Electricity Sales Forecasting Using Hybrid Autoregressive Integrated Moving Average and Soft Computing Approaches in the Absence of Explanatory Variables

Yuehjen E. Shao * and Yi-Shan Tsai

Department of Statistics and Information Science, Fu Jen Catholic University,
New Taipei City 24205, Taiwan (R.O.C); jilltsai13@gmail.com
* Correspondence: stat1003@mail.fju.edu.tw; Tel.: +886-2-2905-2647

Abstract: Electricity is important because it is the most common energy source that we consume and depend on in our everyday lives. Consequently, the forecasting of electricity sales is essential. Typical forecasting approaches often generate electricity sales forecasts based on certain explanatory variables. However, these forecasting approaches are limited by the fact that future explanatory variables are unknown. To improve forecasting accuracy, recent hybrid forecasting approaches have developed different feature selection techniques (FSTs) to obtain fewer but more significant explanatory variables. However, these significant explanatory variables will still not be available in the future, despite being screened by effective FSTs. This study proposes the autoregressive integrated moving average (ARIMA) technique to serve as the FST for hybrid forecasting models. Aside from the ARIMA element, the proposed hybrid models also include artificial neural networks (ANN) and multivariate adaptive regression splines (MARS) because of their efficient and fast algorithms and effective forecasting performance. ARIMA can identify significant self-predictor variables that will be available in the future. The significant self-predictor variables obtained can then serve as the inputs for ANN and MARS models. These hybrid approaches have been seldom investigated on the electricity sales forecasting. This study proposes several forecasting models that do not require explanatory variables to forecast the industrial electricity, residential electricity, and commercial electricity sales in Taiwan. The experimental results reveal that the significant self-predictor variables obtained from ARIMA can improve the forecasting accuracy of ANN and MARS models.

Keywords: forecast; electricity sales; autoregressive integrated moving average (ARIMA); artificial neural networks; multivariate adaptive regression splines; hybrid

1. Introduction

Electricity is one of the most important sources of energy on earth. Typical approaches use single-stage or single models to forecast electricity sales. In general, single models use two kinds of forecasting techniques: statistical modeling and soft computing modeling. For example, the impacts of population and weather-sensitive parameters on electricity consumption in Delhi were investigated [1]. Different multiple linear regression models were developed for the various seasons. The residential demand for electricity in Taiwan from 1955 to 1996 was investigated [2]. In [2], it was assumed that the electricity demand was a function of population growth, electricity prices, the degree of urbanization, and household disposable income. Different long-term regression models for the prediction of electricity consumption in Italy have been proposed [3]. The explanatory variables included gross domestic product (GDP), GDP per capita, and population. Additionally, selected

economic and demographic variables that affect annual electricity consumption in New Zealand were studied [4]. Forecasting models that use multiple linear regression analysis have been proposed. The regression model, a Holt-Winters exponential smoothing method, and the trigonometric gray model with a rolling mechanism were proposed to forecast electricity consumption in Romania [5].

However, regression analysis has its own limitations, such as variation homogeneity for the regression models. Consequently, autoregressive integrated moving average (ARIMA) modeling has been used to forecast electricity sales. ARIMA modeling is appropriate because seasonal effects are included in the forecasts. ARIMA is a well-known forecasting technique that consists of different types of processes. Those processes include autoregressive (AR), moving average (MA), and combined AR and MA. Electricity demand load was forecasted using ARIMA modeling [6]. Comparisons between ARIMA, seasonal ARIMA, and a regression model for forecasting electricity demand have been conducted [7]. The forecasting of electric prices was performed using ARIMA [8]. A combined AR and a finite impulse response filter were proposed for forecasting electricity demand in Lebanon [9]. Several linear or ARIMA-like models and nonlinear models were studied for forecasting electricity consumption in Taiwan [10]. Although ARIMA modeling is widely used for forecasting electricity demand or sales, the linear form of ARIMA models may have difficulty capturing the nonlinear pattern of electricity sales data [11–13].

In contrast to statistical modeling, the artificial neural network (ANN) and multivariate adaptive regression splines (MARS) techniques contain fewer a priori assumptions. ANN and MARS can also serve as alternative modeling schemes for electricity sales forecasting since both can model nonlinearity and provide good forecasting characteristics [14–18]. Determining Turkey's electricity consumption using ANN training algorithms has been proposed [19]. A methodology based on ANN was studied to forecast day-ahead electricity spikes and prices in Ontario, Canada [20]. Three techniques for forecasting electricity consumption (regression analysis, decision trees, and ANN) were investigated [21]. Least squares support vector machines were used to forecast electricity consumption in Turkey [22]. Traditional regression analysis and ANN have also been used for comparison purposes; gross electricity generation, installed capacity, total subscribership, and population were used as explanatory variables. ANN and regression analyses were used to forecast energy consumption in Turkey based on socio-economic and demographic variables [23]. The ANN with four explanatory variables (GDP, population, and the amount of imports and exports) was considered to be the most suitable model. ANN, linear regression, and nonlinear regression approaches were considered in forecasting the electricity consumption of residential and industrial sectors in Turkey [24]. The explanatory variables for the models included installed capacity, gross electricity generation, population, and the total number of customers. However, ANN has been criticized for its long training process when designing the optimal structure [25,26].

In recent years, hybrid forecasting models with feature selection techniques (FSTs) have been studied for electricity sales forecasting. A hybrid model with fast ensemble empirical mode decomposition, variational mode decomposition, and ANN was investigated in forecasting electricity prices [27]. A hybrid forecasting approach for managing and scheduling electric power generation has also been studied [28]. A combined model, including wavelet transforms, ARIMA, and ANN, was used to simultaneously forecast electricity demand and price [29]. A hybrid approach that integrated wavelet transforms, a kernel extreme learning machine, and an autoregressive moving average was developed to forecast electricity prices [30]. A combined forecasting approach based on ANN, an adaptive network-based fuzzy inference system, and seasonal ARIMA was proposed to forecast short-term electricity demand [31]. A hybrid forecasting model that combined a support vector machine and ARIMA with external input modules was proposed to forecast mid-term electricity market clearing prices [32]. In [32], it was assumed that all forecasting input data were given, which may not be practical.

The work used a tree-based machine learning technique to predict the electricity price [33]. Additionally, the hybridization of ARIMA and soft computing techniques to enhance the accuracy of

food crop price prediction was incorporated in the design [34,35]. Additionally, they employed an ARIMA approach to predict conventional electrical load and charging demand of electric vehicles [36]. The study improved the accuracy of ARIMA forecaster by optimizing the integrated and AR order parameters. In [37], they considered the users' long-term electricity load scheduling problem and model the changes of the price information and load demand as a Markov decision process. This study developed an online load scheduling learning algorithm based on the actor-critic method to determine the users' Markov perfect equilibrium policy.

After the literature review for electricity sales and demand forecasting was completed in this study, we made the following observations. The first and most important observation was that the forecasting of electricity sales is extremely worthy of investigation. The second observation was that a significant number of forecasting models generate forecasts based on certain known explanatory variables, which may not be feasible in practical applications. The third observation was that traditional approaches, such as regression, econometric, and ARIMA, as well as soft computing techniques, are widely used for electricity sales forecasting [38]. However, a good technique, MARS, has not been adopted for electricity sales forecasting. The final observation is that electricity sales data series may possess linear and/or nonlinear features and that hybrid forecasting techniques should be used to improve forecasting accuracy. However, although most FSTs in hybrid models can provide a good selection of significant explanatory variables, these variables may not be available in the future and those significant explanatory variables may thus not be appropriate for electricity sales forecasting.

This study proposed the single ANN, single MARS, hybrid ARIMA and ANN (ARIMA-ANN), and hybrid ARIMA and MARS (ARIMA-MARS) models to forecast the industrial electricity, residential electricity, and commercial electricity sales (IERECES) in Taiwan without using any explanatory variables. The reason for using ANN is that it is one of the most widely used soft computing methods for electricity sales forecasting. The reason for using MARS is that it has not been used for electricity sales forecasting. Additionally, both the ANN and MARS techniques are suitable for modeling the nonlinear features of electricity sales data series. Finally, the reason for using the hybrid ARIMA-ANN and ARIMA-MARS is that they possess linear and nonlinear forecasting abilities. The fundamental concept of the hybrid forecasting scheme is to capture various characteristics in the data by making use of individual model's dominance [39,40]. Therefore, the hybrid technique can improve forecasting accuracy using each model's unique features. Since single ANN and MARS were used to forecast the IERECES in Taiwan, this study integrated ARIMA and ANN, and ARIMA and MARS as the hybrid models. Most importantly, since most FSTs select important explanatory variables that may not be available in the future, this study logically employed ARIMA modeling as the FST for the proposed hybrid forecasting models. ARIMA can identify significant self-predictor variables that will be available in the future. In this study, a self-predictor variable refers to the future values of the variable that can be computed using the preceding or available values of the variable itself.

The real dataset of IERECES in Taiwan was collected from January 2006 to July 2016. The IERECES dataset contains three monthly data series: industrial electricity (IE), residential electricity (RE), and the commercial electricity (CE) sales series. This study used the first 115 monthly data records to build different forecasting models and used the last 12 monthly data records for confirmation. The rest of this paper is structured as follows. Section 2 presents the design of different forecasting methodologies. Section 3 introduces the real IERECES data series in Taiwan. Three types of electricity sales series are fitted to four single and two hybrid forecasting models, without using any explanatory variables. Section 4 provides the discussion about forecasting comparisons for four single models and two hybrid models. The final section addresses the research findings and conclusions inferred from this study.

2. Materials and Methods

In this section, the concept and the modeling processes for ARIMA, ANN and MARS are briefly introduced.

2.1. Autoregressive Integrated Moving Average (ARIMA) Modeling

A time series consists of a set of observations on the values that a variable takes at different times, ordered sequentially. Because seasonal effects are involved in electricity sales forecasting, time series forecasting techniques should be used. A well-known ARIMA forecasting technique was developed to predict time series data [41]. A general seasonal ARIMA model is described as follows [41]:

$$\phi_p(B)\Phi_P(B^L)(1-B)^d(1-B^L)^D Z_t = \delta + \theta_q(B)\Theta_Q(B^L)a_t, \tag{1}$$

where

B:	the backward shift operator, defined as: $B^j y_t = y_{t-j}$,
L:	the number of seasons in a year, and $L = 12$ for monthly data,
d:	the values of non-seasonal difference transformations,
D:	the values of seasonal difference transformations,
Z_t:	the working series, which are stationary after fitting a suitable difference transformation from original time series Y_t
δ:	an unknown constant,
a_t:	white noise at time t, which is independent and identical (iid) with normal distribution,
p:	the order of non-seasonal autoregressive (AR) models,
P:	the order of seasonal AR models,
q:	the order of non-seasonal moving average (MA) models,
Q:	the order of seasonal MA models,
$\phi_p(B):$	a polynomial function for a non-seasonal AR model, defined as: $\phi_p(B) = (1 - \phi_1 B - \phi_2 B^2 - \ldots\ldots - \phi_p B^p),$
$\Phi_P(B^L):$	a polynomial function for a seasonal AR model, defined as: $\Phi_P(B^L) = (1 - \Phi_{1,L}B^L - \Phi_{2,L}B^{2L} - \ldots\ldots - \Phi_{P,L}B^{PL}),$
$\theta_q(B):$	a polynomial function for a non-seasonal MA model, defined as: $\theta_q(B) = (1 - \theta_1 B - \theta_2 B^2 - \ldots\ldots - \theta_q B^q),$
$\Theta_Q(B^L):$	a polynomial function for a seasonal MA model, defined as: $\Theta_Q(B^L) = (1 - \Theta_{1,L}B^L - \Theta_{2,L}B^{2L} - \ldots\ldots - \Theta_{P,L}B^{QL}),$

The original nonstationary time series should be transformed into a stationary working series by differencing. Typically, the transformation is carried out using four sets of *d* and *D*. Those four sets include $(d, D) = (0, 0)$, $(d, D) = (1, 0)$, $(d, D) = (0, 1)$, and $(d, D) = (1, 1)$. After obtaining stationary working series, we can apply the sample autocorrelation function (ACF) and sample partial autocorrelation function (PACF) to determine the values of *p*, *P*, *q*, and *Q* for the seasonal ARIMA models.

Additionally, diagnostic checks for testing the lack of fit of a time series model must be performed. The Ljung–Box test was applied to the residuals of a time series after fitting a seasonal ARIMA model to the working series [42]. Instead of testing randomness at each distinct lag, the Ljung–Box statistic was used to test the overall randomness based on several lags. The Ljung–Box statistic is described as follows:

$$Q = n_d (n_d + 2)\sum_{l=1}^{k} (n' - l)^{-1} r_l^2(\hat{a}), \text{ and } n_d = n - d, \tag{2}$$

where

Q:	Ljung–Box statistics, and the asymptotic distribution of Q follows a chi-square distribution with k-n_p degrees of freedom,
n_p:	number of parameters other than δ that must be estimated in the ARIMA model under consideration,
n:	the number of observations,
$r_l^2(\hat{a})$:	the square of the sample autocorrelation of the residuals at lag *l*.

The Ljung–Box test rejects the null hypothesis, which indicates that the underlying model has significant lack of fit if:

$$Q > x_\alpha^2(k - n_p),$$

where $x_\alpha^2(k - n_p)$ is the chi-square distribution table value with k-n_p degrees of freedom and significance level α.

2.2. Artificial Neural Network (ANN) Modeling

An ANN is a massively parallel system composed of highly interconnected and interacting processing units that are based on neurobiological models. ANNs handle information via the interactions of neurons [43]. ANNs have been extensively used in modeling nonlinear processes because of their associated memory characteristics and generalization capabilities. In this study, an ANN was used as an alternative to the single modeling approach to forecast IERECES data series in Taiwan.

The input, hidden, and output layers consist of the ANN. Those three layers contain several nodes or neurons connected by links. An ANN contains nodes that are connected to themselves, which enables a node to influence other nodes and itself. An external source provides input signals to the nodes in the input layer, and the output signals are produced by the nodes in the output layer. The hidden nodes are nonlinear functions of the input variables. The predicted output variable is a function of the nodes in the hidden layer.

ANN modeling can be simply described as follows. For an ANN model, the relationship between inputs $(x_1, x_2, \ldots, x_a)$ and output (y) is represented as [16]

$$y = \alpha_0 + \sum_{j=1}^{b} \alpha_j g\left(\delta_{0j} + \sum_{i=1}^{a} \delta_{ij} x_i\right) + \varepsilon, \tag{3}$$

where α_j $(j = 0, 1, 2, \ldots, b)$ and δ_{ij} $(i = 0, 1, 2, \ldots a; j = 0, 1, 2, \ldots, b)$ are the connection weights of the model; a denotes the number of input nodes; b stands for the number of hidden nodes; and ε denotes the error term. In the hidden layer, the transfer function is often used with a logistic function [16]:

$$g(z) = \frac{1}{1 + \exp(-z)}, \tag{4}$$

Consequently, the ANN transports nonlinear functional mapping from the inputs $(I_1, I_2, \ldots, I_a)$ to the output O; namely, [16]

$$O = f(I_1, I_2, \ldots, I_a, w) + \varepsilon, \tag{5}$$

where w is a vector of parameters and f is a function determined by the ANN structure and the connection weights.

2.3. Multivariate Adaptive Regression Splines (MARS) Modeling

MARS was initially developed as a procedure that modeled the relationships between a few variables [44]. The modeling procedure for MARS is based on a divide-and-conquer strategy where training sets are partitioned into separate regions. Those regions are determined by the regression equation. MARS can usually reveal the important data characteristics that often hide in high-dimensional data series. It designs the models by fitting piecewise linear regressions, and the nonlinearity of a model is approximated by using separate regression slopes at distinct intervals in the explanatory variable space. Therefore, the slope of the regression line changes from one interval to the next as the two "knot" points are crossed. The variables and endpoints of the intervals for each variable can be obtained via a fast but intensive search procedure.

The general MARS model can be described as follows [25]:

$$f(x) = b_0 + \sum_{m=1}^{M} b_m \prod_{j=1}^{j_m} [S_{jm}(x_{v(j,m)} - l_{jm})], \tag{6}$$

where b_0 and b_m are the parameters, M is the number of basis functions (BFs), J_m is the number of knots, S_{jm} takes on values of either 1 or -1 and shows the right or left sense of the corresponding step function, $v(j,m)$ is the label of the independent variable, and l_{jm} is the knot location. The optimal MARS model is selected in a two-step procedure. The first step is to build many BFs to initially fit the data. The BFs are deleted in the order of least contributions to most contributions using the generalized cross-validation (GCV) criterion in the second step. When a variable is removed from the model, the measure of variable importance is provided by computing the GCV values. The GCV is described as follows [25]:

$$GCV(M) = \frac{1}{N} \sum_{i=1}^{N} [y_i - f_M(x_i)]^2 \bigg/ [1 - \frac{C(M)}{N}]^2, \tag{7}$$

where N is the number of observations and $C(M)$ is the set of cost penalty measures of a model containing M BFs.

2.4. Research Framework

Figure 1 shows a generalized depiction of the research framework. As shown in Figure 1, since no explanatory variables were used to forecast the electricity sales in this study, we proposed two types of schemes for ANN and MARS modeling.
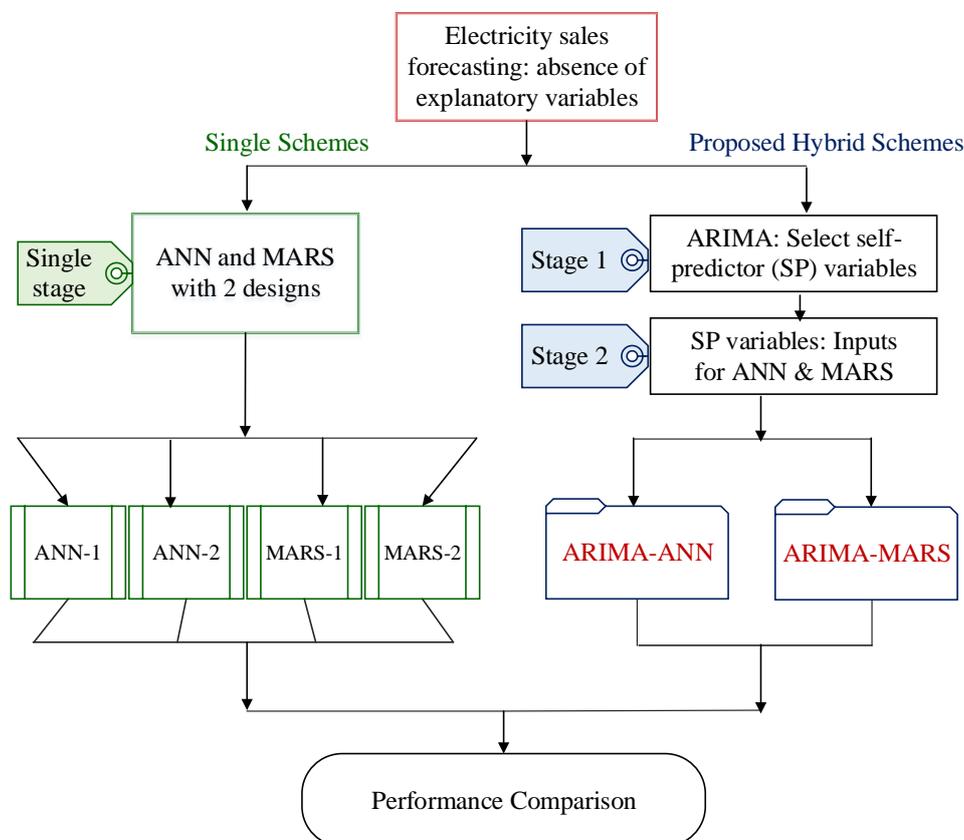

Figure 1. The research framework. ANN: artificial neural network; MARS: multivariate adaptive regression splines; ARIMA: autoregressive integrated moving average.

The first (single) scheme contains two designs. In ANN modeling, the first design arbitrarily selects the preceding two observations at time t-1 and t-2 (i.e., y_{t-1} and y_{t-2}) to serve as the self-predictor variables. Thus, the two variables y_{t-1} and y_{t-2} are considered as two input variables and the variable y_t is deemed the output variable. Since seasonal effects may influence IERECES forecasts, the second ANN design in this study used the preceding 12 observations to forecast electricity sales at time t, and the second design used 12 variables, y_{t-1}, y_{t-2}, y_{t-3}, ... , and y_{t-12}, to serve as input variables and used the single variable y_t to serve as the output for ANN modeling; that is, the second design used 12 input nodes and one output node for the ANN structure. The first and the second designs of ANN are denoted as ANN-1 and ANN-2, respectively. In the modeling of MARS, this study also used the same input structure as ANN modeling, that is, it used two designs for the input variables of MARS modeling. The first design used y_{t-1} and y_{t-2} as two input variables and y_t as the output variable. The second design considered 12 variables, y_{t-1}, y_{t-2}, y_{t-3}, ... , and y_{t-12} as the input variables and y_t as the output variable. The first and the second designs of MARS are denoted as MARS-1 and MARS-2, respectively.

The second (proposed hybrid) scheme included the hybridization of ARIMA and ANN (i.e., ARIMA-ANN) and ARIMA and MARS (i.e., ARIMA-MARS). For the proposed hybrid modeling, this study incorporated FSTs with ANN and MARS to develop forecasting models. The proposed hybrid modeling used the ARIMA as an FST to extract significant self-predictor variables in the first stage. In the second stage, the selected significant self-predictor variables are then served as the inputs for the ANN and MARS forecasting models. Finally, the performance comparison between the single stage models and the proposed hybrid models is discussed.

3. Results

3.1. Datasets and Performance Criteria

Three types of electricity sales data were provided by the Taiwan Power Company to verify the effectiveness of the proposed hybrid modeling. These three datasets covered IERECES from January 2006 to July 2016. The datasets consisted of 127 records for each type of electricity sales. Among them, 115 data records were used to develop the different forecasting models, and the remaining 12 data records were used to perform model confirmation.

In this study, we use the mean absolute percentage error (MAPE), root mean square error (RMSE), and mean absolute difference (MAD) to measure the prediction capability of the models. These prediction measurements are defined as follows [26]:

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\frac{|e_t|}{Y_t} \times 100\%, \tag{8}$$

$$\text{MAE} = \frac{1}{n}\sum_{t=1}^{n}|e_t|, \tag{9}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(e_t)^2}. \tag{10}$$

where e_t is the value of the residual at time t.

MAPE, MAD, and RMSE are measures of the deviation between actual and forecasted values. These measurements can also be used to evaluate forecasting error. The lower the MAPE, MAD, and RMSE values, the closer the forecasted values were to the actual values. Notice that the MAPE metric may be considered to be biased, being relevant only when there are no extreme values in the data sets (including zeros). In this study, since the values of electricity sales are not zeros or near-zeros, MAPE metric can provide a meaningful performance measure.

Additionally, in here, the ANN models were constructed using Qnet97 simulator, which was designed by Vesta Services Inc. (Tampa, FL, USA). For MARS modeling, the package "MARS", which was developed by Salford Systems, was employed in this study.

3.2. Forecasting Results for Industrial Electricity (IE) Sales

Figure 2 displays the original time plot for industrial electricity (IE) sales data in Taiwan. After performing ARIMA modeling by using an SAS package (SAS Institute Inc., Cary, NC, USA), this study obtained the parameter estimates reported in Table 1. Column 2 in Table 1 indicates that the estimates of the model parameters are 0.62838, −0.94744, and −0.50233. All three values are significantly different from 0 because the corresponding p-values or the "Approx Pr > |t|" values (i.e., Column 5) are all greater than or equal to the type I error, 0.05. Additionally, since the corresponding p-values for the Ljung-Box test are all greater than or equal to 0.05, the conclusion on the appropriateness of the ARIMA model is drawn that the underlying model described in Equation (11) is appropriate for modeling IE sales in Taiwan.

$$Z_t = -0.94744\,Z_{t-1} - 0.50233\,Z_{t-2} + a_t - 0.62838\,a_{t-12} \tag{11}$$

where

$$Z_t = (y_t - y_{t-1}) - (y_{t-12} - y_{t-13}).$$
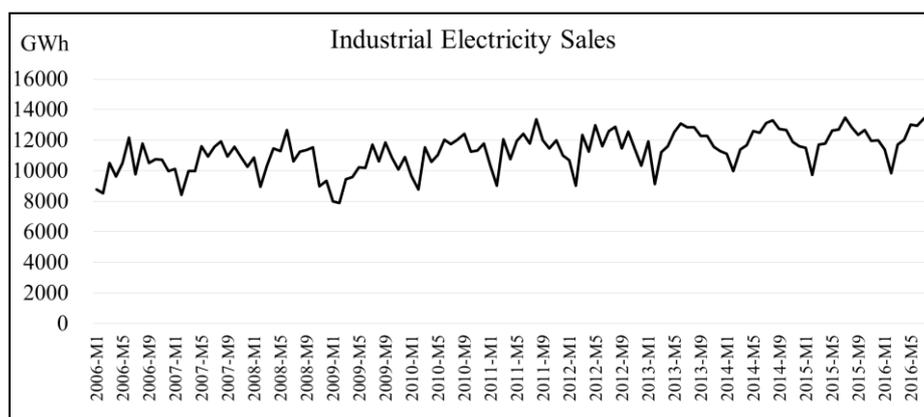
Figure 2. Time plot for industrial electricity (IE) sales in Taiwan (unit: GWh).

Table 1. Parameter estimates for industrial electricity (IE) sales.

The autoregressive Integrated Moving Average (ARIMA) Procedure Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MA1,1	0.62838	0.08064	7.79	<.0001	12
AR1,1	−0.94744	0.08769	−10.80	<.0001	1
AR1,2	−0.50233	0.08732	−5.75	<.0001	2

It has been stated that more than 75% of neural network applications use the back-propagation neural network (BPNN) structure [16]. Therefore, this study used the BPNN in designing the ANN forecasting model. For ANN designs, there is no widely accepted way to determine the number of hidden nodes in the ANN structure. Various experiments and rules of thumb have been reported to determine the number of hidden nodes. Too few hidden nodes confine the network generalization capability, whereas too many hidden nodes may lead to the problem of over-training by the network. Additionally, the term $\{n_i\text{-}n_h\text{-}n_o\}$ represents the number of neurons in the input layer, the number of neurons in the hidden layer, and the number of neurons in the output layer, respectively. The setting

values of hidden nodes were from ($2i - 2$) to ($2i + 2$), where *i* is the number of input variables. Since the learning rate of 0.01 is a very effective setting [16,45], this study used the learning rate value of 0.01 for ANN modeling. Moreover, since MAPE is one of the most important performance measurements of forecasting capability, the smallest MAPE was used as the criterion for selecting the ANN topology.

For the ANN model developed herein, two and twelve input nodes were used for the first and second designs, respectively. The hidden nodes were chosen as 2, 3, 4, 5, and 6 for the first design and 22, 23, 24, 25, and 26 for the second design. In this study, when developing the ANN approach, we used the training algorithms of Levenberg-Marquardt which provided the best forecasting accuracy.

After performing the ANN modeling process, the ANN-1 design showed that the {2-6-1} structure provided the best results and the minimum testing MAPE for IE sales. For the ANN-2 design, the best structure for IE sales was {12-25-1}. Table 2 lists the corresponding MAPE values for different settings of the ANN topologies for the two designs.

Table 2. Mean absolute percentage error (MAPE) values for the two artificial neural networks (ANN) designs (IE sales).

ANN-1 Topology	MAPE	ANN-2 Topology	MAPE
{2-2-1}	5.97	{12-22-1}	2.92
{2-3-1}	5.97	{12-23-1}	2.94
{2-4-1}	5.97	{12-24-1}	2.89
{2-5-1}	5.97	{12-25-1}	2.87
{2-6-1}	5.96	{12-26-1}	2.91

For MARS-1 modeling, the input variables and BFs should initially be selected. The variable selection results and the BFs after modeling IE sales data with MARS are summarized in Table 3. It was observed that the two input variables y_{t-1} and y_{t-2} played crucial roles in building the MARS forecasting models for forecasting IE data. Accordingly, the MARS forecasting function for IE sales is expressed as follows:

$$Y = 11313.8092 + 0.7271(BF1) - 0.3969(BF2). \tag{12}$$

Table 3. The results of MARS-1 for IE sales forecasting.

	Variable Selection Results	
Variable Name	Relative Importance (%)	Basis Functions (BFs)
y_{t-2}	100.0	$BF2 = Max(0, 12076.5 - y_{t-2})$
y_{t-1}	54.3	$BF1 = Max(0, y_{t-1} - 11261.5)$

By observing the MARS forecasting function and BF in Equation (12) and Table 3, we find that BF1 has a positive effect on y_t since the coefficient of BF1 is 0.7271. Therefore, according to the equation of BF1, it can be expected that y_t will increase when the y_{t-1} value is higher than the knot point (11261.5). In contrast, y_{t-1} will give no information about y_t if y_{t-1} stands for lower than the knot point. Additionally, Table 3 shows that y_{t-2} has a negative influence on y_t. The knot point for y_{t-2} is 12076.5.

For MARS-2 modeling, the variable selection results and BFs after modeling IE sales data with MARS are summarized in Table 4. The MARS forecasting function for the IE sales is described as follows:

$$Y = 11469.8347 - 1.4834(BF1) + 0.6711(BF2) - 2.3492(BF3) - 3.0924(BF4) + 13.8920(BF5)$$
$$- 0.3014(BF6) - 10.4358(BF7) - 0.9834(BF8). \tag{13}$$

The proposed hybrid model used the ARIMA as an FST to extract significant self-predictor variables, which served as the inputs for the ANN and MARS models. Based on Equation (11), the ARIMA model indicated that Z_{t-1}, Z_{t-2}, and a_{t-12} would influence Z_t. Since $a_{t-12} = Z_{t-12} - \hat{Z}_{t-12}$,

where $\hat{Z}_{t\text{-}12}$ is the forecast at time t-12, we observed that $Z_{t\text{-}12}$ would influence Z_t. Accordingly, we selected $Z_{t\text{-}1}$, $Z_{t\text{-}2}$, and $Z_{t\text{-}12}$ to serve as the self-predictor variables for the proposed hybrid modeling on IE data series.

The hybrid ARIMA-ANN design used three input nodes and one output node for the ANN structure. The hidden nodes were chosen as 4, 5, 6, 7, and 8 for this design. Table 5 reports the corresponding MAPE values for different settings of the ANN topologies. For the combined ARIMA-ANN design, the best structure was {3-4-1} for IE sales.

Because $Z_{t\text{-}1}$, $Z_{t\text{-}2}$, and $Z_{t\text{-}12}$ were chosen as self-predictor variables, the hybrid ARIMA-MARS design also used three input nodes and one output node for the MARS structure. The variable selection results and BFs are summarized in Table 6. The MARS forecasting function for IE sales is described as follows:

$$Y = 10446.3138 - 1.0310(BF1) + 0.4746(BF2) + 0.3069(BF3) - 0.3974(BF4), \tag{14}$$

where

$$Z_t = (y_t - y_{t\text{-}1}) - (y_{t\text{-}12} - y_{t\text{-}13}).$$

Table 4. The results of MARS-2 for IE sales forecasting.

| Variable Name | Variable Selection Results | |
	Relative Importance (%)	BF
$y_{t\text{-}1}$	100.0	BF1 = Max(0, 9750.84 − $y_{t\text{-}1}$)
$y_{t\text{-}11}$	52.7	BF8 = Max(0, 9986.92 − $y_{t\text{-}11}$)
$y_{t\text{-}2}$	44.8	BF2 = Max(0, $y_{t\text{-}2}$ − 10340)
$y_{t\text{-}7}$	44.8	BF3 = Max(0, $y_{t\text{-}7}$ − 12605)
$y_{t\text{-}10}$	37.8	BF4 = Max(0, $y_{t\text{-}10}$ − 10907.9)
		BF5 = Max(0, $y_{t\text{-}10}$ − 11429.3)
		BF6 = Max(0, 11539.2 − $y_{t\text{-}10}$)
		BF7 = Max(0, $y_{t\text{-}10}$ − 11539.2)

Table 5. MAPE values for proposed ARIMA-ANN design (IE sales).

ARIMA-ANN topology	MAPE
{3-4-1}	2.21
{3-5-1}	2.37
{3-6-1}	2.30
{3-7-1}	2.32
{3-8-1}	2.37

Table 6. Results of proposed ARIMA-MARS for IE sales forecasting.

| Variable Name | Variable Selection Results | |
	Relative Importance (%)	BF
$Z_{t\text{-}1}$	100.0	BF1 = Max(0, 9596.3 − $Z_{t\text{-}1}$)
-	-	BF2 = Max(0, $Z_{t\text{-}1}$ − 9596.3)
$Z_{t\text{-}12}$	48.2	BF4 = Max(0, 10888.2 − $Z_{t\text{-}12}$)
$Z_{t\text{-}2}$	24.3	BF3 = Max(0, $Z_{t\text{-}2}$ − 10234.6)

3.3. Forecasting Results for Residential Electricity (RE) Sales

Figure 3 presents the original time plot for RE sales data in Taiwan. The parameter estimates after performing the ARIMA modeling are listed in Table 7. Column 2 in Table 7 indicates that the estimates of the model parameters are significantly different from 0. Additionally, since the corresponding p-values for the Ljung-Box test are all greater than or equal to 0.05, the conclusion on

the appropriateness of the ARIMA model is drawn that Equation (15) is suitable for modeling the RE sales in Taiwan.

$$Z_t = 25.10872 + 0.5022\,Z_{t-1} - 0.27353\,Z_{t-4} + a_t - 0.77181\,a_{t-12}, \qquad (15)$$

where

$$Z_t = y_{t-12} - y_{t-13}.$$

This study also used two ANN designs for modeling RE sales, as was the case for the structures for modeling IE sales. The first design used y_{t-1} and y_{t-2} as the input variables, and the second design used y_{t-1}, y_{t-2}, y_{t-3}, ... and y_{t-12} as the input variables. Table 8 presents the corresponding MAPE values for different settings of ANN topologies for the two designs after performing ANN modeling. Observing Table 8, we notice that the ANN-1 design with {2-2-1} structure had the smallest MAPE for RE sales. For the ANN-2 design, the {12-25-1} structure was associated with the smallest MAPE for RE sales.
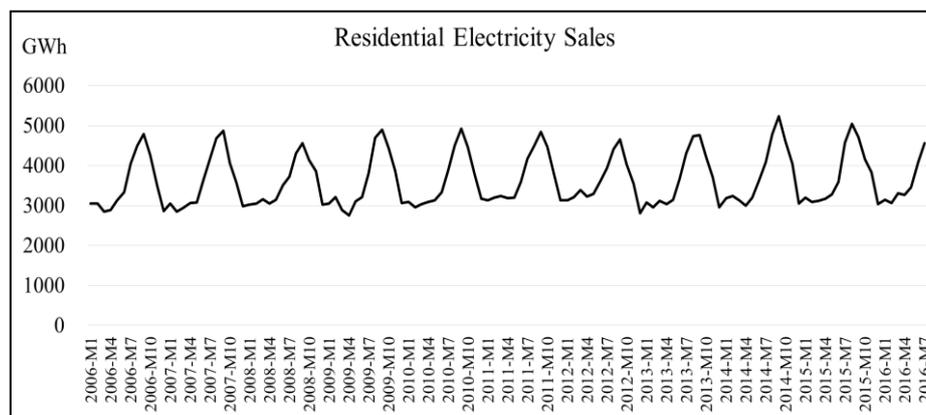

Figure 3. Time plot for residential electricity (RE) sales in Taiwan (unit: GWh).

Table 7. Parameter estimates for residential electricity (RE) sales.

	The ARIMA Procedure				
	Conditional Least Squares Estimation				
Parameter	Estimate	Standard Error	*t* Value	Approx Pr > |t|	Lag
MU	25.10872	6.08970	4.12	<.0001	0
MA1,1	0.77181	0.07085	4.12	<.0001	12
AR1,1	0.50220	0.08919	5.63	<.0001	1
AR1,2	−0.27353	0.08923	−3.07	0.0028	4

Table 8. MAPE values for the two ANN designs (RE sales).

ANN-1 Topology	MAPE	ANN-2 Topology	MAPE
{2-2-1}	6.83	{12-22-1}	4.62
{2-3-1}	6.86	{12-23-1}	4.58
{2-4-1}	6.94	{12-24-1}	4.56
{2-5-1}	6.91	{12-25-1}	4.54
{2-6-1}	6.94	{12-26-1}	4.58

For modeling RE sales with MARS, this study used the same two designs. For MARS-1 modeling, the variable selection results and BFs are summarized in Table 9. The MARS forecasting function for the RE sales is expressed as follows:

$$Y = 5578.3334 - 1.3281(BF1) + 1.0042(BF2) - 1.4247(BF3). \tag{16}$$

For MARS-2 modeling, the variable selection results and the BFs are summarized in Table 10. Moreover, the MARS forecasting function for the RE sales is described as follows:

$$\begin{aligned} Y = {}& 3846.2286 + 0.6741(BF1) - 0.7646(BF2) - 0.3623(BF3) + 0.3844(BF4) + 0.1324(BF5) \\ & + 0.2665(BF6) - 0.6067(BF7) + 0.6001(BF8) - 0.5913(BF9) + 0.7739(BF10). \end{aligned} \tag{17}$$

Table 9. The results of MARS-1 for RE sales forecasting.

| | Variable Selection Results | |
Variable Name	Relative Importance (%)	BF
y_{t-2}	100.0	BF3 = Max(0, 4733.92 − y_{t-2})
y_{t-1}	53.2	BF1 = Max(0, y_{t-1} − 3556.2)
-	-	BF2 = Max(0, y_{t-1} − 4411.95)

Table 10. The results of MARS-2 for RE sales forecasting.

| | Variable Selection Results | |
Variable Name	Relative Importance (%)	BF
y_{t-1}	100.0	
y_{t-2}	21.7	BF1 = Max(0, y_{t-2} − 3509.42)
-	-	BF2 = Max(0, y_{t-2} − 4558.55)
y_{t-10}	21.7	BF7 = Max(0, y_{t-10} − 4268.73)
y_{t-12}	21.7	BF9 = Max(0, 4471.3 − y_{t-12})
-	-	BF10 = Max(0, y_{t-12} − 4471.3)
y_{t-3}	14.1	BF3 = Max(0, 4141.63 − y_{t-3})
y_{t-4}	14.1	BF4 = Max(0, 4153.84 − y_{t-4})
y_{t-11}	9.6	BF8 = Max(0, y_{t-11} − 4616.91)
y_{t-9}	7.7	BF6 = Max(0, 3556.2 − y_{t-9})
y_{t-7}	6.5	BF5 = Max(0, y_{t-7} − 3589.08)

For the hybrid modeling for RE sales forecasting, based on Equation (15), ARIMA modeling selected Z_{t-1}, Z_{t-4}, and Z_{t-12} to serve as the self-predictor variables. Therefore, the hybrid ARIMA-ANN modeling used three input nodes and one output node for the ANN structure. The hidden nodes were chosen as 4, 5, 6, 7, and 8 for this design. Table 11 lists the corresponding MAPE values for different settings of the ANN topologies. For hybrid ARIMA-ANN modeling, the best structure for RE sales was {3-5-1}.

Additionally, the hybrid ARIMA-MARS design used three input nodes and one output node for the MARS structure. The variable selection results and BFs are summarized in Table 12. The MARS forecasting function for the IE sales is described as follows:

$$Y = 3248.8332 + 0.9043(BF1) + 0.1010(BF2) - 0.1605(BF3) + 0.3992(BF4), \tag{18}$$

where

$$Z_t = y_{t-12} - y_{t-13}.$$

Table 11. MAPE values for proposed ARIMA-ANN design (RE sales).

ARIMA-ANN Topology	MAPE
{3-4-1}	4.58
{3-5-1}	4.51
{3-6-1}	4.57
{3-7-1}	4.56
{3-8-1}	4.58

Table 12. Results of proposed ARIMA-MARS for RE sales forecasting.

| Variable Name | Variable Selection Results | |
	Relative Importance (%)	BF
Z_{t-1}	100.0	BF1 = Max(0, Z_{t-1} − 3131.86)
Z_{t-12}	12.2	BF2 = Max(0, Z_{t-4} − 3126.91)
Z_{t-4}	9.4	BF3 = Max(0, 4483.91 − Z_{t-12})
-	-	BF4 = Max(0, Z_{t-12} − 4483.91)

3.4. Forecasting Results for Commercial Electricity (CE) Sales

The original time plot for CE sales data in Taiwan is displayed in Figure 4. Table 13 presents the parameter estimates after performing ARIMA modeling. Column 2 in Table 13 indicates that the estimates of the model parameters are significantly different from 0. Additionally, since the corresponding p-values for the Ljung-Box test are all greater than or equal to 0.05, the conclusion on the appropriateness of the ARIMA model is drawn that Equation (19) is suitable for modeling CE sales in Taiwan.

$$Z_t = -0.68597 Z_{t-1} - 0.46463 Z_{t-2} + a_t - 0.73818\, a_{t-12}, \tag{19}$$

where

$$Z_t = (y_t - y_{t-1}) - (y_{t-12} - y_{t-13}).$$
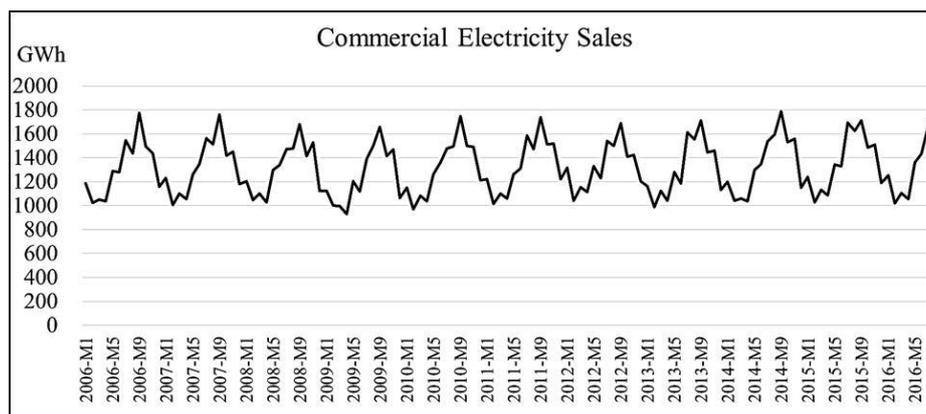
Figure 4. Time plot for commercial electricity (CE) sales in Taiwan (unit: GWh).

The same two designs for modeling IE and RE sales were applied to the ANN modeling of CE sales. The ANN used three and 12 input nodes for the first and second designs, respectively. After performing ANN modeling, Table 14 shows the corresponding MAPE values for different settings of the ANN topologies for the two designs. Based on Table 14, we observed that the ANN-1 design with {2-2-1} structure and ANN-2 design with {12-24-1} structure had the smallest MAPE for CE sales.

Table 13. Parameter estimates for commercial electricity (CE) sales.

| | The ARIMA Procedure | | | | |
| | Conditional Least Squares Estimation | | | | |
Parameter	Estimate	Standard Error	*t* Value	Approx Pr > |t|	Lag
MA1,1	0.73818	0.07617	9.69	<.0001	12
AR1,1	−0.68591	0.09043	−7.58	<.0001	1
AR1,2	−0.46463	0.09121	−5.09	<.0001	2

Table 14. MAPE values for the two ANN designs (CE sales).

ANN-1 Topology	MAPE	ANN-2 Topology	MAPE
{2-2-1}	10.67	{12-22-1}	2.65
{2-3-1}	10.71	{12-23-1}	2.62
{2-4-1}	10.72	{12-24-1}	2.60
{2-5-1}	10.74	{12-25-1}	2.64
{2-6-1}	10.73	{12-26-1}	2.58

In the modeling of MARS for CE sales data, this study also used the same two designs. For MARS-1 modeling, Table 15 reports the variable selection results and BFs. The MARS forecasting function for the CE sales is described as follows:

$$Y = 1113.43356 + 4.26681(BF1) - 4.70790(BF2) - 15.42855(BF3) + 23.48366(BF4) \\ - 17.49202(BF5) + 15.57251(BF6) - 6.48430(BF7) + 1.35723(BF8) - 1.11383(BF9). \tag{20}$$

For MARS-2 modeling, the variable selection results and BFs are reported in Table 16. The MARS forecasting function for the CE sales is described as follows:

$$Y = 1197.44280 + 0.53552(BF1) + 0.36830(BF2) - 1.09315(BF3) + 0.43751(BF4) \\ + 0.62503(BF5) - 0.37300(BF6) - 1.20468(BF7) - 5.55634(BF8) + 8.50438(BF9) - 2.85861(BF10) \\ - 1.00710(BF11) + 0.32421(BF12) + 0.60439(BF13). \tag{21}$$

For the hybrid modeling for CE sales forecasting, we considered Z_{t-1}, Z_{t-2}, and Z_{t-12} as the self-predictor variables, according to the ARIMA model shown in Equation (19). The hybrid ARIMA-ANN modeling used three input nodes and one output node for the ANN structure. The hidden nodes were chosen as 4, 5, 6, 7, and 8 for this design. Table 17 lists the corresponding MAPE values for different settings of the ANN topologies. For hybrid ARIMA-ANN modeling, the best structure was {3-5-1} for CE sales.

The hybrid ARIMA-MARS design used three input nodes and one output node for the MARS structure. The variable selection results and BFs are listed in Table 18. The MARS forecasting function for the CE sales is described as follows:

$$Y = 1291.98446 - 0.89363(BF1) + 1.02147(BF2), \tag{22}$$

where

$$Z_t = (y_t - y_{t-1}) - (y_{t-12} - y_{t-13})$$

Table 15. The results of MARS-1 for CE sales forecasting.

| | Variable Selection Results | |
Variable Name	Relative Importance (%)	BF
y_{t-2}	100.0	BF8 = Max(0, y_{t-2} − 1154.09)
-	-	BF9 = Max(0, y_{t-2} − 1287.99)
y_{t-1}	58.4	BF1 = Max(0, y_{t-1} − 1222.61)

Table 15. *Cont.*

| | Variable Selection Results | |
Variable Name	Relative Importance (%)	BF
-	-	BF2 = Max(0, y_{t-1} − 1279.22)
-	-	BF3 = Max(0, y_{t-1} − 1393.37)
-	-	BF4 = Max(0, y_{t-1} − 1418.06)
-	-	BF5 = Max(0, y_{t-1} − 1471.81)
-	-	BF6 = Max(0, y_{t-1} − 1499.07)
-	-	BF7 = Max(0, y_{t-1} − 1560.28)

Table 16. The results of MARS-2 for CE sales forecasting.

| | Variable Selection Results | |
Variable Name	Relative Importance (%)	BF
y_{t-1}	100.0	BF1 = Max(0, y_{t-1} − 1313.66)
y_{t-8}	25.4	BF5 = Max(0, 1261.19 − y_{t-8})
-	-	BF6 = Max(0, y_{t-8} − 1261.19)
y_{t-11}	24.2	BF11 = Max(0, 1099.97 − y_{t-11})
-	-	BF12 = Max(0, y_{t-11} − 1099.97)
y_{t-4}	21.1	BF2 = Max(0, y_{t-4} − 1346.29)
y_{t-12}	18.1	BF13 = Max(0, y_{t-12} − 1459.56)
y_{t-5}	15.4	BF3 = Max(0, 1102.27 − y_{t-5})
y_{t-10}	12.3	BF7 = Max(0, 1040.75 − y_{t-10})
-	-	BF8 = Max(0, y_{t-10} − 1099.97)
-	-	BF9 = Max(0, y_{t-10} − 1122.64)
-	-	BF10 = Max(0, y_{t-10} − 1148.43)
y_{t-6}	4.5	BF4 = Max(0, 1102.27 − y_{t-6})

Table 17. MAPE values for proposed ARIMA-ANN design (CE sales).

ARIMA-ANN Topology	MAPE
{3-4-1}	2.58
{3-5-1}	2.57
{3-6-1}	2.64
{3-7-1}	2.63
{3-8-1}	2.61

Table 18. Results of proposed ARIMA-MARS for CE sales forecasting.

| | Variable Selection Results | |
Variable Name	Relative Importance (%)	BF
Z_{t-1}	100.0	BF1 = Max(0, 1296.28 − Z_{t-1})
-	-	BF2 = Max(0, Z_{t-1} − 1296.28)

4. Discussion

This study developed various forecasting models to predict three types of electricity sales in Taiwan. Table 19 shows the forecasting results as well as the MAPE, RMSE, and MAD values of the forecasting models. Low MAPE, RMSE, or MAD values are associated with better forecasting accuracy.

In comparison to the forecasting performance of single models in Table 19, we observed that the second design exhibited better performance than the first design of the ANN and MARS modeling techniques. The possible reason may be that the second design has more input information than the first design. The results of IERECE sales forecasting also indicate that the second ANN design has

better forecasting accuracy than the second MARS design. However, the performance of MARS seems to have been better than the ANN approach for the first design.

As shown in Table 19, the proposed hybrid models have better forecasting performance than the single models. For example, the proposed hybrid ARIMA-ANN model was associated with MAPE values of 2.21%, 4.51%, and 2.57% for IE, RE, and CE sales forecasting, respectively; these three MAPE values were also the smallest values. Accordingly, our proposed hybrid models provided more accurate results than the single models. For IE sales forecasting, the MAPE percentage improvements (denoted by MAPE$_{PI}$) of the proposed ARIMA-ANN model over the two single models ANN-1 and ANN-2 were 169.68% and 29.86%, respectively. The MAPE$_{PI}$ of ARIMA-ANN over ANN was computed as

$$\text{MAPE}_{PI} = \frac{(\text{MAPE}_{ANN} - \text{MAPE}_{ARIMA-ANN})}{\text{MAPE}_{ARIMA-ANN}} \times 100\% \qquad (23)$$

For IE sales forecasting, the MAPE$_{PI}$ of the proposed ARIMA-MARS model over the two single models MARS-1 and MARS-2 were 171.73% and 133.33%, respectively. The MAD and RMSE percentage improvements can be calculated based on Equation (23), and are denoted as MAD$_{PI}$ and RMSE$_{PI}$, respectively. Table 20 provides a complete comparison of overall improvement percentage. We observed three negative values in Table 20. For RE sales forecasting, the MAD and RMSE percentage improvements of the ARIMA-ANN over ANN-2 were −3.35% and −17.12%, respectively. For RE sales forecasting, the RMSE percentage improvement of ARIMA-MARS over ANN-1 was −9.83%. A negative percentage improvement means that forecasting performance declined. Nonetheless, the magnitudes of those three negative values are small and only have minor effects on forecasting performance. Additionally, the associated MAPE percentage improvements for those negative values were all positive. That is, if a forecasting model has the minimum MAPE, this does not mean that it has the minimum MAD or RMSE. As shown in Tables 19 and 20, the proposed hybrid models outperform the single models.

Additionally, Figure 5 displays the forecasts of IE sales obtained by using ANN-1, ANN-2, and the proposed ARIMA-ANN models for the last twelve testing records. We notice that the forecasts of the ARIMA-ANN model are closer to the actual observations, whereas the forecasts of the ANN-1 model seem far away from the actual observations. Figure 6 shows the forecasts of IE sales obtained by using MARS-1, MARS-2, and the proposed ARIMA-MARS models for the last twelve testing records. Again, we can observe that the forecasts of the ARIMA-MARS model are closer to the actual observations. Similar results were found for the cases of RE and CE sales. That is, the forecasts of the proposed hybrid models are closer to the actual observations. These findings can be observed in Figures 7–10.

Table 19. Performance comparison of hybrid and single models.

	MAPE (%)	MAD	RMSE
IE sales forecasting			
Single models			
ANN-1	5.96	707.34	874.87
ANN-2	2.87	338.44	440.09
MARS-1	6.44	764.60	907.59
MARS-2	5.53	635.57	854.75
Proposed hybrid models			
ARIMA-ANN	2.21	259.27	336.21
ARIMA-MARS	2.37	271.01	392.70
RE sales forecasting			
Single models			
ANN-1	6.83	241.81	272.50
ANN-2	4.54	175.15	199.51
MARS-1	6.49	234.64	266.71
MARS-2	6.10	243.69	318.48

Table 19. *Cont.*

	MAPE (%)	MAD	RMSE
Proposed hybrid models			
ARIMA-ANN	4.51	181.22	240.73
ARIMA-MARS	5.23	212.30	295.79
CE sales forecasting			
Single models			
ANN-1	10.67	141.49	168.45
ANN-2	2.58	35.10	47.46
MARS-1	10.25	131.67	152.97
MARS-2	3.78	53.02	73.22
Proposed hybrid models			
ARIMA-ANN	2.57	34.24	41.47
ARIMA-MARS	3.27	44.53	51.79

Note: RMSE: root mean square error.

5. Conclusions

For forecasting, typical soft computing approaches need proper explanatory variables to make predictions. However, proper explanatory variables are sometimes difficult to capture, and it is not feasible to obtain the future values of these variables.

Electricity sales forecasting is a challenging task and has drawn considerable attention over the past decades. Consequently, this study investigated hybrid ARIMA-ANN and ARIMA-MARS models for forecasting three types of electricity sales in Taiwan. ARIMA modeling was used as the FST and the significant self-predictor variables were selected by ARIMA. Instead of using unavailable explanatory variables, the self-predictor variables were considered as the inputs for ANN and MARS modeling. The proposed ARIMA-ANN and ARIMA-MARS models were compared with single ANN-1, ANN-2, MARS-1, and MARS-2 models using MAPE, MAD, and RMSE values as performance criteria. The experimental results show that the proposed hybrid ARIMA-ANN and ARIMA-MARS models are good alternatives for electricity sales forecasting since both have good forecasting performance.

Table 20. Improvement of the proposed models in comparison with single models.

	MAPE (%)	MAD (%)	RMSE (%)
IE sales forecasting			
Proposed ARIMA-ANN			
ANN-1	169.68	172.82	160.22
ANN-2	29.86	30.54	30.90
Proposed ARIMA-MARS			
MARS-1	171.73	182.13	131.12
MARS-2	133.33	134.52	117.66
RE sales forecasting			
Proposed ARIMA-ANN			
ANN-1	51.44	33.43	13.20
ANN-2	0.67	-3.35	-17.12
Proposed ARIMA-MARS			
MARS-1	24.09	10.52	-9.83
MARS-2	16.63	14.79	7.67
CE sales forecasting			
Proposed ARIMA-ANN			
ANN-1	315.18	313.23	306.20
ANN-2	0.39	2.51	14.44
Proposed ARIMA-MARS			
MARS-1	213.46	195.69	195.37
MARS-2	15.60	19.07	41.38

Because it is difficult to fully capture the characteristics of real energy data, hybrid modeling should be an acceptable and practical modeling approach. Most importantly, how can soft computing models perform prediction if the proper explanatory variables are not available? Accordingly, the primary contribution of the proposed hybrid models is their ability to provide predictions of electricity sales without requiring extensive effort to obtain the future values of explanatory variables.

Multistep forecasts were considered in this study. The proposed models could handle several months-ahead or one year-ahead forecasts. The modeling process in this study can serve to guide the development of models for electricity sales forecasting in other countries. Although the dataset was collected from January 2006 to July 2016, we believe that the hybrid models will continue to outperform the single-stage models, no matter when our approach will be put in a real production environment.
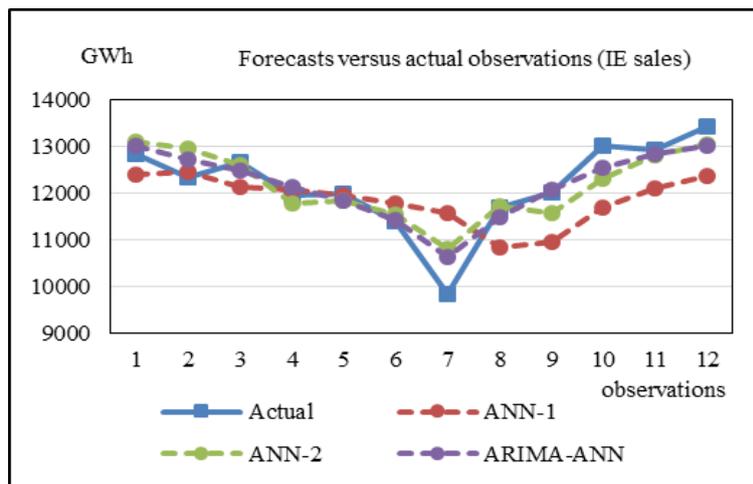

Figure 5. Forecasts of IE sales versus actual observations for the 12 testing records (with the use of ANN-1, ANN-2, and the proposed ARIMA-ANN models).
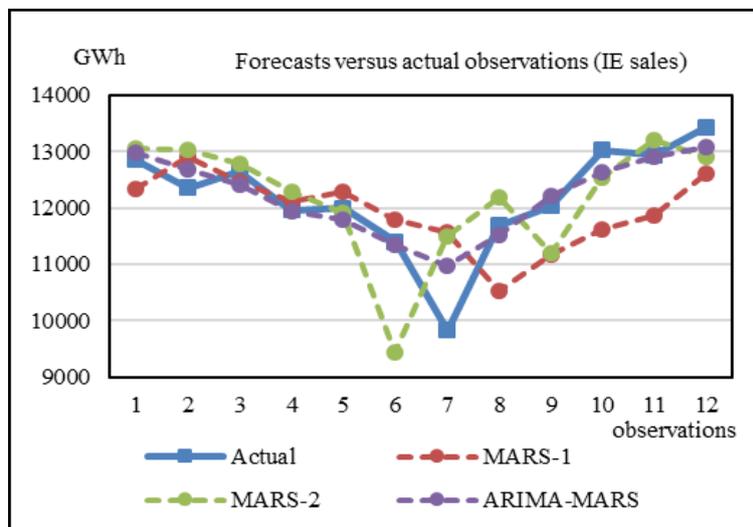

Figure 6. Forecasts of IE sales versus actual observations for the 12 testing records (with the use of MARS-1, MARS-2, and the proposed ARIMA-MARS models).

In this study, the case of d = 0 in ARIMA cay be interpreted as the lack of a trend, which may be because of the economics of an already industrialized country. It is interesting to know if d = 0 still holds for a developing country with a sharp upward trend; this issue needs further investigation by applying the ARIMA technique to electricity sales data for other countries. Additionally, we used 12 inputs (i.e.,

from y_{t-1} to y_{t-12}) for the second ANN and MARS designs in this study. Because considerable inputs may lead to overfitting problems, we did not include more past seasonal components in our design. However, it may be a good practice to use a few of the past seasonal components (i.e., y_{t-24}, y_{t-36}, or y_{t-48}) in future ANN or MARS designs. Also, because soft computing modeling is more powerful in capturing nonlinear relationships between input and output, the future studies may use the White test [46] or Teräsvirta test [47] to see whether a linearity is rejected or not before attempting nonlinear modelling. Finally, the combination of ARIMA and other forecasting techniques, such as support vector regression or extreme learning machine, to improve forecasting performance can be further investigated in future studies.
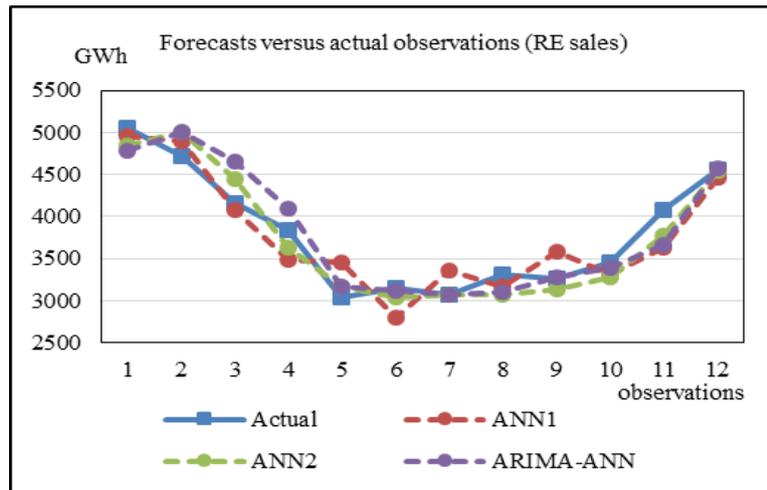
Figure 7. Forecasts of RE sales versus actual observations for the 12 testing records (with the use of ANN-1, ANN-2, and the proposed ARIMA-ANN models).
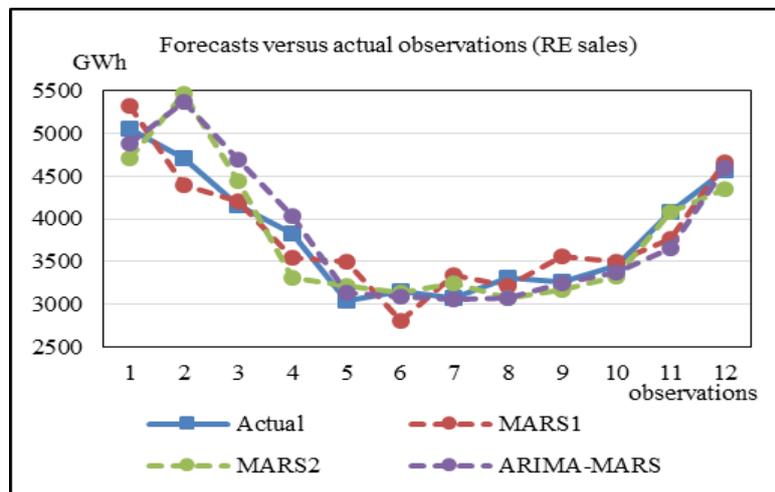
Figure 8. Forecasts of RE sales versus actual observations for the 12 testing records (with the use of MARS-1, MARS-2, and the proposed ARIMA-MARS models).
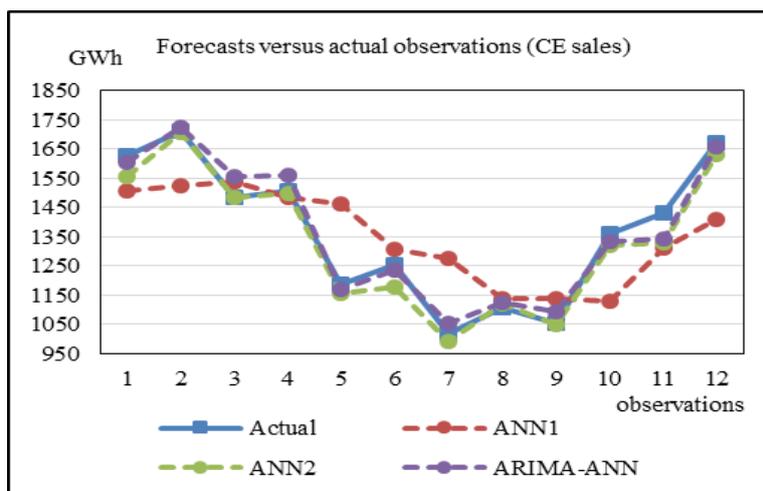
Figure 9. Forecasts of CE sales versus actual observations for the 12 testing records (with the use of ANN-1, ANN-2, and the proposed ARIMA-ANN models).
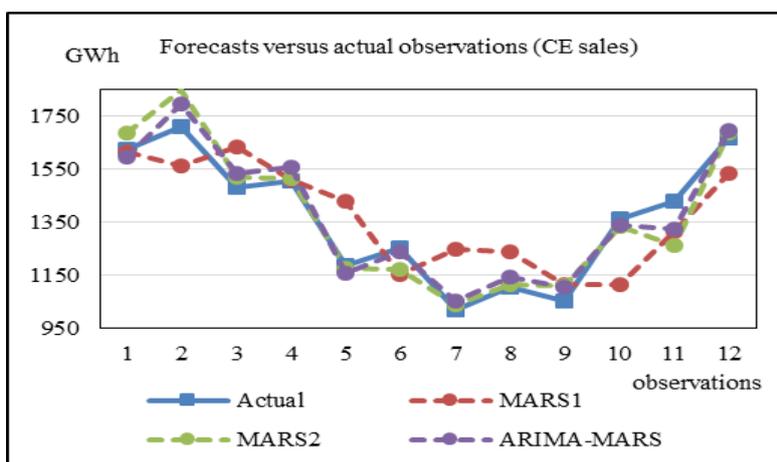
Figure 10. Forecasts of CE sales versus actual observations for the 12 testing records (with the use of MARS-1, MARS-2, and the proposed ARIMA-MARS models).

References

1. Ranjan, M.; Jain, V.K. Modelling of electrical energy consumption in Delhi. *Energy* **1999**, *24*, 351–361. [CrossRef]
2. Holtedahl, P.; Joutz, F.L. Residential electricity demand in Taiwan. *Energy Econ.* **2004**, *26*, 201–224. [CrossRef]
3. Bianco, V.; Manca, O.; Nardini, S. Electricity consumption forecasting in Italy using linear regression models. *Energy* **2009**, *34*, 1413–1421. [CrossRef]
4. Mohamed, Z.; Bodger, P. Forecasting electricity consumption in New Zealand using economic and demographic variables. *Energy* **2005**, *30*, 1833–1843. [CrossRef]

5. Bianco, V.; Manca, O.; Nardini, S.; Minea, A.A. Analysis and forecasting of nonresidential electricity consumption in Romania. *Appl. Energy* **2010**, *87*, 3584–3590. [CrossRef]

6. Pappas, S.S.; Ekonomou, L.; Karamousantas, D.C.; Chatzarakis, G.E.; Katsikas, S.K.; Liatsis, P. Electricity demand loads modeling using auto regressive moving average (ARMA) models. *Energy* **2008**, *33*, 1353–1360. [CrossRef]

7. Sumer, K.K.; Goktas, O.; Hepsag, A. The application of seasonal latent variable in forecasting electricity demand as an alternative method. *Energy Policy* **2009**, *37*, 1317–1322. [CrossRef]

8. Conejo, A.J.; Plazas, M.A.; Espinola, R.; Molina, A.B. Day-ahead electricity price forecasting using the wavelet transform and ARIMA models. *IEEE T. Power Syst.* **2005**, *20*, 1035–1042. [CrossRef]

9. Saab, S.; Badr, E.; Nasr, G. Univariate modeling and forecasting of energy consumption: the case of electricity in Lebanon. *Energy* **2001**, *26*, 1–14. [CrossRef]

10. Pao, H.T. Forecasting energy consumption in Taiwan using hybrid nonlinear models. *Energy* **2009**, *34*, 1438–1446. [CrossRef]

11. Pai, P.F.; Lin, C.S. A hybrid ARIMA and support vector machines model in stock price forecasting. *OMEGA* **2005**, *33*, 497–505. [CrossRef]

12. Khashei, M.; Bijari, M. An artificial neural network (p, d, q) model for time series forecasting. *Expert Syst. Appl.* **2010**, *37*, 479–489. [CrossRef]

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

14. Barak, S.; Sadegh, S.S. Forecasting energy consumption using ensemble ARIMA–ANFIS hybrid algorithm. *Int. J. Electr. Power Energy Syst.* **2016**, *82*, 92–104. [CrossRef]

15. Abedinia, O.; Amjady, N.; Shafie-khah, M.; Catalão, J.P.S. Electricity price forecast using combinatorial neural network trained by a new stochastic search method. *Energy Convers. Manag.* **2015**, *105*, 642–654. [CrossRef]

16. Shao, Y.E.; Chiu, C.C. Applying emerging soft computing approaches to control chart pattern recognition for an SPC-EPC process. *Neurocomputing* **2016**, *201*, 19–28. [CrossRef]

17. Cheng, M.Y.; Cao, M.T. Accurately predicting building energy performance using evolutionary multivariate adaptive regression splines. *Appl. Soft Comput.* **2014**, *22*, 178–188. [CrossRef]

18. Li, Y.; He, Y.; Su, Y.; Shu, L. Forecasting the daily power output of a grid-connected photovoltaic system based on multivariate adaptive regression splines. *Appl. Energy* **2016**, *180*, 392–401. [CrossRef]

19. Kavaklioglu, K.; Ceylan, H.; Ozturk, H.K.; Canyurt, O.E. Modeling and prediction of Turkey's electricity consumption using artificial neural networks. *Energy Convers. Manag.* **2009**, *50*, 2719–2727. [CrossRef]

20. Sandhu, H.S.; Fang, L.; Guan, L. Forecasting day-ahead price spikes for the Ontario electricity market. *Electr. Power Syst. Res.* **2016**, *141*, 450–459. [CrossRef]

21. Tso, G.K.F.; Yau, K.K.W. Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. *Energy* **2007**, *32*, 1761–1768. [CrossRef]

22. Kaytez, F.; Taplamacioglu, M.C.; Cam, E.; Hardalac, F. Forecasting electricity consumption: A comparison of regression analysis, neural networks and least squares support vector machines. *Int. J. Electr. Power Energy Syst.* **2015**, *67*, 431–438. [CrossRef]

23. Kankal, M.; Akpınar, A.; Kömürcü, M.I.; Özşahin, T.S. Modeling and forecasting of Turkey's energy consumption using socio-economic and demographic variables. *Appl. Energy* **2011**, *88*, 1927–1939. [CrossRef]

24. Bilgili, M.; Sahin, B.; Yasar, A.; Simsek, E. Electric energy demands of Turkey in residential and industrial sectors. *Renew. Sustain. Energy Rev.* **2012**, *16*, 404–414. [CrossRef]

25. Chou, S.M.; Lee, T.S.; Shao, Y.E.; Chen, I.F. Mining the breast cancer pattern using artificial neural networks and multivariate adaptive regression splines. *Expert Syst. Appl.* **2004**, *27*, 133–142. [CrossRef]

26. Dai, W.; Shao, Y.E.; Lu, C.J. Incorporating feature selection method into support vector regression for stock index forecasting. *Neural Comput. Appl.* **2013**, *23*, 1551–1561. [CrossRef]

27. Wang, D.; Luo, H.; Grunder, O.; Lin, Y.; Guo, H. Multi-step ahead electricity price forecasting using a hybrid model based on two-layer decomposition technique and BP neural network optimized by firefly algorithm. *Appl. Energy* **2017**, *190*, 390–407. [CrossRef]

28. Li, W.; Yang, X.; Li, H.; Su, L. Hybrid forecasting approach based on GRNN neural network and SVR Machine for electricity demand forecasting. *Energies* **2017**, *10*, 44. [CrossRef]

29. Voronin, S.; Partanen, J. Forecasting electricity price and demand using a hybrid approach based on wavelet transform, ARIMA and neural networks. *Int. J. Energy Res.* **2014**, *38*, 626–637. [CrossRef]

30. Yang, Z.; Ce, L.; Lian, L. Electricity price forecasting by a hybrid model, combining wavelet transform, ARMA and kernel-based extreme learning machine methods. *Appl. Energy* **2017**, *190*, 291–305. [CrossRef]
31. Yang, Y.; Chen, Y.; Wang, Y.; Li, C.; Li, L. Modelling a combined method based on ANFIS and neural network improved by DE algorithm: A case study for short-term electricity demand forecasting. *Appl. Soft Comput.* **2016**, *49*, 663–675. [CrossRef]
32. Yan, X.; Chowdhury, N.A. Mid-term electricity market clearing price forecasting utilizing hybrid support vector machine and auto-regressive moving average with external input. *Int. J. Electr. Power* **2014**, *63*, 64–70. [CrossRef]
33. Pórtoles, J.; González, C.; Moguerza, J.M. Electricity price forecasting with dynamic trees: A benchmark against the random forest approach. *Energies* **2018**, *11*, 1588. [CrossRef]
34. Co, H.C.; Boosarawongse, R. Forecasting Thailand's rice export: statistical techniques vs. artificial neural networks. *Comput. Ind. Eng.* **2007**, *53*, 610–627. [CrossRef]
35. Zou, H.F.; Xia, G.P.; Yang, F.T.; Wang, H.Y. An investigation and comparison of artificial neural network and time series models for Chinese food grain price forecasting. *Neurocomputing* **2007**, *70*, 2913–2923. [CrossRef]
36. Amini, M.H.; Kargarian, A.; Karabasoglu, O. ARIMA-based decoupled time series forecasting of electric vehicle charging demand for stochastic power system operation. *Electr. Power Syst. Res.* **2016**, *140*, 378–390. [CrossRef]
37. Bahrami, S.; Wong, V.W.; Huang, J. An online learning algorithm for demand response in smart grid. *IEEE Trans. Smart Grid* **2017**, *16*, 2983–2999. [CrossRef]
38. Suganthi, L.; Samuel, A.A. Energy models for demand forecasting-a review. *Renew. Sustain. Energy Rev.* **2012**. [CrossRef]
39. Shao, Y.E.; Hou, C.D. Change point determination for a multivariate process using a two-stage hybrid scheme. *Appl. Soft Comput.* **2013**, *13*, 1520–1527. [CrossRef]
40. Shao, Y.E.; Hou, C.D.; Chiu, C.C. Hybrid intelligent modeling schemes for heart disease classification. *Appl. Soft Comput.* **2014**, *14*, 47–52. [CrossRef]
41. Box, G.E.P.; Jenkins, G.M.; Reisel, G.C. *Time Series Analysis: Forecasting and Control*, 4th ed.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2008; pp. 95–100. ISBN 9780470272848.
42. Ljung, G.M.; Box, G.E.P. On a measure of lack of fit in time series models. *Biometrika* **1978**, *65*, 297–303. [CrossRef]
43. Cheng, B.; Titterington, D.M. Neural networks: a review from a statistical perspective. *Stat. Sci.* **1994**, *9*, 2–30. [CrossRef]
44. Friedman, J.H. Multivariate adaptive regression splines. *Ann. Stat.* **1991**, *19*, 1–67. [CrossRef]
45. Shao, Y.E.; Chang, P.Y.; Lu, C.J. Applying two-stage neural network based classifiers to the identification of mixture control chart patterns for an SPC-EPC process. *Complexity* **2017**, *2017*, 1–10. [CrossRef]
46. Teräsvirta, T.; Lin, C.F.; Granger, C.W.J. Power of the neural network linearity test. *J. Time Ser. Anal.* **1993**, *14*, 209–220. [CrossRef]
47. Lee, T.H.; White, H.; Granger, C.W.J. Testing for neglected nonlinearity in time series models: A comparison of neural network methods and alternative tests. *J. Econ.* **1993**, *56*, 269–290. [CrossRef]