This is a repository copy of *On sparse nonlinear system identification using orthogonal matching pursuit, orthogonal least squares and LASSO*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/214149/

Version: Accepted Version

Proceedings Paper:

Wei, H.-L. orcid.org/0000-0002-4704-7346 (2024) On sparse nonlinear system identification using orthogonal matching pursuit, orthogonal least squares and LASSO. In: 2024 32nd Mediterranean Conference on Control and Automation (MED). 2024 32nd Mediterranean Conference on Control and Automation (MED 2024) June 11-14, 2024 | Chania, Crete, Greece., 11-14 Jun 2024, Chania, Crete, Greece. Institute of Electrical and Electronics Engineers (IEEE) , pp. 935-940. ISBN 9798350395457

https://doi.org/10.1109/MED61351.2024.10566162

© 2024 The Authors. Except as otherwise noted, this author-accepted version of a paper published in 2024 32nd Mediterranean Conference on Control and Automation (MED) is made available via the University of Sheffield Research Publications and Copyright Policy under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/

Reuse
This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here: https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

2024 32nd Mediterranean Conference on Control and Automation (MED)
June 11-14, 2024 | Chania, Crete, Greece. pp. 935-940.

Final accepted manuscript

On Sparse Nonlinear System Identification Using Orthogonal Matching Pursuit, Orthogonal Least Squares and LASSO*

Hua-Liang Wei

***Abstract*— System identification, as a rich and vital discipline, provides a practical and general methodology and tool for quantitatively modelling the input-output relationships of dynamical systems. Sparse nonlinear system identification (SNSI), especially parametric sparse nonlinear system identification (PSNSI), is an important and vital field of research with a wide range of applications. This work is concerned with PSNSI and particular attention is paid to the assessment of three well-known mainstream sparse learning methods, namely, orthogonal least squares (OLS), orthogonal matching pursuit (OMP) and least absolute shrinkage and selection operator (LASSO). The performances of these methods are tested and evaluated through three case studies relating to PSNSI problems. The research results and findings of this work provide practical useful information and guidance for researchers to better choose or adapt methods when solving PSNSI problems.**

I. Introduction

System identification (SysID) is a general scientific technique that is widely used for determining mathematical models of dynamical systems or processes based on input and output measurements. There are many cases where the inputs and outputs of a system of interest can be observed or measured, but nothing or just little of the internal structure and inherent dynamics of the system is known; such a system or process is often referred to as a black box or a gray box. System identification provides a vitally important and powerful tool to find a model or a set of models that can well represent the system input and output relationship embedded in the data available for solving the associated modelling tasks. While the internal structure of the system to be identified is usually unknown, the general structures of the models to be used to represent the system can be prespecified. Commonly used system identification models include ARX (AutoRegressive with eXogenous inputs) and ARMAX (AutoRegressive Moving Average with eXogenous inputs) [1], NARX (Nonlinear AutoRegressive with eXogenous inputs) and NARMAX (Nonlinear AutoRegressive Moving Average with eXogenous inputs) [2], neural networks [3], among other models and approaches [4]-[7].

System identification models can be linear or nonlinear, parametric or nonparametric, time-invariant or time-varying, transparent or opaque (black-box), and so on. For many system identification tasks, especially those tasks relating to complex nonlinear systems, the initial models (if not properly addressed and refined), can potentially be very complex and redundant due to the existence of many irrelevant model elements. This is more than true for models defined in high dimensional space (e.g., involving a great number of inputs and outputs), requiring a huge number of unknown parameters. For such cases, model refinement, a process of determining best or sub-best model subsets, becomes significantly important for obtaining meaningful parsimonious or sparse models that best represent the system input and output relationship [8],[9]. Parsimonious sparse models have several attractive advantages, e.g., they can overcome the overfitting problem, have better generalization ability, and save the cost measuring the irrelevant model input variables when carrying out similar modelling tasks in future. For these reasons, sparse system identification has attracted considerable attention and many sparse methods have been developed over the past years including the three well-known methods: orthogonal least squares (OLS) [2],[10],[11], orthogonal matching pursuit (OMP) [12], [13], and least absolute shrinkage and selection operator (LASSO) [14], [15]. These methods and their variations have found extensive applications in a wide range of fields, including nonlinear dynamical system identification and modelling (see, e.g., [2], [16]-[21]).

As mentioned earlier, the internal structure of a system to be identified may be completely unknown (a black box). However, the overall structure of the models used to represent or approximate the system can be prespecified. For illustration purposes, take a simple linear system case as example here. For a set of data of a dynamical system of interest, we can use an ARX model to represent the system as follows: $y(k) = a_1y(k-1) + \ldots + a_py(k-p) + b_1u(k-1) + \ldots + b_qu(k-q)$, where $u(k)$ and $y(k)$ are measured sequences of the system input and output variables, respectively; $\{a_1, \ldots, a_p\}$ and $\{b_1, \ldots, b_q\}$ are model parameters; p and q are the ARX model order. To sufficiently represent the inherent dynamics and the input-output relationship of the system, the initially prespecified models, or the elements to be used to build a model, are usually overcomplete. For example, the true model order of the ARX system may be p=3 and q=2. However, without any a priori knowledge about the system (this is the case for many applications), the model order p and q may be chosen to be much larger than 3 and 2, respectively, in the initial ARX model. Such models are usually not applicable for future use due to their lack of generalization ability caused by overfitting (this is especially true for complex dynamical nonlinear system modelling tasks involving many inputs). That is why model refinement including model subset selection is a crucial process for any system identification task.

This work is concerned with sparse nonlinear system identification problem. While there are a huge number of publications on sparse system identification using the three mainstream methods, OLS, OMP and LASSO, in the literature,

*Resrach supported by STFC (Ref. ST-Y001524-1), NERC (Ref. NE/V002511/1), EPSRC (Ref. EP/H00453X/1) and EPSRC (Ref. EP/I011056/1).

H. L. Wei is with the Department of Automatic Control and Systems Engineering, The University of Sheffield, Sheffield, S1 3JD (Corresponding author phone: 0044 1142225198; e-mail: w.hualiang@sheffield.ac.uk).

2024 32nd Mediterranean Conference on Control and Automation (MED)
June 11-14, 2024 | Chania, Crete, Greece. pp. 935-940.

very little work has been done to comprehensively evaluate and compare their overall performance and efficiency when applied to solve sparse nonlinear system identification tasks. This motivates us to conduct investigations in this area, aiming to bridge the existing research gap. In this work, the NARAMX method is used to approximate and represent the underlying nonlinear dynamical systems, this is because the method can usually lead to transparent and parsimonious models, which us enable to reveal and understand black-box systems through transparent models.

The main contributions of the work include:

- Not much work has been done to comprehensively evaluate the performances of OMP and LASSO for parametric sparse nonlinear system identification (PSNSI) problems. This work fills the gap.
- The overall performances of LASSO, OMP and OLS are evaluated through three case studies, two on simulation data and one on real data of a continuous-time two-tank system.
- The results and findings are useful and provide practical guidance for better choosing or adapting methods when solving PSNSI problems.

II. Nonlinear System Identification Using Parametric Models

Parametric models, due to their good properties (e.g., transparent, easy-to-interpret, and easy-to-communicate), are preferred for many system identification applications in particular when the primary modelling task is to investigate and establish a quantitative relationship between the system inputs and outputs. For example, more than often, the major study interest of system identification is to understand how a system output is explicitly related to the system inputs, and how the involved model input variables interact with each other. For such application scenario, parametric models, especially parametric sparse models may be particularly preferred to nonparametric models.

Previous experiences show that a large class of nonlinear systems can be represented using NARMAX models [2]. Taking the case of single-input, single-output (SISO) systems as an example, the general form of NARMAX models is as follows:

$$\begin{aligned} y(k) = f\,(& y(k-1), y(k-2)..., y(k-n_y), \\ & u(k-\tau), u(k-\tau-1),..., u(k-n_u), \\ & e(k-1), e(k-2),..., e(k-n_e)) + e(k). \end{aligned} \tag{1}$$

where $u(k)$ is an input sequence, $y(k)$ is the system output sequence, and $e(k)$ is noise sequence; n_y, n_u and n_e are the associated maximum time lags; τ is the time delay between the response and the model input variables, and usually $\tau = 0$ or $\tau = 1$; $f(\bullet)$ is some unknown function that needs to be built from available training data. In the literature, a variety of basis functions can be used to appreciate the to approximate the unknown function $f(\bullet)$; in many applications, polynomial models, due to their attractive properties, are commonly employed [2].

The form (1) can be easily extended to multi-input and single-output (MISO) and multi-input and multi-output (MIMO) cases in a straightforward way. For MISO and MIMO cases, interested readers are referred to [2],[22],[23].

The nonlinear degree (i.e., the degree of nonlinearity) of a NARMAX model is defined as the highest order of all model terms. For example, the nonlinear degree of the model $y(k) = a_1 y(k-1) + a_2 u(k-2)$ is 1, whereas the nonlinear degree of the model $y(k) = b_1 u(k-1) + b_2 u^3(k-2)$ is 4.

In most real applications, the actual model structure of the system of interest is unknown (and may never be known). The main task of system identification is to find a model that can best represent the input-output relationship of the system through learning from data. In doing so, an appropriate number of experimental settings need to be specified prior to model identification. For NARMAX models, the settings include the nonlinear degree, maximum lags for inputs and outputs, model size (the total number of model terms), and so on. For convenience of description, take a simple scenario as an example: Assume that the actual model of a SISO system is $y(k) = 0.5y(k\text{-}1) + 0.8u(k\text{-}1) + 0.2u(k\text{-}1)u(k\text{-}2)$, which is assumed to be unavailable from first principle. If we set the maximum lags as n_y=2, n_u=2, n_e=0, the time delay τ =1, and the nonlinear degree as $\ell = 2$, then a collection of candidate model terms to be used to represent the system may be defined as follows:

$$D = \left\{ \begin{array}{l} y(k-1),\;\; y(k-2),\; u(k-1),\; u(k-2), \\ y^2(k-1),\;\; y(k-1)y(k-2), \\ \qquad\qquad y(k-1)u(k-1),\; y(k-1)u(k-2), \\ y^2(k-2),\; y(k-2)u(k-1),\; y(k-2)u(k-2), \\ u^2(k-1),\;\; u(k-1)u(k-2),\; u^2(k-2) \end{array} \right\} \tag{2}$$

It is expected that a good system identification algorithm should be able to correctly identify the three actual model terms, $y(k\text{-}1)$, $u(k\text{-}1)$ and $u(k\text{-}1)u(k\text{-}2)$, from the above dictionary consisting of 14 candidate terms.

For most applications, only a relatively small number of important model terms are needed in the final models. Methods that can efficiently select the most significant model terms and result in good solutions for system identification and parametric modelling problems are always highly needed.

III. A Brief Introduction of Orthogonal Least Squares, Orthogonal Matching Pursuit and LASSO

This section does not intend to provide detailed descriptions of the three methods: orthogonal least squares (OLS), orthogonal matching pursuit (OMP) and LASSO. Instead, it focuses on a quick introduction to the basic mechanism behind these methods.

These methods are used to find sparse solutions to the following generalized linear regression problem of the form:

$$y(k) = a_0 + a_1 x_1(k) + a_2 x_2(k) + ... + a_p x_p(k) + e(k)\,. \tag{3}$$

where y is a target signal, x_1, x_2, …, x_p are regressors, and e is noise. Note that a large class of discrete-time nonlinear

2024 32nd Mediterranean Conference on Control and Automation (MED)
June 11-14, 2024 | Chania, Crete, Greece. pp. 935-940.

dynamic models can be formulated using Eq. (1). For example, the nonlinear difference equation $y(k) = ay(k\text{-}1) + bu(k\text{-}1) + cy(k\text{-}1)u(k\text{-}2)$ can be written as $y(k) = a_1x_1(k) + a_2x_2(k) + a_3x_3(k)$, with $x_1(k) = y(k\text{-}1)$, $x_2(k) = u(k\text{-}1)$, and $x_3(k) = y(k\text{-}1)u(k\text{-}2)$. Also note that often not all the regressors are equally important for representing the response; some regressors may play little or no role in explaining the change of the response. The main objective of finding a sparse solution amounts to finding a best subset consisting of m variables from the p candidate variables (in many applications $m << p$), such that the m variables can well present the target signal y.

A. Orthogonal Least Squares

The commonly used orthogonal least squares (also known as forward regression with orthogonal least squares, shortly, FROLS [2]) method was originally developed for dynamical model construction based on measured input-output data, by selecting the most important model terms from a specified dictionary consisting of a sufficiently large number candidate linear and nonlinear model terms. The selection process is implemented through an iterative forward orthogonalization procedure: the most important model terms are selected step by step; in each iteration, one important term is selected and is added to a subset comprising those terms selected in all the previous iterations.

In each iteration, the importance of the newly selected model term is measured using a simple but efficient and interpretable index, called error-reduction-ratio (ERR) [2]. ERR is calculated based on the overall model residual, by projecting the original target signal onto a set of bases (orthogonal vectors) generated from the model subset through an orthogonalization procedure. Readers are directed to [2], [8] and [11] for detailed descriptions of FROLS.

B. Orthogonal Matching Pursuit

Orthogonal matching pursuit (OMP) is an improved version of the well-known matching pursuit method [25], which was originally designed to solve the following sparse representation problem: There is a target signal y, together with a finite number of candidates basis signals, $x_1, x_2, \ldots, x_p$; find a best subset consisting of m basis signals from the p candidates (in many applications $m << p$), such that the m variables can well present the target signal y.

The overall implementation procedure of OMP is similar to that of OLS. The main difference between OMP and OLS is that, in each iteration, the subset selected by the former best represents the current model residual, while the subset selected by the latter ensures to minimize the overall model residual, when the target signal y is approximated by a subset consisting of m model terms selected so far. Interested readers are referred to [12] and [13] for details of OMP.

C. Least Absolute Shrinkage and Selection Operator

Note that finding the exact solution to the sparse modelling problem mentioned in Section B can be formulated as the following L_0-norm regularized optimization problem:

$$\hat{\beta} = \arg\min_{\beta} \| \beta \|_0 \quad \text{subject to } X\beta = \mathbf{y}, \tag{4}$$

or

$$\min_{\beta} \left\{ \frac{1}{2} \| \mathbf{y} - X\beta \|^2 + \lambda \| \beta \|_0 \right\}. \tag{5}$$

where $\mathbf{y}$ is a target signal vector, X is the regression matrix with proper dimension (e.g., $n \times p$), β is the regression model parameter vector, λ is a penalty coefficient, and the symbol $\| \cdot \|_0$ represents the L_0-norm of a vector defined as the total number of nonzero entries in the vector.

Solving the optimization problem defined in (5) is NP-hard [26], which is very difficult to solve. To overcome the difficulty encountered by the NP-hard problem, the L_0-norm in (2) is relaxed to L_1-norm in the following manner [14],[15]:

$$\min_{\beta} \| \mathbf{y} - X\beta \|_2^2 \quad \text{subject to } \|\beta\|_1 \le c, \tag{6}$$

or

$$\arg\min_{\beta} \left\{ \frac{1}{2} \| \mathbf{y} - X\beta \|_2^2 + \lambda \| \beta \|_1 \right\}. \tag{7}$$

where c is a parameter that controls the range of the constraint set, and the symbol $\| \cdot \|_1$ represents the L_1-norm of a vector defined as the sum of absolute values of all entries of the vector. The problem defined by (7) is commonly referred to as LASSO regression [14].

IV. Case Studies

This section provides three case studies, two simulation cases and one real data modelling case, to evaluate the performances of the three methods, OLS, OMP and LASSO, from different perspectives.

A. Case 1: Model Identification with an Overcomplete Dictionary of Candidate Model Terms

This is a case where the prespecified dictionary contains all the true model terms of the system to be identified. As a simple example, the dictionary, D, defined by (7) is an overcomplete dictionary.

Now, consider a system with two inputs and one output described by the following model:

$$\begin{cases} x(k) = 0.5x(k-1) - 0.8u_1(k-2) + u_2^2(k-1) \\ \qquad -0.25x(k-2)u_1(k-2) + \xi(k) \\ y(k) = x(k) + \eta(k) \end{cases} \tag{8}$$

where $x(k)$ is the system state signal, $y(k)$ is the system output signal, $u_1(k)$ and $u_2(k)$ are two input signals, $\xi(k)$ and $\eta(k)$ are noise signals.

Numerical experiments were conducted to test the performance of OLS, OMP and LASSO for recovering the system model given by (8) from simulation data. Experimental settings are as follows.

1) Data acquisition

Data were collected through simulations. The system (8) was simulated 100 times independently; in each simulation, the two inputs u_1 and u_2 were set to be Gaussian signal $N(0, 1)$, and the two noise signals were set to be $\xi \sim N(0, 0.1^2)$ and $\eta \sim N(0, 1)$, respectively. A total of 100 datasets were collected;

2024 32nd Mediterranean Conference on Control and Automation (MED)
June 11-14, 2024 | Chania, Crete, Greece. pp. 935-940.

each dataset contains a total of 200 samples (input and output data points).

2) Candidate model settings

The model settings are as follows: $n_y = 5$, $n_u = 5$, $n_e = 0$, $\tau = 1$, and $\ell = 3$. This resulted in a dictionary, D, consisting of 816 candidate model terms in total, including the three actual model terms (for model (8)). The three methods, LASSO, OMP and OLS are applied to the 100 datasets, based on which to select their own best models independently.

3) The determination of model size

The determination of model size (i.e., the number of model terms) is important to control the model complexity and avoid overfitting.

For LASSO, the penalty coefficient λ plays a crucial for determining the model sparsity (size). Following the recommendation from [14], in this study the value of λ is chosen to be such that is within one standard error of the minimum, as such a value can often lead to a more parsimonious model with a minimal loss in mean square error (MSE).

For OMP, we use LASSO as a baseline method. The model size (the number of model terms) selected by LASSO for each individual datasets is used as a reference, to see whether OMP can show comparable performance if it is allowed to select exactly the same numbers of model terms as that of LASSO.

For OLS, the well-known Bayesian Information Criterion (BIC) [27] is used to determine model size. The identification results on the 100 datasets for LASSO, OMP and OLS are summarised in Table I.

TABLE I. PERFORMANCE COMPARISON OF LASSO, OMP AND OLS FOR RECOVERING THE SYSTEM MODEL GIVEN BY (8)

Information of the Identified Models Based on the 100 Datasets					
LASSO		OMP		OLS	
T[a]	Model size[b]	T[a]	Model size[b]	T[a]	Model size[b]
89	4, 49, 14.32	0	4, 49, 14.32	87	4, 8, 6.33

a. Times that all the five actual model terms are included in the selected models by the method out of 100 runs.
b. The minimum, maximum and average numbers of model terms in the 100 identified models.

From Table I, the following can be observed:

- Both LASSO and OLS work well for recovering the system model. While the two methods apparently show comparable performance, OLS can produce much compact models than LASSO for the problem.
- OMP shows no skills for solving the problem, in that the time that all the five actual model terms are included in the selected models by the method out of 100 runs is zero. This may be explained that many candidate model terms in the dictionary are very highly correlated. Perhaps the inherent mechanism of the method makes it struggling in dealing with these highly correlated regressors.

B. *Case 2: Model Identification with an Undercomplete Dictionary of Candidate Model Terms*

This is a case where some or none of the actual system model terms are not included in the pre-specified dictionary. This is a typical scenario in many real applications.

Consider a SISO system designed in [26] and described by the following model:

$$y(k) = -u(k-1)\sqrt{|y(k-1)|} + 0.5u^3(k-1) + u(k-2) + \xi(k) \quad (9)$$

where $u(k)$ is the system input and $\xi(k)$ is noise. The model was simulated by making the input u be formally distributed on [-1, 1], and the noise ξ be following Gaussian distribution, $\xi \sim N(0, 0.1^2)$. A total of 1000 data points were recorded: the first 500 samples were used for model identification and the second 500 samples were used to test the model performance.

The model settings are as follows: $n_y = 2$, $n_u = 2$, $n_e = 0$, $\tau = 1$, and $\ell = 3$. This resulted in a dictionary, D, consisting of 35 candidate model terms. The three methods (LASSO, OLS and OMP) were respectively applied to the 500 training data points, and three models were obtained accordingly. A brief summary of the three models are reported in Table II. Note that the MSE values listed in Table II were calculated based on the model simulation prediction (i.e., model simulation output), which is different from the conventionally used one-step-ahead prediction.

TABLE II. MODEL PERFORMANCE COMPARISON FOR THE SYSTEM DESCRIBED BY MODEL (9)

Performance of the Models Over The Training and Test Dataset								
LASSO			OMP			OLS		
S[a]	Tr[b]	Te[c]	S[a]	Tr[b]	Te[c]	S[a]	Tr[b]	Te[c]
14	0.0384	0.0196	24	0.0211	0.0202	4	0.0362	0.0170

a. S – model size (the number of model terms).
b. Tr – mse (mean square error) on the training data.
c. Te – mse (mean square error) on the test data.

More details of the two models identified by OLS and LASSO are respectively presented below:

$$\begin{aligned} y(k) = &-0.4805u(k-1) + 1.009u(k-2) \\ &-0.4866u(k-1)y^2(k-1) + 0.5042u^3(k-1) \end{aligned} \quad (10)$$

$$\begin{aligned} y(k) = &-0.3431u(k-1) + 0.9167u(k-2) \\ &+0.0208u(k-1)u(k-2) + 0.0091u(k-2)y(k-2) \\ &+0.0117y^2(k-1) + 0.0165y^2(k-2) \\ &+0.2944u^3(k-1) + 0.0960u^2(k-1)u(k-2) \\ &+0.0076u^2(k-1)y(k-1) - 0.0126u^2(k-1)y(k-2) \\ &-0.0011u(k-1)u(k-2)y(k-2) \\ &-0.4809u(k-1)y^2(k-1) + 0.0656u^3(k-2) \\ &+0.0107u(k-2)y^2(k-2) \end{aligned} \quad (11)$$

To save space, the details of the 24-term model identified by OMP is not shown here.

2024 32nd Mediterranean Conference on Control and Automation (MED)
June 11-14, 2024 | Chania, Crete, Greece. pp. 935-940.

To visually demonstrate the performance of the identified models, we take the model (10) as an example to display how well the model represents the original system. Fig. 1 displays the model simulation output (model prediction output) and the corresponding actual output on part of the testing data (samples from 951 to1000). Note that there is no special consideration that only the last 50 data points are displayed here; the main purpose is to give a clear, zoomed-in comparison and visualization between the model simulation output and the actual output.

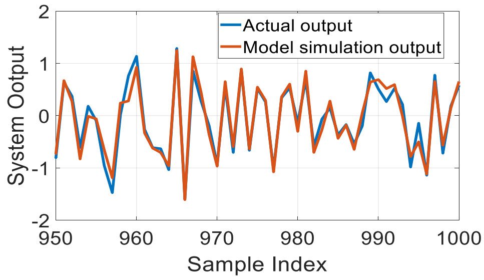

Figure 1. A comprison between the model simulation output from(10) and the corresponding actual output for the system described by (9).

C. Case 3: Identification of a two-tank system

This is a case where data were collected from experiments on a continuous-time two-tank system. The dataset contains a total of 3000 input and output samples recorded with a sampling period of 0.2 second. The input $u(k)$ (units: voltage [V]) is applied to a pump, which produces an inflow to the upper tank. There is a small hole at the bottom of the upper tank, leading to an outflow going into the lower tank. The overall output $y(k)$ of the two-tank system is measured as the liquid level (units: [m]) of the lower tank. The input and output samples are shown in Fig. 2. More details of the system may be found in [28].

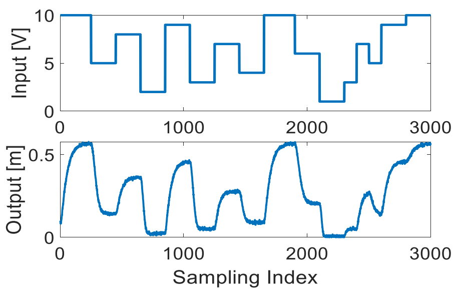

Figure 2. The input and output samples of the two-tank system.

The model settings are as follows: $n_y = 5$, $n_u = 5$, $n_e = 0$, τ =1, and ℓ = 2, that is, the basic variables are chosen to be $y(k\text{-}1)$, …, $y(k\text{-}5)$, $u(k\text{-}1)$, …, $u(k\text{-}5)$. The resulting dictionary, D, consists of 66 candidate model terms in total. The three methods, LASSO, OMP and OLS were applied to the dataset, where the first 2000 samples were used for model identification and the remaining 1000 samples were used to test the model performance. The same model size determination scheme described in Case 1 (Section IV-A) was used to determine the model sizes. The information about the identified models is summarized in Table III.

TABLE III. MODEL PERFORMANCE COMPARISON FOR THE TWO-TANK SYSTEM

Performance of the Models Over the Training and Test Dataset								
LASSO			OMP			OLS		
S[a]	Tr[b]	Te[c]	S[a]	Tr[b]	Te[c]	S[a]	Tr[b]	Te[c]
12	0.0064	0.0095	5	0.0110	0.0046	10	0.00046	0.00029

d. S – model size (the number of model terms)
e. Tr – mse (mean square error) on the training data
f. Te – mse (mean square error) on the test data

The model simulation outputs produced by the models identified by the three methods are shown in Fig. 3.

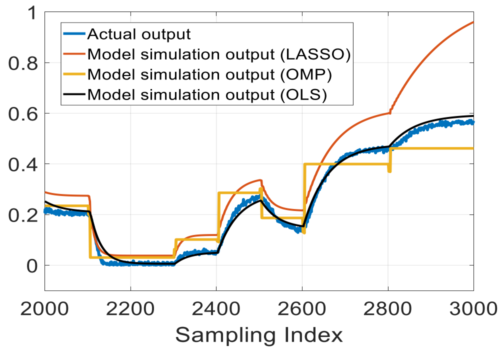

Figure 3. The model simulation outputs generated from the models identified by the three methods for the two-tank system.

D. *A Brief Summary*

From the three case studies, the following observations and findings are noted:

- For Case 1, with an overcomplete dictionary of candidate model terms, both LASSO and OLS correctly identified all the actual system model terms; the models produced by OLS are much more compact than those of LASSO. The OMP method does not show any skill towards correctly identifying all the actual model terms.
- For Case 2, with an undercomplete dictionary of candidate model terms, the three methods produced comparable results; OLS showed the best performance.
- For Case 3, involving a real continuous-time system, OMP did not work at all. Although LASSO showed some good performance, its power seemed not to be fully brought into play.

CONCLUSION

Modeling and identification of nonlinear dynamical systems, especially parametric sparse nonlinear system identification (PSNSI) is a challenging task. Many good algorithms have been developed and are available for dealing with difficult PSNSI tasks, including the three mainstream

2024 32nd Mediterranean Conference on Control and Automation (MED)
June 11-14, 2024 | Chania, Crete, Greece. pp. 935-940.

methods, LASSO, OMP and OLS and their variants. A lot of work has been done on evaluating the overall performances of OMP in the application areas of sparse signal and image reconstruction, and of LASSO in multiple linear regression (and with variable selection) and related application areas, but not much work has been done to comprehensively evaluate the performances of these two types of methods for PSNSI problems where candidate regressors (model terms) are formed by lagged variables of the system inputs and outputs; many of these candidate terms are potentially highly correlated and may be indistinguishable when the sampling period is small [29]. To fill this gap, three case studies have been carried out in this paper.

Hopefully the results and findings presented in this paper will provide practical useful guidance for better choosing or adapting methods when solving PSNSI problems. It is worth mentioning that the work of the paper is limited to a small number of case studies. In future, more studies will be carried out to further explore the power of LASSO and OLS, and where possible to integrate and give play to each other's strengths, so as to offer better solutions for difficult and challenging PSNSI problems.

REFERENCES

[1] L. Ljung, System Identification Toolbox User's Guide. The MathWorks, Inc., 2022.
[2] S. A. Billings, Nonlinear System Identification: NARMAX Methods in the Time, Frequency, and Spatio-Temporal Domains. Wiley, 2013.
[3] O. Nelles, Nonlinear System Identification: From Classical Approaches to Neural Networks and Fuzzy Models. Berlin: Springer-Verlag, 2020.
[4] D. Gedon, N. Wahlström, T. B. Schön, and L. Ljung, "Deep state space models for nonlinear system identification," *IFAC-PapersOnLine*, vol. 54, no. 7, pp. 481–486, 2021.
[5] H. Zhou, C. Ibrahim, W. X. Zheng, and W. Pan, "Sparse Bayesian deep learning for dynamic system identification," Automatica, vol. 144, p. 110489, Oct. 2022.
[6] Y. Sun and H. L. Wei, "Efficient mask attention-based NARMAX (MAB-NARMAX) model identification," in *Proc. 27th International Conference on Automation and Computing (ICAC)*, Bristol, UK, 2022, pp. 1–6.
[7] R. Sasaki, S. Sugimoto, and Y. Shinya et al., "A deep neural network with module architecture for model reduction and its application to nonlinear system identification," *IFAC PapersOnLine*, vol. 56, no.2, pp.10650–10655, 2023.
[8] H. L. Wei, S. A. Billings, and J. Liu, "Term and variable selection for nonlinear system identification," *Int. J. Control*, vol. 77, pp. 86–110, Jan. 2004.
[9] H. L. Wei and S. A. Billings, "Model structure selection using an integrated forward orthogonal search algorithm assisted by squared correlation and mutual information," *Int. J. Model., Identificat. Control*, vol. 3, no. 4, pp. 341–356, Oct. 2008.
[10] S. Chen, S. A. Billings, and W. Luo, "Orthogonal least squares methods and their application to non-linear system identification," *Int. J. Control*, vol. 50, no. 5, pp. 1873–1896, Nov. 1989.
[11] S. Chen, C. F. N. Cowan, and P. M. Grant, "Orthogonal least squares learning algorithm for radial basis function networks," *IEEE Trans. Neural Netw*., vol. 2, no. 2, pp. 302–309, Mar. 1991.
[12] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. 27th Asilomar Conf. Signals, Syst. Comput.*, vol. 1. Pacific Grove, CA, USA, 1993, pp. 40–44.
[13] J. A. Tropp and A. C. Gilbert, ''Signal recovery from random measurements via orthogonal matching pursuit,'' *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.
[14] R. Tibshirani, "Regression shrinkage and selection via the LASSO," *J. Roy. Stat. Soc., Ser. B (Methodol.)*, vol. 58, no. 1, pp. 267–288, 1996.
[15] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning, 2nd ed. New York: Springer, 2009.
[16] S. A. Billings and H. L. Wei, "A new class of wavelet networks for nonlinear system identification," *IEEE Trans. Neural Netw.*, vol. 16, no. 4, pp. 862–874, Jul. 2005.
[17] H. L. Wei, S. A. Billings, and M. A. Balikhin, "Wavelet based non-parametric NARX models for nonlinear input–output system identification," *Int. J. Syst. Sci.*, vol. 37, no. 15, pp. 1089-1096, 2006.
[18] Y. Sun, I. Simpson, H. L. Wei, and E. Hanna, "Probabilistic seasonal forecasts of North Atlantic atmospheric circulation using complex systems modelling and comparison with dynamical models," Meteorological Applications, vol. 31, no. 1, e2178, 2024.
[19] S. Fattahi and S. Sojoudi, "Data-driven sparse system identification," in *Proc. Allerton Conference on Communication, Control, and Computing*, Monticello, IL, USA, 2018, pp. 462–469.
[20] A. Oraiopoulos and B. Howard,"One-step ahead modelling of building thermal dynamics: A comparison of backward selection and LASSO approaches," in *Proc. 5th IBPSA-England Conference on Building Simulation and Optimization*, Loughborough, UK, 2020, pp. 277-285.
[21] C. Cheng, E.-W. Bai, and Z. Peng, "Identification of sparse Volterra systems: An almost orthogonal matching pursuit approach," *IEEE Trans. Autom. Control*, vol. 67, no. 4, pp. 2027–2032, Apr. 2022.
[22] H. L. Wei, "System identification-informed transparent and explainable machine learning with application to power consumption forecasting," in *Proc. 3rd International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*, Tenerife, Spain, 2023, pp. 1-6.
[23] H. L. Wei, M. A. Balikhin, R. J. Boynton, and S. N. Walker, "Assessing uncertainty in space weather forecasting using quantile regression and complex nonlinear systems identification techniques," in *Proc. Int. Conference on Computing, Electronics & Communications Engineering (iCCECE)*, Swansea, UK, 2023, pp79-84.
[24] S. A. Billings and H. L. Wei, "An adaptive orthogonal search algorithm for model subset selection and non-linear system identification," *Int. J. Control*, vol. 81, no. 5, pp. 714-724, 2008.
[25] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.
[26] B. K. Natarajan, "Sparse approximate solutions to linear systems," SIAM Journal on Computing, vol. 24, no. 2, pp. 227–234, 1995.
[27] G. Schwarz, "Estimating the dimension of a model," Ann. Statist., vol. 6, no. 2, pp. 461-464, 1978.
[28] Ö. Alpan, "PID control of identified two tank system," [Online] https://www.mathworks.com/matlabcentral/fileexchange/39852-pid-control-of-identitified-two-tank-system), MATLAB Central File Exchange. Retrieved 8 Dec. 2023.
[29] L. A. Aguirre and S. A. Billings, "Improved structure selection for nonlinear models based on term clustering", *Int. J. Control*, 62, pp. 569-587, 1995.