Research Article

Forecasting SO$_2$ Pollution Incidents by means of Elman Artificial Neural Networks and ARIMA Models

Antonio Bernardo Sánchez,[1] Celestino Ordóñez,[2] Fernando Sánchez Lasheras,[1] Francisco Javier de Cos Juez,[2] and Javier Roca-Pardiñas[3]

[1] *Departamento de Construcción e Ingeniería de Fabricación, Universidad de Oviedo, 33203 Gijón, Spain*
[2] *Departamento de Explotación y Prospección de Minas, Universidad de Oviedo, Oviedo, 33004 Asturias, Spain*
[3] *Departamento de Estadística, Universidad de Vigo, 36310 Vigo, Spain*

Correspondence should be addressed to Celestino Ordóñez; ordonezcelestino@uniovi.es

An SO$_2$ emission episode at coal-fired power station occurs when the series of bihourly average of SO$_2$ concentration, taken at 5-minute intervals, is greater than a specific value. Advance prediction of these episodes of pollution is very important for companies generating electricity by burning coal since it allows them to take appropriate preventive measures. In order to forecast SO$_2$ pollution episodes, three different methods were tested: Elman neural networks, autoregressive integrated moving average (ARIMA) models, and a hybrid method combining both. The three methods were applied to a time series of SO$_2$ concentrations registered in a control station in the vicinity of a coal-fired power station. The results obtained showed a better performance of the hybrid method over the Elman networks and the ARIMA models. The best prediction was obtained 115 minutes in advance by the hybrid model.

1. Introduction

Coal-fired power stations are a major source of atmospheric pollutants, SO$_2$ being one of the most significant. Mixed with rain, with SO$_2$ is transformed into sulfuric acid producing acid rain. The wind helps to transport this element thousands of kilometers until it settles on the ground causing various negative effects. Sulfuric acid causes respiratory irritation, sometimes leading to damage to lung tissue.

European legislation on air pollution from coal-fired power stations establishes limits for the emissions of SO$_2$. Specifically, it imposes a limit on the average of 24 consecutive concentrations of SO$_2$ taken at 5-minute intervals. An emission episode is said to occur when the series of biannual averages is greater than a threshold set by current regulations. The interest is in predicting emission episodes with at least one hour in advance.

Forecasting SO$_2$ levels can be addressed through mathematical models such as autoregressive-moving-average (ARMA) or artificial neural networks (ANNs) models. An ARMA model was used by Hassanzadeh et al. in 2009 [1]

to forecast SO$_2$ levels for five stations. According to their results, an ARMA (2,2) model provides reliable predictions. Kandya and Mohan [2] study the forecasting of SO$_2$ (and other pollutants) using five statistical techniques. Although each technique has its own advantages and limitations, they found that the one-day prediction autoregressive integrated moving average (ARIMA) [3] technique scores well over the other techniques. Goyal et al. [4] pointed out that linear models such as multilinear regression MLR and ARIMA fail to predict extreme concentrations of pollutants.

Neural networks have also been used to predict concentrations of SO$_2$ and other pollutants emitted by a power plant. Mok and Tam [5] used three-layered feed-forward artificial neural networks to predict the daily SO$_2$ concentration 5 days in advance. Nunnari et al. [6] compared multilayer perceptron (MLP) models with a neurofuzzy approach and an autoregressive-moving-average model with exogenous inputs (ARMAX) model. The results confirmed the superiority of the MLP model over the other. Pérez et al. [7] compared the forecasting produced by three different methods (MLP, multiple linear regression (MLR), and persistence methods).

They concluded that the MLP models achieved more accurate regression results and better predictions. Fernández de Castro et al. [8] found good results for predicting SO$_2$ levels half an hour in advance in the neighborhood of a power plant using neural networks. Cortina et al. (2008) compared an adaptive linear neural network (ADALINE) and a generalized regression neural network (GRNN) for the prediction of pollution levels due to the chemical industry and electricity generation in Salamanca (Spain). Prediction experiments were carried out for 1, 12, and 24 hours in advance. They concluded that a linear regression network needs less adjustment of parameters than a nonlinear regression network, thus facilitating its implementation; however, to obtain better results with a linear regression network, they need to search for a pattern scheme. Abdul-Wahab and Al-Alawi [9] used a neural network model to forecast SO$_2$ concentration levels at a refinery in Oman; they also analyzed the effect of five meteorological parameters that were expected to affect the SO$_2$ concentrations. Zhang et al. [10] compared the performance of several statistical methods in SO$_2$ forecasting. The results showed that all the methods mentioned can be used in time series analysis of air pollutants although the denoising BP neural network has some advantages, mainly relating to its strong memory and learning ability.

Hybrid ARIMA-ANN models have also been applied in atmospheric pollutant forecasting. Tseng et al. [11] showed that the hybrid performed better than ARIMA or ANN alone. Díaz-Robles et al. [12] developed a hybrid model combining ARIMA and ANN to predict extreme events of particle emission in a city in Chile. They concluded that the hybrid model performs better than either of the models used separately.

In this study we analyzed the usefulness of Elman recurrent neural networks in forecasting SO$_2$ emission episodes in a coal-fired power station, given the capacity of this ANN to work with temporary data and compared to an ensemble method that improves the prediction of an ARIMA model using Elman recurrent neural networks. The method proposed is completely different from those combining ARIMA and ANN models previously exposed. The differences are twofold: (1) the kind of neural network used is different (Elman neural network instead of back-propagation recurrent neural network) and (2) the way the network is used, not for an adjustment of the ARIMA model residuals but for the prediction of SO$_2$ concentration in half an hour before the time of interest.

2. Materials and Methods

2.1. The Database. The data for the present research was obtained from a coal-fired power station in northern Spain. It contains the average 5-minute concentrations of SO$_2$ measured during the year 2012 in a control station located in the neighborhood of the power station. Figure 1 shows the time series for the whole year. As be appreciated, there are some SO$_2$ peaks in the line representing the SO$_2$ concentrations registered in the control station, some of them corresponding to emission episodes.
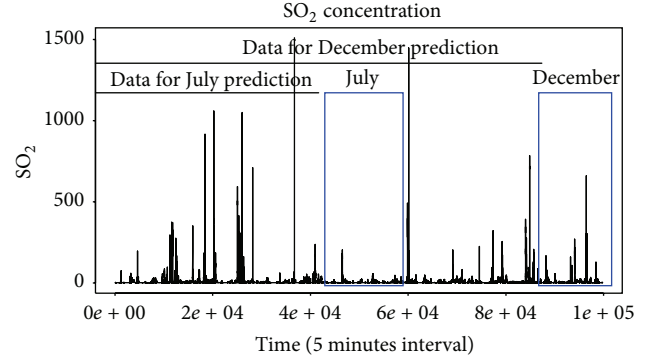

Figure 1: Time series of SO$_2$ concentrations registered on the control station. Data from July and December were used to validate the models.

The data corresponding from January 1 to June 30 was used to construct one model, while that corresponding to July was used for the referred model validation. Also, another test was carried out. For this test, models were trained with data from January 1 to November 30 and SO$_2$ records for December were used for validation. The purpose of using these two sets of data for the models training was to compare the results obtained with different amounts of training information and also to perform the SO$_2$ concentration prediction in two different seasons (summer and winter, resp.).

2.2. Methods

2.2.1. The ARIMA Models. ARIMA models are the most generally used class of models for forecasting time series that can be stationarized by transformations such as differencing and logging [13]. The acronym ARIMA stands for auto-regressive integrated moving average. lags of the differenced series appearing in the forecasting equation are called *autoregressive* terms, while lags of the forecast errors are called *moving average* terms. A time series which needs to be differenced to be made stationary is said to be an *integrated* version of a stationary series.

A nonseasonal ARIMA model [13] is classified as an ARIMA (p, d, q) model, where

(i) p is the number of autoregressive terms,

(ii) d is the number of non-seasonal differences,

(iii) q is the number of lagged forecast errors in the prediction equation.

The generalized form of ARIMA can be described as follows [14]:

$$\emptyset(B) \cdot \Phi(B^S) \cdot (1-B)^d \cdot (1-B)^D \cdot Y_t = \theta(B) \cdot \Theta(B^S) \cdot Z_t, \tag{1}$$

where B is the backward shift operator, d is non-seasonal order of differences, D is seasonal order of differences, and $\emptyset, \Phi, \theta$, and Θ are polynomials in B and B^S.
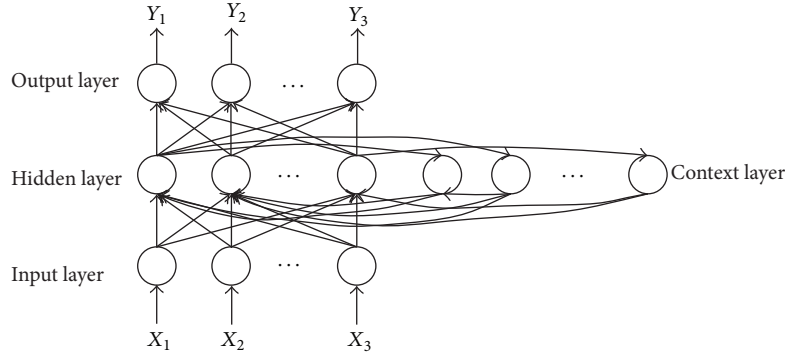
FIGURE 2: Scheme of a recursive Elman neural network with one hidden layer.

Forecasting based on ARIMA (autoregressive integrated moving averages) models, commonly known as the Box-Jenkins approach, comprises the following stages:

(i) model identification,

(ii) parameter estimation,

(iii) diagnostic checking.

These stages are repeated until a suitable model for the given data has been identified. In this research we have used a variation of the Hyndman and Khandakar algorithm [15] which combines unit root tests and minimization of the AIC and MLE to obtain the ARIMA models. The use of these algorithms speeds up the model identification process.

2.2.2. Recurrent Neural Networks Models. For the present research a kind of partially recurrent neural network (RNN) called an Elman network [16] was employed. An Elman RNN is a network with an initial configuration based on a regular feedforward neural network.

An Elman network has a layer called a context layer. The neurons in the context layer, which are called context neurons, hold a copy of the output given by the neurons of the hidden layer to the output one (Figure 2). In the following computing step, information that was given as an output by the hidden layer is used as a new input information for this layer.

The strength of the relationships between neurons in an Elman RNN is indicated by their weights. For this kind of neural network, the weight values of the neurons are chosen randomly when the process is initiated, and their values are changed during the model training in order to optimize them, with the exception of the weights from the hidden layer which do not change during the training process because the values of the context neurons need to receive the output information of the hidden layer as it is calculated.

The dynamics of the Elman RNN are described by the following equations [17]:

$$X_h \cdot (k+1) = S \cdot [W_c \cdot X_c(k+1) \cdot W_1 \cdot U(k)],$$

$$X_c \cdot (k+1) = W_t \cdot X_h, \qquad (2)$$

$$Y \cdot (k+1) = W_o \cdot X_h(k+1),$$

where S is hyperbolic tangent function, $U(k)$ is an input of the network at a discrete time k, $Y(k)$ is output of the network at a discrete time k, X_c is nodes of the context layer, X_h nodes of the hidden layer, W_c is weight matrix of the context-hidden layer, W_I is weight matrix of the input-hidden layer, W_r is weight matrix of the hidden-context layer, and W_o is weight matrix of the hidden-output layer.

The training of the Elman recurrent neural networks models was made using the Levenberg-Marquardt algorithm. This procedure is a modification of Gauss-Newton's method, which was designed in order to minimize the sum of squares of nonlinear functions combining this technique with the steepest-descent algorithm. The Levenberg-Marquardt algorithm, whose application is currently very common for RNN, was chosen as it does not suffer the slow convergence problems that were reported in the methods [18] from which it is derived.

2.2.3. The Proposed Hybrid Model. Since their introduction in the 1970s [3], ARIMA models have been used for the forecasting of linear time series. As has already been reported in the previous studies, these kinds of models have a poor performance as regards prediction in nonlinear problems [19]. In order to overcome their limitations, a hybrid model is proposed in the present research. This hybrid model consists of two main steps:

(i) training of an Elman recurrent neural network in order to mimic the temporal linear behavior of the SO_2 time series, predicting some output values that will be the input of the ARIMA model,

(ii) selection of an ARIMA model that will model the time variation of the SO_2 concentration, using as input values the prediction of the recurrent neural network.

Therefore, as may be observed in Figure 3, the proposed hybrid model uses the SO_2 concentrations predicted by the Elman recurrent neural network model two to three hours in advance as input for an ARIMA model in order to achieve a more accurate prediction of the SO_2 concentration. The main goal of this hybrid model is not only to achieve a general more accurate prediction of the SO_2 concentration at all times, but also to improve the detection of pollution incidents. In other
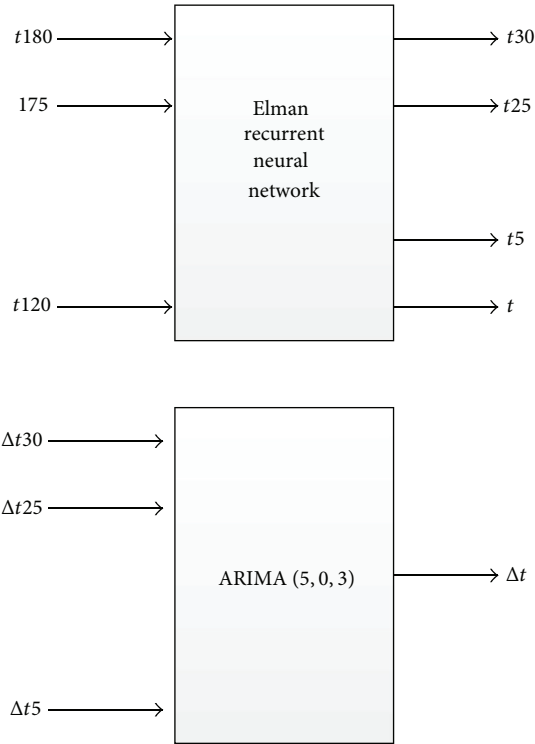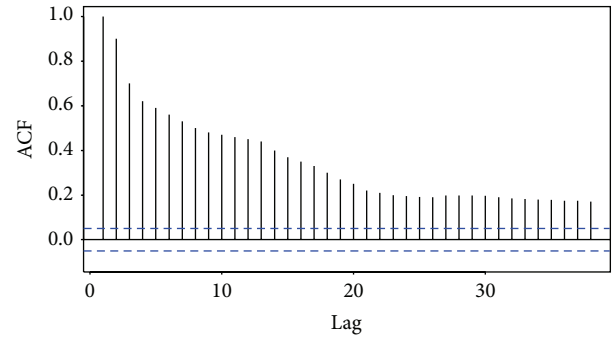
FIGURE 3: Scheme of the hybrid model.

(a)

(b)

FIGURE 4: Autocorrelation function variable SO_2 (a) and residual partial correlation (b).
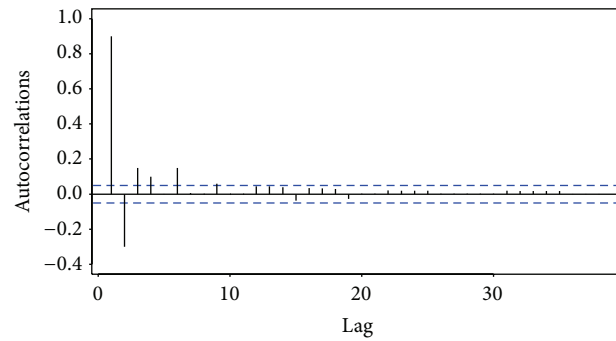
words, this hybrid model was chosen in order to improve the capability of the previous model to detect incidences of pollution as early as possible. The SO_2 concentration in each moment is calculated as the sum of the concentration 5 minutes before ($t5$) and in the mentioned moment (t) given by the recurrent neural network, plus the increase in the SO_2 concentration from the previous moment, and divided by two.
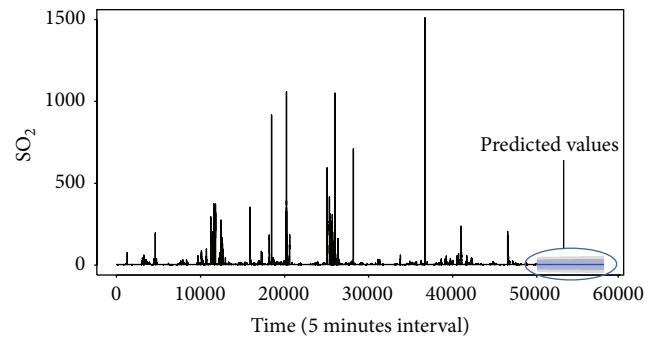
3. Results and Discussion

3.1. Results of the ARIMA Model. The best model found using the Hyndman and Khandakar algorithm was the ARIMA $(5, 0, 3)$. Figure 4 shows the autocorrelation functions (a) and the residual partial autocorrelation (b). The root-mean-square error (RMSE) obtained for the set of training data (data from January to November) applying the mentioned ARIMA model was 8.7446. The Ljung-Box statistic [20] was used to check the adequacy of model. The P value for the Ljung-Box statistic was 0.9821, and therefore it can be stated that the data in the residuals were independently distributed or, in other words, that the residuals from the ARIMA model have no correlation. The error obtained when the ARIMA $(5, 0, 3)$ model was applied to the set of validation data set corresponding to the month of December was of 8.9629. Similarly, when the equivalent model with data from January to June was trained, the RMSE obtained was 8.9101. In this case, the P value of the Ljung-Box statistic was 0.8347 while the RMSE value of the model applied to the data of July gave a value of 10.60121. As it can be observed in Figure 5, the main

FIGURE 5: ARIMA $(5,0,3)$: December 2012. Predicted SO_2 concentration.

weakness of the ARIMA model is that in spite of its good average, it seems to be unable to predict pollution incidents because it is unable to reproduce the real peaks of SO_2 concentration (see Figure 1).

3.2. Results of the Recurrent Neural Network Model. Some Elman recurrent neural networks architectures were tested in order to find the best generalization characteristics of the data. The input variables employed were the SO_2 concentrations 2 and 3 hours before the current time and recorded each five minutes.

The best configuration resulted in an Elman neural network with 11 neurons in the hidden layer. The activation
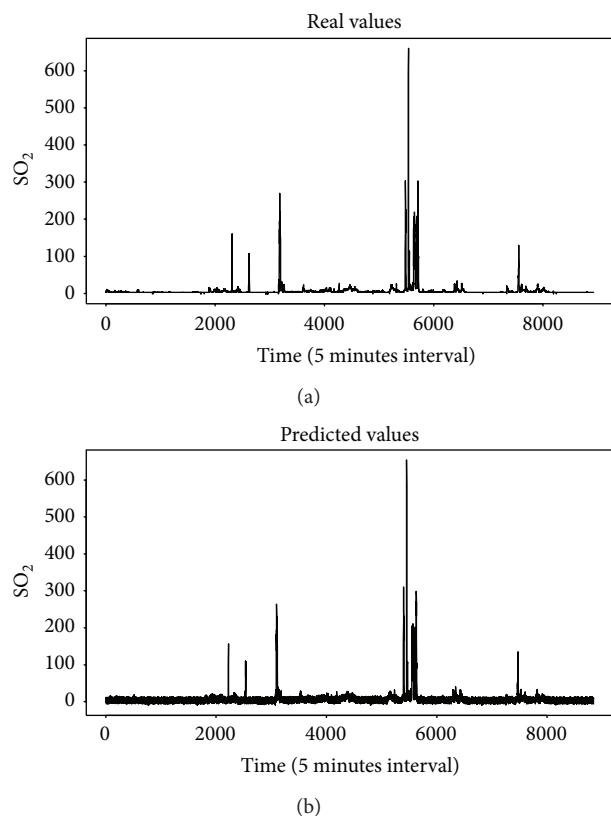
FIGURE 6: Hybrid model: December 2012. Real SO_2 concentration (a) versus predicted concentration (b).

function employed was the hyperbolic tangent [21], while the learning rate was 0.1 and the momentum 0.9. The results obtained using the information from January to November as training data gave us an RMSE of 5.5225 while convergence was achieved after 934 epochs in the case of the model trained with data from January to June; convergence was achieved after 632 epochs with an RMSE of the training data set of 7.8013. The validation results corresponding to the dataset of December, trained with the data from January to November were 7.6121 while the results of the data corresponding to the month of July using the model trained with the data from January to June gave us an RMSE result of 8.0347.

3.3. Results of the Hybrid Model. The proposed hybrid model was applied to the same database. First of all, a hybrid model using the values of the SO_2 concentration from January to November as input data was trained. The RMSE value of the training data set was 5.0238 while the RMSE obtained using the validation subset corresponding to the month of December was 6.6850. Finally, another model was trained using data from January to June, obtaining an RMSE of 6.6014 while the RMSE obtained when the model was validated with data from the month of July, the value obtained was 6.5356. In this case, the number of epochs that were necessary for the Elman recurrent neural networks convergence was from 1023 and 712, respectively. Figure 6 shows the results obtained of the application of the hybrid model to the

month of December. This figure represents a total of 8,928 measurements, showing their real value and the predicted one. It can be observed how the hybrid model is able to predict the SO_2 pollution incidences although in some cases it is not able to predict their maximum values.

4. Conclusions

In the present research, the utility of three different mathematical models (ARIMA, Elman recurrent neural network, and a hybrid model) to predict SO_2 emission episodes of a coal-fired station was analyzed. Emission episodes correspond to peaks in the time series of SO_2 concentration.

The ARIMA model was not able to reproduce emission episodes, just the general trend of the time series. The Elman recurrent neural network performed better given its capacity to detect emission episodes. However, the best results were obtained with a hybrid model that applies the ARIMA model to the Elman neural network output.

The results obtained with the hybrid model made it possible to predict emission episodes 115 minutes in advance, which is a sufficient response time to take preventive measures.

We would like to remark that the main advantages of the method proposed in the present research are, on the one hand, linked with the capability of Elman recurrent neural networks to perform sequence prediction that is beyond the power of a standard backpropagation recurrent neural network, and, on the other hand, with how its capabilities are used in our hybrid model. Therefore, in the proposed hybrid model, the predictions obtained from the Elman recurrent neural network are used as input values for an ARIMA model that, having information corresponding from 30 to 5 minutes in advance, is able to predict the SO_2 concentration 55 minutes in advance.

References

[1] S. Hassanzadeh, F. Hosseinibalam, and R. Alizadeh, "Statistical models and time series forecasting of sulfur dioxide: a case study Tehran," *Environmental Monitoring and Assessment*, vol. 155, no. 1–4, pp. 149–155, 2009.

[2] A. Kandya and M. Mohan, "Forecasting the urban air quality using various statistical techniques," in *Proceedings of the 7th International Conference on Urban Climate*, Yokohama, Japan, 2009.

[3] G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco, Calif, USA, 1976.

[4] P. Goyal, A. T. Chan, and N. Jaiswal, "Statistical models for the prediction of respirable suspended particulate matter in urban cities," *Atmospheric Environment*, vol. 40, no. 11, pp. 2068–2077, 2006.

[5] K. M. Mok and S. C. Tam, "Short-term prediction of SO_2 concentration in Macau with artificial neural networks," *Energy and Buildings*, vol. 28, no. 3, pp. 279–286, 1998.

[6] G. Nunnari, A. F. M. Nucifora, and C. Randieri, "The application of neural techniques to the modelling of time-series of atmospheric pollution data," *Ecological Modelling*, vol. 111, no. 2-3, pp. 187–205, 1998.

[7] P. Pérez, A. Trier, and J. Reyes, "Prediction of PM2.5 concentrations several hours in advance using neural networks in Santiago, Chile," *Atmospheric Environment*, vol. 34, no. 8, pp. 1189–1196, 2000.

[8] B. M. Fernández de Castro, J. M. Prada Sánchez, W. González Manteiga, M. Febrero Bande, J. L. Bermúdez Cela, and J. J. Hernández Fernández, "Prediction of SO_2 levels using neural networks," *Journal of the Air & Waste Management Association*, vol. 53, no. 5, pp. 532–539, 2003.

[9] S. A. Abdul-Wahab and S. M. Al-Alawi, "Prediction of sulfur dioxide (SO_2) concentration levels from the Mina Al-Fahal Refinery in Oman using Artificial Neural Networks," *American Journal of Environmental Sciences*, vol. 4, no. 5, pp. 473–481, 2008.

[10] J. Zhang, H. Jiang, Z. Chen, X. Li, and Y. Lu, "The comparison of environmental time series statistical prediction methods," in *Proceedings of the Asia Pacific Conference on Environmental Science and Technology Advances in Biomedical Engineering*, vol. 6, pp. 267–272, 2012.

[11] F.-M. Tseng, H.-C. Yu, and G.-H. Tzeng, "Combining neural network model with seasonal time series ARIMA model," *Technological Forecasting and Social Change*, vol. 69, no. 1, pp. 71–87, 2002.

[12] L. A. Díaz-Robles, J. C. Ortega, J. S. Fu et al., "A hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban areas: the case of Temuco, Chile," *Atmospheric Environment*, vol. 42, no. 35, pp. 8331–8340, 2008.

[13] T. C. Mills, *Time Series Techniques for Economists*, Cambridge University Press, Cambridge, UK, 1990.

[14] C.-S. Ong, J.-J. Huang, and G.-H. Tzeng, "Model identification of ARIMA family using genetic algorithms," *Applied Mathematics and Computation*, vol. 164, no. 3, pp. 885–912, 2005.

[15] R. J. Hyndman and Y. Khandakar, "Automatic time series forecasting: the forecast package for R," *Journal of Statistical Software*, vol. 27, no. 3, pp. 1–22, 2008.

[16] J. L. Elman, "Finding structure in time," *Cognitive Science*, vol. 14, no. 2, pp. 179–211, 1990.

[17] S. Ren and L. Gao, "Resolve of overlapping voltammetric signals in using a wavelet packet transform based Elman recurrent neural network," *Journal of Electroanalytical Chemistry*, vol. 586, no. 1, pp. 23–30, 2006.

[18] M. T. Hagan and M. B. Menhaj, "Training feedforward networks with the Marquardt algorithm," *IEEE Transactions on Neural Networks*, vol. 5, no. 6, pp. 989–993, 1994.

[19] P. G. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, vol. 50, pp. 159–175, 2003.

[20] G. M. Ljung and G. E. P. Box, "On a measure of lack of fit in time series models," *Biometrika*, vol. 65, no. 2, pp. 297–303, 1978.

[21] A. C. Mathias and P. C. Rech, "Hopfield neural network: the hyperbolic tangent and the piecewise-linear activation functions," *Neural Networks*, vol. 34, pp. 42–45, 2012.