Proceedings Paper:

Gu, Y., Wei, H., Balikhin, M. et al. (2019) Machine learning enhanced NARMAX model for Dst index forecasting. In: Proceedings of the 25th International Conference on Automation and Computing (ICAC). 25th International Conference on Automation and Computing (ICAC' 19), 05-07 Sep 2019, Lancaster, UK. IEEE. ISBN: 9781728125183.

https://doi.org/10.23919/IConAC.2019.8895027

Proceedings of the 25th International Conference on Automation & Computing, Lancaster University, Lancaster UK, 5-7 September 2019

Final accepted manuscript

Machine Learning Enhanced NARMAX Model for Dst Index Forecasting

Yuanlin Gu, Hua-Liang Wei*, Michael A. Balikhin, Richard J. Boynton, Simon N. Walker
Department of Automatic Control and Systems Engineering
University of Sheffield, Sheffield
(*corresponding author: w.hualiang@sheffield.ac.uk)

***Abstract*—As many systems and equipment are sensitive to magnetic disturbances, it is important to understand the magnetosphere system, to reduce the negative effect caused by severe space weather situations. The disturbance storm time (Dst) index is used to measure the magnetic disturbances and it is correlated with solar wind variables. This study presents a new machine learning enhanced NARMAX (MLE-NARMAX) model for 3 hours ahead forecasting of Dst index. An important advantage of the MLE-NARMAX model is that it provides a transparent and explainable model structure. The model performance is tested over three typical strong storm periods, where the prediction skills are 0.9734, 0.9598 and 0.9206 in terms of correlation, and 0.9474, 0.9173, and 0.8333 in terms prediction efficiency (PE). Compared to the conventional NARX model, the MLE-NARMAX produces better model predictions.**

***Keywords*-Space Weather; Dst Forecast; NARMAX model; Machine Learning;**

I. Introduction

The magnetosphere is a very complex system. To understand the magnetosphere system, the Dst index was developed to measure the magnetic disturbances and it is known to be correlated with a number of solar wind variables [1]-[3][4]. There are plenty of studies aiming to forecast Dst index from solar wind measurements using different methods including machine learning techniques [5]. In [6],[7], Elman recurrent neural networks were first proposed for Dst index prediction. Since then, many other neural network models have been introduced for Dst index prediction [8-13]. The NARMAX method has also been applied to Dst index forecasting [14-16]. Other methods, for example, wavelets models were also used to forecast Dst index [1-2]. A comparison study of the Dst index forecast models suggests that the neural network by Temerin and Li produces the best predictions when all the events are considered [17]. It is not surprising as the advantage of neural networks is that they can describe the behaviours of complex systems and can usually achieve excellent model prediction performance. The performance of neural networks could become even more powerful when increasing the number of the hidden layers. However, some common issues can arise with neural networks, for example, the interpretability of the model and the potential overfitting problem. Another issue is that complicated neural networks usually require a huge amount of computational time.

Compared to neural networks, the nonlinear autoregressive moving average with exogenous (NARMAX) model provides a much simple representation of nonlinear systems [18-19]. It employs an orthogonal forward regression (OFR) algorithm to measure and rank the significance of each candidate model terms, so that the most significant model terms can be selected accordingly [18-21]. More importantly, the NARMAX model provides a transparent and parsimonious model structure, which is very useful for understanding and interpreting the system behaviour. In general, the neural network model is better at generating model prediction and the NARMAX model is better at providing a transparent and explainable model structure. The NARMAX method and its variants [22] has been successfully applied to various research fields including ecological [23], environmental [24], geophysical [2,14,25,26], medical [27], societal [28] and neurophysiological sciences [29].

Based on the above considerations, this study presents a new modelling framework, which consists of two sub-models, namely, the NARX model component and the neural network model component. The NARX sub-model is established to capture and represent the most important system dynamics in a transparent way, while the neural network sub-model is established to accommodate the error details that are not captured by the NARX model. In this way, both the advantages of the NARMAX model (e.g. transparent, interpretable, simple) and the neural networks (e.g. general strong learning ability) can be well exploited and combined.

The proposed framework is therefore referred to as machine learning enhanced NARMAX (MLE-NARMAX) model. In this study, the MLE-NARMAX model is used to generate 3 hours ahead predictions for Dst index, on three typical test periods of strong storms. The results are compared with those produced by the conventional NARX and neural networks. The main features of the MLE-NARMAX model are: 1) the resulting models are transparent and easy to interpret, and 2) the model possesses good prediction performance.

The paper is organized as follows. The proposed MLE-NARMAX model is introduced in section II. A brief description of the dataset is presented in section III. A case study of modelling and forecasting of the Dst index is presented in section IV, and finally the work is concluded in section V.

II. THE PROPOSED MLE-NARMAX METHOD

In this section, the NARMAX method is briefly reviewed and the proposed MLE-NARMAX model is introduced.

A. A brief review of NARMAX method

Consider the case of multiple-input single-output (MISO) systems, for which a NARMAX model can be written as:

$$y(t) = f[y(t-\mathrm{d}), \dots, y(t-n_y), x_1(t-d) \dots, x_1(t-n_u), x_2(t), \dots, x_2(t-n_u) \ \dots, x_r(t-d), \dots, x_r(t-n_u), e(t-\mathrm{d}), \dots e(t-n_e)] + e(t) \quad (1)$$

where $y(t), x_1(t), \dots, x_r(t), e(t)$ are the system output, input and noise signals, respectively; n_y, n_u and n_e are the maximum time lags for the system output, input and noise; d is a time delay, and in this study we assume that the time delay d =3, meaning that the model are used for 3 hour ahead forecasting; $f[\cdot]$ is some nonlinear function. Without including the noise moving-averaging model elements $e(t-\mathrm{d}), \dots, e(t-n_e)$, the NARMAX model is reduced to the NARX model, which can be written in a linear-in-the-parameters form as [19]:

$$y(t) = \sum_{m=1}^{M} \theta_m \varphi_m(t) + e(t) \quad (2)$$

where $\varphi_1(t) \dots \varphi_r(t)$ are the model terms generated from the regressor vector $[y(t-\mathrm{d}), \dots, y(t-n_y), x_1(t-d) \dots, x_1(t-n_u), x_2(t-d), \dots, x_2(t-n_u) \ \dots, x_r(t-d), \dots, x_r(t-n_u)]^T$, θ_m are the unknown parameters and M is the number of candidate model terms. The identification of NARMAX model consists of two main steps. The first step is to establish a NARX model using OFR algorithm [19]. However, the polynomial NARX structure might not perfectly describe the system behaviors, but it can usually provide a good approximation if both the NARX model structure and model parameters are well estimated. To reduce bias of the NARX model, NARMAX method introduces a second step which is a noise modelling process [18], where the noise signal (model residual) is treated as an extra "input" (explanatory variable), referred to as the moving average (MA) elements. In each search step, a candidate NARX model is established first, based on which the model error (residual) is calculated and used to estimate an associated candidate NARMAX model. The procedure repeats many times until a NARMAX model with good performance is established.

B. The MLE-NARMAX model

In this study, the MLE-NARMAX framework, combining NARX and neural network approaches, which can effectively deal with correlated or colored noise, is proposed. It is a two-stage training method. In the first stage, a NARX model is established to represent the main dynamics of the system of interest. Note that the NARX might not be sufficient for representing the system dynamics in every detail. In the second stage, a neural network is designed and trained to accommodate the model residual of NARX model and the potential correlations (interactions) between model residuals (of the first stage), system inputs and outputs variables. So, the final MLE-NARMAX model consists of two sub-models, namely, the NARX sub-model and neural network sub-model.

(i). First-stage NARX sub-model

Using the OFR algorithm (the details of the implementation can be found in [19-21]), the first-stage NARX sub-model can be established as:

$$y_{NARX}(t) = \theta_{l_1}\varphi_{l_1}(t) + \dots + \theta_{l_n}\varphi_{l_n}(t) + e(t) \quad (3)$$

where $\varphi_{l_1}(t), \varphi_{l_2}(t), \dots, \varphi_{l_n}(t)$ are the selected model terms and $\theta_{l_1}, \theta_{l_2}, \dots, \theta_{l_n}$ are the estimated parameters. Note that the NARX sub-model is a linear-in-the-parameters representation, where individual model terms are fully transparent, and their contributions are measurable and interpretable. The significant model terms are selected from a pre-specified dictionary and then ranked based on the values of the ERR index. While in most situations, NARX model can provide a good representation of the underlying system dynamics of interest, NARX model might not sufficiently capture all the details of the system. This motivates the use of a neural network sub-model in the second stage to improve the prediction performance.

(ii). Second-stage neural network sub-model

In the second stage, a neural network is used to approximate the model residual of the NARX model. Note that the output (desired signal) of the neural network is the model error, while the inputs of the neural network include not only the original input variables of the NARX model, but also the lagged versions of the model residual variable. The motivation of introducing a neural network model to approximate the model error is to take advantage of neural network approximation capability to accommodate the dependent and correlated relations between the model error and all the candidate explanatory variables that are sufficiently explained by the NARX model.

To avoid any confusion, in this study we use e(t) and $\varepsilon(t)$ to represent noise (of a general sense) and model error (residual). The model error of the NARX model (3) is:

$$\varepsilon(t) = y(t) - \hat{y}_{NARX}(t) \quad (4)$$

The signal ε (t) is used as the desired output signal to train the neural network sub-model of the form:

$$\varepsilon(t) = g[\omega_1(t), \omega_2(t), \dots, \omega_{M'}(t)] \quad (5)$$

where $g[\cdot]$ represents the constructed neural network sub-model, and the input vectors $\omega_k(t)$, with $k = 1,2, \dots, M'$, are defined as $y(t-\mathrm{d}), \dots, y(t-n_y), x_1(t-d) \dots, x_1(t-n_u), x_2(t-d), \dots, x_2(t-n_u) \ \dots, x_r(t-d), \dots, x_r(t-n_u), \varepsilon(t-\mathrm{d})), \dots, \varepsilon(t-n_z)$, where n_ρ is the time lag for the error signal. Note that the neural network sub-model uses all the model terms $\omega_1(t), \omega_2(t), \dots, \omega_{M'}(t)$, meaning that the model structure can be extremely complicated and the modelling process can therefore time-consuming.

The general structure of the MLE-NARMAX model is presented in Fig. 1, where the MLE-NARMAX model can be explicitly expressed as:

$$y(t) = \theta_{l_1}\varphi_{l_1}(t) + \theta_{l_2}\varphi_{l_2}(t) + \cdots + \theta_{l_n}\varphi_{l_n}(t) + g[\omega_1(t), \omega_2(t), \ldots, \omega_M{}'(t)] \quad (6)$$

where the $\theta_{l_1}\varphi_{l_1}(t) + \theta_{l_2}\varphi_{l_2}(t) + \cdots + \theta_{l_n}\varphi_{l_n}(t)$ is the NARX sub-model and $g[\omega_1(t), \omega_2(t), \ldots, \omega_{M'}(t)]$ is the neural network sub-model. The model prediction of the MLE-NARMAX model can be calculated as:

$$\hat{y}_{MLE-NARMAX}(t) = \hat{y}_{NARX}(t) + \widehat{\varepsilon_{NN}}(t) \quad (7)$$

where $\hat{y}_{NARX}(t)$ is the model prediction of NARX sub-model and $\widehat{\varepsilon_{NN}}(t)$ is the model prediction of neural network sub-model.

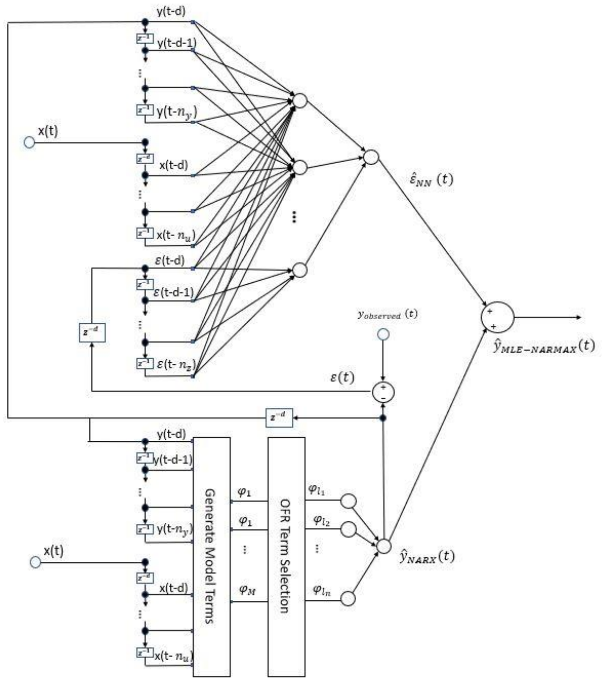

Figure 1. The MLE-NARMAX model structure.

III. Data

The process of Dst is treated to be a dynamic nonlinear system, where the system inputs are solar wind variables and the system output is the Dst index. The description of the solar wind variables and Dst index is given in Table I. All the variables were sampled every 1 hour. It should be noted that $VBst = V \times B_T \sin^6\left(\frac{\theta}{2}\right)/1000$ is a multiplied input which was suggested to be included in the model inputs [30].

TABLE I. Dst Index and Solar Wind Variables

Name	Description
Dst	Dst index [nT]
V	solar wind speed/velocity (flow speed) [km/s]
p	solar wind pressure (flow pressure) [nPa]
n	solar wind density (proton density) [cm^{-3}]
B	interplanetary magnetic field (IMF) [nT]
Bz	the north-south IMF [nT]
Bst	$Bst = B_T \sin^6(\theta/2)$ [nT] [8]

The Dst and solar wind data of year 2014 were used for training the model. Three time periods of intense storms (Dst<-100nT), Mar 2015, Jun 2015 and Sep 2017 were used to evaluate the model. The time series of the Dst index and solar wind variables of the three interested periods are shown in Fig. 2. For Dst index forecast, negative peak values are important. From the figure, there are strong storms in these periods. In total, there are around 8700 data points in the training dataset and around 2200 data points in the three test datasets.

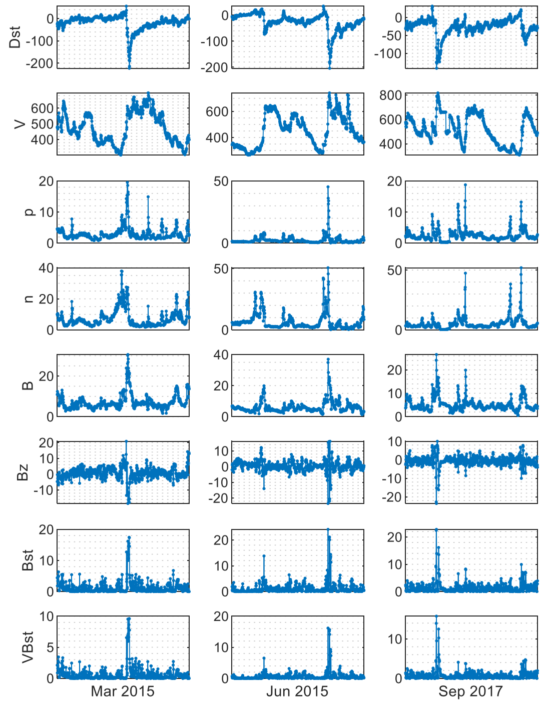

Figure 2. Observations of sampled Dst index and solar wind variables of the three test periods.

IV. The MLE-NARMAX Model for 3 Hour Ahead Dst Index Forecasting

This section presents the identified MLE-NARMAX model that predict Dst index 3 hours ahead. The model was evaluated on three test periods.

A. Forecast Dst index 3 hours ahead

The 3 hours ahead prediction of Dst can be defined as:

$$Dst(t) = F_{MLE-NARMAX}[Dst(t-3) \ldots Dst(t-n_y), V(t-3) \ldots V(t-n_u), p(t-3) \ldots p(t-n_u), n(t-3) \ldots n(t-n_u), B(t-3) \ldots B(t-n_u), Bz(t-3) \ldots Bz(t-n_u), Bst(t-3) \ldots Bst(t-n_u), VBst(t-3) \ldots VBst(t-n_u)] \quad (8)$$

where $F_{MLE-NARMAX}$ is the MLE-NARMAX framework. To evaluate the prediction of the model, the correlation coefficient, prediction efficiency (PE), and normalized root-mean square error (NRMSE) are calculated. The value of PE is defined as

$$PE = 1 - \frac{\sigma^2_{error}}{\sigma^2_{observed}} \quad (9)$$

where $\sigma^2_{observed}$ is the variance of the observed Dst values and σ^2_{error} is the variance of the error between the predicted and observed Dst values.

TABLE II. SELECTED MODEL TERMS OF THE NARX SUB-MODEL

No	Model Term	ERR (100%)	Parameter	t-statistics
1	Dst(t-03)	78.1229	0.8462	103.3129
2	B(t-04) *VBst(t-03)	3.3731	-0.1680	4.6679
3	B(t-04) *VBst(t-04)	0.4205	0.1528	6.1903
4	p(t-03) *p(t-04)	0.3519	-0.2623	16.1041
5	Bz(t-03) *Bst(t-03)	0.2090	0.2002	13.3520
6	p(t-04) *n(t-03)	0.1400	0.0728	11.8165
7	n(t-04) *Dst(t-03)	0.1077	-0.0064	7.0959
8	Bst(t-03)	0.0645	-0.6475	6.9029
9	Bz(t-03) *Bz(t-03)	0.0590	0.0346	8.5650
10	V(t-04) *Bz(t-04)	0.0844	-0.0006	6.5766

TABLE III. COMPARISON OF THE PERFORMNACES OF NARX MODEL, NEURAL NETWORK AND MLE-NARMAX MODEL OF THE THREE TEST PERIODS

Period	Model Type	Correlation Coefficient	Prediction Efficiency	NRMSE
Mar 2015	NARX	0.9502	0.9029	0.0353
	Neural Network*	0.9716	0.9439	0.0269
	MLE-NARMAX*	0.9734	0.9474	0.0260
Jun 2015	NARX	0.8907	0.7368	0.0678
	Neural Network*	0.9599	0.9212	0.0364
	MLE-NARMAX*	0.9598	0.9173	0.0368
Sep 2017	NARX	0.8828	0.7735	0.0642
	Neural Network*	0.9295	0.8487	0.0500
	MLE-NARMAX*	0.9206	0.8333	0.0529

* The algorithm was run for ten times and the averaged statistics are recorded

B. The identified MLE-NARMAX model

In order to determine the maximum time lags for both the input and output variables, following the approach described in [21] we have carried out pre-modelling experiments and simulations [21], the results suggest that the maximum time lags of the input and output were chosen to be $nu = 4$ and $ny = 4$. The initial full model was chosen to be a polynomial form with nonlinear degree of 2.

In the first step, a 10-term bilinear NARX sub-model was identified. The 10 model terms, together with their corresponding ERR values and t-statistics, are shown in Table II. The t-statistics given in the table indicates that all the selected model terms are significant. Note the first-stage bilinear NARX model reported in Table II can be written as:

$$Dst(t) = 0.8462\, Dst(t-3) - 0.1680\, B(t-4)\, VBst(t-3) + \cdots \quad (10)$$

In the second step, the neural network sub-model was estimated to fit the error of NARX sub-model. According to the trial experiments and simulations, the number of neurons is chosen to be 10. The estimation algorithm was run for 10 times and the averaged performances were recorded. Then, the final MLE-NARMAX model is obtained with the NARX sub-model and the neural network sub-model. As the neural network contains too many nodes and connections, the details of the model are not presented here.

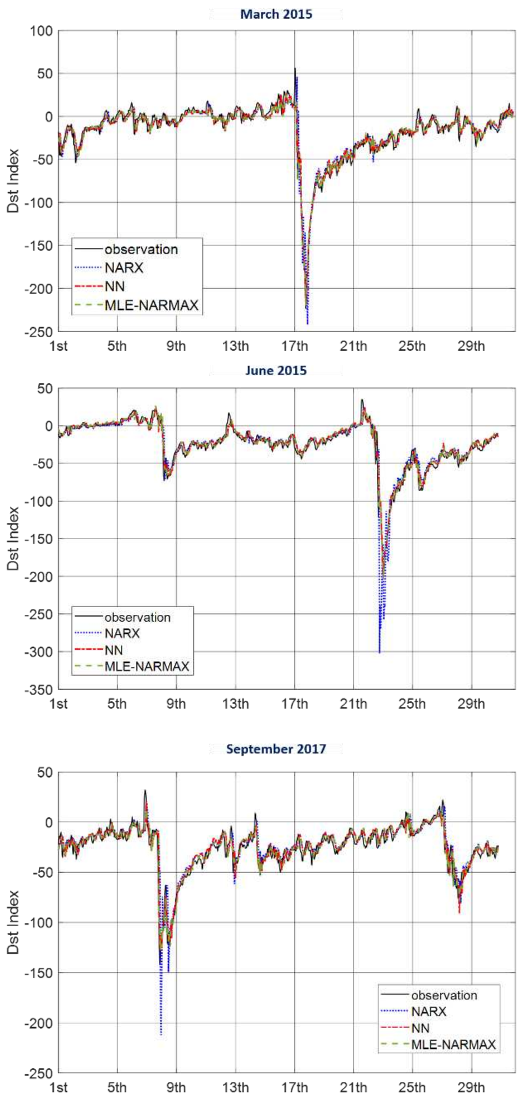

Figure 3. Comparison of the predictions of the NARX model, neural network model and MLE-NARMAX model of the three test datasets.

C. Performance and advantage of the MLE-NARMAX model

The MLE-NARMAX model was used to generate 3 hours ahead Dst predictions for the three test periods: Mar 2015, Jun 2015 and Sep 2017. Fig. 3 presents graphical comparisons of the observed and predicted Dst index of the three test periods. The statistical performances of the NARX model, neural network and the MLE-NARMAX model on the three test periods are presented in Table III. From the statistics, the performances of the MLE-NARMAX model are similar to those of the neural networks and better than those of the NARX models for all the three test periods. It can be seen that for the test period of June 2015, the improvement is obviously more significant than those for the other two periods. From Table III and Fig. 3, it can be noticed that while the bilinear NARX structure can well capture the features and dynamics of the Dst process at most times of the test periods of June 2015 and September 2017, the model does not sufficiently capture the system dynamics at the severe situation times. The neural network sub-model, however, can help improve the model performance.

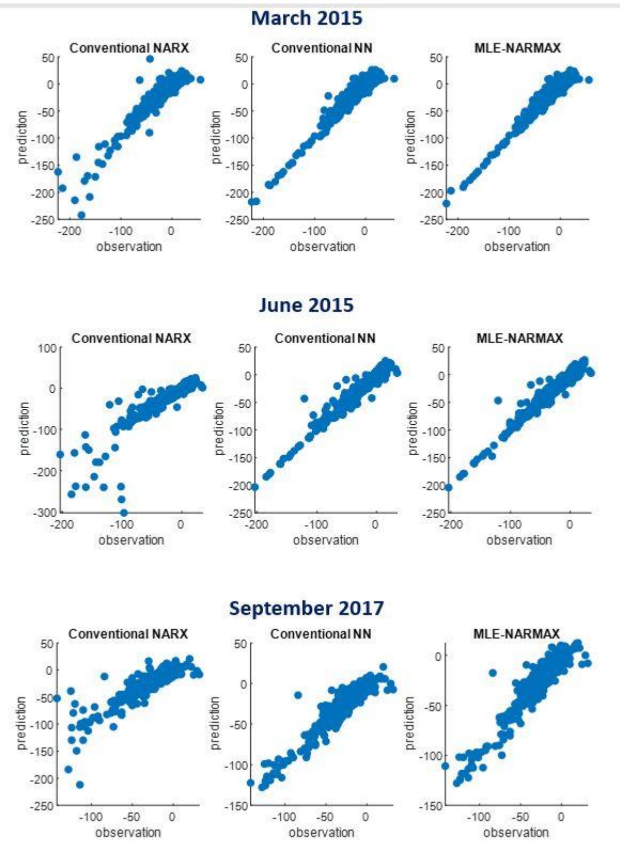

Figure 4. Scatter plots of the 3 hours ahead NARX model, neural network model and MLE-NARMAX model of the three test datasets.

Fig. 4 shows the scatter plots of the NARX model, neural network and MLE-NARMAX model. From these plots, it can be seen that the MLE-NARMAX model produces better predictions for strong storms (Dst< -100nT) than the bilinear NARX model alone. In other words, the results show that the combination of the NARX sub-model and neural network sub-model can better predict change of the Dst index during strong storm periods.

From the results shown in Table II, the NARX sub-models only consists of 10 significant terms. Obviously, the model provides a parsimonious and transparent representation, where contributions of the selected model terms are clear. However, such a simple bilinear NARX model may not always be sufficient to capture the underlying dynamics of the process, and the model prediction performance may be improved by introducing a sub-model to characterize some dynamics of the system hidden in the model residuals that is not captured by the sub-NARX model. This explains why the two-stage MLE-NARMAX performs better than the NARX model.

Note that although the neural network sub-model improves the model performance, the model structure itself cannot be written down. Fig. 5 shows the training state of one of the neural network sub-models. The two variables 'gradient' and 'gamk' indicate the values of the associated gradient and the effective number of parameters at each iteration, respectively. The figure shows that during the training process of the MLE-NARMAX models, the neural network sub-model contains over 150 parameters. The model complexity of the neural network sub-model is much higher than that of the NARX sub-model. In addition, the neural network sub-model takes many steps to train. For 'big' data modeling problems where the data size is much larger, the training of neural network can take quite long time. On the contrary, the training of the NARX sub-model only takes a few steps and use relatively much less time. Therefore, the MLE-NARMAX model is developed so that it can take the advantages of both NARX model and neural network model. For example, it provides a transparent representation of complex nonlinear systems, which helps to understand the systems behaviors. Meanwhile, it provides good model prediction performance.

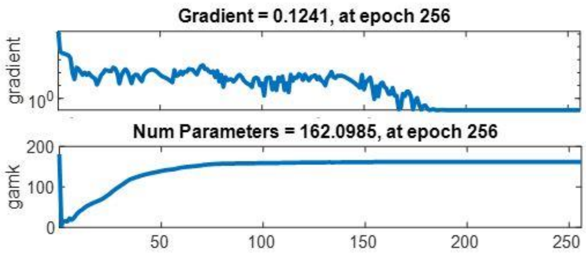

Figure 5. Training state of the neural network sub-model of the MLE-NARMAX model.

V. CONCLUSION

In this paper, a new MLE-NARMAX method is proposed for 3 hours ahead prediction of the Dst index. Three periods with typical strong storms were used to test the model performance. The MLE-NARMAX model outperforms the conventional NARX model in terms of correlation coefficient and prediction efficiency. More importantly, the MLE-NARMAX model is capable to provide an interpretable representation of the system, which can reveal the most significant model terms and in the meantime show good generalization properties. For many real data modelling problems, where the central modelling task and objective is not is not only for prediction but also for understanding and explaining the input-output behaviour or cause-effect relationships of the systems, the proposed MLE-NARMAX model is a good choice.

For future work, we intend to further develop the neural network sub-model by employing deep learning methods [31][32], to seek further improvement of the prediction performance of the MLE-NARMAX model.

ACKNOWLEDGMENT

The authors gratefully acknowledge that the work was supported by the Engineering and Physical Sciences Research Council (EPSRC) under Grant EP/I011056/1 and Platform Grant EP/H00453X/1, and EU Horizon 2020 Research and Innovation Programme Action Framework under the grant agreement 637302.

REFERENCES

[1] H. L. Wei, S. A. Billings, and M. Balikhin, “Prediction of the Dst index using multiresolution wavelet models,” J. Geophys. Res. Sp. Phys., vol. 109, no. A07212, 2004.

[2] H. L. Wei, D. Q. Zhu, S. A. Billings, and M. A. Balikhin, “Forecasting the geomagnetic activity of the Dst index using multiscale radial basis function networks,” Adv. Sp. Res., vol. 40, no. 12, pp. 1863–1870, 2007.

[3] Y. Kamide, W. Baumjohann, I. a. Daglis, W. D. Gonzalez, M. Grande, J. a. Joselyn, R. L. McPherron, J. L. Phillips, E. G. D. Reeves, G. Rostoker, a. S. Sharma, H. J. Singer, B. T. Tsurutani, and V. M. Vasyliunas, “Current understanding of magnetic storms: Storm-substorm relationships,” J. Geophys. Res. Sp. Phys., vol. 103, no. A8, pp. 17705–17728, 1998.

[4] T. Podladchikova, A. Petrukovich, and Y. Yermolaev, “Geomagnetic storm forecasting service StormFocus: 5 years online,” Journal of Space Weather and Space Climate, no. A22, 2018

[5] E. Camporeale, “The challenge of machine learning in space weather nowcasting and forecasting, ” arXiv:1903.05192v2 [physics.space-ph], unpublished.

[6] J. G. Wu and H. Lundstedt, “Prediction of geomagnetic storms from solar wind data using Elman recurrent neural networks,” Geophys. Res. Lett., vol. 23, no. 4, pp. 319-322, 1996.

[7] J. G. Wu and H. Lundstedt, “Geomagnetic storm predictions from solar wind data with the use of dynamic neural networks,” J. Geophys. Res. A Sp. Phys., vol. 102, no. 47, pp. 14255-14268, 1997.

[8] W. D. Gonzalez, A. Dal Lago, A. L. Clúa de Gonzalez, L. E. A. Vieira, and B. T. Tsurutani, “Prediction of peak-Dst from halo CME/magnetic cloud-speed observations,” J. Atmos. Solar-Terrestrial Phys., vol. 66, no. 2, pp. 161–165, 2004.

[9] E. Amata, G. Pallocchia, G. Consolini, M. F. Marcucci, and I. Bertello, “Comparison between three algorithms for Dst predictions over the 2003-2005 period,” J. Atmos. Solar-Terrestrial Phys., vol. 70, no. 2–4, pp. 496–502, 2008.

[10] M. Temerin and X. Li, “Dst model for 1995-2002,” J. Geophys. Res. Sp. Phys., vol. 111, no. A04221, 2006.

[11] M. Temerin and X. Li, “A new model for the prediction of Dst on the basis of the solar wind,” J. Geophys. Res. Sp. Phys., vol. 107, no. A12, 2002.

[12] J. Y. Lu, Y. X. Peng, M. Wang, S. J. Gu, M. X. Zhao, “Support Vector Machine combined with Distance Correlation learningforDstforecasting during intense geomagnetic storms, “Planetary and Space Science, Vol. 120, pp. 48-55, 2016.

[13] P. Wintoft, M. Wik, “Evaluation of Kp and Dst Predictions Using ACE and DSCOVR Solar Wind Data,” Space Weather, Vol. 16, no. 12, 2018

[14] R. J. Boynton, M. A. Balikhin, S. A. Billings, H. L. Wei, and N. Ganushkina, “Using the NARMAX OLS-ERR algorithm to obtain the most influential coupling functions that affect the evolution of the magnetosphere,” J. Geophys. Res. Sp. Phys., vol. 116, no. 5, 2011.

[15] R. J. Boynton, M. A. Balikhin, S. A. Billings, A. S. Sharma, and O. A. Amariutei, “Data derived narmax dst model,” Ann. Geophys., vol. 29, no. 6, pp. 965–971, 2011.

[16] Y Gu, H. L. Wei, M. Balikhin, “Nonlinear predictive model selection and model averaging using information criteria,” Vol. 6 , no. 1, pp. 319-328, 2018

[17] E. Y. Ji, Y. J. Moon, N. Gopalswamy, and D. H. Lee, “Comparison of Dst forecast models for intense geomagnetic storms,” J. Geophys. Res. Sp. Phys., vol. 117, no. 3, 2012.

[18] S. Chen and S. A. Billingst, “Representations of non-linear systems: The narmax model,” Int. J. Control, vol. 49, no. 3, pp. 1013–1032, 1989.

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

[20] 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, 1989.

[21] H. L. Wei, S. A. Billings, and J. Liu, “Term and variable selection for non-linear system identification,” *Int. J. Control*, vol. 77, no. 1, pp. 86–110, 2004.

[22] J. R. A. Solares, H. L. Wei, S. A. Billings, “A novel logistic-NARX model as a classifier for dynamic binary classification,” Neural Computing and Applications, Vol. 31, no. 1, pp. 11-25, 2019

[23] A. M. Marshall, G. R. Bigg, S. M. Leeuwen, J. K. Pinnegar, H. L. Wei, T. J. Webb, et al., "Quantifying heterogeneous responses of fish community size structure using novel combined statistical techniques," Global change biology, vol. 22, pp. 1755-1768, 2016.

[24] G. Bigg, H. Wei, D. Wilton, Y. Zhao, S. Billings, E. Hanna, et al., "A century of variation in the dependence of Greenland iceberg calving on ice sheet surface mass balance and regional climate change," in Proc. R. Soc. vol. 470, 2014.

[25] M. Balikhin, R. Boynton, S. Walker, J. Borovsky, S. Billings, and H. Wei, "Using the NARMAX approach to model the evolution of energetic electrons fluxes at geostationary orbit," Geophysical Research Letters, vol. 38, L18105, 2011.

[26] Y. Gu, H. L. Wei, R. J. Boynton, S. N. Walker, and M. A. Balikhin, “System Identification and Data-Driven Forecasting of AE Index and Prediction Uncertainty Analysis Using a New Cloud-NARX Model,” *J. Geophys. Res. Sp. Phys.*, vol. 124, no. 1, pp. 248–263, 2019.

[27] C. Billings, H. Wei, P. Thomas, S. Linnane, and B. Hope-Gill, "The prediction of in-flight hypoxaemia using non-linear equations," Respiratory medicine, vol. 107, pp. 841-847, 2013.

[28] Y. Gu and H.-L. Wei, "Significant Indicators and Determinants of Happiness: Evidence from a UK Survey and Revealed by a Data-Driven Systems Modelling Approach," Social Sciences, vol. 7, no. 53, 2018.

[29] Y. Li, H. L. Wei, S. A. Billings, and P. Sarrigiannis, "Identification of nonlinear time-varying systems using an online sliding-window and common model structure selection (CMSS) approach with applications to EEG," International Journal of Systems Science, vol. 47, pp. 2671-2681, 2015.

[30] W. D. Gonzalez, J. A. Joselyn, Y. Kamide, H. W. Kroehl, G. Rostoker, B. T. Tsurutani, and V. M. Vasyliunas, “What is a Geomagnetic Storm?,” *J. Geophys. Res.*, vol. 99, no. 44, pp. 5771-5792, 1994

[31] J. Chen, G. Q. Zeng, W. Zhou, W. Du and K. D. Lu, “Wind speed forecasting using nonlinear-learning ensemble of deep learning time series prediction and extremal optimization,” *Energy Conversion and Management*, vol. *165*, pp. 681-695, 2018.

[32] X. Huang, H. Wang, L. Xu, L. Liu, R. Li, and X. Dai, “Deep learning based solar flare forecasting model. I. Results for line-of-sight magnetograms,” *The Astrophysical Journal*, vol, *856*, no. 1, p.7, 2018.