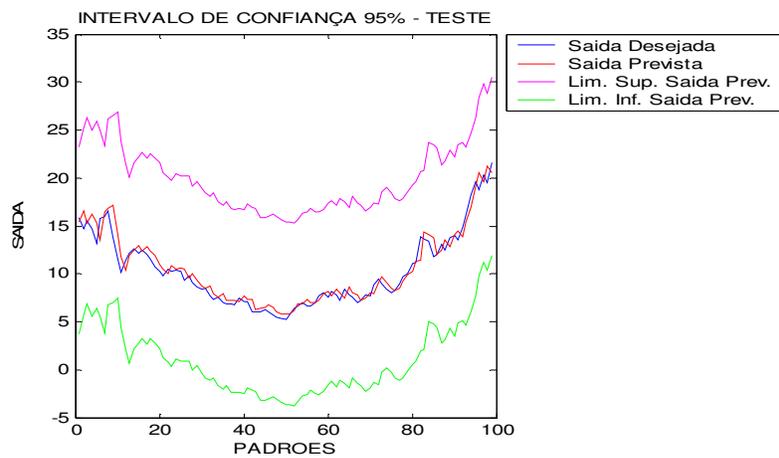


Figura 4.28 - Intervalo de confiança obtido estatisticamente com a saída prevista pelo modelo NFHB-T2 para o conjunto de teste da base de dados MGB – Série 1.

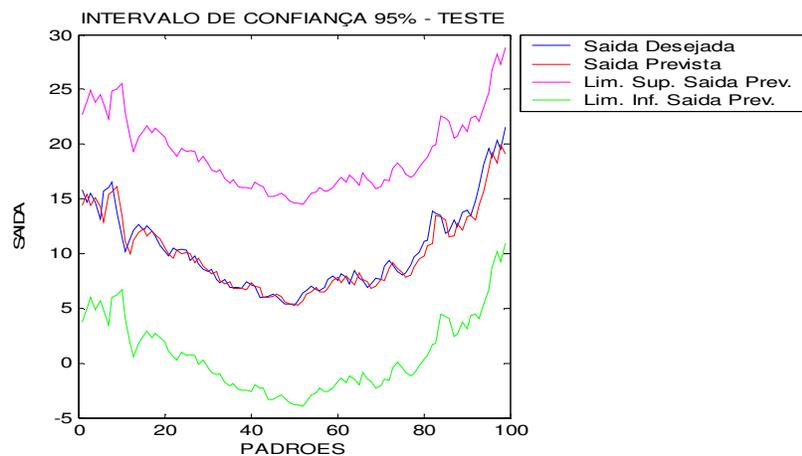


Figura 4.29 - Intervalo de confiança obtido estatisticamente com a saída prevista pelo modelo NFHB que usa conjuntos fuzzy do tipo 1 para o conjunto de teste da base de dados MGB – Série 1.

- Previsão de séries temporais: Série que segue um MGB – Série 2

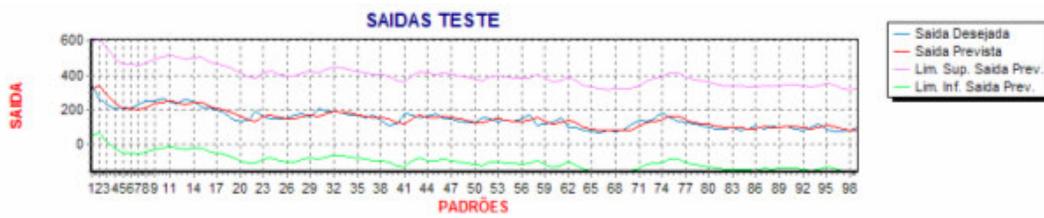


Figura 4.30 – Intervalo de confiança obtido automaticamente pelo modelo NFHB-T2 para o conjunto de teste da base de dados MGB – Série 2.

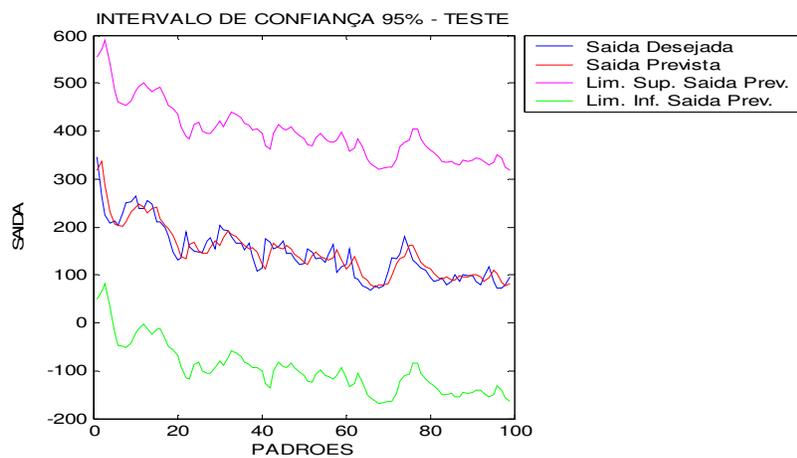


Figura 4.31 - Intervalo de confiança obtido estatisticamente com a saída prevista pelo modelo NFHB-T2 para o conjunto de teste da base de dados MGB – Série 2.

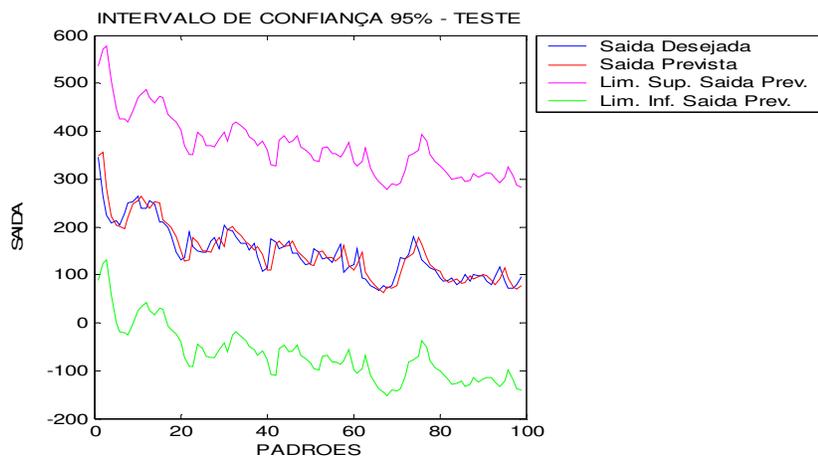


Figura 4.32 - Intervalo de confiança obtido estatisticamente com a saída prevista pelo modelo NFHB que usa conjuntos fuzzy do tipo 1 para o conjunto de teste da base de dados MGB – Série 2.

A partir das Figuras 4.24, 4.25, 4.26, 4.27, 4.28, 4.29, 4.30, 4.31 e 4.32 pode-se observar que os intervalos de confiança obtidos de forma automática pelo modelo NFHB-T2 são coerentes quando comparados com intervalos de confiança obtidos por métodos tradicionais. Observa-se também que estes intervalos oferecem maior credibilidade na maioria de casos, uma vez que são mais estreitos que os intervalos de confiança tradicionais.

## 4.5

### Conclusão

O modelo NFHB-T2 mostrou boa aplicabilidade na tarefa de previsão e aproximação de funções em casos onde a complexidade e a incerteza estão presentes. Este modelo apresentou resultados semelhantes e em vários casos superiores ao de modelos como o NFHB com conjuntos fuzzy do tipo 1 e as RN MLP. Com o NFHB-T2, o tempo computacional foi semelhante ao requerido pelo modelo NFHB com conjuntos fuzzy do tipo 1 e muito menor do que o requerido pelas RN MLP para realizar a mesma tarefa. O modelo NFHB-T2 tem a vantagem de modelar incertezas de forma totalmente nova, sendo capaz de fornecer um intervalo de confiança – importante em aplicações reais – para a suas saídas, contrariamente ao modelo NFHB e às RN MLP.

No capítulo a seguir apresentam-se as conclusões finais e as sugestões para trabalhos futuros.