

Identificação de Sistemas Não Lineares a partir de Modelos de *Hammerstein* Utilizando Aprendizagem Profunda[★]

Matheus Ferreira da Silva^{*} Péricles Rezende Barros^{**}
George Acioli Júnior^{**}

^{*} Programa de Pós-Graduação em Engenharia Elétrica, Departamento de Engenharia Elétrica, Universidade Federal de Campina Grande, PB, (e-mail: matheus.silva@ee.ufcg.edu.br)

^{**} Departamento de Engenharia Elétrica, Universidade Federal de Campina Grande, PB, (e-mail: [prbarros, georgeacioli]@dee.ufcg.edu.br)

Abstract: In this work, connections between the areas of systems identification and deep learning are explored. A method of nonlinear systems identification is proposed based on deep neural networks and Hammerstein models. The proposed approach is formulated from the separation of the linear and non-linear parts of the models. Thus, the used network structures reflect this separation strategy and the identification is performed through the training of deep neural networks. As a reference, was adopted an identification method in which orthonormal and radial basis functions are used. In general, the proposed approach was able to provide better identification results compared to the adopted reference method.

Resumo: Neste trabalho são exploradas conexões entre as áreas de identificação de sistemas e aprendizagem profunda. Um método de identificação de sistemas não lineares é proposto com base em redes neurais profundas e modelos de *Hammerstein*. A abordagem proposta é formulada a partir da separação das partes linear e não linear dos modelos. Assim, as estruturas de rede utilizadas refletem essa estratégia de separação e a identificação é realizada por meio do treinamento de redes neurais profundas. Como referência, foi adotado um método de identificação no qual são empregadas funções de base ortonormais e radiais. De maneira geral, a abordagem proposta foi capaz de fornecer melhores resultados de identificação em comparação com o método de referência adotado.

Keywords: nonlinear system identification; Hammerstein models; deep neural networks.

Palavras-chaves: identificação de sistemas não lineares; modelos de Hammerstein; redes neurais profundas.

1. INTRODUÇÃO

A identificação de sistemas diz respeito à obtenção de modelos para sistemas dinâmicos a partir de dados. Em particular, a modelagem de sistemas dinâmicos não lineares é um problema importante em muitas áreas da ciência e engenharia. Estes modelos são especialmente importantes para muitos algoritmos de controle modernos, pois possibilitam a predição do comportamento dos sistemas modelados (Schüssler et al., 2019).

Por outro lado, a aprendizagem profunda (em inglês, *deep learning*), que integra o chamado aprendizado de máquina, é um tópico bastante presente em pesquisas atuais sobre modelagem e estimação. Dessa forma, abordagens baseadas em aprendizagem profunda têm sido amplamente aplicadas para a modelagem de sistemas complexos. Isso se deve, sobretudo, à sua capacidade de identificar e capturar padrões em problemas não lineares (Gedon et al., 2021).

Assim como a identificação de sistemas, a aprendizagem profunda também lida com o problema fundamental de inferir modelos a partir de dados observados. Portanto, existem várias conexões entre esses dois tópicos. Isso possibilita que avanços importantes na área de *deep learning* possam ser empregados no desenvolvimento e no aprimoramento de técnicas de identificação de sistemas, especialmente para sistemas não lineares (Ljung et al., 2020).

Com relação à modelagem não linear orientada a blocos, os modelos de *Hammerstein* são os principais pontos de partida para a identificação de sistemas não lineares (Nelles, 2020). Em particular, nos trabalhos de Latawiec et al. (2003), Stanislawski et al. (2008) e Stanislawski e Latawiec (2008) são propostos esquemas de identificação baseados na separação das partes linear e não linear de modelos orientados a blocos. Nestas abordagens, as partes lineares dos sistemas são aproximadas por funções de bases ortonormais (*Orthonormal Basis Functions* - OBF). Já para a aproximação da parte não linear, são utilizadas funções de bases radiais (*Radial Basis Functions* - RBF).

[★] Este trabalho foi realizado com o apoio financeiro do CNPq.

Por outro lado, existem na literatura de identificação de sistemas, alguns trabalhos voltados para a aplicação de técnicas aprendizagem profunda em problemas de identificação, sobretudo para sistemas não lineares. Em (Ljung et al., 2020) e (Schüssler et al., 2019) são exploradas relações entre a área de *deep learning* e abordagens convencionais de modelagem de sistemas dinâmicos. Os resultados obtidos, de maneira geral, indicam que técnicas de aprendizagem profunda podem alcançar desempenhos comparáveis ao estado da arte dos métodos de identificação convencionais.

Em (Andersson et al., 2019) e em (Gedon et al., 2021) considera-se que os desenvolvimentos recentes na área de *deep learning* são relevantes para a modelagem de sistemas não lineares. Os resultados alcançados indicam que as técnicas de aprendizagem profunda representam uma alternativa competitiva em relação aos métodos clássicos de identificação. O que se deve principalmente à sua alta flexibilidade e capacidade de capturar uma ampla gama de dinâmicas de um sistema.

Neste artigo é proposta a aplicação de redes neurais profundas para a identificação de sistemas não lineares a partir de modelos não lineares orientados a blocos, em particular, modelos de *Hammerstein*. Esta abordagem é proposta com o intuito de avaliar os efeitos da associação de técnicas de aprendizagem profunda com estratégias clássicas identificação de sistemas não lineares.

Este artigo está organizado da seguinte maneira. Na seção 2 é apresentada a formulação do problema abordado. A seção 3 diz respeito à metodologia proposta. Já na seção 4 são apresentados e discutidos os resultados obtidos. Por fim, a seção 5 se refere à conclusão do trabalho.

2. FORMULAÇÃO DO PROBLEMA

Como consideração inicial, é ressaltado que redes neurais profundas comuns, tais como as redes *feedforward* são modelos ARX não lineares (NARX). Assim, podem facilmente serem incorporadas em códigos e práticas de identificação de sistemas. Além disso, o uso de aprendizagem profunda, em particular para a estimação de sistema dinâmicos, tem fortes ligações com a prática de identificação de sistemas (Ljung et al., 2020).

Considerando um sistema com estrutura de *Hammerstein* formado pela função linear $G_1(q)$ e por um mapeamento não linear estático da entrada $f_1(u(t))$, além de um perturbação $e_H(t)$ assim como apresentado na Figura 1, é possível escrever:

$$y(t) = G_1(q)[f_1(u(t)) + e_H(t)] . \quad (1)$$

Partindo de (1), busca-se simplificar o processo de identificação por meio da representação do modelo como uma soma de partes lineares e não lineares. Assim como em (Stanisławski e Latawiec, 2008), essa estratégia é realizada com o intuito de permitir a aplicação de técnicas mais simples para a identificação tanto do comportamento não linear do sistema, quanto da dinâmica linear.

Considerando um modelo de *Hammerstein* como mostrado em (1), e sendo $G_1(q)$ uma função de primeira ordem com um atraso conhecido, d , tal que

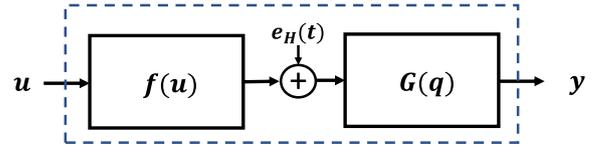


Figura 1. Representação esquemática de um sistema de *Hammerstein*.

$$G_1(q) = q^{-d} \frac{K}{q - p_1} = \frac{K}{1 - p_1 q^{-1}} q^{-(d+1)} . \quad (2)$$

Assumindo que $G_1^{-1}(q)$ existe, de modo que

$$G_1^{-1}(q)y(t) = [f_1(u(t)) + e_H(t)] . \quad (3)$$

Assim,

$$y(t) - p_1 q^{-1} y(t) = K q^{-(d+1)} [f_1(u(t))] + e_1(t) , \quad (4)$$

onde $e_1 = K q^{-(d+1)} e_H(t)$.

Dessa forma,

$$y(t) - e_1(t) = p_1 q^{-1} y(t) + K [f_1(u(t-d-1))] , \quad (5)$$

e, por fim, fazendo $\hat{y}(t) = y(t) - e_1(t)$

$$\hat{y}(t) = p_1 q^{-1} y(t) + K [f_1(u(t-d-1))] . \quad (6)$$

Reescrevendo (6) utilizando $\gamma_1 = p_1$ e $\Psi(\cdot) = K [f_1(\cdot)]$, tem-se

$$\hat{y}(t) = \underbrace{\gamma_1 y(t-1)}_{\text{parte linear}} + \underbrace{\Psi(u(t-d-1))}_{\text{parte não linear}} . \quad (7)$$

A equação (7) é uma estrutura de modelo onde há a separação entre as partes linear e não linear. Um desenvolvimento semelhante pode ser realizado para modelos de *Hammerstein* com parte linear de ordens mais elevadas. Considerando, por exemplo, a função $G_2(q)$ de segunda ordem com atraso dada por

$$G_2(q) = q^{-(d+2)} \frac{K}{1 + a_1 q^{-1} + a_0 q^{-2}} , \quad (8)$$

é obtida a seguinte estrutura de modelo (9).

$$\hat{y}(t) = \underbrace{\gamma_1 y(t-1) + \gamma_2 y(t-2)}_{\text{parte linear}} + \underbrace{\Psi(u(t-d-2))}_{\text{parte não linear}} . \quad (9)$$

Da maneira similar, a partir de uma função $G_3(q)$ de segunda ordem com atraso e com um zero dada por

$$G_3(q) = q^{-(d+1)} \frac{K(1 - b_1 q^{-1})}{(1 + a_1 q^{-1} + a_0 q^{-2})} , \quad (10)$$

é obtida a seguinte estrutura de modelo

$$\hat{y}(t) = \underbrace{\gamma_1 y(t-1) + \gamma_2 y(t-2)}_{\text{parte linear}} + \underbrace{\Psi(u(t-d-1), u(t-d-2))}_{\text{parte não linear}}. \quad (11)$$

Dessa forma, o problema considerado consiste na identificação de sistemas não lineares a partir de modelos com estruturas de *Hammerstein*. Assumindo um modelo na forma da equação (7), (9) ou (11), propõe-se o ajuste de parâmetros por meio de uma rede neural profunda. Tendo em vista a separação das partes linear e não linear dos modelos, são propostas estruturas particulares de rede neurais para a realização do ajuste de parâmetros.

3. METODOLOGIA PROPOSTA

Considerando um sistema simulado com estrutura de *Hammerstein* a metodologia proposta consiste nos seguintes passos:

- (1) obtenção de conjuntos de dados de entrada e saída do sistema de interesse;
- (2) seleção de uma estrutura de modelo dentre as apresentadas nas equações (7), (9) e (11);
- (3) escolha de uma estrutura de rede correspondente à estrutura de modelo adotada;
- (4) realização do procedimento de treinamento para diferentes escolhas estruturais em relação à parte multicamadas;
- (5) seleção do modelo mais preciso para representar o sistema.

No passo 1 é gerado um conjunto de treinamento a ser utilizado para a identificação do sistema de interesse. Além disso, é necessário também gerar um conjunto de dados de teste destinado à avaliação da precisão dos modelos ajustados.

No passo 3 são definidas as quantidades de entradas associadas às partes linear e multicamadas da rede a ser utilizada. Esta definição é realizada a partir da estrutura de modelo selecionada no passo 2.

Uma vez definida a estrutura geral da rede a partir do passo 3, o passo 4 é realizado para que se possa selecionar uma parte multicamadas capaz de representar a não linearidade do sistema. Para isso o treinamento é repetido para diferentes combinações de números de camadas ocultas e neurônios por camada.

Para a execução do passo 5, os diferentes modelos obtidos são aplicados a um conjunto de dados de teste. Foi adotado como critério de comparação o valor do erro quadrático médio (EQM) sobre os dados dos conjuntos de teste. O EQM pode ser calculado como

$$EQM = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2. \quad (12)$$

Onde, para N pontos de dados de teste, y_i é o valor real da saída e \hat{y}_i é o valor estimado para o ponto i

3.1 Sinal de excitação

Como o intuito é obter dados de entrada e saída para a identificação de sistemas não lineares, é necessário que os sinais de excitação utilizados contemplem diversos valores de amplitude dentro de uma faixa de interesse. A excitação de um sistema em diferentes pontos de operação permite que sejam obtidas informações a respeito do seu comportamento não linear, o que possibilita a sua identificação.

Dessa forma, foram adotadas excitações do tipo APRBS (*Amplitude Modulated PRBS* - PRBS Modulado em Amplitude). O sinal PRBS (*Pseudo-Random Binary Sequence* - Sequência Binária Pseudo-Aleatória), por sua vez, é amplamente utilizado para a obtenção de dados para aplicações de identificação de sistemas lineares. Entretanto, por ser um sinal binário, isto é, só apresentar duas amplitudes possíveis, o sinal PRBS não é adequado para a obtenção de dados de sistemas não lineares para fins de identificação.

Dessa maneira, para se ter um sinal com características semelhantes ao PRBS, mas que permita a identificação de comportamentos não lineares, é construído, a partir de um PRBS, um sinal com variações aleatórias de amplitude dentro de uma faixa estabelecida, resultando assim no chamado APRBS, o qual é amplamente utilizado em aplicações de identificação de sistemas não lineares (Isermann e Münchhof, 2010).

3.2 Estrutura de rede

Para exemplificar o procedimento de escolha das estruturas de rede neural podemos considerar a equação (11). Neste caso, tem-se que a saída estimada $\hat{y}(t)$ é dada pela combinação linear de valores passados da saída, $y(t-1)$ e $y(t-2)$, ponderados pelos coeficientes γ_1 e γ_2 , respectivamente. Além disso, $\hat{y}(t)$ é composto ainda por um mapeamento não linear $\Psi(\cdot)$ aplicado a valores passados da entrada, $u(t-d-1)$ e $u(t-d-2)$.

Dessa forma, para a obtenção de um modelo com a forma mostrada em (11), propõe-se utilizar uma estrutura de rede neural com uma parte multicamadas capaz de aprender o mapeamento não linear $\Psi(\cdot)$ e uma outra parte destinada a fazer o ajuste dos coeficientes γ_1 e γ_2 da parte linear. Na Figura 2 pode ser vista uma representação da estrutura de rede utilizada para o modelo em questão.

Na parte multicamadas é utilizada uma estrutura *feed-forward* com F camadas ocultas C^k , $k = 1, 2, \dots, F$, além das camadas de entrada, C^{in} , e de saída, C^{out} . Além disso, a rede é composta pelos vetores de polarização b^i e pelas matrizes de pesos W^i , $i = 1, 2, \dots, F-1$, bem como pelos conjuntos de pesos de entrada, W^{in} , e de saída, W^{out} .

No que se refere ao processo de treinamento, isto é, o ajuste dos pesos que compõem a rede como um todo, os valores passados de $u(t)$ e $y(t)$ são fornecidos como entradas da parte multicamadas e da parte linear da rede, respectivamente. Para a realização do treinamento, foi adotado um procedimento iterativo de atualização dos pesos baseado no gradiente descendente de uma função de custo $J(t)$ definida como:

$$J(t) = \frac{1}{2} [y(t) - \hat{y}(t)]^2. \quad (13)$$

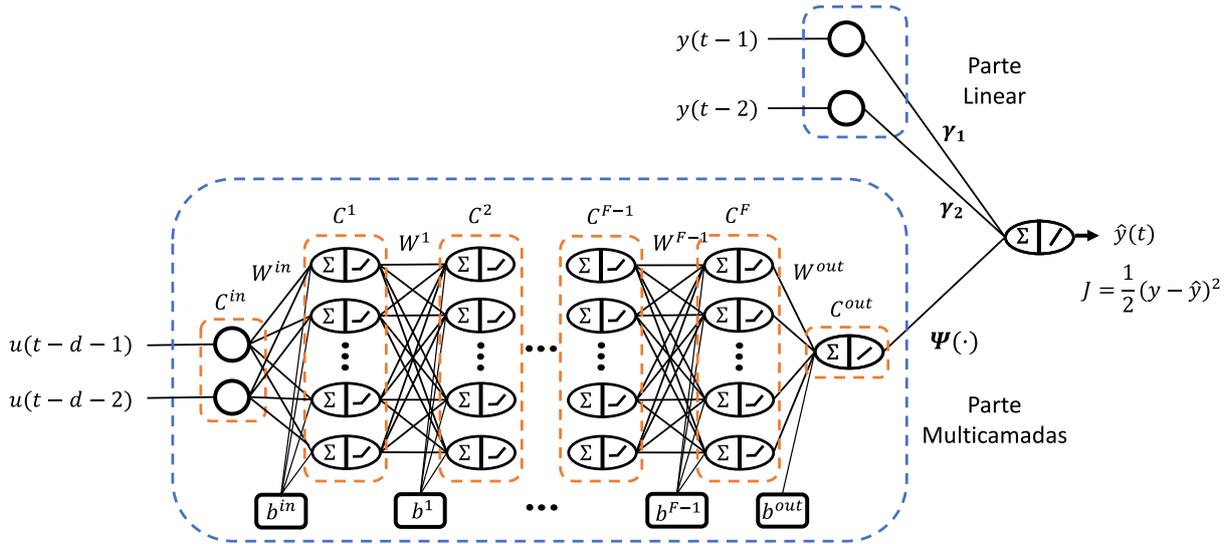


Figura 2. Ilustração da estrutura de rede adotada.

Dessa forma, para o passo de treinamento $n + 1$, cada coeficiente γ e cada peso w da parte multicamadas são atualizados conforme (14) e (15), respectivamente

$$\gamma[n + 1] = \gamma[n] - \frac{\partial J}{\partial \gamma}, \quad (14)$$

$$w[n + 1] = w[n] - \frac{\partial J}{\partial w}. \quad (15)$$

3.3 Método de referência

Para fins de comparação, foi adotado como referência o método de identificação apresentado por Stanisławski e Latawiec (2008). Nesta abordagem são utilizadas funções de bases ortonormais (*Orthonormal Basis Functions* - OBF) e funções de bases radiais (*Radial Basis Functions* - RBF) para a identificação de sistemas não lineares com estrutura de *Hammerstein* a partir da seguinte formulação

$$\hat{y}(t) = - \sum_{i=1}^M a_i L_i(q^{-1}, p) y(t) + \sum_{j=1}^m z_j \phi_j(t-d). \quad (16)$$

Onde as funções de base ortonormal, $L_i(q, p)$, são funções de *Laguerre* definidas em termos de um polo p como

$$L_i(q, p) = \frac{\sqrt{1-p^2}}{q-p} \left[\frac{1-pq}{q-p} \right]^{i-1} \quad i = 1, \dots, M. \quad (17)$$

Já as funções de base radial, $\phi_j(u(t))$, são funções gaussianas da seguinte forma

$$\phi_j(u(t)) = \exp(-b_j \|u(t)\|^2), \quad j = 1, \dots, m. \quad (18)$$

Para a obtenção de um modelo a partir da equação (16) é possível ajustar os coeficientes de ponderação a_i e z_j a partir de um conjunto de dados de entrada e saída utilizando o algoritmo dos *mínimos quadrados recursivos*.

4. RESULTADOS DE IDENTIFICAÇÃO

Nesta seção são apresentados os resultados de identificação obtidos a partir do método proposto e por meio do método de referência. De acordo com as estruturas de modelo estabelecidas, foram adotados três diferentes casos de estudo de sistemas simulados.

Para a simulação de sistemas de interesse foi utilizado o esquema geral apresentado na Figura 3 formado por uma não linearidade $f_H(u)$ e um bloco linear dinâmico $G_i(q)$, além de um sinal de ruído adicionado à saída.

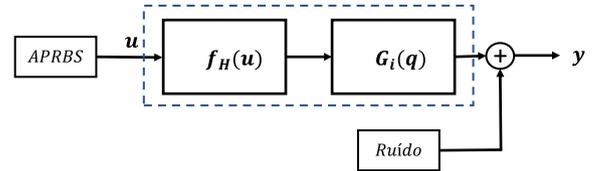


Figura 3. Esquema geral utilizado para as simulações dos sistemas de *Hammerstein*.

- No **caso 1** é considerado um sistema de *Hammerstein* formado por uma função linear $G_1(q)$ de primeira ordem com atraso e uma não linearidade $f_H(\cdot)$ dadas de acordo com (19) e (20), respectivamente

$$G_1(q) = (q^{-1}) \frac{0,5}{1 - 0,97q^{-1}}, \quad (19)$$

$$f_H(v) = v + 3v^2 + 2v^3. \quad (20)$$

- Já no **caso 2** foi adotado um sistema de *Hammerstein* formado por uma função linear $G_2(q)$, de segunda ordem com um atraso conhecido d e a mesma não linearidade $f_H(\cdot)$ empregada no caso anterior. A função de $G_2(q)$ escolhida é:

$$G_2(q) = (q^{-2}) \frac{0,5}{1 - 1,5q^{-1} + 0,65q^{-2}}. \quad (21)$$

- Por fim, no **caso 3** foi adotado um sistema de *Hammerstein* formado por uma função linear $G_3(q)$, de segunda ordem com atraso e com um zero, além da não linearidade $f_H(\cdot)$. Foi utilizada a seguinte função $G_3(q)$:

$$G_3(q) = (q^{-1}) \frac{1 + 0,5q^{-1}}{1 - 1,5q^{-1} + 0,7q^{-2}} . \quad (22)$$

A partir dos sistemas com estruturas de *Hammerstein* definidos, foram gerados conjuntos de treinamento composto por 6000 pares de valores de entrada e saída, além de conjuntos de teste com 1000 pontos de dados. Para os casos de estudo 1, 2 e 3 foram realizadas simulações utilizando-se perturbações do tipo ruído branco gaussiano (média zero) com uma variâncias iguais a 0, 1, 0,05 e 0,08, respectivamente. Os resultados obtidos para cada um dos casos são apresentados a seguir.

4.1 Caso 1

Para este caso, na Figura 4 são mostrados trechos do sinal de excitação APRBS para uma faixa de amplitude [0, 1], bem como a saída do sistema incluindo o ruído adicionado. Para este caso, o melhor ajuste para o conjunto

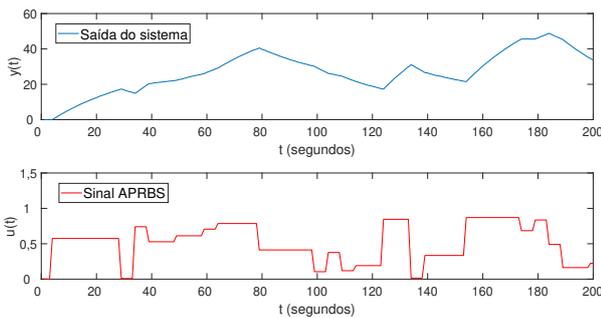


Figura 4. Exemplo de resposta do sistema para uma excitação APRBS (Caso 1).

de teste foi obtido para uma estrutura multicamadas composta por 10 camadas ocultas e 6 neurônios por camada oculta. Foi obtido, assim, o modelo expresso em (23) por meio da rede treinada. Desse modo, para a entrada $y(t - 1)$ na parte linear e $u(t - d - 1)$ na parte multicamadas, a rede é capaz de fornecer uma estimativa $\hat{y}(t)$ para a saída do sistema

$$\hat{y}(t) = 0,9709y(t - 1) + \Psi(u(t - d - 1)) . \quad (23)$$

Para este caso de estudo, tem-se que o coeficiente da parte linear, γ_1 , corresponde ao negativo do coeficiente de primeira ordem do denominador da função $G_1(q)$, isto é, 0,97. Por outro lado, o valor estimado para este parâmetro foi de 0,9709, assim como apresentado em (23).

O erro quadrático médio obtido com a estrutura de rede escolhida foi de 0,2166. Já o valor de erro obtido por meio do método de referência foi de 1,8185. Uma comparação gráfica entre a saída real do sistema e as saídas estimadas por meio de ambos os métodos pode ser vista na Figura 5.

Na Figura 6 são apresentados graficamente os valores de erro de estimação ponto a ponto para ambos os métodos.

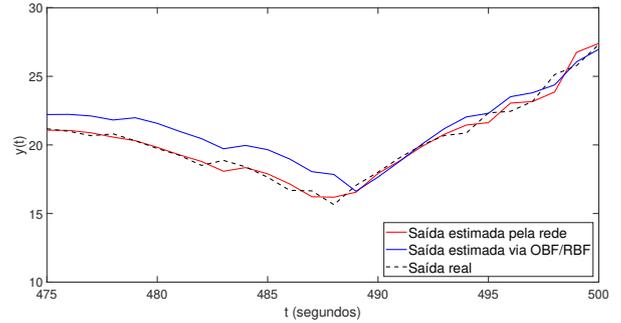


Figura 5. Comparação entre a saída real e as saídas estimadas utilizando os métodos proposto e de referência (Caso 1).

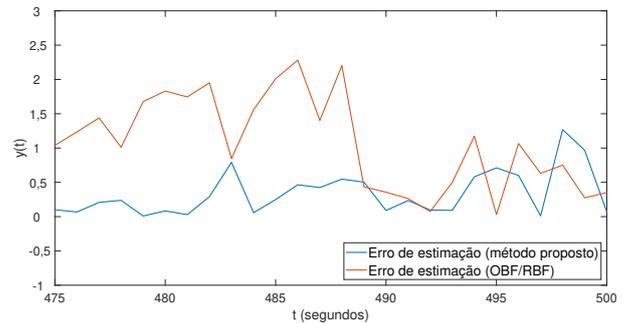


Figura 6. Erros de estimação utilizado ambos os métodos para um subconjunto de teste (Caso 1).

4.2 Caso 2

Para este caso foi selecionada uma rede com a parte multicamadas composta por 8 camadas ocultas e 7 neurônios por camada, de modo que a rede completa corresponde ao modelo dado por

$$\hat{y}(t) = 1,4528y(t - 1) - 0,6008y(t - 2) + \Psi(u(t - d - 2)) . \quad (24)$$

Tendo em vista que, para este caso, os coeficientes γ_1 e γ_2 correspondem, respectivamente, ao negativo dos coeficientes de primeira e segunda ordem do denominador da função $G_2(q)$, na Tabela 1 são mostrados os valores real e estimado desses coeficientes.

	Valor real	Valor estimado
γ_1	1,5000	1,4528
γ_2	-0,6500	-0,6008

Tabela 1. Valores reais e estimados para os coeficientes da parte linear (Caso 2).

O valor de erro quadrático médio obtido com o método proposto foi 0,2013. Por outro lado, a aplicação do método OBF/RBF de referência levou a um valor de EQM de 1,1578 sobre o conjunto de teste.

Na Figura 7 são apresentados graficamente os valores das saídas real e estimadas por meio dos métodos em questão para um subconjunto dos dados de teste. Já na Figura 8

são mostrados os erros de estimação obtidos por meio de ambos os métodos.

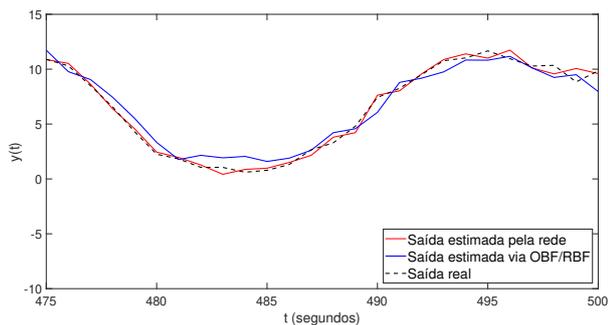


Figura 7. Comparação entre a saída real e as saídas estimadas utilizando os métodos proposto e de referência (Caso 2).

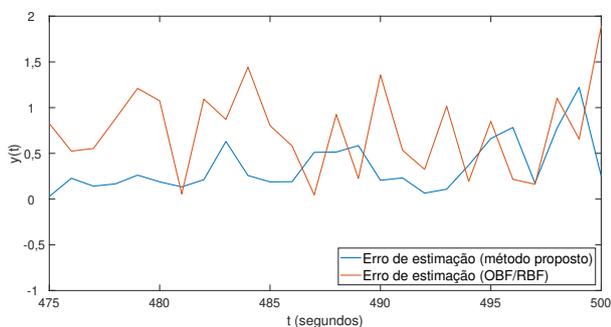


Figura 8. Erros de estimação utilizado ambos os métodos para um subconjunto de teste (Caso 2).

4.3 Caso 3

Neste caso, o modelo mais preciso obtido, mostrado na equação (25), corresponde a uma de rede com a parte multicamadas composta por 6 camadas ocultas e 10 neurônios por camada

$$\hat{y}(t) = 1,4878y(t-1) - 0,6902y(t-2) + \Psi(u(t-d-1), u(t-d-2)). \quad (25)$$

Na Tabela 2 são mostrados os valores reais e estimados dos coeficientes γ_1 e γ_2 do modelo.

	Valor real	Valor estimado
γ_1	1,5000	1,4878
γ_2	-0,7000	-0,6902

Tabela 2. Valores reais e estimados para os coeficientes da parte linear (Caso 3).

O valor de erro médio quadrático obtido com o método proposto foi 0,3370. Por outro lado, a aplicação do método OBF/RBF de referência levou a um valor de EQM de 5,6931 sobre o conjunto de teste.

Na Figura 9 são apresentados os gráficos das saídas real e estimadas por meio de ambos os métodos para um intervalo específico dos dados de teste. Já na Figura 10

são apresentados os erros de estimação obtidos a partir dos dois métodos.

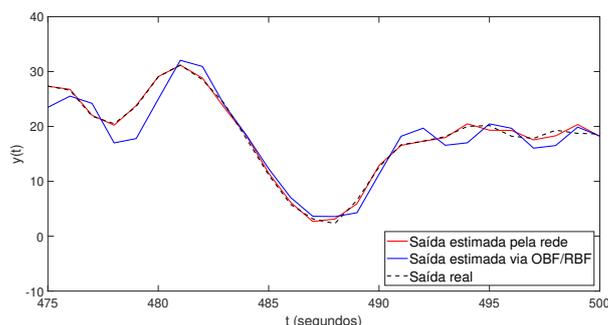


Figura 9. Comparação entre a saída real e as saídas estimadas utilizando os métodos proposto e de referência (Caso 3).

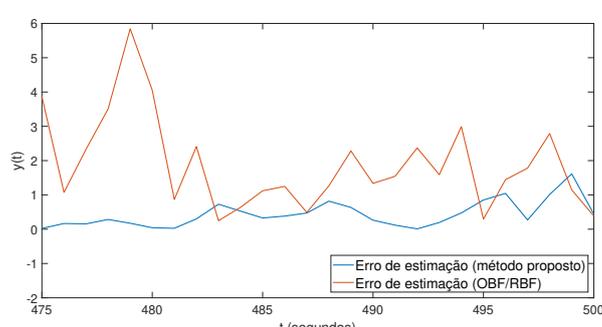


Figura 10. Erros de estimação utilizado ambos os métodos para um subconjunto de teste (Caso 3).

4.4 Discussão

Na Tabela 3 são reunidos os valores de erro médio quadrático obtidos sobre os conjuntos de teste para cada um dos casos de estudo adotados.

	Caso 1	Caso 2	Caso 3
Método proposto	0,2166	0,2013	0,3370
Método OBF/RBF	1,8185	1,1578	5,6931

Tabela 3. Valores de erro quadrático médio (EQM) obtidos.

A comparação de resultados do método proposto e o método baseado em OBF/RBF indica que o método proposto foi capaz de fornecer modelos mais precisos para os sistemas de interesse. Dessa forma, é possível afirmar que a abordagem proposta apresenta uma maior capacidade de modelagem de dinâmicas não lineares quando comparado com o método de referência.

No que se refere às estruturas de rede escolhidas no método proposto, na Tabela 4 estão reunidas as combinações selecionadas de quantidades de camadas e neurônios. Apesar de terem sido testadas estruturas com até 10 camadas ocultas, durante a seleção dos melhores modelos foi observado que o incremento da quantidade de camadas nem

sempre leva ao aumento da precisão dos modelos obtidos em relação aos dados de teste.

	Caso 1	Caso 2	Caso 3
Número de camadas ocultas	8	8	6
Número de neurônios por camada	9	7	10

Tabela 4. Quantidades de camadas e neurônios para cada caso de estudo.

Nesse sentido, é importante destacar que redes com maiores capacidades de aprendizagem conseguem se ajustar muito bem aos dados de treinamento, mesmo com a presença de ruído. Porém, este comportamento faz com que elas não alcancem bom desempenho sobre dados de teste não vistos. Esse comportamento caracteriza a ocorrência de um sobreajuste (*overfitting*) aos dados de treinamento.

5. CONCLUSÃO

Tendo em vista o objetivo de avaliar efeitos da incorporação de técnicas de aprendizagem profunda em um contexto de identificação de sistemas, a formulação proposta consistiu na integração de uma técnica de aprendizagem profunda com uma estratégia básica de modelagem amplamente empregada para tarefas de identificação de sistemas não lineares. Para fins de comparação foi adotado um método de referência no qual não são aplicadas técnicas baseadas em aprendizagem.

A partir dos testes comparativos realizados, foram observadas melhorias com relação aos valores de erro quadrático médio obtidos. Estes resultados indicam que os métodos baseados em aprendizagem podem promover melhoras nos resultados de identificação em comparação com técnicas convencionais. Entretanto, há que se considerar que a aplicação do método proposto torna o procedimento de identificação mais complexo e custoso em razão dos processos de treinamento.

Dessa forma, considera-se que a aplicação de técnicas de aprendizagem profunda para a identificação de sistemas tende a ser mais vantajosa para problemas mais complexos, nos quais se esteja mais preocupado com a efetividade dos modelos obtidos do que com o custo da sua obtenção. Assim, para situações nas quais é mais difícil obter modelos precisos por meio de técnicas convencionais, a aplicação de técnicas baseadas em aprendizagem torna-se uma alternativa mais viável.

AGRADECIMENTOS

Os autores agradecem à Universidade Federal de Campina Grande (UFCG) pelo suporte à realização deste trabalho.

REFERÊNCIAS

Andersson, C., Ribeiro, A.H., Tiels, K., Wahlström, N., e Schön, T.B. (2019). Deep convolutional networks in system identification. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, 3670–3676. IEEE.

Gedon, D., Wahlström, N., Schön, T.B., e Ljung, L. (2021). Deep state space models for nonlinear system identification. *IFAC-PapersOnLine*, 54(7), 481–486.

Isermann, R. e Münchhof, M. (2010). *Identification of dynamic systems: an introduction with applications*. Springer Science & Business Media.

Latawiec, K.J., Marciak, C., Rojek, R., e Oliveira, G.H. (2003). Linear parameter estimation and predictive control of wiener/hammerstein systems. *IFAC Proceedings Volumes*, 36(16), 345–350.

Ljung, L., Andersson, C., Tiels, K., e Schön, T.B. (2020). Deep learning and system identification. *IFAC-PapersOnLine*, 53(2), 1175–1181.

Nelles, O. (2020). *Nonlinear System Identification: From Classical Approaches to Neural Networks, Fuzzy Models, and Gaussian Processes*. Springer Nature.

Schüssler, M., Munker, T., e Nelles, O. (2019). Deep recurrent neural networks for nonlinear system identification. In *2019 IEEE Symposium Series on Computational Intelligence (SSCI)*, 448–454. IEEE.

Stanislawski, R., Hunek, W.P., e Latawiec, K.J. (2008). Modeling of nonlinear block-oriented systems using orthonormal basis and radial basis functions. In *2008 19th International Conference on Systems Engineering*, 55–58. IEEE.

Stanislawski, R. e Latawiec, K.J. (2008). Orthonormal basis and radial basis functions in modeling and identification of nonlinear block-oriented systems. *Automation and Robotics*, 277.