Article

Stacked Auto-Encoder Modeling of an Ultra-Supercritical Boiler-Turbine System

Hao Zhang [1], Xiangjie Liu [1], Xiaobing Kong [1],* and Kwang Y. Lee [2]

[1] The State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources, North China Electric Power University, Beijing 102206, China; zhanghaooa@163.com (H.Z.); liuxj@ncepu.edu.cn (X.L.)

[2] The Department of Electrical and Computer Engineering, Baylor University, Waco, TX 76798-7356, USA; Kwang_Y_Lee@baylor.edu

* Correspondence: kongxiaobing@ncepu.edu.cn; Tel.: +86-10-6177-2103

check for updates

Abstract: The ultra-supercritical (USC) coal-fired boiler-turbine unit has been widely used in modern power plants due to its high efficiency and low emissions. Since it is a typical multivariable system with large inertia, severe nonlinearity, and strong coupling, building an accurate model of the system using traditional identification methods are almost impossible. In this paper, a deep neural network framework using stacked auto-encoders (SAEs) is presented as an effective way to model the USC unit. In the training process of SAE, maximum correntropy is chosen as the loss function, since it can effectively alleviate the influence of the outliers existing in USC unit data. The SAE model is trained and validated using the real-time measurement data generated in the USC unit, and then compared with the traditional multilayer perceptron network. The results show that SAE has superiority both in forecasting the dynamic behavior as well as eliminating the influence of outliers. Therefore, it can be applicable for the simulation analysis of a 1000 MW USC unit.

Keywords: ultra-supercritical unit; deep neural network; stacked auto-encoder; maximum correntropy

---

1. Introduction

With the fast development of China's economy in the 21st century, the demand for electricity is growing rapidly. Although the installed capacity of renewable energy, such as wind power and solar power, have increased in recent years, coal-fired power generation still accounts for a large proportion of the power generation. In China, the coal-fired installed capacity reached 921.2 GW by the end of 2017, accounting for nearly 72% of the total electricity generation [1]. In the process of coal burning, many air pollutants may be released, e.g., sulfur dioxide (SO_2), nitrogen oxides (NO_X), and carbon dioxide (CO_2), which are extremely dangerous to the global climate [2]. In this case, the Chinese government pledges to reduce the CO_2 emissions per unit of GDP by 60–65% in 2030 compared to 2005 levels [3]. To meet the above requirement, it is an inevitable trend to develop coal-fired power generation technology with large-capacity, low-pollution, and high-efficiency.

At present, most power plant designers are attempting to improve the boiler-turbine efficiency by increasing steam parameters [4]. Thus, ultra-supercritical (USC) coal-fired power plants operating at higher temperature and pressure levels have been gaining increasing attention worldwide. Theoretically, every 20 °C rise in the main steam temperature can result in an approximately 1% increase in efficiency [5]. The cycling heat efficiency of the USC units is up to 49%, which is approximately 10% higher than that of subcritical units. Meanwhile, the release of CO_2 and SO_2 can be reduced by 145 g/kWh and 0.4 g/kWh, respectively [6]. In the past decades, the USC power plants have been greatly promoted in China, with more than one hundred 1000 MW USC units put into operation by the end of 2017.

While the USC units enjoy higher efficiency and lower emissions, they are also more complicated than subcritical units. For instance, there is no obvious boundary between water and steam under the once-through operation, resulting in the strong coupling effect among boiler parameters. In addition, the load-cycling operation of USC units lead to the operating point changing in a wide range, making the nonlinearity of the plant variables even more serious. Due to the high-complexity as well as the nonlinearity, multi-variable, strong-coupling characteristics, the modeling and control of USC units faces greater challenges.

The modeling of power plants can be categorized into two groups: first-principle modeling [7–9] and experimental modeling [10]. A classical nonlinear dynamic model derived from first-principles for a natural circulation 160 MW drum-boiler was presented in [11], which was developed on the basis of several fundamental physical laws. Due to its clear physical structure, this model has been widely used for controller design [12–14]. In [15], a mathematical simulation model was developed to study the stability of a steam boiler drum subjected to all of the possible initial operating conditions, including both stable and unstable. Papers [16,17] present the static and dynamic mathematic model of a supercritical power plant and its application to improve the load changes and start-up processes. In [18], three different flexible dynamic models of the same single-pressure combined-cycle power plant have been successfully developed, and based on these models, an evaluation of the drum lifetime reduction was performed. However, owing to the complexity of the USC unit, it is hardly possible to build an accurate first-principle model, and experimental modeling offers a good framework. In [6], the dynamic model of a 1000 MW power plant was established by combining the experimental modeling approach and the first-principle modeling approach, which can be feasible and applicable for simulation analysis and testing control algorithms. Based on this model, a sliding mode predictive controller was proposed in [19] to achieve excellent load tracking ability under wide-range operation. In [20], this model was further improved with added closed-loop validations and more reasonable structure.

In 1995, Irwin originally developed a feedforward neural network (NN) to model a 200 MW oil-fired and drum-type turbo-generator unit [21]. Due to its practicability and flexibility, NNs become useful tools for power plant modeling [22,23]. In [24], an effective NN modeling method for a steam boiler was proposed: this model maps the influence of flue gas losses and energy losses due to unburned combustibles on the main operational parameters of the boiler. In [25], two separate NN models were developed for the boiler and the steam turbine, which are eventually integrated into a single NN model representing a 210 MW real power plant. Subsequently, some other methods were also introduced in NN, such as fuzzy logic. In [26], Liu et al. firstly presented a model of a USC unit using a fuzzy neural network, the results showing that the model's built-in fuzzy neural network had satisfactory accuracy and performance. In [27], a fuzzy model of the USC unit was firstly developed, and then based on the model, an extended state observer-based model predictive control was proposed. In [28], an improved Takagi-Sugeno fuzzy framework was applied to the modeling of a 1000 MW USC unit, the parameters were identified by a k-means++ algorithm and an improved stochastic gradient algorithm.

During the past decades, computer technology has been widely used in USC power plants. The supervisory information system, which provides comprehensive optimization for the plant's real-time production, collecting all the process data and storing the data in the historical database. These massive datasets are of great value since they can reflect the actual operational condition of the USC unit and embody the unit's complex physical and chemical characteristics. The big data generated in the USC power plant are generally characterized by massiveness, multi-source, heterogeneity, and high-dimension. Developing an advanced modeling technology based on big data is of great significance to the USC unit. Traditional NNs with shallow architecture have low efficiency in digging and extracting effective information from big data, since they often suffer from uncontrolled convergence speed and local optima. Meanwhile, it is more difficult to optimize the parameters of NNs as the number of hidden layers and the training sample size increase.

The deep neural network (DNN), proposed by Hinton et al. in 2006 [29], provides an effective tool to deal with the big data modeling problem. In the DNN, layer-by-layer unsupervised learning is performed for pre-training before the subsequent supervised fine-tuning [30]. The lower layers represent the low-level features from inputs while the upper layers extract the high-level features that explain the input samples. Through layer-wise-greedy-learning, DNNs can effectively extract the compact, hierarchical, and inherently abstract features in the original data and, thus, are able to achieve high-performance modeling with big data. As one of the commonly used DNN architectures, a stacked auto-encoder (SAE) is constructed by stacking several shallow auto-encoders (AE) [31], which learns features by first encoding the input data and then reconstructing it. Due to its remarkable representation ability, SAE has been successfully applied in fault diagnosis [32], electricity price forecasting [33], and wind speed forecasting [34].

During the training procedure of SAE, the mean square error (MSE) has been widely used as the loss function, owing to its simplicity and efficiency. SAE under MSE usually performs well when the training data are not disturbed by outliers. However, in practical application, the dataset obtained from a USC power plant will inevitably contain outliers due to various reasons, which makes the performance of the SAE deteriorate rapidly. Therefore, it is quite important to develop a new loss function. Unlike MSE, maximum correntropy (MC) [35] is a Gaussian-like weighting function, it is a local criterion of similarity and thus can be very useful for cases when the measurement data contains outliers. Since it could attenuate the large error terms effectively, the outliers would have a less impact.

Accordingly, the main contributions of this paper are summarized as follows:

(1)   In order to establish an accurate USC unit model using generated big data, SAE is adopted as the DNN model structure in this paper. The SAE model can generalize very well and yield better performance when compared to conventional shallow architectures. The SAE model is concise and suitable for big data analysis.
(2)   In order to reduce the bad influence of outliers on the modeling, a loss function using MC is developed in this paper.

The rest of the paper is organized as follows: Section 2 presents a brief description for the USC unit. Section 3 proposes the USC power plant modeling using SAE. The simulation results are given in Section 4. Finally, conclusions are drawn in Section 5.

2. The Ultra-Supercritical Coal-Fired Boiler-Turbine Unit

2.1. Brief Description of USC Unit

The power plant considered in this paper is a pulverized coal firing, once-through steam-boiler generation unit with a power rate of 1000 MW. The maximum steam consumption of the power plant is 2980 T/h with a superheated steam pressure and temperature of 26.15 MPa and 605 °C, respectively. Figure 1 shows the simplified diagram of the USC boiler-turbine unit. The boiler mainly includes the economizer, the waterwall, the separator, the superheater, and the reheater. The tandem compound triple turbine consists of a high-pressure (HP) turbine, an intermediate-pressure (IP) turbine, and a low-pressure (LP) turbine.

As shown in Figure 1, the pulverizing system transforms the raw coal into the pulverized coal so that it can fully burn in the furnace. The feedwater is first warmed by the economizer, and further heated in the waterwall which surrounds the furnace vertically and spirally. Eventually, it turns into steam with high temperature and pressure. There is a separator on top of the furnace, and the steam passes through the separator and superheats in the superheater. The superheater consists of four parts: primary, division, platen, and finish. The turbine governor valve controls the quantity of superheated steam delivered to the HP turbine. The extraction steam from the HP turbine goes to the reheater. The reheated steam is used to drive the IP/LP turbine.
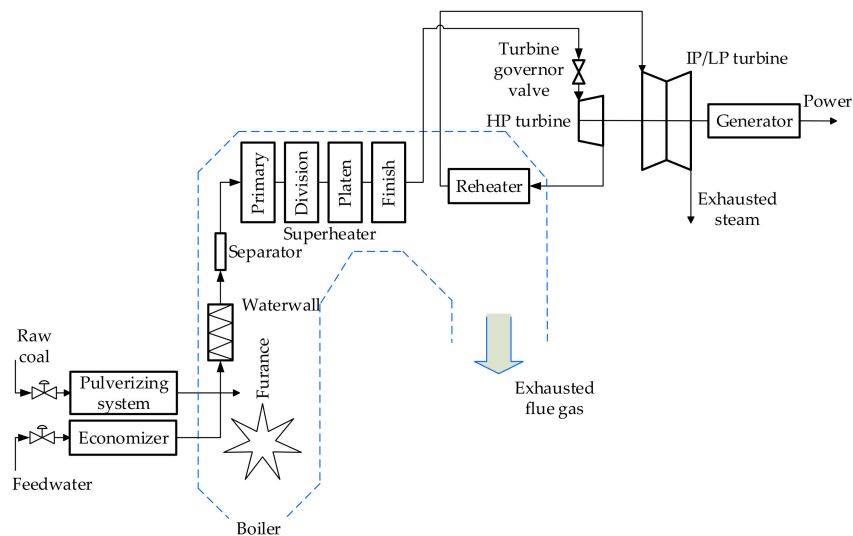
Figure 1. The layout of the 1000 MW boiler-turbine unit.

2.2. Determination of Input-Output Variables

In the once-through operation of the USC unit, there is no obvious boundary between water and steam. The feedwater is continuously heated, evaporated, and superheated from the inlet of the economizer. Without the buffering of the steam drum, the USC unit will suffer greater disturbances than a subcritical unit. This leads to the strong non-linearity and coupling of the USC unit, which can be seen as a complicated system with multiple inputs and multiple outputs.

In order to reduce the impact of external disturbances and simplify the model structure of the USC unit, the following assumptions are made:

1. The fuel flow and the forced draft volume are balanced to ensure the combustion stability.
2. The ratio between the forced draft volume and the induced draft volume remains constant, to ensure that the pressure in the furnace is stable.
3. The control of the main steam temperature is relatively independent.

If the above assumptions are satisfied, the USC unit can be depicted as a three-input, three-output nonlinear system, as shown in Figure 2. The inputs u_1, u_2, u_3 are the fuel flow rate, the turbine governor valve opening, and the feedwater flow rate, respectively. The outputs y_1, y_2, y_3 are the electric power, the main steam pressure, and the separator outlet steam temperature, respectively. The direct correlation property between water and steam causes the strong coupling between the inputs and the outputs.
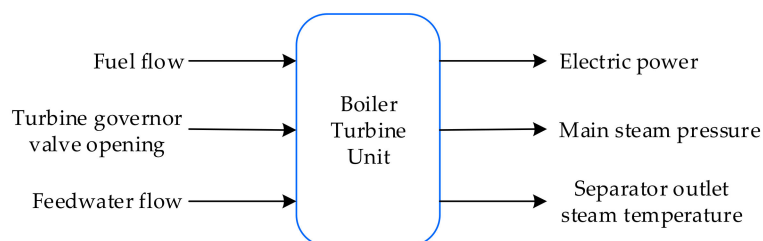

Figure 2. The three-input, three-output system of the USC unit.

3. Stacked Auto-Encoder

The SAE adopts a multi-layer structure, which is hierarchically stacked by a series of AEs, as shown in Figure 3. Denote the kth hidden layer to be h^k, and then the AE associated with h^{k-1} and $h^k (k = 1, 2, \cdots, l)$ is indicated as AE^k.
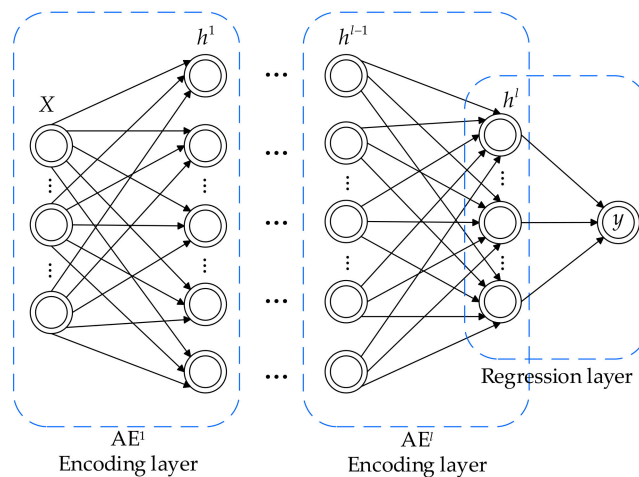
Figure 3. The architecture of the SAE.

3.1. Auto-Encoder

The AE is a one-hidden-layer feedforward NN with an encoder and a decoder. The encoder converts the input data from a high-dimensional representation into a low-dimensional abstract representation. Then the decoder reconstructs the input data from the corresponding codes. The main purpose of the AE is to learn an approximation in the hidden layer so that the input data can be perfectly reconstructed in the output layer. The structure of AE^k is shown in Figure 4.
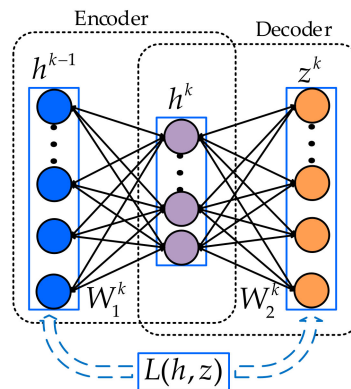

Figure 4. The *k*th auto-encoder.

Given the input h^{k-1} of AE^k, the hidden representation h^k can be obtained through the encoder based on Equation (1), and then maps back to a reconstructed vector z by the decoder as in Equation (2):

$$h^k = f(W_1^k h^{k-1} + b_1^k), \tag{1}$$

$$z^k = g(W_2^k h^k + b_2^k), \tag{2}$$

where the function $f(x) = g(x) = 1/(1 + e^{-x})$, W_2^k and b_1^k represent the weight matrix and bias term of the encoder, and W_2^k and b_2^k represent the weight matrix and bias term of the decoder, respectively. The parameter set in AE^k is $\theta^k = \left\{ W_1^k, b_1^k, W_2^k, b_2^k \right\}$.

The parameter set θ^k can be optimized by minimizing the reconstruction error:

$$J_{AE}(\theta^k) = \frac{1}{m} \sum_{i=1}^{m} L(h_i^{k-1}, z_i^k), \tag{3}$$

where *m* is the sample size and *L* is the mean square error (MSE) expressed as:

$$L_{AE-MSE}\left(h_i^{k-1}, z_i^k\right) = \frac{1}{2}\left\|h_i^{k-1} - z_i^k\right\|^2,$$ (4)

3.2. New Loss Function Design Using Maximum Correntropy

Usually large amounts of operating data are captured continuously by the online plant data acquisition system in the USC power plant. Before using these data for network training, data preprocessing is required, since they will always contain some outliers. The generation of outliers may include faulty sensors, human errors, errors in data capturing system, etc. However, it is very difficult to remove all outliers manually since the sample size is too large.

During the training procedure of AE^k ($k = 1, 2, \cdots, l$), the MSE is used as the loss function, owing to its simplicity and efficiency. The AE under MSE usually performs well when the training data are not disturbed by outliers. However, when the outliers are mixed within the training data, the performance of the AE under MSE may deteriorate greatly.

Notice that the MSE function is a quadratic function in the joint space with a valley along the $h^{k-1} = z^k$ line. The quadratic term has the net effect of amplifying the contribution of samples which are far away from the $h^{k-1} = z^k$ line, so that the outliers would have a great impact on the normal training of the model.

Unlike MSE, MC [35] uses a Gaussian-like weighting function so that it is a local criterion of similarity and, thus, can be very useful for cases when the measurement data contains large outliers. Since it could attenuate the large error terms effectively, the outliers would have less of an impact. The MC function to be maximized is expressed as:

$$L_{AE-MC}(h_i^{k-1}, z_i^k) = \frac{1}{d}\sum_{i=1}^{d} K_\sigma(h_i^{k-1}, z_i^k),$$ (5)

where *d* is the number of the output units, and $K_\sigma(\cdot, \cdot)$ is the Gauss kernel, which is defined as:

$$K_\sigma(a, b) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(a-b)^2}{2\sigma^2}\right),$$ (6)

where σ is the kernel size.

Owing to the effectiveness of MC function, it is chosen as the loss function of each AE in this paper.

3.3. SAE Model Structure and Learning Algorithm

The SAE model can be established by stacking several AEs. Figure 5 shows the structure of the SAE used for USC modeling. Due to the inertia and delay of the system, the historical data of $\{u_1(k), u_2(k), u_3(k), y_1(k-1), y_2(k-1), y_3(k-1)\}$ in the last two steps are also adopted as the inputs of the model. Thus, the total number of inputs and outputs are 18 and three, respectively. In this model, multiple AEs are used to obtain the intrinsic features from the original USC data, while the regression layer is responsible for outputting the expected normal behaviors of the system along the time axis.
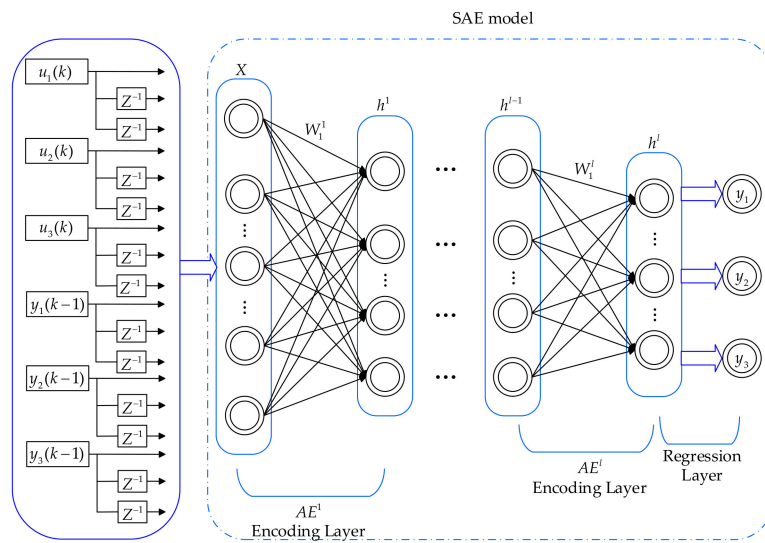
Figure 5. Structure of the SAE used for USC modeling.

The training of the SAE includes two steps: an unsupervised layer-wise pre-training step and a supervised fine-tuning step, as shown in Figure 6. In Figure 6a, with the original training data, the AE at the bottom layer is first trained by minimizing the reconstruction error in Equation (3) using the gradient descent method. Then, the generated hidden representation can be used as the input for training the higher-level AE. In this way, multiple AEs can be stacked hierarchically. After the layer-wise pre-training, all the obtained hidden layers are stacked, and the regression layer can be added on top of the SAE to generate the final outputs, as shown in Figure 6b. The parameters of the whole SAE network can be fine-tuned in a supervised way using the gradient descent method.
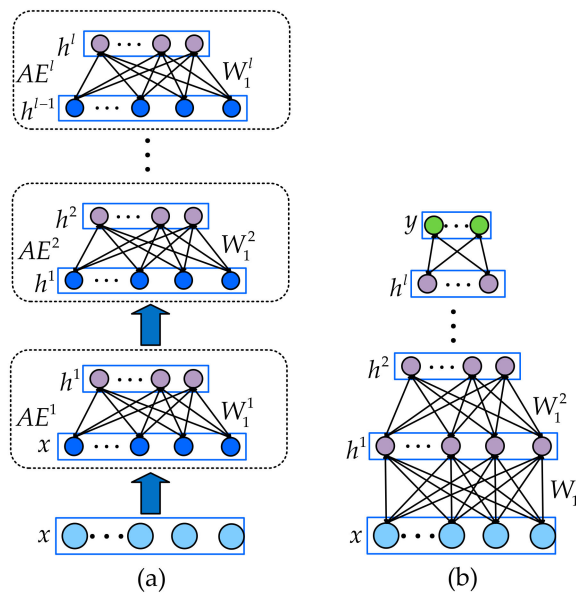

Figure 6. (**a**) Unsupervised layer-wise pre-training of the SAE. (**b**) Supervised fine-tuning of the SAE.

4. USC Unit Modeling

4.1. Experimental Settings

Training of the SAE with a dataset including all possible variations in the range of working conditions is very crucial. The dataset used for training was carefully selected from the very large amount of data logged in the historical database, during which the working condition varies frequently.

Twenty thousand sets of continuous I/O data with 1 s sampling were selected for training, with load changing conditions ranging from 550 MW to 1000 MW. Another 3000 sets of I/O data were selected for validation.

Within the datasets, there exist outliers that need to be removed in advance. In practice, the outliers can be identified in different ways. Usually, the data points which deviate substantially from the general trend of variations of its neighboring points can be considered as outliers. Additionally, the outliers can be found by checking the relationship between trends of data for highly correlated parameters. For example, the increase of fuel flow must correspond to the effect on that of the electric power, with a regular correspondence. By using these method, the detected outliers are listed in Table 1. As for the identified outliers in the dataset, they are replaced by the data from their neighboring points. These datasets used for training and validation, after preprocessing the outliers, are shown in Figure 7. All the preprocessed data are normalized in the range of [0,1] before establishing the SAE model.
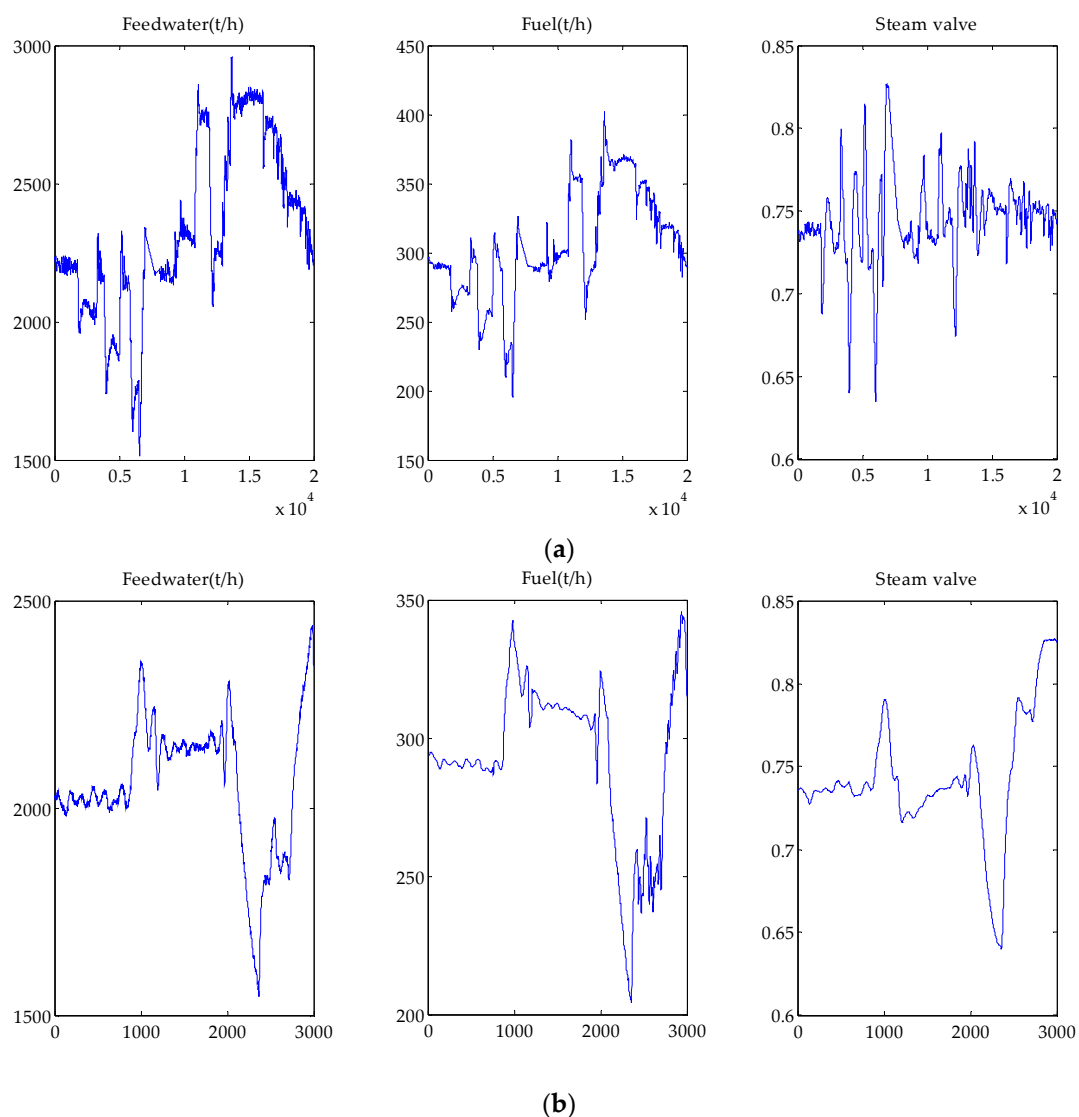

Figure 7. The datasets used for training (**a**) and validating (**b**).

Table 1. The number of outliers in training and validating sample sets.

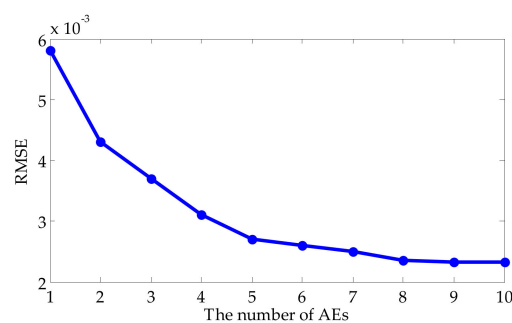| Parameter	Training Sample Set	Validating Sample Set
u_1	36/20,000	14/3000
u_2	24/20,000	8/3000
u_3	34/20,000	11/3000
y_1	28/20,000	14/3000
y_2	23/20,000	12/3000
y_3	29/20,000	9/3000

The root mean squared error (*RMSE*) in Equation (7) is employed as the evaluation metric:

$$RMSE = \sqrt{\frac{\sum_{k=1}^{K} (y_k^* - y_k)^2}{K}}, \tag{7}$$

where y^* is the model output and y is the plant output, K is the total data number. Notice that y^* and y are normalized values. There are several parameters that have to be defined for the SAE model, such as the nodes in each layer, the number of AEs, learning rate, momentum, etc. These parameters are determined through cross-validation only on the training set. The initial weights and biases of each AE were chosen to be small random values sampled from a zero-mean Gauss distribution with a standard deviation of 0.01. The maximum number of epochs is set to 100 and the fine-tuning stage terminates when the variation in *RMSE* of the validation set is less than 10^{-3}. This criterion will reduce the model complexity and, thus, result in a better generalization by avoiding overfitting.

In order to determine the optimal structure of SAE model, i.e., the number of AEs and hidden layer units in each AE, experiments were repeatedly done by choosing the number of AEs ranging from 1 to 10, while the number of units in hidden layers from $\varphi = [18, 17, \cdots, 5, 4]$. The optimal structure is found from different configurations considering the *RMSE* value.

The relationship between the number of AEs and the *RMSE* of the learning network is shown in Figure 8. The network is unable to generalize well when the number of AEs is too small because of the insufficient number of tunable parameters in the model. The performance of the network gradually improves as the number of AEs increases, especially when the number reaches 8. However, when the number of AEs increase further, the improvement seems to be very little, as using more AEs will lead to more complex structures that are prone to overfitting. Moreover, the vanishing gradient problem also imposes negative impacts on the fine-tuning of the SAE when the number of AEs increases. Therefore, the network structure of SAE is set to be eight hidden layers.

Figure 8. The relationship between the number of AEs and the *RMSE* value.

4.2. The Modeling Results

Figure 9 shows the modeling results with the 20,000 sets of training data. Then, this training model was validated by the 3000 sets of validating data, as shown in Figure 10. When adopting very different sets of operating data, the SAE model is still able to achieve good performance. From both the

training and validating, it is clearly seen that the SAE model can predict the USC dynamic accurately over a wide range of loads.
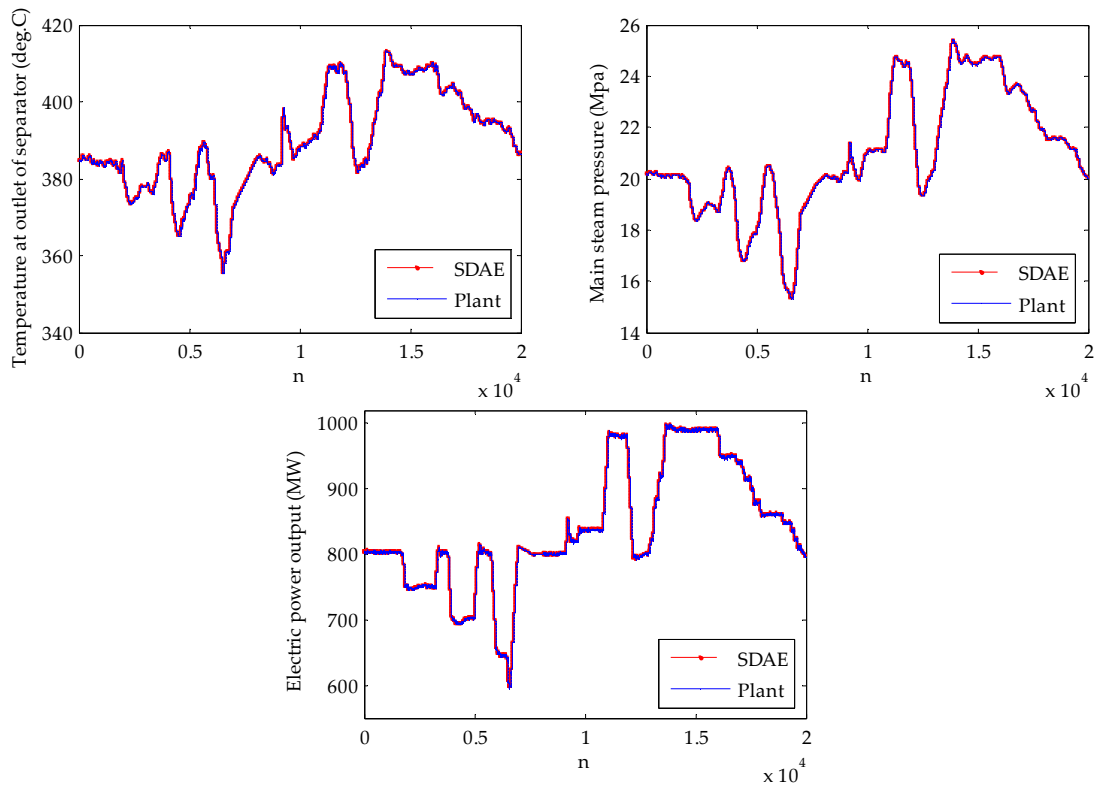

Figure 9. A comparison of the boiler system and the SAE model (training).
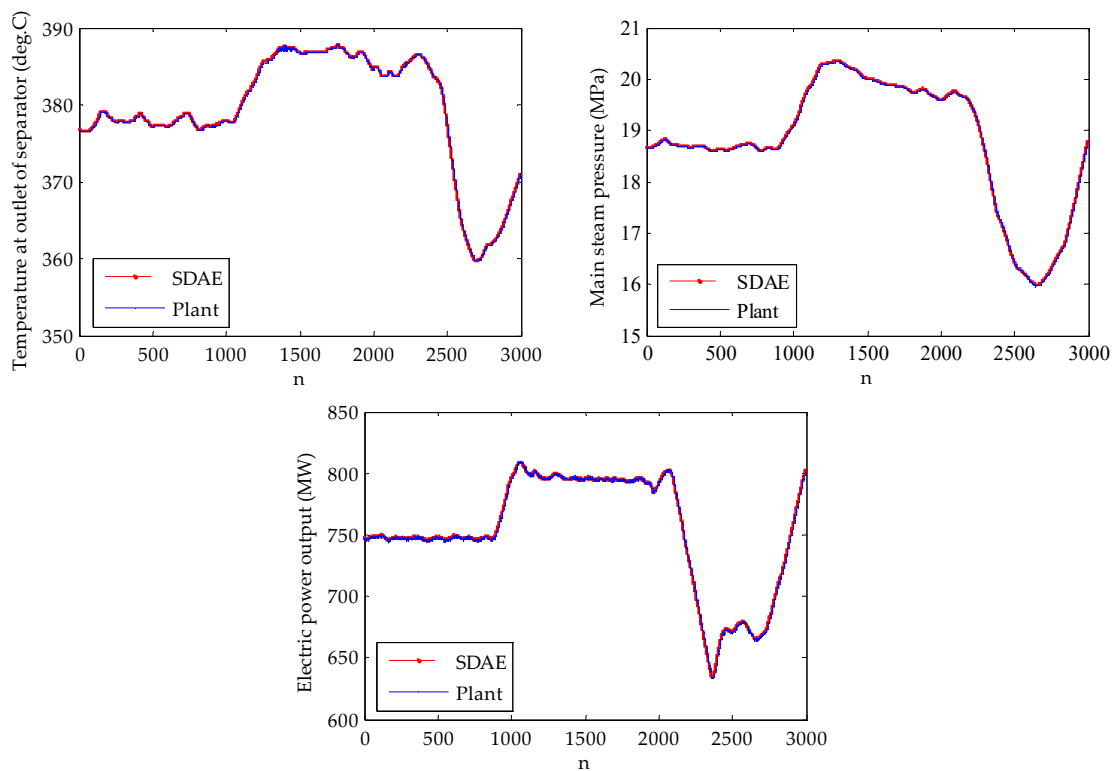

Figure 10. A comparison of the boiler system and the SAE model (validating).

This SAE model was then compared with a general multi-layer perceptron (MLP) network adopting the structure in [26], under the same I/O data. Table 2 lists the comparing results. It is found that the performance of the MLP network is lower than that of the SAE network, as it is a shallow architecture which often suffers from uncontrolled convergence speed and local optimality, especially when the training sample size grows too large.

Table 2. The root mean square errors of three adopted models.

		Temperature	Pressure	Power
MLP	Training	0.0039	0.0022	0.0034
	Validating	0.0076	0.0065	0.0072
SAE	Training	0.0016	0.0007	0.0015
	Validating	0.0031	0.0019	0.0034

4.3. The Modeling Using Maximum Correntropy

As listed in Table 1, there exists outliers in both the training and validating sample sets. In order to disclose the essential influence of these outliers on the modeling effect, the simulation is repeated without the preprocessing process. It can be found that the modeling performance soon deteriorates, as shown by the green line in Figures 11 and 12.
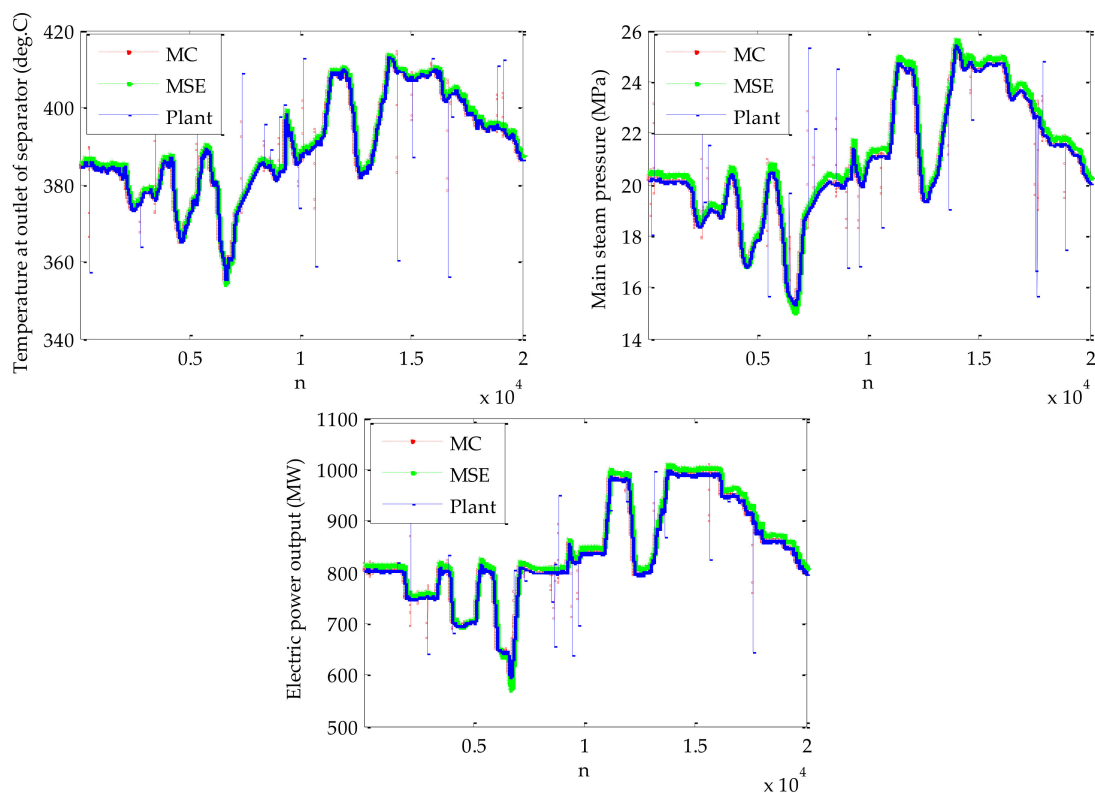

Figure 11. A comparison of SAE models using different loss functions without the preprocessing process (training).
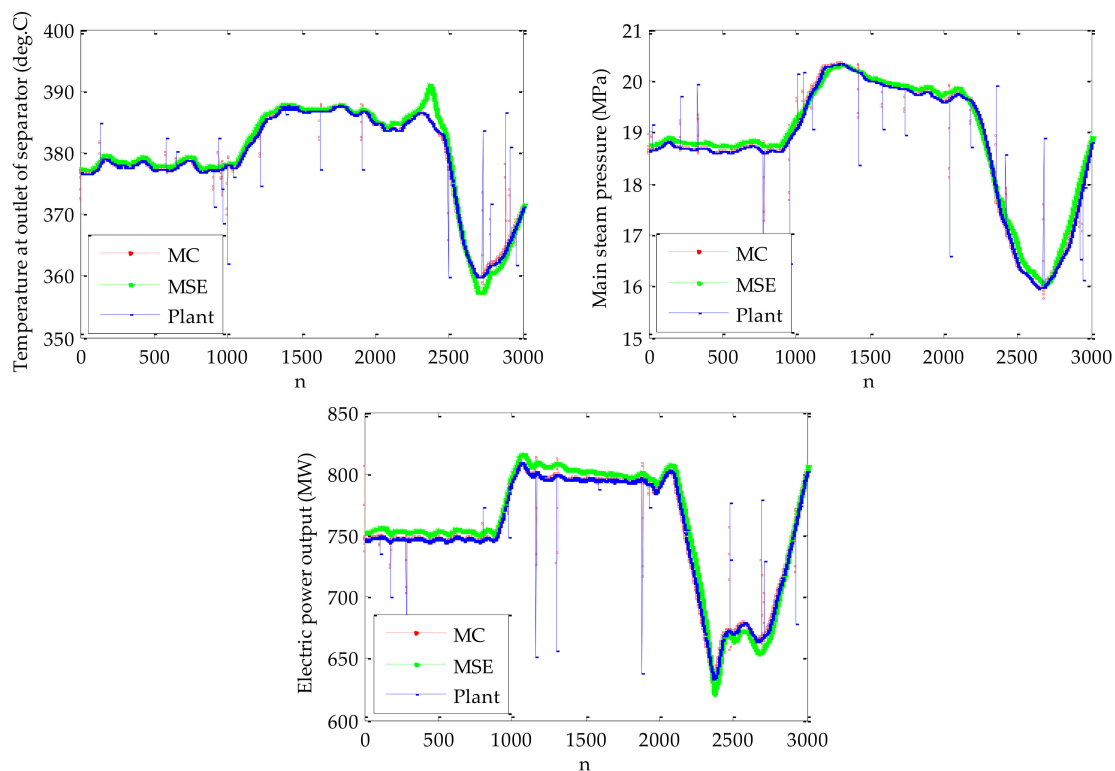
Figure 12. A comparison of SAE models using different loss functions without the preprocessing process (validating).

The simulation using the MC loss function is shown by the red line of Figures 11 and 12. Table 3 lists the *RMSE* values under these two methods, which clearly shows the advantage of the SAE model incorporating the MC function in alleviating the influence of outliers.

Table 3. The root mean square errors of SAE without the preprocessing process.

	RMSE	Temperature	Pressure	Power
MSE	Training	0.0231	0.0301	0.0307
	Validating	0.0529	0.0476	0.0559
MC	Training	0.0185	0.0209	0.0169
	Validating	0.0217	0.0275	0.0248

5. Conclusions

For the modeling of a 1000 MW USC coal-fired boiler-turbine unit, a DNN framework using an SAE was proposed in this paper. Real-time measurement big data generated from a wide range of operating points were used for the network training and validating. To evaluate the effectiveness of the proposed model, a comparative analysis of the SAE and MLP network was constructed. From the results, the following conclusions can be summarized.

(1) Compared with the shallow-layer NN, the DNN architecture adopting the SAE model trained by an unsupervised greedy layer-by-layer pre-training and a supervised fine-tuning is very efficient to deal with the big data modeling problem since it can effectively extract the compact, hierarchical, and inherent abstract features in the original USC unit data through the layer-wise-greedy-learning.

(2) MC is a local criterion of similarity which could attenuate the large error terms effectively. The proposed SAE model adopting MC as the loss function reduces the poor influence of outliers effectively, compared with adopting MSE as the loss function.

In summary, the proposed SAE model can be suitably applied for analyzing the dynamic behaviors of the 1000 MW USC boiler-turbine system.

Abbreviations

The following abbreviations are used in this manuscript:

USC	Ultra-supercritical
SAE	Stacked auto-encoder
CO_2	Carbon dioxide
SO_2	Sulfur dioxide
NO_X	Nitrogen oxides
NN	Neural network
DNN	Deep neural network
AE	Auto-encoder
MSE	Mean square error
MC	Maximum correntropy
HP	High-pressure
IP	Intermediate-pressure
LP	Low-pressure
RMSE	Root mean squared error
MLP	Multi-layer perceptron

References

1. National Bureau of Statistics of China. Available online: http://www.stats.gov.cn/tjsj/zxfb/201803/t2018039_1588744.html (accessed on 25 August 2019).
2. Zhao, Y.; Wang, S.; Duan, L.; Lei, Y.; Cao, P.; Hao, J. Primary air pollutant emissions of coal-fired power plants in China: Current status and future prediction. *Atmos. Environ.* **2008**, *42*, 8442–8452. [CrossRef]
3. Zhao, Z.; Chen, Y.; Chang, R. How to stimulate renewable energy power generation effectively? China's incentive approaches and lessons. *Renew. Energy* **2016**, *92*, 147–156. [CrossRef]
4. Wang, C.; Zhao, Y.; Liu, M.; Qiao, Y.; Chong, D.; Yan, J. Peak shaving operational optimization of supercritical coal-fired power plants by revising control strategy for water-fuel ratio. *Appl. Energy* **2018**, *216*, 212–223. [CrossRef]
5. János, M.B. High efficiency electric power generation: The environmental role. *Prog. Energy Combust.* **2007**, *33*, 107–134.
6. Liu, J.; Yan, S.; Zeng, D.; Hu, Y.; Lv, Y. A dynamic model used for controller design of a coal fired once-through boiler-turbine unit. *Energy* **2015**, *93*, 2069–2078. [CrossRef]
7. Alobaid, F.; Mertens, N.; Starkloff, R.; Lanz, T.; Heinze, C.; Epple, B. Progress in dynamic simulation of thermal power plants. *Prog. Energy Combust.* **2016**, *59*, 79–162. [CrossRef]
8. Sun, L.; Hua, Q.; Li, D.; Pan, L.; Xue, Y.; Lee, K.Y. Direct energy balance based active disturbance rejection control for coal-fired power plant. *ISA Trans.* **2017**, *70*, 486–493. [CrossRef]
9. Sun, L.; Li, D.; Lee, K.Y.; Xue, Y. Control-oriented modeling and analysis of direct energy balance in coal-fired boiler-turbine unit. *Control Eng. Pract.* **2016**, *55*, 38–55. [CrossRef]
10. Kocaarslan, I.; Cam, E. Experimental modeling and simulation with adaptive control of power plant. *Energy Convers. Manag.* **2007**, *48*, 787–796. [CrossRef]
11. Åström, K.J.; Bell, R.D. Simple Drum-Boiler Models. In Proceedings of the IFAC Power Systems Modeling and Control Applications, Brussels, Belgium, 5–8 September 1988; Volume 21, pp. 123–127.

12. Liu, X.; Guan, P.; Chan, C. Nonlinear multivariable power plant coordinate control by constrained predictive scheme. *IEEE Trans. Control Syst. Technol.* **2010**, *18*, 1116–1125. [CrossRef]

13. Wu, X.; Shen, J.; Li, Y.; Lee, K.Y. Data-driven modeling and predictive control for boiler-turbine unit. *IEEE Trans. Energy Convers.* **2013**, *28*, 470–481. [CrossRef]

14. Moradi, H.; Alasty, A.; Vossoughi, G. Nonlinear dynamics and control of bifurcation to regulate the performance of a boiler-turbine unit. *Energy Convers. Manag.* **2013**, *68*, 105–113. [CrossRef]

15. Bracco, S.; Troilo, M.; Trucco, A. A simple dynamic model and stability analysis of a steam boiler drum. *J. Power Energy* **2009**, *223*, 809–820. [CrossRef]

16. Alobaid, F.; Postler, R.; Ströhle, J.; Epple, B.; Gee, K.H. Modeling and investigation start-up procedures of a combined cycle power plant. *Appl. Energy* **2008**, *85*, 1173–1189. [CrossRef]

17. Alobaid, R.; Ströhle, J.; Epple, B.; Gee, K.H. Dynamic simulation of a supercritical once-through heat recovery steam generator during load changes and start-up procedures. *Appl. Energy* **2009**, *86*, 1274–1282. [CrossRef]

18. Benato, A.; Bracco, S.; Stoppato, A.; Mirandola, A. Dynamic simulation of combined cycle power plant cycling in the electricity market. *Energy Convers. Manag.* **2016**, *107*, 76–85. [CrossRef]

19. Tian, Z.; Yuan, J.; Zhang, X.; Kong, L.; Wang, J. Modeling and sliding mode predictive control of the ultra-supercritical boiler-turbine system with uncertainties and input constraints. *ISA Trans.* **2018**, *76*, 43–56. [CrossRef]

20. Fan, H.; Zhang, Y.; Su, Z.; Wang, B. A dynamic mathematical model of an ultra-supercritical coal fired once-through boiler-turbine unit. *Appl. Energy* **2017**, *189*, 654–666. [CrossRef]

21. Irwin, G.; Brown, M.; Hogg, B.; Swidenbank, E. Neural network modeling of a 200 MW boiler system. *IEE Proc. Control Theory Appl.* **1995**, *142*, 529–536. [CrossRef]

22. Suresh, M.V.J.J.; Reddy, K.S.; Kolar, A.K. ANN-GA based optimization of a high ash coal-fired supercritical power plant. *Appl. Energy* **2011**, *88*, 4867–4873. [CrossRef]

23. Lu, S.; Hogg, B.W. Dynamic nonlinear modeling of power plant by physical principles and neural networks. *Int. J. Electr. Power Energy Syst.* **2000**, *22*, 67–78. [CrossRef]

24. Rusinowski, H.; Stanek, W. Neural modelling of steam boilers. *Energy Convers. Manag.* **2007**, *48*, 2802–2809. [CrossRef]

25. Smrekar, J.; Pandit, D.; Fast, M.; Assadi, M.; De, S. Prediction of power output of a coal-fired power plant by artificial neural network. *Neural Comput. Appl.* **2010**, *19*, 725–740. [CrossRef]

26. Liu, X.; Kong, X.; Hou, G.; Wang, J. Modeling of a 1000 MW power plant ultra super-critical boiler system using fuzzy-neural network methods. *Energy Convers. Manag.* **2013**, *65*, 518–527. [CrossRef]

27. Zhang, F.; Wu, X.; Shen, J. Extended state observer based fuzzy model predictive control for ultra-supercritical boiler-turbine unit. *Appl. Therm. Eng.* **2017**, *118*, 90–100. [CrossRef]

28. Hou, G.; Du, H.; Yang, Y.; Huang, C.; Zhang, J. Coordinated control system modeling of ultra-supercritical unit based on a new T-S fuzzy structure. *ISA Trans.* **2018**, *74*, 120–133. [CrossRef]

29. Hinton, G.E.; Salakhutdinov, R.R. Reducing the dimensionality of data with neural networks. *Science* **2006**, *313*, 504–507. [CrossRef]

30. Bengio, Y.; Lamblin, P.; Popovici, D.; Larochelle, H. Greedy layer-wise training of deep networks. In Proceedings of the 20th Annual Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 3–6 December 2006; Volume 19, pp. 153–160.

31. Schmidhuber, J. Deep learning in neural networks: An overview. *Neural Netw.* **2015**, *61*, 85–117. [CrossRef]

32. Sun, W.; Shao, S.; Zhao, R.; Yan, R.; Zhang, X.; Chen, X. A sparse auto-encoder-based deep neural network approach for induction motor faults classification. *Measurement* **2016**, *89*, 171–178. [CrossRef]

33. Wang, L.; Zhang, Z.; Chen, J. Short-term electricity price forecasting with stacked denoising autoencoders. *IEEE Trans. Power Syst.* **2016**, *32*, 2673–2681. [CrossRef]

34. Khodayar, M.; Kaynak, O.; Khodayar, M.E. Rough deep neural architecture for short-term wind speed forecasting. *IEEE Trans. Ind. Inform.* **2017**, *13*, 2770–2779. [CrossRef]

35. Liu, W.; Pokharel, P.P.; Principe, J.C. Correntropy: Properties and applications in non-Gaussian signal processing. *IEEE Trans. Signal Process.* **2007**, *55*, 5286–5298. [CrossRef]