





NonSysId: Nonlinear System Identification with Improved Model Term Selection for NARMAX Models

Rajintha Gunawardena ¹, Zi-Qiang Lang ², and Fei He ¹

¹ Centre for Computational Science and Mathematical Modelling, Coventry University, Coventry CV1 5FB, UK ² School of Electrical and Electronic Engineering, The University of Sheffield, Western Bank, Sheffield S10 2TN, UK

DOI: [10.21105/joss.08028](https://doi.org/10.21105/joss.08028)

Software

- [Review](#) 
- [Repository](#) 
- [Archive](#) 

Editor: [Sébastien Boisgérault](#) 

Reviewers:

- [@JAuriol81](#)
- [@brivadis](#)

Submitted: 16 December 2024

Published: 21 October 2025

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#)).

Summary

System identification lies at the intersection of control theory, dynamic systems theory and machine learning, involving the construction of mathematical models of complex dynamic linear or nonlinear systems using input-output data, enabling the prediction of system behaviour and analysis in both time and frequency domains. This approach can model the entire system or capture specific dynamics within it. For meaningful analysis, it is essential that the model accurately reflects the underlying system behaviour. This paper introduces NonSysId, an open-source MATLAB software package designed for nonlinear system identification, specifically focusing on NARMAX models. The software incorporates an advanced term selection methodology that prioritises simulation (free-run) accuracy while preserving model parsimony. A key feature is the integration of iterative Orthogonal Forward Regression (iOFR) with Predicted Residual Sum of Squares (PRESS) statistic-based term selection, facilitating robust model generalisation without the need for a separate validation dataset. Furthermore, techniques for reducing computational overheads are implemented. These features make NonSysId particularly suitable for real-time applications, such as structural health monitoring, fault diagnosis, and biomedical signal processing, where capturing signals under consistent conditions is challenging, resulting in limited or no validation data.

Statement of Need

The package focuses on discrete-time polynomial nonlinear auto-regressive with exogenous input (NARX) models, with ARX as the linear special case. Here, the polynomial functional mapping relates past input/output terms (input/output lagged terms) to the current output. Polynomial NARX models are equivalent to RNNs ([Sum et al., 1999](#)) and are widely used in fields such as control, condition monitoring, advanced manufacturing, and the modelling and analysis of physiological and biological systems ([Chiras et al., 2002](#); [Gao et al., 2023](#); [He et al., 2016](#); [He & Yang, 2021](#); [Liu et al., 2024](#); [Ritzberger & Jakubek, 2017](#); [K. Wang et al., 2024](#); [Zainol et al., 2022](#)). Extending the NARX model with a noise model produces the NARMAX model.

The core challenge is identifying a parsimonious structure from a large set of candidate terms, as redundant terms may induce significant spurious dynamics ([Aguirre & Billings, 1995](#); [Mendes & Billings, 1998](#)). The Orthogonal Forward Regression (OFR) algorithm, also known as FROLS ([Billings et al., 1987](#); [S. Chen et al., 1989](#)), evaluates the independent contribution of each candidate term to the output according to a term selection criterion ([Hong et al., 2003](#); [Korenberg et al., 1988](#); [L. Wang & Cluett, 1996](#)), enabling efficient step-wise forward selection of terms. Using the error reduction ratio (ERR), a common criterion, the OFR incrementally selects terms that maximise the explained variance (model fit). Recently, two open-sourced packages for system identification have been introduced: SysIdentPy ([Junior](#)

et al., 2020) for Python and the narmax package (Ayala et al., 2020) for R. Both packages, based on the original OFR algorithm, are well-developed and comprehensive. However, the traditional OFR method (using ERR) has known limitations (Mao & Billings, 1997; Piroddi & Spinelli, 2003). These include the selection of redundant/incorrect terms, overfitting and inaccuracies/instability in long-horizon simulation (free-run or model-predicted output). These arise from inputs that do not persistently excite the system under consideration (i.e., inputs don't excite all system dynamics effectively) or from complex noise structures (Mao & Billings, 1997; Piroddi & Spinelli, 2003). Furthermore, models must generalise well to unseen data, typically assessed through cross-validation (Little et al., 2017; Stone, 1974). However, in some real-time applications, such as structural health monitoring or fault diagnosis, obtaining separate validation data may be impractical (Gharehbaghi et al., 2022; Vamsikrishna & Gijo, 2024). In neuroscience, rapidly changing complex dynamics between brain regions make it difficult to obtain data precisely containing the same behaviour for validation purposes (T. Chen et al., 2016; Eichenbaum et al., 2021; Kunjan et al., 2021; Lehnertz et al., 2021; Seedat et al., 2024). The following section outlines the features in the NonSysId package, specifically developed to address these challenges.

Features in NonSysId

The NonSysId package introduced in this paper implements an OFR-based system identification methodology designed to address the key issues outlined in the previous section. This is achieved by integrating and extending several OFR variants already available in the literature (Guo et al., 2015; Hong et al., 2003; L. Wang & Cluett, 1996), and a simulation-based model selection procedure proposed here. At the time of writing, NonSysId is the only open-source package that implements previously proposed solutions to overcome the limitations of the original OFR algorithm.

To address these limitations, the NonSysId package incorporates the iterative OFR (iOFR) algorithm (Guo et al., 2015), simulation-based stability tests (bounded-input bounded-output), and model selection techniques for improved long-horizon prediction and stability. Specifically, the candidate models generated at each iteration of iOFR are simulated using constant inputs of zeros and ones separately. If the corresponding model outputs remain bounded with a small variance, typically less than 10^{-2} , the model is regarded as stable. From this set of stable models, the Bayesian Information Criterion (BIC) (Schwarz, 1978; Stoica & Selen, 2004), based on the mean squared simulated error (MSSE) (Piroddi & Spinelli, 2003), is used to select the best model for the current iteration. Furthermore, the package integrates the PRESS-statistic-based (Allen, 1974) term selection criterion (Hong et al., 2003; L. Wang & Cluett, 1996), enabling efficient stepwise forward selection of terms that minimise the one-step-ahead leave-one-out cross-validation error. These enhancements provide robust term selection (compared to OFR-ERR), built-in cross-validation, and models with improved long-horizon prediction capabilities and simulation stability (Aguirre et al., 2010). For NARX models, the candidate term set can be extensive and computationally demanding in the iFRO algorithm (Billings et al., 1987; S. Chen et al., 1989; Guo et al., 2015). NonSysId, following (Wei et al., 2004) and (Guo et al., 2015), incorporates strategies to accelerate the iterative stepwise forward selection process. Additionally, the package includes correlation-based residual analysis techniques for validating nonlinear models (Billings & Voon, 1983). These features make NonSysId particularly suitable for real-time applications where inputs may not be persistently exciting and separate validation datasets may be unavailable. Detailed descriptions of the NonSysId package's methodology, including pseudocode for core procedures and approaches used to reduce computational time, are available in the supplementary documentation.

Examples

This section presents examples from the NonSysId package, which implements iOFR_S with PRESS-based term selection, utilising synthetic data from a NARX model and real data from an electromechanical system.

Synthetic data example

The following example demonstrates how to identify a NARX model using the NonSysId package. In this example, we consider a NARX model of a DC motor (Equation 1) as described in (Lacerda Junior et al., 2019).

$$\begin{aligned} y(t) = & 1.7813y(t-1) - 0.7962y(t-2) + 0.0339u(t-1) + 0.0338u(t-2) \\ & - 0.1597y(t-1)u(t-1) - 0.1396y(t-1)u(t-2) \\ & + 0.1297y(t-2)u(t-1) + 0.1086y(t-2)u(t-2) + 0.0085y(t-2)^2 \end{aligned} \quad (1)$$

In Equation 1, $y(t)$ is the output and $u(t)$ is the input at the time sample t . The NARX model is excited using two inputs: (a) White noise, $u(t) \sim \mathcal{N}(0, 1)$, and (b) a multi-tone sinusoidal, $u(t) = 0.2(4 \sin(\pi t) + 1.2 \sin(4\pi t) + 1.5 \sin(8\pi t) + 0.5 \sin(6\pi t))$. Matlab scripts for this example are available in the code repository, along with documentation in the code repository, which provides a straightforward guide for using iOFR_S in the NonSysId package.

Figure 1 and Figure 2 depict the training and testing data alongside the model-simulated output, $\hat{y}(t)$, for the inputs (a) and (b), respectively. Testing data refers to data not explicitly included during training, as the model is already cross-validated during the identification/training process.

Table 1 and Table 2 list the identified terms and parameter values of the NARX models for inputs (a) and (b), along with the mean squared PRESS error and ERR for each term. The PRESS error estimates the predicted cross-validation error at each step, while sorting the tables by PRESS error reveals the order of term inclusion during forward selection. ERR indicates the portion of the output variance explained by each term.

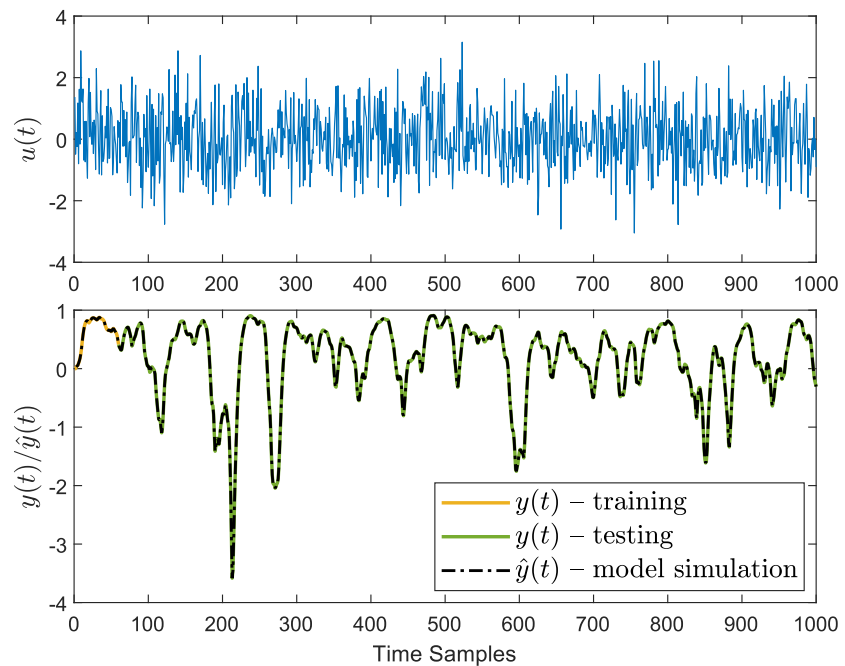


Figure 1: Model identification results under input (a). Only the first 60 samples are used for identifying/training the model. The error variance is 1.6018×10^{-25} .

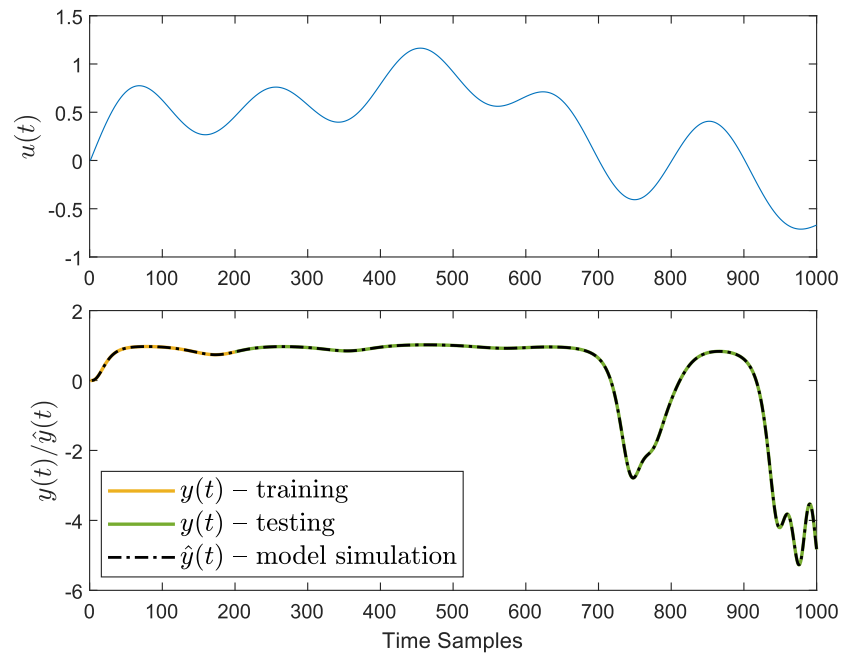


Figure 2: Model identification results under input (b). Compared to input (a), fewer system dynamics are excited—limited frequency components in the input signal. Therefore, to avoid a suboptimal model, up to 200 samples are used to identify the model. The error variance is 8.2178×10^{-18} .

Table 1: The model identified when Equation 1 is excited with input (a), white noise

Model term	Mean squared PRESS error	ERR	Parameters/Coefficients
$y(t-1)$	1.342×10^{-3}	0.95001	1.7813
$y(t-2)$	1.6759×10^{-4}	2.255×10^{-3}	-0.7962
$u(t-1)$	0.47871	4.7434×10^{-2}	0.0339
$u(t-2)$	6.8123×10^{-5}	1.8925×10^{-4}	0.0338
$y(t-1)u(t-1)$	2.2653×10^{-5}	3.6489×10^{-5}	-0.1597
$y(t-1)u(t-2)$	6.1439×10^{-5}	$1.9004e \times 10^{-5}$	-0.1396
$y(t-2)y(t-2)$	3.1515×10^{-30}	$5.3837e \times 10^{-7}$	0.0085
$y(t-2)u(t-1)$	3.7241×10^{-7}	$2.9966e \times 10^{-5}$	0.1297
$y(t-2)u(t-2)$	4.6109×10^{-5}	$2.8901e \times 10^{-5}$	0.1086

Table 2: The model identified when Equation 1 is excited with input (b), a multi-tone sinusoid

Model term	Mean squared PRESS error	ERR	Parameters/Coefficients
$y(t-1)$	1.2209×10^{-4}	0.1035	1.7813
$y(t-2)$	7.0858×10^{-7}	1.7841×10^{-4}	-0.7962
$u(t-1)$	2.8085×10^{-9}	2.5768×10^{-9}	0.0339
$u(t-2)$	3.7183×10^{-8}	3.5856×10^{-7}	0.0338
$y(t-1)u(t-1)$	4.5778×10^{-12}	2.7792×10^{-9}	-0.1597
$y(t-1)u(t-2)$	2.9234×10^{-7}	6.0493×10^{-7}	-0.1396
$y(t-2)y(t-2)$	3.8123×10^{-9}	4.6086×10^{-8}	0.0085
$y(t-2)u(t-1)$	1.9182×10^{-25}	6.4198×10^{-12}	0.1297
$y(t-2)u(t-2)$	7.0559×10^{-2}	0.89632	0.1086

The correlation-based statistical validation tests for nonlinear models (Billings & Voon, 1983), Figure 3 and Figure 4, indicate model bias. However, it is minor, the residual variances, 1.6018×10^{-25} for (a) and 8.2178×10^{-18} for (b), are negligible compared to the output variances of 0.069 and 0.0581, respectively.

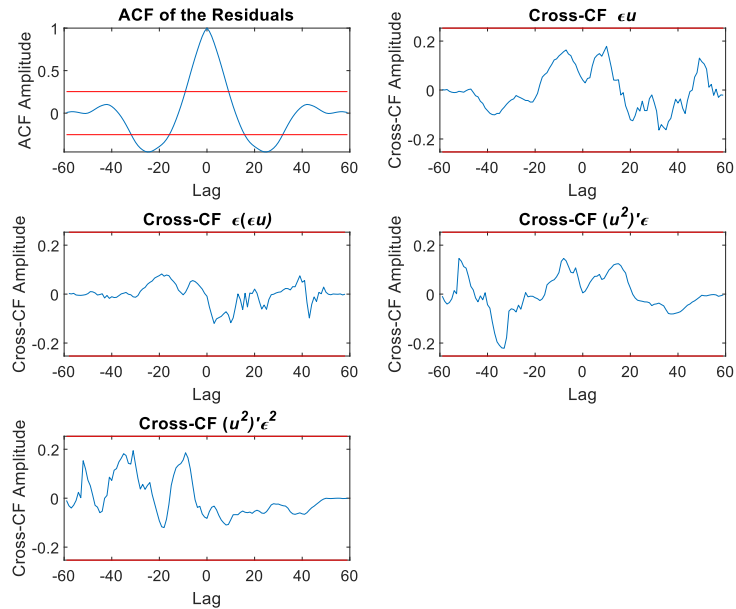


Figure 3: Model validation results for input (a). The red bounds indicate the tolerances the correlation function should stay within for the identified model to be unbiased.

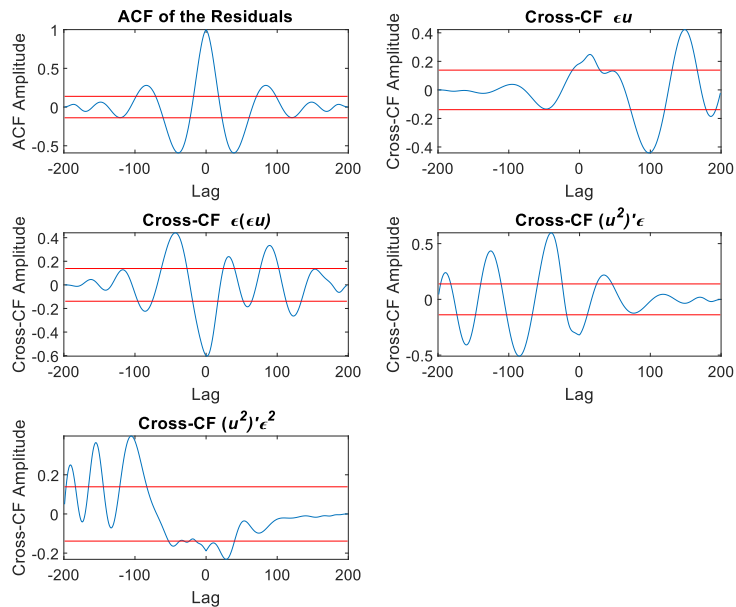


Figure 4: Model validation results for input (b). The red bounds indicate the tolerances the correlation function should stay within for the identified model to be unbiased.

Real data example

The real data in this example is obtained from an electromechanical system described in (Lacerda Junior et al., 2017). The system comprises two DC motors, one as a driver and the

other as a generator, mechanically coupled by a shaft. The input is the voltage applied to the driver motor. This input is a pseudo-random binary signal (PRBS) designed to excite the system over a range of dynamics. The system output is the rotational speed (angular velocity) of the generator motor.

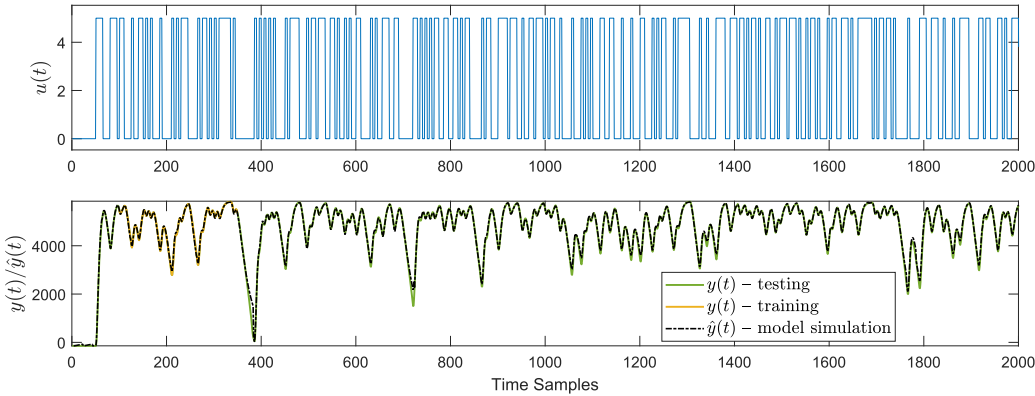


Figure 5: Model identification results from the electro-mechanical system. The model simulation output $\hat{y}(t)$ is presented against the actual output $y(t)$ of the system given in Equation 1. The input $u(t)$ is a PRBS. Only 250 samples are used for identifying/training the model using iOFR_s in the ‘NonSysId’ package.

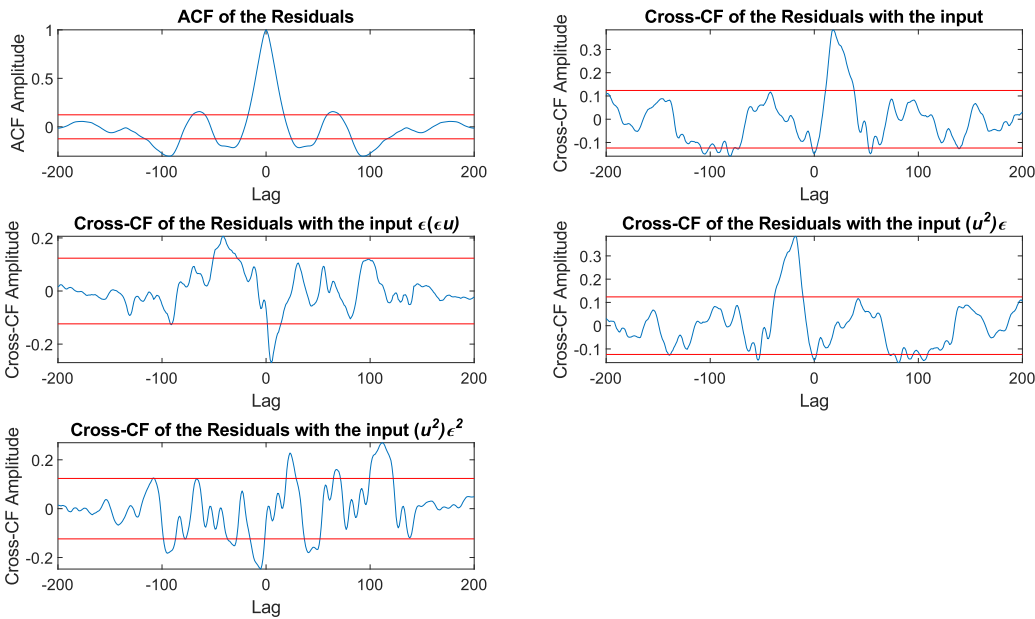


Figure 6: Model validation results for the system in (Lacerda Junior et al., 2017). The red bounds indicate the tolerances the correlation function should stay within for the identified model to be unbiased.

Table 3: The model identified from the data generated from the system in (Lacerda Junior et al., 2017)

Model term	Mean squared PRESS error	ERR	Parameters/Coefficients
$y(t-1)$	8128.5	0.49526	1.7844
$y(t-2)$	975.85	0.00028497	-0.79156
$u(t-1)$	318.88	2.6363×10^{-5}	47.205

Model term	Mean squared PRESS error	ERR	Parameters/Coefficients
$y(t-2)u(t-1)$	158.23	6.211×10^{-6}	-0.037612
$y(t-3)u(t-1)$	1.2306×10^7	0.50441	0.030086
$u(t-2)u(t-2)$	91.271	2.5147×10^{-6}	1.89
$u(t-2)u(t-3)$	71.842	7.2261×10^{-7}	-0.91694

Future Work

The NonSysId package currently supports the identification of single-input single-output (SISO) and multi-input single-output (MISO) models, with correlation-based residual analysis being limited to SISO models. Future updates will extend support to multi-input multi-output (MIMO) identification and enable validation for MISO and MIMO systems. While the package now handles polynomial NARX models, upcoming versions will expand iOFR_S to broader basis functions. A Python open-source release is also planned.

Acknowledgements

RG and FH were supported by EPSRC grant [EP/X020193/1].

References

- Aguirre, L. A., Barbosa, B. H. G., & Braga, A. P. (2010). Prediction and simulation errors in parameter estimation for nonlinear systems. *Mechanical Systems and Signal Processing*, 24(8), 2855–2867. <https://doi.org/10.1016/j.ymssp.2010.05.003>
- Aguirre, L. A., & Billings, S. A. (1995). Dynamical effects of overparametrization in nonlinear models. *Physica D: Nonlinear Phenomena*, 80(1), 26–40. [https://doi.org/10.1016/0167-2789\(95\)90053-5](https://doi.org/10.1016/0167-2789(95)90053-5)
- Allen, D. M. (1974). The relationship between variable selection and data agumentation and a method for prediction. *Technometrics*, 16(1), 125–127. <https://doi.org/10.1080/00401706.1974.10489157>
- Ayala, H. V. H., Gritti, M. C., & dos Santos Coelho, L. (2020). An R library for nonlinear black-box system identification. *SoftwareX*, 11, 100495. <https://doi.org/10.1016/j.softx.2020.100495>
- Billings, S. A., Korenberg, M. J., & Chen, S. (1987). *Identification of non-linear output-affine systems using an orthogonal least-squares algorithm*. <https://doi.org/10.1080/00207728808964057>
- Billings, S. A., & Voon, W. S. F. (1983). Structure detection and model validity tests in the identification of nonlinear systems. *IEE Proceedings D (Control Theory and Applications)*, 130, 193–199. <https://doi.org/10.1049/ip-d.1983.0034>
- Chen, S., Billings, S. A., & Luo, W. (1989). Orthogonal least squares methods and their application to non-linear system identification. *Int. J. Control*, 50(5), 1873–1896. <https://doi.org/10.1080/00207178908953472>
- Chen, T., Cai, W., Ryali, S., Supekar, K., & Menon, V. (2016). Distinct global brain dynamics and spatiotemporal organization of the salience network. *PLOS Biology*, 14(6), 1–21. <https://doi.org/10.1371/journal.pbio.1002469>
- Chiras, N., Evans, C., & Rees, D. (2002). Global Nonlinear Modeling of Gas Turbine Dynamics

- Using NARMAX Structures . *Journal of Engineering for Gas Turbines and Power*, 124(4), 817–826. <https://doi.org/10.1115/1.1470483>
- Eichenbaum, A., Pappas, I., Lurie, D., Cohen, J. R., & D'Esposito, M. (2021). Differential contributions of static and time-varying functional connectivity to human behavior. *Network Neuroscience*, 5(1), 145–165. https://doi.org/10.1162/netn_a_00172
- Gao, Y., Yu, C., Zhu, Y.-P., & Luo, Z. (2023). A NARX model-based condition monitoring method for rotor systems. *Sensors*, 23(15). <https://doi.org/10.3390/s23156878>
- Gharehbaghi, V. R., Noroozinejad Farsangi, E., Noori, M., Yang, T. Y., Li, S., Nguyen, A., Málaga-Chuquitaype, C., Gardoni, P., & Mirjalili, S. (2022). A critical review on structural health monitoring: Definitions, methods, and perspectives. *Archives of Computational Methods in Engineering*, 29(4), 2209–2235. <https://doi.org/10.1007/s11831-021-09665-9>
- Guo, Y., Guo, L. Z., Billings, S. A., & Wei, H.-L. (2015). An iterative orthogonal forward regression algorithm. *International Journal of Systems Science*, 46(5), 776–789. <https://doi.org/10.1080/00207721.2014.981237>
- He, F., Sarigiannis, P. G., Billings, S. A., Wei, H., Rowe, J., Romanowski, C., Hoggard, N., Hadjivassiliou, M., Rao, D. G., Grünewald, R., Khan, A., & Yianni, J. (2016). Nonlinear interactions in the thalamocortical loop in essential tremor: A model-based frequency domain analysis. *Neuroscience*, 324, 377–389. <https://doi.org/10.1016/j.neuroscience.2016.03.028>
- He, F., & Yang, Y. (2021). Nonlinear system identification of neural systems from neurophysiological signals. *Neuroscience*, 458, 213–228. <https://doi.org/10.1016/j.neuroscience.2020.12.001>
- Hong, X., Sharkey, P. M., & Warwick, K. (2003). Automatic nonlinear predictive model-construction algorithm using forward regression and the PRESS statistic. *IEEE Proceedings-Control Theory and Applications*, 150(3), 245–254. <https://doi.org/10.1049/ip-cta:20030311>
- Junior, W. R. L., Andrade, L. P. C. da, Oliveira, S. C. P., & Martins, S. A. M. (2020). SysIdentPy: A python package for system identification using NARMAX models. *Journal of Open Source Software*, 5(54), 2384. <https://doi.org/10.21105/joss.02384>
- Korenberg, M., Billings, S. A., Liu, Y. P., & McIlroy, P. J. (1988). Orthogonal parameter estimation algorithm for non-linear stochastic systems. *Int. J. Control*, 48(1), 193–210. <https://doi.org/10.1080/00207178808906169>
- Kunjan, S., Grummett, T. S., Pope, K. J., Powers, D. M. W., Fitzgibbon, S. P., Bastiampillai, T., Battersby, M., & Lewis, T. W. (2021). The necessity of leave one subject out (LOSO) cross validation for EEG disease diagnosis. In M. Mahmud, M. S. Kaiser, S. Vassanelli, Q. Dai, & N. Zhong (Eds.), *Brain informatics* (pp. 558–567). Springer International Publishing. https://doi.org/10.1007/978-3-030-86993-9_50
- Lacerda Junior, W. R., Almeida, V. M., & Milani, S. A. (2017). Identificação de um motor/gerador cc por meio de modelos polinomiais autorregressivos e redes neurais artificiais. *Proc. XIII Simpósio Brasileiro de Automação Inteligente*, 1–6.
- Lacerda Junior, W. R., Martins, S. A. M., Nepomuceno, E. G., & Lacerda, M. J. (2019). Control of hysteretic systems through an analytical inverse compensation based on a NARX model. *IEEE Access*, 7, 98228–98237. <https://doi.org/10.1109/ACCESS.2019.2926057>
- Lehnertz, K., Rings, T., & Bröhl, T. (2021). Time in brain: How biological rhythms impact on EEG signals and on EEG-derived brain networks. *Frontiers in Network Physiology*, 1. <https://doi.org/10.3389/fnetp.2021.755016>
- Little, M. A., Varoquaux, G., Saeb, S., Lonini, L., Jayaraman, A., Mohr, D. C., & Kording, K. P. (2017). Using and understanding cross-validation strategies. Perspectives on Saeb et al. *GigaScience*, 6(5), gix020. <https://doi.org/10.1093/gigascience/gix020>

- Liu, Z., Lang, Z.-Q., Gui, Y., Zhu, Y.-P., & Laalej, H. (2024). Digital twin-based anomaly detection for real-time tool condition monitoring in machining. *Journal of Manufacturing Systems*, 75, 163–173. <https://doi.org/10.1016/j.jmsy.2024.06.004>
- Mao, K. Z., & Billings, S. A. (1997). Algorithms for minimal model structure detection in nonlinear dynamic system identification. *International Journal of Control*, 68(2), 311–330. <https://doi.org/10.1080/002071797223631>
- Mendes, E. M. A. M., & Billings, S. A. (1998). On overparametrization of nonlinear discrete systems. *International Journal of Bifurcation and Chaos*, 08(03), 535–556. <https://doi.org/10.1142/S0218127498000346>
- Piroddi, L., & Spinelli, W. (2003). An identification algorithm for polynomial NARX models based on simulation error minimization. *International Journal of Control*, 76(17), 1767–1781. <https://doi.org/10.1080/00207170310001635419>
- Ritzberger, D., & Jakubek, S. (2017). Nonlinear data-driven identification of polymer electrolyte membrane fuel cells for diagnostic purposes: A volterra series approach. *Journal of Power Sources*, 361, 144–152. <https://doi.org/10.1016/j.jpowsour.2017.06.068>
- Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2), 461–464. <https://doi.org/10.1214/aos/1176344136>
- Seedat, Z. A., Pier, K. St., Holmes, N., Rea, M., Al-Hilaly, L., Tierney, T. M., Embury, C. M., Pardington, R., Mullinger, K. J., Cross, J. H., Boto, E., & Brookes, M. J. (2024). Simultaneous whole-head electrophysiological recordings using EEG and OPM-MEG. *Imaging Neuroscience*, 2, 1–15. https://doi.org/10.1162/imag_a_00179
- Stoica, P., & Selen, Y. (2004). Model-order selection: A review of information criterion rules. *IEEE Signal Processing Magazine*, 21(4), 36–47. <https://doi.org/10.1109/MSP.2004.1311138>
- Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 36(2), 111–147. <https://doi.org/10.1111/j.2517-6161.1974.tb00994.x>
- Sum, J. P. F., Kan, W.-K., & Young, G. H. (1999). A note on the equivalence of NARX and RNN. *Neural Computing & Applications*, 8(1), 33–39. <https://doi.org/10.1007/s005210050005>
- Vamsikrishna, A., & Gijo, E. V. (2024). New techniques to perform cross-validation for time series models. *Operations Research Forum*, 5(2), 51. <https://doi.org/10.1007/s43069-024-00334-8>
- Wang, K., Zhang, J., Croxford, A. J., & Yang, Y. (2024). Nonlinear autoregressive exogenous method for structural health monitoring using ultrasonic guided waves. In F.-G. Yuan (Ed.), *Structural health monitoring/ management (SHM) in aerospace structures* (pp. 427–452). Woodhead Publishing. <https://doi.org/10.1016/B978-0-443-15476-8.00004-6>
- Wang, L., & Cluett, W. R. (1996). Use of PRESS residuals in dynamic system identification. *Automatica*, 32(5), 781–784. [https://doi.org/10.1016/0005-1098\(96\)00003-9](https://doi.org/10.1016/0005-1098(96)00003-9)
- Wei, H.-L., Billings, S. A., & Liu, J. (2004). Term and variable selection for non-linear system identification. *International Journal of Control*, 77(1), 86–110. <https://doi.org/10.1080/00207170310001639640>
- Zainol, N. M., Damanhuri, N. S., Othman, N. A., Chiew, Y. S., Nor, M. B. M., Muhammad, Z., & Chase, J. G. (2022). Estimating the incidence of spontaneous breathing effort of mechanically ventilated patients using a non-linear auto regressive (NARX) model. *Computer Methods and Programs in Biomedicine*, 220, 106835. <https://doi.org/10.1016/j.cmpb.2022.106835>