

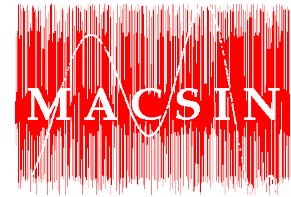
Laboratório de Modelagem, Análise e Controle de Sistemas Não-Lineares

Departamento de Engenharia Eletrônica

Universidade Federal de Minas Gerais

Av. Antônio Carlos 6627, 31270-901 Belo Horizonte, MG Brasil

Fone: +55 31 3409-3470



Uso de Informação Auxiliar em Redes Neurais e Formação de Comitês na Identificação de Sistemas Dinâmicos

Leandro Freitas de Abreu

Orientadores: Prof. Dr. Luis Antônio Aguirre
Prof. Dr. Bruno Henrique Groenner Barbosa

Belo Horizonte, 30 de março de 2013

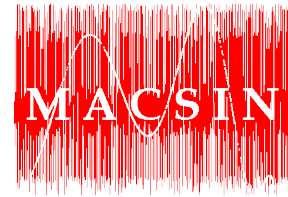
Laboratório de Modelagem, Análise e Controle de Sistemas Não-Lineares

Departamento de Engenharia Eletrônica

Universidade Federal de Minas Gerais

Av. Antônio Carlos 6627, 31270-901 Belo Horizonte, MG Brasil

Fone: +55 31 3409-3470



Uso de Informação Auxiliar em Redes Neurais e Formação de Comitês na Identificação de Sistemas Dinâmicos

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal de Minas Gerais como requisito parcial para a obtenção do grau de Mestre em Engenharia Elétrica

Leandro Freitas de Abreu

Orientadores: Prof. Dr. Luis Antônio Aguirre
Prof. Dr. Bruno Henrique Groenner Barbosa

UNIVERSIDADE FEDERAL DE MINAS GERAIS
ESCOLA DE ENGENHARIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

Belo Horizonte
30 de março de 2013

À minha amada esposa Bárbara Mourão Aranha Freitas.

Agradecimentos

Agradeço a Deus pelas oportunidades, pela infindável fonte de força e, principalmente, pelas pessoas em meu caminho.

Em especial, à minha amada esposa Bárbara, a quem dedico esta dissertação. Agradeço todo o apoio nas decisões que permitiram o desenvolvimento deste trabalho e a constante motivação nos desafios diários. O meu agradecimento e amor eterno.

Agradeço ao meu pai Antônio que, apesar de não estar mais presente no plano físico, há algum tempo plantou uma semente de curiosidade e força de vontade, que cresce a cada dia.

Agradeço ao professor Luis Aguirre pela confiança e incentivo já nos primeiros momentos do curso. Agradeço aos orientadores, professor Luis Aguirre e professor Bruno Barbosa, os instigantes comentários e discussões, que possibilitaram o avanço deste trabalho e me permitiram uma melhor visão do estimulante mundo da ciência.

Agradeço aos membros do Projeto Petrobras: ao engenheiro Alex Teixeira que conduziu com maestria o projeto junto à Petrobras; ao professor Luis Aguirre, coordenador do projeto pela UFMG; aos professores Bruno Barbosa, Eduardo Mendes, Bruno Teixeira e Giovani Rodrigues pelas interessantes discussões e apoio no projeto.

Agradeço aos colegas do MACSIN, em especial, ao Alípio, ao Dimas, ao Víctor, ao Tales e ao Luciano pelas desvairadas discussões, que estimulam ainda mais a curiosidade e a vontade de ir além.

Agradeço à UFMG como instituição promotora do conhecimento. Aos professores do DELT e ao PPGEE pela estrutura e suporte nas atividades acadêmicas. Ao CPH e toda sua equipe, pelo auxílio e assistência prestados no decorrer deste trabalho.

Agradeço todo apoio financeiro proveniente da Petrobras e Fundação Christiano Ottoni. Agradeço também ao CNPq, FAPEMIG e CAPES pela contribuição indireta no trabalho.

*“Quanto a vós, nunca vos deixeis chamar de Mestre,
pois um só é vosso Mestre e todos vós sois irmãos.”
(Mateus 23, 8)*

Resumo

Este trabalho dedica-se ao estudo de duas estratégias de identificação de sistemas dinâmicos não lineares utilizadas, principalmente, nos casos em que os dados disponíveis não são suficientemente representativos. É apresentado um procedimento de uso de informação auxiliar no treinamento de redes neurais e um estudo sobre a formação de comitês na identificação de sistemas dinâmicos. Ambas as abordagens são contextualizadas no cenário de identificação de sistemas e aplicadas em sistemas experimentais.

A primeira abordagem, sobre identificação caixa cinza, utiliza a curva estática do sistema como informação auxiliar, que é adicionada na fase de ajuste dos parâmetros (treinamento) da rede neural. O treinamento é feito por meio de otimização multiobjetivo, que minimiza o erro de simulação livre sobre dados dinâmicos e o erro sobre a curva estática. O procedimento é aplicado em dois processos experimentais: um sistema piloto de bombeamento de água e um processo industrial de extração de petróleo em poço não surgente em alto mar. A abordagem caixa cinza proposta foi comparada à identificação caixa preta, especialmente em regiões de operação não presentes nos dados de identificação. Os resultados experimentais mostram que a abordagem caixa cinza sempre apresenta melhor desempenho em, pelo menos, um dos critérios: erro em relação à curva estática ou erro nos dados dinâmicos de validação, cuja faixa de operação é mais abrangente. É mostrado que a melhoria prometida pela metodologia proposta está fortemente relacionada com o desempenho estático do modelo obtido pela abordagem caixa preta.

A segunda abordagem, sobre a formação de comitês de modelos dinâmicos, é apresentada com foco no conceito de *diversidade*. É feito um estudo sobre as formas de medir diversidade, com foco nas técnicas que não fazem restrições quanto à classe ou estrutura de modelos, i.e. que utilizam apenas dados de entrada e saída. Por meio de simulações, é mostrado que esse tipo de métrica de diversidade sofre forte influência dos dados utilizados nas medições. Assim, baseado no conceito, é proposto um procedimento de medição de diversidade, que especifica, inclusive, as características dos dados dinâmicos usados na medição. É proposto ainda um combinador de modelos, do tipo “média ponderada”, baseado nessa medição de diversidade. Resultados experimentais, no sistema de extração de petróleo em águas profundas, mostram que o procedimento pode reduzir o valor e a variância do erro de estimação, em comparação com o combinador do tipo “média simples”, especialmente nos casos em que a diversidade dos modelos não é garantida.

Abstract

This work is dedicated to study two strategies for identification of nonlinear dynamical systems, specially, when available data are not representative. A procedure to incorporate auxiliary information during the training of a neural network and a study about combining dynamic models are presented. The background of both approaches and experimental applications are shown.

The first approach (gray-box identification) uses static curve as auxiliary information, which is added in the neural network parameters adjustment stage (training). A multi-objective approach is adopted to minimize both free-simulation error on dynamic data and static curve error. The proposed technique was applied to two experimental processes: a pilot hydraulic pumping system and an industrial gas-lift offshore oil well. The proposed gray-box identification technique was compared to black-box approaches, particularly, in operating regimes that were not available in the dynamical identification data sets. Results shows that the gray-box procedure yields models with better performance than the ones obtained by the black-box approach in, at least, one of the objectives: static function error or dynamic test data prediction error, where the dynamic data cover a broader operating range. It is shown that the implementation of this gray-box approach is justifiable when the black-box procedure does not achieve a model with good static performance.

The second approach, on combining dynamic models, is presented with focus on *diversity* concept. Techniques to measure diversity are described, specially those without restrictions about model class or structure, i.e. which uses only input and output data sets. Simulated problems shows strong influence of data set on this kind of diversity metrics. Based on the concept, a procedure to measure diversity is proposed, which specifies appropriate data properties for the measurement. Using this measure, a weighted average model combiner is proposed and applied on an industrial gas-lift offshore oil well. Results shows that combiner can reduce the value and variance of estimation error, in comparison with simple averaging, specially when model's diversity are not assured.

Sumário

1	Introdução	1
2	Extração de Petróleo em Águas Profundas	3
2.1	O Processo de Produção	3
2.2	Instrumentação	6
2.3	Conclusões do Capítulo	8
3	Informação Auxiliar na Identificação de Sistemas	11
3.1	Introdução	11
3.1.1	Contextualização	11
3.1.2	Metodologia Proposta	13
3.2	Procedimento de Inserção de Informação Auxiliar em Redes Neurais	13
3.2.1	Estruturas dos modelos	13
3.2.2	Ajuste dos pesos	13
3.2.3	Uso de informação auxiliar	14
3.3	Resultados Experimentais	16
3.3.1	O Sistema Hidráulico	16
3.3.2	O Sistema de Extração de Petróleo	25
3.3.3	Discussão dos resultados experimentais	30
3.4	Conclusões do Capítulo	31
4	Diversidade em Comitês de Modelos Dinâmicos	33
4.1	Introdução	33
4.1.1	Contextualização	33
4.1.2	Metodologia Proposta	34
4.2	Sobre Diversidade	34
4.2.1	Métricas de Diversidade	37
4.3	Diversidade em Identificação de Sistemas	39
4.3.1	Diversidade em Sistemas Dinâmicos Lineares	40
4.3.2	Diversidade em Modelos Não lineares	47
4.3.3	Procedimento de Medição de Diversidade de <i>Ensembles</i> em Problemas de Identificação	55
4.3.4	Aplicação do Procedimento	61
4.4	Combinadores de Modelos	61
4.4.1	Combinador Baseado em Diversidade	62

4.4.2	Aplicação no Sistema de Extração de Petróleo	63
4.5	Conclusões do Capítulo	66
5	Considerações Finais	69
	Bibliografia	77

Lista de Figuras

2.1	Ilustração de uma plataforma de extração de petróleo <i>off-shore</i> conectada a dois poços de extração. Fonte: [Teixeira, 2011].	4
2.2	Diagrama simplificado da extração de petróleo utilizando a técnica de elevação artificial por <i>gas-lift</i> contínuo. Fonte: [Teixeira, 2011].	5
2.3	Coluna de produção convencional utilizando a técnica de elevação artificial por <i>gas-lift</i> . Fonte: [Thomas, 2004].	6
2.4	Esquema representando a completção submarina (molhada) de um poço de petróleo em águas profundas. Fonte: [Villela, 2005].	7
2.5	Esquema de um separador trifásico, encontrado em plataformas petrolíferas, que separa o fluido do poço em óleo, gás e água. Fonte: [Thomas, 2004].	7
2.6	Diagrama simplificado de instrumentação típica da extração de petróleo em águas profundas.	9
3.1	Sistema de bombeamento de água formado por dois conjuntos moto-bomba de 10cv cada. Fonte: [Barbosa, 2006].	16
3.2	Conjunto turbina-gerador, acoplado ao sistema de bombeamento de água. Fonte: [Barbosa, 2006].	17
3.3	Dados de entrada (a) e saída (b) obtidos nos experimentos dinâmicos, utilizados na identificação do sistema hidráulico.	18
3.4	Dados de entrada (a) e saída (b) obtidos nos experimentos dinâmicos, utilizados para validação dos modelos.	18
3.5	Curva estática medida do sistema (●) e aproximação polinomial de segundo grau (⋯), conforme Equação 3.9.	19
3.6	Curva Pareto obtida do ajuste multiobjetivo dos parâmetros da rede neural (Eq. 3.10). A cor de preenchimento dos pontos representa o erro (PRMSE) sobre o teste dinâmico em faixa ampla: quanto mais próximo da cor branca, menor o erro. Estão marcados, em azul (○ 4, 839) e vermelho (◦ 5, 218), os modelos caixa cinza com melhor desempenho em cada um dos objetivos. Em verde (◉ 4, 797), é marcado o modelo que apresentou o melhor desempenho no teste dinâmico em faixa ampla. O quadrado (◻ 4, 842) representa o modelo caixa preta. Os valores escritos nos quatro pontos marcados referem-se ao PRMSE sobre os dados dinâmicos em faixa ampla. A região cinza, contornada pela linha tracejada (— —), representa a região em que o erro pode ser considerado nulo, inferior à precisão do instrumento (< 0, 25%).	21
3.7	Curva estática de (◻) \mathcal{M}_p , (◦) \mathcal{M}_{c_1} , (◊) \mathcal{M}_{c_2} e (●) medições.	22

3.8	Trecho do sinal de validação que mostra a simulação livre dos modelos $(--)$ \mathcal{M}_p , $(-)$ \mathcal{M}_{c_1} e $(-\cdot-)$ \mathcal{M}_{c_2} . A linha preta contínua $(—)$ é a medição.	23
3.9	Avaliação do desempenho dinâmico em faixa ampla (validação) e da curva estática para os modelos (\blacksquare) \mathcal{M}_p , (\bullet) \mathcal{M}_{c_1} , (\blacklozenge) \mathcal{M}_{c_2} . Os pontos indicados em cinza (\cdot) mostram as demais soluções da curva Pareto na modelagem caixa cinza.	24
3.10	Desempenho do modelo dinâmico \mathcal{M}_p (caixa preta) sobre os dados de validação do sistema de extração de petróleo. Os dados estimados e medidos são apresentados, respectivamente, em preto e vermelho.	26
3.11	Conjunto Pareto de soluções utilizando a abordagem caixa cinza. É mostrado também o desempenho do (\blacksquare) modelo caixa preta \mathcal{M}_p . É destacado em verde $(\circ 2, 751)$ um dos modelos obtidos no procedimento caixa cinza: \mathcal{M}_c . Os números em destaque mostram o desempenho (RMSE) sobre os dados de validação. A intensidade das cores de preenchimento dos círculos também ilustram o valor do RMSE sobre os dados de validação: quanto mais clara a cor, menor o erro sobre os dados de validação.	27
3.12	Curva estática dos modelos dinâmicos (\square) \mathcal{M}_p , (\circ) \mathcal{M}_c e (\bullet) a curva estática desejada, obtida experimentalmente.	28
3.13	Desempenho do modelo dinâmico \mathcal{M}_c (caixa cinza) sobre os dados de validação do sistema de extração de petróleo. Os dados estimados e medidos são apresentados, respectivamente, em preto e vermelho.	29
3.14	Conjunto Pareto de soluções mostrados no espaço do erro (RMSE) dinâmico de validação e erro (RMSE) sobre a curva estática. É mostrado também o desempenho dos modelos (\blacksquare) \mathcal{M}_p e (\bullet) \mathcal{M}_c	30
4.1	Diagramas de Venn ilustrando o papel da diversidade na identificação de sistemas. É apresentado um caso em que se deseja aproximar F , utilizando apenas as informações contidas em Z e T . É mostrado que um <i>ensemble</i> com alta diversidade (b) pode ajustar-se melhor à função original do que um <i>ensemble</i> com baixa diversidade (a).	36
4.2	Comportamento dos sistemas dinâmicos lineares apresentados na forma de (a) resposta ao degrau e (b) diagrama de Bode para o sistema alvo da modelagem $G(s)$ e seus respectivos modelos $G_1(s)$, $G_2(s)$ e $G_3(s)$	41
4.3	Diagramas de Venn que ilustra a disposição dos (\times) polos e (\circ) zeros dos modelos G_1 , G_2 , G_3 e no sistema alvo da modelagem $G(s)$	43
4.4	Representação das massas de dados Z_1 , Z_2 e Z_3 no âmbito dos diagramas de Venn (apresentados anteriormente na fig. 4.3) e o espectro de frequências desses dados. A área destacada na cor cinza, à esquerda, ilustra qual região do <i>ensemble</i> é excitada por cada conjunto de dados.	44
4.5	Medidas de diversidade para os três tipos de <i>ensembles</i> : E_1 , E_2 , E_3 , utilizando os dados de entrada Z_1 , em diferentes métricas.	45
4.6	Medidas de diversidade para os três tipos de <i>ensembles</i> : E_1 , E_2 , E_3 , utilizando os dados de entrada Z_2 , em diferentes métricas.	46
4.7	Valores de medidas de diversidade para os três tipos de <i>ensembles</i> : E_1 , E_2 , E_3 , utilizando os dados de entrada Z_3 , em diferentes métricas.	46
4.8	Valores da raiz do erro quadrático médio (RMSE) dos <i>ensembles</i> E_1 , E_2 , E_3 , para duas massas de dados Z_1 e Z_3	47

4.9	Curva estática do sistema não linear, alvo da modelagem, apresentado na Equação 4.19.	48
4.10	Resposta do sistema ao degrau unitário em dois pontos de operação distintos: (a) em $u(k) = h(k - 5)$ e (b) $u(k) = 9 + h(k - 5)$, em que $h(k)$ é a função de Heaviside (degrau unitário em $k = 0$).	49
4.11	Módulo da transformada discreta de Fourier dos sinais de entrada de (a) Z_1 e (b) Z_2 , usados no ajuste dos parâmetros dos modelos. Os gráficos mostram uma escala entre as frequências mínima e máxima representáveis: $[-\pi, \pi]$	50
4.12	Trechos dos sinais de entrada e saída dos conjuntos de dados Z_1 e Z_2 usados no treinamento.	51
4.13	Trechos dos sinais de entrada utilizados nas medições de diversidade no (a) (c) (e) domínio do tempo e no (b) (d) (f) domínio da frequência. A escala vertical no domínio do tempo está ajustada para a faixa de operação do sistema.	52
4.14	Saída do sistema S_{nl} e dos modelos M_1 , M_2 , M_3 para uma série de degraus de entrada $U = [0, 0, \dots, 0, 1, 1, \dots, 1, 5, 5, \dots, 5, 6, 6, \dots, 6, 9, 9, \dots, 9, 10, 10, \dots, 10]$. É destacada, no centro da imagem, a resposta dos sistemas na mesma faixa de operação dos dados de treinamento. Nessa imagem destacada, a saída do sistema S_{nl} está sobreposta à saída de M_3	53
4.15	Características sugeridas dos sinais de entrada, no domínio da frequência, para medição de diversidade. Em que Z_1 representa dados de ajuste do primeiro modelo, Z_2 dados de ajuste do segundo modelo e D_1 dados a serem usados na medição da diversidade entre os modelos. Todos representam apenas sinais de entrada dos modelos.	56
4.16	Ilustração da curva estática de um sistema S_1 aproximado por dois modelos, M_1 e M_2 , em que, em destaque, é apresentada a faixa de operação usada no treinamento desses modelos.	57
4.17	Ilustração dos passos do procedimento para medição da diversidade, apresentados na Seção 4.3.3. Os sinais aqui apresentados são apenas ilustrativos.	60
4.18	Simulação livre de dois <i>ensembles</i> compostos pelos mesmos modelos (vide Tabela 4.3), mas com ponderações distintas. O sinal indicado por E_1 representa a média simples das saídas dos modelos; E_2 representa a média ponderada pela diversidade, conforme Eq. 4.27; medido representa o valor da pressão PDG-P. Usando a métrica de erro definida na Eq. 4.29 (MAPE), tem-se: $e_{E_1} = 6,4\%$, $e_{E_2} = 3,5\%$	65
4.19	Pesos atribuídos aos modelos de acordo com a diversidade. A ordem dos modelos é a mesma apresentada na Tabela 4.3.	66

Lista de Tabelas

3.1	Desempenho de modelos identificados para o sistema de bombeamento.	23
4.1	Parâmetros estruturais das redes neurais utilizadas na modelagem do sistema dinâmico não linear apresentado na Eq. 4.19.	49
4.2	Medidas de diversidade nos <i>ensembles</i> $E_1 = \{M_1M_2\}$ e $E_2 = \{M_1M_3\}$, usando quatro métricas distintas: correlação, ambiguidade, descorrelação e informação mútua (vide Seção 4.2.1 para detalhes); e usando três conjuntos de dados de entrada distintos: D_1 , D_2 e D_3	54
4.3	Características dos modelos que compõem o <i>ensemble</i> de estimação da pressão de fundo de poço de extração de petróleo, erro (MAPE) sobre os dados de validação e o peso atribuído pelo combinador.	63

Lista de Siglas

ANM	Árvore de Natal Molhada
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
CPH	Centro de Pesquisas Hidráulicas e Recursos Hídricos
DELT	Departamento de Engenharia Eletrônica da UFMG
FAPEMIG	Fundação de Amparo à Pesquisa do Estado de Minas Gerais
FCO	Fundação Christiano Ottoni
FFT	Transformada rápida de Fourier - <i>Fast Fourier Transform</i>
GLC	<i>Gas-Lift</i> Contínuo
iFFT	Transformada rápida de Fourier inversa - <i>Inverse Fast Fourier Transform</i>
LIT	Linear Invariante no Tempo
MA	Média móvel - <i>Moving Average</i>
MACSIN	Grupo de Pesquisa em Modelagem, Análise e Controle de Sistemas Não-lineares
MAPE	Erro Percentual Absoluto Médio - <i>Mean Absolute Percentage Error</i>
mlc	metros de coluna de líquido
NARMAX	modelo não linear autorregressivo, de média móvel com entradas exógenas - <i>Nonlinear AutoRegressive Moving Average model with eXogenous inputs</i>
NARX	modelo não linear autorregressivo, com entradas exógenas - <i>Nonlinear AutoRegressive model with eXogenous inputs</i>
PDF	Função densidade de probabilidade - <i>Probability Density Function</i>
PDG	<i>Permanent Downhole Gauge</i>
PPGEE	Programa de Pós-Graduação em Engenharia Elétrica
PRMSE	Percentual da raiz do erro quadrático médio - <i>Percentage Root Mean Square Error</i>
RMSE	Raiz do erro quadrático médio - <i>Root Mean Squared Error</i>
rpm	rotações por minuto

- SDV** Válvula de Desligamento - *ShutDown Valve*
- SISO** Única entrada e única saída - *Single Input, Single Output*
- UFMG** Universidade Federal de Minas Gerais

Introdução

A compreensão humana se dá por meio de modelos. Reduzimos o real em partes de interesse, ideias menores, mais tangíveis ao nosso entendimento e mais facilmente manipuláveis, baseando-nos em observações ou compreensões que temos da realidade. Elaboramos uma ideia em escala reduzida da complexa dimensão do real. Criamos modelos.

Seja um desenho de rascunho ou uma maquete, os modelos têm como função principal a boa representatividade do sistema real, naquele objetivo específico para o qual foi desenvolvido. Um modelo de vestuário tem diferentes requisitos de um modelo arquitetônico, por exemplo. Portanto, além de fácil manipulação, os modelos devem atender aos objetivos de sua construção.

Em engenharia não é diferente. Busca-se modelos matemáticos melhor manipuláveis, mais compreensíveis, que expliquem e representem bem, com determinado objetivo, os mais complexos fenômenos reais, das mais variadas maneiras. Quando esses fenômenos evoluem com o tempo (parâmetro de evolução), são chamados de *sistemas dinâmicos*. A área que busca explicação matemática para esses fenômenos é a *modelagem de sistemas dinâmicos*, principal área tocada por esta dissertação.

Equacionar um sistema real por completo, com todas as funções e parâmetros que o regem, é, por vezes, uma tarefa bastante afanosa. Ante a essa dificuldade, surgiu a área de *identificação de sistemas*, que busca construir modelos por meio de dados. Por outro lado, com o desenvolvimento da inteligência artificial, outra área que ganha espaço na identificação de sistemas é o *aprendizado de máquina* aliado a modelos do tipo *redes neurais artificiais*. Principalmente com as redes Perceptron multicamadas (MLP, do termo em inglês *multilayer perceptron*), que, com apenas uma camada escondida, são, comprovadamente, aproximadoras universais de funções [Cybenko, 1989].

Este trabalho aborda duas possíveis tendências do uso de redes neurais em identificação de sistemas dinâmicos. A primeira é a identificação de sistemas usando informação auxiliar, procedimento presente em vários trabalhos com outros tipos de modelos [Karplus, 1977, Bohlin, 1991, Tulleken, 1993, Johansen, 1996, Corrêa, 2001, Nepomuceno et al., 2003], mas incomum em redes neurais [Joerding and Meador, 1991, Thompson and Kramer, 1994, Cubillos et al., 1996, Amaral, 2001]. Essa abordagem ainda utiliza o erro de simulação livre no ajuste dos parâmetros, que, em geral, gera resultados mais interessantes [Aguirre et al., 2010]. A segunda tendência é o uso de máquinas de comitê, ou *ensembles* de modelos (neurais ou não), na modelagem de sistemas dinâmicos. Procedimento comum na área de previsão de séries temporais [Bates and Granger, 1969, Newbold and Granger, 1974, Winkler and Makridakis, 1983], mas não muito comum em identificação de sistemas.

O objetivo específico deste trabalho é a análise de ambas as abordagens (i.e. uso de informação auxiliar em redes neurais e formação de comitês de modelos dinâmicos) em problemas que dispõem de pouca informação presente nos dados de identificação. O processo motivador deste trabalho é a extração de petróleo em águas profundas (descrito no Capítulo 2), que reflete bem a situação em que

os dados disponíveis são pouco representativos para o comportamento global do sistema. Tanto o uso de informação auxiliar quanto a combinação de modelos (*ensembles*) são tentativas de extrapolar as informações presentes nos dados, são maneiras de fazer com que os modelos aprimorem sua *capacidade de generalização*. Enquanto a primeira abordagem tenta inserir um conhecimento prévio (comportamento estático) do comportamento global do sistema, a segunda é uma tentativa de combinar melhor os modelos identificados (e validados), de modo a não reforçar erros comuns aos modelos. Ficará claro que o objetivo aqui não é obter o melhor modelo para um conjunto (e.g. limitado) de dados de teste ou validação, mas sim obter uma resposta mais aceitável em *toda faixa de operação do sistema*.

O primeiro capítulo faz uma introdução, apresenta o contexto e demarca os temas abordados nesta dissertação. O segundo capítulo descreve o processo motivador desse trabalho: a extração de petróleo em águas profundas, apresentando o funcionamento do processo, os desafios associados e aplicabilidade da identificação de sistemas nesse contexto. O terceiro capítulo mostra um estudo, autocontido, sobre identificação de sistemas dinâmicos em modelos neurais, usando a curva estática como informação auxiliar e o erro de simulação livre no ajuste dos parâmetros. O quarto capítulo faz um estudo, também autocontido, sobre o uso de máquinas de comitê (*ensembles*) em identificação de sistemas, com foco no conceito de *diversidade* de modelos dinâmicos, em que são apresentadas formas de mensurar diversidade e os fatores mais influentes na quantificação desse conceito. O quinto capítulo traz as considerações finais, com as possíveis contribuições deste trabalho e algumas perspectivas de trabalhos futuros.

Extração de Petróleo em Águas Profundas

Este capítulo descreve o processo motivador desta dissertação: a extração de petróleo em águas profundas, que está inserida no contexto do projeto, fruto da parceria entre a UFMG e a PETROBRAS, com o objetivo de investigar, definir, desenvolver e testar modelos matemáticos e técnicas de processamento de sinais no contexto de poços de produção de petróleo para a estimação da pressão de fundo de poço [Aguirre, 2011]. Projeto esse que tem importância estratégica no cenário nacional, dado o crescimento da indústria petrolífera nos últimos anos e a importância da aplicação de novas tecnologias na área, especialmente na extração de petróleo [Jansen et al., 2008]. Outros trabalhos foram realizados no mesmo contexto e podem ser consultados para melhor compreensão do projeto e das ferramentas utilizadas. São eles: [Pankiewicz, 2011, Teixeira et al., 2012, Abreu et al., 2012, Araújo, 2012, Lima, 2012, Castro, 2013].

Dentre os diversos desafios na extração de petróleo em águas profundas, destaca-se a otimização de seus processos de produção, que busca reduzir custos e elevar a produtividade dos poços. Porém, para alcançar tais metas, faz-se necessário um bom conhecimento do processo em questão, especialmente da dinâmica do poço durante sua vida útil. Nesse cenário, a modelagem matemática de poços produtores pode contribuir na previsibilidade de sua vida útil e na otimização de seus processos. Dentre as variáveis que compõem o complexo cenário de extração, a pressão do fundo do poço tem importante papel na determinação de sua produtividade. Contudo, a medição dessa pressão é mais um desafio e, agravado pelo longo período de operação do sensor (vida útil do reservatório) que pode chegar a algumas décadas, a falha desse sensor é, praticamente, dada como certa. Por outro lado, a manutenção do sensor não é uma alternativa, devido aos elevados custos de parada de produção e eventuais riscos ambientais.

Assim, a estimação da pressão de fundo de poço, baseada nas demais medições disponíveis (sensor virtual), se mostra uma alternativa interessante. A seguir, são apresentados maiores detalhes sobre o processo mencionado.

2.1 O Processo de Produção

A partir de substâncias orgânicas, depositadas junto com sedimentos, que, sob certas condições termoquímicas e após uma cadeia de processos, é formado o composto de hidrocarbonetos que chamamos de petróleo. Apesar de ter origem em rochas ditas *geradoras*, o petróleo se desloca, por motivos ainda questionáveis [Thomas, 2004], e acumula-se nas *rochas-reservatório*, alvo da extração.

Um reservatório de petróleo *off-shore* é constituído por poços produtores (vide Figura 2.1), que podem ser classificados em *surgentes* e *não surgentes*. O primeiro refere-se àqueles cuja pressão do reservatório

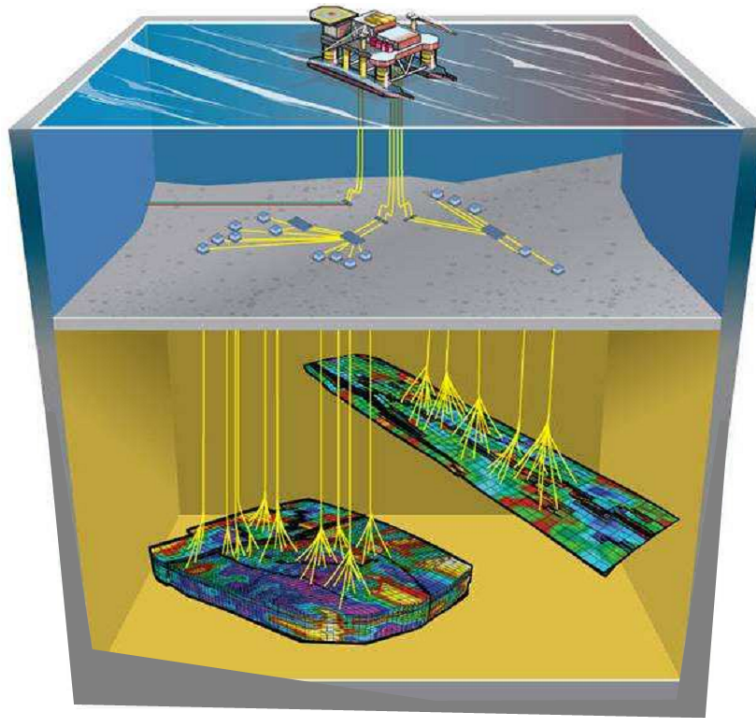


Figura 2.1: Ilustração de uma plataforma de extração de petróleo *off-shore* conectada a dois poços de extração. Fonte: [Teixeira, 2011].

é suficientemente elevada para que os fluidos alcancem livremente a superfície, por *elevação natural*. O segundo refere-se àqueles cuja pressão do reservatório é, relativamente, baixa e os fluidos contidos nele não chegam à superfície de forma natural, fazendo-se necessária alguma técnica de *elevação artificial*. Tais técnicas também são empregadas em poços surgentes ao final de sua vida produtiva ou quando a vazão do poço está muito abaixo do que deveria produzir [Thomas, 2004].

Na elevação artificial, destaca-se a técnica de injeção de *gas-lift* contínuo (GLC), pela sua versatilidade (em termos de vazão e profundidade) e por exigir investimentos relativamente baixos para poços profundos [Thomas, 2004]. O GLC consiste, basicamente, na injeção do gás, em alta pressão, de forma controlada e contínua, com o objetivo de gaseificar o fluido e aumentar a produção por meio de dois mecanismos físicos: energia de expansão do gás injetado, que desloca o líquido para a superfície, e a redução da massa específica do fluido, que faz com que a pressão do reservatório seja suficiente para deslocá-lo à superfície [Teixeira, 2011]. Na produção ocorre então o escoamento bifásico (líquido e gás) que são separados na plataforma. A figura 2.2 mostra um diagrama simplificado sobre a injeção de *gas-lift* na extração de petróleo.

Neste trabalho, são considerados os poços produtores de petróleo *off-shore*, não surgentes, que utilizam da técnica de elevação por injeção de *gas-lift* contínuo. A seguir, são apresentados os sistemas de instrumentação utilizados nesse tipo de extração.

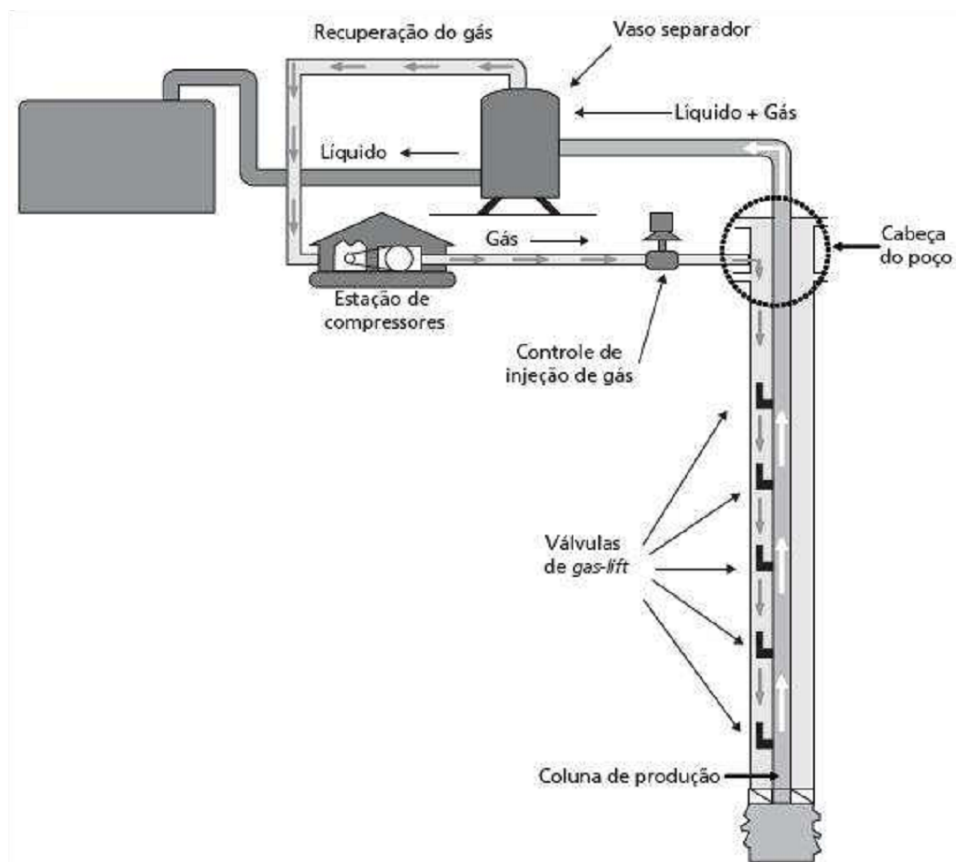


Figura 2.2: Diagrama simplificado da extração de petróleo utilizando a técnica de elevação artificial por *gas-lift* contínuo. Fonte: [Teixeira, 2011].

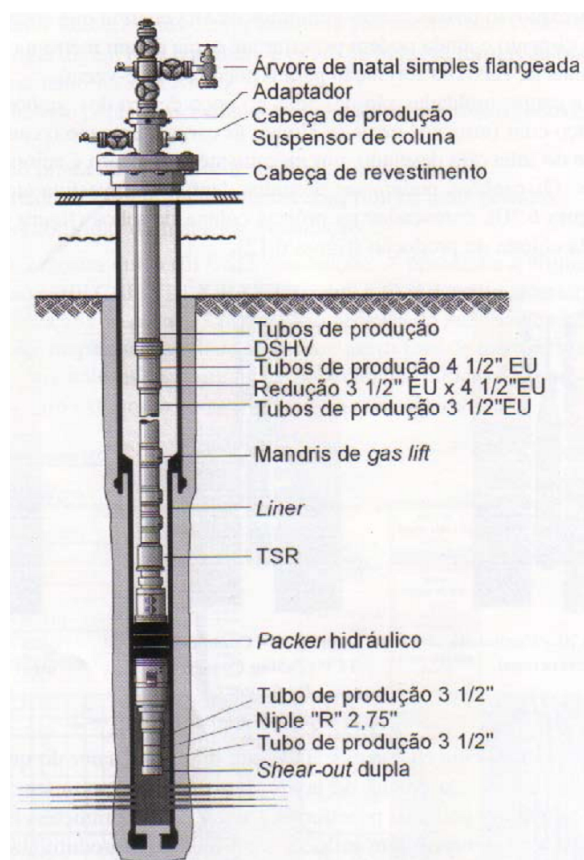


Figura 2.3: Coluna de produção convencional utilizando a técnica de elevação artificial por *gas-lift*. Fonte: [Thomas, 2004].

2.2 Instrumentação

Nos poços de produção são inseridas tubulações responsáveis por conduzir o petróleo da reserva até a cabeça do poço, chamadas de *coluna de produção*. A Figura 2.3 mostra uma coluna de produção convencional, onde situa-se o sensor *permanent downhole gauge* (PDG), que disponibiliza importantes informações de pressão e temperatura do poço. Os *dutos de produção* completam o curso do petróleo, conduzindo-o da cabeça do poço até a plataforma, em que o trecho que toca o leito marinho é chamado de *flowline* e parte suspensa até a plataforma é chamada de *riser*. O petróleo é extraído de vários poços que são conectados à um mesmo sistema de válvulas, *manifold submarino de produção*, para então ser conduzido à plataforma.

Na cabeça de cada poço é instalado um conjunto de válvulas, conexões e adaptadores que permitem o fluxo do fluido à superfície de forma segura e controlada. Esse sistema é chamado de *árvore de natal* que, no caso de poços submersos são do tipo ANM, sigla para *árvore de natal molhada*. A Figura 2.4 mostra um esquema desse complexo sistema denominado, para este caso, *completação submarina* ou *completação molhada*.

Na plataforma, o fluido produzido alimenta um separador, em geral, trifásico, que separa o fluido em óleo, gás e água. A Figura 2.5 mostra um esquema desse tipo de separador.

O diagrama simplificado de instrumentação típica de um sistema de extração de petróleo em águas profundas é mostrado na Figura 2.6. Verifica-se a presença das válvulas de desligamento *SDV* (*shutdown*

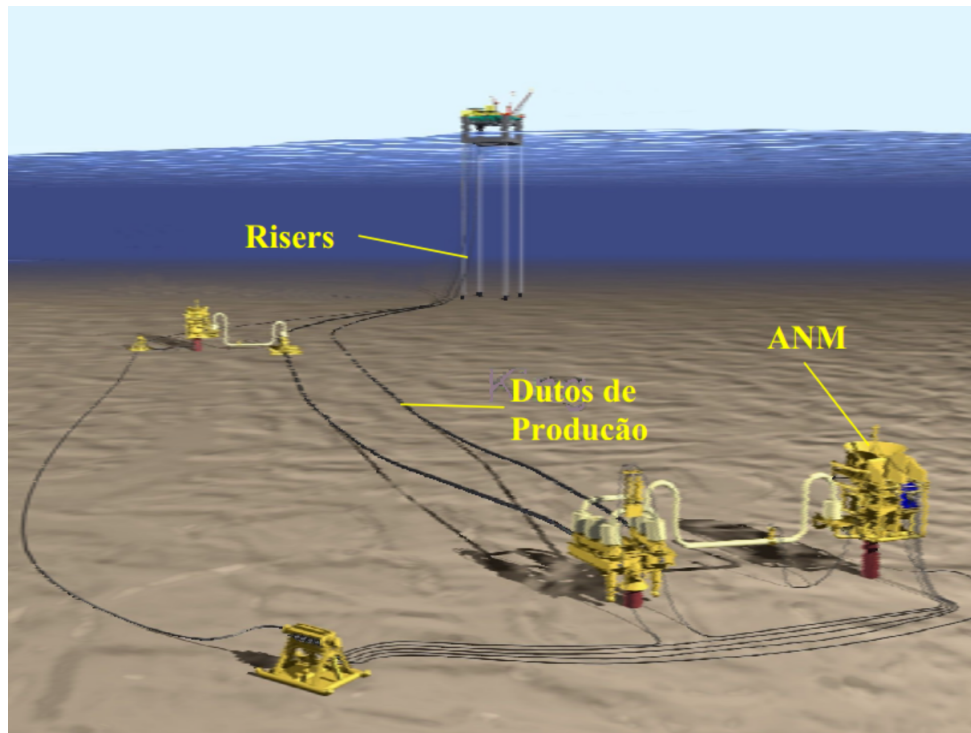


Figura 2.4: Esquema representando a completação submarina (molhada) de um poço de petróleo em águas profundas. Fonte: [Villela, 2005].

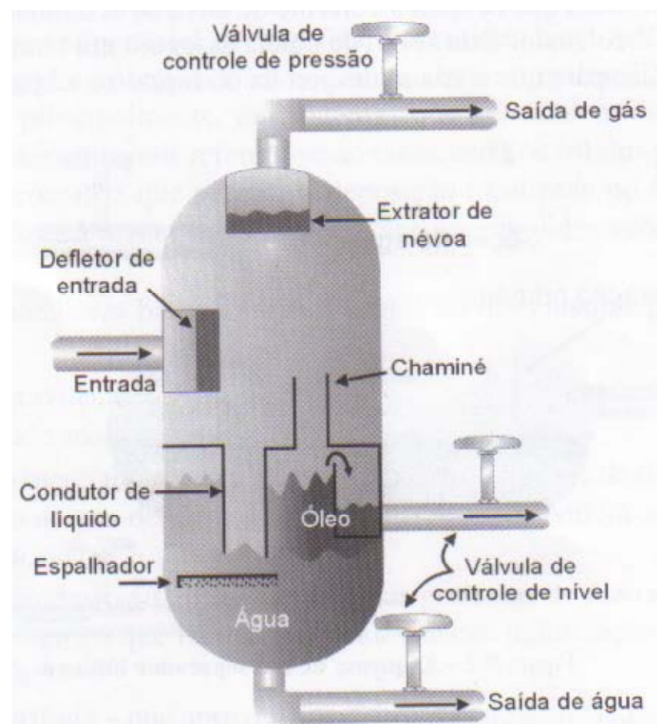


Figura 2.5: Esquema de um separador trifásico, encontrado em plataformas petrolíferas, que separa o fluido do poço em óleo, gás e água. Fonte: [Thomas, 2004].

valve) e de estrangulamento *choke*. A primeira é operada em situações de emergência, enquanto a segunda é operada, em malha aberta, para fazer restrições de vazão nas linhas de produção e injeção de *gas-lift*. No diagrama são apresentadas as seguintes medições:

- Vazão instantânea de *gas-lift*,
- Abertura da válvula *choke* de *gas-lift*,
- Temperatura à montante de SDV de *gas-lift*,
- Pressão à montante de SDV de *gas-lift*,
- Pressão à jusante de SDV de *gas-lift*,
- Pressão de fundo de poço no PDG,
- Temperatura de fundo de poço no PDG,
- Pressão na árvore de natal,
- Temperatura na árvore de natal,
- Pressão à montante de SDV,
- Pressão à montante de *choke*,
- Temperatura à montante de *choke*,
- Abertura da válvula *choke* de produção,
- Pressão à jusante de *choke*.

Nem todas essas quatorze medições estão disponíveis em todos os poços. Na obtenção de modelos dinâmicos para estimação da pressão no PDG, foram utilizados, principalmente, os instrumentos da plataforma, devido às melhores condições de manutenção e maior confiabilidade das medições.

As medições foram coletadas, manualmente, nos sistemas historiadores da planta de extração. No contexto do projeto, foram realizados estudos sobre a influência da compressão de dados realizada nos historiadores que concluíram que, para os objetivos do projeto (e deste trabalho), não há influência significativa. A coleta de dados foi realizada de forma a priorizar trechos que tenham comportamentos dinâmicos representativos e evitando trechos que apresentam casos particulares ou situações peculiares que não representam o funcionamento normal da planta. Foram observados vários tipos de dinâmica de poços, em várias fases de sua vida útil. Uma importante característica que varia de poço para poço (e com seu tempo de vida) é o regime de *golfadas*, que está relacionado com a forma que se dá o escoamento do fluido à superfície. Maiores informações são encontradas no trabalho de [Thomas, 2004].

2.3 Conclusões do Capítulo

Esse foi um breve capítulo que descreveu um pouco do processo que motivou essa dissertação: a extração de petróleo em águas profundas. Foi apresentado um pouco do processo, as medições disponíveis no sistema e a coleta de dados. Nos próximos capítulos, o processo aqui descrito é utilizado na aplicação das técnicas desenvolvidas.

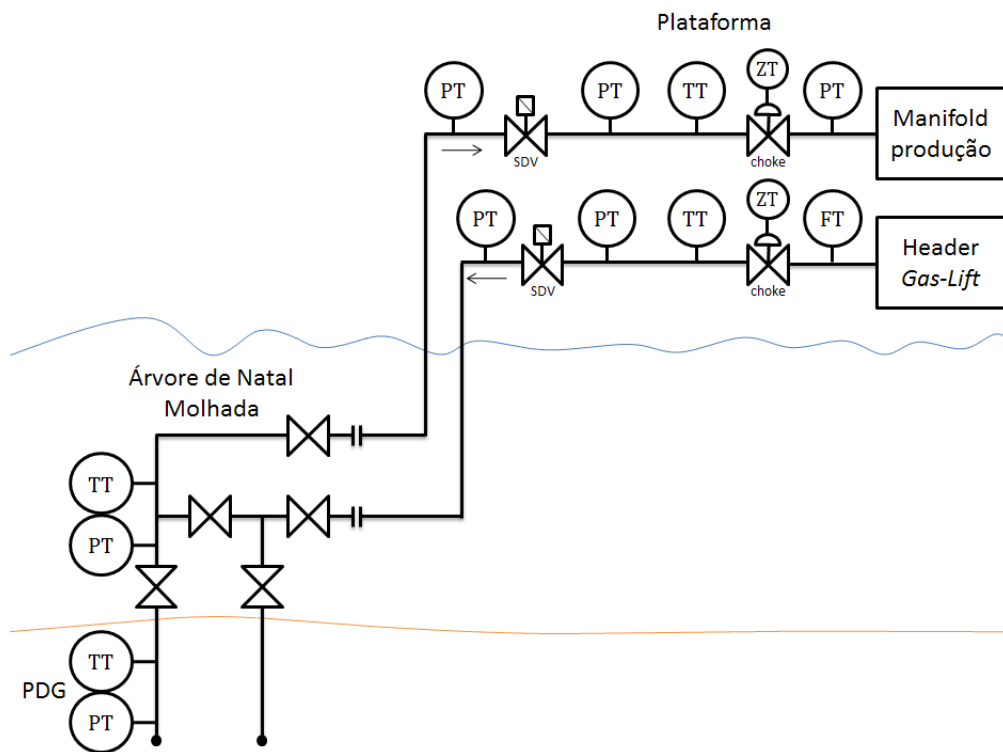


Figura 2.6: Diagrama simplificado de instrumentação típica da extração de petróleo em águas profundas.

Informação Auxiliar na Identificação de Sistemas

3.1 Introdução

¹ O uso de redes neurais artificiais vem ganhando espaço em diversos campos do conhecimento [Paliwal and Kumar, 2009], dentre outros motivos, devido ao seu bom desempenho ante às mais variadas espécies de problemas. Outro fator motivador da disseminação das redes é o amplo desenvolvimento das áreas relacionadas, tais como aprendizado de máquina e otimização. Esses e outros fatores fizeram com que as redes neurais se tornassem uma solução de prateleira: aplicável em várias áreas, dispensando vastos conhecimentos do processo em questão (a rede “aprende” o processo) e, com a incorporação dos algoritmos de treinamento em diversas ferramentas, sem a necessidade de grandes conhecimentos sobre aprendizado de máquina por parte do usuário. Além da incitante ideia de que foram inspiradas nas estruturas neurais de organismos inteligentes.

Na modelagem de sistemas dinâmicos, as redes neurais também marcam presença há décadas [Narendra and Parthasarathy, 1990, Chen et al., 1990], especialmente em identificação de sistemas em que a informação modelada do sistema é extraída de dados. Nesse contexto, foram publicados inúmeros trabalhos de identificação de sistemas (e.g. não lineares) que utilizam estruturas neurais em seus modelos. Um dos maiores desafios na área de identificação de sistemas não lineares é o não conhecimento do funcional que faz o mapeamento da entrada para a saída, em que, informações auxiliares do processo fornecem grande ajuda na obtenção de modelos mais precisos [Kerschen et al., 2006]. Contudo, nem sempre é fácil incorporar informação extra em estruturas neurais e, por vezes, opta-se por deixar de acrescentar a informação ou partir para uma estrutura mais simples, e.g. linear, em que a informação é mais facilmente incorporada.

3.1.1 Contextualização

Modelagem é a área do conhecimento que estuda maneiras de desenvolver e implementar modelos matemáticos de sistemas reais [Aguirre, 2007]. Uma das formas de obter esses modelos é pela *modelagem caixa branca*, em que as equações do modelo são baseadas nos processos físicos, químicos ou fenomenológicos do sistema real [Garcia, 2005]. Nesse tipo de modelagem, é necessário um vasto conhecimento do processo envolvido e, por vezes, um longo tempo disponível para obtenção dos modelos — essa forma de modelagem não será tratada nesse trabalho. *Identificação de sistemas*, por outro lado, estuda técnicas

¹Partes desse capítulo foram apresentados em [Abreu et al., 2012].

alternativas de modelagem matemática. Dentre elas, temos as técnicas de identificação *caixa preta* (ou *modelagem caixa preta*), em que os modelos matemáticos são obtidos baseando-se apenas nas medições (e.g. entradas e saídas) de processo. Nesse procedimento, nenhum conhecimento prévio sobre o processo é necessário, o que torna a identificação aplicável aos mais diversos tipos de sistemas, desde previsão do tempo até a modelagem de sistemas sociais e financeiros. Pode-se dizer que a identificação caixa preta é fruto dessa busca por generalidade de aplicação do método.

Uma alternativa à modelagem caixa preta é denominada *modelagem caixa cinza* [Aguirre, 2007], sendo que esta se diferencia daquela pelo fato de que, no processo de obtenção do modelo, além do conjunto de dados dinâmicos é também utilizada *informação auxiliar*. Nesse procedimento, a maior dificuldade encontrada é a inserção dessa informação durante o treinamento dos modelos dinâmicos.

Além dessa distinção entre modelos caixa branca, preta ou cinza, existem várias outras formas de classificação e categorização de modelos. Para as discussões que seguem, este trabalho se restringe a modelos dinâmicos, determinísticos, discretos, invariantes no tempo, paramétricos, com parâmetros concentrados, não autônomos e não lineares. Maiores detalhes sobre o significado de cada uma dessas classificações pode ser encontrada em [Aguirre, 2007].

A fim de facilitar a discussão, os seguintes aspectos relacionados ao problema de identificação serão considerados: i) a classe de modelos; ii) a classe de algoritmos utilizados no treinamento e iii) o tipo de informação auxiliar utilizada, sendo que este último se aplica apenas na identificação caixa cinza.

A *classe de modelos* refere-se a estrutura do modelo utilizada na identificação, tais como modelos lineares, modelos não lineares polinomiais, modelos RBF (*Radial Basis Function*), redes neurais. Vários exemplos de identificação de sistemas utilizando essas estruturas são encontrados na literatura, tanto para modelagem caixa preta [Corrêa et al., 2000, Aguirre et al., 2007, Xie et al., 2009] como para modelagem caixa cinza [Herbert and Tulleken, 1993, Johansen, 1996, Corrêa et al., 2002, Aguirre et al., 2004b, Aguirre et al., 2004a, Aguirre et al., 2007, Chen et al., 2008, Barbosa, 2009, Barbosa et al., 2011].

Quanto ao tipo de *algoritmo de treinamento*, para o ajuste dos parâmetros do modelo aos dados, existem exemplos de identificação, caixa preta e caixa cinza, que utilizam de técnicas bayesianas [Herbert and Tulleken, 1993], otimização não linear [Corrêa et al., 2002], estimador de mínimos quadrados com restrições [Aguirre et al., 2004a], e otimização multiobjetiva [Johansen, 1996, Nepomuceno et al., 2007, Barbosa et al., 2011].

Finalmente, com respeito ao *tipo de informação auxiliar* utilizada, tem-se: estabilidade e sinal algébrico do ganho (para modelos lineares) [Herbert and Tulleken, 1993]; valores do sistema em estado estacionário [Nepomuceno et al., 2007], curva estática [Barbosa et al., 2011], simetria [Amaral, 2001, Aguirre et al., 2004b]. Vale mencionar que em alguns casos a informação auxiliar pode ser incorporada no modelo na etapa de determinação de estrutura (topologia) seguido de treinamento com algoritmos convencionais [Aguirre et al., 2004b, Connally et al., 2007].

Especificamente no que concerne a modelos do tipo redes neurais [Silva et al., 2010], informação sobre a curva estática de um sistema foi incorporada por meio de um treinamento em duas etapas em [Amaral, 2001]. Por outro lado, algoritmos multiobjetivos foram utilizados para melhorar o treinamento de redes neurais em [Teixeira et al., 2000]. No procedimento aqui descrito, esses dois elementos são combinados, ou seja, modelos neurais (*multilayer perceptron*) dinâmicos serão treinados utilizando a curva estática como informação auxiliar, por meio de otimização multiobjetivo.

3.1.2 Metodologia Proposta

Este capítulo descreve um procedimento para se adicionar informação auxiliar em modelos do tipo rede neural. A informação auxiliar aqui tratada é a curva estática do sistema e a rede neural do tipo MLP (*multilayer perceptron*). A adição de informação auxiliar proposta é feita na fase de treinamento (i.e. ajuste dos parâmetro), utilizando o erro de simulação livre de dados dinâmicos e o erro sobre a curva estática. Considera-se que o – importante – problema de escolha de estrutura do modelo já tenha sido resolvido. O procedimento gera um conjunto de modelos, cuja validação/escolha do(s) modelo(s) final(is) não é discutida aqui.

O procedimento é avaliado em um sistema hidráulico de bombeamento [Barbosa, 2006] e, posteriormente, aplicado a um sistema de extração de petróleo em águas profundas (vide Capítulo 2). Modelos caixa preta e caixa cinza são comparados a fim de salientar os principais aspectos da metodologia. Finalmente, são feitas considerações acerca da aplicabilidade e vulnerabilidade do procedimento.

3.2 Procedimento de Inserção de Informação Auxiliar em Redes Neurais

A seguir, serão mostrados os principais pontos do procedimento utilizado para treinar redes neurais de maneira a incorporar informação auxiliar sobre a curva estática.

3.2.1 Estruturas dos modelos

Neste trabalho, foram obtidos modelos NARMAX (*Nonlinear AutoRegressive Moving Average with exogenous inputs*), cuja estrutura geral é mostrada a seguir:

$$y(k) = F[y(k-1), \dots, y(k-n_y), u(k-1), \dots, u(k-1-n_u), e(k-1), \dots, e(k-n_e)] + e(k), \quad (3.1)$$

em que n_y , n_u e n_e são os atrasos máximos considerados de saída $y(k)$, entrada $u(k)$ e ruído $e(k)$, respectivamente, e F uma função não linear, que será implementada por meio de modelos perceptron multicamadas (MLP).

Uma das etapas mais difíceis na construção de modelos não lineares, qualquer que seja a classe de modelos considerada, é a determinação da estrutura, ou topologia. Em consequência disso, diversos trabalhos descrevem formas de determinar uma topologia adequada para redes neurais [Mezard and Nadal, 1989, Fahlman and Lebiere, 1990, Baffes and Zelle, 1992, Romaniuk and Hall, 1993, Bartlett, 1994, Garcia-Pedrajas et al., 2003]. Em particular, alguns aspectos da topologia da rede podem ser ajustados para incorporar certos tipos de informação auxiliar [Amaral, 2001, Aguirre et al., 2004b]. Uma abordagem diferente a esse importante problema é o treinamento multiobjetivo, a fim de melhorar a capacidade de generalização da rede treinada [Teixeira et al., 2000]. Essa etapa é igualmente importante tanto no treinamento caixa preta, quanto no treinamento utilizando informação auxiliar.

3.2.2 Ajuste dos pesos

Para o ajuste dos parâmetros nos modelos neurais caixa preta foi utilizado o algoritmo de Levenberg-Marquadt, implementado no pacote *Norgaard Toolbox* [Norgaard, 1997]. Esse algoritmo ajusta os pesos

da rede de maneira a minimizar a norma L_2 do vetor de resíduos, ou seja, o vetor de erros de predição de um passo à frente ao longo dos *dados dinâmicos de treinamento* Z . A seguinte nomenclatura será utilizada no restante deste trabalho.

O conjunto de dados dinâmicos é dado por

$$Z = \begin{bmatrix} y(1) & u(1) \\ y(2) & u(2) \\ \vdots & \vdots \\ y(N) & u(N) \end{bmatrix}.$$

A predição de um passo à frente é

$$\tilde{y}(k) = F[\mathbf{y}(k-1), \mathbf{u}(k-1)], \quad (3.2)$$

sendo $\mathbf{y}(k-1) = [y(k-1), \dots, y(k-n_y)]^T$ e o vetor de entrada $\mathbf{u}(k-1) = [u(k-1), \dots, u(k-n_u)]^T$.

Assim, o treinamento caixa preta é dado pela solução do seguinte problema de otimização

$$\hat{\theta} = \arg \min_{\theta} J_1(\theta, Z), \quad (3.3)$$

em que θ é o vetor de pesos da rede e a função de custo a ser minimizada é

$$J_1(\theta, Z) = \sum_{k=1}^N [y(k) - \tilde{y}(k)]^2. \quad (3.4)$$

A expressão (3.4) mostra que no presente tipo de treinamento *unicamente* os dados dinâmicos Z , de comprimento N , são utilizados. Além disso, pelo fato de a relação entre θ e $\tilde{y}(k)$ ser não linear, o problema (3.3) é não convexo.

Além dos dados de treinamento Z , são utilizados conjuntos de dados de *teste* e *validação*, ambos diferentes dos dados de treinamento. O desempenho sobre os dados de teste foi utilizado como critério para a escolha da rede neural caixa preta, apresentada nos resultados. Os dados de validação tem como função a avaliação do desempenho final do procedimento. Após a análise do desempenho sobre os dados de validação, não se faz qualquer tipo de ajuste de parâmetros ou escolha do modelo final.

3.2.3 Uso de informação auxiliar

Deseja-se treinar uma rede, mas de maneira diferente àquela descrita na Seção 3.2.2. Primeiramente, define-se a predição livre como

$$\hat{y}(k) = F[\hat{\mathbf{y}}(k-1), \mathbf{u}(k-1)], \quad (3.5)$$

sendo $\hat{\mathbf{y}}(k-1) = [\hat{y}(k-1), \dots, \hat{y}(k-n_y)]^T$.

Além disso, assume-se que existe o seguinte conjunto de dados em estado estacionário

$$\bar{Z} = \begin{bmatrix} \bar{y}(1) & \bar{u}(1) \\ \bar{y}(2) & \bar{u}(2) \\ \vdots & \vdots \\ \bar{y}(M) & \bar{u}(M) \end{bmatrix},$$

sendo que a barra indica valores constantes em estado estacionário e, tipicamente, $M \ll N$. Assim, é possível definir as seguintes funções de custo:

$$J_S(\theta, Z) = \sqrt{\frac{1}{N} \sum_{k=1}^N [y(k) - \hat{y}(k)]^2} \quad e \quad (3.6)$$

$$J_{SF}(\theta, \bar{Z}) = \sqrt{\frac{1}{M} \sum_{k=1}^M [\bar{y}(k) - \hat{y}(k)]^2}, \quad (3.7)$$

sendo que $\hat{y}(k) = F[\hat{\mathbf{y}}, \bar{\mathbf{u}}]$ é a saída da rede em estado estacionário.

A justificativa para o uso de J_S em identificação de modelos, em vez da função objetivo “clássica” J_1 foi discutida por [Aguirre et al., 2010]. Portanto, a informação dinâmica do sistema, contida em Z , é aprendida pelo modelo ao minimizar J_S . Se a informação estática não estiver bem representada em Z , é de se esperar que o desempenho do modelo em estado estacionário seja ruim.

Considera-se que *também* se tem acesso ao conjunto de dados \bar{Z} . O treinamento que minimize apenas J_{SF} resultará em um modelo dinâmico muito ruim, pois não há informação dinâmica em \bar{Z} , por definição. Percebe-se que uma solução de compromisso é procurar utilizar Z e \bar{Z} *simultaneamente*. Uma maneira de alcançar esse alvo é por meio de técnicas multiobjetivas. Assim, ao definir um objetivo “composto” $\mathbf{J}(\theta, Z, \bar{Z}) = [J_S(\theta, Z) \ J_{SF}(\theta, \bar{Z})]$, o treinamento resulta na solução do seguinte problema não convexo para o caso de redes neurais:

$$\begin{cases} \hat{\theta} & = \arg \min_{\theta} \mathbf{J}(\theta, Z, \bar{Z}) \\ \text{sujeito a} & : \theta \in \mathbb{R}^n. \end{cases} \quad (3.8)$$

A solução do problema (3.8) é, na verdade, um conjunto de soluções, em que cada uma das soluções apresenta uma melhoria em relação a qualquer outra solução em, pelo menos, um dos objetivos. Em outras palavras, o conjunto será composto por soluções *não dominadas* e é denominado conjunto de *soluções eficientes*, soluções *Pareto-ótimas* ou simplesmente *conjunto Pareto* [Takahashi, 2007]. Esse conjunto de soluções pode ser conseguido utilizando o algoritmo evolucionário multiobjetivo NSGA-II, como também implementado em [Barbosa, 2009]. A principal justificativa para a escolha de um algoritmo evolucionário é devido à não convexidade do problema de otimização proposto, devido a vários fatores, dentre eles o uso do erro de simulação livre e a alta complexidade do problema. Foi utilizado o algoritmo NSGA-II por questões de implementação, porém, a princípio, qualquer outro algoritmo de otimização multiobjetivo não muito susceptível a mínimos locais poderia ser utilizado. A próxima seção ilustra o procedimento com os resultados experimentais.

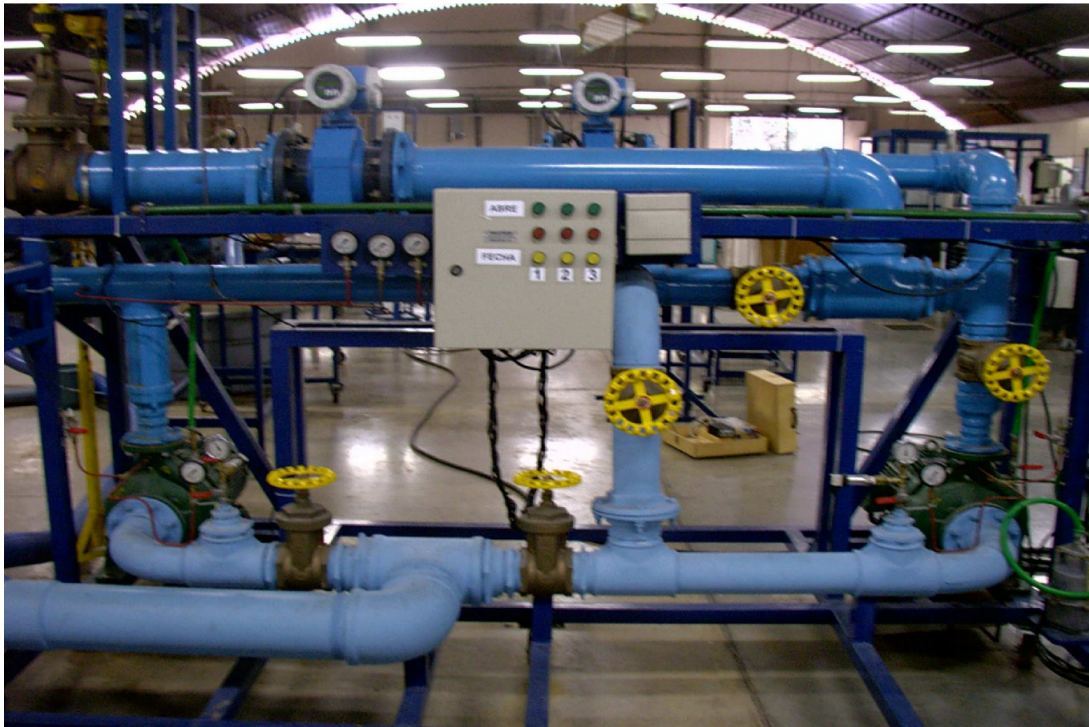


Figura 3.1: Sistema de bombeamento de água formado por dois conjuntos moto-bomba de 10cv cada. Fonte: [Barbosa, 2006].

3.3 Resultados Experimentais

Nesta seção serão apresentados resultados experimentais em duas aplicações: um sistema hidráulico de bombeamento e um sistema de extração de petróleo em poços não surgentes.

3.3.1 O Sistema Hidráulico

O sistema hidráulico utilizado está situado no Centro de Pesquisas Hidráulicas e Recursos Hídricos (CPH) da UFMG e tem por objetivo emular o comportamento de uma queda d'água acoplada a uma turbina hidráulica de uma usina hidrelétrica [Barbosa, 2006]. O sistema é composto por duas bombas, que simulam a queda d'água e um conjunto turbina-gerador, que simula o sistema de geração de uma usina.

A Figura 3.1 mostra o sistema de bombeamento, composto por duas bombas de 10cv, de características muito semelhantes, acionadas por dois inversores de frequência. Nos testes aqui descritos, foi utilizada a configuração paralela, em que as bombas recebem a mesma referência de velocidade. Juntas, as bombas alimentam uma turbina hidráulica *Francis* acoplada a um gerador. O conjunto turbina-gerador é mostrado na Figura 3.2.

Das várias medições instaladas no sistema, foram utilizadas, neste trabalho, a referência de velocidade das bombas, em rpm (rotações por minuto), e a pressão de recalque das bombas, em mlc (metros de coluna de líquido). A referência de velocidade é dada pelos inversores de frequência e a pressão é fornecida pelo transmissor, ambos coletados pelo sistema supervisor da planta, desenvolvido por [Barbosa, 2006]. Informações mais detalhadas sobre o sistema são encontradas nos trabalhos de [Caixeiro, 2003, Araújo, 2003, Duarte, 2004, Barbosa, 2006, Barbosa et al., 2011].

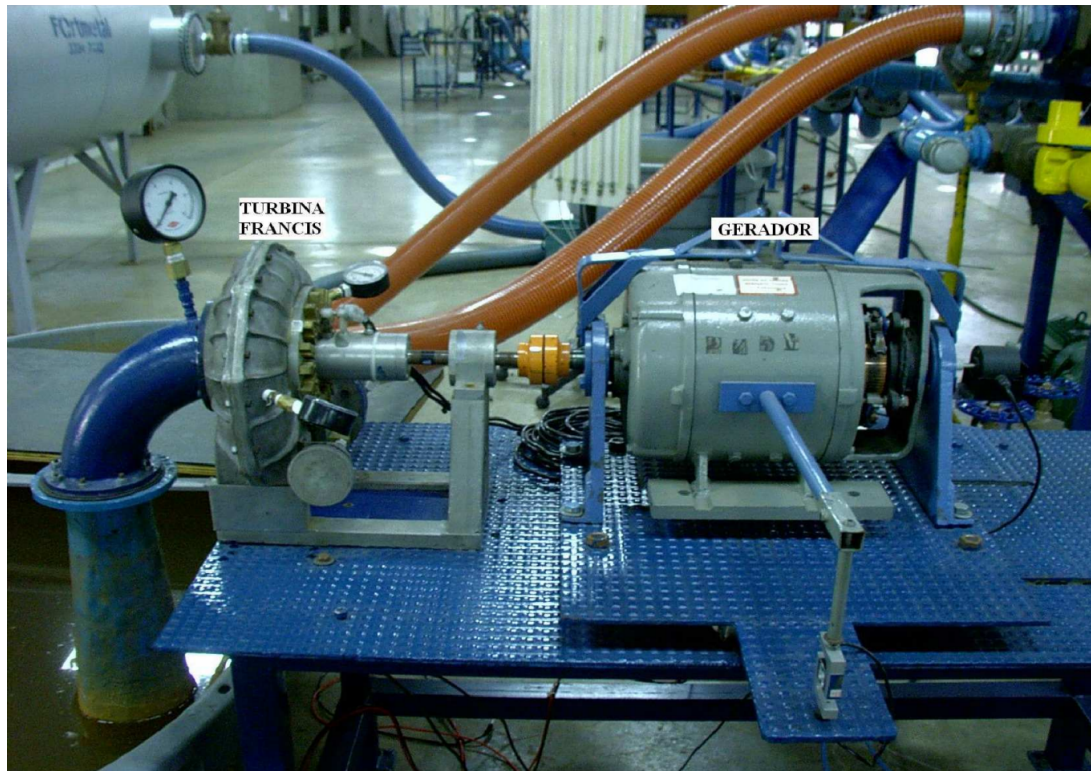


Figura 3.2: Conjunto turbina-gerador, acoplado ao sistema de bombeamento de água. Fonte: [Barbosa, 2006].

Ensaios

Como entrada do sistema, foi utilizado o valor de referência de velocidade das bombas hidráulicas, em rpm (rotações por minuto). Como saída, foi considerada a medição de pressão, a jusante das bombas, cuja precisão é de $\pm 0,175$ mlc (metros de coluna de líquido), ou 0,25% do fundo de escala do instrumento. Os testes foram realizados em malha aberta.

Para a realização dos *testes dinâmicos*, foi escolhido um período de amostragem $T_s = 50$ ms, de acordo com critérios adotados em [Barbosa et al., 2011]. A forma do sinal de entrada foi escolhida de forma a excitar o sistema em variados pontos de operação, com diferentes valores de amplitudes, de modo que o comportamento dinâmico do sistema fique bem representado no conjunto de dados. Para a identificação do sistema foram coletadas $N = 3100$ amostras, para teste e validação foram coletadas 750 amostras.

Os dados de identificação e teste do sistema produziram valores de saída compreendidos entre, aproximadamente, 13 e 19,2 mlc, sendo que a faixa de operação pode variar entre, aproximadamente, 6 e 25 mlc. Isso mostra certa limitação da faixa de operação nesses dados, o que é típico nos casos em que há restrições de várias naturezas durante a execução dos testes. Os dados de validação, por outro lado, foram coletados numa ampla faixa de operação, compreendida entre 6 e 25 mlc. Os sinais de entrada e saída utilizados para identificação e validação são mostrados nas Figuras 3.3 e 3.4, respectivamente, em que é notória a diferença entre a faixa de operação dos dados de identificação e validação, sendo o primeiro muito mais restrito do que o segundo. Esses dados servem para análise da capacidade dos modelos em apresentar boas respostas em regiões de operação desconhecidas.

Além dos testes dinâmicos, foi realizado um experimento para obtenção da curva estática do sistema.

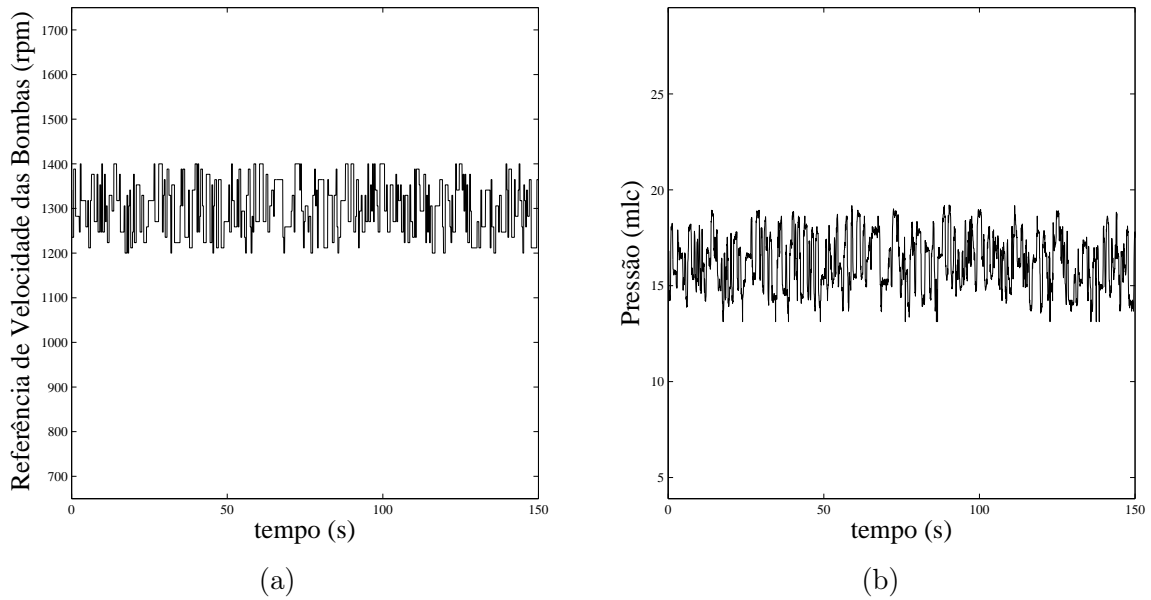


Figura 3.3: Dados de entrada (a) e saída (b) obtidos nos experimentos dinâmicos, utilizados na identificação do sistema hidráulico.

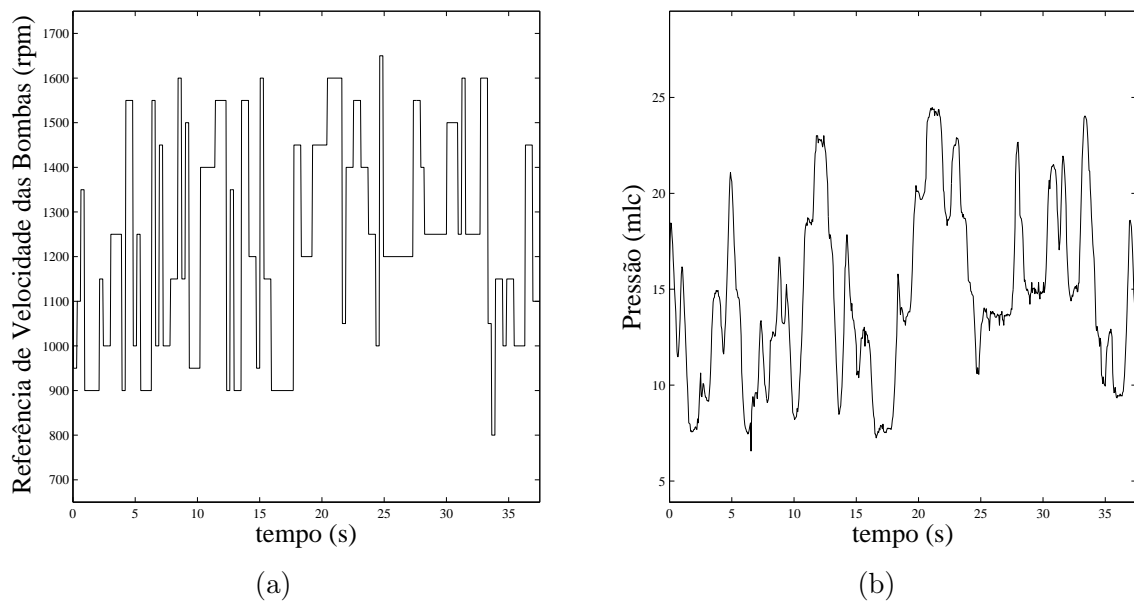


Figura 3.4: Dados de entrada (a) e saída (b) obtidos nos experimentos dinâmicos, utilizados para validação dos modelos.

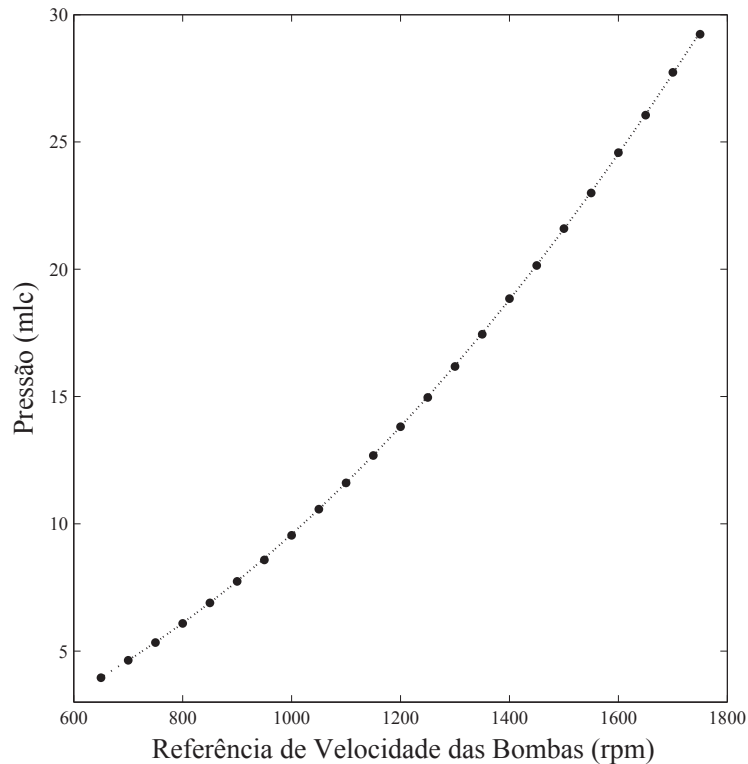


Figura 3.5: Curva estática medida do sistema (●) e aproximação polinomial de segundo grau (⋯), conforme Equação 3.9.

Nesse experimento, a referência de velocidade da bomba foi mantida em um valor fixo até o sistema atingir o regime permanente. Uma vez em regime permanente, foram amostrados 240 valores de pressão e sua média foi registrada juntamente com o valor de entrada de referência de velocidade. Cada procedimento desse corresponde a um ponto na curva estática, cuja abscissa é a variável de entrada (referência de rotação das bombas) e a ordenada, a saída (valor da pressão a jusante das bombas). Utilizando esse procedimento, foram gerados $M = 23$ pontos da curva estática com valores de entrada igualmente espaçados entre 650 e 1750 rpm. A curva estática obtida é mostrada na Figura 3.5.

Além dos pontos medidos da curva estática, foi realizada uma aproximação polinomial de segundo grau, conforme a seguir:

$$H(\bar{u}) = \beta \bar{u}^2 + \alpha \bar{u} + \kappa. \quad (3.9)$$

Na Figura 3.5 também é mostrada a curva obtida pela regressão polinomial de segunda ordem, conforme Equação 3.9, com $\beta = 9,3514 \times 10^{-6}$, $\alpha = 6,0492 \times 10^{-4}$ e $\kappa = 0,3825$. Pela figura, verifica-se a característica não linear da curva estática.

Nota-se que, nesta aplicação, a curva estática foi obtida por medições feitas no sistema real, em toda a faixa de operação. No entanto, para o procedimento de modelagem caixa cinza aqui apresentado, nada impediria que essa curva estática fosse obtida por outros meios, tais como equações de processo ou informações empíricas, como é o caso da aplicação mostrada na Seção 3.3.2.

Modelos

O primeiro grande desafio é a escolha da topologia da rede. Como o objetivo deste trabalho é comparar o desempenho de modelos neurais caixa preta e caixa cinza, decidiu-se escolher uma estrutura previamente utilizada em [Barbosa et al., 2011]:

$$\hat{y} = \theta_0 + \sum_{i=1}^7 \theta_i \tanh \left(\sum_{j=1}^6 \theta_{i,j} y(k-j) + \sum_{j=0}^6 \theta_{i,(j+7)} u(k-j) + \sum_{j=1}^2 \theta_{i,(j+13)} e(k-j) \right),$$

em que a parte média móvel ($e(k-1)$ e $e(k-2)$) do modelo foi considerada somente no treinamento caixa preta. No treinamento bi-objetivo, devido ao fato de estar otimizando a função J_S , que utiliza a simulação livre, foi utilizada apenas a parte autorregressiva com entrada exógena (NARX) do modelo.

Como o algoritmo de treinamento possui características estocásticas, foram realizados 100 treinamentos e foi escolhido o modelo que deteve o menor erro de aproximação nos dados de teste.

Resultados

A modelagem caixa cinza produziu o conjunto Pareto mostrado na Fig. 3.6, que também mostra o desempenho do modelo caixa preta. Na figura é mostrado o percentual da raiz do erro médio quadrático (PRMSE) em relação ao fundo de escala do instrumento, igual a 70 mlc. Verifica-se que a solução do Pareto com o menor erro de simulação livre (\circ 4, 839) domina a solução caixa preta, ou seja, o modelo caixa preta estimado não faz parte do conjunto Pareto. Isso ocorre devido ao fato de que o ajuste dos parâmetros feito no modelo caixa preta utiliza do erro de estimação um passo à frente como critério de minimização, sendo que, na Fig. 3.6, é apresentado o erro de simulação livre.

Alguns modelos obtidos por identificação caixa cinza têm melhores desempenhos de validação em faixa ampla. À medida que a solução começa a considerar mais a curva estática, observa-se uma melhoria do desempenho *dinâmico* em faixa ampla. Porém, a partir de certo ponto, quando esse erro sobre a curva estática fica bastante reduzido pelo otimizador, os modelos começam a ter desempenho muito inferior sobre os dados dinâmicos de validação.

Para as discussões que seguem, serão comparados três modelos: \mathcal{M}_p (modelo obtido por identificação caixa preta - indicado na Fig. 3.6 por \square 4, 842), \mathcal{M}_{c_1} (modelo obtido por identificação caixa cinza que obteve o melhor desempenho dinâmico em faixa ampla - indicado na Fig. 3.6 por \circ 4, 797) e \mathcal{M}_{c_2} (modelo obtido por identificação caixa cinza que obteve o melhor desempenho sobre a curva estática - indicado na Fig. 3.6 por \circ 5, 218).

A Fig. 3.7 mostra a curva estática dos modelos \mathcal{M}_p , \mathcal{M}_{c_1} e \mathcal{M}_{c_2} . Pode-se observar que a curva é bem ajustada por todos os modelos para a faixa de entrada entre 1200 e 1400 rpm, mesma região de operação dos dados de identificação. Nas demais regiões (i.e. < 1200 e > 1400 rpm), ocorrem erros de aproximação da curva. É possível notar que o modelo caixa preta (\mathcal{M}_p) obteve o pior desempenho, especialmente na faixa acima de 1500 rpm. Nota-se que os modelos com ajuste bi-objetivo se aproximam mais da curva estática real.

A Figura 3.8 mostra um trecho da simulação dos três modelos para os dados de validação. É possível verificar que, em geral, o desempenho dinâmico (transitório), fora da faixa em que o sistema foi identificado, apresenta comportamento bem diferente do sistema real. Por outro lado, o transitório compreendido entre 19 e 20s, dentro da faixa em que o sistema foi identificado, apresenta bom desempenho. É possível verificar que o desempenho do modelo \mathcal{M}_{c_2} se mostra muito oscilatório, com características dinâmicas

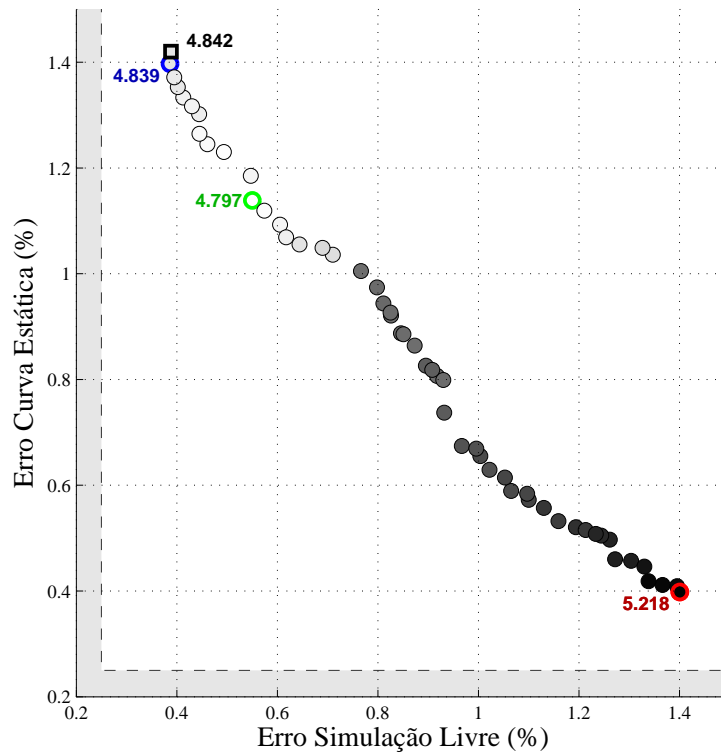


Figura 3.6: Curva Pareto obtida do ajuste multiobjetivo dos parâmetros da rede neural (Eq. 3.10). A cor de preenchimento dos pontos representa o erro (PRMSE) sobre o teste dinâmico em faixa ampla: quanto mais próximo da cor branca, menor o erro. Estão marcados, em azul (\circ 4,839) e vermelho (\circ 5,218), os modelos caixa cinza com melhor desempenho em cada um dos objetivos. Em verde (\circ 4,797), é marcado o modelo que apresentou o melhor desempenho no teste dinâmico em faixa ampla. O quadrado (\square 4,842) representa o modelo caixa preta. Os valores escritos nos quatro pontos marcados referem-se ao PRMSE sobre os dados dinâmicos em faixa ampla. A região cinza, contornada pela linha tracejada (—), representa a região em que o erro pode ser considerado nulo, inferior à precisão do instrumento ($< 0,25\%$).

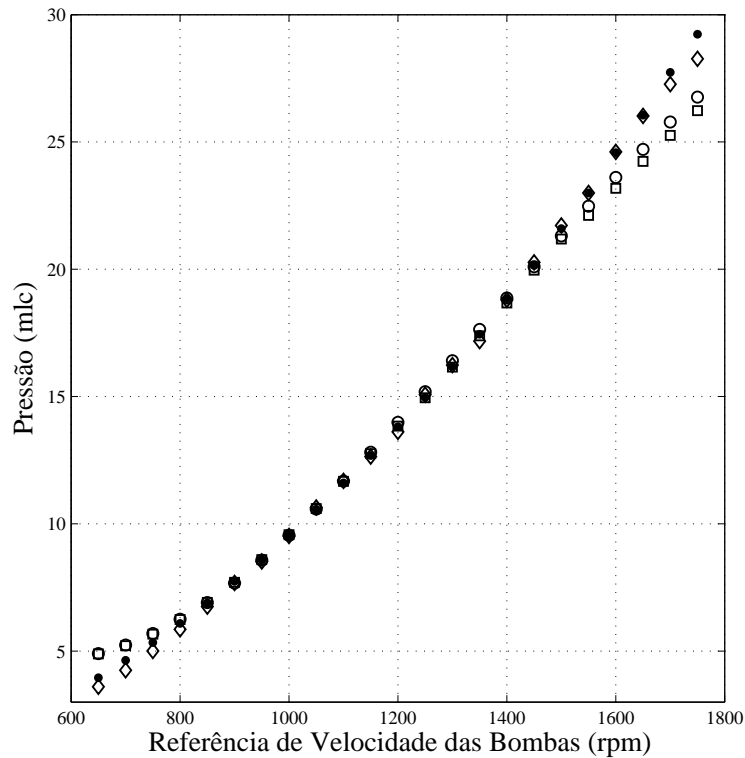


Figura 3.7: Curva estática de (\square) \mathcal{M}_p , (\circ) \mathcal{M}_{c_1} , (\diamond) \mathcal{M}_{c_2} e (\bullet) medições.

bem diferentes das apresentadas pelo sistema real. Por outro lado, o modelo caixa cinza \mathcal{M}_{c_1} , apesar de ter um comportamento dinâmico semelhante ao modelo \mathcal{M}_p (em azul), verifica-se que no patamar mais alto, em torno de 24 mlc, fora da faixa de operação dos dados de identificação, houve uma redução do *offset* na pressão, mostrando que sua função estática está mais adequada ao sistema real.

De fato, a informação adicional implementada na identificação caixa cinza traz consigo informações sobre o comportamento estático do sistema. A implementação do método tende a reduzir erros de *offset* em regiões conhecidas apenas pela curva estática do sistema. Não há informação alguma com relação ao comportamento dinâmico do sistema em outras regiões de operação.

Outro ponto de interesse na análise dos resultados do procedimento bi-objetivo é o fato de que o modelo que pior se ajustou à curva estática obteve o melhor desempenho nos ensaios dinâmicos, e vice-versa. Isso sugere a existência de dois objetivos conflitantes: o ajuste da curva estática vs. ajuste do comportamento dinâmico do sistema. Esses objetivos não são conflitantes *per se*, uma vez que ambos coexistem no sistema sendo identificado. O ponto central, que justifica o procedimento bi-objetivo, é que a informação dinâmica e estática nem sempre são igualmente balanceadas em um *único* conjunto de dados [Barbosa et al., 2011]. Nesse contexto, a abordagem caixa cinza ajuda a balancear ambas informações a fim de obter um modelo parcimonioso, com bom desempenho estático e dinâmico.

A Figura 3.9 mostra o desempenho dos modelos no espaço do erro dinâmico em faixa ampla (validação) e erro da curva estática. É possível verificar que, para uma faixa maior de operação, a modelagem caixa cinza pode encontrar soluções que dominam o modelo caixa preta. É interessante notar ainda que o modelo caixa preta não domina nenhuma solução. Ou seja, utilizando a abordagem caixa cinza, para este caso, haverá sempre uma melhoria de desempenho, seja no comportamento estático, no dinâmico ou em ambos.

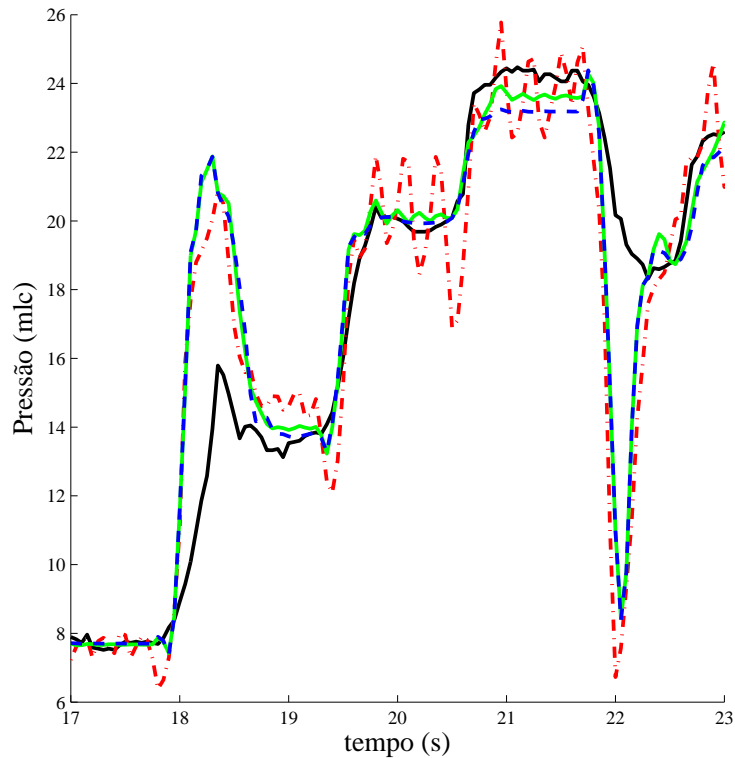


Figura 3.8: Trecho do sinal de validação que mostra a simulação livre dos modelos ($---$) \mathcal{M}_p , ($-$) \mathcal{M}_{c_1} e ($- \cdot -$) \mathcal{M}_{c_2} . A linha preta contínua ($—$) é a medição.

Tabela 3.1: Desempenho de modelos identificados para o sistema de bombeamento.

Modelo	PRMSE (%)			
	$J_S(\text{id.})$	$J_S(\text{teste})$	$J_S(\text{val.})$	J_{SF}
\mathcal{M}_p	0,3885	0,5548	4,8423	1,4200
\mathcal{M}_{c_1}	0,5500	0,8325	4,7971	1,1387
\mathcal{M}_{c_2}	1,4071	1,5603	5,2178	0,3987

Comparação entre resultados

A Tabela 3.1 mostra o resultado comparativo entre os modelos apresentados, para todas as massas de dados utilizadas. Nesses resultados, é possível verificar que a rede neural caixa preta (\mathcal{M}_p) é a que mais se ajusta aos dados de identificação e obtém o melhor resultado, entre todos os modelos, para os dados de teste. Numa tarefa comum de modelagem, em que a faixa de operação estaria restrita a uma região menor (como nos dados de identificação e teste), poderia dar a falsa impressão de que o modelo neural caixa preta seria o mais adequado. Porém, pelo desempenho nos dados de validação em faixa ampla, verifica-se que o modelo neural caixa preta não detém o melhor resultado.

Ao comparar a rede neural caixa preta (\mathcal{M}_p) com a rede caixa cinza (\mathcal{M}_{c_1}), verifica-se que, ao adicionar a informação da curva estática no modelo neural, o modelo pode responder melhor para dados não conhecidos. Em outras palavras, com o uso da modelagem caixa cinza, é possível encontrar modelos com melhor desempenho, tanto dinâmico (global) quanto estático. No presente exemplo, essa melhora não é dramática, pois mesmo no conjunto de dados dinâmicos de identificação Z e de teste, a informação

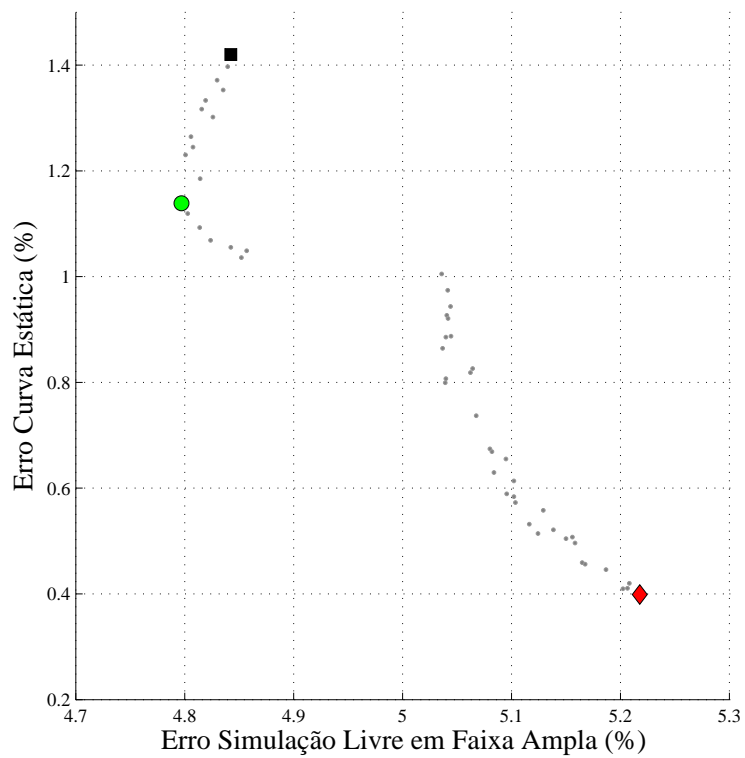


Figura 3.9: Avaliação do desempenho dinâmico em faixa ampla (validação) e da curva estática para os modelos (■) \mathcal{M}_p , (●) \mathcal{M}_{c_1} , (◆) \mathcal{M}_{c_2} . Os pontos indicados em cinza (·) mostram as demais soluções da curva Pareto na modelagem caixa cinza.

do comportamento estático do sistema está razoavelmente bem representada, fazendo com que a adição da informação auxiliar do comportamento estático global não gere ganhos expressivos. Esse não é o caso do próximo exemplo, que servirá para ilustrar uma situação em que a vantagem de se usar o procedimento caixa cinza é maior.

3.3.2 O Sistema de Extração de Petróleo

O processo de extração de petróleo em poços não surgentes consiste, basicamente, na injeção de gás (*gas-lift*), que forma uma mistura gás-óleo, com densidade menor, e faz com que o óleo suba até a superfície. Nesse contexto, a identificação de sistemas dinâmicos se aplica à estimação da pressão de fundo de poço (PDG-P), que é um importante fator na determinação da produtividade e tempo de vida do poço de petróleo. Porém, essa informação nem sempre está disponível, devido às más condições do ambiente de operação desse sensor e ao alto custo relacionado à troca e à manutenção do dispositivo. Esses fatores motivam a estimação dessa pressão de fundo de poço a partir de medições feitas na superfície. Maiores informações sobre o processo são fornecidas no Capítulo 2.

Modelos

Para realizar a estimação da pressão de fundo de poço são utilizadas medições disponíveis na plataforma, quais sejam: vazão de injeção de *gas-lift*, pressão na linha de injeção de *gas-lift*, pressão à montante da *choke* de produção, temperatura na linha de produção e pressão à montante de SDV de produção.

As medições foram amostradas a uma taxa $T_s = 1\text{min}$ e, com a adição de alguns termos de atraso, servem de entrada para a rede neural. Foi utilizada uma estrutura NARMAX (os termos MA foram utilizados apenas na abordagem caixa preta), com uma camada escondida com 10 neurônios, função de ativação $\tanh()$ e função linear na camada de saída. A saída da rede representa a pressão PDG-P.

Os procedimentos de treinamento foram os mesmos apresentados na Seção 3.2. Para a obtenção do modelo caixa preta foram treinadas 100 redes utilizando o algoritmo de Levenberg-Marquadt, implementado no pacote *Norgaard Toolbox* [Norgaard, 1997]. Para fins de comparação, foi escolhida a rede neural caixa preta com o melhor desempenho em dados de teste.

Para a identificação do sistema, foram coletadas duas janelas de dados: identificação ($N = 5000$) e validação ($N = 95000$). A curva estática do processo foi obtida a partir de regressão polinomial de segundo grau baseada em pontos de operação conhecidos previamente. Foram selecionadas, manualmente, em várias massas de dados, regiões de operação em que o sistema aparenta estar em estado estacionário. Para cada região selecionada, foram tomadas as médias dos valores de entrada e saída, obtendo assim os pontos de operação conhecidos do sistema. A partir desses pontos de operação e baseado no conhecimento prévio (fornecido por comunicação pessoal com engenheiros de processo da Petrobras) de que a curva estática tem características de uma função de segundo grau (parábola), foi feita a interpolação desses pontos e, assim, obtida a curva estática do sistema de extração de petróleo. Apesar do modelo ter cinco medições na entrada da rede, a curva estática foi obtida em duas dimensões apenas (pressão de fundo PDG-P vs vazão de *gas-lift*), considerando as demais medições constantes.

Resultados

A Figura 3.10 mostra o desempenho da rede neural caixa preta \mathcal{M}_p para os dados de validação. É possível verificar, a menos do primeiro estreito patamar, um *offset* significativo entre os dados medidos

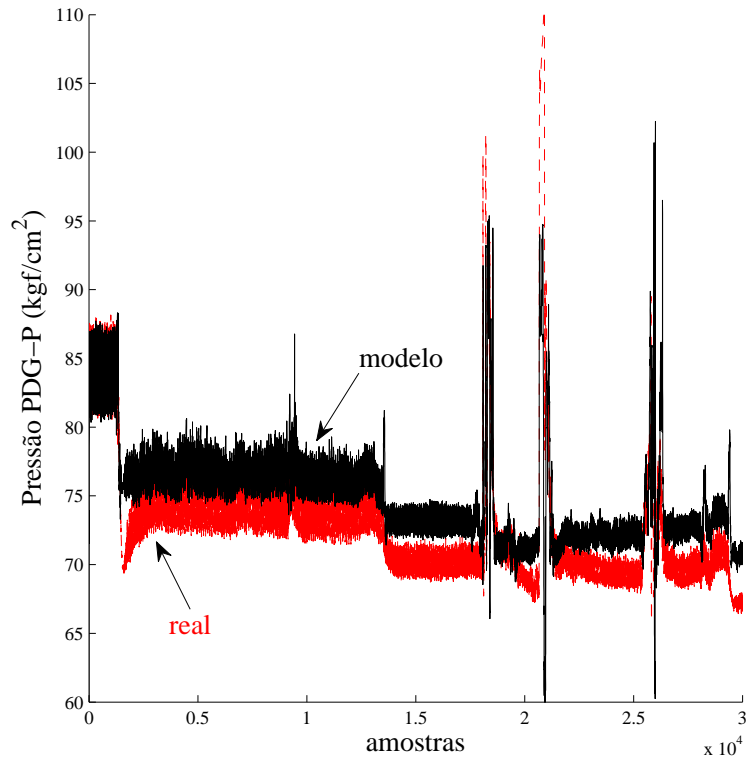


Figura 3.10: Desempenho do modelo dinâmico \mathcal{M}_p (caixa preta) sobre os dados de validação do sistema de extração de petróleo. Os dados estimados e medidos são apresentados, respectivamente, em preto e vermelho.

e estimados. Nesse caso, é esperado que a identificação caixa cinza reduza essa diferença com a inserção da informação sobre o comportamento estático do sistema.

Assim, utilizando a abordagem caixa cinza, produziu-se o conjunto Pareto de soluções mostrado na Fig. 3.11, em que é mostrado também o desempenho do modelo \mathcal{M}_p . Nota-se que o conjunto Pareto contém soluções com desempenho melhor em relação a curva estática. É possível verificar também, pela intensidade das cores, que o desempenho sobre os dados de validação do modelo \mathcal{M}_p é um dos piores, comparado ao conjunto de soluções caixa cinza. O modelo \mathcal{M}_c (\circ 2, 751) apresenta desempenho bem superior ao modelo \mathcal{M}_p (\blacksquare 3, 498).

As curvas estáticas dos modelos \mathcal{M}_p e \mathcal{M}_c são mostradas na Fig. 3.12. Verifica-se que a curva estática desejada tem a forma de uma parábola, o que torna difícil o bom ajuste de modelos que utilizam um conjunto de dados com faixa de operação limitada que, nesse caso, está compreendida entre 2600 e 4300m³/h (região em que a curva estática de \mathcal{M}_p se aproxima mais do desejado). Porém, especialmente na faixa entre 5000 e 10000m³/h, verifica-se o mau ajuste do modelo \mathcal{M}_p , o que explica o *offset* nos dados de validação (Fig. 3.10). O modelo \mathcal{M}_c , por outro lado, tem, em um dos objetivos, o compromisso de se ajustar a essa curva estática, apresentando uma aproximação melhor em toda a região de operação.

A aproximação melhor da curva estática do modelo \mathcal{M}_c resulta numa redução do erro sobre os dados de validação, como pode ser observado na Figura 3.13. Vale ressaltar que não é observada melhoria no comportamento dinâmico, e nem era de se esperar, pois a informação adicional inserida contém apenas informação sobre o comportamento estático do sistema.

A Figura 3.14 mostra o desempenho dos modelos no espaço do erro dinâmico de validação e erro sobre

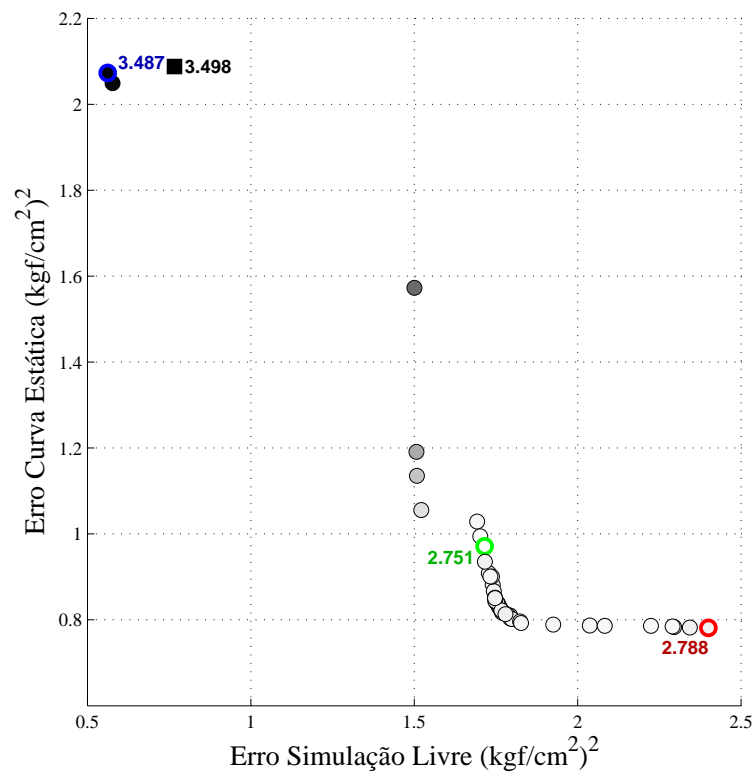


Figura 3.11: Conjunto Pareto de soluções utilizando a abordagem caixa cinza. É mostrado também o desempenho do (■) modelo caixa preta \mathcal{M}_p . É destacado em verde (○ 2,751) um dos modelos obtidos no procedimento caixa cinza: \mathcal{M}_c . Os números em destaque mostram o desempenho (RMSE) sobre os dados de validação. A intensidade das cores de preenchimento dos círculos também ilustram o valor do RMSE sobre os dados de validação: quanto mais clara a cor, menor o erro sobre os dados de validação.

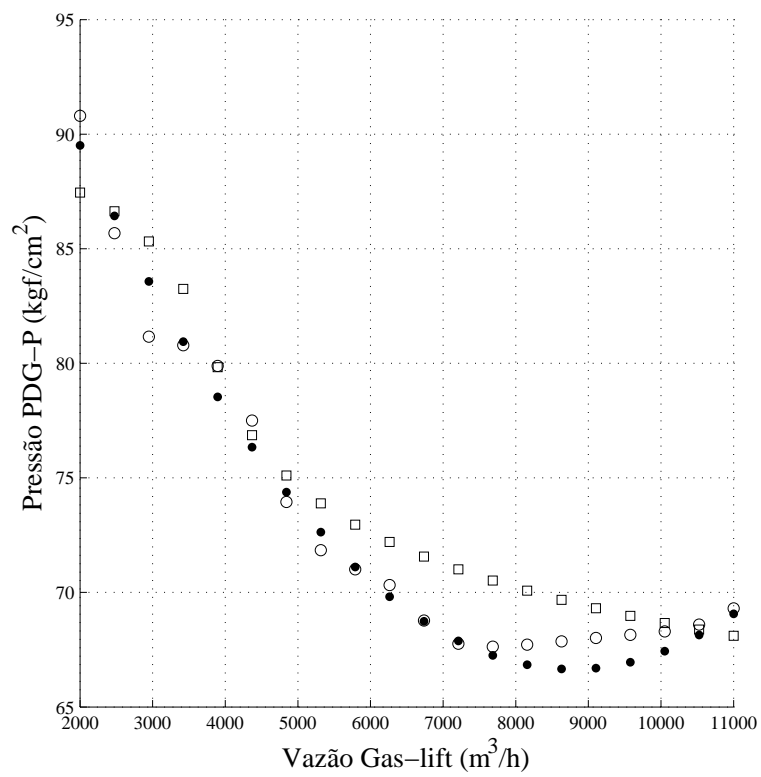


Figura 3.12: Curva estática dos modelos dinâmicos (□) \mathcal{M}_p , (○) \mathcal{M}_c e (●) a curva estática desejada, obtida experimentalmente.

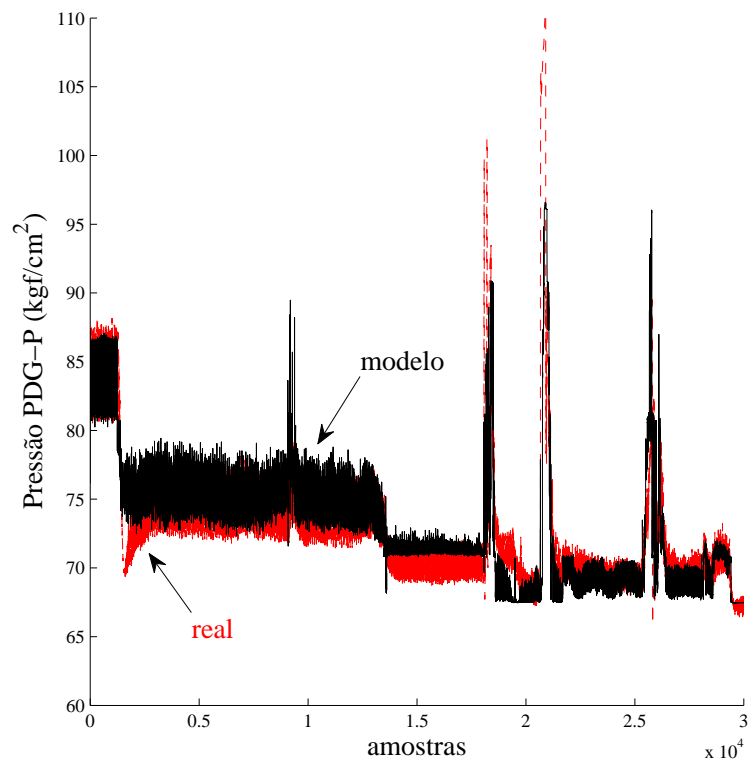


Figura 3.13: Desempenho do modelo dinâmico \mathcal{M}_c (caixa cinza) sobre os dados de validação do sistema de extração de petróleo. Os dados estimados e medidos são apresentados, respectivamente, em preto e vermelho.

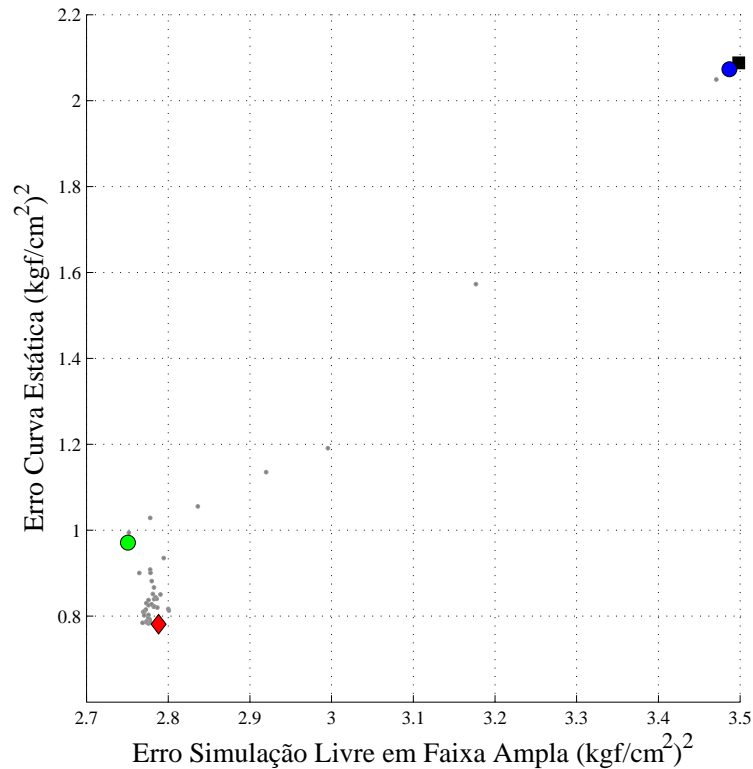


Figura 3.14: Conjunto Pareto de soluções mostrados no espaço do erro (RMSE) dinâmico de validação e erro (RMSE) sobre a curva estática. É mostrado também o desempenho dos modelos (■) \mathcal{M}_p e (●) \mathcal{M}_c .

a curva estática. É possível verificar que o modelo \mathcal{M}_p é dominado por todas as soluções do conjunto Pareto. Seja qual for o decisor, para este caso, a solução caixa cinza, para os dados de validação, sempre apresentará melhor desempenho estático e dinâmico sobre o modelo caixa preta. Esse resultado pode ser explicado, em parte, pela característica fortemente não linear da curva estática, sendo que os dados de validação ocupam exatamente a região em torno do ponto mais baixo da parábola na curva estática, região onde o ganho do sistema sofre mais alterações com pequenas variações no ponto de operação.

3.3.3 Discussão dos resultados experimentais

A abordagem caixa cinza mostrou-se uma alternativa interessante para o caso do sistema de extração de petróleo, melhorando significativamente o desempenho dos modelos neurais. Já no sistema de bombeamento, as melhorias não foram tão expressivas. Essa diferença pode ser explicada por diversos fatores. Comparando as curvas estáticas, por exemplo, é possível verificar que o sistema de extração de petróleo opera em uma região que dificilmente pode ser aproximada por uma reta, enquanto que, no sistema de bombeamento, essa aproximação é mais fácil.

Outro fator relevante pode estar relacionado às características dos dados de identificação. Os testes no sistema de bombeamento foram realizados em laboratório, em boas condições de ensaio, e os sinais de identificação foram escolhidos de forma a manter o sistema persistentemente excitado (dentro da faixa limitada de operação). Por outro lado, no sistema de extração de petróleo, foi escolhida uma janela de identificação contendo apenas um degrau (com oscilações). É claro que um sinal persistentemente

excitante é muito mais rico em informação, estática e dinâmica, que um único degrau.

Assim, como procedimento de identificação caixa cinza, em geral, sugere-se que, primeiro, seja feito o levantamento da curva estática do modelo caixa preta e do sistema real. Caso a diferença entre as curvas seja expressiva, sugere-se a utilização da metodologia proposta.

3.4 Conclusões do Capítulo

Esse capítulo descreveu um procedimento de identificação caixa cinza usando redes neurais, cuja informação auxiliar é a curva estática do sistema. A informação auxiliar é adicionada na fase de treinamento, em que é empregada a otimização multiobjetivo em duas métricas de erro: raiz do erro quadrático médio (RMSE) de simulação livre sobre dados dinâmicos e RMSE da curva estática. O procedimento é aplicável a qualquer sistema que disponha de dados dinâmicos de treinamento e da curva estática do sistema. Foram mostradas duas aplicações do procedimento em sistemas reais: um sistema hidráulico de bombeamento (10cv) e um sistema de extração de petróleo em águas profundas.

O sistema hidráulico mostrou o desempenho do procedimento em um sistema relativamente bem comportado quanto à ruídos e com dados dinâmicos bem representativos. Nesse caso, foi mostrado que o procedimento não gera ganhos significativos, já que a informação do comportamento estático está bem representada nos dados dinâmicos de treinamento. Sistemas cuja curva estática não adiciona muita informação ao comportamento do processo podem não obter ganhos com a aplicação do procedimento.

O sistema de extração de petróleo em águas profundas, por outro lado, apresentou ganhos expressivos quando adicionada a informação da curva estática usando o método proposto. Essa aplicação mostrou que o procedimento pode gerar melhoria significativa de desempenho quando as características estáticas não estão bem representadas nos dados.

O procedimento mostrou uma possibilidade de adicionar informação extra em modelos neurais. Mesmo sendo uma ferramenta de uso genérico, foi mostrado que as redes neurais podem incorporar informações específicas de um sistema, sem modificações na estrutura do modelo. Foi mostrado que essa “customização” de uma ferramenta tão geral pode trazer ganhos significativos. Compreender a estrutura do modelo neural e interferir no processo de aprendizagem para se adaptar melhor a cada uso das redes neurais pode ser um caminho promissor.

Diversidade em Comitês de Modelos Dinâmicos

4.1 Introdução

Com a evolução das técnicas de modelagem e identificação de sistemas, hoje é possível obter modelos dinâmicos de várias maneiras distintas, utilizando das mais variadas técnicas e procedimentos, tornando-se, às vezes, difícil optar por um modelo ou outro. A estratégia de combinar vários modelos pode ser uma solução para este caso, em que vários modelos são utilizados para a formação de um modelo maior, uma *máquina de comitê*. Um comitê pode utilizar-se de uma arquitetura *modular* (e.g. mistura de especialistas), em que cada componente soluciona parte do problema e a solução final é dada pela mistura desses; ou uma arquitetura paralela (e.g. *ensemble*), em que todos os componentes tentam gerar, concorrentemente, uma solução completa para o problema em questão e a combinação destes gera uma resposta mais adequada ao problema. É possível ainda que ambas as abordagens sejam utilizadas num problema, contudo, este trabalho se restringe aos *ensembles* de modelos dinâmicos, numa arquitetura unicamente paralela. Na literatura, *ensembles* são utilizados, principalmente, com a justificativa de aprimorar a capacidade de generalização do modelo [Krogh and Vedelsby, 1995, Sharkey, 1996] e reduzir o sobre-ajuste (*overfitting*) aos dados [Sollich and Krogh, 1995]. A combinação de modelos, por sua vez, traz consigo alguns questionamentos: vale a pena combinar todos os modelos obtidos? Como selecionar os modelos? Qual a melhor forma de combiná-los?

4.1.1 Contextualização

Para iniciar a discussão, é importante salientar que essa ideia de combinar a resposta de componentes, que resolvem um mesmo problema, não é nova. A natureza utiliza, por exemplo, na reprodução das espécies, um combinador – aleatório talvez – do código genético de dois indivíduos para formar um terceiro, que pode ser mais adaptado ao meio do que cada um dos indivíduos que o gerou. Em sistemas governamentais ou judiciais é comum que decisões importantes sejam tomadas por conselhos ou comitês, que combinam várias opiniões e apresentam uma única resposta ao problema. Em engenharia, essa ideia também não é nova, sendo que alguns registros datam de 1931, em que sistemas receptores são combinados de forma a recuperar melhor um sinal de comunicação [Beverage and Peterson, 1931]. Em identificação de sistemas, trabalhos combinando preditores foram realizados por volta da década de 60 [Buell, 1958, Graybill and Deal, 1959, Brennan, 1959, Nilsson, 1965, Ablow and Kaylor, 1965, Seshadri, 1966, Bates and Granger, 1969], mas foi a partir da década de 90 que os trabalhos nessa área se intensificaram [Barbosa, 2009] e, no início da década, vários trabalhos justificaram o uso de *ensembles* a partir de resultados empíricos

[Hansen and Salamon, 1990, Wolpert, 1992, Perrone, 1993, Drucker et al., 1994, Hashem, 1994, Hashem and Schmeiser, 1995]. [Krogh and Vedelsby, 1995] formalizaram melhor a justificativa de utilizar *ensembles* em problemas de regressão em que, por meio da decomposição de ambiguidade, foi provado que o erro quadrático do *ensemble* é, garantidamente, menor ou igual à média dos erros médios quadráticos de seus componentes [Krogh and Vedelsby, 1995]. Desde então, vários trabalhos envolvendo *ensembles* foram desenvolvidos, muitos deles utilizando redes neurais aplicadas em problemas de classificação.

Um ponto comum para se obter uma boa combinação de membros, seja na evolução das espécies ou nos conselhos governamentais, é o princípio da *diversidade*. Para que a evolução seja bem sucedida, por exemplo, a biodiversidade pode ser um fator determinante para que seres vivos prosperem nas próximas gerações após uma grande calamidade. A diversidade de membros de um comitê também é uma forma de garantir que as decisões do grupo sejam satisfatórias para a maioria dos representados. Na engenharia não é diferente. O trabalho de [Beverage and Peterson, 1931] já referenciava o termo como uma importante característica para o sistema de recepção de sinais. [Hansen and Salamon, 1990], apesar de não fazerem referência explícita ao termo “diversidade”, apresentaram um método que baseia-se no *N-version programming*¹, cuja diversidade já era um princípio conhecido nesse contexto [Avizienis, 1984]. Trabalhos mais recentes mostram que o uso do conceito de diversidade nos *ensembles* pode melhorar muito seu desempenho [Schapire, 1990, Croll et al., 1995, Ueda and Nakano, 1996, Opitz and Shavlik, 1996, Rosen, 1996, Sharkey and Sharkey, 1997, Freund and Schapire, 1997, Liu and Yao, 1999, Melville and Mooney, 2003, Canuto et al., 2007]. Esses e vários outros trabalhos não citados aqui apontam que o conceito de *diversidade* pode ser a chave para responder as perguntas do início deste capítulo.

4.1.2 Metodologia Proposta

Este capítulo dedica-se ao estudo do conceito de diversidade e à sua aplicação na combinação de modelos, aplicados ao problema de identificação de sistemas. É objetivo que se desenvolva uma metodologia para a medição da diversidade de modelos e sua utilização, direta ou indireta, na formação de um *ensemble*. Espera-se que também seja possível verificar a influência da diversidade no desempenho do *ensemble*.

Inicialmente, é feito um breve estudo do conceito de diversidade, suas aplicações na identificação de sistemas e suas formas de cálculo. Em seguida, são feitos testes preliminares acerca das formas de cálculo da diversidade em modelos dinâmicos. A partir dos resultados dos testes preliminares, são realizados testes em sistemas experimentais e os resultados são avaliados.

4.2 Sobre Diversidade

Conforme já dito, diversidade é uma preocupação recorrente quando o assunto é combinação de componentes, em várias áreas do conhecimento. [Beverage and Peterson, 1931] definiram diversidade como o princípio pelo qual vantagem é obtida pelo fato de que erros não ocorrem simultaneamente em dois sistemas receptores. [Avizienis, 1984], em seu trabalho de engenharia de software, cunhou o termo *design diversity* para referenciar ao desenvolvimento de softwares (ou hardwares), com a mesma finalidade, mas de forma independente, mais tolerante a falhas. Segundo [Sharkey, 1996], o termo provém da ideia de combinar programas, em paralelo, que possam falhar, mas de forma independente (ou não correlacionada). Quando no contexto de redes neurais, a diversidade remete à ideia do erro de um componente

¹Método utilizado em engenharia de software, em que múltiplos programas são desenvolvidos, de forma independente, mas com as mesmas funcionalidades.

ser compensado pelo acerto dos demais. Para clarear essa ideia, uma ilustração feita por [Sharkey and Sharkey, 1997] aplicada à aproximação de funções, em forma de diagramas de Venn, é reproduzida na Figura 4.1, mas com alterações acerca da interpretação da figura que, nesse trabalho, é feita no âmbito de identificação de sistemas dinâmicos.

Considere a Figura 4.1. Seja \mathcal{D} o conjunto universo, fictício, que contém todos os comportamentos dinâmicos possíveis. Considere um sistema dinâmico que pode ser completamente descrito pela função $f(\cdot)$ que, no conjunto universo \mathcal{D} , abrange o subconjunto F de “comportamentos dinâmicos”. Um modelo ideal desse sistema deveria, portanto, possuir o mesmo tamanho e ocupar a mesma posição de F no conjunto \mathcal{D} . Mas, para a obtenção do modelo, é disponibilizado um conjunto de dados dinâmicos de treinamento que representa bem o conjunto Z de comportamentos dinâmicos. Verifica-se que o conjunto Z é muito menor que F ($Z \subseteq F$), mostrando a situação, comum, de que os dados utilizados na identificação não representam bem todo o comportamento dinâmico do sistema. Assume-se que é disponibilizada também uma massa de dados de teste, que representa bem os comportamentos dinâmicos contidos no conjunto T . De forma semelhante a Z , $T \subseteq F$. Essa figura ilustra bem o cenário de identificação de sistemas, em que os dados de identificação e teste não representam bem todo o comportamento dinâmico do sistema.

Seja $P = \{P_1, P_2, \dots\}$ o conjunto de todos modelos identificados a partir dos dados de treinamento Z e teste T , supondo ajuste perfeito a esses dados: $(Z \cup T) \subset P_i \quad \forall i \in \mathbb{N}^*$; e que os modelos em P têm a mesma complexidade de F , i.e. têm o mesmo número de elementos: $n(P_i) = n(F) \quad \forall i \in \mathbb{N}^*$, em que $n(X)$ representa o número de elementos no conjunto X . Na Fig. 4.1, isso significa que todos os modelos contidos em P abrangem os conjuntos Z e T , e possuem o mesmo tamanho de F . Nessa interpretação, a “complexidade” está associada à quantidade de “comportamentos dinâmicos” abrangida pelo modelo, ou seja, ao tamanho dos conjuntos P_1, P_2, P_3, \dots . Observa-se que $F \subset P$, porém, como os dados usados no ajuste dos parâmetros não são representativos para todo o comportamento dinâmico de F (Z e $T \subseteq F$), há inúmeros conjuntos em P , que representam bem Z e T , mas não abrangem bem a região hachurada F . Modelos que representam bem essa região são modelos com boa *capacidade de generalização*, i.e. boa capacidade de extrapolar as informações contidas nos dados de treinamento, Z e T , e se ajustar a F .

Na Figura 4.1 é possível observar dois *ensembles*: $\{P_1, P_2, P_3\}$ e $\{P_4, P_5, P_6\}$. Individualmente, os modelos representados por P_4, P_5, P_6 parecem se ajustar melhor a F . Porém, quando considera-se o grupo (*ensemble*), verifica-se que $\{P_1, P_2, P_3\}$ pode explicar melhor F . O *ensemble* $\{P_1, P_2, P_3\}$ possui modelos com pouca informação em comum e é dito *diverso*. Idealmente, apenas as informações presentes nos dados de ajuste Z deveriam ser comuns aos componentes do *ensemble*. Por outro lado, os modelos com baixa diversidade $\{P_4, P_5, P_6\}$ têm grande interseção. Pode-se dizer que a diversidade está associada com a diferença de comportamento dinâmico entre os componentes. Medidas de diversidade (vide Seção 4.2.1) tentam, essencialmente, calcular o quão diferente é a resposta de um modelo em relação ao outro, dado um sinal de entrada². Porém, é importante frisar que essa medida também depende do sinal de entrada aplicado nos modelos. *Sinais que excitam apenas comportamentos dinâmicos presentes nos dados de treinamento, Z e T , não ajudam a mensurar a diversidade*. Na Fig. 4.1, verifica-se que todos os modelos possuem comportamentos dinâmicos dos conjuntos Z e T . Utilizando esses dados, o cálculo da diversidade indicaria, falsamente, que os *ensembles* $\{P_1, P_2, P_3\}$ e $\{P_4, P_5, P_6\}$ possuem baixa diversidade, já que os dados excitam apenas os comportamentos dinâmicos em comum a todos os modelos. Esse caso é uma

²Apesar de existir outras formas de calcular diversidade, neste trabalho, considera-se apenas as medidas que não fazem restrições quanto à estrutura ou classe dos modelos, que utilizam apenas dados de entrada e saída.

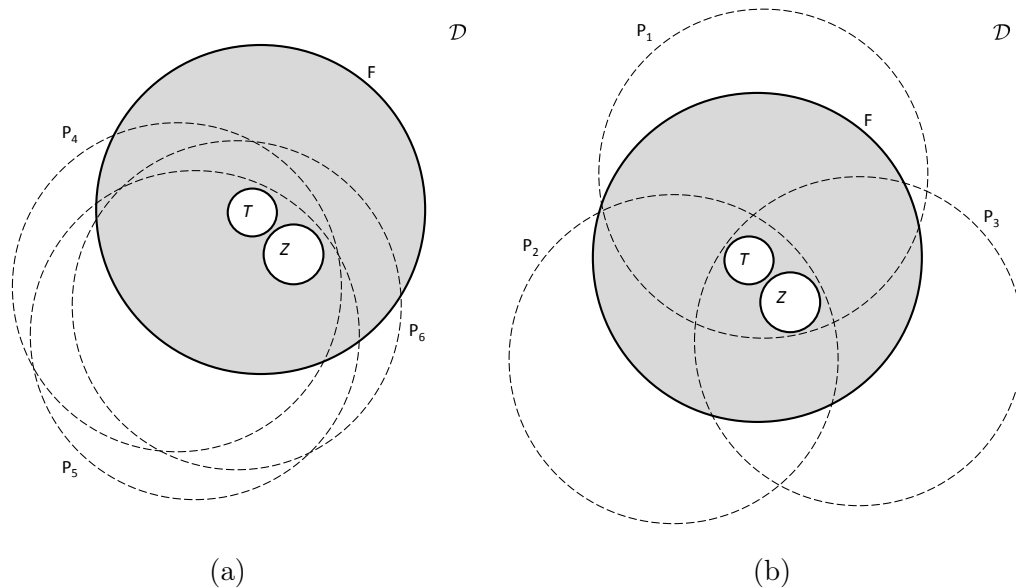


Figura 4.1: Diagramas de Venn ilustrando o papel da diversidade na identificação de sistemas. É apresentado um caso em que se deseja aproximar F , utilizando apenas as informações contidas em Z e T . É mostrado que um *ensemble* com alta diversidade (b) pode ajustar-se melhor à função original do que um *ensemble* com baixa diversidade (a).

típica situação de *consenso*, em que os componentes apresentam respostas com pouca variabilidade entre si e um baixo erro de aproximação da resposta do sistema real. Talvez uma das maiores dificuldades em se mensurar a diversidade de um *ensemble* seja distinguir o *consenso* de *baixa diversidade*.

Idealmente, seria interessante avaliar a diversidade extraindo toda a informação dinâmica dos modelos, com exceção das informações presentes nos dados de treinamento. No diagrama da Figura 4.1, isso representaria avaliar a diversidade no conjunto complemento de $Z \cup T$, ou seja, em $\mathcal{D} - \{Z \cup T\}$. Nota-se que os dados utilizados para a avaliação da diversidade devem ser mais representativos do que os dados utilizados no treinamento, mais ricos em comportamentos dinâmicos. Dados persistentemente excitantes, que exploram todos os possíveis comportamentos dinâmicos de um sistema, podem ser uma boa alternativa para avaliação da diversidade entre modelos dinâmicos.

Vale lembrar que, na Figura 4.1, assumiu-se que os modelos em P se ajustam perfeitamente à Z e T . Porém, na prática, não se tem a garantia de que todas as informações presentes nos dados de ajuste estejam nos modelos. Uma das consequências ao não atendimento à essa consideração seria uma redução no tamanho dos conjuntos Z e T no diagrama, que ilustraria apenas as informações contidas nos modelos. O caso de dados ruidosos, também não considerado na análise, faria com que Z e T tivessem dinâmicas não presentes em F , i.e. $\{Z \cup T\} \cap F \neq \emptyset$. Isso representa um desafio extra, uma vez que os modelos – especialmente em casos de *overfitting* – tenderiam a abranger também as regiões de Z e T que não pertencem a F (ruído). Nesses casos, a diversidade entre os modelos se torna um desafio ainda maior, tanto na medição quanto na obtenção.

Grandes regiões de interseção entre modelos representam baixa diversidade. Por outro lado, modelos com nenhuma região de interseção não se justificam na arquitetura paralela de *ensemble* propriamente dito. Talvez, nesse caso, seja mais adequada uma arquitetura “mistura de especialistas” (definição na Seção 4.1), em que cada modelo explica uma parte do sistema alvo. Tratando-se de *ensembles*, é impor-

tante também que haja regiões de interseção, em que mais de um modelo possa explicar determinadas dinâmicas e elevar a exatidão da resposta do *ensemble*.

Na próxima seção, é apresentado um estudo sobre medidas de diversidade, em que é feito um apanhado das principais formas de cálculo e suas interpretações.

4.2.1 Métricas de Diversidade

Conforme visto anteriormente, as medidas de diversidade tentam quantificar o quão diferentes são os modelos dinâmicos. Um procedimento usual utilizado é apresentar a dois modelos um conjunto de dados de entrada e, a partir das respostas dos modelos, são utilizadas métricas de comparação entre os sinais de saída. Nota-se que a escolha do conjunto de dados é um passo importante para esse cálculo de diversidade, que foi discutido na seção anterior. Por outro lado, o uso de métricas apropriadas para a comparação entre os sinais de saída dos modelos também se faz necessário. Algumas das principais métricas de diversidade utilizadas são discutidas nesta seção.

Os dados de entrada e saída utilizados nas medidas de diversidade são colocados conforme a seguir. Seja um conjunto de dados dinâmicos de entrada dado por $U = [u(1) u(2) \cdots u(N)]$, em que $u(k)$ representa o k -ésimo valor de um total de N amostras. Sejam dois modelos dinâmicos³ que, apresentados ao conjunto U , geram as saídas dadas por $Y_i = [y_i(1) y_i(2) \cdots y_i(N)]$ e $Y_j = [y_j(1) y_j(2) \cdots y_j(N)]$. As médias temporais desses valores são denotadas por μ_{Y_i} e μ_{Y_j} , respectivamente. Essa nomenclatura será utilizada nas medidas de diversidade a seguir.

Covariância e Índice de Correlação

[Ueda and Nakano, 1996], por meio da decomposição polarização-variância-covariância, mostraram que, considerando um *ensemble* cuja saída é a média simples das respostas dos componentes, o erro de generalização depende, também, da covariância entre os componentes. Essa covariância pode ser utilizada para medição da diversidade entre componentes e é definida como:

$$\text{cov}(Y_i, Y_j) = E[(Y_i - \mu_{Y_i})(Y_j - \mu_{Y_j})'], \quad (4.1)$$

em que $E[\]$ representa o operador esperança matemática. Um alto valor de covariância pode indicar diversidade entre modelos.

Outra medida utilizada, semelhante à covariância, é o índice de correlação, definido como:

$$\rho_{i,j} = \frac{\sum_{k=1}^N (y_i(k) - \mu_{Y_i})(y_j(k) - \mu_{Y_j})}{\sqrt{\sum_{k=1}^N (y_i(k) - \mu_{Y_i})^2 \sum_{k=1}^N (y_j(k) - \mu_{Y_j})^2}}. \quad (4.2)$$

Esse índice é uma medida normalizada entre -1 e 1 , em que $\rho_{i,j} = 1$ representa total correlação proporcional entre Y_i e Y_j e $\rho_{i,j} = -1$ indica total correlação inversamente proporcional. O valor $\rho_{i,j} = 0$ mostra que não há correlação linear entre os dados avaliados. Assim, o valor de $\rho_{i,j}$ próximo de zero pode indicar alta diversidade entre componentes. No entanto, alguns autores defendem a ideia de que o treinamento

³Note que, neste trabalho, não são feitas considerações sobre as condições iniciais, o que, provavelmente, invalida as análises para sistemas com alta sensibilidade às condições iniciais (e.g. sistemas caóticos).

de *ensembles* de redes neurais usando correlação negativa (e.g. *negative correlation learning*) é eficaz [Liu and Yao, 1999], o que é contestado por outros [Nguyen, 2006].

O Termo de Ambiguidade

[Krogh and Vedelsby, 1995] provaram, para um *ensemble* de média ponderada, que seu erro quadrático é menor que a média do erro quadrático médio de seus componentes. Para isso, foi utilizada a decomposição de ambiguidade, conforme a seguir:

$$E_e = \bar{E} - \bar{A}, \quad (4.3)$$

em que E_e representa o erro quadrático do *ensemble*:

$$E_e = (\hat{y}_e - y)^2. \quad (4.4)$$

\bar{E} a média ponderada do erro de generalização do *ensemble*:

$$\bar{E} = \sum_i w_i (\hat{y}_i - y)^2 \quad (4.5)$$

e \bar{A} o termo de ambiguidade:

$$\bar{A} = \sum_i w_i (\hat{y}_i - \hat{y}_e)^2, \quad (4.6)$$

sendo $\hat{y}_e = \sum_i w_i \hat{y}_i$ a saída do *ensemble*, \hat{y}_i a saída do i -ésimo componente, w_i o peso do i -ésimo componente, y a saída desejada. O termo de ambiguidade apresentado é usado em diversos trabalhos como medida de diversidade [Opitz and Shavlik, 1996, Brown, 2004, Brown et al., 2005b], em que um alto valor de ambiguidade pode indicar uma alta diversidade do *ensemble*. É válido notar que essa medida é aplicável apenas em combinadores lineares (e.g. média ponderada) e o cálculo é feito *a posteriori*, ou seja, após a inclusão do i -ésimo componente no *ensemble* e após a obtenção dos pesos w_i .

Índice de Descorrelação

[Lima, 2004] propôs a medida de um “índice de descorrelação”, baseada no índice de correlação de Pearson. O índice de descorrelação entre dois modelos i e j é dado por:

$$\mathcal{D}(i, j) = 1 - \left| \frac{\sigma_{ij}}{\sigma_i \sigma_j} \right|, \quad (4.7)$$

em que:

$$\sigma_i = \sqrt{\frac{1}{N-1} \sum_{k=1}^N (y_i(k) - \mu_{Y_i})^2}, \quad (4.8)$$

$$\sigma_j = \sqrt{\frac{1}{N-1} \sum_{k=1}^N (y_j(k) - \mu_{Y_j})^2}, \quad (4.9)$$

$$\sigma_{ij} = \frac{1}{N-1} \sum_{k=1}^N (y_i(k) - \mu_{Y_i})(y_j(k) - \mu_{Y_j}). \quad (4.10)$$

Essa medida de diversidade é normalizada no intervalo $[0, 1]$, em que o valor máximo pode representar alta diversidade entre os componentes. Nota-se que essa medida não tem informações sobre o sinal

algébrico da correlação, ou seja, se é direta ou inversamente correlacionados. Essa informação pode ser relevante na formação de *ensembles* efetivos [Barbosa, 2009].

Informação Mútua

Informação mútua também é usada no estudo de *ensembles* [Yao and Liu, 2004] e pode ser definida como:

$$I(y_i, y_j) = H(y_i) + H(y_j) - H(y_i, y_j), \quad (4.11)$$

em que $H(y_i)$ é a entropia de y_i , $H(y_j)$ a entropia de y_j , $H(y_i, y_j)$ a entropia conjunta entre y_i e y_j . Supondo y_i e y_j variáveis aleatórias de distribuição gaussiana, com variância σ_i^2 e σ_j^2 , respectivamente, sua entropia é dada por:

$$H(y_i) = \frac{1}{2}[1 + \log(2\pi\sigma_i^2)], \quad (4.12)$$

$$H(y_j) = \frac{1}{2}[1 + \log(2\pi\sigma_j^2)]. \quad (4.13)$$

A informação mútua é dada por:

$$H(y_i, y_j) = 1 + \log(2\pi) + \frac{1}{2}[\sigma_i^2\sigma_j^2(1 - \rho_{ij}^2)], \quad (4.14)$$

em que ρ_{ij} é o coeficiente de correlação entre y_i e y_j , definido na Equação 4.2.

A informação mútua pode mostrar, aproximadamente, a similaridade entre duas redes neurais [Yao and Liu, 2004] e, assim, um baixo valor pode indicar alta diversidade entre os componentes i e j .

Outras Métricas

Apesar deste trabalho se ater às métricas supracitadas, existem outras possíveis formas de medição de diversidade. Entre elas, tem-se a correlação entre resíduos, em que se faz necessário o uso de dados rotulados (*labeled data*) para a medição da diversidade, e seu uso remete a ideia de independência entre falhas dos modelos; medidas de verossimilhança; decomposição em valores singulares e coerência.

Observa-se que, nesse trabalho, são analisadas apenas as medidas de diversidade mais genéricas, que utilizam dados nas medições e são facilmente aplicadas ao problema de regressão. Medições específicas, que consideram variações estruturais, análises de agrupamento de termos e outras não são exploradas nesse trabalho.

4.3 Diversidade em Identificação de Sistemas

Nesta seção, são realizados experimentos que objetivam a aplicação das métricas de diversidade em modelos dinâmicos, no contexto de identificação de sistemas. Primeiro, são feitos experimentos utilizando modelos dinâmicos lineares, a fim de compreender melhor a diversidade nesse contexto e, em seguida, são realizados estudos utilizando modelos não lineares com estrutura tipo rede neural. Por fim, é proposto um procedimento, prático, para a medição de diversidade em modelos dinâmicos aplicados à identificação de sistemas.

4.3.1 Diversidade em Sistemas Dinâmicos Lineares

As medidas de diversidade aqui apreciadas necessitam de dados para que o cálculo seja feito. Assim, um bom conjunto de dados dinâmicos é parte essencial para se obter uma medida condizente com os modelos analisados. A seguir, são discutidas questões acerca da escolha dos dados na aplicação de métricas de diversidade.

Definição dos Testes

Seja o sistema $G(s)$, linear invariante no tempo (LIT), alvo da modelagem, dado por:

$$G(s) = \frac{0,1737(s^2 + 0,2s + 9,01)(s + 20)}{(s + 1)(s + 5)(s^2 + 0,25s + 6,26)}. \quad (4.15)$$

Na modelagem do sistema $G(s)$, supõe-se que foram obtidos três modelos, diferentes entre si, dados por:

$$G_1(s) = \frac{0,05(s + 20)}{(s + 1)}, \quad (4.16)$$

$$G_2(s) = \frac{0,5549(s^2 + 0,2s + 9,01)}{(s + 1)(s + 5)}, \quad (4.17)$$

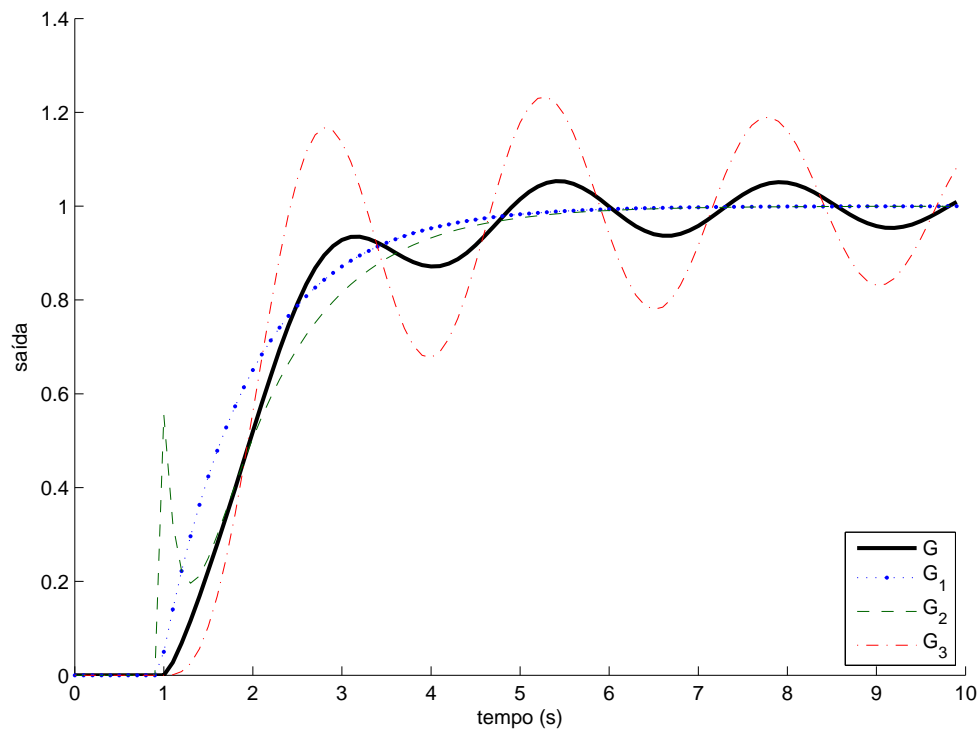
$$G_3(s) = \frac{6,26}{(s + 1)(s^2 + 0,25s + 6,26)}. \quad (4.18)$$

Note que os modelos se diferenciam tanto em estrutura (número de polos e zeros) quanto nas características dinâmicas de comportamento (localização dos polos e zeros). Porém, o polo dominante de $G(s)$, $s = -1$, está presente em todos os modelos. Isso ilustra uma situação, comum, de que o comportamento mais “óbvio” do sistema, ditado pelo polo dominante (mais próximo de $s = 0$), é facilmente detectado por todos os modelos dinâmicos que formarão o *ensemble*. Porém, outras características do sistema são detectadas apenas por alguns modelos, tais como a localização dos zeros e demais polos. Assim, a formação do *ensemble* se faz necessária para descrever o sistema por completo e também para que comportamentos dinâmicos triviais tenham sua precisão reforçada.

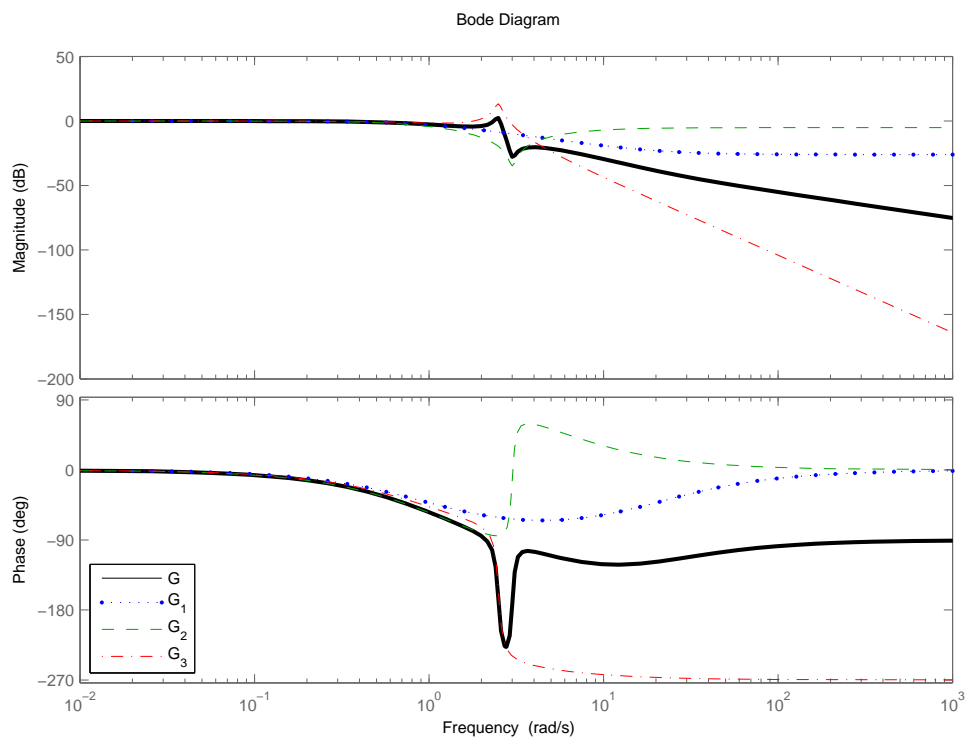
A Figura 4.2 mostra a resposta ao degrau unitário no tempo 1s e o diagrama de Bode dos quatro sistemas. A resposta ao degrau mostra, claramente, a diferença no comportamento dinâmico dos três modelos. Enquanto $G_1(s)$ apresenta uma resposta típica de um sistema de primeira ordem; $G_2(s)$ apresenta um pico inicial, provocado pela presença dos zeros complexos conjugados e amortecimento suave; já o modelo $G_3(s)$ mostra um comportamento bastante oscilatório. As diferenças entre os sistemas também são notórias no diagrama de Bode, principalmente nas frequências mais elevadas. A amplitude e fase dos sistemas até o valor aproximado de 10^0rad/s são bastante parecidas, devido à coexistência do polo dominante $s = -1$ em todos os sistemas.

Apesar de os modelos mencionados não serem completos *per se*, seus polos e zeros estão precisamente localizados de acordo com o sistema alvo $G(s)$, situação bastante incomum na prática. Incertezas, de várias naturezas, associadas à modelagem de sistemas dinâmicos, por vezes, não permitem uma precisa alocação de tais polos e zeros. Assim, considere três tipos de modelos dinâmicos descritos a seguir:

- **Modelos tipo A (\mathcal{M}_A):** obtidos a partir de $G_1(s)$, cujos polos e zeros são deslocados, aleatoriamente, com distribuição uniforme, de $\pm 15\%$ do valor original;
- **Modelos tipo B (\mathcal{M}_B):** obtidos a partir de $G_2(s)$, cujos polos e zeros são deslocados, aleatoria-



(a) resposta ao degrau em 1s



(b) diagrama de Bode

Figura 4.2: Comportamento dos sistemas dinâmicos lineares apresentados na forma de (a) resposta ao degrau e (b) diagrama de Bode para o sistema alvo da modelagem $G(s)$ e seus respectivos modelos $G_1(s)$, $G_2(s)$ e $G_3(s)$.

mente, com distribuição uniforme, de $\pm 15\%$ do valor original;

- **Modelos tipo C** (\mathcal{M}_C): obtidos a partir de $G_3(s)$, cujos polos e zeros são deslocados, aleatoriamente, com distribuição uniforme, de $\pm 15\%$ do valor original.

Dessa forma, tem-se três tipos de modelos que, em média, possuem a mesma localização de polos e zeros que os modelos $G_1(s)$, $G_2(s)$ e $G_3(s)$.

De posse de tais modelos, formam-se três tipos de *ensembles*, contendo três modelos cada um, conforme a seguir:

- **Ensembles tipo 1** (E_1): formados por três modelos de mesmo tipo (e.g. $\{\mathcal{M}_A, \mathcal{M}_A, \mathcal{M}_A\}$ ou $\{\mathcal{M}_B, \mathcal{M}_B, \mathcal{M}_B\}$ ou $\{\mathcal{M}_C, \mathcal{M}_C, \mathcal{M}_C\}$);
- **Ensembles tipo 2** (E_2): formados por três modelos, sendo dois de mesmo tipo e outro de tipo diferente dos demais (e.g. $\{\mathcal{M}_A, \mathcal{M}_A, \mathcal{M}_B\}$ ou $\{\mathcal{M}_A, \mathcal{M}_A, \mathcal{M}_C\}$ ou $\{\mathcal{M}_B, \mathcal{M}_B, \mathcal{M}_A\}, \dots$);
- **Ensembles tipo 3** (E_3): formados por três modelos de tipos diferentes (e.g. $\{\mathcal{M}_A, \mathcal{M}_B, \mathcal{M}_C\}$).

Ensembles do tipo 1 não conseguem explicar muito do sistema original, já que possuem apenas parte dos polos e zeros próximos aos de $G(s)$. *Ensembles* do tipo 2 explicam melhor o sistema original, mas, ainda assim, não representam todo seu comportamento. Somente *ensembles* do tipo 3 pode modelar $G(s)$ por completo, já que os modelos que o compõe são complementares. Em outras palavras, *ensembles* E_3 são, claramente, mais *diversos* que *ensembles* E_1 . Em analogia com a Figura 4.1, mostrada na seção anterior, a Figura 4.3 ilustra a disposição dos polos e zeros em cada modelo.

Conforme discutido na Seção 4.2, é necessário que os dados utilizados na medição de diversidade consigam expressar bem as diferenças entre modelos, especialmente em regiões em que essas diferenças sejam esperadas, ou seja, fora da região de interseção entre os modelos (vide Fig. 4.3). Assim, os testes de medição da diversidade entre modelos utilizarão três conjuntos de dados de entrada, conforme a seguir:

- **Dados tipo A** (Z_1): ruído branco gaussiano, filtrado por um passa-baixa Chebyshev tipo I, ordem $n = 5$, ripple de 1dB na banda de passagem e frequência de corte $f_0 = 0,5\text{rad/s}$;
- **Dados tipo B** (Z_2): ruído branco gaussiano;
- **Dados tipo C** (Z_3): ruído branco gaussiano, filtrado por um passa-alta Chebyshev tipo I, ordem $n = 5$, ripple de 1dB na banda de passagem e frequência de corte $f_0 = 0,5\text{rad/s}$.

O conjunto de dados Z_1 representa dados limitados, cujo conteúdo espectral é restrito a regiões de baixa frequência. Tais dados permitem apenas a verificação do efeito de polos mais lentos, i.e. do polo comum entre os três tipos de modelos. Por outro lado, Z_2 representa dados persistentemente excitantes, que exploram todas as frequências, possibilitando a verificação do comportamento em todo conteúdo espectral nos três modelos. O conjunto Z_3 , por sua vez, representa dados com conteúdo espectral apenas nas frequências mais elevadas ($f > 0,5\text{rad/s}$), regiões em que os modelos possuem menos em comum (vide diagrama de Bode, Fig. 4.2b). Este cenário pode ser associado aos conjuntos de Venn apresentados na Fig. 4.3, em que Z_1 , Z_2 e Z_3 expressam melhor o comportamento dos modelos em determinadas regiões espectrais. A Figura 4.4 ilustra essa situação, em que, à esquerda é destacado, em cinza, as regiões excitadas pelos dados nos modelos, ou seja, quais comportamentos são melhores representados com cada massa de dados. À direita é mostrado o espectro de frequências das massas de dados Z_1 , Z_2 e Z_3 , respectivamente.

A fim de avaliar as métricas de diversidade em *ensembles* E_1 , E_2 e E_3 , e utilizando as massas de dados Z_1 , Z_2 e Z_3 , foi realizado o seguinte procedimento de teste:

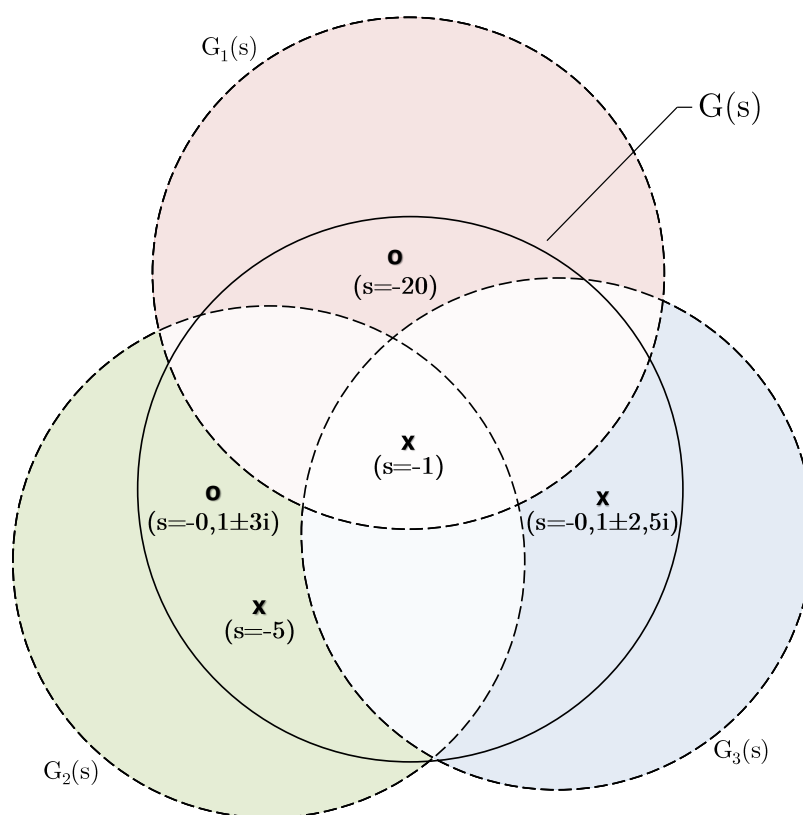


Figura 4.3: Diagramas de Venn que ilustra a disposição dos (\times) polos e (\circ) zeros dos modelos G_1 , G_2 , G_3 e no sistema alvo da modelagem $G(s)$.

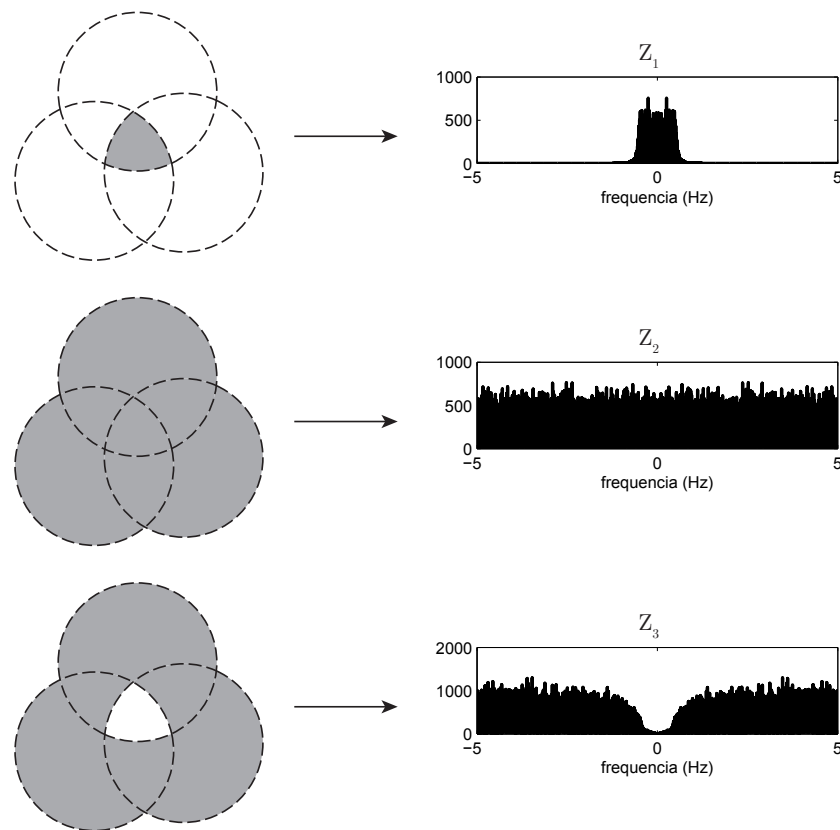


Figura 4.4: Representação das massas de dados Z_1 , Z_2 e Z_3 no âmbito dos diagramas de Venn (apresentados anteriormente na fig. 4.3) e o espectro de frequências desses dados. A área destacada na cor cinza, à esquerda, ilustra qual região do *ensemble* é excitada por cada conjunto de dados.

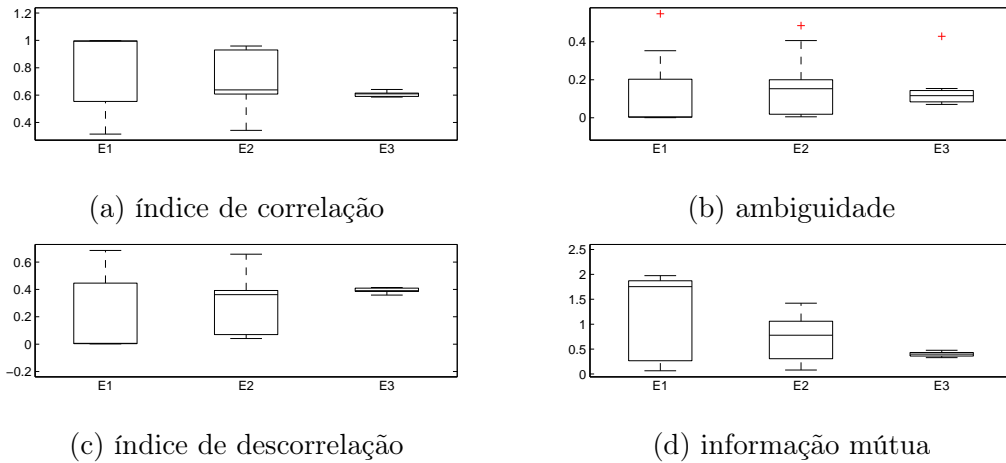


Figura 4.5: Medidas de diversidade para os três tipos de *ensembles*: E_1 , E_2 , E_3 , utilizando os dados de entrada Z_1 , em diferentes métricas.

1. Tomam-se 12 *ensembles* de cada tipo E_1 , E_2 , E_3 ;
2. Em cada *ensemble*, a diversidade é calculada, utilizando as métricas da Seção 4.2.1, par a par entre os componentes. A média desses valores é considerada a “diversidade do *ensemble*”;
3. Essa diversidade do *ensemble* é calculada com o uso dos conjuntos de dados Z_1 , Z_2 , Z_3 ;
4. Tomam-se as médias (μ_{E_1} , μ_{E_2} , μ_{E_3}) e variâncias ($\sigma_{E_1}^2$, $\sigma_{E_2}^2$, $\sigma_{E_3}^2$) dessas medidas, para cada tipo de *ensemble*, em cada conjunto de dados;
5. As medidas são avaliadas em termos de distinção da diversidade entre os tipos de *ensembles*, ou seja, o quão “distinguíveis” são os três tipos E_1 , E_2 e E_3 .

Como resultado desse procedimento, dispõem-se, para cada métrica, de três conjuntos de medições de diversidade (usando Z_1 , Z_2 , Z_3), com seus respectivos valores de média e variância. Os valores de diversidade são avaliados a seguir.

Resultados e Discussão

A Figura 4.5 mostra as medidas de diversidade para cada tipo de *ensemble*, utilizando o conjunto de dados de entrada Z_1 . Nessa figura, é importante avaliar, principalmente, a distinção entre os *ensembles* E_3 e E_1 , sendo o primeiro dito *diverso* e o segundo não, por razões já mencionadas. Nota-se uma difícil distinção entre *ensembles* diversos e não diversos, com sobreposições entre os valores. Neste caso, observa-se que a métrica que utiliza informação mútua apresenta melhor resultado, i.e., melhor distinção entre *ensembles* E_3 e E_1 .

A Figura 4.6 mostra as medidas de diversidade para cada tipo de *ensemble*, utilizando o conjunto de dados de entrada Z_2 . Nesse caso, nota-se uma melhoria expressiva na distinção entre *ensembles* diversos e não diversos. As métricas mostram resultados parecidos entre si, exceto pelo termo de ambiguidade, que apresenta menor distinção entre os *ensembles*.

A Figura 4.7 mostra as medidas de diversidade para cada tipo de *ensemble*, utilizando o conjunto de dados de entrada Z_3 . Desta vez, nota-se, claramente, uma distinção mais nítida entre *ensembles* diversos E_3 e não diversos E_1 . A métrica de informação mútua, que havia apresentado melhor resultado

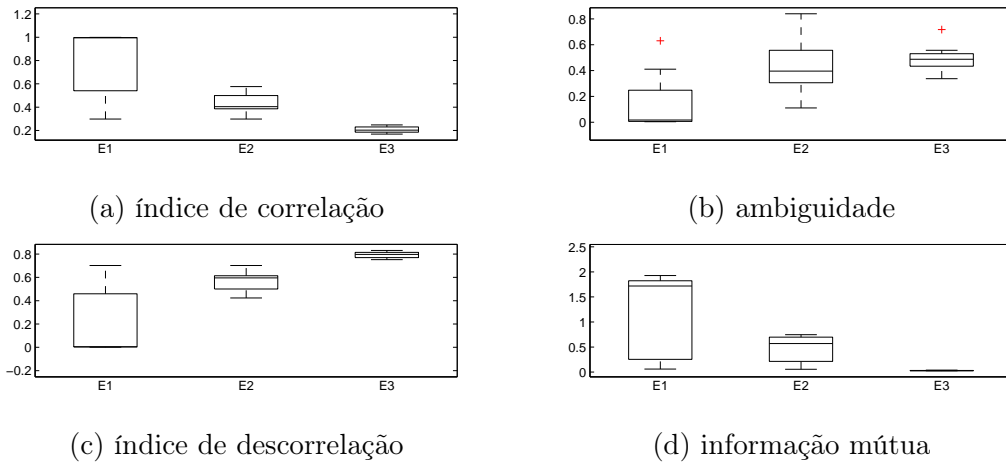


Figura 4.6: Medidas de diversidade para os três tipos de *ensembles*: E_1 , E_2 , E_3 , utilizando os dados de entrada Z_2 , em diferentes métricas.

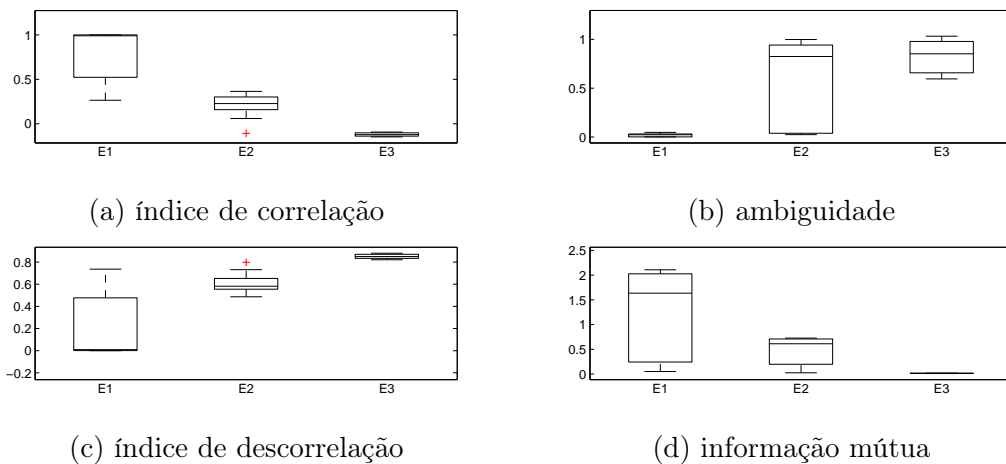


Figura 4.7: Valores de medidas de diversidade para os três tipos de *ensembles*: E_1 , E_2 , E_3 , utilizando os dados de entrada Z_3 , em diferentes métricas.

utilizando Z_1 , desta vez em Z_3 , mostra-se menos distinguível dentre as métricas. Em contrapartida, o termo de ambiguidade, que havia mostrado os piores resultados até então, utilizando Z_3 apresentou um dos melhores dentre as métricas. É interessante observar também as medidas baseadas em índices de correlação, que, para Z_1 e Z_2 , mostraram resultados muito semelhantes, porém, desta vez, observa-se uma diferença, especialmente na medida de E_2 , em que o índice de descorrelação mostra uma melhor distinção da medida de E_3 .

Além da diversidade, é interessante avaliar também as medidas de erro dos *ensembles* em relação à resposta do sistema alvo da modelagem. A Figura 4.8 mostra os valores de raiz do erro quadrático médio (RMSE) para os três tipos de *ensembles*, utilizando a massa de dados Z_1 e Z_3 . Comparando a média dos RMSEs em Z_1 e Z_3 , nota-se uma maior exatidão dos *ensembles* no primeiro conjunto de dados. Isso mostra que comportamentos explicados por vários modelos (i.e. Z_1) elevam o nível de exatidão do *ensemble*, seja ele diverso ou não. Contudo, sobre os dados Z_3 são mais claros os benefícios relacionados à diversidade de *ensembles*, em que E_3 se mostra mais preciso em relação aos demais. Ademais, a

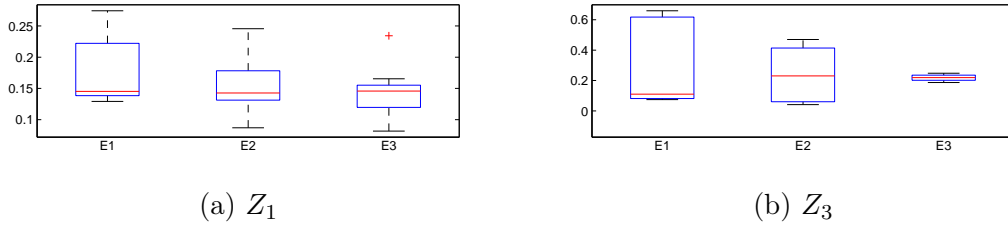


Figura 4.8: Valores da raiz do erro quadrático médio (RMSE) dos *ensembles* E_1 , E_2 , E_3 , para duas massas de dados Z_1 e Z_3 .

baixa variância observada contribui para a robustez do *ensemble*, que apresenta respostas satisfatórias a diferentes conjuntos de dados.

Conclusões do Teste

O teste utilizando sistemas lineares mostrou que a medida de diversidade está mais relacionada ao conjunto de dados do que à métrica utilizada. O uso de um bom conjunto de dados, que excita os modelos em regiões em que as diferenças de comportamento são esperadas (e necessárias) para o bom desempenho do *ensemble*, pode ajudar muito na distinção entre *ensembles* diversos e não diversos. Não é interessante medir diversidade, por exemplo, utilizando dados que excitam apenas os comportamentos triviais, bem representados em todos os modelos.

Foi mostrado também que o uso de *ensembles* pode aprimorar a exatidão e a precisão da resposta dos modelos, sendo que a última é melhor alcançada com o uso de *ensembles* diversos. Melhorias na exatidão também foram observadas com o uso de *ensembles* diversos (como discutido em [Brown, 2004]).

Apesar dos testes terem sido aplicados em sistemas lineares, boa parte das considerações também podem ser aplicadas a sistemas não lineares. Na próxima seção, são realizados experimentos usando sistemas e modelos não lineares.

4.3.2 Diversidade em Modelos Não lineares

Nesta seção são realizados experimentos utilizando modelos não lineares, com estrutura de redes neurais artificiais. O objetivo aqui é investigar em que medida o que foi mostrado usando sistemas lineares, na seção anterior (Seção 4.3.1), se aplica a modelos não lineares, mais especificamente, redes neurais.

Definição do problema

Seja o sistema dinâmico S_{nl} , não linear, discreto, alvo da modelagem, dado por:

$$y(k) = \frac{2,5y(k-1)y(k-2)}{1 + y(k-1)^2 + y(k-2)^2} + 0,3\cos(0,5y(k-1) + 0,5y(k-2)) + 1,2u(k-1). \quad (4.19)$$

Esse sistema também foi usado, com propósitos ilustrativos, na referência [Bravo and Normey-Rico, 2009]. Trata-se de um sistema de segunda ordem, SISO (única entrada e única saída), com diferentes tipos de não linearidades, tais como termos quadráticos, divisão desses termos e função cosseno. Para se ter ideia da não linearidade do sistema, a Figura 4.9 mostra sua curva estática. Nota-se uma oscilação do ganho do sistema, que pode observada pela variação da inclinação da curva ao longo da faixa de operação

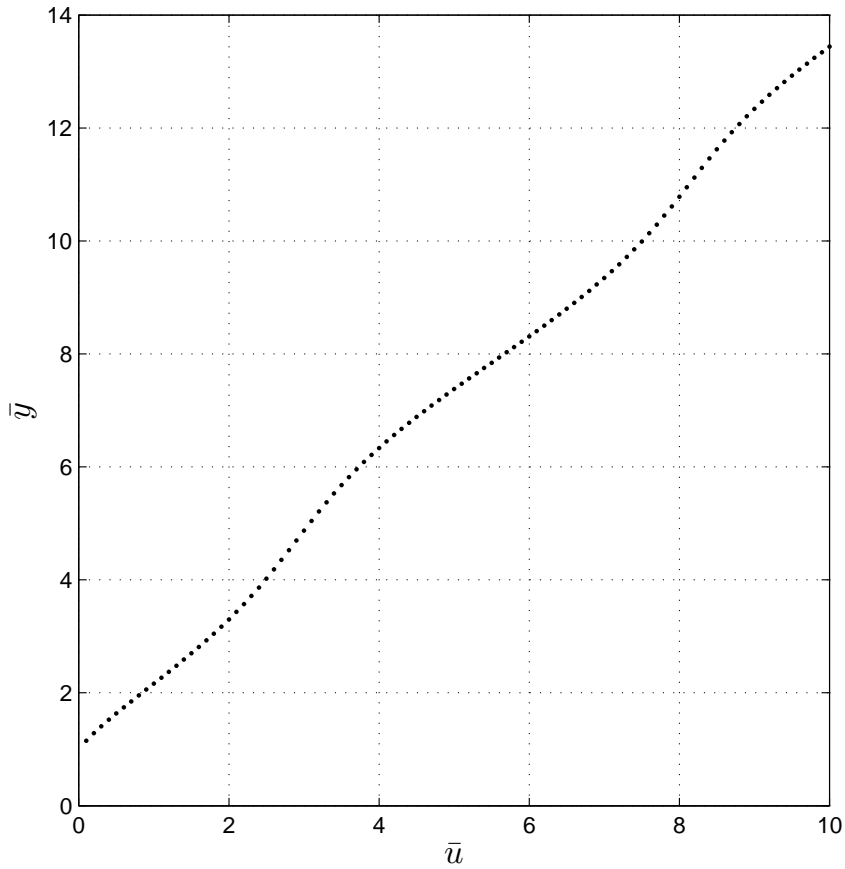


Figura 4.9: Curva estática do sistema não linear, alvo da modelagem, apresentado na Equação 4.19.

apresentada. Não linearidades podem ser observadas também no comportamento dinâmico do sistema. A Figura 4.10 mostra a resposta ao degrau unitário em dois pontos de operação distintos. Nota-se diferenças no comportamento transitório do sistema, e.g., nos valores de sobre-elevação (*overshoot*) e frequência de oscilação.

São utilizados na identificação, três modelos dinâmicos, não lineares, com a seguinte estrutura geral:

$$\hat{y}(k) = \sum_{i=1}^q w_i f_i \left(\sum_{j=1}^m w_{ij} \psi_j(k) + w_{i0} \right) + w_0, \quad (4.20)$$

em que q é o número de neurônios na camada escondida, m o número de regressores ψ_i , w_i os pesos na camada de saída, w_{ij} os pesos na camada escondida e f_i a i -ésima função de ativação, que pode ser linear ou $\tanh(\cdot)$. O vetor de regressores $\Psi = [\psi_1 \psi_2 \dots \psi_m]$ é dado por:

$$\Psi = [y(k-1)y(k-2)\dots y(k-n_y)u(k)u(k-1)\dots u(k-n_u)]^T, \quad (4.21)$$

em que n_y é o número de atrasos nos regressores de saída e n_u o número de atrasos nos regressores de

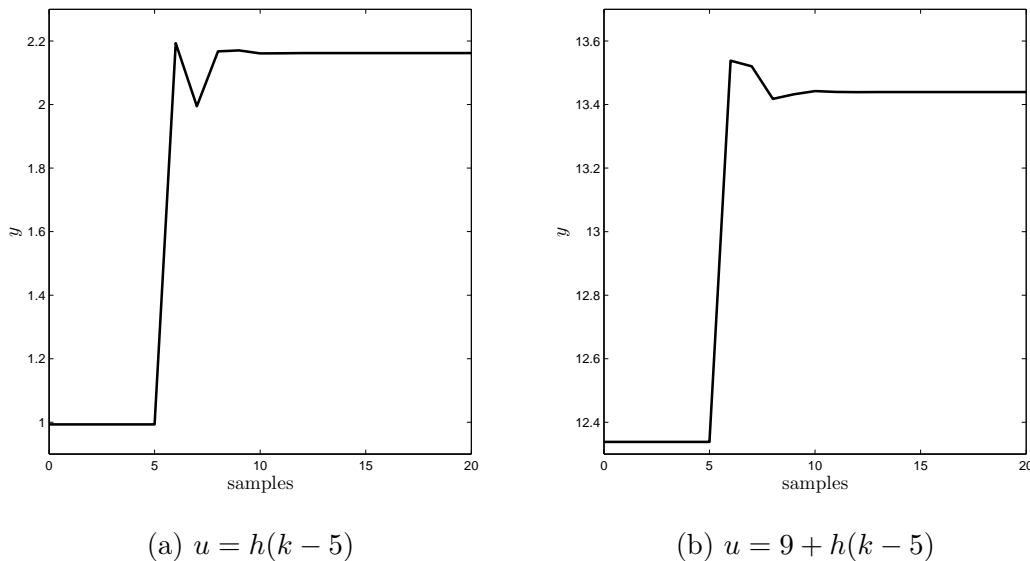


Figura 4.10: Resposta do sistema ao degrau unitário em dois pontos de operação distintos: (a) em $u(k) = h(k - 5)$ e (b) $u(k) = 9 + h(k - 5)$, em que $h(k)$ é a função de Heaviside (degrau unitário em $k = 0$).

entrada.

Os três modelos dinâmicos M_1 , M_2 e M_3 obtidos possuem variações estruturais nos parâmetros de atraso dos regressores e na quantidade de neurônios na camada escondida, que estão mostradas na Tabela 4.1. Os três modelos possuem uma camada escondida, com função de ativação $\tanh(x)$, e uma camada de saída com função linear.

Ajuste dos parâmetros

No ajuste dos parâmetros, foram utilizados dois conjuntos de dados, $Z_1 = \{[u(k)y(k)]|k = 1, \dots, N_1\}$ e $Z_2 = \{[u(k)y(k)]|k = 1, \dots, N_2\}$, com $N_1 = 3000$ e $N_2 = 100$ amostras, respectivamente. Esses dados foram obtidos a partir de dois sinais distintos de ruído branco gaussiano, filtrados por um passa-baixas Chebyshev tipo I, com ripple de 1dB na banda de passagem, ordem $n = 20$ e frequência de corte $w_0 = 0, 2\pi$ rad/amostra. A Figura 4.11 mostra o módulo da transformada discreta de Fourier dos sinais de entrada. Isso ilustra um fato comum em que dados disponíveis para identificação possuem espectro limitado. Nesse caso, a generalização ou extrapolação do comportamento dinâmico do sistema, para frequências mais elevadas, se torna um desafio na identificação de sistemas. A Figura 4.12 mostra os

Tabela 4.1: Parâmetros estruturais das redes neurais utilizadas na modelagem do sistema dinâmico não linear apresentado na Eq. 4.19.

	q	m	n_u	n_y
M_1	20	13	5	8
M_2	10	30	10	20
M_3	6	3	1	2

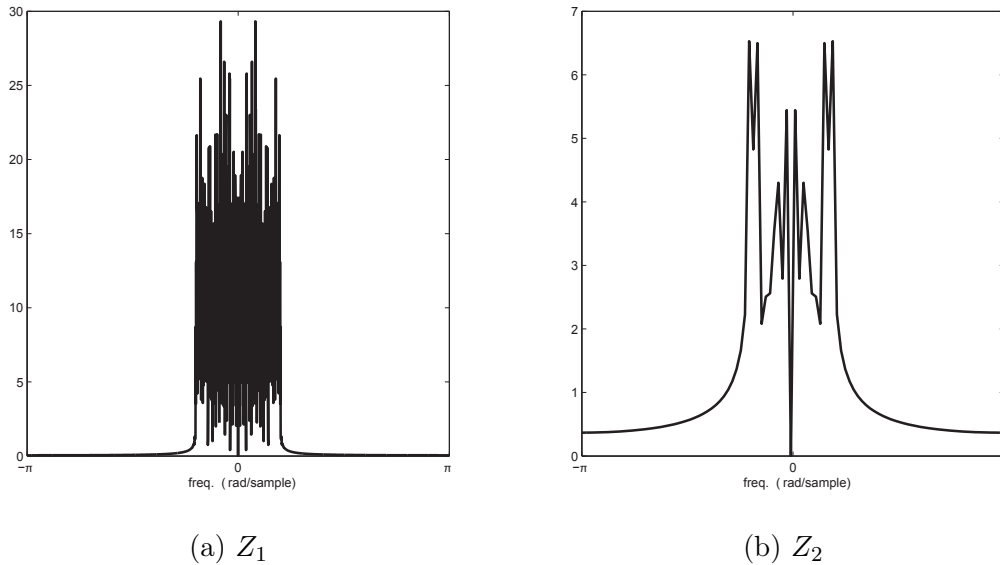


Figura 4.11: Módulo da transformada discreta de Fourier dos sinais de entrada de (a) Z_1 e (b) Z_2 , usados no ajuste dos parâmetros dos modelos. Os gráficos mostram uma escala entre as frequências mínima e máxima representáveis: $[-\pi, \pi]$.

sinais de entrada e saída dos dois conjuntos de dados, em que, o primeiro, Z_1 , foi utilizado no ajuste dos modelos dinâmicos M_1 e M_2 , enquanto Z_2 foi utilizado no ajuste dos parâmetros de M_3 , todos por meio do algoritmo de otimização de Levenberg-Marquadt.

Sabe-se que uma das formas de se obter diversidade é, justamente, utilizando diferentes conjuntos de dados para o ajuste dos parâmetros dos modelos [Sharkey and Sharkey, 1997]. Assim, espera-se que as medidas indiquem menor diversidade nos modelos $[M_1M_2]$ em relação aos modelos $[M_1M_3]$.

Medição de diversidade

Será mensurada a diversidade de $E_1 = \{M_1M_2\}$ e $E_2 = \{M_1M_3\}$. Serão calculadas as métricas apresentadas anteriormente na Seção 4.2.1, usando-se as seguintes massas de dados de entrada:

- D_1 : sinal de entrada de média $\bar{u} = 6$, tamanho $N = 5000$, obtido pela filtragem de um ruído branco gaussiano por um passa-baixas Chebyshev tipo I, ordem $n = 20$, ripple de 1dB na banda de passagem e frequência de corte $f_0 = 0,5\pi$ rad/amostra;
- D_2 : ruído branco gaussiano de média $\bar{u} = 5$ e tamanho $N = 5000$;
- D_3 : sinal de entrada de média $\bar{u} = 2$, tamanho $N = 5000$, obtido pela filtragem de um ruído branco gaussiano por um passa-altas Chebyshev tipo I, ordem $n = 20$, ripple de 1dB na banda de passagem e frequência de corte $f_0 = 0,5\pi$ rad/amostra.

Esses dados remetem-se à situação ilustrada, anteriormente, na Fig. 4.4, mas, desta vez, com sistema e modelos não lineares. A Figura 4.13 mostra trechos dos sinais nos domínios do tempo e da frequência. A massa de dados D_1 representa dados com características muito semelhantes aos dados utilizados no treinamento: com espectro de frequências limitado e mesma faixa de operação em torno do valor $\bar{u} = 6$. O conjunto de dados D_2 representa dados com conteúdo espectral e faixa de operação mais amplos,

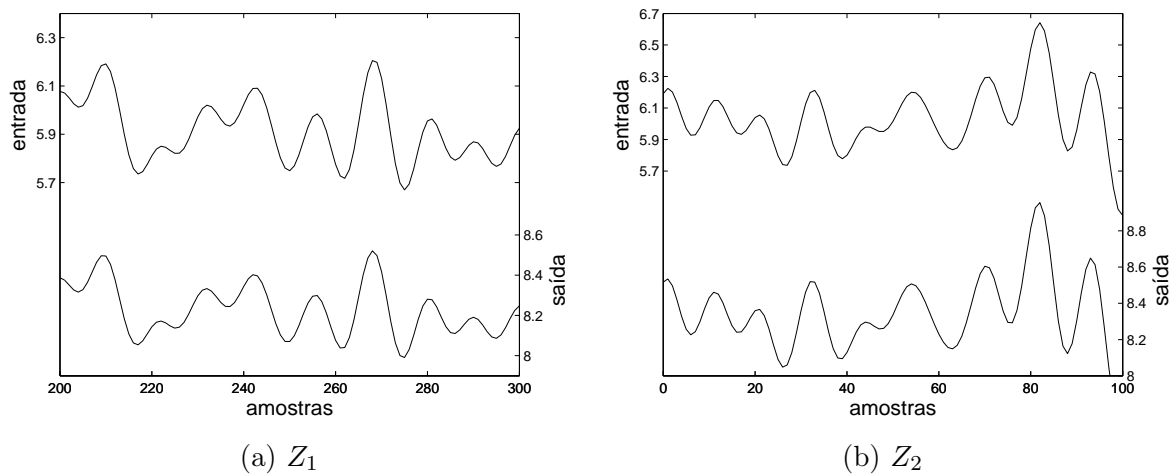


Figura 4.12: Trechos dos sinais de entrada e saída dos conjuntos de dados Z_1 e Z_2 usados no treinamento.

entre $\bar{u} = 0$ e $\bar{u} = 10$. Os dados D_3 , por sua vez, representam dados com conteúdo espectral e faixa de operação complementar aos dados de treinamento, i.e., excita o sistema somente em frequências não excitadas durante a fase de treinamento e em faixas de operação não visitadas pelos dados de treinamento.

Resultados e Discussão

A Figura 4.14 mostra a saída dos modelos e do sistema S_{nl} para uma série de degraus de entrada, a saber, $U = [0, 0, \dots, 0, 1, 1, \dots, 1, 5, 5, \dots, 5, 6, 6, \dots, 6, 9, 9, \dots, 9, 10, 10, \dots, 10]$. Verifica-se que os modelos se aproximam da resposta desejada (S_{nl}) apenas na região em que $5 \leq u(k) \leq 6$, cuja faixa de operação é a mesma dos dados de treinamento. A imagem destacada ao centro mostra o modelo M_3 com comportamento bastante similar ao sistema real, o que não ocorre com M_1 e M_2 . Vale lembrar que esses modelos foram obtidos utilizando dados limitados na frequência, enquanto que, um sinal de entrada em degrau, possui todas as frequências. Isso explica, em parte, as diferenças entre os modelos e o sistema. O bom desempenho de M_3 é justificado pelos regressores serem idênticos aos do sistema S_{nl} .

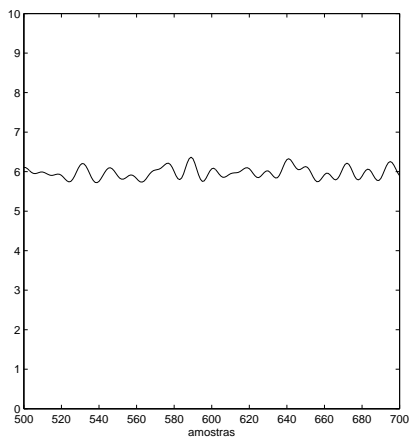
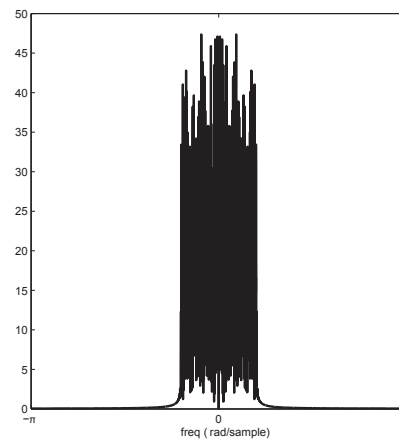
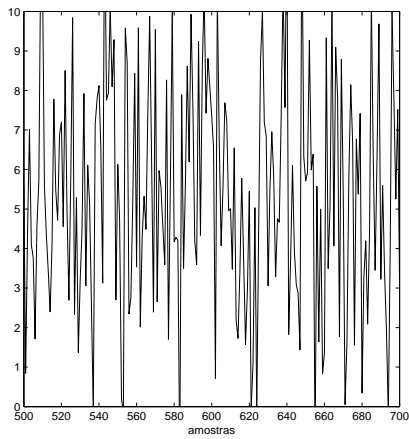
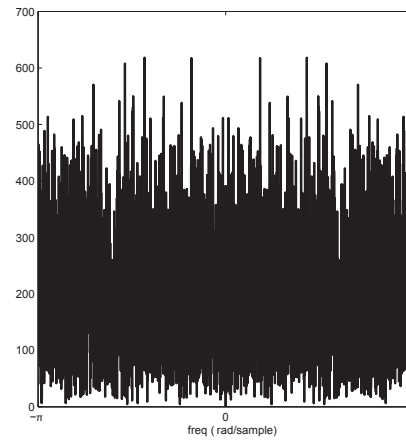
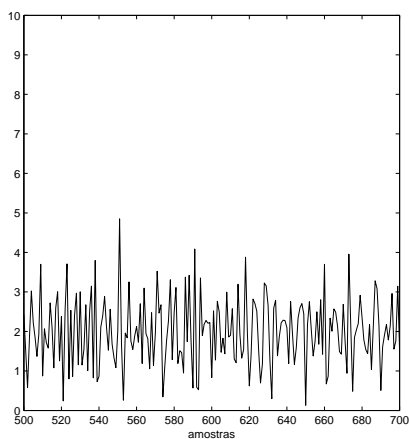
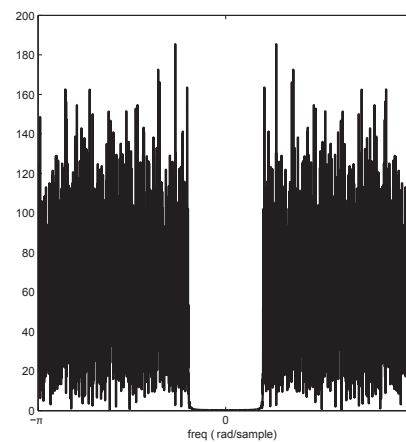
(a) D_1 (b) D_1 (c) D_2 (d) D_2 (e) D_3 (f) D_3

Figura 4.13: Trechos dos sinais de entrada utilizados nas medições de diversidade no (a) (c) (e) domínio do tempo e no (b) (d) (f) domínio da frequência. A escala vertical no domínio do tempo está ajustada para a faixa de operação do sistema.

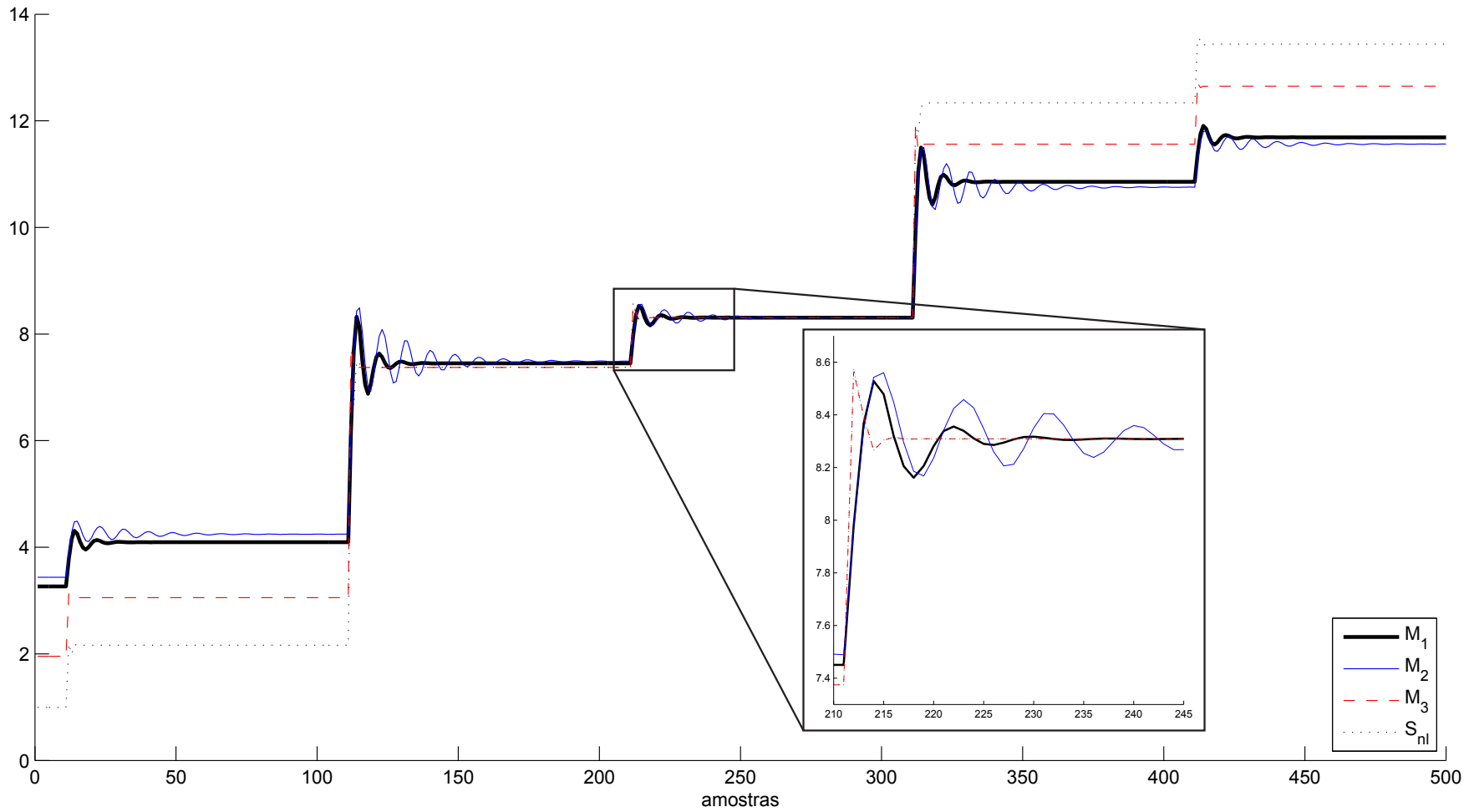


Figura 4.14: Saída do sistema S_{nl} e dos modelos M_1 , M_2 , M_3 para uma série de degraus de entrada $U = [0, 0, \dots, 0, 1, 1, \dots, 1, 5, 5, \dots, 5, 6, 6, \dots, 6, 9, 9, \dots, 9, 10, 10, \dots, 10]$. É destacada, no centro da imagem, a resposta dos sistemas na mesma faixa de operação dos dados de treinamento. Nessa imagem destacada, a saída do sistema S_{nl} está sobreposta à saída de M_3 .

Analisando, ainda na Figura 4.14, nota-se a diferença do comportamento dinâmico dos modelos, em que M_2 se mostra muito oscilatório, M_3 menos e M_1 intermediário. Os valores em estado estacionário também se diferenciam bastante em trechos fora da faixa de treinamento, sendo que os modelos mais próximos entre si são M_1 e M_2 . Numa análise qualitativa, de maneira geral, é possível dizer que os modelos apresentam certo grau de diversidade, com diferentes padrões de generalização, em que, aparentemente, $E_2 = \{M_1M_3\}$ é mais diverso que $E_1 = \{M_1M_2\}$, tanto no comportamento dinâmico quanto estático. Isso será verificado nas medidas de diversidade.

A Tabela 4.2 mostra os resultados das medidas de diversidade realizadas. É interessante notar nessas medidas que, usando a massa de dados D_1 , parece que E_1 é um pouco mais diverso que E_2 , para todas as métricas. Porém, ao analisar o resultado para as massas de dados D_2 e D_3 , verificamos exatamente o oposto, e com uma diferença de valores muito maior. O termo de ambiguidade, por exemplo, em D_3 , chega a ser mais de dez vezes maior em E_2 , em relação a E_1 . Mostra que o modelo M_1 é mais parecido com M_2 do que com M_3 , o que condiz com as análises qualitativas e a expectativa inicial dos experimentos.

Tabela 4.2: Medidas de diversidade nos *ensembles* $E_1 = \{M_1M_2\}$ e $E_2 = \{M_1M_3\}$, usando quatro métricas distintas: correlação, ambiguidade, descorrelação e informação mútua (vide Seção 4.2.1 para detalhes); e usando três conjuntos de dados de entrada distintos: D_1 , D_2 e D_3 .

	Dados	Diversidade			
		Correlação	Ambiguidade	Descorrelação	Informação Mútua
E_1	D_1	0,8879	0,0019	0,1121	0,7614
	D_2	0,9290	0,3588	0,0710	0,9050
	D_3	0,8960	0,0401	0,1040	0,7283
E_2	D_1	0,9140	0,0014	0,0860	0,9405
	D_2	0,6899	3,5942	0,3101	0,2982
	D_3	0,6370	0,5972	0,3630	0,2415

Nesses resultados, apresentados na Tabela 4.2, não foram observadas diferenças significativas entre as métricas utilizadas, contudo, observou-se diferenças expressivas na escolha de dados para a medição da diversidade. Foi mostrado que o uso de dados de entrada semelhantes aos utilizados no treinamento, mesmo que descorrelacionados, não apresentam bons resultados na medição da diversidade. O uso de dados que visitam toda a faixa de operação do sistema e com amplo conteúdo espectral (D_2) ou dados complementares (D_3) se mostra mais eficaz na averiguação da diversidade em *ensembles*. Pouca diferença foi observada entre o uso de D_2 ou D_3 .

Conclusões do Teste

Foi realizado experimento sobre a medição da diversidade num problema de identificação de sistemas não lineares, utilizando modelos neurais. Os resultados mostraram que a averiguação da diversidade está muito associada aos dados de entrada usados no cálculo das métricas. Foi mostrado que dados parecidos com os de treinamento, em termos de conteúdo espectral e faixa de operação, não ajudam muito na avaliação da diversidade. Nesse caso, é preferível utilizar dados ricos em conteúdo espectral, com ampla faixa de operação; ou ainda dados com conteúdo espectral e faixa de operação complementares aos do treinamento.

Esse experimento mostrou, em sistemas não lineares, o que foi discutido para sistemas lineares na Seção 4.3.1. Em vista desses resultados, cabe aqui, uma proposta de um procedimento, geral, para medição da diversidade entre modelos em problemas de identificação de sistemas dinâmicos. Esse procedimento será discutido na próxima seção.

4.3.3 Procedimento de Medição de Diversidade de *Ensembles* em Problemas de Identificação

À luz dos experimentos realizados, cabe aqui a definição de um procedimento, geral, para a medição da diversidade de *ensembles* aplicados em problemas de identificação de sistemas dinâmicos. Primeiro são sintetizados os conceitos e, em seguida, o procedimento é estabelecido.

Conceitos

Dados similares aos utilizados no ajuste dos parâmetros dos modelos não ajudam muito na medição de diversidade. Porém, comparando-se dois modelos, é possível que cada modelo tenha usado dados de ajuste com diferentes características. Nesse caso, é interessante que a diversidade seja mensurada com dados que não contenham características que são *comuns* a ambos os dados de treinamentos dos modelos. A Figura 4.15 ilustra melhor essa ideia, mostrando, no domínio da frequência, dados de ajuste do primeiro modelo, Z_1 , e dados de ajuste do segundo modelo, Z_2 . O sinal D_1 representa o conjunto de dados a ser usado na medição da diversidade.

A faixa de operação percorrida pelo conjunto de dados também é importante. É interessante que, na medição de diversidade, os dados cubram regiões de operação distintas das visitadas pelos dados de treinamento. A Figura 4.16 ilustra o motivo, já discutido anteriormente, em que M_1 e M_2 representam dois modelos diversos em um problema de identificação do sistema S_1 . Em geral, a faixa de operação utilizada no treinamento é bem aproximada pelos modelos dinâmicos, como pode ser observado na região destacada da figura. Note que, nessa região, não é possível observar o quão diferentes são os modelos, mas, por outro lado, fora da região de treinamento, é possível observar a diferença no padrão de generalização. A diversidade, com relação ao comportamento estático do sistema, é melhor observada em regiões fora dos limites de operação dos dados de treinamento. Nos casos em que os modelos utilizam diferentes conjuntos de dados de ajuste de parâmetros, com diferentes faixas de operação, é interessante que a região de operação *em comum* entre os dados de treinamento não componha o conjunto de dados usado na medição da diversidade.

As diretrizes apresentadas até então, usando análise na frequência e faixa de operação dos sinais de entrada, são apenas algumas características que podem ajudar na medição da diversidade. Sistemas não lineares podem apresentar padrões de comportamento que variam também com outros aspectos não analisados aqui, tais como amplitude do sinal de entrada, combinação de diferentes entradas, entre outros. O ideal seria que o sinal utilizado na medição de diversidade explorasse todo o comportamento do sistema não presente nos dados de treinamento dos modelos. Porém, acredita-se que, usando dados que se diferenciam dos de ajuste apenas no conteúdo espectral e na faixa de operação, é possível obter uma boa medição da diversidade de modelos dinâmicos.

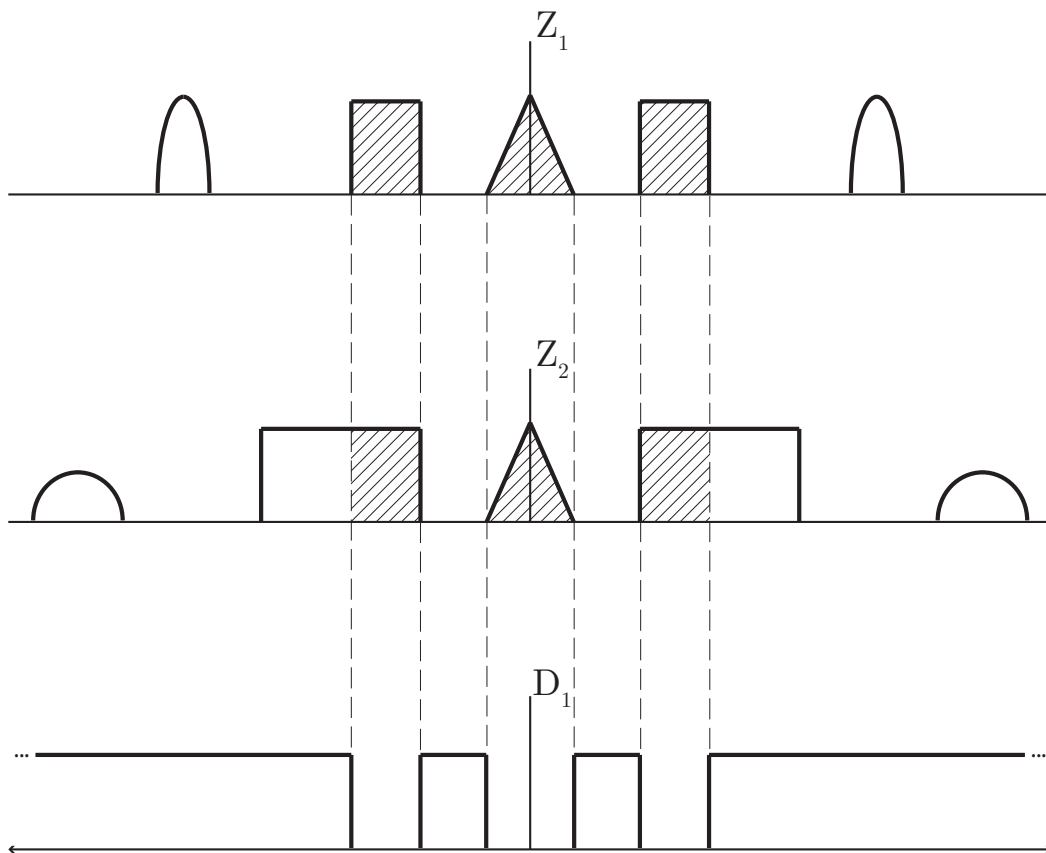


Figura 4.15: Características sugeridas dos sinais de entrada, no domínio da frequência, para medição de diversidade. Em que Z_1 representa dados de ajuste do primeiro modelo, Z_2 dados de ajuste do segundo modelo e D_1 dados a serem usados na medição da diversidade entre os modelos. Todos representam apenas sinais de entrada dos modelos.

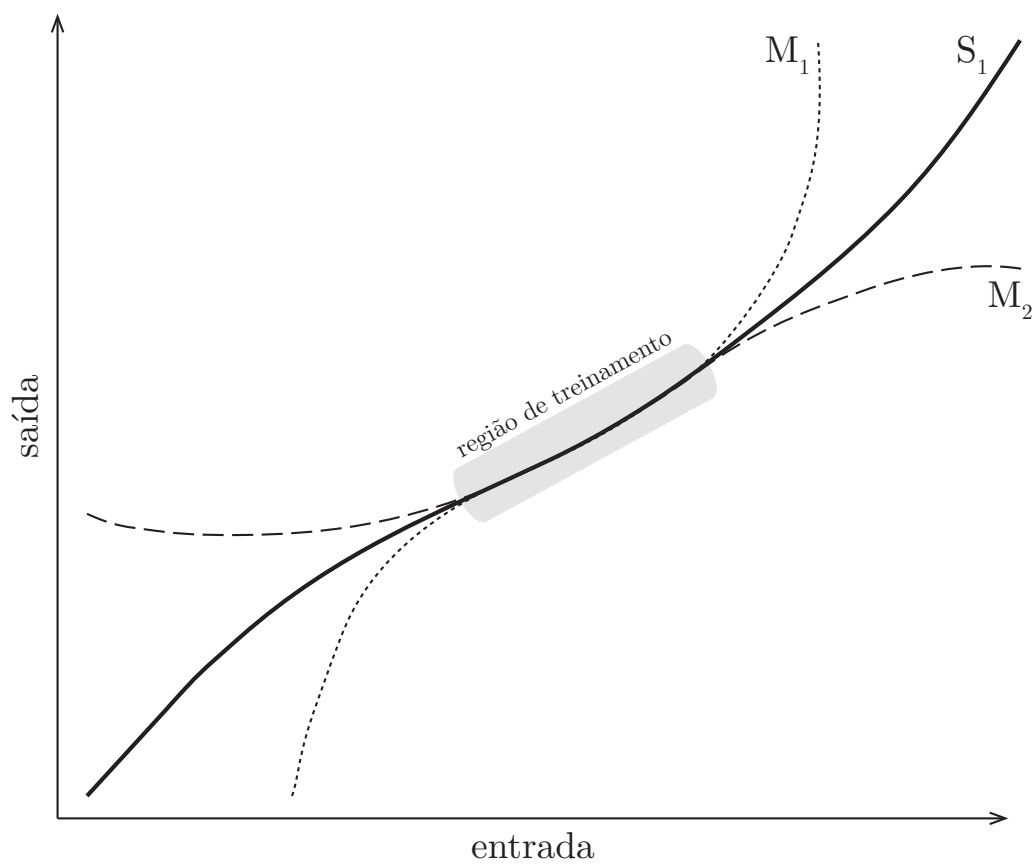


Figura 4.16: Ilustração da curva estática de um sistema S_1 aproximado por dois modelos, M_1 e M_2 , em que, em destaque, é apresentada a faixa de operação usada no treinamento desses modelos.

Procedimento de Medição da Diversidade

Considere um sistema S_1 , com N_u entradas e N_y saídas, aproximado por dois modelos M_1 e M_2 , que utilizaram, no ajuste de seus parâmetros, os conjuntos de dados Z_1 e Z_2 , respectivamente, com N_{Z_1} e N_{Z_2} amostras cada, dado por:

$$Z_1 = \left[\begin{array}{cccc|cccc} u_1^1(1) & u_1^2(1) & \cdots & u_1^{N_u}(1) & y_1^1(1) & y_1^2(1) & \cdots & y_1^{N_y}(1) \\ u_1^1(2) & u_1^2(2) & \cdots & u_1^{N_u}(2) & y_1^1(2) & y_1^2(2) & \cdots & y_1^{N_y}(2) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ u_1^1(N_{Z_1}) & u_1^2(N_{Z_1}) & \cdots & u_1^{N_u}(N_{Z_1}) & y_1^1(N_{Z_1}) & y_1^2(N_{Z_1}) & \cdots & y_1^{N_y}(N_{Z_1}) \end{array} \right], \quad (4.22)$$

$$Z_2 = \left[\begin{array}{cccc|cccc} u_2^1(1) & u_2^2(1) & \cdots & u_2^{N_u}(1) & y_2^1(1) & y_2^2(1) & \cdots & y_2^{N_y}(1) \\ u_2^1(2) & u_2^2(2) & \cdots & u_2^{N_u}(2) & y_2^1(2) & y_2^2(2) & \cdots & y_2^{N_y}(2) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ u_2^1(N_{Z_2}) & u_2^2(N_{Z_2}) & \cdots & u_2^{N_u}(N_{Z_2}) & y_2^1(N_{Z_2}) & y_2^2(N_{Z_2}) & \cdots & y_2^{N_y}(N_{Z_2}) \end{array} \right]. \quad (4.23)$$

Alternativamente, numa forma mais resumida, Z_1 e Z_2 pode ser denotado por:

$$Z_1 = \left[\mathbf{u}_1^1 \quad \mathbf{u}_1^2 \quad \cdots \quad \mathbf{u}_1^{N_u} \mid \mathbf{y}_1^1 \quad \mathbf{y}_1^2 \quad \cdots \quad \mathbf{y}_1^{N_y} \right], \quad (4.24)$$

$$Z_2 = \left[\mathbf{u}_2^1 \quad \mathbf{u}_2^2 \quad \cdots \quad \mathbf{u}_2^{N_u} \mid \mathbf{y}_2^1 \quad \mathbf{y}_2^2 \quad \cdots \quad \mathbf{y}_2^{N_y} \right], \quad (4.25)$$

em que \mathbf{u}_j^i representa a série temporal de dados da i -ésima entrada do j -ésimo conjunto de treinamento.

Sistemas experimentais são limitados, em que, na maioria dos casos, existem restrições nos valores mensuráveis dos sinais de entrada e saída. Esses limites serão considerados na medição da diversidade, sendo u_i^l e u_i^h , respectivamente, os limites inferior e superior máximo atingíveis da i -ésima entrada.

Estabelecida a notação, o procedimento para medição da diversidade pode ser sintetizado nos passos mostrados a seguir e ilustrados na Figura 4.17.

1. Para cada série temporal de entrada dos dados de treinamento \mathbf{u}_1^1 e \mathbf{u}_2^1 , faça: $\mathbf{u}_a = \mathbf{u}_1^1 * \mathbf{u}_2^1$, em que $*$ é o operador *convolução*. A fim de facilitar o entendimento, a partir desse ponto, até o passo 7, o índice i será omitido. Vale ressaltar que a intenção dessa convolução entre os sinais é reforçar, no domínio da frequência, o conteúdo espectral comum às entradas em Z_1 e Z_2 .
2. Retira-se a média do sinal $\mathbf{u}_b = \mathbf{u}_a - \mu_{u_a}$, sendo μ_{u_a} a média temporal de \mathbf{u}_a .
3. Faz-se a transformada de Fourier (FFT) de \mathbf{u}_b e obtém-se o módulo A e fase Θ no domínio da frequência.
4. A fim de obter um sinal, na frequência, com espectro complementar ao de A , faz-se a seguinte manipulação: $A_c = 1 - A/\max(A)$.
5. Com o sinal, na frequência, de módulo A_c e fase Θ , faz-se então a transformada inversa de Fourier (iFFT) e obtém-se $u_c = \text{iFFT}(A_c, \Theta)$. Note que, nesse momento, se tem um sinal com pouco conteúdo espectral nas frequências comuns entre \mathbf{u}_1 e \mathbf{u}_2 . Resta ajustar sua faixa de operação.
6. Aqui, considera-se que a faixa de operação de \mathbf{u}_1 e \mathbf{u}_2 sejam próximas, com média μ_u e variância σ_u^2 . Com esses valores, obtém-se a faixa de operação complementar por meio do seguinte procedimento, apresentado em pseudo-código:

Se $|\mu_u - u_l| > |u_h - \mu_u|$ **então:**

$$\mu_d = \frac{\mu_u - 2\sigma_u - u_l}{2} + u_l$$

$$\sigma_d = \frac{\mu_d - u_l}{2}$$

senão

$$\mu_d = \frac{u_h - \mu_u - 2\sigma_u}{2} + \mu_u + 2\sigma_u$$

$$\sigma_d = \frac{\mu_d - \mu_u - 2\sigma_u}{2}$$

fim se

7. Ajusta-se a média e variância de \mathbf{u}_c para os valores obtidos μ_d e σ_d^2 , respectivamente, e obtém-se o sinal de entrada \mathbf{u}_d .
8. O procedimento descrito até aqui deve ser realizado para todas as entradas em Z_1 e Z_2 . Assim, obtém-se o conjunto de entradas $\mathbf{U}_d = [\mathbf{u}_d^1 \ \mathbf{u}_d^2 \ \dots \ \mathbf{u}_d^{N_u}]$ para medição da diversidade.
9. Aplica-se os sinais de entrada U_d em M_1 e M_2 , e obtém-se as respostas dos modelos \mathbf{Y}_1 e \mathbf{Y}_2 , respectivamente.
10. Escolhe-se uma métrica de diversidade (vide Seção 4.2.1) e aplica-se às saídas dos modelos: $D_{1,2} = (\textit{diversidade})(\mathbf{Y}_1, \mathbf{Y}_2)$.

Esse procedimento é ilustrado na Figura 4.17, que faz referência aos mesmos passos supracitados. Vale lembrar que o sinal U_d , obtido para medição da diversidade, não precisa ser “rotulado”, i.e. não precisa haver sinal de saída do sistema real para esses dados, apenas as saídas dos *modelos* são utilizadas.

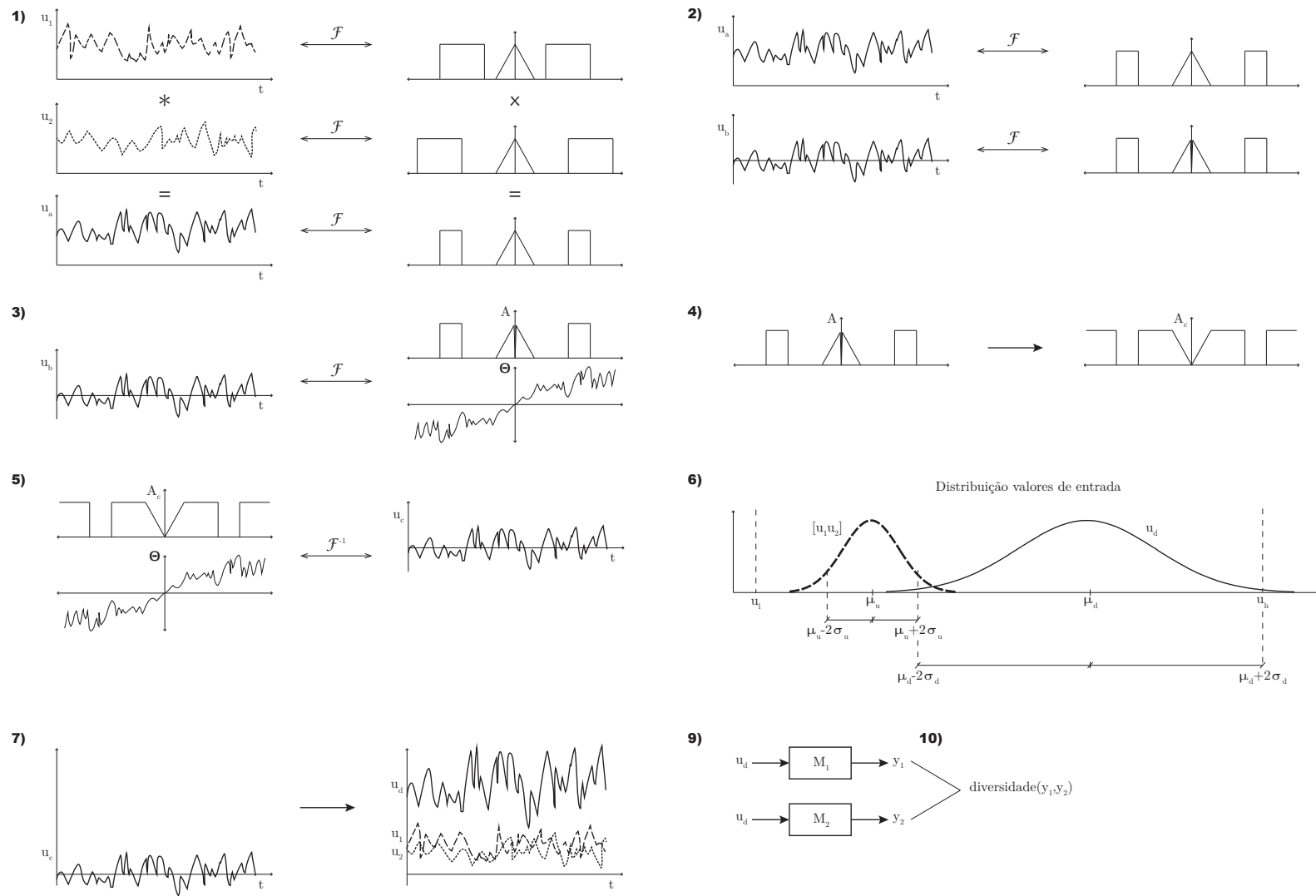


Figura 4.17: Ilustração dos passos do procedimento para medição da diversidade, apresentados na Seção 4.3.3. Os sinais aqui apresentados são apenas ilustrativos.

4.3.4 Aplicação do Procedimento

O procedimento apresentado mostrou formas de mensurar a diversidade entre dois modelos dinâmicos. No contexto de identificação de sistemas, essa medida de diversidade pode ser útil nas etapas de treinamento ou validação, por exemplo (para relação das etapas de um problema de identificação de sistemas vide [Aguirre, 2007]). Porém, neste trabalho, é considerado apenas o problema em que todos os modelos a compõem o *ensemble* já tenham sido, previamente, validados. Ou seja, são modelos que já possuem estrutura definida, parâmetros ajustados e seu desempenho validado. Assim, a medida de diversidade será aplicada apenas na formação do *ensemble*, no combinador de modelos. Não será feita nenhuma restrição quanto ao número de parâmetros, estrutura ou a ordem de cada modelo. Será considerado que é possível o acesso, para cada modelo, apenas aos dados utilizados no ajuste de seus parâmetros (i.e. treinamento, teste e validação) e sinais de saídas para um sinal de entrada qualquer.

Diversidade é uma medida relativa. Só é possível dizer que o modelo M_1 é mais diverso que M_2 , se houver algum modelo M_0 que sirva de base para essa comparação. Assim, diz-se que M_1 é mais diverso que M_2 , *em relação a M_0* (como realizado na Seção 4.3.2). É claro que, se for alterado o modelo base de comparação, já não se pode dizer mais nada a respeito, até que se faça uma nova medição. Por outro lado, é possível comparar a *diversidade média* entre dois *ensembles*, i.e. a média da diversidade de todos os modelos que compõem o *ensemble*, comparados dois a dois (como realizado na Seção 4.3.1). A informação obtida, nesse caso, seria, por exemplo, que o *ensemble* E_1 tem modelos mais diversos que o *ensemble* E_2 . Observe que isso não, necessariamente, diz respeito ao desempenho dos *ensembles*.

Outra possível aplicação da medida de diversidade é a inclusão de um novo membro em um *ensemble* já formado. Nesse caso, recomenda-se a medição da diversidade do modelo a ser adicionado em relação à resposta de cada componente do *ensemble*, individualmente, e, assim, verificar se o novo modelo é redundante ou não com algum componente. Nessa situação, não se recomenda medir diversidade entre o novo membro e a resposta do *ensemble* já formado, pois o combinador de modelos pode influenciar na medida de diversidade.

A medida de diversidade também pode auxiliar na construção do combinador de modelos. Esse assunto é abordado na próxima seção, mantendo as premissas, feitas aqui, de combinar apenas modelos válidos.

4.4 Combinadores de Modelos

A diversidade também pode ser útil na construção do *ensemble*. Métodos que otimizam, explicitamente, alguma métrica de diversidade durante a formação do *ensemble*, por meio do treinamento de seus membros (e.g. redes neurais), têm se mostrado eficazes (e.g. [Brown et al., 2005a]). Porém, neste trabalho, para a formação de *ensembles*, será considerado o problema em que os modelos já foram, previamente, treinados e são considerados modelos *válidos* (i.e. modelos que atingiram um desempenho aceitável em dados de validação). Nesse tipo de problema não há restrições quanto à estrutura do modelo, número de parâmetros ou ordem dos modelos. É pressuposto apenas o conhecimento prévio de seus dados de ajuste dos parâmetros (i.e. treinamento, teste e validação).

Vários combinadores de modelos foram propostos, desde métodos triviais (e.g. média simples) [Clemen, 1989, Bunn, 1985] a mais sofisticados [Mirchev et al., 2012]. Contudo, grande parte dos trabalhos – especialmente aqueles que usam medidas de diversidade – considera o problema em que a construção do *ensemble* é feita juntamente com o ajuste dos parâmetros dos modelos (e.g. [Brown et al., 2005a]). Traba-

lhos que consideram apenas modelos válidos na composição do *ensemble* são encontrados mais facilmente na área de previsão de séries temporais (*forecasting*), em que combinadores de modelos são usados há décadas [Bates and Granger, 1969, Newbold and Granger, 1974, Winkler and Makridakis, 1983]. Porém, nesse contexto, é mais comum a utilização de modelos autônomos (i.e. sem entradas exógenas), que não é contemplado pelo conceito de diversidade aqui trabalhado.

O combinador de modelos aqui proposto considera apenas modelos *válidos* e usa, explicitamente, uma medida de *diversidade* (proposta na Seção 4.3). Primeiro, é proposto um combinador do tipo média ponderada. Em seguida o combinador de modelos proposto é aplicado a um problema real de estimação da pressão de fundo de poço de petróleo.

4.4.1 Combinador Baseado em Diversidade

Considere $E = \{\mathcal{M}_1, \mathcal{M}_2, \dots, \mathcal{M}_M\}$ um conjunto de M modelos, *válidos*, que formarão um *ensemble*, cuja saída \hat{y}_e é dada pela média ponderada das saídas dos modelos $\{\hat{y}_1, \hat{y}_2, \dots, \hat{y}_M\}$:

$$\hat{y}_e = \sum_{i=1}^M w_i \hat{y}_i, \quad (4.26)$$

em que a soma dos pesos é $\sum w_i = 1$.

Esse combinador de modelos é usado em diversos trabalhos, em que os pesos w_i são ajustados de diversas maneiras, entre elas, de modo a minimizar a variância do erro [Bates and Granger, 1969, Newbold and Granger, 1974], a correlação entre componentes [Perrone, 1993], o erro médio quadrático [Hashem and Schmeiser, 1995], dentre outros. Neste trabalho, esses pesos serão escolhidos de acordo com a *diversidade* média de cada modelo no *ensemble*, conforme metodologia para medição de diversidade proposta na Seção 4.3, utilizando a métrica de *correlação*. Vale lembrar que a medida é feita apenas com dados de entrada, elaborados especialmente para este fim, aplicados aos modelos.

Assim, considerando a métrica de *correlação* para medição da diversidade, a atribuição dos pesos é feita conforme a seguir:

$$w_i = \frac{\bar{d}_i}{\sum_{j=1}^M \bar{d}_j}, \quad (4.27)$$

sendo \bar{d}_i a diversidade média do modelo \mathcal{M}_i em relação aos demais, dado por:

$$\bar{d}_i = \frac{1}{M-1} \sum_{i \neq j}^M d(\hat{y}_i, \hat{y}_j), \quad (4.28)$$

em que $d(\)$ é a métrica de diversidade empregada (vide Seção 4.2.1), \hat{y}_i a saída do i -ésimo modelo para os dados de medida de diversidade aplicados na entrada. Nessa ponderação, considera-se que a métrica $d(\)$ utilizada apresenta valores crescentes de acordo com o aumento da diversidade. Assim, a ponderação maior será naqueles modelos que, em média, são mais diversos que os demais.

Esse tipo de ponderação é uma tentativa de aprimoramento da capacidade de generalização do *ensemble* e, adicionalmente, nas regiões utilizadas no ajuste dos parâmetros dos modelos, eleva-se a precisão da resposta. A seguir, esse combinador de modelos é aplicado a um problema real de estimação da pressão de fundo de poço de extração de petróleo em águas profundas.

Tabela 4.3: Características dos modelos que compõem o *ensemble* de estimação da pressão de fundo de poço de extração de petróleo, erro (MAPE) sobre os dados de validação e o peso atribuído pelo combinador.

	Id. do Modelo	Estrutura	Erro (validação)
1	narxpoli2-ny46tg	polinomial	2,9842
2	rbf-ny5nul16t	RBF	2,8081
3	rbf-ny5nul18t	RBF	2,3127
4	rbf-ny5nul19t	RBF	3,2566
5	rbf-ny5nul110t	RBF	2,4370
6	rbf-ny5ny128t	RBF	2,1493
7	mlp-1h12n	MLP	4,8864
8	mlp-1h13n	MLP	4,4297
9	mlp-1h15n	MLP	3,7713
10	mlp-1h17n	MLP	3,2676
11	mlp-1h110n	MLP	2,6783
12	mlp-1h113n	MLP	2,2695
13	mlp-1h115n	MLP	2,1188
14	mlp-1h120n	MLP	2,0464
15	mlp-1h125n	MLP	2,2477
16	mlp-1h130n	MLP	2,6556
17	mlp-1h12nL	MLP	6,7834
18	mlp-1h13nL	MLP	6,7707
19	mlp-1h15nL	MLP	6,7441
20	mlp-1h17nL	MLP	6,6846
21	mlp-1h110nL	MLP	6,6451
22	mlp-1h113nL	MLP	6,6119
23	mlp-1h115nL	MLP	6,5905
24	mlp-1h120nL	MLP	6,4864
25	mlp-1h125nL	MLP	6,5207
26	mlp-1h130nL	MLP	6,4609

4.4.2 Aplicação no Sistema de Extração de Petróleo

Um problema comum na extração de petróleo em águas profundas é a estimação da pressão de fundo de poço (PDG-P), a partir das medições feitas na plataforma (vide Capítulo 2). Foram concebidos modelos dinâmicos cujas entradas são as medidas disponíveis na plataforma e a saída a pressão de fundo. Diferentes tipos de estrutura, números de entradas, valor dos atrasos utilizados como entrada e técnicas de modelagem foram empregados. Esse conjunto de modelos foi submetido ao procedimento descrito na Seção 4.3.3, para a formação do *ensemble*. A Tabela 4.3 mostra algumas características dos modelos utilizados.

No *ensemble*, a grande maioria dos modelos é composta por redes neurais do tipo MLP (*Multilayer Perceptron*), com estruturas bastante semelhantes. Em contrapartida, há apenas um modelo polinomial. É natural imaginar que modelos com estruturas muito parecidas geram padrões de generalização semelhantes, o que pode indicar, por exemplo, que o modelo polinomial tem uma diversidade média maior em relação aos demais modelos desse *ensemble*. Observa-se que, neste caso, uma média simples faria com que a saída do *ensemble* ficasse muito mais parecida com a dos modelos MLPs do que com o modelo

polinomial, por exemplo.

A diversidade dos modelos foi avaliada conforme procedimento citado na Seção 4.3.3, em que os dados usados para avaliação da diversidade dos modelos são complementares aos utilizados no ajuste dos parâmetros (i.e. treinamento e teste). Uma terceira massa de dados foi utilizada na avaliação do desempenho do *ensemble*. Esse desempenho foi avaliado segundo a métrica de erro conhecida como MAPE (*Mean Absolute Percentage Error*), que representa o erro percentual absoluto médio, dado por:

$$e = \frac{100}{N} \sum_{t=1}^N \left| \frac{y(t) - \hat{y}_e(t)}{y(t)} \right|, \quad (4.29)$$

em que N é o número de amostras, $y(t)$ o valor real da variável e \hat{y} o valor estimado.

Resultados

A Figura 4.18 mostra a pressão de fundo de poço real e a estimada por dois diferentes métodos: média simples (E_1) e a média ponderada pela diversidade (E_2). Nota-se que o *ensemble* ponderado pela diversidade supera, em desempenho, a média simples das saídas dos modelos. Enquanto o MAPE de E_1 é de 6,4%, o de E_2 é de 3,5%. Apesar dessa diferença de desempenho, observa-se que as respostas de ambos são, dinamicamente, parecidas, o que não é surpresa, já que a diferença entre os dois *ensembles* está apenas na ponderação da saída dos modelos. Assim, não se espera disparidades muito significativas no comportamento dinâmico do *ensemble*.

O *ensemble* de média simples pode fazer com que os erros dos modelos não diversos sejam reforçados. A busca por diversidade tende a reduzir a correlação dos ruídos, evitando que esse tipo de erro se eleve. Como os modelos, em sua grande maioria, são formados por redes neurais MLP, é possível que erros intrínsecos a essa estrutura sejam reforçados na média simples (E_1), ocasionando o pior desempenho em relação à ponderação que busca a diversidade (E_2). Por outro lado, se todos os modelos componentes do *ensemble* forem diversos, a ponderação proposta pouco (ou nada) melhoraria em relação à média simples das saídas.

A Figura 4.19 mostra um gráfico com a ponderação feita pelo combinador baseado em diversidade, cuja ordem dos modelos é a mesma apresentada na Tabela 4.3. Nota-se que o combinador atribui uma importância maior aos modelos polinomiais e RBFs, onde verifica-se valores maiores dos pesos. Esse resultado corrobora a ideia de que as redes neurais MLP possuem padrões de generalização semelhantes, com baixa diversidade entre si e, nesse caso, recebem uma ponderação menor.

Outro ponto interessante de se ressaltar é que a ponderação proposta não assegura o melhor desempenho possível do *ensemble* em quaisquer dados de entrada/saída — o que é extremamente difícil de se garantir. Nos dados apresentados na Fig. 4.18, por exemplo, o modelo `mlp-1h120n` obteve o melhor desempenho, com erro de aproximadamente 2,0%, índice que não foi atingido por nenhum dos dois combinadores apresentados. Porém, isso não significa que esse modelo seja o mais exato em todas as situações, sua escolha poderia causar erros mais elevados em outros intervalos. Por outro lado, o erro médio cometido pelos modelos foi de 4,3%, cujo combinador E_1 fica aquém, mas é superado pelo combinador que busca a diversidade, E_2 , com erro de 3,5%.

O papel do método aplicado é, principalmente, evitar que os erros simultâneos (i.e. correlacionados) em modelos não diversos sejam reforçados, o que aprimora a capacidade de generalização do *ensemble*. Nos casos em que todos os modelos a serem combinados forem, garantidamente, diversos, o método não proverá ganhos significativos.

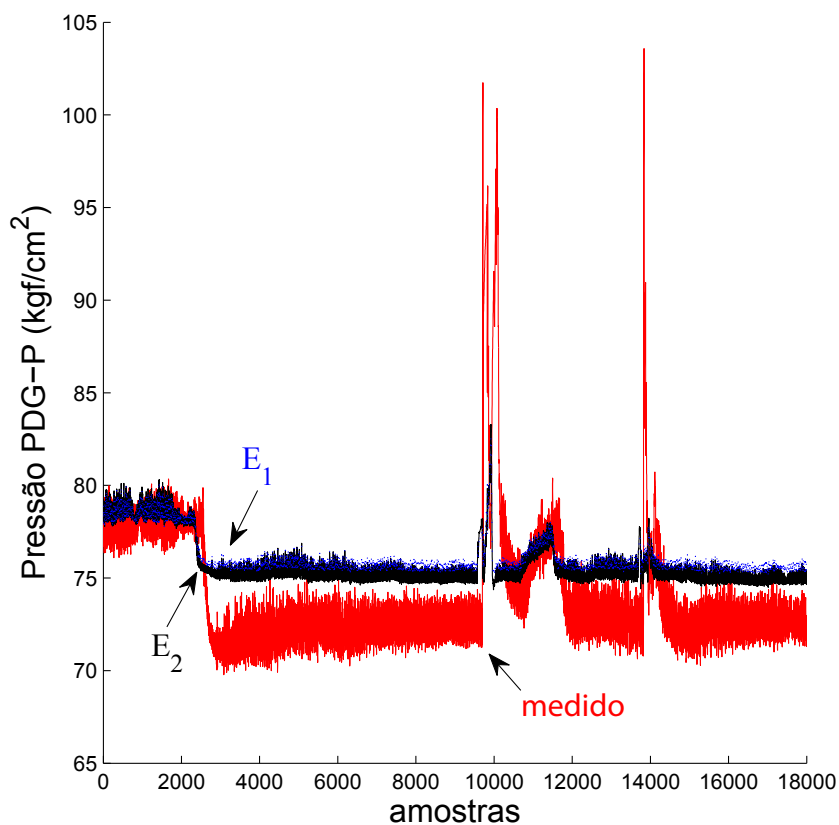


Figura 4.18: Simulação livre de dois *ensembles* compostos pelos mesmos modelos (vide Tabela 4.3), mas com ponderações distintas. O sinal indicado por E_1 representa a média simples das saídas dos modelos; E_2 representa a média ponderada pela diversidade, conforme Eq. 4.27; medido representa o valor da pressão PDG-P. Usando a métrica de erro definida na Eq. 4.29 (MAPE), tem-se: $e_{E_1} = 6,4\%$, $e_{E_2} = 3,5\%$.

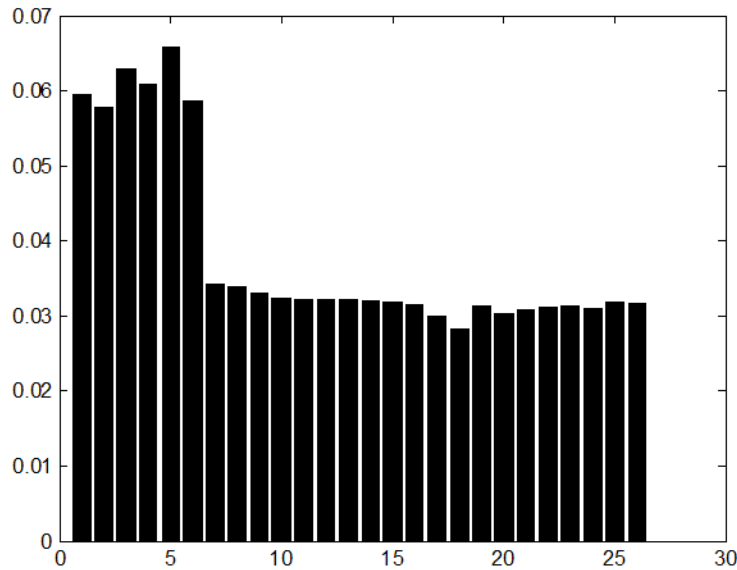


Figura 4.19: Pesos atribuídos aos modelos de acordo com a diversidade. A ordem dos modelos é a mesma apresentada na Tabela 4.3.

4.5 Conclusões do Capítulo

Este capítulo discutiu questões sobre a diversidade de modelos dinâmicos. Foram discutidas questões conceituais, desde o uso do termo *diversidade* até algumas das principais formas de medi-la. A fim de ampliar o entendimento sobre a questão, foram realizados ensaios em modelos dinâmicos lineares e, posteriormente, em sistemas não lineares. Como resultado dos ensaios, foi derivado um método de medição da diversidade entre dois modelos dinâmicos, aplicados a um problema de identificação. A partir desse método, foi sugerido um combinador de modelos baseado nessa medida de diversidade, que distribui pesos num *ensemble* de modelos dinâmicos *válidos* de acordo com sua diversidade no ensemble. Esse combinador de modelos foi aplicado em um sistema de extração de petróleo em águas profundas. Como resultado, nos dados de avaliação, obteve-se um *ensemble* que superou o erro médio dos modelos e também o combinador trivial, i.e. média simples da saída dos modelos.

Foi mostrado que, conceitualmente, o termo *diversidade* está mais associado ao erro cometido pelos modelos do que às regiões de grande acerto. A diversidade seria uma forma de evitar que falhas ocorram simultaneamente em dois sistemas, que, no contexto de identificação de sistemas, seria evitar a correlação dos erros dos modelos dinâmicos.

Foram mostradas algumas das métricas mais utilizadas para quantificar diversidade. Nos testes com sistemas dinâmicos lineares, foi visto que os dados utilizados na medida de diversidade é mais determinante do que a métrica utilizada. Assim, a escolha dos dados para medição da diversidade se torna peça chave para determinação do procedimento. Os testes com sistemas não lineares mostraram que dados semelhantes aos dados de ajuste dos parâmetros (treinamento) dos modelos não ajudam muito na medição da diversidade, que é mais interessante utilizar dados com diferentes faixas de operação e conteúdo espectral.

Baseado nessas conclusões preliminares, foi descrito um procedimento de medição de diversidade entre dois modelos. O procedimento foi utilizado na concepção de um combinador de modelos do tipo

média ponderada — pela diversidade. Este combinador foi aplicado a um sistema real de extração de petróleo em águas profundas. Nesse ensaio, o combinador de modelos obteve melhor desempenho, em termos de erro percentual absoluto médio (MAPE), em relação ao combinador *média simples* e ao MAPE médio dos modelo individuais.

O combinador de modelos proposto não assegura o melhor desempenho possível (em termos de MAPE) em todas as situações, e.g. há um modelo que sobressai dentre todos os componentes e supera o desempenho do combinador. Foi mostrado também que, nos casos em que o *ensemble* é, garantidamente, diverso, o combinador proposto não provê ganhos significativos, em relação ao combinador trivial (*média simples*), mas também não ocasiona piora expressiva no desempenho. Porém, quando a diversidade não é garantida, o combinador de modelos tem melhor desempenho do que o combinador do tipo *média simples*. O experimento mostrou ainda que, em média, o combinador de modelos tem melhor desempenho do que a simples escolha de um dos componentes. Assim, o combinador baseado em diversidade pode melhorar a *robustez* do sistema, em termos de *capacidade de generalização* e menor variabilidade do erro, principalmente em *ensembles* cuja diversidade dos componentes não é garantida.

Considerações Finais

Esse trabalho fez um estudo de duas possíveis aplicações de redes neurais na identificação de sistemas dinâmicos, especialmente na identificação de sistemas cujos dados de identificação não são representativos para o comportamento global do sistema. Foram tratados, principalmente, casos sub-ótimo de identificação de sistemas, em que um dos maiores desafios é obter respostas aceitáveis em toda faixa de operação do sistema, com maior robustez e melhor *capacidade de generalização*. A primeira abordagem foi a adição de informação extra (comportamento estático do sistema) em modelos neurais e, a segunda, a aplicação de máquinas de comitê (*ensembles*) na modelagem de sistemas. Ambas abordagens foram motivadas pelo sistema de extração de petróleo em águas profundas, que é caracterizado pela pouca informação dinâmica disponível nos dados e um elevado grau de não linearidade do sistema. O sistema foi utilizado para aplicação dos procedimentos desenvolvidos.

No Capítulo 3 foi mostrada uma possibilidade de adicionar informação auxiliar sobre o comportamento estático do sistema em modelos neurais, mantendo-se a estrutura. Foram discutidas questões acerca dos comportamentos estático e dinâmico do modelo que, embora não sejam objetivos comprovadamente conflitantes, o ajuste aos dados pode ser balanceado através do método proposto.

A aplicação da metodologia em dois sistemas experimentais mostrou, na prática, uma situação em que o método não ocasionou melhorias significativas (sistema hidráulico de bombeamento) e outro em que a informação extra trouxe benefícios consideráveis (sistema de extração de petróleo em águas profundas). Concluiu-se que o treinamento com adição da informação extra sobre a curva estática é mais vantajoso quando o comportamento em regime permanente do sistema não está bem representado nos dados de treinamento dinâmico. Caso contrário, a adição da informação não provê ganhos expressivos, mas também, nos casos apresentados, não acarreta perdas significativas de desempenho.

No Capítulo 4 foi apresentado um estudo sobre a combinação de modelos dinâmicos, de qualquer natureza, com foco na diversidade dos componentes. Foi realizada a busca de uma possível origem do termo no contexto de identificação de sistemas e em áreas relacionadas. Baseado nessa busca da origem conceitual do termo, foi mostrado que a diversidade de componentes está mais associada ao erro dos modelos do que às regiões em que os modelos têm bom desempenho. A ideia principal da diversidade é a de evitar que falhas ocorram simultaneamente.

Foram estudadas algumas métricas de diversidade que, essencialmente, utilizam dados na medição. Foi mostrado, utilizando sistemas lineares, que a escolha dos dados para medição da diversidade é mais determinante para a medição da diversidade do que a própria escolha da métrica. Foi mostrado ainda, utilizando sistemas não lineares, que dados semelhantes aos dados de treinamento dos modelos não ajudam muito na aferição da diversidade entre modelos. Dados complementares aos dados de treinamento, i.e. diferentes faixas de operação e conteúdo espectral, são mais adequados à medição da diversidade.

Baseado nos conceitos de diversidade e nos estudos com sistemas lineares e não lineares, foi concebido

um procedimento de medição da diversidade entre dois modelos. A metodologia de medição de diversidade ainda foi utilizada para dar origem a um combinador de modelos, que consiste em um ponderador baseado nas medições de diversidade. Esse combinador de modelos foi aplicado no sistema de extração de petróleo em águas profundas e obteve desempenho superior à média simples dos modelos. Porém, sabe-se que, nos casos em que os modelos forem, comprovadamente, diversos, a aplicação do combinador proposto não acarretará perdas nem ganhos significativos. O combinador proposto não assegura o melhor desempenho possível em todas as situações, mas mostrou-se melhor do que a simples escolha de um dos modelos ou a média simples dos modelos.

Este trabalho mostrou duas vertentes do desenvolvimento e inserção das redes neurais em problemas de identificação de sistemas dinâmicos, essencialmente nos casos em que os dados de identificação não são representativos. Foram realizados estudos nas áreas envolvidas e ensaios práticos que visavam comprovar os conceitos aqui defendidos. Mostrou-se que, ainda que seja utilizada uma solução “genérica”, a inclusão do conhecimento específico de um sistema pode ser muito bem vinda. Mostrou-se também que a combinação de modelos dinâmicos pode ser refinada, aplicando-se conceito de diversidade, que pode aprimorar a robustez e a capacidade de generalização do modelo.

Possíveis trabalhos futuros podem ser desenvolvidos mesclando-se os dois métodos aqui estudados, em que a diversidade dos modelos resultantes do procedimento com uso de informação auxiliar seja averiguada e que, possivelmente, sejam utilizados na formação de um *ensemble*. Nesse caso, a validação dos modelos com uso de informação auxiliar deveria ser realizada, já que o segundo procedimento trabalha com modelos válidos.

Os conceitos de diversidade estudados poderiam ser desenvolvidos para outros casos, não abordados aqui, como, por exemplo, para sistemas caóticos, autônomos ou contínuos. Aplicações em sistemas experimentais, utilizando esses tipos de modelos, também poderiam ser desenvolvidas.

Outra possível abordagem do conceito de diversidade seria a aplicação da metodologia aqui proposta no treinamento de componentes para um *ensemble*. Nessa abordagem, a diversidade seria utilizada para treinar modelos que comporão um *ensemble*, ou seja, os modelos seriam formados utilizando o conceito de diversidade. Não seriam trabalhados apenas modelos válidos, como é o caso deste trabalho.

Referências Bibliográficas

- [Ablow and Kaylor, 1965] Ablow, C. and Kaylor, D. (1965). A committee solution of the pattern recognition problem (corresp.). *Information Theory, IEEE Transactions on*, 11(3):453 – 455.
- [Abreu et al., 2012] Abreu, L. F., Barbosa, B. H. G., Teixeira, A. F., and Aguirre, L. A. (2012). Uso de informação auxiliar no treinamento de redes neurais: um procedimento bi-objetivo. In *CBA 2012*, pages 1719–1726, Campina Grande, PB.
- [Aguirre, 2007] Aguirre, L. A. (2007). *Introdução à Identificação de Sistemas – Técnicas Lineares e Não-Lineares Aplicadas a Sistemas Reais*. Editora UFMG, 3 edition.
- [Aguirre, 2011] Aguirre, L. A. (2011). Plano de trabalho. Technical report, Universidade Federal de Minas Gerais.
- [Aguirre et al., 2007] Aguirre, L. A., Alves, G. B., and Corrêa, M. V. (2007). Steady-state performance constraints for dynamical models based on rbf networks. *Engineering Applications of Artificial Intelligence*, 20(7):924 – 935.
- [Aguirre et al., 2010] Aguirre, L. A., Barbosa, B. H., and Braga, A. P. (2010). Prediction and simulation errors in parameter estimation for nonlinear systems. *Mechanical Systems and Signal Processing*, 24(8):2855 – 2867.
- [Aguirre et al., 2004a] Aguirre, L. A., Barroso, M. F. S., Saldanha, R. R., and Mendes, E. M. A. M. (2004a). Imposing steady-state performance on identified nonlinear polynomial models by means of constrained parameter estimation. *Control Theory and Applications, IEE Proceedings -*, 151(2):174 – 179.
- [Aguirre et al., 2004b] Aguirre, L. A., Lopes, R. A. M., Amaral, G. F. V., and Letellier, C. (2004b). Constraining the topology of neural networks to ensure dynamics with symmetry properties. *Phys. Rev. E*, 69(026701).
- [Amaral, 2001] Amaral, G. F. V. (2001). *Uso de Conhecimento A Priori e Redes Neurais na Identificação de Sistemas Dinâmicos Não-Lineares*. Dissertação de Mestrado, PPGEE, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil.
- [Araújo, 2003] Araújo, G. C. S. (2003). Instrumentação de uma plataforma de bombeamento para ensaios de máquinas hidráulicas.
- [Araújo, 2012] Araújo, K. C. P. (2012). Desenvolvimento de uma ferramenta para identificação de sinais correlacionados: Aplicação em dados provenientes de uma plataforma de petróleo.

- [Avizienis, 1984] Avizienis, A. (1984). Design diversity: an approach to fault tolerance of design faults. In *Proceedings of the July 9-12, 1984, national computer conference and exposition, AFIPS '84*, pages 163–171, New York, NY, USA. ACM.
- [Baffes and Zelle, 1992] Baffes, P. T. and Zelle, J. M. (1992). Growing layers of perceptrons: Introducing the extenatron algorithm. In *Proceedings of 1992 IEEE/INNS International Joint Conference on Neural Networks—Baltimore*, pages 392–397.
- [Barbosa, 2006] Barbosa, B. H. G. (2006). Instrumentação, modelagem, controle e supervisão de um sistema de bombeamento de água e módulo turbina-gerador. Master’s thesis, Programa de Pós-graduação em Engenharia Elétrica da UFMG.
- [Barbosa, 2009] Barbosa, B. H. G. (2009). *Computação Evolucionária e Máquinas de Comitê Na Identificação de Sistemas Não-lineares*. Tese de doutorado, PPGEE, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil.
- [Barbosa et al., 2011] Barbosa, B. H. G., Aguirre, L. A., Martinez, C. B., and Braga, A. P. (2011). Black and gray-box identification of a hydraulic pumping system. *Control Systems Technology, IEEE Transactions on*, 19(2):398–406.
- [Bartlett, 1994] Bartlett, E. B. (1994). Dynamic node architecture learning: An information theoretic approach. *Neural Networks*, 7(1):129–140.
- [Bates and Granger, 1969] Bates, J. M. and Granger, C. W. J. (1969). The Combination of Forecasts. *OR*, 20(4):451–468.
- [Beverage and Peterson, 1931] Beverage, H. H. and Peterson, H. O. (1931). Diversity receiving system of r.c.a. communications, inc., for radiotelegraphy. *Proceedings of the Institute of Radio Engineers*, 19(4):529–561.
- [Bohlin, 1991] Bohlin, T. (1991). *Interactive system identification: prospects and pitfalls*. Communications and control engineering series. Springer-Verlag.
- [Bravo and Normey-Rico, 2009] Bravo, C. O. A. and Normey-Rico, J. E. (2009). Controle de plantas não lineares utilizando controle preditivo linear baseado em modelos locais. *Sba: Controle & Automação Sociedade Brasileira de Automatica*, 20:465–481.
- [Brennan, 1959] Brennan, D. (1959). Linear Diversity Combining Techniques. *Proceedings of the IRE*, 47(6):1075–1102.
- [Brown, 2004] Brown, G. (2004). *Diversity in Neural Network Ensembles*. PhD thesis, School of Computer Science, The University of Birmingham.
- [Brown et al., 2005a] Brown, G., Wyatt, J., Harris, R., and Yao, X. (2005a). Diversity creation methods: A survey and categorisation. *Journal of Information Fusion*, 6(1):5–20.
- [Brown et al., 2005b] Brown, G., Wyatt, J. L., and Tiño, P. (2005b). Managing diversity in regression ensembles. *J. Mach. Learn. Res.*, 6:1621–1650.
- [Buell, 1958] Buell, C. E. (1958). Meaning of combined climate and persistence forecasts. *Journal of Meteorology*, 15(6):564–565.
- [Bunn, 1985] Bunn, D. W. (1985). Statistical efficiency in the linear combination of forecasts. *International Journal of Forecasting*, 1(2):151–163.

- [Caixeiro, 2003] Caixeiro, L. F. A. (2003). Análise, pré-projeto, projeto e re-projeto e orçamento de uma plataforma de bombeamento.
- [Canuto et al., 2007] Canuto, A. M. P., Abreu, M. C. C., de Melo Oliveira, L., Xavier, Jr., J. a. C., and Santos, A. d. M. (2007). Investigating the influence of the choice of the ensemble members in accuracy and diversity of selection-based and fusion-based methods for ensembles. *Pattern Recogn. Lett.*, 28(4):472–486.
- [Castro, 2013] Castro, W. S. (2013). Estimação recursiva da pressão de fundo de poço para extração de petróleo em águas profundas.
- [Chen et al., 1990] Chen, S., Billings, S. A., and Grant, P. M. (1990). Non-linear system identification using neural networks. *International Journal of Control*, 51(6):1191–1214.
- [Chen et al., 2008] Chen, S., Wolfgang, A., Harris, C. J., and Hanzo, L. (2008). Symmetric RBF classifier for nonlinear detection in multiple-antenna-aided systems. *IEEE Trans. Neural Netw.*, 19(5):737–745.
- [Clemen, 1989] Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, 5(4):559–583.
- [Connally et al., 2007] Connally, P., Li, K., and Irwing, G. W. (2007). Prediction- and simulation-error based perceptron training: Solution space analysis and a novel combined training scheme. *Neurocomputing*, 70:819–827.
- [Corrêa, 2001] Corrêa, M. V. (2001). *Identificação Caixa-Cinza de Sistemas Não-Lineares Utilizando Representações NARMAX Racionais e Polinomiais*. PhD thesis, UFMG.
- [Corrêa et al., 2000] Corrêa, M. V., Aguirre, L. A., and Mendes, E. M. A. M. (2000). Modeling chaotic dynamics with discrete nonlinear rational models. *International Journal of Bifurcation & Chaos in Applied Sciences & Engineering*, 10(5):1019.
- [Corrêa et al., 2002] Corrêa, M. V., Aguirre, L. A., and Saldanha, R. R. (2002). Using prior knowledge to constrain parameter estimates in nonlinear system identification. *IEEE Trans. Circuits Syst. I*, 49(9):1376–1381.
- [Croll et al., 1995] Croll, P. R., Sharkey, A. J. C., Bass, J. M., Sharkey, N. E., and Fleming, P. J. (1995). Dependable, intelligent voting for real-time control software. *Engineering Applications of Artificial Intelligence*, 8(6):615 – 623.
- [Cubillos et al., 1996] Cubillos, F. A., Alvarez, P. I., Pinto, J. C., and Lima, E. L. (1996). Hybrid-neural modeling for particulate solid drying processes. *Powder Technology*, 87(2):153–160.
- [Cybenko, 1989] Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, 2:303–314.
- [Drucker et al., 1994] Drucker, H., Cortes, C., Jackel, L. D., LeCun, Y., and Vapnik, V. (1994). Boosting and other ensemble methods. *Neural Comput.*, 6(6):1289–1301.
- [Duarte, 2004] Duarte, E. A. C. (2004). Projeto e construção dos componentes de uma turbina hidráulica do tipo francis.
- [Fahlman and Lebiere, 1990] Fahlman, S. E. and Lebiere, C. (1990). The cascade-correlation learning architecture. In *Advances in Neural Information Processing Systems 2*, pages 524–532. Morgan Kaufmann.

- [Freund and Schapire, 1997] Freund, Y. and Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119 – 139.
- [Garcia, 2005] Garcia, C. (2005). *Modelagem e Simulação de Processos Industriais e de Sistemas Eletromecânicos Vol. 1*. EDUSP.
- [Garcia-Pedrajas et al., 2003] Garcia-Pedrajas, N., Hervás-Martinez, C., and Muñoz-Perez, J. (2003). Covnet: a cooperative coevolutionary model for evolving artificial neural networks. *Neural Networks, IEEE Transactions on*, 14(3):575 – 596.
- [Graybill and Deal, 1959] Graybill, F. A. and Deal, R. B. (1959). Combining unbiased estimators. *Biometrics*, 15(4):pp. 543–550.
- [Hansen and Salamon, 1990] Hansen, L. K. and Salamon, P. (1990). Neural network ensembles. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12:993–1001.
- [Hashem, 1994] Hashem, S. (1994). Optimal linear combinations of neural networks. *Neural Networks*, 10(4):599–614.
- [Hashem and Schmeiser, 1995] Hashem, S. and Schmeiser, B. (1995). Improving model accuracy using optimal linear combinations of trained neural networks. *Neural Networks, IEEE Transactions on*, 6(3):792 – 794.
- [Herbert and Tulleken, 1993] Herbert, J. A. and Tulleken, A. F. (1993). Grey-box modeling and identification using physical knowledge and bayesian techniques. *Automatica*, 29(2):285–308.
- [Jansen et al., 2008] Jansen, J.-D., Bosgra, O. H., and den Hof, P. M. V. (2008). Model-based control of multiphase flow in subsurface oil reservoirs. *Journal of Process Control*, 18(9):846 – 855.
- [Joerding and Meador, 1991] Joerding, W. H. and Meador, J. L. (1991). Encoding a priori information in feedforward networks. *Neural Netw.*, 4(6):847–856.
- [Johansen, 1996] Johansen, T. A. (1996). Identification of non-linear systems using empirical data and prior knowledge - an optimization approach. *Automatica*, 32:337.
- [Karplus, 1977] Karplus, W. J. (1977). The spectrum of mathematical modeling and systems simulation. *SIGSIM Simul. Dig.*, 9(1):32–38.
- [Kerschen et al., 2006] Kerschen, G., Worden, K., Vakakis, A. F., and Golinval, J.-C. (2006). Past, present and future of nonlinear system identification in structural dynamics. *Mechanical Systems and Signal Processing*, 20(3):505 – 592.
- [Krogh and Vedelsby, 1995] Krogh, A. and Vedelsby, J. (1995). Neural network ensembles, cross validation, and active learning. In *Advances in Neural Information Processing Systems*, pages 231–238. MIT Press.
- [Lima, 2004] Lima, C. A. M. (2004). *Comitê de Máquinas: Uma Abordagem Unificada Empregando Máquinas de Vetores-Suporte*. PhD thesis, Faculdade de Engenharia Elétrica e de Computação, Universidade Estadual de Campinas.
- [Lima, 2012] Lima, G. G. (2012). Integração de modelos matemáticos ao sistema br-optimus: Implementação de um sensor virtual de pressão.

- [Liu and Yao, 1999] Liu, Y. and Yao, X. (1999). Ensemble learning via negative correlation. *Neural Networks*, 12:1399–1404.
- [Melville and Mooney, 2003] Melville, P. and Mooney, R. J. (2003). Constructing diverse classifier ensembles using artificial training examples. In *Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence (IJCAI-2003)*, pages 505–510, Acapulco, Mexico.
- [Mezard and Nadal, 1989] Mezard, M. and Nadal, J.-P. (1989). Learning in feedforward layered networks: the tiling algorithm. *Journal of Physics A: Mathematical and General*, 22(12):2191.
- [Mirchev et al., 2012] Mirchev, M., Duane, G. S., Tang, W. K. S., and Kocarev, L. (2012). Improved modeling by coupling imperfect models. *Communications in Nonlinear Science and Numerical Simulation*, 17(7):2741 – 2751.
- [Narendra and Parthasarathy, 1990] Narendra, K. S. and Parthasarathy, K. (1990). Identification and control of dynamical systems using neural networks. *Neural Networks, IEEE Transactions on*, 1(1):4–27.
- [Nepomuceno et al., 2007] Nepomuceno, E. G., Takahashi, R. H. C., and Aguirre, L. A. (2007). Multiobjective parameter estimation: Affine information and least-squares formulation. *Int. J. Control*, 80(6):863–871.
- [Nepomuceno et al., 2003] Nepomuceno, E. G., Takahashi, R. H. C., Amaral, G. F. V., and Aguirre, L. A. (2003). Nonlinear identification using prior knowledge of fixed points: a multiobjective approach. *I. J. Bifurcation and Chaos*, 13(5):1229–1246.
- [Newbold and Granger, 1974] Newbold, P. and Granger, C. W. J. (1974). Experience with forecasting univariate time series and the combination of forecasts. *Journal of the Royal Statistical Society. Series A (General)*, 137(2):pp. 131–165.
- [Nguyen, 2006] Nguyen, M. H. (2006). *Cooperative Coevolutionary Mixture of Experts A Neuro Ensemble Approach for Automatic Decomposition Of Classification Problems*. PhD thesis, School of Information Technology and Electrical Engineering.
- [Nilsson, 1965] Nilsson, N. J. (1965). *Learning machines: foundations of trainable pattern-classifying systems*. McGraw-Hill series in systems science. McGraw-Hill.
- [Norgaard, 1997] Norgaard, M. (1997). Neural network based system identification - toolbox. Technical Report 97-E-851, Technical University of Denmark.
- [Opitz and Shavlik, 1996] Opitz, D. W. and Shavlik, J. W. (1996). Generating accurate and diverse members of a neural-network ensemble. In *Advances in Neural Information Processing Systems*, pages 535–541. MIT Press.
- [Paliwal and Kumar, 2009] Paliwal, M. and Kumar, U. A. (2009). Neural networks and statistical techniques: A review of applications. *Expert Systems with Applications*, 36(1):2 – 17.
- [Pankiewicz, 2011] Pankiewicz, L. M. A. (2011). Análise de sinais e modelagem matemática aplicadas a dados de um poço de extração de petróleo.
- [Perrone, 1993] Perrone, M. P. (1993). *Improving Regression Estimation: Averaging Methods for Variance Reduction with Extensions to General Convex Measure Optimization*. Copy, Department of Physics at Brown University.

- [Romaniuk and Hall, 1993] Romaniuk, S. G. and Hall, L. O. (1993). Divide and conquer neural networks. *Neural Networks*, 6(8):1105 – 1116.
- [Rosen, 1996] Rosen, B. (1996). Ensemble learning using decorrelated neural networks. *Connection Science*, 8:373–384.
- [Schapire, 1990] Schapire, R. E. (1990). The strength of weak learnability. *Mach. Learn.*, 5(2):197–227.
- [Seshadri, 1966] Seshadri, V. (1966). Comparison of combined estimators in balanced incomplete blocks. *The Annals of Mathematical Statistics*, 37(6):pp. 1832–1835.
- [Sharkey, 1996] Sharkey, A. J. C. (1996). On combining artificial neural nets. *Connection Science*, 8:299–313.
- [Sharkey and Sharkey, 1997] Sharkey, A. J. C. and Sharkey, N. E. (1997). Combining diverse neural nets. *Knowl. Eng. Rev.*, 12(3):231–247.
- [Silva et al., 2010] Silva, I. N., Spatti, D. H., and Flauzino, R. A. (2010). *Redes Neurais Artificiais para Engenharia e Ciências Aplicadas*. Artliber, São Paulo.
- [Sollich and Krogh, 1995] Sollich, P. and Krogh, A. (1995). Learning with ensembles: How overfitting can be useful. In *NIPS*, pages 190–196.
- [Takahashi, 2007] Takahashi, R. H. C. (2007). Otimização escalar e vetorial.
- [Teixeira, 2011] Teixeira, A. F. (2011). Metodologia para a identificação de modelos matemáticos de poços de produção de petróleo. documentos de projeto, Petrobras.
- [Teixeira et al., 2012] Teixeira, B. O. S., Teixeira, A. F., Aguirre, L. A., Gomes, L. P., and Barbosa, B. H. G. (2012). Ukf-based data-driven soft sensing: A case study of a gas-lifted oil well. In *16th IFAC Symposium on System Identification*, volume 16 Part 1, pages 918–923.
- [Teixeira et al., 2000] Teixeira, R. A., Braga, A. P., Takahashi, R. C. H., and Saldanha, R. R. (2000). Improving generalization of MLPs with multi-objective optimization. *Neurocomputing*, 35:189–194.
- [Thomas, 2004] Thomas, J. E. (2004). *Fundamentos de Engenharia de Petróleo*. Interciencia.
- [Thompson and Kramer, 1994] Thompson, M. L. and Kramer, M. A. (1994). Modeling chemical processes using prior knowledge and neural networks. *AIChE J.*, 40(8):1328–1340.
- [Tulleken, 1993] Tulleken, H. J. A. F. (1993). Grey-box modelling and identification using physical knowledge and bayesian techniques. *Automatica*, 29(2):285–308.
- [Ueda and Nakano, 1996] Ueda, N. and Nakano, R. (1996). Generalization error of ensemble estimators. In *Neural Networks, 1996., IEEE International Conference on*, volume 1, pages 90 –95 vol.1.
- [Villela, 2005] Villela, M. J. R. (2005). Análise do comportamento da temperatura em sistemas de produção de petróleo: Comparação entre completação seca e molhada. Master’s thesis, Pontifícia Universidade Católica do Rio de Janeiro - PUC-Rio.
- [Winkler and Makridakis, 1983] Winkler, R. L. and Makridakis, S. (1983). The combination of forecasts. *Journal of the Royal Statistical Society. Series A (General)*, 146(2):pp. 150–157.
- [Wolpert, 1992] Wolpert, D. H. (1992). Stacked generalization. *Neural Networks*, 5:241–259.

-
- [Xie et al., 2009] Xie, W. F., Zhu, Y. Q., Zhao, Z. Y., and Wong, Y. K. (2009). Nonlinear system identification using optimized dynamic neural network. *Neurocomputing*, 72:3277–3287.
- [Yao and Liu, 2004] Yao, X. and Liu, Y. (2004). Evolving neural network ensembles by minimization of mutual information. *Int. J. Hybrid Intell. Syst.*, 1(1-2):12–21.

