The Application of Cyber Physical System for Thermal Power Plants: Data-Driven Modeling

Yongping Yang [1], Xiaoen Li [1], Zhiping Yang [1], Qing Wei [1], Ningling Wang [1] and Ligang Wang [2,*]

[1] National Research Center for Thermal Power Engineering and Technology, North China Electric Power University, Beinong Road 2, Beijing 102206, China; yyp@ncepu.edu.cn (Y.Y.); shawnli89@163.com (X.L.); yzprr@163.com (Z.Y.); weiqingmail@163.com (Q.W.); epower2004@163.com (N.W.)

[2] Industrial Process and Energy Systems Engineering, Swiss Federal Institute of Technology in Lausanne, Rue de l'Industrie 17, 1951 Sion, Switzerland

[*] Correspondence: lgwangeao@163.com; Tel.: +41-21-69-34208

check for
updates

Abstract: Optimal operation of energy systems plays an important role to enhance their lifetime security and efficiency. The determination of optimal operating strategies requires intelligent utilization of massive data accumulated during operation or prediction. The investigation of these data solely without combining physical models may run the risk that the established relationships between inputs and outputs, the models which reproduce the behavior of the considered system/component in a wide range of boundary conditions, are invalid for certain boundary conditions, which never occur in the database employed. Therefore, combining big data with physical models via cyber physical systems (CPS) is of great importance to derive highly-reliable and -accurate models and becomes more and more popular in practical applications. In this paper, we focus on the description of a systematic method to apply CPS to the performance analysis and decision making of thermal power plants. We proposed a general procedure of CPS with both offline and online phases for its application to thermal power plants and discussed the corresponding methods employed to support each sub-procedure. As an example, a data-driven model of turbine island of an existing air-cooling based thermal power plant is established with the proposed procedure and demonstrates its practicality, validity and flexibility. To establish such model, the historical operating data are employed in the cyber layer for modeling and linking each physical component. The decision-making procedure of optimal frequency of air-cooling condenser is also illustrated to show its applicability of online use. It is concluded that the cyber physical system with the data mining technique is effective and promising to facilitate the real-time analysis and control of thermal power plants.

Keywords: cyber physical system; thermal power plants; data mining; physical knowledge; online monitoring and control

1. Introduction

Thermal power plants convert chemical energies of fossil fuels, e.g., coal, natural gas, petroleum or oil, into electricity. Fossil fuels still play the most important role in energy generation sector and contributes 80% of world's power generation [1]. Nowadays, more and more aspects [2] are considered for the operation strategy of thermal power plants to improve key performance indicators, e.g., stability, safety, efficiency [3,4], flexibility [5,6] and emission level. The controlling and optimization of thermal power plants are facing new challenges under various boundary conditions and operation modes, which may be potentially handled by the application of cyber physical systems (CPS).

The concept of CPS has been proven to be an effective method for design and development of future systems [7]. CPS refers to the conjoining and coordination between the cyber layers

(e.g. information, models, simulation) and the physical layers (e.g., components, sensors and actuators). The cyber and physical layers can be operated at different spatial and temporal scales and can be tightly interacted with each other. There are three issues to be addressed for the application of CPS (as illustrated in Figure 1): (1) The cyber layers can receive real-time information of the state and performance of the physical layers; (2) CPS should include well-established models of each component, so that the physical layers can be accurately represented in the cyber layers; (3) Decision-making of the physical layers should be based on the analysis and results of the cyber layers. The handling of these issues relies on data mining technique for data analysis as well as the communication network for bidirectional data transmission. The whole resolution procedure involves the analysis of big data, the method of machine learning, and the concept of intellectual control and analysis. With such features, CPS has been widely applied in the field of energy systems [8], automotive systems [9], medical systems [10] and transportation systems [11]. Specific to the applications in energy systems, CPS has been intensively applied to, e.g., the smart grid [12,13], distribute power systems [14], smart buildings and cities [15–17], and peak shaving of renewable energy [18,19].
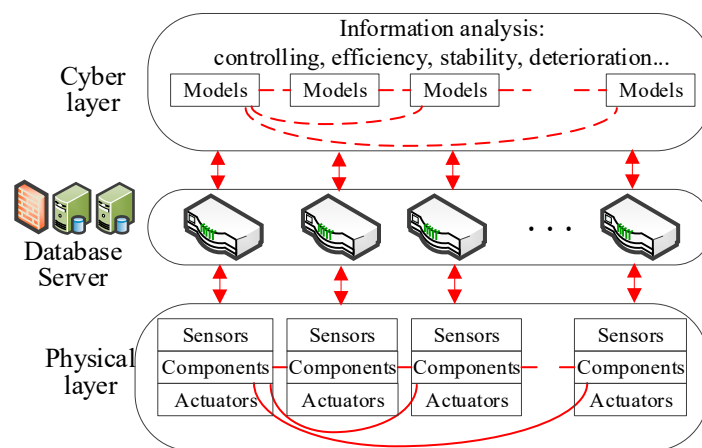

Figure 1. The generic structure of cyber physical system for the application of thermal power plants.

However, there are few papers on the application of CPS in power generation sector and there is a lack of general framework and standard application procedure. In this paper, the CPS for the application of thermal power plants is intensively discussed. Typical CPS structure is illustrated in Figure 1: The components, such as boiler, turbines, heat exchangers and pumps, are linked in the physical layer. For each component, corresponding submodules are established and linked to the cyber layer. The interaction between both layers relies on the communication network. The motivation of CPS application for thermal power plants is, based on accurate simulation models in the cyber layer, to deeply investigate the information generated in the physical layer for the online analysis and decision-making under any operation mode or any boundary conditions. The iterative procedure between the physical and cyber layers continuously gathers information from the physical layer, processes and analyzes data in the cyber layer, and sends analysis results to the physical layer to support the decision-making.

This paper is organized as follows: In Section 2, we propose a general framework for applying the CPS to thermal power plants after a review of various types of modeling methodologies and a follow-up discussion on how to choose the most proper method. Section 3 introduces a case study to describe the physical and cyber layers specifically for a traditional supercritical coal-fired power plant. A data-driven model of air-cooling condenser is established to further support the application of the proposed idea. Section 4 discusses the relationships between the theoretical modeling and machine learning based modeling. Finally, in Section 5, the conclusions and key innovations are further emphasized.

2. General Methodology for Applying CPS to Thermal Power Plants

In this section, the general approach for applying cyber physical system to thermal power plants is discussed. The concept of cyber physical system emphasizes the integration between the knowledge derived by machine learning algorithms from field data and theoretical analysis. The cyber layer first uses the historical data collected from the physical layer (e.g., the actual system layout and measurements) to establish the accurate regression model to represent the behavior of the considered components and systems under the system boundary and component states occurred in the past and to predict closely the system/component performance under different conditions. Then, the established models are employed as a parallel simulation platform, which receives real-time measured boundary conditions to obtain the optimal controlling strategy and/or the analysis of the efficiency, stability or safety of the physical layer. The general procedure of such CPS for thermal power plants is shown in Figure 2, where the steps A1–A3 represent the offline phase, while the steps B1–B3 stand for the online phase. The offline and online phases can be combined to provide an iterative online application.
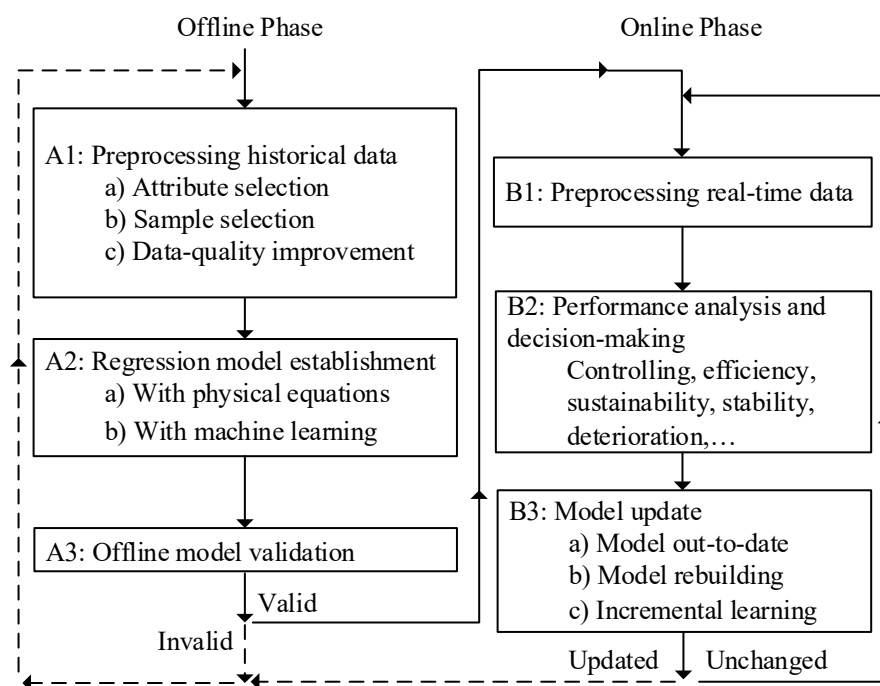

Figure 2. The procedure to establish a data-driven model based on the concept of cyber physical systems (CPS).

2.1. Preprocessing Historical Data (A1)

The CPS relies highly on the accuracy and quality of data sources. The preprocessing of historical data mainly includes the following three aspects:

2.1.1. Attribute Selection Based on the Business Target

A good understanding of the business target is the cornerstone for the proper selection of necessary input attributes to build the cyber-representation model for the considered components or subsystems. The physical layer also needs to be considered for the selection of the attributes. The sensors and actuators need to be well placed and, particularly, additional sensors or actuators for offering enough measurement redundancy might be added for monitoring and controlling of the components. With the redundancy measurements, some difficult-to-measure but necessary attributes can be obtained indirectly based on the basic mass and energy balance established with physical knowledge. For instance, most Rankine-cycle based thermal power plants do not equip sensors for

direct measurement of mass flowrate of the main steam, which is usually calculated based on the mass balance (e.g., the outlet mass flowrate of the condensate pump) and the energy balance of the feedwater preheaters. Therefore, the measurement system needs to be well defined and optimized to ensure that all the information required for the cyber layer is obtained with certain redundancy.

2.1.2. The Selection of the Historical Samples Based on Timeliness

For establishing a data-driven cyber physical model, it is important to select a proper period of historical samples. The latest samples should be considered with a high priority, since they represent the current component/system states more reasonably. There are mainly two issues for the selection of the historical samples: (1) How large the training dataset should be and (2) when the model should be updated to accurately reflect the current system/component states. Generally, a larger training dataset would cover more different boundary conditions so that the model established based on this data is expected to be valid for various situations and does not need to be updated frequently. However, a larger training dataset not only means a large computation time but also involves the representativeness of the data outdated. For a small dataset, it might not contain enough information for calibrating a model for those conditions not covered in the dataset.

2.1.3. The Enhancement of Data Quality

The data-driven model established at the cyber phase is based on the measurements, therefore the data quality may not be good enough due to measurement noises and sensors faults. However, the accuracy and validity of the model highly relays on the quality of the employed training data. Therefore, it is highly recommended to perform rigorous and conservative preprocessing of measurement data to ensure high data quality. Several data processing techniques are available with their advantages and disadvantages highlighted below.

(1) Data reconciliation

The well-known data reconciliation method is an excellent choice to enhance data quality. Data reconciliation technique utilizes redundant measurements from the employed sensor network to correct the measured value of each sensor by considering basic mass and energy balance of the flowsheet represented by the sensor network. Data reconciliation enhances the reliability of the original measured values by reconciled values obtained by systematically minimizing the system error of the sensor network, which is formulated as Equations (1) and (2):

$$\min_{\hat{x},\hat{u}} \gamma = \sum_{i=1}^{m} \left(\frac{\hat{x}_i - x_i}{\sigma_i} \right)^2, \tag{1}$$

$$\text{s.t. } f_l(\hat{x}_i, \hat{u}_k) = 0 \quad i = 1, \ldots, m; \ k = 1, \ldots, p; \ l = 1, \ldots, q, \tag{2}$$

where the terms $\hat{x}_i$ and x_i are the reconciled and measured values, respectively. The term σ_i is the standard deviation of the corresponding measured variable, while the objective function γ is the sum of all errors between the reconciled and measured values. The constraint functions f_l mainly refer to the mass balance equations, energy balance equations, and other physical equations or even the empirical equations. In the constraint functions f_l, the term $\hat{u}_k$ is the calculated (also named as reconciled) value of the unmeasured variables. The terms m, p and q represent the number of measured and unmeasured variables, and the number of constraint equations, respectively. As measurement redundancy is required for data reconciliation, the number of constraint equations should be larger than that of unmeasured variables ($p > q$) to solve the optimization problem of Equation (1), while the degrees of freedom can be described as $r = p - q$. In addition, data reconciliation is not compulsory but is highly recommended for building a highly-reliable model of the considered thermal power plant.

(2)   Outlier detection

The outliers, wrong measurements, in the recorded operating data may significantly reduce the reliability of the reconciled values. The outliers occur mainly because of the faults of the sensors themselves, the errors during data transmission, and the changes of component states due to, e.g., unit maintenance. Many useful approaches based on, e.g., statistical hypothesis [20], distance, density or frequency characteristics of the dataset [21–23], or machine learning [24–26], have been developed to detect outliers. As the outliers may occur at different locations and may not share the common natures, the choice of a suitable detection method based on the data characteristics is of great significance.

(3)   Normalization

Most attributes for modeling thermal power plants are continuous variables. Each of these attributes has its own unit and scaling. The normalization is a way of ensuring that the structure of the dataset is appropriate for the employed mathematical algorithms (discussed in detail in Section 2.2). For example, for the normalization in support vector regression method, the original data is usually scaled into [0,1] or [−1,1], while in Auto-Associative Kernel Regression (AAKR) method [27], the dataset is usually normalized to have zero mean and standard deviation equal to one. The lack of normalization may lead to longer computation time or even computational divergence.

2.2. Regression Model Establishment (A2)

This step essentially regresses the component/system states, which can predict the corrected measured variables of components of subsystems for given boundary conditions. The real component and the cyber layer can be directly connected with the regression model and the feedback from the cyber layer can assist the CPS to perform more reasonable and beneficial controlling of physical components. With advanced artificial intelligence to determine the optimal operating conditions for current boundary conditions and component states, the physical controlling procedure can be optimized to enhance the competitiveness.

2.2.1. Physical Modeling or Thermodynamic Modeling

If the behavior of the considered component has already been studied intensively in the literature, a physical model can usually be established in the cyber layer with physical equations or empirical equations for simulating the components [28–30]. Component characteristics, such as efficiency curve, heat transfer coefficient or pressure drop, can be acquired by analyzing the historical data recorded by the sensor network with parameter estimation technique. This is of great importance since the component performance usually decays after a certain time of operation and the original design performance cannot be valid for long-term operation. This ensures that the component states are always up-to-date.

2.2.2. Machine Learning

Machine learning is also a technique to identify and correct the parameters representing the component behaviors. Generally, machine learning is classified as supervised and unsupervised learning algorithms. The monitoring and controlling of thermal power plants always depend on various attributes and the relationships between the input variables and the targets (outputs). Thus, data mining in cyber physical system mainly refers to supervised learning algorithms, among which the support vector regression, neural network, fuzzy rough sets and decision tree are the most recommended methods for the application of thermal power plants. Here we focus on the support vector regression for machine learning.

Support vector regression (SVR) investigates mathematically the best-possible relationship between inputs attributes x and target variable y. The regression function of SVR algorithm, which is more concise compared with neural network, is formulated as Equation (3):

$$f(x) = \sum_{i=1}^{l} w_i \, \varphi_i(x_i) + b = w^{\mathrm{T}} \varphi(x) + b, \tag{3}$$

where the terms $w = [w_1, w_2, \ldots, w_l]^{\mathrm{T}}$ and b are the weight factor of each support vector and the function bias, respectively. The function $\varphi(x) = [\varphi_1(x_1), \ldots, \varphi_l(x_l)]^{\mathrm{T}}$ maps the input attributes x into a high-dimension space $\varphi(x)$. The training process of SVR model contains a trade of between the structural risk and confidence interval, which means SVR method usually has a good generalization performance. Besides, the concept of support vectors refers to the historical samples on the boundary of the decision surface. These samples are similar as the baseline, which shows the generalization of the established model and gives a further understanding of the component performance. These features make SVR method highly suitable for modelling thermal power plants. However, the shortage of SVR method is that it becomes computational expensive with an increasing data size.

2.3. Offline Model Validation (A3)

Before real-time optimization to offer the optimal controlling strategies, the established regression model should be validated to ensure high accuracy for various conditions. A commonly used technique is to employ 90% of the chosen dataset (after step A1) for the modeling calibration (step A2) and to use the remaining of the dataset for model validation. If the established model is with high accuracy for different operating conditions, it is then possible to be applied for the online phase.

2.4. Preprocessing of the Real-Time Data (B1)

This step is in fact similar to the step A1: preprocessing of historical data. However, only one sample or several samples are needed for each iteration, depending on whether the temporal information is needed in the model, thus less computation time is possible to achieve for real-time control.

2.5. Online Performance Analysis and Decision Making (B2)

With the established regression models, CPS can mainly use for performance analysis and decision making [31]. Specific to thermal power plants, CPS is available for monitoring the component conditions or system performances (e.g., thermal efficiency [4,5] or heat transfer efficiency), fault diagnosis (e.g., online advanced exergy analysis [32,33]) and life cycle analyze (e.g., [34]). Moreover, CPS can also support decision making and parameter optimization [35] by automatically determining the optimal controlling strategy and the corresponding operating parameters according to historical data and present state, which is further explained by the case study in Section 3.2.2.

2.6. Model Updates (B3)

The offline model was built by the recent, most representative historical data, while various situations, e.g., significant changes of boundary conditions, the failure of key components and continuous performance degradation of components, may cause a significant mismatch between the prediction of the regressed model and the actual component/system performance. Regular inspections of the target variables are needed to check when the regressed model should be updated to reflect the current component/system status:

(1)　Model rebuilding

Component overhaul or the change of operation modes leads to the requirement of rebuilding the model. The old datasets are discarded, and the current information is needed to rebuild the model. For example, the flowsheets of turbine subsystem are different in the heating period and non-heating period, thus the input attributes of extracted steam are different for modeling the turbine subsystem.

(2)　Incremental learning

Incremental learning can employ newly-added information and discard the outdated to prevent the burden of the huge dataset. This method is mainly used for the situation of performance degradation, under which the efficiency or the value of character parameter decays gradually and slowly. The difference between the predicted and the actual values increases if without a regular update. Incremental learning and support vector regression is the most popular combined techniques. Incremental support vector machine uses partial support vectors as the "candidate vectors", while the final support vectors are selected depending on the newly-added samples [36]. Incremental learning is useful for support vector regression method, especially giving the consideration that the computational memory of the SVR increases significantly for a large dataset. Similarly, the incremental supervised learning method with neural network is also an effective method for a large dataset (e.g., [37]).

(3)　Model unchanged

The incremental learning and the rebuild of the model are optional. Usually, new information just continuously passes through the steps B1 and B2, and then the online analysis and decision making are realized in real time.

3. Case Study: The Application to Turbine Subsystem and Air Cooling Condenser

Section 2 discussed the general steps and the methodologies for establishing a cyber physical system of thermal power plants. However, for different components or subsystems, the design of data-driven analysis systems can be complicated. Take the optimal controlling strategy of air-cooling condenser (ACC) as an example: The increase in fan frequency enhances the cycle efficiency and accordingly boosts the gross power output $\dot{E}_{gross}$; however, the air fan power consumption $\dot{E}_{fan}$ is also increased. Thus, there exists a maximum net power gain with a corresponding optimal fan frequency. Below the optimal fan frequency, the net power gain increases with the increase in the fan frequency. With the fan frequency increases continuously, the net power gain decreases as the fan consumption $\dot{E}_{fan}$ increases dramatically.

The optimal controlling strategy of air fans under various load rates, ambient temperatures and wind conditions is one of the most important issues of such power plants. In this section, the procedure of how to solve this decision-making problem based on the proposed CPS idea in Figure 2 is discussed. The turbine subsystem of a traditional 660 MW single-reheat supercritical coal-fired power plant is described first in Section 3.1. The maximization of the net power gain is carried out in Section 3.2 based on the real-time data. Only the modeling of air cooling condenser, the key component of such a system, is given in detail below rather than all the turbine subsystem components.

3.1. The Physical Layer

The physical layer of the thermal power plant is described in terms of components, flowsheet, sensors and actuators. The major physical components involved are the boiler, generator (G), turbine (divided into several stages), feedwater preheaters, air cooling condenser (ACC), feedwater pump (FWP), condensate pump (CP) and deaerator (DA). The flowsheet of the turbine subsystem is illustrated in Figure 3. The main steam generated in the boiler is expanded in the high-pressure turbine (HPT1-2) and is reheated afterward. The reheated steam then expands through the intermediate-pressure (IPT1-2) and low-pressure turbines (LPT1-4) and is finally condensed in the air-cooling condenser (ACC). For each turbine stage, certain amount of steam is extracted for

heating up feedwater in the *i*th feedwater preheater (FWPH*i*) before entering the boiler. In addition, a de-aerator (DA) is also equipped to remove dissolved gas in the feedwater.

The sensor network placed in the turbine subsystem is also drawn in Figure 3, where the red, green and white marks represent the sensors for pressure, temperature and mass flowrate, respectively. A temperature sensor and a pressure sensor are applied for each extracted steam. The inlet and outlet positions of the feedwater pump are applied with temperature sensors. The ambient temperature and ambient pressure are also recorded, which is mainly used for analyzing the performance of air cooling condenser. Mass flowrate sensors are placed at three different locations: the outlet of condensate pump, the inlet of the deaerator and the outlet of the feedwater pump. The electric output of the generator and the power consumption of air fans are also measured. Besides, there are multiple sensors of each measurement point for a mutual check. The main actuators in turbine subsystem include the control valve of turbines and pumps, and the frequency of air fans, etc.
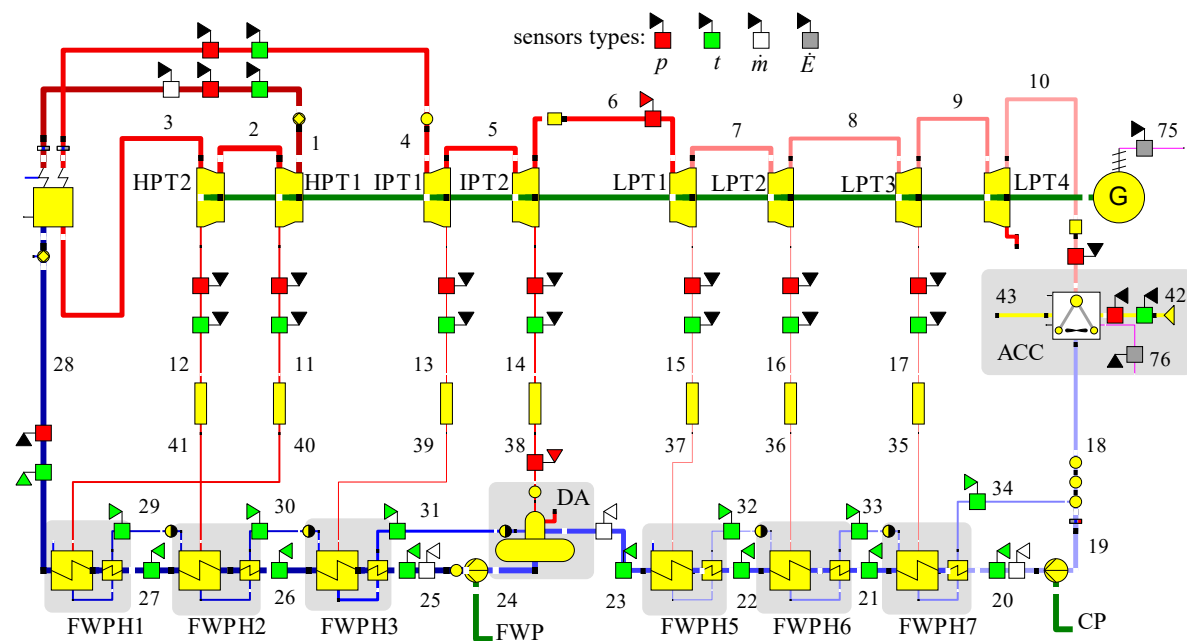

Figure 3. The flowsheet of the turbine subsystem investigated.

3.2. The Cyber Layer

The cyber system is responsible for performance analysis and decision making of each component. The modeling of turbine subsystem, especially the ACC, in the offline phase is given in Section 3.2.1. Based on the offline model, the decision making of optimal air fan frequency in online phase is solved and discussed in Section 3.2.2.

With the analysis of two-week real operation data (Figure 4), we have found that the present controlling strategy of fan frequency can only deliver qualitative control: The higher the load factor and the higher the ambient temperature, the larger the fan frequency will be. When the ambient temperature is higher than 27 °C or the load factor is higher than 90%, the fan frequency is adjusted to the maximum (55 Hz). For low load factors and low ambient temperatures, the fan frequency is varied within a wide range. Such control strategy follows the normal logics but is not necessarily the optimal in terms of overall-plant performance. A quantitative guide is needed to make sure the optimal plant operation under any combinations of ambient temperature and load factor. Therefore, following the procedure described in Figure 2, the quantitative relationship between the optimal fan frequency and the boundary conditions is established and discussed.
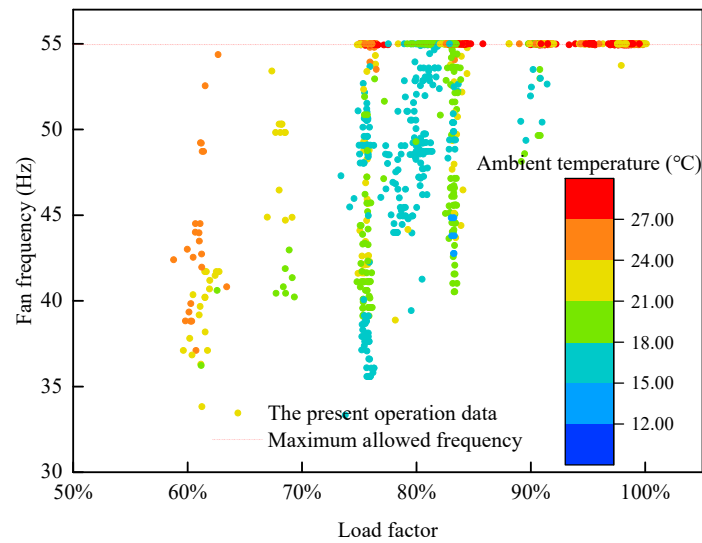
Figure 4. The present operation strategy of fan frequency.

3.2.1. The Offline Phase for Performance Prediction of the Turbine Subsystem and ACC

The cyber layer of the turbine subsystem is illustrated in Figure 5. Following the actual flowsheet in the physical layer, the models of each turbine stage ($\mathcal{F}_{\text{turbine}}$), feedwater preheaters ($\mathcal{F}_{\text{FWPH}}$), pumps ($\mathcal{F}_{\text{pump}}$), ACC ($\mathcal{F}_{\text{ACC}}$), boiler, and DA are linked in the cyber layer. For given boundary conditions, the gross power output of the turbine $\dot{E}_{\text{gross}}$ and the power consumption of air fans $\dot{E}_{\text{fan}}$ can be calculated by $\mathcal{F}_{\text{ACC}}$. The difficulty of accurate calculation of both terms is the determination of backpressure p_{back}, a crucial parameter for performance prediction of the ACC.
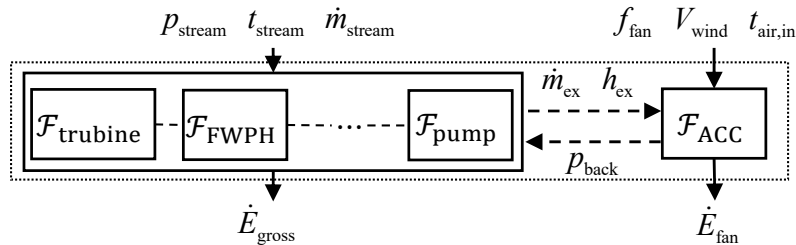

Figure 5. The cyber layer of turbine subsystem and the inputs for calculating $\dot{E}_{\text{gross}}$ and $\dot{E}_{\text{fan}}$.

Here, only the modeling of $\mathcal{F}_{\text{ACC}}$ by machine learning (support vector regression) is discussed in detail. Air cooling condenser aims at removing heat from the exhaust steam. For ACC controlling, the main actuator is the AC asynchronous motors for adapting the air-fan frequency. There are sensors recording ambient temperature and pressure, ambient wind speed and direction, inlet and outlet thermodynamic state of the ambient air and exhaust steam, etc. The key issue is to accurately calculate the heat transfer between exhausted steam and cooling air under various conditions, e.g., air flowrate and temperature.

Step A1: Preprocessing of the historical data. One-month operation data, which was recorded by Supervisory Information System (SIS) every 60 s, is employed and only quasi-steady state points are selected. The attributes for modeling p_{back} are selected according to the physical knowledge, including f_{fan}, $t_{\text{air,in}}$, V_{wind}, $\dot{m}_{\text{ex}}$ and h_{ex}. Data reconciliation is also applied to enhance the data quality.

Step A2: Regression model establishment. The traditional physical equations for modeling p_{back} (e.g., [38,39]) lack the consideration of ambient wind and fouling. In this case, the ε-SVR method [40] is applied for training the back pressure p_{back}:

$$p_{\text{back}} = \mathcal{F}_{\text{ACC}}\left(f_{\text{fan}}, t_{\text{air,in}}, V_{\text{wind}}, \dot{m}_{\text{ex}}, h_{\text{ex}}\right), \tag{4}$$

where the input attributes, f_{fan}, $t_{\text{air,in}}$, V_{wind}, $\dot{m}_{\text{ex}}$ and h_{ex} are fan frequency, ambient temperature, wind speed, the mass flowrate and enthalpy of the exhaust steam.

Step A3: Offline validation. 90% of the selected samples are employed for training, while the remaining 10% is used to validate the accuracy of the trained SVR model. The comparison between the model prediction and measured values for these 10% samples is illustrated in Figure 6. Around full load operation, the fan frequency reaches its maximum and the ambient temperature is the major driving force for such a control, since around that load the back pressure increases with the increase in the ambient temperature. For instance, under full-load conditions with an ambient temperature higher than 27 °C, the back pressure is higher than 20 kPa even if the air fans run at full speed. Under partial-load condition or with a low ambient temperature, however, it is much easier to maintain a lower back pressure by using a relatively low fan frequency, which potentially reduces the total fan power and enhances the overall plant performance. The validation in Figure 6 illustrates that the root means square error (RMSE) between the actual value and the prediction of back pressure is around 0.8 kPa under different operating conditions.
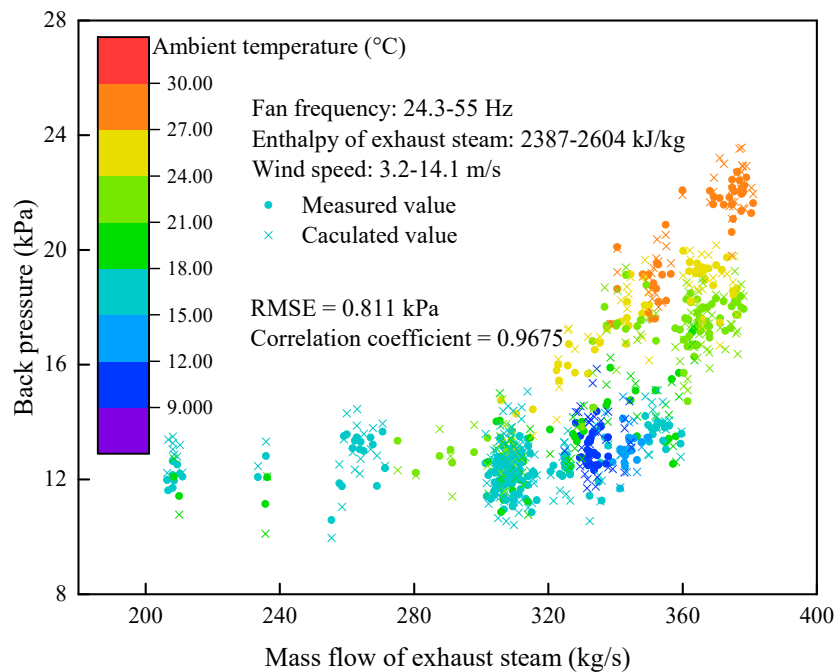

Figure 6. The validation between the calculated and the measured values (also discussed in [41]).

3.2.2. The Online Phase and Decision Making of Optimal Air-Fan Frequency

The online phase continuously executes the steps B1-B3 for real-time performance analysis and decision making. Preprocessing of the real-time data (Step B1) is similar to preprocessing of the historical data (step A1) and the corresponding input attributes are shown in Figure 5. With these inputs and the regression function of Equation (4), $\dot{E}_{\text{fan}}$ and $\dot{E}_{\text{gross}}$ can be predicted under any boundary condition. For online analysis and decisions (step B2), the optimal controlling strategy of air fan frequency is an optimization problem and can be formulated as follows:

$$\max_{f_{\text{fan}}} \Delta\dot{E}_{\text{net}} = (\dot{E}_{\text{gross}} - \dot{E}_{\text{gross,0}}) - (\dot{E}_{\text{fan}} - \dot{E}_{\text{fan,0}})$$
$$= \mathcal{F}(f_{\text{fan}}, t_{\text{air,in}}, V_{\text{wind}}, p_{\text{stream}}, t_{\text{stream}}, \dot{m}_{\text{stream}}), \tag{5}$$

$$\text{s.t. } f_{\text{fan}} \in [0, 55] \text{ (Hz)}, \tag{6}$$

where the term $\Delta\dot{E}_{net}$ is the net power gain, which should be maximized. $\dot{E}_{gross,0}$ and $\dot{E}_{fan,0}$ are the reference power with a fan frequency of 35 Hz. The terms p_{stream}, t_{stream} and $\dot{m}_{stream}$ represent the thermodynamic state of the streams for calculating turbine performance, feedwater preheaters, pumps, etc.

Figure 7 shows the optimal fan frequency and the corresponding net power gain $\Delta\dot{E}_{net}$ under different load factors. Note that the input values of $t_{air,in}$ and V_{wind} are set as constant, and the p_{stream}, t_{stream} and $\dot{m}_{stream}$ are set as the design values in each load factor. Figure 7 illustrates that the optimal fan frequency increases with the increasing in the load factor. When the load factor reaches around 90%, the fan frequency has reached the maximum allowed frequency (55 Hz). Note that the optimal controlling strategy can also be calculated for any given $t_{air,in}$, V_{wind}, p_{stream}, t_{stream} and $\dot{m}_{stream}$.
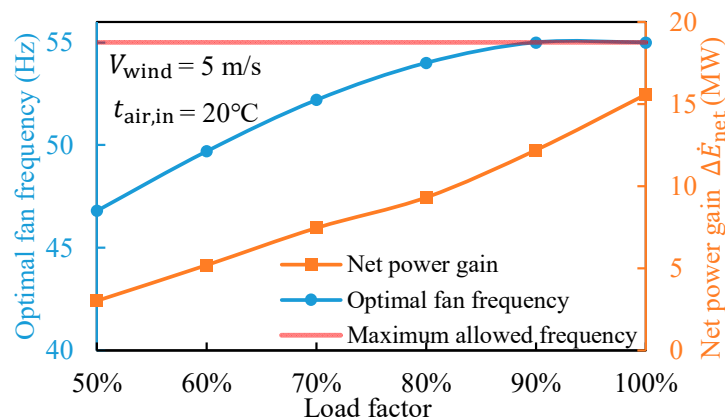

Figure 7. Schematic of the optimal control of air fan frequency based on the cyber layer.

A further comparison of actual and the optimal controlling strategy at the conditions (load factor 59.6%, air inlet temperature 23.8 °C) is given in Table 1. It is found that a higher f_{fan} of 49.0 Hz is preferred compared to the actual applied frequency of 37.1 Hz to enhance the heat transfer effect, which is mostly due to the performance degradation of ACC after years of running. With the optimal fan frequency, although the $\dot{E}_{fan}$ increases from 3.2 MW (37.1 Hz) to 7.3 MW (49.0 Hz) when reaching the optimal condition, the gross power output increases more than the increment of power consumption.

Table 1. A comparison between the real operation value and optimal value.

	Load factor	$t_{air,in}$	V_{wind}	f_{fan}	p_{back}	$\dot{E}_{gross}$	$\dot{E}_{fan}$	$\dot{E}_{gross} - \dot{E}_{fan}$
Real	59.6%	23.8 °C	6.9 m/s	37.1 Hz	15.15 kPa	421.3 MW	3.2 MW	418.1 MW
Optimal	59.6%	23.8 °C	6.9 m/s	49.0 Hz	10.98 kPa	429.6 MW	7.3 MW	422.3 MW

The trained model of $\mathcal{F}_{ACC}$ can be kept as long as the boundary conditions of newly recorded data have already been considered in the training data. However, the $\mathcal{F}_{ACC}$ model should be rebuilt with the corresponding record data, when the boundary conditions have been changed dramatically, e.g., in winter, one or more columns of air-fan clusters will be stopped for anti-freezing and the ambient temperature is decreases significantly (even out of the temperature range covered by the currently-employed data).

4. Discussion: The Combination of Mathematical Algorithms and Physical Knowledge

With data mining technique and big data in the cloud, the CPS enables performance analysis and the feedback of the controlling strategy in a quantitative manner from the cyber layer. However, there are some drawbacks associated with the mathematical algorithms, where physical knowledge is needed as the complementary.

When establishing a complex model, mathematical algorithms with, e.g., gray relation analysis, correlation coefficients and fuzzy rough sets, are often used for selecting the input attributes. However, it is worth noting that once the target is clear, the related attributes are determined, regardless which algorithms are used. In this case, the most proper algorithm for calculating the related input attributes can be chosen according to the prior physical knowledge. For example, approach with correlation coefficient does not consider the combination of the attributes, which makes it not proper when there is high interaction between the input attributes. The method with fuzzy rough sets considers the combination of the attributes, which indicates its applicability for the dataset with classification characteristics. For the models with well-established physical knowledge or empirical equations, necessary attributes can be directly determined. A collection of literature with different methods for selecting the input attributes is listed in the 3rd column of Table 2.

The data-driven models are mostly employed as a black-box, which hardly illustrates the physical relationships behind the inputs and outputs, since machine learning algorithms rely highly on the data set rather than the mass and energy balance or constraint equations. A review of recent modeling methods for energy systems is listed in the 4th column of Table 2. Based on these experience, it is advised that machine learning algorithm should be used for a complex thermodynamic process, of which the internal characteristics cannot be straightforwardly revealed by a set of equations. A hybrid model combining machine learning and thermodynamics is suggested for modeling the components in thermal power plants.

Table 2. A review of recent data-driven modeling for components or subsystems in energy systems.

Authors	Application Field	Method for Selecting Attributes	Method for Building the Regression Model
Jiang et al. [42]	Modeling feedwater preheater and extraction steam pipe	Component physics	Dominant factor modeling (Belongs to physical modeling)
Li et al. [43]	Modeling air cooling-condenser	Component physics	Physical modeling
Rossi et al. [44]	Determining the baseline energy consumption of combined heat and power plant	System physics	A comparison between physical model and neural network
Naik [45]	Modeling heat rejection capacity of air cooling condenser	Variance analysis	Correlation-based models (similar to linear regression)
Azadeh [46]	Electrical power generated from thermal power plants	Knowledge based	Adaptive-network-based fuzzy inference system (ANFIS)
Du et al. [47]	Modeling direct air cooling condenser of coal-fired power plant	Grey correlation degree	Neural network
Hernández et al. [48]	Modeling Water purification process, single-stage heat transformer and compressor.	Component physics and experiment	Neural network
Yoo et al. [49]	Modeling loss-of-coolant-accident situation of Nuclear power plants	Correlation coefficient (Pearson)	Support vector regression
Wang et al. [50]	Modeling coal consumption rate of coal-fired power plant	Fuzzy rough set method	Support vector regression
Li et al. [41]	Modeling air cooling-condenser	Component physics	Support vector regression
Xu et al. [51]	Modeling power supply and coal consumption rate of coal-fired power plant	Grey correlation degree	Particle Swarm Optimization based Support Vector Machine
Capozzoli et al. [52]	Energy fault detection analysis for a cluster of buildings	Knowledge based	A combination of Classification and Regression Trees (CART) and Artificial neural networks
Yan et al. [53]	Modeling ground source heat pump	Correlation coefficient (Pearson)	A comparison between Support vector machine, neural network and decision tree

5. Conclusions

In this paper, we introduced a general framework and the standard procedure for applying CPS to thermal power plants. According to the characteristics of the data recorded for the considered

thermal power plants, the proposed CPS is designed with offline modeling phase and online analysis phase. This work concerns the basic modeling and connection of each component in the cyber layer, and performs a case study to further explain the details of each involved step. The idea of hybrid modeling combining physical knowledge with data mining techniques is highlighted.

The case study focuses on the application of CPS to investigate the optimal controlling strategy of ACC in an air-cooled supercritical thermal power plant with the proposed data-driven approach. Prior knowledge of the ACC system is needed for the selection of the input attributes and for the enhancement of the data quality. For the case study, fan frequency, ambient temperature, wind speed, the mass flowrate and enthalpy of the exhaust steam are selected as the input attributes. The data preprocessing is performed with data reconciliation method by combining mass and energy balance equations. Then, the off-design model to predict the back pressure under different conditions is established with SVR. The RMSE of the back-pressure prediction model is only 0.81 kPa, which is accurate enough power plant application. Finally, the optimal fan frequency under various boundary condition is acquired for online application. The CPS model gives a quantitative relationship between the optimal fan frequency and the boundary conditions rather than a qualitative controlling strategy according to the operators' experience. The proposed framework is also capable of the performance analysis or decision making of any other component in thermal power plants.

Author Contributions: Xiaoen Li, Ligang Wang, Zhiping Yang, Ningling Wang and Yongping Yang conceived and designed the procedure for the application of the cyber physical system. Xiaoen Li performed the case study. Xiaoen Li and Ligang Wang analyzed the results and wrote the paper. Qing Wei contributed to the design of the figures.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kumar, R. A critical review on energy, exergy, exergoeconomic and economic (4-E) analysis of thermal power plants. *Eng. Sci. Technol. Int. J.* **2017**, *20*, 283–292. [CrossRef]
2. Wang, L.; Yang, Y.; Dong, C.; Morosuk, T.; Tsatsaronis, G. Multi-objective optimization of coal-fired power plants using differential evolution. *Appl. Energy* **2014**, *115*, 254–264. [CrossRef]
3. Yang, Y.; Wang, L.; Dong, C.; Xu, G.; Morosuk, T.; Tsatsaronis, G. Comprehensive exergy-based evaluation and parametric study of a coal-fired ultra-supercritical power plant. *Appl. Energy* **2013**, *112*, 1087–1099. [CrossRef]
4. Wang, L.; Wu, L.; Xu, G.; Dong, C.; Yang, Y. Calculation and analysis of energy consumption interactions in thermal systems of large-scale coal-fired steam power generation units. *Proc. Chin. Soc. Electr. Eng.* **2012**, *32*, 9–14.
5. Wang, L.; Yang, Y.; Morosuk, T.; Tsatsaronis, G. Advanced Thermodynamic Analysis and Evaluation of a Supercritical Power Plant. *Energies* **2012**, *5*, 1850–1863. [CrossRef]
6. Huang, S.; Li, C.; Tan, T.; Fu, P.; Wang, L.; Yang, Y. Comparative Evaluation of Integrated Waste Heat Utilization Systems for Coal-Fired Power Plants Based on In-Depth Boiler-Turbine Integration and Organic Rankine Cycle. *Entropy* **2018**, *20*, 89. [CrossRef]
7. Khaitan, S.K.; McCalley, J.D. Design techniques and applications of cyberphysical systems: A survey. *IEEE Syst. J.* **2015**, *9*, 350–365. [CrossRef]
8. Arghandeh, R.; Meier, A.V.; Mehrmanesh, L.; Mili, L. On the definition of cyber-physical resilience in power systems. *Renew. Sustain. Energy Rev.* **2016**, *58*, 1060–1069. [CrossRef]
9. Xie, Y.; Zeng, G.; Ryo, K.; Xie, G.; Dou, Y.; Zhou, Z. An optimized design of CAN FD for automotive cyber-physical systems. *J. Syst. Archit.* **2017**, *81* (Suppl. C), 101–111. [CrossRef]

10. Jezewski, J.; Pawlak, A.; Horoba, K.; Wrobel, J.; Czabanski, R.; Jezewski, M. Selected design issues of the medical cyber-physical system for telemonitoring pregnancy at home. *Microprocess. Microsyst.* **2016**, *46*, 35–43. [CrossRef]

11. Liu, H.; Sun, D.; Liu, W. Lattice hydrodynamic model based traffic control: A transportation cyber–physical system approach. *Phys. A: Stat. Mech. Appl.* **2016**, *461* (Suppl. C), 795–801. [CrossRef]

12. Wan, Y.; Cao, J.; Zhang, S.; Tu, G.; Lu, C.; Xu, X.; Li, K. An integrated cyber-physical simulation environment for smart grid applications. *Tsinghua Sci. Technol.* **2014**, *19*, 133–143.

13. Saber, A.Y.; Venayagamoorthy, G.K. Efficient utilization of renewable energy sources by gridable vehicles in cyber-physical energy systems. *IEEE Syst. J.* **2010**, *4*, 285–294. [CrossRef]

14. Al Faruque, M.A.; Ahourai, F. A model-based design of cyber-physical energy systems. In Proceedings of the 2014 19th Asia and South Pacific, Design Automation Conference (ASP-DAC), Singapore, 20–23 January 2014; IEEE: Piscataway, NJ, USA, 2014; pp. 97–104.

15. Cheng, Z.; Shein, W.W.; Tan, Y.; Lim, A.O. Energy efficient thermal comfort control for cyber-physical home system. In Proceedings of the 2013 IEEE International Conference on Smart Grid Communications (SmartGridComm), Vancouver, BC, Canada, 21–24 October 2013; IEEE: Piscataway, NJ, USA, 2013; pp. 797–802.

16. Schmidt, M.; Moreno, M.V.; Schülke, A.; Macek, K.; Mařík, K.; Pastor, A.G. Optimizing legacy building operation: The evolution into data-driven predictive cyber-physical systems. *Energy Build.* **2017**, *148* (Suppl. C), 257–279. [CrossRef]

17. Hamdaoui, Y.; Maach, A. Energy Efficiency Approach for Smart Building in Islanding Mode Based on Distributed Energy Resources. In Proceedings of the International Conference on Advanced Information Technology, Services and Systems, Tangier, Morocco, 14–15 April 2017; Springer: Berlin, Germany, 2017; pp. 36–49.

18. Maffei, A.; Srinivasan, S.; Meola, D.; Palmieri, G.; Iannelli, L.; Holhjem, Ø.H.; Mafafioti, G.; Mathisen, G.; Glielmo, L. A Cyber-Physical Systems Approach for Implementing the Receding Horizon Optimal Power Flow in Smart Grids. *IEEE Trans. Sustain. Comput.* **2017**. [CrossRef]

19. Gomes, I.L.R.; Pousinho, H.M.I.; Melíco, R.; Mendes, V.M.F. Bidding and Optimization Strategies for Wind-PV Systems in Electricity Markets Assisted by CPS. *Energy Procedia* **2016**, *106* (Suppl. C), 111–121. [CrossRef]

20. Yamanishi, K.; Takeuchi, J.-I. Discovering outlier filtering rules from unlabeled data: Combining a supervised learner with an unsupervised learner. In Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 26–29 August 2001; ACM: New York, NY, USA, 2001; pp. 389–394.

21. Fawzy, A.; Mokhtar, H.M.O.; Hegazy, O. Outliers detection and classification in wireless sensor networks. *Egypt. Inform. J.* **2013**, *14*, 157–164. [CrossRef]

22. Parimala, M.; Lopez, D.; Senthilkumar, N.C. A Survey on Density Based Clustering Algorithms for Mining Large Spatial Databases. *Int. J. Adv. Sci. Technol.* **2011**, *31*, 59–66.

23. Said, A.M.; Dominic, D.D.; Samir, B.B. Outlier Detection Scoring Measurements Based on Frequent Pattern Technique. *Res. J. Appl. Sci. Eng. Technol.* **2013**, *6*, 1341–1347. [CrossRef]

24. Maciá-Pérez, F.; Berna-Martinez, J.V.; Fernández Oliva, A.; Abreu Ortega, M.A. Algorithm for the detection of outliers based on the theory of rough sets. *Decis. Support Syst.* **2015**, *75* (Suppl. C), 63–75. [CrossRef]

25. Mohamed, M.S.; Kavitha, T. Outlier Detection Using Support Vector Machine in Wireless Sensor Network Real Time Data. *Int. J. Soft Comput. Eng.* **2011**, *1*, 81–86.

26. Nam, H.; Sugiyama, M. Direct density ratio estimation with convolutional neural networks with application in outlier detection. *IEICE Trans. Inf. Syst.* **2015**, *98*, 1073–1079. [CrossRef]

27. Rigamonti, M.; Baraldi, P.; Zio, E.; Alessi, A.; Astigarraga, D.; Galarza, A. Identification of the degradation state for condition-based maintenance of insulated gate bipolar transistors: A self-organizing map approach. *Microelectron. Reliab.* **2016**, *60* (Suppl. C), 48–61. [CrossRef]

28. Wang, L.; Lampe, M.; Voll, P.; Yang, Y.; Bardow, A. Multi-objective superstructure-free synthesis and optimization of thermal power plants. *Energy* **2016**, *116*, 1104–1116. [CrossRef]

29. Wang, L.; Voll, P.; Lampe, M.; Yang, Y.; Bardow, A. Superstructure-free synthesis and optimization of thermal power plants. *Energy* **2015**, *91*, 700–711. [CrossRef]

30. Wang, L.; Yang, Y.; Dong, C.; Morosuk, T.; Tsatsaronis, G. Systematic Optimization of the Design of Steam Cycles Using MINLP and Differential Evolution. *ASME J. Energy Resour. Technol.* **2014**, *136*, 031601. [CrossRef]

31. Negri, E.; Fumagalli, L.; Macchi, M. A Review of the Roles of Digital Twin in CPS-based Production Systems. *Procedia Manuf.* **2017**, *11* (Suppl. C), 939–948. [CrossRef]

32. Fu, P.; Wang, N.; Wang, L.; Morosuk, T.; Yang, Y.; Tsatsaronis, G. Performance degradation diagnosis of thermal power plants: A method based on advanced exergy analysis. *Energy Convers. Manag.* **2016**, *130*, 219–229. [CrossRef]

33. Wang, L.; Fu, P.; Wang, N.; Morosuk, T.; Yang, Y.; Tsatsaronis, G. Malfunction diagnosis of thermal power plants based on advanced exergy analysis: The case with multiple malfunctions occurring simultaneously. *Energy Convers. Manag.* **2017**, *148*, 1453–1467. [CrossRef]

34. Martín-Gamboa, M.; Iribarren, D.; Dufour, J. Environmental impact efficiency of natural gas combined cycle power plants: A combined life cycle assessment and dynamic data envelopment analysis approach. *Sci. Total Environ.* **2018**, *615*, 29–37. [CrossRef] [PubMed]

35. Wang, L. Thermo-Economic Evaluation, Optimization and Synthesis of Large-Scale Coal-Fired Power Plants. Ph.D. Thesis, Technical University of Berlin, Berlin, Germany, 2016.

36. Losing, V.; Hammer, B.; Wersing, H. Incremental on-line learning: A review and comparison of state of the art algorithms. *Neurocomputing* **2018**, *275* (Suppl. C), 1261–1274. [CrossRef]

37. Fontenla-Romero, O.; Pérez-Sánchez, B.; Guijarro-Berdiñas, B. An incremental non-iterative learning method for one-layer feedforward neural networks. *Appl. Soft Comput.* **2017**. [CrossRef]

38. Kumar, A.; Joshi, J.B.; Nayak, A.K.; Vijayan, P.K. A review on the thermal hydraulic characteristics of the air-cooled heat exchangers in forced convection. *Sadhana* **2015**, *40*, 673–755. [CrossRef]

39. Liu, J.; Hu, Y.; Zeng, D.; Wang, W. Optimization of an air-cooling system and its application to grid stability. *Appl. Therm. Eng.* **2013**, *61*, 206–212. [CrossRef]

40. Chang, C.-C.; Lin, C.-J. LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol. (TIST)* **2011**, *2*, 27. [CrossRef]

41. Li, X.; Wang, N.; Wang, L.; Kantor, I.; Robineau, J.-L.; Yang, Y.; Maréchal, F. A data-driven model for the air-cooling condenser of thermal power plants based on data reconciliation and support vector regression. *Appl. Therm. Eng.* **2018**, *129* (Suppl. C), 1496–1507. [CrossRef]

42. Jiang, X.; Liu, P.; Li, Z. A data reconciliation based framework for integrated sensor and equipment performance monitoring in power plants. *Appl. Energy* **2014**, *134*, 270–282. [CrossRef]

43. Li, X.; Wang, N.; Wang, L.; Yang, Y.; Maréchal, F. Identification of optimal operating strategy of direct air-cooling condenser for Rankine cycle based power plants. *Appl. Energy* **2018**, *209*, 153–166. [CrossRef]

44. Rossi, F.; Velázquez, D.; Monedero, I.; Biscarri, F. Artificial neural networks and physical modeling for determination of baseline consumption of CHP plants. *Expert Syst. Appl.* **2014**, *41*, 4658–4669. [CrossRef]

45. Naik, B.K.; Muthukumar, P. Empirical Correlation Based Models for Estimation of Air Cooled and Water Cooled Condenser's Performance. *Energy Procedia* **2017**, *109*, 293–305. [CrossRef]

46. Azadeh, A.; Saberi, M.; Anvari, M.; Azaron, A.; Mohammadi, M. An adaptive network based fuzzy inference system-genetic algorithm clustering ensemble algorithm for performance assessment and improvement of conventional power plants. *Expert Syst. Appl. Int. J.* **2011**, *38*, 2224–2234. [CrossRef]

47. Du, X.; Liu, L.; Xi, X.; Yang, L.; Yang, Y.; Liu, Z.; Zhang, X.; Yu, C.; Du, J. Back pressure prediction of the direct air cooled power generating unit using the artificial neural network model. *App. Therm. Eng.* **2011**, *31*, 3009–3014. [CrossRef]

48. Hernández, J.A.; Colorado, D.; Cortés-Aburto, O.; El Hamzaoui, Y.; Velazquez, V.; Alonso, B. Inverse neural network for optimal performance in polygeneration systems. *Appl. Therm. Eng.* **2013**, *50*, 1399–1406. [CrossRef]

49. Yoo, K.H.; Back, J.H.; Na, M.G.; Kim, J.H.; Hur, S.; Kim, C.H. Prediction of golden time using SVR for recovering SIS under severe accidents. *Ann. Nucl. Energy* **2016**, *94*, 102–108. [CrossRef]

50. Wang, N.; Zhang, Y.; Zhang, T.; Yang, Y. Data Mining-Based Operation Optimization of Large Coal-Fired Power Plants. *AASRI Procedia* **2012**, *3*, 607–612. [CrossRef]

51. Xu, J.; Gu, Y.; Chen, D.; Li, Q. Data mining based plant-level load dispatching strategy for the coal-fired power plant coal-saving: A case study. *Appl. Therm. Eng.* **2017**, *119*, 553–559. [CrossRef]

52. Capozzoli, A.; Lauro, F.; Khan, I. *Fault Detection Analysis Using Data Mining Techniques for a Cluster of Smart Office Buildings*; Pergamon Press, Inc.: Oxford, UK, 2015; pp. 4324–4338.
53. Yan, L.; Hu, P.; Li, C.; Yao, Y.; Xing, L.; Lei, F.; Zhu, N. The performance prediction of ground source heat pump system based on monitoring data and data mining technology. *Energy Build.* **2016**, *127*, 1085–1095. [CrossRef]