Proceedings Paper:

Wei, H.-L. orcid.org/0000-0002-4704-7346 (2023) System identification-informed transparent and explainable machine learning with application to power consumption forecasting. In: 2023 3rd International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME) Proceedings. 2023 3rd International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME), 19-21 Jul 2023, Tenerife, Canary Island, Spain. Institute of Electrical and Electronics Engineers (IEEE) . ISBN 9798350322989

https://doi.org/10.1109/ICECCME57830.2023.10252535

Final Accepted Manuscript 06/04/2023.
Proc. of the International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME 2023)
19-20 July 2023, Tenerife, Canary Islands, Spain

System Identification-informed Transparent and Explainable Machine Learning with Application to Power Consumption Forecasting

Hua-Liang Wei [1,2,3,4]
[1] *Dept. Automatic Control and Systems Engineering*
[2] *Energy Institute;* [3] *Centre of Machine Intelligence*
[4] *Centre for Signal Processing and Cpmplex Systems*
The University of Sheffield
Sheffield, UK
w.hualiang@sheffield.ac.uk; ORCID: 0000-0002-4704-7346

***Abstract*— System identification (SysID) is the art and science of dealing with dynamic data modelling problems from systems science perspectives. It has been an active field and is still very active today, due to its wide range of applications, especially its basic principles of finding transparent, interpretable and parsimonious models for different purposes. The past decades have witnessed the explosive growth in machine learning (ML) and its applications in all areas of science and engineering. Meanwhile, there has been an increasing demand for the development of transparent, explainable and/or interpretable ML models. This paper proposes a new framework for developing System Identification-informed Transparent and Explainable MAchine Learning (SITEMAL) models. A case study, involving a real power consumption dataset, is presented to demonstrate the application of the proposed modelling framework and its performance for power consumption forecasting.**

Keywords—system identification, machine learning, transparent model, explainable model, power consumption.

I. Introduction

System identification (SysID) is the art and science of dealing with data modelling problems from systems science perspectives, as stated in the pioneering paper [1]: "identification is the determination, on the basis of input and output, of a system within a specified class of systems, to which the system under test is equivalent." There has been much overlap and interplay between SysID and machine learning (ML) during their entire evolution, from the beginning until now. Both of them have their own main specific focuses. SysID pays more attention to best model identification, or more specifically, best model structure detection, which are usually transparent and parsimonious so as to facilitate late-stage analysis, explanation, interpretation or other purposes. Machine learning is more concerned with achieving best prediction performances than with model explanation and interpretation. Most ML models, especially deep learning (DL) models, are black-box, which may produce good predictions but lack interpretability [2]-[5].

In system identification, it is usually assumed that a priori knowledge about the system's inherent structure is not available. For a given system, if the number of observations or measurements of the input and output signals is sufficiently large and the data are sufficiently informative (e.g., the input data are rich enough, or ideally persistently exciting), then most of the state-of-the-art system identification methods (see e.g. [6]-[8]) can provide good models for the system.

Machine learning has a broad spectrum, involving a large group of computational methods and algorithms for building models for solving a variety of tasks, which can be roughly categorized into two main streams: classification and regression. Most tasks, where ML is commonly used, only involve static data; some others involve data from dynamical systems but the central objectives are to make predictions, with little attention being paid to model transparency, explanation and/or interpretation. In comparison with ML, SysID covers a relatively narrower spectrum of model types, which are built based on data from dynamical systems and usually used for a variety of purposes including control [9]-[11], system analysis [12][13], interpretation and simulations [14]-[18], and prediction or forecasting as well [19][20]. It is worth mentioning that there has been continuing interplay or synergy between SysID and ML over the past years. For example, in [21]-[23], ML models were used for system identification but the main focus was on forecasting. In [24]-[26], deep neural networks were used as a tool for nonlinear system identification, and in [27][28], the link, interplay and synergy between SysID and ML were investigated.

This paper is mainly concerned with modelling and analysis of multi-input systems, aiming to develop a new framework for building System Identification-informed Transparent and Explainable MAchine Learning (SITEMAL) models. The main contributions of the paper include the proposal of a SITEMAL framework, and the identification of sparse and transparent predictive models which are useful when interpretation and/or explanation becomes necessary or highly desirable.

II. A Brief Overview of System Identification and Machine Learning

This section presents a brief overview of SysID and ML, and the focus will be on modelling methods for multi-input systems.

A. System Identification

Mathematical models play central roles in almost all areas of science and engineering. For many complex systems, especially complex nonlinear dynamical systems, it is almost impossible to establish accurate mathematical models using first-principle modelling approaches. Data-based or data-driven modelling methods, can therefore provide a highly effective and attractive alternative for obtaining system models. For dynamical systems, the technique that is used to derive models from measured input and output signals is called system identification [6]-[8], where it is usually assumed that a priori knowledge about the system's inherent structure is not available. As an example, consider the case of multiple-input, single-output (MISO) systems, with an assumption that the system true model structures are not known, but the system input and output signals are available. A simple illustration graph is shown in Fig. 1.

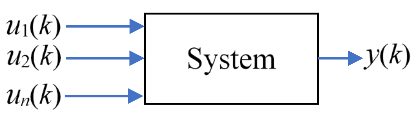

Fig. 1. A multiple-input, single-output system.

Such a system can usually be represented by a Nonlinear Lagged Inputs and Outputs (NLIO) model [9]-[13] as follows:

$$\begin{aligned} y(k) = f\,(& y(k-1), y(k-2)\ldots, y(k-n_y), \\ & u_1(k-\tau), u_1(k-\tau-1), \ldots, u_1(k-n_u), \ldots, \\ & u_v(k-\tau), u_v(k-\tau+1), \ldots, u_v(k-n_u)) + e(k) \end{aligned} \tag{1}$$

where $u_1(k)$, $u_2(k)$,…, $u_v(k)$ are v input sequences, $y(k)$ 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 $\tau = 1$; $f(\cdot)$ is some unknown function that needs to be built from available training data. Model (1) is a special case of the well-known Nonlinear AutoRegressive Moving Average with eXogenous inputs (NARMAX) model [7][29][30].

In many applications, only a small number of important model terms are needed for characterizing the system behaviours. An efficient model structure detection method is highly needed to select the most significant model terms. One of the most efficient and commonly used algorithms for model term selection is orthogonal least squares (OLS) [31] and its variants, such as forward regression with orthogonal least squares (FROLS)[7]. A detailed description and pseudo-code of the OLS algorithm can be found in [32] and [33]. Models produced by these algorithms are usually transparent, interpretable, parsimonious and simulatable (TIPS) [14]-[16].

The following well-known single-input, single-output (SISO) AutoReressive with eXogenous (ARX) model is a special case of the NARL model:

$$\begin{aligned} y(k) = {} & a_0 + a_1 y(k-1) + a_2 y(k-2) + \ldots + a_{n_y} y(k-n_y) \\ & + b_1 u(k-1) + b_2 u(k-2) + \ldots + b_{n_u} u(k-n_u)) + e(k) \end{aligned} \tag{2}$$

B. Machine Learning

Machine learning, as a main branch of artificial intelligence (AI), aims to enable computers to learn from data, and gradually improve the leaning ability and power with time, without needing to be explicitly programmed. ML algorithms are able to find patterns (e.g., inherent dynamics, features, anomalies, links and relationships between different variables) in training data and learn from data, and gradually are able to make predictions or even decisions by machines themselves.

Fig.2(a) shows the structure of an artificial neural network (ANN) machine learning model for a system with 10 inputs and 1 output. The structure of the ANN model is assumed to be known here, but in practice the following information is not known: what happens within the three hidden layers, how the input variables interplay or interact with each and how the output is related to these input variables. So, unlike in SysID, where models are usually transparent, most ML models, especially ANN (including deep neural network) models, are opaque. That is why ANN is usually referred to as a black-box modelling approach, as shown in Fig.2(b), where the relationships between the inputs and output are not explicitly known.

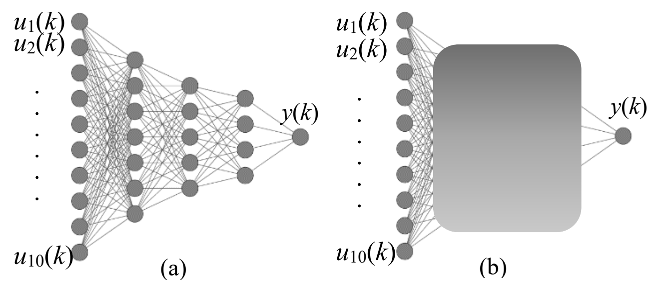

Fig. 2. A neural network model for a 10 inputs, 1 output system.

C. A Simple Comparison Between System Identification and Machine Learning

A basic comparison between SysID and ML is presented in Table 1.

TABLE I. Simple Comparison Between SysID and ML

Features	**SysID**	**ML**
Model transparency	Yes (generally)	No (generally)
Model interpretability	Yes (generally)	No (generally)
Does it need large data?	No	Yes (generally)
Does it work for small data?	Yes	No (generally)
Nonlinear presentation ability	It depends	Very strong
Generalization ablity	Good	It depends
For regression problems?	Yes	Yes
For images (e.g. classification)?	No	Yes

III. The Proposed SITEMAL Framework

For linear modelling problems, there exist mature methods and algorithms [6], so this section focuses on nonlinear system modelling using both nonlinear SysID methods and ML techniques.

A. Nonlinear System Identification

For simplicity and convenience of description, we start with a simple special case of the NLIO model (1), where v, n_y, n_u, and τ are set to v=2, $n_y = n_u$ =1, v=1 and $\tau = 0$, therefore, model (1) reduces to

$$y(k) = f(y(k-1), u(k), u(k-1)) + e(k) \tag{3}$$

In practical nonlinear SysID, the unknown function $f(\cdot)$ can be approximated by using a set of basis functions. The most commonly used basis functions are polynomials [7]. To construct NLIO polynomial models, a dictionary consisting of a good number of candidate model terms is usually defined first. For example, for the above system (3), a dictionary of nonlinear degree 2 is as follows:

$$D = \begin{Bmatrix} y(k-1),\ u(k),\ u(k-1),\ y^2(k-1), \\ y(k-1)u(k),\ \ y(k-1)u(k-1), \\ u^2(k),\ \ \ u(k)u(k-1),\ \ \ u^2(k-1) \end{Bmatrix} \tag{4}$$

The orthogonal least squares [28] and its variants (see e.g. [12][13][29][30]) can be used to determine which model terms are important and should be included in the model. The final identified model could be quite simple as follows:

$$y(k) = 0.6y(k-1) + 1.5u(k-1) + 25y(k-1)u(k-1) \tag{5}$$

Note that the nonlinear degree of the above model is ℓ =2, which is determined by the highest degree of all the model terms.

B. System Identification-informed Machine Learning

A multi-input nonlinear system identification task can be described as follows. There is a response variable y that is dependent on a set of explanatory variables $\mathbf{x} = \{x_1, x_2, ..., x_n\}$. A collection of observations of the input and output variables are available, which are denoted by $\{\mathbf{x}(k), y(k)\}$ (k =1, 2,…, N). The true dependent relationship between y and $\mathbf{x}$ is not known. The main task of SysID is to find a model $\hat{\mathbf{y}} = f(\mathbf{x})$ that approximates the input-output relationship as close as possible. Note that in many real applications, the value of the output y at time instant k depends on previous values of the input and output signals. For example, if a system can be well represented by model (3), then its output value y at instant k depends on $u(k)$, $u(k-1)$ and $y(k-1)$. So, the set of $\mathbf{x}$ is composed of these lagged input and output variables, that is, $\mathbf{x}(k)=\{x_1(k), x_2(k), x_3(k)\} = \{u(k), u(k-1), y(k-1)\}$.

The basic idea of the proposed SITEMAL framework is to take advantage of the models produced by nonlinear system identification techniques, which have TIPS (transparent, interpretable, parsimonious and simulatable) properties [14]-[16]. Some powerful nonlinear system identification methodologies, such as the NARMAX methodology [7], provide useful information on how many lagged variables are involved in the modelling procedure, which variables interplay or interact with others, which regressors (terms) are important, and so on. Such information is useful for building ML models.

The implementation process of SITEMAL is as follows.

1) Building a NLIO model using NARMAX methods;
2) Building an ML model using the obtained NLIO model;
3) The NLIO model can be used to explain which input variables are important and how these important input variables interactively and collectively determine the system response. In the meantime, the NLIO+ML model can be used to predict the system future behaviour.

Note that in the above step 2), the ML model is built as follows. Let the NLIO model built in step 1) is $\hat{\mathbf{y}} = f(\mathbf{x}_s)$, where $\mathbf{x}_S$ is a subset of $\mathbf{x}$, whose elements are those variables appearing in the NLIO model built in step 1). The error or residual of the model is:

$$\mathbf{r} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - f(\mathbf{x}_s) \tag{6}$$

Theoretically, if the NLIO model is good enough to represent the input-output relationship of the system of interest, then $\mathbf{r}$ cannot be predicted using the predictors in $\mathbf{x}_S$; otherwise, another ML method can be used to predict the error $\mathbf{r}$. In this way, features and patterns that are not captured by the NLIO model may be further exploited and revealed by the ML model.

Let the ML model built in step 2) is $\hat{\mathbf{r}} = g(\mathbf{x}_s)$, then

$$\mathbf{r} = \hat{\mathbf{r}} + \xi = g(\mathbf{x}_s) + \xi \tag{7}$$

Combining (6) and (7), yields,

$$\mathbf{y} = f(\mathbf{x}_s) + g(\mathbf{x}_s) + \xi \tag{8}$$

where ξ is the new error when $g(\mathbf{x}_s)$ is added to $f(\mathbf{x}_s)$.

IV. Case Study on Power Consumption Prediction

To evaluate the performance of the proposed approach, a case study is performed on a real power consumption dataset, measured at Tetouan, a city in northern Morocco.

A. Data

The energy distribution network at Tetouan is powered by three source stations. Data in three power consumption zones, namely, Zones 1, 2 and 3, were collected every 10 minutes, between 00.00.00, 01-01-2017 and 23.50.00, 30-12-2017. There are a total of 52416 observations. The dataset is a good comprehensive benchmark for analysing the power consumption patterns at Tetouan. More details about the dataset can be found in [34]. As an illustration, the first 4464 values of the power consumption in Zone 1 (measure in January 2017) are shown in Fig. 3.

The dataset involves a total of 8 variables as follows: 1) Five explanatory variables; 2) Three responses, corresponding to three power distribution networks for three consumption zones, Zones 1,2, and 3. These variables are shown in Table II.

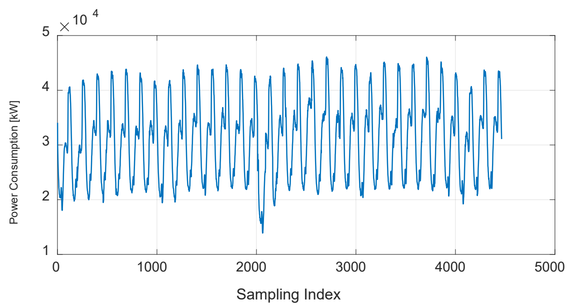

Fig. 3. Power consumption in Zone 1 of Tetouan city in Janury 2017.

TABLE II. EXPLANATORY AND RESPONSE VARIABLES INVOLVED IN THE POWER CONSUMPTION DATASET

Variables	Description
Temperature (x_1)	Weather temperature of Tetouan city (ºC)
Humidity (x_2)	Weather humidity of Tetouan city (g·m^{-3})
Wind speed (x_3)	Wind speed of Tetouan city (unit: unclear)
General diffuse flows (x_4)	See below
Diffuse flows (x_5)	A catchall term to describe low-temperature (< 0.2° to ~ 100°C) fluids (unit: unclear)
Zones 1,2,3 (y_1, y_2, y_3)	Power consumption (units: kW) in the three zones

B. Models Obtained Through System Identification

In the literature, the power consumption data described in the previous section have been analysed using different ML models, where it was assumed that the value of a response y at the present time instant k is only dependant on the values of the five predictors at the same instant, that is, $x_1(k)$, $x_2(k)$, …, $x_5(k)$. In this study, however, the data are analysed using dynamic models such as the NLIO model presented by (1), where the value of a response y at time instant k is assumed to be potentially dependent on the historical values of itself, as well as historical values of these explanatory variables.

Some basic information about the modelling experiments are as follows:

- The first 4464 samples of the power consumption in January 2017 are used for model identification, the remaining 47982 samples (February - December) are used for testing.
- The behaviour of some explanatory variables may potentially affect the response variable (power consumption) in a long period (e.g. within 24 hours), so the maximum time lag was set to $6 \times 24 = 144$.
- For 10 minutes ahead prediction, the time delay between inputs and outputs was set to $\tau = 1$; for 60 minutes ahead prediction, the time delay was set to $\tau = 6$; and for 120 minutes ahead prediction, the time delay was set to $\tau = 12$.

For each of the three response variables y_1, y_2, and y_3, three separate models were built for 10 minutes, 60 minutes and 120 minutes ahead predictions, respectively. So, a total of 9 models were obtained. Some information about the 9 models is given in Table III. The performances of these 9 models, measured by root mean square error (RMSE) and mean absolute error (MAE), are shown in Table IV. It is worth stressing that no cross-product interaction (between variables) was identified to be significantly important by the system identification algorithms; this probably suggests that interplays or interactions between these candidate predictors may not play a strong role in predicting the power consumption.

TABLE III. NUMBER OF MODEL TERMS IN THE IDENFIED 9 MODLES

Models	Zone 1 No of model terms	Zone 2 No of model terms	Zone 3 No of model terms
10 minutes ahead prediction	8	8	5
60 minutes ahead prediction	14	9	12
120 minutes ahead prediction	16	17	17

TABLE IV. MODEL PERFORMANCES OF THE 9 IDENTIFIED MODELS

Models	Zone 1 RMSE and MAE		Zone 2 RMSE and MAE		Zone 3 RMSE and MAE	
	Train	Test	Train	Test	Train	Test
10 minutes ahead	360.97 253.26	462.31 304.74	268.98 183.45	311.04 216.33	230.15 166.18	301.85 197.76
1 hour ahead	2184.42 1588.29	2560.82 1866.34	1236.14 929.98	1698.68 1280.52	1626.41 1136.79	2591.68 2013.07
2 hours ahead	3610.84 2752.94	4496.40 3259.81	2067.42 1659.11	3291.41 2527.90	2661.52 1970.96	4696.88 3655.37

Taking the case of Zone 1 as an example, the three models, for 10, 60 and 120 minutes ahead predictions, are as follows:

$$\begin{aligned} y_1(k) = &1.4566y_1(k-1) - 0.4004y_1(k-2) \\ &-0.4347x_4(k-59) - 0.0716y_1(k-4) \\ &+0.4320x_2(k-62) + 0.8648x_4(k-31) \\ &+0.4383x_4(k-126) + 360.2650 \end{aligned} \tag{9}$$

$$\begin{aligned} y_1(k) = &1.9916y_1(k-6) - 1.5649x_4(k-63) \\ &-1.2144y_1(k-7) - 2.2550x_1(k-30) \\ &+4.3326x_4(k-35) + 3.8127x_4(k-130) \\ &+6634.53 + 3.3817x_4(k-40) \\ &-2.7710x_4(k-99) - 4.8578x_4(k-58) \\ &+3.2431x_4(k-45) - 2.3600x_4(k-71) \\ &-4.8578x_5(k-9) - 1.9348x_4(k-9) \end{aligned} \tag{10}$$

$$\begin{aligned} y_1(k) = &1.3122y_1(k-12) - 3.6741x_4(k-66) \\ &+6.2341x_4(k-40) + 13086.46 \\ &+4.2443x_4(k-135) - 1.5941x_4(k-96) \\ &-0.7589y_1(k-13) + 6.7947x_4(k-37) \\ &+7.0727x_4(k-45) - 7.4548x_4(k-61) \\ &-4.9683x_4(k-72) + 5.113x_5(k-12) \\ &-4.5868x_4(k-12) - 3.4912x_4(k-102) \\ &+2.1710x_5(k-48) + 4.3478x_4(k-137) \end{aligned} \tag{11}$$

Note that the terms in the above models are arranged according to their importance for explaining the variation of the response.

Also note that in comparison to other ML methods, these models show far better prediction performance. For example, for the case of 60 minutes ahead prediction, the values of RMSE and MAE are 2184.42 and 1588.29, respectively, which are far smaller than the results produced by RF (random forest) [34], where the values of RMSE and MAE were 21109.7 and 15442.0, respectively. It is worth mentioning that the results, 21109.7 and 15442.0, produced by RF are the best ones among five ML methods including RF, DT (decision tree), SVR (support vector regression), FFNN (feed-forward neural network), and LR (linear regression) [34].

A comparison between the 60 minute ahead predictions from model (10) and the true observations for Zone 1 are shown in Fig. 4, where for a clear visualization, only 4320 data points of period between 00.00.00, 01-12-2017 and 23.50.00, 30-12-2017, are shown. The scatter plot between the model predicted values and the measurements is shown in Fig. 5.

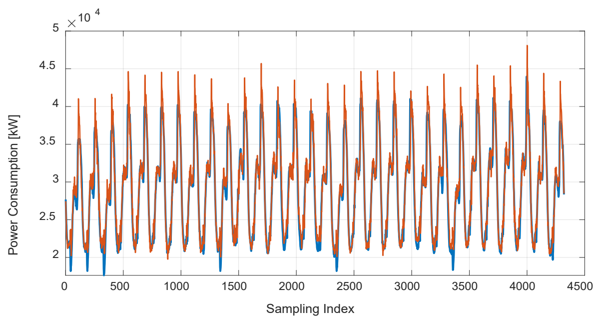

Fig. 4. One hour ahead prediction of the power consumption in Zone 1 of Tetouan city December 2017.

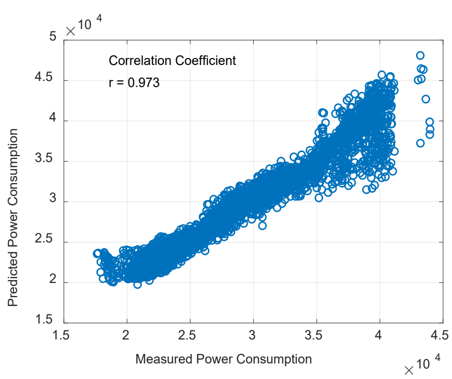

Fig. 5. One hour ahead prediction of the power consumption in Zone 1 of Tetouan city in December 2017.

C. Potential Enhancement of Prediction with Deep Learning

As discussed in Section IV-B, models obtained by using system identification techniques may not always be able to sufficiently capture the input-output relationships of the data. One way to know if the obtained models are good enough is to model the modelling error, i.e., $\mathbf{r}$ defined in (6), using another ML method. In doing so, long-short term memory (LSTM) neural networks were used to model the errors of the nine identified models, and it tuned out that modelling the errors of these nine models using LSTM networks did not help improve their prediction performances. This may suggest that these nine models sufficiently capture the input-output relationships of the data, and their errors are unpredictable. Therefore, the nine models can now be used to make predictions of power consumption of Tetouan city, as well as analysing how power consumption depends on the available influential predictors.

Taking model (11) as an example, the prediction result given by LSTM for the error signal of model (11) is shown in Fig. 6, where it can be observed that the LSTM model just simply tracks the mean of the error signal, meaning that the error signal is unpredictable.

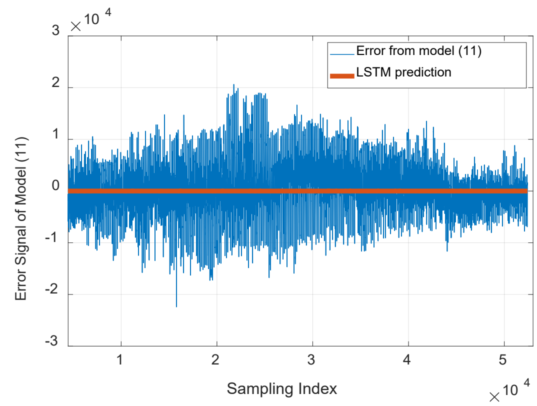

Fig. 6. LSTM prediction result of the error signal of model (11), over the test data, 00.00.00, 01-01-2017 – 23.50.00, 30-12-2017.

V. Conclusion

A system identification-informed transparent and interpretable machine learning modelling framework, called SITEMAL, is proposed. The prediction results, in comparison to other ML methods, are far better and are really promising and encouraging. The excellent performances partly benefit from the use of some long-lagged explanatory variables such as $x_4(k\text{-}137)$ (general diffuse flows), which were detected by the nonlinear system identification algorithms (e.g., FROLS), suggesting that the behaviour of some explanatory variables may potentially affect the electric power consumption in a long period. In addition to transparency and interpretability, the framework has several other attractive advantages such as sparsity, parsimony, no requirement for a large training dataset, and working fast. In this study, LSTM was considered to explore the useful information from the modelling error signals of the models identified by means of system identification techniques. In future, we will try more state-of-the-art deep learning methods, especially those that are powerful in prediction but weak in explanation so that the ability of the proposed framework can be further enhanced.

Acknowledgment

This work was supported in part by NERC (Ref. NE/V002511/1, Ref. NE/V001787/1, and Ref. NE/W005875/1), EPSRC (Ref. EP/H00453X/1 and Ref. EP/I011056/1), and EU Horizon 2020 Research and Innovation Programme Action Framework under the grant Ref. 637302.

REFERENCES

[1] L. A. Zadeh, "From circuit theory to system theory," Proc. I.R.E., vol. 50, pp. 856-865, May 1962.

[2] C. Voyant, G. Notton, S. Kalogirou, M.-L. Nivet, C. Paoli, F. Motte, and A. Fouilloy, “Machine learning methods for solar radiation forecasting: A review,” Renew. Energy, vol. 105, pp. 569–582, May 2017.

[3] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, “Statistical and machine learning forecasting methods: Concerns and ways forward,” PLoS ONE, vol. 13, art. no. e0194889, Mar. 2018.

[4] A. Mosavi et al., “State of the art of machine learning models in energy systems, a systematic review,” Energies, vol. 12, 1301, April 2019.

[5] H. Demolli, A. S. Dokuz, A. Ecemis, and M. Gokcek, “Wind power forecasting based on daily wind speed data using machine learning algorithms,” Energy Convers. Manag., vol. 198, 111823, October 2019.

[6] L. Ljung, System Identification Toolbox User's Guide. The MathWorks, Inc., 2022.

[7] S. A. Billings, Nonlinear System Identification: NARMAX Methods in the Time, Frequency, and Spatio-Temporal Domains, Wiley, 2013.

[8] O. Nelles, Nonlinear System Identification: From Classical Approaches to Neural Networks, Fuzzy Models, and Gaussian Processes. Springer Nature: Berlin/Heidelberg, Germany, 2020.

[9] C. Deiler, “Aerodynamic modeling, system identification, and analysis of iced aircraft configurations,” J. Aircr. vol. 55, pp.145–161. July 2017.

[10] P. Lichota, F. Dul, A. Karbowski, “System identification and LQR controller design with incomplete state observation for aircraft trajectory tracking,” Energies, vol. 13, 5354, October 2020.

[11] A. Simorgh, A. Razminia, and V. I. Shiryaev, “System identification and control design of a nonlinear continuously stirred tank reactor,” Math. Comput. Simul., vol. 173, pp. 16–31, February 2020.

[12] R.J. Hall, H.-L. Wei, and E. Hanna, “Complex systems modelling for statistical forecasting of winter North Atlantic atmospheric variability: A new approach to North Atlantic seasonal forecasting,” Q. J. R. Meteorol Soc., vol. 145, pp. 2568-2585, June 2019.

[13] Y. Gu et al., "System identification and data-driven forecasting of AE index and prediction uncertainty analysis using a new cloud‐NARX model," J. Geophys. Res., vol. 124, pp. 248-263, December 2018.

[14] S. Billings and H.-L. Wei, “NARMAX model as a sparse, interpretable and transparent machine learning approach for big medical and healthcare data analysis,” in Proc. IEEE 21st Int. Conf. High Perform. Comput. Commun.; IEEE 17th Int. Conf. Smart City; IEEE 5th Int. Conf. Data Sci. Syst. (HPCC/SmartCity/DSS), Aug. 2019, pp. 2743–2750.

[15] H.-L. Wei, “Sparse, interpretable and transparent predictive model identification for healthcare data analysis,” in Advances in Computational Intelligence. Springer International Publishing, pp. 103–114, May 2019.

[16] H.-L. Wei and S. A. Billinsg, "Modelling COVID-19 pandemic dynamics using transparent, interpretable, parsimonious and simulatable (TIPS) machine learning models: A case study from systems thinking and system identification perspectives," in Recent Advances in AI-enabled Automated Medical Diagnosis, CRC Press, 2022, pp. 13-28.

[17] H. Chen et al., “Model-based method with nonlinear ultrasonic system identification for mechanical structural health assessment,” Trans. Emerg. Telecommun. Technol., vol. 31, e3955, March 2020.

[18] T. Weber et al., “Machine learning based system identification tool for data-based energy and resource modeling and simulation,” Procedia CIRP, vol. 80, pp. 683–688, May 2019.

[19] B. Hadid, E. Duviella, and S. Lecoeuche, “Data-driven modelling for river flood forecasting based on a piecewise linear ARX system identification,” J. Process Control, vol. 86, pp. 44-56, February 2020.

[20] C. McHugh, S. Coleman, and D. Kerr, “Hourly electricity price forecasting with NARMAX,” Mach. Learn. Appl., vol. 9, 100383, 2022.

[21] Y. Gu et al., “Machine learning enhanced narmax model for dst index forecasting,” in 25th International Conference on Automation and Computing (ICAC), September 2019.

[22] N. Nikentari and H.-L.Wei, “Multi-task learning for time series forecasting using narmx-lstm,” in IEEE 27th Int. Confer. Automation and Computing (ICAC), September 2022.

[23] T. de Almeida Ushikoshi, L. A. Aguirre, “Guidelines for choosing hyperparameters of echo state networks for system identification: two case studies,” Int. J. Model. Identif. Control., vol. 42, pp. 105-115, March 2023.

[24] M. Schussler, T. Munker, and O. Nelles, “Deep recurrent neural networks for nonlinear system identification,” in 2019 IEEE Symposium Series on Computational Intelligence (SSCI), December 2019, pp. 448–454.

[25] C. Andersson et al., “Deep convolutional networks in system identification,” in Proc. IEEE 58th IEEE Conference on Decision and Control. Nice, France, March 2019.

[26] H. Zhou, C. Ibrahim, W.X. Zheng, and W. Pan, “Sparse bayesian deep learning for dynamic system identification,” Automatica, vol. 144, 110489, October 2022.

[27] L. Ljung, C. Andersson, K. Tiels, and T. B. Schon, “Deep learning and system identification,” IFAC-PapersOnLine, vol. 53, 1175–1181, July 2020.

[28] K. Fujimoto, S. Sakaino, and T. Tsuji, ‘‘Time series motion generation considering long short-term motion,’’ in Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS), Nov. 2019, pp. 6842–6848.

[29] I. J. Leontaritis and S.A. Billings, "Input-output parametric models for non-linear systems Part I: Deterministic non-linear systems," Int. J. Control, vol. 41, pp. 303–328, 1985.

[30] S. Chen and S.A. Billings, "Representations of non-linear systems: the NARMAX model," Int. J. Control, vol. 49, pp. 1013-1032, 1989.

[31] 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, pp. 1873–1896, 1989.

[32] H.-L. Wei, S. A. Billings, and J. Liu, “Term and variable selection for non-linear system identification,” Int. J. Control, vol. 77, pp. 86-110, January 2004.

[33] 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. Identif. Control., vol. 3, pp. 341-356, October 2008.

[34] A. Salam and A. Hibaoui, “Comparison of machine learning algorithms for the power consumption prediction—Case study of Tetouan city,” in Proc. 2018 6th Int. Renewable and Sustainable Energy Conference (IRSEC), Rabat, Morocco, 5–8 December 2018.