A Fast Algorithm for Training Dynamical Neural Networks Using Steady-State Prior Information of Offshore Oil Platform

Leandro Freitas[1], Bruno H. G. Barbosa[2], Luis A. Aguirre[3]

[1]*Departamento de Automação Industrial e Tecnologia da Informação, Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais Campus Betim*
Rua Itaguaçu, 595, 32677-564, Betim, MG, Brazil
leandro.freitas@ifmg.edu.br
[2]*Departamento de Automática, Universidade Federal de Lavras*
CP 3037, 37200-000, Lavras, MG, Brazil
brunohb@ufla.br
[3]*Departamento de Engenharia Eletrônica, Universidade Federal de Minas Gerais*
Av. Antônio Carlos 6627, 31270-901, Belo Horizonte, MG, Brazil
aguirre@ufmg.br

Abstract. In systems identification, the use of auxiliary information configures a grey-box approach. This paper describes a methodology to estimate model parameters including auxiliary information about the static behavior of the system in a bi-objective approach and discusses a decision maker based on the auxiliary information. The procedure can be applied to many model structures, such as polynomial models or multilayer perceptron (MLP) neural networks, without the need of computing the model fixed points. The grey-box modeling procedure was applied to design a soft-sensor for the downhole pressure of a real gas-lifted deep-water offshore oil well. To this end, steady-state values of the downhole pressure were estimated from historical data from (almost) stationary conditions. The available training and validation (dynamical) data had information over a limited operating range, while test data had operating ranges not present in the training and validation data. The identified dynamic models used only platform variables with a fixed MLP structure. The results indicate that the procedure yields suitable models with good static and dynamic performance. Besides, the use of auxiliary information helped to find models with better dynamical performance on operating regimes not originally represented in the dynamical data. Whereas an identified black-box MLP model obtained a root mean squared error (RMSE) of 6.7 kgf/cm^2 in a free-run simulation over test data, the proposed approach achieved an RMSE of 3.7 kgf/cm^2. This is very relevant for many practical situations where the available dynamical data does not cover all operating regimes of the system. The procedure described in this work can be applied with different model classes with greatly reduced computing time.

Keywords: artificial intelligence, grey-box identification, soft-sensors, steady-state information

1 Introduction

The main purpose of *system identification* is to build dynamic models based on experimental data. To this end, related fields of knowledge like *statistics*, *optimization*, *machine learning*, have become important tools to extract information about system dynamics from data. The use of artificial neural networks (ANN) for systems identification is another successful tool [1] especially dealing with nonlinear systems due to its properties (e.g. universal approximator) and good data fitting. In most cases, model parameters are estimated using just one source of information: the dynamical data. This will be referred to as *black box* identification.

Determining a nonlinear model from a finite set of observations without any *prior knowledge* about the system is an ill-posed problem [2], because a unique model may not exist, or it may not depend continuously on the observations. This issue is worsened when dealing with noisy signals, non-informative data (e.g. non-persistently exciting inputs) and high-dimensional systems. From an optimization point of view, the search space and the number of local minima grows indefinitely, generating an extra challenge. Hence, physical insights about the system is almost a requirement to achieve suitable models.

The use of information apart from the dynamical data constitutes a *grey-box* approach. Here, the additional information, referred to as *auxiliary information*, is assumed to be a set of steady-state data, although other alternatives exist [3–6]. A central challenge in grey-box identification is how to efficiently employ auxiliary information. This can be done in a number of ways by means of constraints, linguistic rules and others [7].

Grey-box identification was first implemented using linear structures [2–4], but it seems more powerful for nonlinear structures, including polynomial models [5, 8–10], RBFs [11–13], fuzzy systems [7, 14, 15], multilayer perceptron (MLP) networks [16–19]. Due to being nonlinear in the parameters, procedures for including auxiliary information in MLP networks seems to be less explored, although some methods have been put forward for semi-physical modeling [17, 19], for static information [20] (exact matching) and symmetry [18] for neural models.

Barbosa and colleagues [10] described another way to include auxiliary information, using bi-objective parameter estimation, where one objective is to improve the fitness to empirical data and the other is to improve the fitness to the auxiliary information. This approach was implemented using polynomial models and compared to another techniques (e.g. constrained polynomial models, ANNs). The method was applied also in [21, 22], using MLP networks, and the second objective was the minimization of the *root mean squared error* (RMSE) of the process static curve. The main drawbacks of those methods are the high computational cost, due to the calculation of the model static curve to evaluate the objective function, and the use of evolutionary algorithms, due to the non-convexity of the problem. Models that are linear with respect to the parameters lead to convex problems, that are easier to deal with and the static curve can be sometimes determined analytically [5].

A method for estimating grey-box models was proposed in a previous work [21]. However, the procedure generate a family of solutions that are not trivial to chose just one suitable model between them. The aim of this paper is to discuss the choice of a solution using *auxiliary information* about static information in dynamic models.

This paper is organized as follows. The problem is defined in Section 2 and a background is provided in Section 3. In Section 4 the proposed procedure is presented and the results and discussions are provided in Section 5. Section 6 shows the main conclusions and suggestions for future work.

2 Problem Statement

The problem is to estimate a NARX (Nonlinear AutoRegressive with eXogenous inputs) model given by

$$y(k) = F\left(\boldsymbol{\psi}(k-1),\, \boldsymbol{\theta}\right), \tag{1}$$

where $F(\,.\,)$ is a nonlinear function; $\boldsymbol{\psi}(k-1) = [y(k-1)\cdots y(k-n_y)\ \ u(k-1)\cdots u(k-n_u)]^T$ is the vector of independent variables, n_u and n_y are the maximum lags of the input and output signals, respectively; $\boldsymbol{\theta} \in \mathbb{R}^q$ is the vector of parameters to be estimated from measured data. The model structure, defined by $F(\,.\,)$, is assumed to be known. For a given model structure, $\boldsymbol{\theta}$ is estimated from two sources of information: i) training data $\boldsymbol{Z}_{\mathrm{d}}$; and ii) auxiliary information $\boldsymbol{Z}_{\mathrm{s}}$. Training data $\boldsymbol{Z}_{\mathrm{d}}$ is organized as

$$\boldsymbol{Z}_{\mathrm{d}} = [\boldsymbol{\psi}(k-1)\ \ y(k)],\ k = 1, \ldots, N_{\mathrm{d}} :\ \boldsymbol{Z}_{\mathrm{d}} \in \mathbb{R}^{N_{\mathrm{d}}\times(n_u+n_y)} \tag{2}$$

obtained from dynamical tests or historical data where the system is clearly in transient regime (non stationary) to produce informative data. Ideally, it is obtained from persistently exciting inputs, like PRBS (pseudorandom binary sequence) or noisy input signals, ranging over a wide operating range. However, in many practical applications the available training data covers a narrow operating range, producing (black-box) models that have good performance on that narrow range, but with poor generalization performance.

To improve generalization, an auxiliary source of information about the steady-state behavior of the system $\boldsymbol{Z}_{\mathrm{s}}$, expressed as a set of pairs $(\bar{u},\, \bar{y})$, is used. For a pair given by $(\bar{u}_0,\, \bar{y}_0)$, this means that if the input $u(k) = \bar{u}_0$ is held for a sufficiently long time, then $\lim_{k\to\infty} y(k) = \bar{y}_0$. Hence $\bar{y}_0$ is an asymptotically stable fixed point for a constant input $\bar{u}_0$. Using this static information $(\bar{u}_0,\, \bar{y}_0)$, the vector of independent variables $\bar{\boldsymbol{\psi}}(k-1)$ in steady state is defined from $\boldsymbol{\psi}(k-1)$ as:

$$\begin{aligned}
\boldsymbol{\psi}(k-1) &= \Big[y(k-1) \quad \cdots \quad y(k-n_y) \quad u(k-1) \quad \cdots \quad u(k-n_u)\Big]^T, \\
\bar{\boldsymbol{\psi}}(k-1) &= \Big[\bar{y}(k-1) \quad \cdots \quad \bar{y}(k-n_y) \quad \bar{u}(k-1) \quad \cdots \quad \bar{u}(k-n_u)\Big]^T, \\
\bar{\boldsymbol{\psi}}(k-1) &= \Big[\bar{y}_0 \quad \cdots \quad \bar{y}_0 \quad \bar{u}_0 \quad \cdots \quad \bar{u}_0\Big]^T.
\end{aligned} \tag{3}$$

The dataset with auxiliary information about the steady-state behavior of the system is organized as

$$\boldsymbol{Z}_{\mathrm{s}} = [\bar{\boldsymbol{\psi}}(k-1)\ \ \bar{y}(k)],\ k = 1, \ldots, N_{\mathrm{s}} :\ \boldsymbol{Z}_{\mathrm{s}} \in \mathbb{R}^{N_{\mathrm{s}}\times(n_u+n_y)}, \tag{4}$$

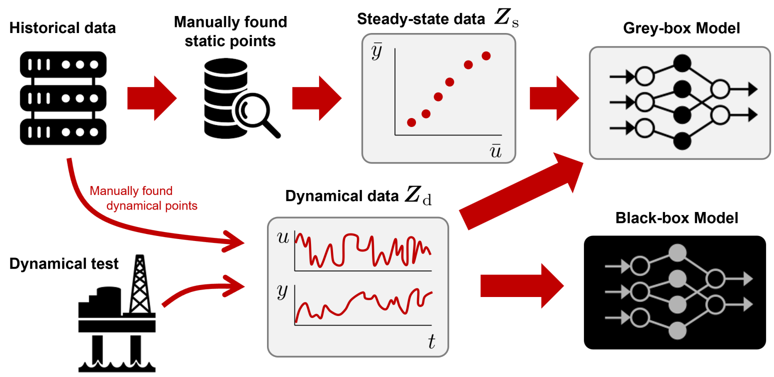

Figure 1. Datasets used to estimate grey- and black-box models. While black-box model uses only Z_{d} to estimate parameters, the grey-box take advantage of two sources of information: Z_{d} *and* Z_{s}.

where N_{s} is the number of static points in $\boldsymbol{Z}_{\mathrm{s}}$.

The estimated model (1) should exhibit "good" *dynamical* and *static* behavior. It is known that a *dynamical* dataset (e.g. $\boldsymbol{Z}_{\mathrm{d}}$) contains some *static* information – the opposite is not true. Nevertheless, it is assumed that $\boldsymbol{Z}_{\mathrm{d}}$ is less informative than $\boldsymbol{Z}_{\mathrm{s}}$ in terms of static behavior. A narrow operating range of dynamical data $\boldsymbol{Z}_{\mathrm{d}}$ is a typical practical example of such a case, hence both datasets are complementary.

Figure 1 illustrates the information used in model estimation (1) using grey- and black-box approaches. The aim is to estimate $\boldsymbol{\theta}$ from $\boldsymbol{Z}_{\mathrm{d}}$ *and* $\boldsymbol{Z}_{\mathrm{s}}$ simultaneously.

3 Background

Neural networks models can achieve good performance in dynamical systems [1, 10, 21]. The NARX dynamical model considered in this work has the following structure:

$$y(k) = \theta_0 + \sum_{i=1}^{n_h} \theta_i \tanh\left(\theta_{i,0} + \sum_{j=1}^{n_y} \theta_{i,j} y(k-j) + \sum_{j=1}^{n_u} \theta_{i,(j+n_y)} u(k-j)\right), \tag{5}$$

where n_h denotes the number of hidden layers, a structural parameter assumed to be known.

Using standard optimization algorithms (e.g., Levenberg–Marquardt) it is possible to fit model (5) to dynamical data $\boldsymbol{Z}_{\mathrm{d}}$, but typically $\boldsymbol{Z}_{\mathrm{s}}$ is not be used during the training. To include steady-state information, a bi-objective approach was proposed in [22], where $\hat{\boldsymbol{\theta}}$ was estimated by minimizing

$$J_1 = \frac{1}{N_{\mathrm{d}}} \sum_{k=1}^{N_{\mathrm{d}}} \sqrt{[y(k) - \hat{y}(k)]^2}, \qquad J_2 = \frac{1}{N_{\mathrm{s}}} \sum_{k=1}^{N_{\mathrm{s}}} \sqrt{[\bar{y}(k) - \hat{\bar{y}}(k)]^2}, \tag{6}$$

where J_1 and J_2 were computed over $\boldsymbol{Z}_{\mathrm{d}}$ and $\boldsymbol{Z}_{\mathrm{s}}$ datasets, respectively and $\hat{\bar{y}}(k)$ is the the k-th fixed point of the model for the input $u(k) = \bar{u}(k),\ \forall k$. An evolutionary algorithm was used for parameters estimation due to the nonconvexity of the optimization problem.

The main drawback of the approach proposed in [22] is the explicit calculation of $\hat{\bar{y}}(k)$. To see this, consider the measured static point $(\bar{u}_0,\, \bar{y}_0)$. In this case, the calculation of the fixed point of the model consists of finding a $\hat{\bar{y}}$ that satisfies the equation $F\left(\bar{u}_0,\, \hat{\bar{y}},\, \hat{\boldsymbol{\theta}}\right) - \hat{\bar{y}} = 0$, in which $F(\,.\,)$ is given by the right hand side of (5) with $u(k-j) = \bar{u}_0,\ \forall j = 1, \ldots, n_u$. Note that the solution of this equation is not trivial and may not be found analytically, due to nonlinearity. Hence, $\hat{\bar{y}}$ was obtained numerically by simulation. The computational cost to evaluate the objective function and solve the recursive optimization problem implicit in $\hat{\bar{y}}$, i.e. $\hat{\bar{y}} = F(F(F(\ldots F(\bar{u}_0,\, \hat{\bar{y}},\, \hat{\boldsymbol{\theta}}))))$ is quite high. The method presented in the next section aims at circumventing this shortcoming.

4 Methodology

Black-box procedures estimate the parameters that minimize the cost function

$$J_{\mathrm{d}} = \|y(k) - F\left(\boldsymbol{\psi}(k-1),\, \boldsymbol{\theta}\right)\|, \tag{7}$$

over the training data set $\boldsymbol{Z}_{\mathrm{d}}$. Grey-box procedures would also use information in the static data $\boldsymbol{Z}_{\mathrm{s}}$. One way of doing this is by means of bi-objective where another cost function, J_2 in (6), is simultaneously considered. However, computing J_2 requires finding the fixed points of the model, which is computationally expensive.

The key feature in the methodology is that the fixed points need *not* be explicitly computed neither analytically nor numerically. Instead, the one-step-ahead error from a desired "static target value" $\bar{y}(k)$ is minimized

$$J_{\mathrm{s}} = \|\bar{y}(k) - F\left(\bar{\boldsymbol{\psi}}(k-1),\, \hat{\boldsymbol{\theta}}\right)\|. \tag{8}$$

In the proposed algorithm, $\bar{y}(k)$ is chosen to be a point on the static curve. In (8) the model fixed point is not computed as for J_2 (see Eq. 6). Fortunately, both J_2 and J_{s} reach minima at a fixed point [21]. While J_2 only uses static data (measured and from the model), J_{s} uses both: the target is static but the model output is used at all time. Hence there is no need to reach steady-state in order to use J_{s}. It is worth noting that J_{s} (8) resembles J_{d} (7) but with the calculation done over $\boldsymbol{Z}_{\mathrm{s}}$, instead of $\boldsymbol{Z}_{\mathrm{d}}$. Despite of the simplicity, its great advantage is to permit the use of ordinary minimization algorithms to also fit the steady state data. To estimate the parameters of the model using both J_{s} and J_{d}, a convex combination of such cost functions is used [10]:

$$\boldsymbol{J_{\mathrm{sd}}} = (1-\lambda)J_{\mathrm{d}} + \lambda J_{\mathrm{s}}, \tag{9}$$

where the $\lambda \in [0,\, 1]$ is the parameter that weights the balance between *static* and *dynamical* information. When $\lambda = 0$ the estimation algorithm only considers the dynamical information (e.g. *black-box* approach) and the addition of auxiliary information about the steady state increases as $\lambda \to 1$. In the extreme case of $\lambda = 1$, the parameter estimation totally disregards the dynamical information. Generally, competitive models can be achieved by a suitable balance of the information in $\boldsymbol{Z}_{\mathrm{s}}$ and $\boldsymbol{Z}_{\mathrm{d}}$.

The problem of finding the right balance for λ is addressed in this paper and it is known as a *decision making* problem. For a given set of M solutions, denoted by $\Theta = \{\theta_i\}, i = 1, 2, \ldots, M$, the problem is to find a θ_i that provides a suitable solution. Considering the plane formed by J_{d} and J_{s}, a simple decision maker can be computed as the minimal distance (L2 norm) to the ideal solution (the origin) as:

$$\theta_a = \{\theta | \sqrt{e_{\mathrm{d}}(\theta_a)^2 + J_{\mathrm{s}}(\theta_a)^2} < \sqrt{e_{\mathrm{d}}(\theta_i)^2 + J_{\mathrm{s}}(\theta_i)^2}, \forall \theta_i \in \Theta\}, \tag{10}$$

where $e_{\mathrm{d}}(\theta_a)$ is the free-run simulation root mean squared error (RMSE) computed over $\boldsymbol{Z}_{\mathbf{d}}$. Note that this decision maker use the auxiliary information in J_{s} to select the solution.

A second decision maker is compared:

$$\theta_b = \{\theta | \sqrt{e_{\mathrm{t}}(\theta_b)^2 + J_{\mathrm{s}}(\theta_b)^2} < \sqrt{e_{\mathrm{t}}(\theta_i)^2 + J_{\mathrm{s}}(\theta_i)^2}, \forall \theta_i \in \Theta\}, \tag{11}$$

where $e_{\mathrm{t}}(\theta_b)$ is the free-run RMSE computed over $\boldsymbol{Z}_{\mathbf{t}}$.

Other decision maker can be computed as the parameters in Θ that provide the minimum RMSE of the free-run simulation over $\boldsymbol{Z}_{\mathbf{t}}$. A decision maker, proposed by Barroso et al. [23], considers the free-run simulation error over $\boldsymbol{Z}_{\mathbf{t}}$, in which the minimally correlated with the model output is chosen.

5 Results and Discussion

Grey-box modeling is used to estimate the downhole pressure in deep water oil production plants, which is an important measure for well productivity and operational management [21, 24–26]. These models should represent well the dynamic and static behavior of the process in order to provide valuable information about long service life of the oil well and to avoid harmful events like severe slugging [21]. The methodology presented in Section 4 is used to obtain models with good balance between dynamic and static performance, by including the auxiliary information in a bi-objective fashion. The auxiliary information (static points) was estimated manually by inspection of the historical data. Figure 2 compares the stationary points (Z_{s}) with the training (Z_{d}) and test (Z_{t}) data sets.

Figure 3 shows instantaneous gas-lift flow rate ($\mathrm{FT4}, u_1$) and downhole pressure ($\mathrm{PT1}, y$) over the *training* and *validation* data. Like the previous numerical examples, the training and test data have information over a limited operating range, while the *validation* data has operating ranges not present in the training data.

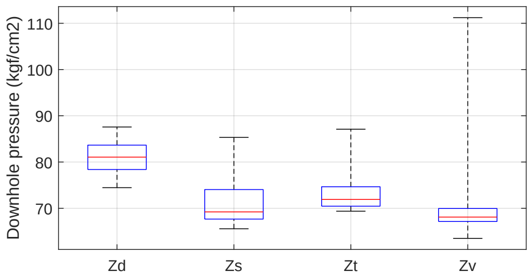

Figure 2. Comparison between the output range, the downhole pressure, in datasets Z_d, Z_s, Z_t and Z_v. Information below 70 kgf/cm^2 is present only in auxiliary information Z_s and in the validation dataset Z_v.

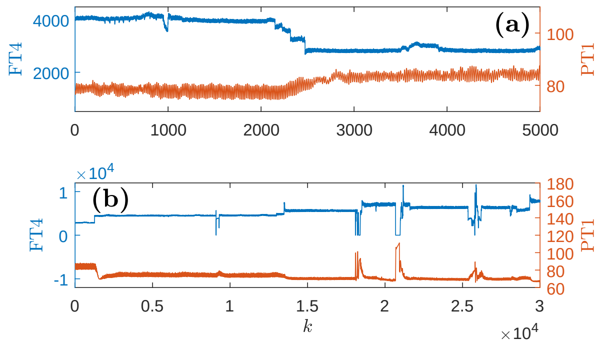

Figure 3. Instantaneous gas-lift flow rate FT4 (u_1) and the downhole pressure PT1 (y) from (a) training $\boldsymbol{Z}_\mathrm{d}$; and (b) validation $\boldsymbol{Z}_\mathrm{v}$ datasets. The fast oscillations are due to severe slugging. From [21].

The identified dynamic models use only platform variables [26, 27] with fixed MLP structure [21]

$$\begin{aligned} y(k) = \theta_0 + \sum_{i=1}^{10} \theta_i \tanh \Big(& \theta_{i,0} + \theta_{i,1} y(k-1) + \theta_{i,2} y(k-2) + \theta_{i,3} y(k-3) \\ & + \theta_{i,4} u_1(k-1) + \theta_{i,5} u_1(k-42) + \theta_{i,6} u_1(k-136) \\ & + \theta_{i,7} u_2(k-1) + \theta_{i,8} u_2(k-42) + \theta_{i,9} u_2(k-136) \\ & + \theta_{i,10} u_3(k-1) + \theta_{i,11} u_3(k-5) + \theta_{i,12} u_3(k-22) \\ & + \theta_{i,13} u_4(k-1) + \theta_{i,14} u_4(k-5) + \theta_{i,15} u_4(k-22) \\ & + \theta_{i,16} u_5(k-1) + \theta_{i,17} u_5(k-5) + \theta_{i,18} u_5(k-22) \Big), \end{aligned} \tag{12}$$

that has 1 hidden layer with 10 nodes with activation function $\tanh(\cdot)$, and linear function in the output node. In (12), the signals $u_i(k)$ are variables available at the platform, and $y(k)$ is the downhole pressure. For comparison, a model $\mathcal{M}_0$ was estimated using only the dynamic training data ($\boldsymbol{Z}_\mathrm{d}$), without auxiliary information, in a black-box approach with the Levenberg-Marquardt algorithm. The proposed procedure is applied in training model family $\mathcal{M}_3$, with the Levenberg-Marquardt algorithm and different λ values (9), $\lambda \in [0.02, 0.98]$. The whole training of the 201 parameters of models in $\mathcal{M}_3$ spend about 105 seconds.

Table 1 shows the RMSE over the validation dataset. The best model obtained by the proposed procedure $\mathcal{M}_{3c}$ achieved improved results (Figure 4). $\mathcal{M}_{3c}$ reached better performance than the black-box model $\mathcal{M}_0$, especially at operating points for which the only source of information was the *auxiliary* data (e.g. $y \approx 70$). This is very relevant for many practical situations where the available dynamical data does not cover all operating regimes of the system. Obtaining static data from historical records is normally a straightforward task that may help to find more representative models as shown in this real-world example.

Table 1 shows the RMSE over the validation dataset. t is relevant to point out that for none of the examples, the best models were obtained for $\lambda = 0$ (which would mean to say that there was no gain in using static data). In particular, for the downhole soft-sensor, the relative importance of dynamical and static data is very well balanced. Hence it seems fair to conclude that the use of Proposition 1 makes good use of steady-state information while keeping computational costs quite moderate.

The main advantage of the method lies in its simplicity of application, which allows the insertion of an *auxil-*

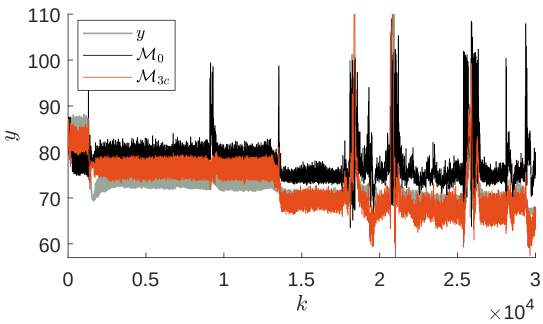

Figure 4. Free-run simulation over validation dataset $\boldsymbol{Z}_{\mathrm{v}}$. From [21].

Table 1. Root mean squared error (RMSE) of each model evaluated over validation dataset $\boldsymbol{Z}_{\mathrm{v}}$ in a free-run simulation. Adapted from [21].

Model	**RMSE over $\boldsymbol{Z}_{\mathrm{v}}$ [26]**	**Decision Maker**
$\mathcal{M}_0$	6.7420	–
$\mathcal{M}_{3a}$	16.7807 ($\lambda = 0.08$)	θ_a using $\boldsymbol{Z}_{\mathrm{d}}$ and $\boldsymbol{Z}_{\mathrm{s}}$ (10)
$\mathcal{M}_{3b}$	10.9986 ($\lambda = 0.46$)	θ_b using $\boldsymbol{Z}_{\mathrm{t}}$ and $\boldsymbol{Z}_{\mathrm{s}}$ (11)
$\mathcal{M}_{3c}$	23.9496 ($\lambda = 0.68$)	min corr. [23]
$\mathcal{M}_{3d}$	10.9986 ($\lambda = 0.46$)	min RMSE over $\boldsymbol{Z}_{\mathrm{t}}$
$\mathcal{M}_{3e}$	3.7285 ($\lambda = 0.54$)	min RMSE over $\boldsymbol{Z}_{\mathrm{v}}$
$\mathcal{M}_{3f}$	173.949 ($\lambda = 0.98$)	max RMSE over $\boldsymbol{Z}_{\mathrm{v}}$

iary information about steady state conditions in models with rather complex structures, such as neural networks. The estimation was done with ordinary algorithms, as weighted least-squares (polynomial models) and backpropagation (MLPs models). We argue that the procedure can be applied to other nonlinear model classes as well.

6 Conclusion

This work presented a method for including *auxiliary information* about steady state behavior of the system in parameter estimation stage, by adding a new objective function (weighted problem). The method allows to choose the right balance between static and dynamic information.

The main difference between the proposed procedure and previous grey-box procedures is that here it is not necessary to calculate or even estimate the fixed-points of the model. So, it was possible to add auxiliary information in rather complex structures as dynamical neural networks – which can be very difficult to calculate its fixed-points explicitly, leading to a lower computational cost of the procedure. This work opens the possibility of applying the grey-box procedure to models of other classes, like *auto regressive integrated moving average* (ARIMA), wavenets, *radial basis functions* (RBF), *smoothing spline models* (SSM), *fuzzy models*.

7 ACKNOWLEDGMENTS

This work has been supported by the Brazilian agency CAPES. The authors gratefully acknowledge financial support by CNPq: 303412/2019-4 (LAA); CNPq: 304201/2018-9 (BHGB).

References

[1] S. Chen, S. A. Billings, and P. M. Grant. Non-linear system identification using neural networks. *International Journal of Control*, vol. 51, n. 6, pp. 1191–1214, 1990.

[2] T. A. Johansen. Identification of non-linear systems using empirical data and prior knowledge - an optimzation approach. *Automatica*, vol. 32, pp. 337, 1996.

[3] H. J. A. F. Tulleken. Grey-box modeling and identification using physical knowledge and bayesian techniques. *Automatica*, vol. 29, n. 2, pp. 285–308, 1993.
[4] E. Eskinat, S. H. Johnson, and W. L. Luyben. Use of auxiliary information in system identification. *Ind. Eng. Chem. Res.*, vol. 32, pp. 1981–1992, 1993.
[5] L. A. Aguirre, M. F. S. Barroso, R. R. Saldanha, and E. M. A. M. Mendes. Imposing steady-state performance on identified nonlinear polynomial models by means of constrained parameter estimation. *Control Theory and Applications, IEE Proceedings* -, vol. 151, n. 2, pp. 174 – 179, 2004a.
[6] E. G. Nepomuceno, R. H. C. Takahashi, and L. A. Aguirre. Multiobjective parameter estimation: Affine information and least-squares formulation. *Int. J. Control*, vol. 80, n. 6, pp. 863–871, 2007.
[7] L. Sánchez, I. Couso, and M. González. A design methodology for semi-physical fuzzy models applied to the dynamic characterization of LiFePO4 batteries. *Applied Soft Computing*, vol. 14, pp. 269–288, 2014.
[8] M. V. Corrêa, L. A. Aguirre, and R. R. Saldanha. Using prior knowledge to constrain parameter estimates innonlinear system identification. *IEEE Trans. Circuits Syst. I*, vol. 49, n. 9, pp. 1376–1381, 2002.
[9] E. G. Nepomuceno, R. H. C. Takahashi, G. F. V. Amaral, and L. A. Aguirre. Nonlinear identification using prior knowledge of fixed points: a multiobjective approach. *International Journal of Bifurcation and Chaos*, vol. 13, n. 5, pp. 1229–1246, 2003.
[10] B. H. G. Barbosa, L. A. Aguirre, C. B. Martinez, and A. P. Braga. Black and gray-box identification of a hydraulic pumping system. *Control Systems Technology, IEEE Transactions on*, vol. 19, n. 2, pp. 398 –406, 2011.
[11] L. A. Aguirre, G. B. Alves, and M. V. Corrêa. Steady-state performance constraints for dynamical models based on rbf networks. *Engineering Applications of Artificial Intelligence*, vol. 20, n. 7, pp. 924 – 935, 2007.
[12] S. Chen, C. J. Harris, and X. Hong. Grey-box radial basis function modelling: The art of incorporating prior knowledge. In *2009 IEEE/SP 15th Workshop on Statistical Signal Processing*, number 2, pp. 377–380, 2009.
[13] J. B. de Almeida Rego, A. de Medeiros Martins, and E. B. Costa. Deterministic System Identification Using RBF Networks. *Mathematical Problems in Engineering*, vol. 2014, pp. 1–10, 2014.
[14] J. Abonyi, R. Babuska, and F. Szeifert. Fuzzy modeling with multivariate membership functions: gray-box identification and control design. *IEEE transactions on systems, man, and cybernetics. Part B, Cybernetics : a publication of the IEEE Systems, Man, and Cybernetics Society*, vol. 31, n. 5, pp. 755–67, 2001.
[15] T. Abdelazim and O. Malik. Identification of nonlinear systems by Takagi-Sugeno fuzzy logic grey box modeling for real-time control. *Control Engineering Practice*, vol. 13, n. 12, pp. 1489–1498, 2005.
[16] D. C. Psichogios and L. H. Ungar. A hybrid neural network-first principles approach to process modeling. *AIChE Journal*, vol. 38, n. 10, pp. 1499–1511, 1992.
[17] M. L. Thompson and M. A. Kramer. Modeling chemical processes using prior knowledge and neural networks. *AIChE Journal*, vol. 40, n. 8, pp. 1328–1340, 1994.
[18] L. A. Aguirre, R. A. M. Lopes, G. F. V. Amaral, and C. Letellier. Constraining the topology of neural networks to ensure dynamics with symmetry properties. *Phys. Rev. E*, vol. 69, n. 026701, 2004b.
[19] Y. Oussar and G. Dreyfus. How to be a gray box: dynamic semi-physical modeling. *Neural networks : the official journal of the International Neural Network Society*, vol. 14, n. 9, pp. 1161–72, 2001.
[20] G. F. V. Amaral. Use of neural networks and a priori knowledge in the identification of nonlinear systems. Master's thesis, UFMG. Document in Portuguese, 2001.
[21] L. Freitas, B. H. Barbosa, and L. A. Aguirre. Including steady-state information in nonlinear models: An application to the development of soft-sensors. *Engineering Applications of Artificial Intelligence*, vol. 102, n. March, pp. 104253, 2021.
[22] L. F. Abreu, B. H. G. Barbosa, A. F. Teixeira, and L. A. Aguirre. Use of auxiliary information in neural network training: a bi-objective procedure. In *CBA 2012*, pp. 1719–1726, Campina Grande, PB. Document in Portuguese, 2012.
[23] M. S. F. Barroso, R. H. C. Takahashi, and L. A. Aguirre. Multi-objective parameter estimation via minimal correlation criterion. *J. Process Control*, vol. 17, n. 4, pp. 321–332, 2007.
[24] B. O. Teixeira, B. H. Barbosa, L. P. Gomes, A. F. Teixeira, and L. A. Aguirre. Ukf-based data-driven soft sensing: A case study of a gas-lifted oil well*. *IFAC Proceedings Volumes*, vol. 45, n. 16, pp. 918–923. 16th IFAC Symposium on System Identification, 2012.
[25] B. H. G. Barbosa, L. P. Gomes, A. F. Teixeira, and L. A. Aguirre. Downhole pressure estimation using committee machines and neural networks. *IFAC-PapersOnLine*, vol. 48, n. 6, pp. 286 – 291. 2nd IFAC Workshop on Automatic Control in Offshore Oil and Gas Production {OOGP}, 27-29 May 2015, 2015.
[26] L. A. Aguirre, B. O. S. Teixeira, B. H. G. Barbosa, A. F. Teixeira, M. C. M. M. Campos, and E. M. A. M. Mendes. Development of soft sensors for permanent downhole gauges in deepwater oil wells. *Control Engineering Practice*, vol. 65, n. Supplement C, pp. 83 – 99, 2017.
[27] de G. A. P. Morais, B. H. G. Barbosa, D. D. Ferreira, and L. S. Paiva. Soft sensors design in a petrochemical process using an evolutionary algorithm. *Measurement*, vol. 148, pp. 106920, 2019.