ORIGINAL RESEARCH Open access

A Data-Driven Adaptive Updating Framework for Integrated Energy System Components Using Knowledge Distillation

Lina Hassan[1*], Omar Khalaf[2], Reem Jaber[1]

Abstract

As societies move toward carbon neutrality, legacy power generation paradigms are giving way to integrated energy systems (IES), where model-guided dispatch proves indispensable when both generation and consumption fronts exhibit manifold uncertainties. The rise of AI-driven algorithms has overcome previous obstacles concerning model fidelity. That said, under circumstances marked by a high share of renewables, the load modulation of individual components calls for superior agility; accordingly, a component's mathematical formulation must keep pace with ceaselessly shifting operational states. Hence, pinpointing alterations in operating regimes and carrying out prompt model renewals stand as critical imperatives. The present work puts forward a knowledge-distillation-centered framework for IES component model construction and revision. At the heart of this framework lies model lightweighting, executed through a teacher-student arrangement inherent to knowledge distillation, thereby compacting elaborate neural network structures. Model renewal is initiated by principal component analysis. Additionally, this study probes the extent to which model inaccuracies arising from deferred revisions affect overall IES dispatching. Case investigations focus on pivotal IES elements, specifically coal-fired boilers and turbines. Findings demonstrate that the time required for model refreshment is slashed by 76.67% under the proposed scheme. Across changing operating settings and when set against two conventional benchmarks, the introduced method lowers average deviation figures by 12.61% and 3.49%, respectively, thereby heightening model adaptability. The vital importance of periodic component model revision is further substantiated, with evidence that a 1.00% mean squared error residing in a component's model may translate to a 0.075 MW power deviation. The presented methodology furnishes real-time, flexible support for data-driven modeling and renewal efforts within IES.

Keywords Data-driven, Component modeling, Adaptive update, Knowledge distillation, Variable operating conditions, Integrated energy system

*Correspondence:
Lina Hassan
lina.hassan@ju.edu.jo
[1] Department of Intelligent Energy Engineering, Faculty of Engineering, University of Jordan, Amman, Jordan
[2] Department of Artificial Intelligence and Smart Infrastructure, Faculty of Engineering, Jordan University of Science and Technology, Irbid, Jordan

Introduction

Background

The unrelenting outpouring of carbon dioxide, chiefly traceable to the combustion of fossil-based fuels, propels worldwide climatic disruption and escalating temperatures [1, 2]. For this reason, a cross-societal understanding prevails that the transformation of energy frameworks toward green and replenishable sources must be hastened [3, 4]. Taking China as a case in point, by the conclusion of June 2023, the installed capacity attributable to renewable sources had, for the first time, eclipsed that of coal- and gas-fired units, comprising roughly 48.80% of overall installed capacity [5]. Within this

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC BY-NC-SA 4.0), which permits non-commercial use, distribution, reproduction, and adaptation in any medium or format, provided that appropriate credit is given to the original author(s) and source, a link to the license is provided, and any changes made are indicated. Adapted material must be distributed under the same license.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

milieu, time-honored energy production is progressively morphing into integrated energy systems (IES), driven by the goals of assimilating larger quantities of renewable power, boosting both the effectiveness and ecological soundness of energy exploitation, and bringing down carbon discharges [6]. Broadly defined multi-energy arrangements, among them the multi-energy systems described by Mancarella [7], the standalone IES outlined by Wu *et al*. [8], the district-level IES presented by Ramachandra [9], and the virtual power plants set forth by Saboori *et al*. [10], have risen to prominence as key themes in scholarly research and industrial practice. Throughout this study, IES is delineated as the generation-oriented side of multi-energy systems, distinguished by added layers of complication in terms of spatial reach, temporal dynamics, entity variety, and operating contexts. Within this framing, the IES takes on a stakeholder identity, pursuing the highest possible return on assets. Consequently, dispatch decision-making anchored in an IES optimization construct carries critical weight.

A substantial body of research has tackled the coordinated dispatching of multi-energy systems. These inquiries span the treatment of uncertainty, the simultaneous optimization of multi-energy interplay, and the optimization involving several interested parties. To illustrate, Aghdam *et al*. [11] examined the uncertainty-conscious scheduling of electric-hydrogen configurations that incorporate virtual energy storage capacities. They minimized operational costs amid price ambiguity by applying stochastic optimization routines. Oskouei *et al*. [12] introduced a blended robust-stochastic optimization framework to address the unpredictability of electricity market tariffs and consumption needs in setups with combined heat and power generation and power-to-X storage. Mirzaei *et al*. [13] likewise turned to hybrid robust-stochastic optimization instruments to cope with the indeterminacy present in multi-energy systems. Mobasseri *et al*. [14] confronted the uncertainty optimization challenge for electricity-heat-hydrogen systems and constructed a model that draws upon Hong 2m+1, information gap decision theory, and scenario-centered methodologies. Li *et al.* [15] assembled a distributed collaborative optimization construct for multi-energy systems that is activated by events. These bodies of work deliver scientifically grounded models that underpin the realization of high-efficiency, low-carbon operation within IES.

That said, the precision of intra-system component representations leaves a deep imprint on the formulation and execution of dispatch-related decisions.

Literature review

Compact linear component characterizations are routinely adopted in present-day system-level optimization efforts. For example, Zheng *et al*. [16] employed energy efficiency ratios to represent IES component models covering combined heat and power, power-to-gas conversion, and compression equipment. Similarly, Cesena and Mancarella [17] advanced a linear approximation methodology tailored for nonlinear network equations, whereas Li and Xu [18] recast a nonlinear formulation into a linear equivalent to ease computational demands. Zhang *et al*. [19] portrayed internal machinery using an energy hub conceptualization, complete with immutable model parameters. While the conversion coefficient strategy renders component modeling more straightforward and heightens model execution speed, some sacrifice of exactness is unavoidable.

Apart from drawing on streamlined data-oriented models, segments of the research community lean on mechanistic, physics-grounded modeling pathways. Nittaya *et al*. [20] delved into the mechanical dynamics representation of post-combustion CO2 capture apparatus, while Sani *et al*. [21] harnessed mechanistic models for tanker trucks and cooling installations. Gabrielli *et al.* [22] offered a modeling structure for electrochemical conversion devices that springs from first-principles thermodynamic dynamic modeling and spawns reduced-order linear approximations. Ma *et al.* [23] pieced together a granular wind farm model incorporating free dissipation energy, resonant dissipation energy, and forced dissipation energy. That being the case, mechanistic modeling adds a heavy layer of complexity and presents obstacles to solving optimal dispatch decision problems.

To improve model utility, some scholars have broadened simple data models by integrating deep learning methodologies. Abbasimehr *et al*. [24] developed a demand forecasting approach for energy that relies on convolutional neural network architectures. Ajbar et *al*. [25] put together a predictive representation for the thermal efficiency metric of parabolic trough

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

collectors by leaning on artificial neural networks. Wang *et al.* [26] brought forward a recurrent neural network-based method for projecting the thermal energy usage rate of electric thermal storage heating units. These explorations make clear that data models rooted in neural network techniques can surface patterns from prior operational records, lifting the generalization power of the coefficient-based approach and trimming complexity when juxtaposed with classic mechanistic representations.

Yet, model trustworthiness is situation-dependent, conditional upon the steadiness of operating environments. As renewable electricity becomes more thoroughly woven into the grid and component loads undergo constant reconfiguration, fortifying model resilience to such oscillations becomes a formidable challenge. Kämper *et al*. [27] advanced a data-driven modeling modality for multi-energy systems that uses piecewise linear regression. Mannsor *et al.* [28] underscored the benefits of piecewise mapping functions when contrasted with models relying on non-piecewise fixed-cost parameters. Benzaama *et al*. [29] laid out a procedure for crafting thermal behavior models of residential buildings via piecewise affine autoregressive genetic identification. That said, these piecewise linear constructs remain offline in nature. Pushing the envelope, Qin *et al*. [30] subsequently rolled out an adaptive piecewise linear modeling scheme to flexibly reproduce the nonlinear operating signatures of diverse energy conversion devices. Rahbar *et al*. [31] boosted online optimization performance by appending constraints aimed at pulling the model into tighter congruence with the actual physical system, alongside improvements to the optimization process itself.

Various researchers have proposed rolling update paradigms grounded in model predictive control (MPC) that use error feedback correction. For instance, based on projected variations in power consumption, Labidi *et al*. [32] executed operational optimization of multi-energy district boiler installations through the MPC framework. *Vasilj et al.* [33] assembled nonlinear thermal models with algebraic representations for heating networks by means of nonlinear expressions and subsequently proposed a nonlinear MPC methodology to deliver operational optimization for multi-energy microgrids. Building on these foundations, Jimenez et *al.* [34] proposed a model-free technique that leverages deep reinforcement learning (DRL) to mitigate the effects of external uncertainty. Ye et *al*. [35] advanced a strategy for real-time autonomous energy administration within residential multi-energy systems, using the DRL paradigm. Yang *et al.* [36] formulated a DRL-centered island energy system construct through the prism of adaptive energy governance, one that proves capable of handling the variability introduced by renewable generation and load fluctuations. Because of its inherently dynamic reward structure, DRL offers a measure of resilience when confronted with the unpredictability of surrounding conditions. That being said, even though DRL has succeeded in delivering model-free optimization, it remains reliant on a static coefficient approach when representing IES component models. The work undertaken by Wen and Aziz [37] illuminates this particular shortcoming. The task of accommodating numerous constraints within DRL frameworks warrants further investigation. For these reasons, exploring how component models are built and subsequently refreshed under fluctuating operating regimes holds considerable practical value.

Beyond the techniques employed for refreshing, the twin challenges of deciding precisely when a model refresh ought to be triggered and how to properly appraise the resulting model loom large. Currently prevailing approaches can be grouped into those that hinge on prediction accuracy thresholds, those that rest upon specific indicator benchmarks, and those that carry out refreshment continuously without any form of pre-screening. Brauns *et al*. [38] designed a scheme wherein models for short-term vertical power flow forecasting undergo periodic retraining and refreshing, driven by the accumulation of condition-shift data. Wang *et al*. [39] outlined a model refresh protocol for rotor systems centered on an examination of parameter sensitivity. Zhang *et al.* [40] unpacked the refresh architecture and accompanying techniques from a multi-dimensional modeling standpoint, selecting the cutting tool of machine tools as their focal object. Their approach relied on physical data to drive model refreshes, working to narrow the gap between the physical artifact and its digital counterpart. In addition, Schreiber and Sick [41] laid out an offline approach to model refreshment and adjustment, bolstering model flexibility through a blend of model ensembling and fine-tuning. Huang *et al*. [42] embraced a strategy of direct, on-the-fly refreshing, employing a recursive least squares formulation equipped with a forgetting factor to continuously re-estimate the model. Yu *et al*. [43] crafted a real-time refresh scheme that combines an enhanced Gaussian particle filter with a Dirichlet process mixture model. The foregoing investigations offer partial guidance regarding refresh strategies. Nevertheless, scholarship focused squarely on the refresh procedures for component-level

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

models stays relatively sparse, particularly as applied to IES contexts. Hence, further inquiry remains necessary to settle on the most advantageous timing and appraisal protocols.

Motivation and contributions

As the review of prior work lays bare, the operational reality of IES—wherein a large share of renewable integration forces near-constant adjustment of component loads—progressively diminishes the fitness of component models. Although data-driven models based on neural networks can improve generalization and cut through complexity, they are less adjustable amid shifting operational conditions and are cumbersome to refresh. Accordingly, the ambition of this study is to seek out a means of preserving model precision while simultaneously enabling rapid, web-based refreshes.

Two threads within computer science—meta-learning and incremental learning—open promising paths toward adaptive model refreshment. Meta-learning equips models with the capacity to learn [44], and incremental learning retains awareness of previously encountered data through knowledge extraction [45]. Investigations by Sarmas *et al*. [46] and Wu *et al*. [47] validated the edge that online incremental learning holds in the forecasting of energy demand. However, when standard neural architectures are trained, they tend to be computationally demanding, require extensive memory footprints and impede the real-time renewal of models.

As a technique for model compression, knowledge distillation permits the transfer of learned representations from cumbersome architectures into leaner surrogates [48]. It has already been applied in industrial settings to reduce the complexity and computational expense of predictive models. By way of illustration, Li *et al*. [49] grappled with the power consumption problem inherent to additive manufacturing workflows by teasing out latent information or highly salient attributes from energy-related data streams and slashing the complexity and computational overhead of prediction models through knowledge distillation. Ai *et al*. [50] assembled a hybrid learning framework dedicated to feature representation, one that draws jointly upon knowledge distillation and supervised contrastive learning for industrial process oversight. Even so, the deployment of this technique within energy systems remains rare. Considering the dynamic and nonlinear character of IES, elaborate models yield excellent predictive fidelity yet prove impractical because of the oversized demands they place on computing cycles and time. Knowledge distillation supplies an answer by migrating the learned knowledge housed within complex models into simpler ones, thus paring back computational load while safeguarding predictive accuracy.

Beyond this, although the knowledge distillation route settles the mechanism by which refreshes are carried out, the timing of such refreshes stays an open problem. Thus, the current work further pursues approaches for sensing when operating conditions have shifted. Principal component analysis (PCA) is a dimensionality reduction instrument of wide application, prized for its ability to preserve key informational content. It has been pressed into service for fault surveillance in industrial contexts. Elsamanty *et al*. [51] directed PCA toward the diagnostic challenges of rotating machinery faults. Swiercz and Mroczkowska [52] probed the suitability of PCA-based methods for the prompt identification of leaks within steam boiler piping circuits at thermal power facilities. This line of work sparks the notion of applying operational surveillance tools to gauge the necessity for model refreshment. Within the tangled structure of IES, PCA holds the capacity to meaningfully collapse data dimensionality and streamline model intricacy. By treating standard operational profiles as baseline references and divergent profiles as off-nominal, streams of operational monitoring data can be exploited to detect instances where the model no longer aligns.

To synthesize, present-day component models have absorbed—to varying extents—approaches such as piecewise linear formulations and rolling bias correction to dampen inaccuracies under shifting operational profiles. Several difficulties persist unsolved: (1) The temporal targeting of refreshes: The determination of when a refresh should occur is seldom addressed in existing constructs. Refreshes tend to follow a fixed cadence, which introduces lag, or are gated by thresholds, which can provoke excessively frequent renewals and inflate the appetite for computational capacity. (2) The mechanism underpinning refreshes: Piecewise linearization succeeds in decomposing operating envelopes under specific conditions, but it neglects the coupled, multivariable essence of real operating regimes; error correction schemes are only

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

workable when the deviation stays modest; and multi-condition models currently on offer remain products of offline training. (3) The vetting of refreshed models: The assurance that a post-refresh model constitutes an improvement over its predecessor suffers from a deficit of rigorous, empirically grounded adjudication.

Table 1 sums up the prevailing state of inquiry into component modeling and refresh methodologies.

Table 1. Comparison of research status.

Reference	Component Modeling Approach	Research Scope			Research Subject/Area	Research Limitations
		Real-Time Updating	Updating Technique	Model Implementation		
[16, 18]	Coefficient-based method	×	×	×	Energy system	Insufficient model accuracy
[20, 22]	Mechanism modeling method	×	×	×	Energy system	Excessive model complexity
[24, 26]	Neural networks	×	×	×	Energy system	Models are offline and unable to adapt to operating conditions
[27, 29]	Piece-wise linear method	×	×	×	Energy system	Models are offline and unable to adapt to operating conditions
[30]	Adaptive piece-wise linear method	√	√	×	Energy system	Inability to manage multi-variable coupling
[31]	Coefficient-based method	√	√	×	Energy system	Lagging due to reliance on deviation correction
[32, 33]	MPC	√	√	×	Energy system	Lagging due to reliance on deviation correction
[38]	LSTM	√	√	×	Energy system	Frequent and delayed updates triggered by threshold-based judgments
[42]	LSTM	√	√	×	Machine tool feed systems	Frequent and delayed updates triggered by threshold-based judgments
[34, 37]	DRL	×	√	×	Energy system	Utilization of fixed coefficient methods for component models

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

[41]	Neural networks and Bayesian linear regression	×	√	√	Energy system	Combining multiple model complex models and not online adaptation
Proposed method	Neural networks	√	√	√	Energy system	/

This investigation puts forward a method for modeling and refreshing components within IES. The contributions it delivers are threefold:

(1) A data-driven scheme for modeling and refreshing IES components is advanced, bringing together the triaging of refresh criteria, instantaneous model renewal, and model selection under one unified banner.

(2) A knowledge-distillation-grounded methodology is harnessed to enable real-time renewal of intricate models. At the same time, a PCA-anchored component condition detection construct is erected to resolve the dilemma of refresh timing.

(3) Building upon a quantitative dissection of how growing renewable penetration shapes boiler load behavior within IES, the influence that model renewal exerts upon aggregate dispatch outcomes is scrutinized.

Materials and Methods

Illustrated in **Figure 1** is the sequence of steps constituting the component modeling and refresh approach built upon knowledge distillation that this investigation puts forward. The workflow begins with pre-training a foundational knowledge distillation setup, which pairs a high-capacity teacher neural network with an elementary student neural network. The teacher, being a computationally heavy neural design, achieves high accuracy under constant operating regimes—something documented in the work of Sakiewicz *et al*. [53]—but encounters difficulties regarding flexibility and imposes steep computational costs [54]. The student, in stark contrast, embodies a far more parsimonious neural design, sacrificing a measure of precision in exchange for heightened adjustability and frugal resource demands.

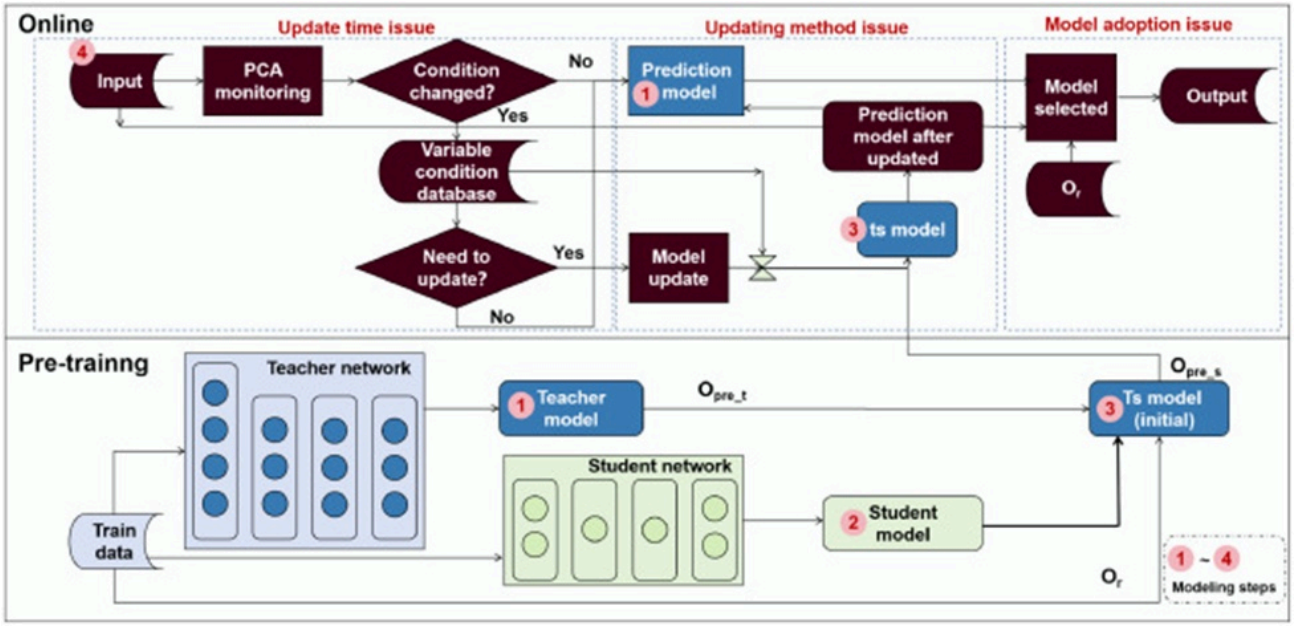


Figure 1. Component modeling and updating method based on knowledge distillation.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

Outputs projected by the teacher network on the training corpus are routed into the student network alongside the actual measured outputs, after which an additional round of tuning is executed to arrive at a fresh model—the ts model—that soaks up and condenses the expertise held by the teacher. The ts model inherits the structurally lean footprint of the student yet operates at a level of accuracy exceeding that of the student alone. When placed into live service, a PCA-driven module continuously monitors the operational state, flagging any drifts that are then folded into subsequent ts model updates. The refreshed model and the incumbent model then run in parallel, issuing predictions; if the refreshed variant demonstrates persistently better accuracy over a sufficiently lengthy interval, it displaces the older component model.

Operating condition detection method based on PCA

PCA falls within the family of multivariate statistical process surveillance techniques and is widely used for fault monitoring throughout the process sector [55]. The method functions by compressing the dimensionality of multivariable, high-dimensional data, mapping it onto a lower-dimensional manifold, and converting it into single-variable statistical metrics. Energy production plants and process industry installations share common ground, both being elaborate, multivariable, intertwined operations. Taking a cue from the domain of fault detection, this study repurposes PCA for discerning component operating states, employing it as a preliminary screen to judge whether a model refresh has become necessary.

Normalization of the input data must first be carried out.

$$\mathbf{X} = (\mathbf{X}_0 - \mathbf{X}_M)/\mathbf{X}_V \quad (1)$$

In the notation above,$\mathbf{X}_0$ stands for the original data matrix comprising n observations and m variables; $\mathbf{X}_m$ denotes the mean vector of $\mathbf{X}_0$;$\mathbf{X}_V$ indicates the variance of $\mathbf{X}_0$; **X** is the resulting normalized data matrix.

The covariance matrix is then derived using **Eq. 2.**

$$\mathbf{Z} = \mathbf{X}\mathbf{X}^T/(n-1) \quad (2)$$

Within this step, Z represents the covariance matrix.

Thereafter, the eigenvalues and their associated eigenvectors are computed.

$$\mathbf{Z}\vec{\mathbf{v}} = \lambda\vec{\mathbf{v}} \quad (3)$$

In the given expression, λ refers to the eigenvalues and $\vec{\mathbf{V}}$ to the eigenvectors.

The eigenvalues are ranked in descending order, the leading k components are chosen such that their cumulative contribution exceeds d (a user-adjustable parameter), and the accompanying k eigenvectors are gathered to form the loading matrix P.

$$\mathbf{X} = \mathbf{T}\mathbf{P}^T + \mathbf{E} \quad (4)$$

Here, T signifies the score matrix, while E corresponds to the residual matrix.

Two statistical indices, namely $\mathbf{T}^2$ and SPE, are employed to track the process condition.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

$$T^2 = X_n^T P^T \lambda^{-1} P X_n \quad (5)$$

$$T^2_{lim} = k\frac{n-1}{n-k}F_\alpha(k, n-k) \quad (6)$$

For this expression, $\mathbf{X}_n$ designates the newly arriving data batch; 1-α establishes the confidence band, with α typically fixed at 0.05; F stands for the F distribution; and $\mathbf{T}^2_{lim}$ marks the upper control threshold for T2.

$$SPE = X_n^T\left(1 - PP^T\right)X_n \quad (7)$$

$$SPE_{lim} = \theta_1\left[\frac{c_\alpha h_0\sqrt{2\theta_2}}{\theta_1} + 1 + \frac{\theta_2 h_0(h_0-1)}{\theta_1^2}\right]^{1/h_0} \quad (8)$$

$$\theta_s = \sum_{j=k+1}^{m}\lambda_j^s (s = 1, 2, 3) \quad (9)$$

$$h_0 = 1 - \frac{2\theta_1\theta_3}{3\theta_2^2} \quad (10)$$

Within the formula, $\mathbf{SPE}_{lim}$ defines the upper control threshold for SPE, and ca is the critical value drawn from the standard normal distribution.

Component prediction model based on knowledge distillation

To enable the rapid online renewal of intricate models, this research turns to a strategy rooted in knowledge distillation. The central tenet of knowledge distillation is to build a lightweight model and train it using supervisory cues extracted from a higher-performing counterpart, thereby securing a distilled model of notable quality, with the broader ambition of elevated capability alongside maintained accuracy. In live application, the distilled model issues predictions, thus delivering on the promise of model lightness.

Knowledge distillation unfolds across three primary forms: offline distillation, online distillation, and self-distillation [56]. The contrasts separating these three knowledge distillation modalities are laid out in **Figure 2**. Offline distillation typically follows a two-stage path: the teacher network is trained beforehand, and the student network is trained afterwards. The descriptor “offline” underscores the requirement to build a fairly elaborate teacher network outside the live loop, which is subsequently tapped to educate the student network through knowledge extraction. Online distillation, on the other hand, calls for just one stage, with the teacher and student networks being trained in tandem. Self-distillation, a special offshoot of online distillation, erases the distinction between teacher and student, necessitating only a single model; in this setup, the student draws lessons from the outputs it itself generates.

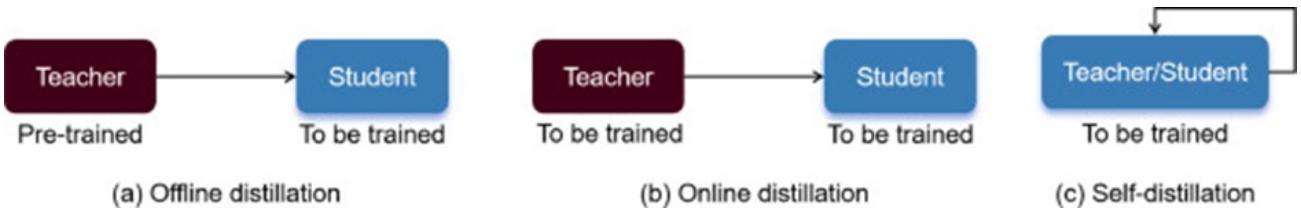


Figure 2. Three knowledge distillation methods.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

The offline distillation variant—the most prevalent—is selected for this study, and the traits and relative merits of the three modalities are compiled in **Table 2**. From a deployment perspective, offline distillation is the most straightforward. The teacher may encompass a suite of models trained via disparate software environments, possibly housed on separate machines. On the accuracy front, online distillation holds an advantage over its offline counterpart by refreshing both teacher and student simultaneously, a feature that can lift precision. Yet this gain brings added computational cost, working against the aim of model compression. Regarding domain suitability, online distillation aligns more naturally with knowledge fusion or contexts such as multi-modal and cross-domain applications [57]. Self-distillation, for its part, predominantly serves to strengthen the performance of models aimed at downstream tasks rather than to fulfill model compression or transfer learning roles.

Table 2. Comparison of three distillation methods.

Distillation approach	Key feature	Strength	Typical application area
Offline distillation (This study)	Requires pre-training	Flexibility in choosing large pre-trained models as teachers; Simplicity and low resource consumption	Model compression
Online distillation	Teacher and student models can be updated simultaneously	Removes the need for pre-training; Enables mutual learning among models	Knowledge fusion
Self-distillation	Able to transfer deep-layer information to shallow layers	Eliminates the need for a separate teacher model	Model performance enhancement

Given the above considerations, offline distillation is selected here to realize model compaction and to raise the efficiency of model renewals during live operation. Additionally, in the current landscape of IES operational optimization, pre-trained teacher models already exist for the majority of components, further validating the choice of offline distillation in this work.

Pre-training model construction

The pre-training phase hinges upon first schooling an intricate neural network and then approximating it with a leaner stand-in, ultimately yielding a model that is architecturally sparse yet rivals the original in predictive fidelity.

Teacher model

Before assembling the teacher network, the dataset must be prepared, which includes addressing outliers and imputing missing entries. Moreover, the selected data should be normalized to temper the sway of features spanning dissimilar numerical ranges on model formation. During the build-out of the teacher, assorted neural designs can be tested to construct several candidate models, and the most suitable configuration should then be selected. Let us posit an elaborate neural network configured with NT hidden layers, each populated by Ni neurons (i = 1, 2, ..., NT).

Furnished with the input collection I and the paired output collection O, the elaborate neural network is put through training. This cycle includes both the forward computation pass and the error backpropagation pass. Backpropagation is the most common algorithm in neural network training; the gap between the desired output and the computed output is propagated backward layer by layer, guiding weight modifications until the error contracts within a tolerable margin [58].

The optimization target takes the form of **Eq. 11**.

$$L_t = \mathrm{MSE}\left(\mathbf{O}_{pre_t}, \mathbf{O}_t\right) \quad (11)$$

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

$$\mathrm{MSE}(\mathbf{y}_1, \mathbf{y}_2) = \frac{(\mathbf{y}_1 - \mathbf{y}_2)^2}{n} \quad (12)$$

In the definition above, Lt quantifies the loss sustained by the teacher network; MSE captures the average of the squared gaps between actual and forecasted values across all data points; $\mathbf{O}_{pre_t}$ signifies the prediction issued by the teacher model; Or denotes the observed ground truth; and y1, y2 constitute the two variable inputs to the MSE function.

The activation leaving neuron j housed in hidden layer m is expressed as:

$$\mathbf{O}_{m,j} = \sigma_m \left(\sum_{i=1}^{In} w_{i,j} x_i + b_j \right) \quad (13)$$

Within this equation, Omj corresponds to the output emitted by neuron j situated in hidden layer m; σm stands for the activation function employed by hidden layer m; In indicates the tally of input features arriving at the hidden layer; wi,j is the weight connecting input xi to neuron j of hidden layer m; xi is the i-th input directed to neuron j of hidden layer m; and bj marks the bias term belonging to neuron j of hidden layer m.

Weights are adjusted according to the following rule:

$$w_{i,j}(t+1) = w_{i,j}(t) + \eta[(1-\mu)D(t) + \mu D(t-1)] \quad (14)$$

$$D(t) = -\frac{\partial L_t}{\partial w_{i,j}(t)} \quad (15)$$

In this rule, η represents the learning rate, and μ denotes the momentum factor, constrained by $0 \le \mu < 1$.

The pseudo-code for the complete neural network procedure is given as **Algorithm 1.**

Executing **Algorithm 1** furnishes the fully trained teacher network.

Student model

When shaping the student network, the architecture must be stripped down as much as possible. A ceiling on permissible error should be defined, and the barest student structure that still respects this error bound should be fashioned. Let the student network be arranged with NS hidden layers (where NS ≤ NT), each housing Nj neurons (j = 1, 2,..., NS), with Nj ≤ Ni. Employing the very same input collection I and output collection O that were used to train the teacher, the preliminary student network is assembled following **Algorithm 1.**

The loss function that governs the student model is defined in **Eq. 16.**

$$L_s = \mathrm{MSE}(\mathbf{O}_{pre_s}, \mathbf{O}_t) \quad (16)$$

In the expression, Ls represents the loss incurred by the student network, and $\mathbf{O}_{pre_s}$ denotes the forecast delivered by the student model.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

Ts model (Model after knowledge distillation)

Harnessing the predictions cast by the teacher network and abiding by the overall loss formulation, the student network is updated to realize knowledge distillation, with the end product being the ts model.

The loss function tailored to the ts model is detailed in **Eq. 18.**

$$\mathbf{O}_s = (1-\beta)\mathbf{O}_{pre_t} + \beta\mathbf{O}_r \quad (17)$$

$$L = \mathrm{MSE}\left(\mathbf{O}_{pre_s}, \mathbf{O}_s\right) \quad (18)$$

Within this formulation, L corresponds to the aggregate loss; Os serves as the target value guiding the model update; and β acts as a tunable coefficient, termed the beta parameter.

In building the ts model, careful thought must be given to the setting of the beta parameter as well as to the scheduling of training iterations.

For any given IES component model, the feature variables tied to the component's output make up the aforementioned input collection I, and the component's output itself constitutes the collection O. The legacy predictive model before any refresh is typically an intricate one, labeled f1, which serves as the teacher; the structurally trimmed predictive model obtained through further training is f2, i.e., the student; and the model emerging from the knowledge distillation process is referred to as the ts model.

Online update strategy

The deliverable of knowledge distillation is the ts model. When put into live service, the freshly updated model fnew (which starts as the ts model) and the legacy model f1 both generate forecasts at the same time. Should the forecast error of the updated model, accumulated across a time window, drop below that of the f1 model, or should the error span exceed the acceptable boundary ε, then a model swap is enacted, replacing f1 with fnew.

The protocol that governs when a model refresh should be triggered is stated as follows:

$$\left(\sum\left(\left|\mathrm{f}_1(\mathbf{I}) - \mathbf{O}_r\right| > \left|\mathrm{f}_{new}(\mathbf{I}) - \mathbf{O}_r\right|\right)\right) > d_{t1} \quad (19)$$

$$d_{t2} < \left(\sum\left(\left|\mathrm{f}_1(\mathbf{I}) - \mathbf{O}_r\right| > \left|\mathrm{f}_{new}(\mathbf{I}) - \mathbf{O}_r\right|\right)\right) < d_{t1} \quad \text{and} \sum\left(\left|\mathrm{f}_1(\mathbf{I}) - \mathbf{O}_r\right| - \left|\mathrm{f}_{new}(\mathbf{I}) - \mathbf{O}_r\right|\right) > \varepsilon \quad (20)$$

Within the above, ε defines the acceptable boundary for the cumulative error range; dt1 and dt2 specify threshold durations for the accumulation period governing model refreshes, with dt2 < dt1.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

When faced with prolonged anomalous operating states, fresh input Inew and output Onew are drawn from the anomalous condition bank, and the ts model is fine-tuned. The fine-tuning procedure relies on **Algorithm 1**, with all other parameters—including the starting weights and bias terms—being carried over directly from the ts model. The loss function selected is the prediction error of the ts model alone.

Case Introduction

Notwithstanding the ongoing rebalancing of the energy portfolio and the diminishing relative contribution, coal-fired boilers retain an indispensable function in meeting China's energy and power needs. This owes to the inherent variability of renewable sources and the likelihood that coal burn in power stations may increase as electricity consumption climbs [59]. Attaining efficient and ecologically sound coal utilization is of particular consequence for the long-term sustainability of energy. As depicted in **Figure 3**, IES are set to emerge as the principal pathway for amplifying the effectiveness of traditional coal-fired boilers [60]. Inside the IES framework, boilers are charged with converting fossil-derived energy into steam destined for further transformation and use. Prudent manipulation of boiler load, minimization of energy wastage, elevation of the renewable energy capture rate, and a gradual energy transition all stand as pivotal elements in the reshaping of coal-fired power facilities. Given both the installed footprint and the enduring relevance of coal-fired boilers, the present work employs the coal-fired boiler situated within an IES as the demonstration case to verify the proposed component modeling and refresh methodology. The data stem from the actual operational logs of a 220.00 t/h high-temperature, ultra-high-pressure circulating fluidized bed boiler located in Zhejiang Province.

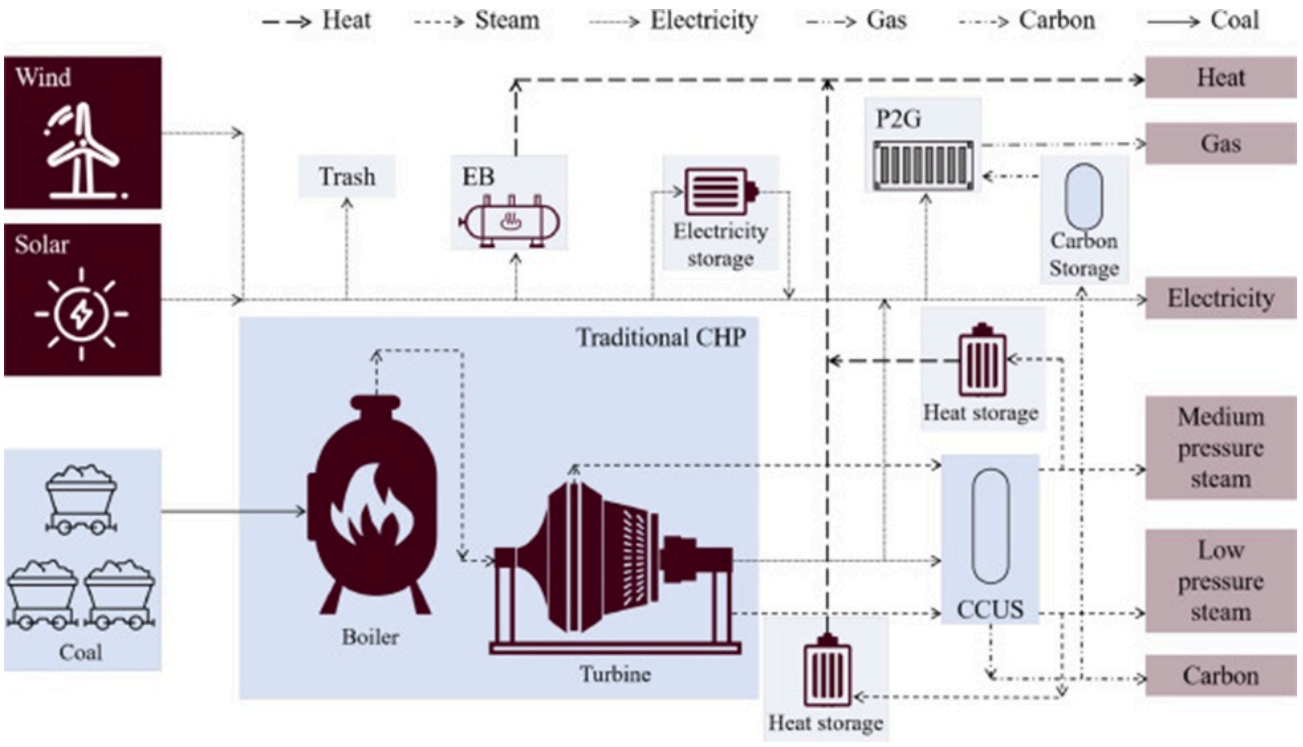


Figure 3. The key role of coal-fired boilers in future low-carbon IES.

In specific terms, the historical record stretches from March 10, 2023, to May 20, 2023, recorded at a one-minute cadence, amounting to 86,793 valid records in total. Boiler efficiency is shaped by the following factors: main steam temperature, main steam pressure, main steam flow rate, feedwater temperature, and coal properties (covering moisture content, ash content, volatile matter, and net calorific value on an as-received basis). The inputs and outputs of the boiler efficiency prediction model are tabulated in **Table 3.**

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

Table 3. Component prediction model input-output.

ID	Parameter type	Parameter	Unit
1	Output	Efficiency	%
2	Input	Main steam temperature	°C
3	Input	Main steam pressure	MPa
4	Input	Main steam flow rate	t/h
5	Input	Feed water temperature	°C
6	Input	Moisture	%
7	Input	Ash	%
8	Input	Volatile matter	%
9	Input	Net calorific value as received basis	J/kg

Alongside the boiler efficiency-related inputs and outputs catalogued in **Table 3,** a further set of parameters, as detailed in **Table 4**, is incorporated to serve as surveillance variables for tracking boiler operating states. A descriptive summary of the dataset utilized throughout this investigation is furnished in **Table 5.**

Table 4. Supplementary parameters for boiler operating conditions.

Identifier	Parameter	Unit
1	Furnace temperature	°C
2	Flue gas oxygen concentration	%
3	Primary and secondary air ratio	/
4	Primary air volume	Nm^3/h
5	Secondary air volume	Nm^3/h
6	Coal feeder frequency	Hz
7	Induced draft fan frequency	Hz
8	Forced-draft blower	Hz

Table 5. Training and condition testing dataset.

Dataset	Time Period	Explanation
Train	2023/3/10–2023/4/30	Basic modeling conditions
Test: case #0	2023/5/4–2023/5/11	Scenario involving partial efficiency variations
Test: case #1	2022/2/16–2022/2/25	Scenario with incorrect calorific value unit

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

Test: case #2	2023/3/27–2023/4/1	Scenario where a portion of data is drawn from training data
Test: case #3	2022/2/16–2022/2/25	Scenario with calorific value unit adjusted based on case #1
Test: case #4	2022/4/20–2022/4/29	Scenario with altered operating conditions

To more robustly establish the broad transferability of the component modeling and refreshment framework being advanced, operational measurements from a steam turbine have been enlisted for supplementary proof. A forecasting model for turbine power generation has been formulated by leveraging data on inflow, outflow, and extraction flow rates, together with temperature and pressure signals from the turbine. The computational environment comprises an Apple M1 processor with 8 GB of memory, operating under macOS.

Results and Discussion

Construction of the baseline component prediction model

The baseline component prediction model is a neural network organized into five layers, with neuron complements of 500, 500, 500, 500, and 1. The “relu” activation function is uniformly applied across all layers. Training relies on the “adam” optimizer, with a randomly partitioned 70% share of the full dataset reserved for this purpose. The model’s forecasting performance is visualized in **Figure 4**. On the y-axis of **Figure 4**, the label “Deviation” captures the absolute difference between forecasted and observed quantities, expressed as a percentage of the observed quantity. Across the test subset, the mean squared error (MSE) amounts to 0.14%. The mean absolute percentage error computed over the training subset registers at 0.12%, never exceeding 0.67%. Over the test subset, the corresponding mean absolute percentage error edges upward to 0.18%, yet stays beneath a ceiling of 1.14%. Such outcomes confirm that the baseline model’s predictive precision meets the commonly accepted industrial forecasting tolerance of 5%.

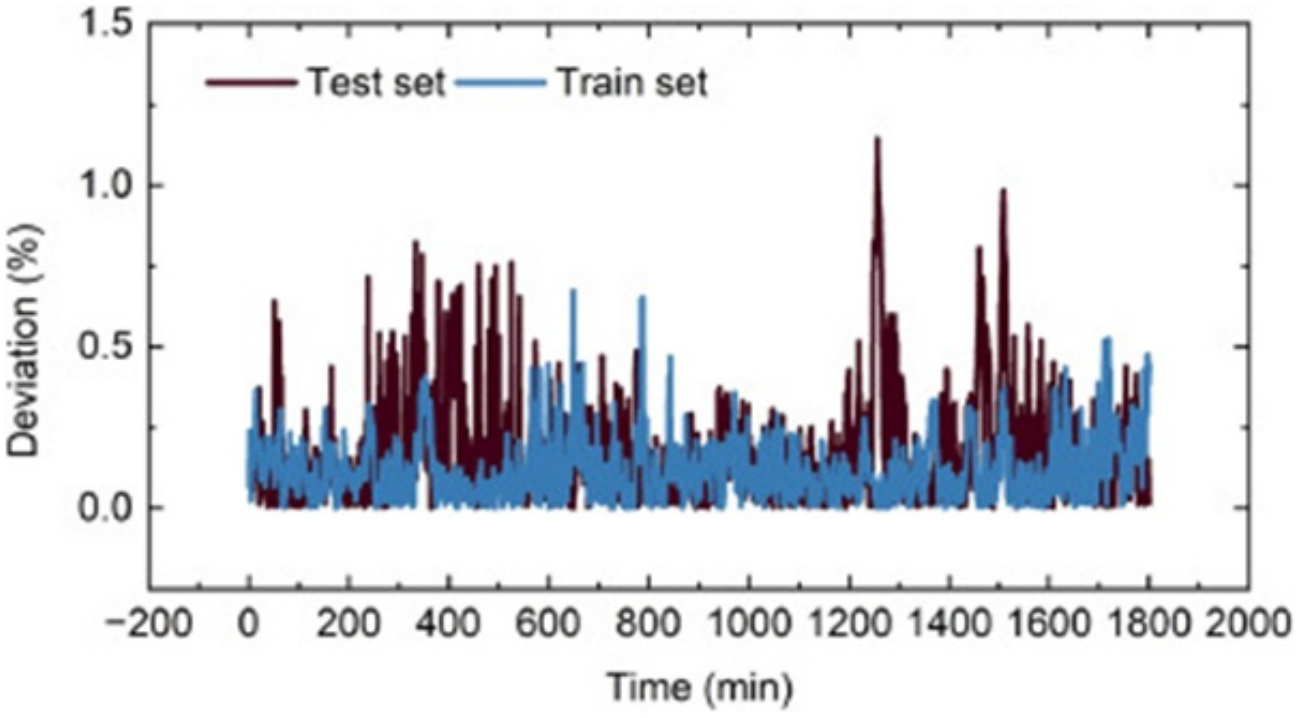


Figure 4. Results of baseline component prediction model.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

Validation of the modeling and updating method

The modeling and refresh scheme laid out in this paper spans three constituent parts: the signal that activates a model refresh, the mechanism by which the refresh is executed, and the protocol for endorsing the refreshed model. To begin with, the soundness of both PCA and the knowledge distillation construct is confirmed. Subsequently, real-world online application and interpretative discussion are undertaken, building upon this verification. As a final step, a separate IES element—namely, the steam turbine—is drawn upon to test the generalizability of the framework.

Effectiveness of operating condition detection

A confidence threshold of 95.00% is selected, which translates to setting α equal to 0.05. The interval employed for training matches that specified, whereas the testing stretch runs from 13:36 on May 4, 2023, to 14:37 on May 11, 2023. After computing cumulative contribution proportions, retaining seven principal components pushes the cumulative contribution to 85.00%, as shown in **Figure 5.**

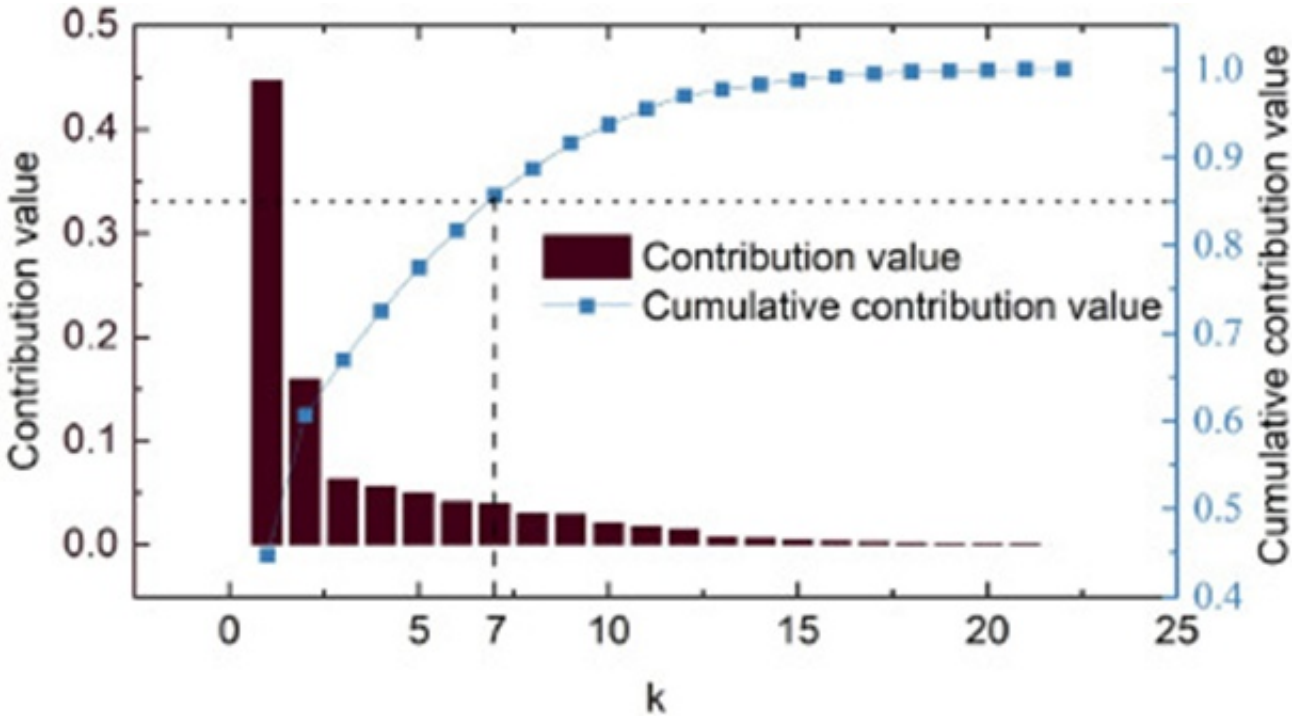


Figure 5. Principal components contribution.

Figures 6 and 7 exhibit the outcomes of operating state surveillance applied to the test data (case #0). The respective control boundaries for T^2 and SPE stand at 14.07 and 8.07. Inspection of **Figures 6 and 7** makes it clear that the PCA-centered detection approach introduced here began sensing a departure in the operating regime of the test sequence from the 3315th observation onward. The operating conditions clustered around the 3315th observation have been isolated and graphed.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

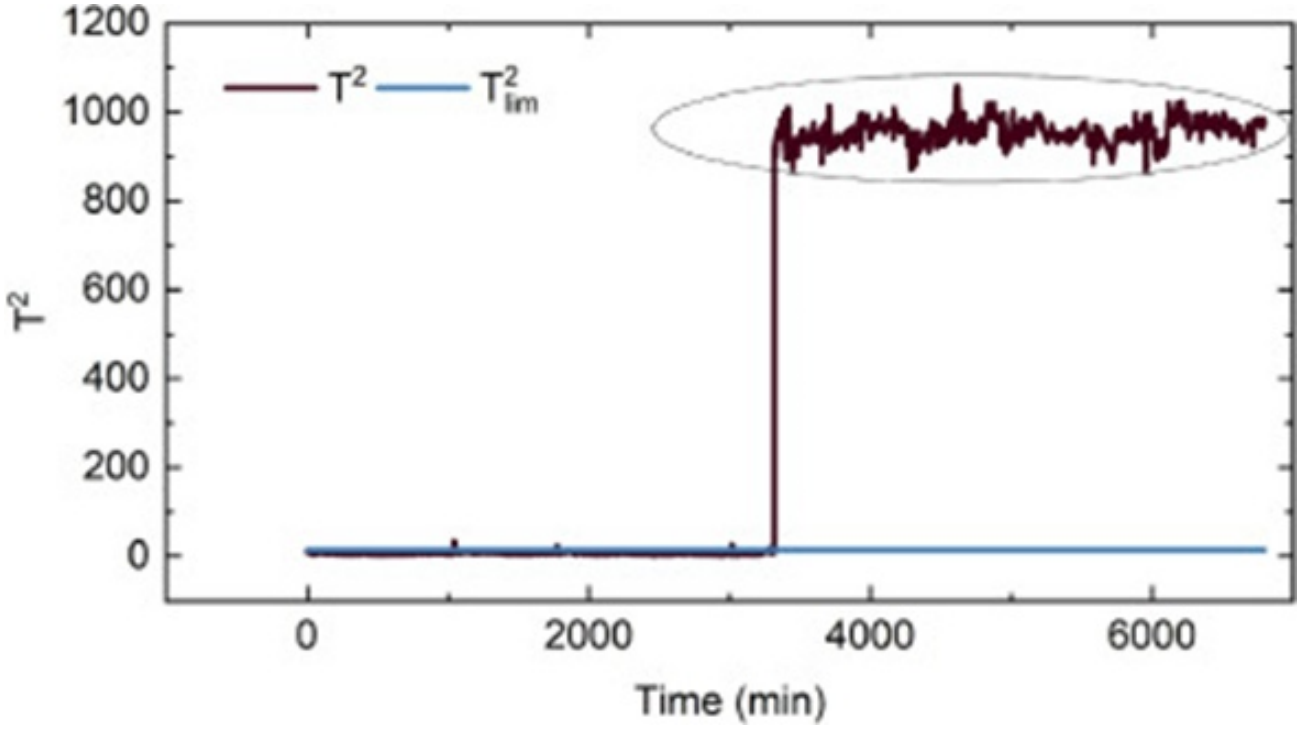


Figure 6. Condition detection results - T² of case #0.

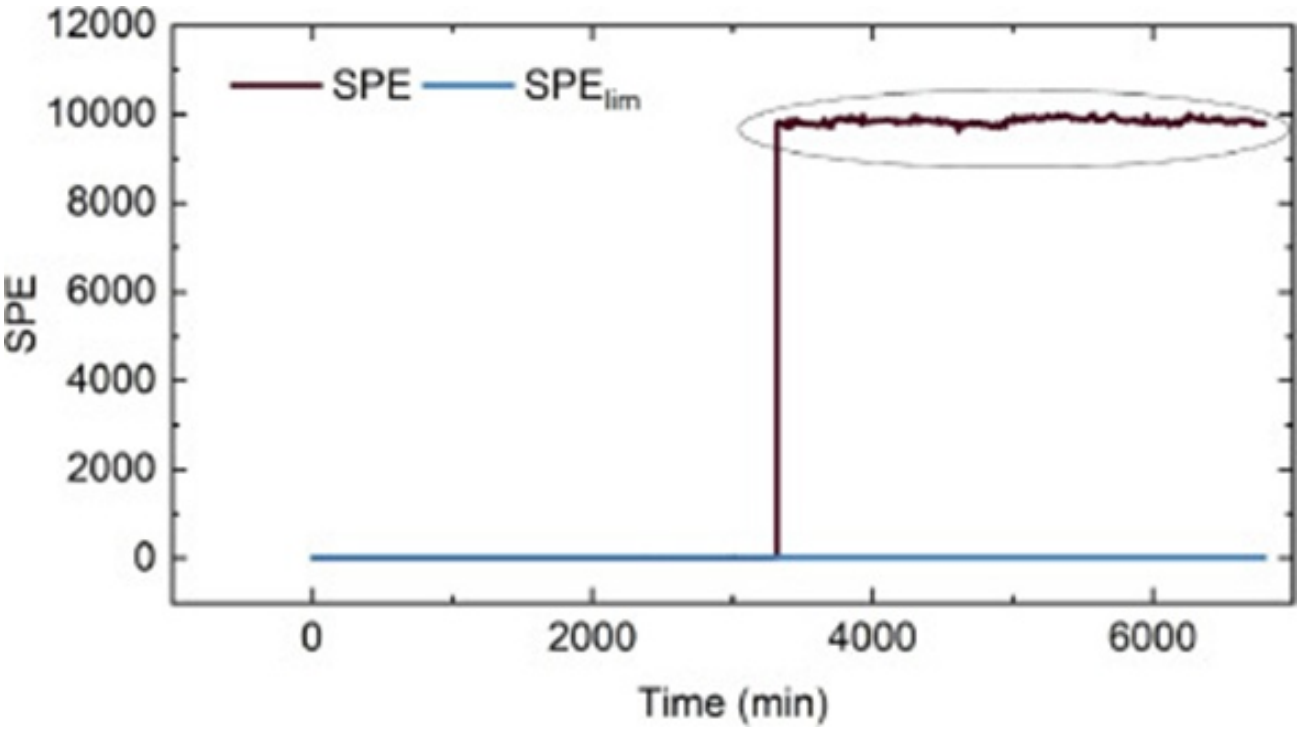
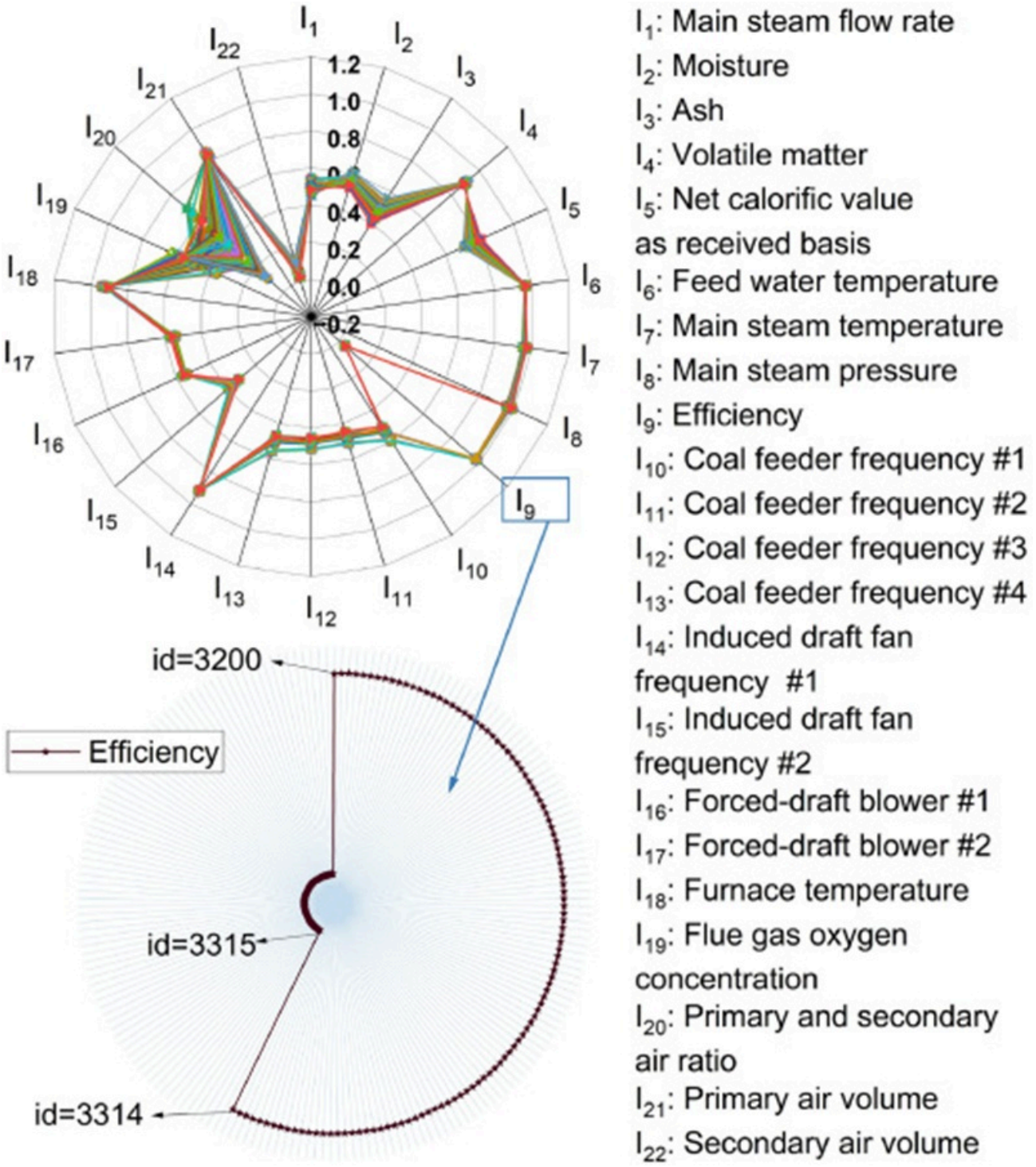


Figure 7. Condition detection results - SPE of case #0.

What **Figure 8** reveals is that, commencing at the 3315th observation, boiler efficiency readings undergo a steep downturn. This downturn subsequently provokes movements in both T² and SPE. The upper portion of **Figure 8** lays out the detailed operating state values at successive time instants, spanning main steam flow rate, moisture, ash, and additional variables. Since each variable has been normalized, the radial span from the origin to the periphery of the plot

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

corresponds to a range stretching from -0.2 to 1.2, with every polygon standing for a single time-sliced operating state. A considerable degree of overlap among the edges of most polygons can be discerned. Among them, the flue gas oxygen concentration and the primary-secondary air ratio produce edge geometries that differ visibly from the rest, signaling that the operating regime has changed. The edge traced by boiler efficiency displays a sharp discontinuity, where a value abruptly shifts to a position near 0. Zooming in further, the boiler efficiency readings spanning observations 3200 through 3398 have been extracted and plotted in the lower portion of **Figure 8**. What becomes evident is that boiler efficiency truly experiences a sudden collapse to near zero across the band from observation 3315 to observation 3398, with the break initiating precisely at the 3315th observation.



© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

Figure 8. Actual distribution of conditions in case #0.

The capacity of the operating state detection technique outlined in this work is further tested on cases #1, #2, and #3. As portrayed in **Figure 9**, within case #1, all observations are marked as representing an operational shift owing to a sizeable deviation in heating value. Turning to case #2, the presented scheme successfully registers select alterations in the operating state. The anomalous patterns flagged in case #3 do, upon examination, indeed stray from the training regime. Collectively, these observations bear out the ability of the introduced method to furnish prompt monitoring of evolving operating states.

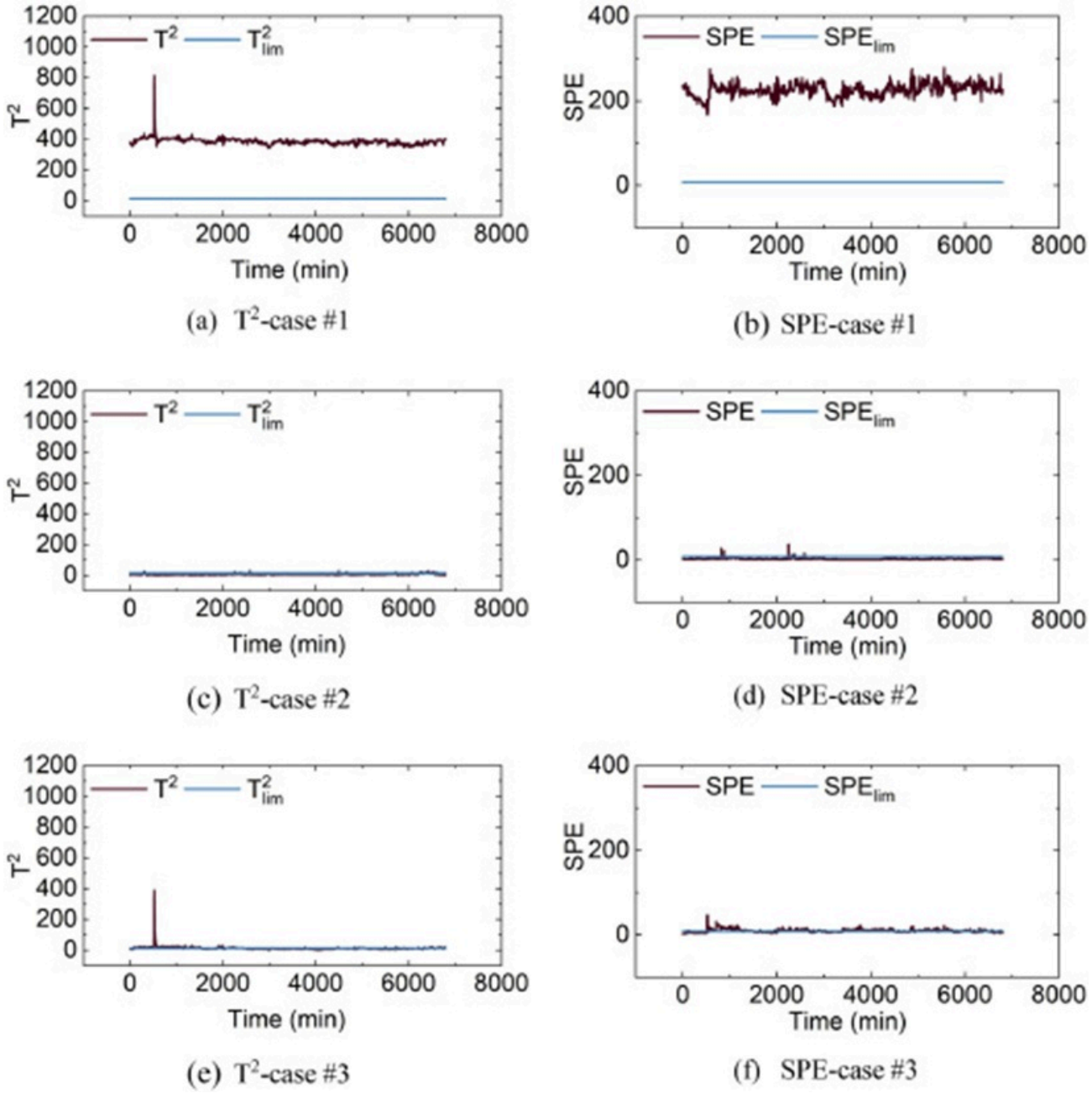


Figure 9. Test results of condition detection for the other three test sets.

Parameter selection and modeling of knowledge distillation

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

The component forecasting model erected is appointed as the teacher model. In parallel, a structurally reduced student model is fabricated. While the student model preserves the five-hidden-layer blueprint of the teacher, the number of processing units within each hidden layer is scaled down to 10, 10, 10, and 1, in that sequence. All other configuration parameters are kept identical. Once trained, the student model records an MSE of 0.30 %, a figure that exceeds the MSE of the baseline model.

Once the student model stands ready, the knowledge distillation operation is performed. As indicated by **Eq. 17,** the effectiveness of the distillation process is modulated by the beta parameter. As illustrated in **Figure 10**, a sweep of experiments is conducted with a range of beta values to gauge their bearing on the distillation outcome. The evidence suggests that, for iteration counts under 100, the model's error trajectory exhibits a downward slope as beta is raised from 0 to 0.50. Once beta surpasses 0.50, however, additional increments in beta leave the MSE largely unaltered. Looking at the iteration dimension, the jump from 1 iteration to 10 iterations delivers a meaningful boost to MSE. When the iteration count lies between 10 and 1000, the MSE curves are tightly grouped, spanning an MSE bracket of 0.28% to 0.30%.

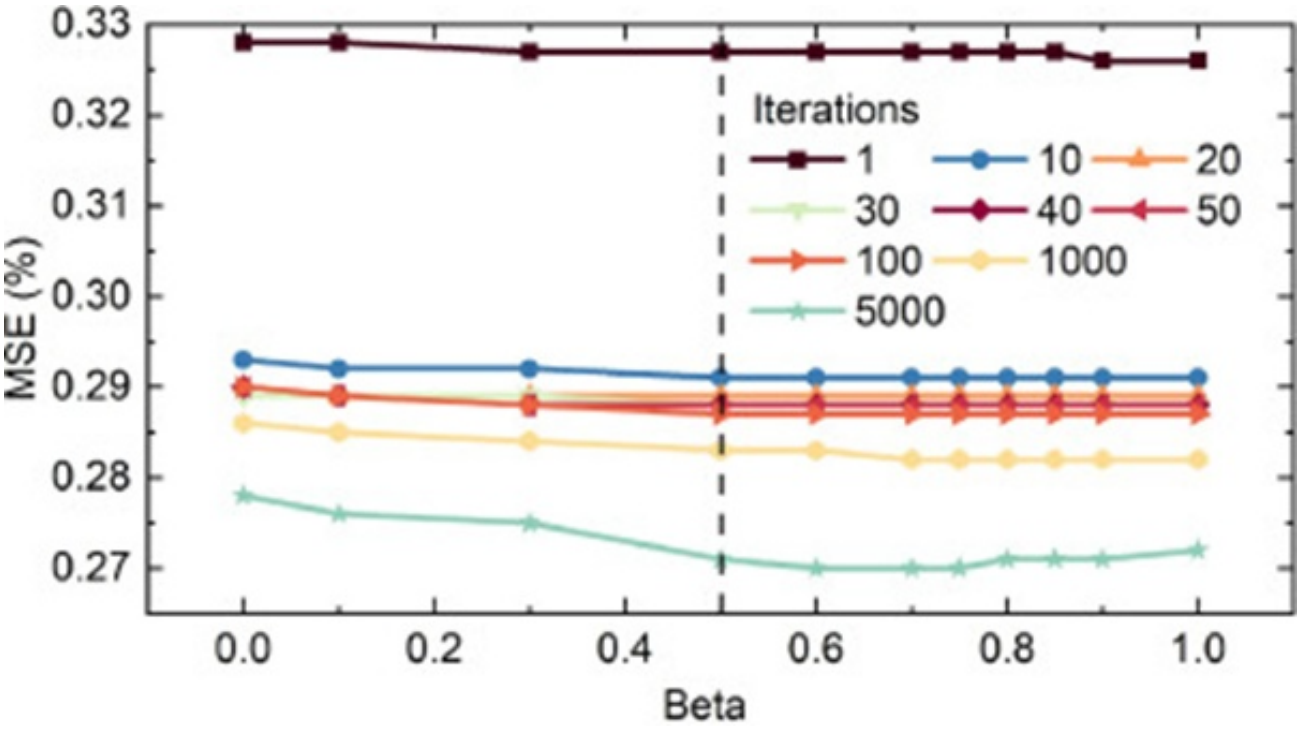


Figure 10. Parameter selection of knowledge distillation model – beta coefficients.

Figure 11 captures the evolution of time expenditure under varying distillation iteration numbers, with beta held steady at 0.50. As the iteration count ascends, so too does the consumed time. For iteration counts ranging from 1 to 100, the time expenditure remains fairly similar, yet a pronounced surge in time consumption manifests once the iteration count hits 1000.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

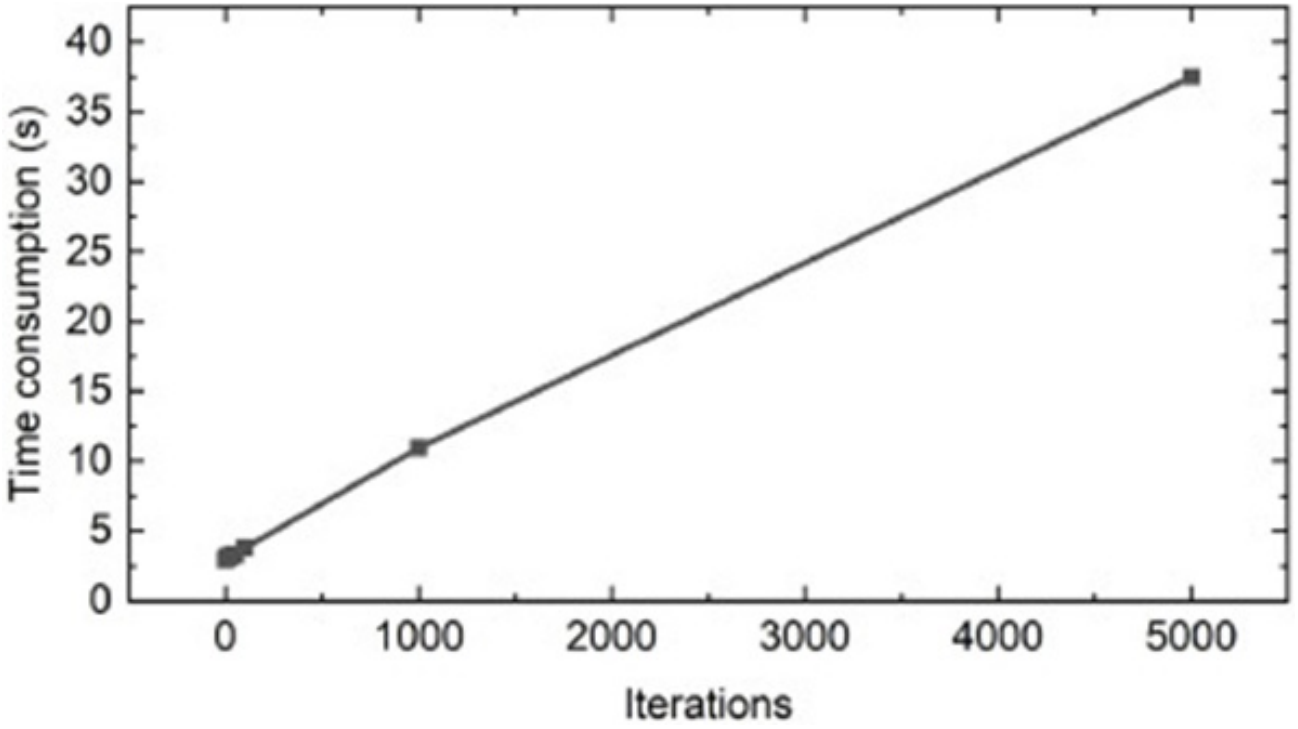
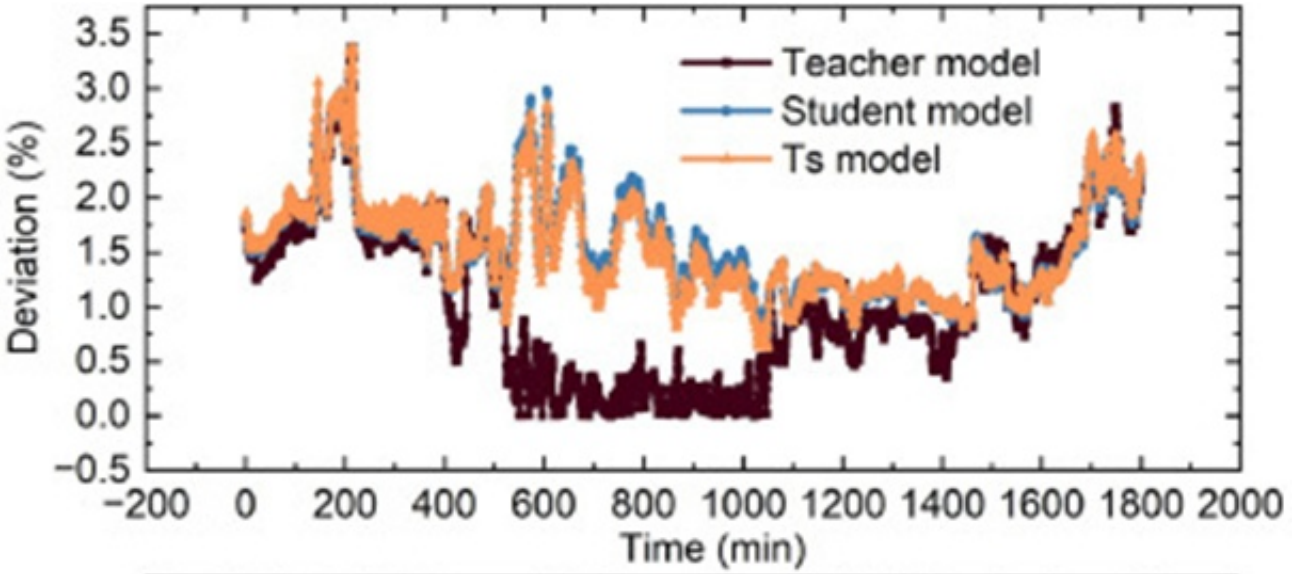


Figure 11. Parameter selection of knowledge distillation model – iterations.

Weighing both the degree of MSE enhancement and the associated time outlay, the present study opts for a beta value of 0.50 paired with 10 iterations when performing knowledge distillation to prime the initial ts model.

Drawing upon case #3, **Figure 12** showcases the predictive discrepancies exhibited by the freshly distilled initial model. Notably, the model emerging from knowledge distillation achieves a discernible gain over the original student model, with the average deviation contracting by 1.39%. This observation suggests that knowledge distillation can absorb the experiential knowledge housed within the teacher even as the model architecture is streamlined, thereby sharpening predictive accuracy.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press



Deviation (%)	Ts model	Teacher model	Student model
Average	1.54	1.05	1.56

Figure 12. Results of constructing the initial model of knowledge distillation.

Online application based on knowledge distillation model

The distilled model yielded before is rolled out under a variety of operating scenarios, where ten rounds of refresh computation are performed, and the resulting metrics are averaged. Case #3 serves as the repository of anomalous operating state data. At moments when a model refresh becomes warranted, **Figure 13** depicts the iterative refreshment behavior produced by three alternative models: the teacher, the student, and the ts model.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

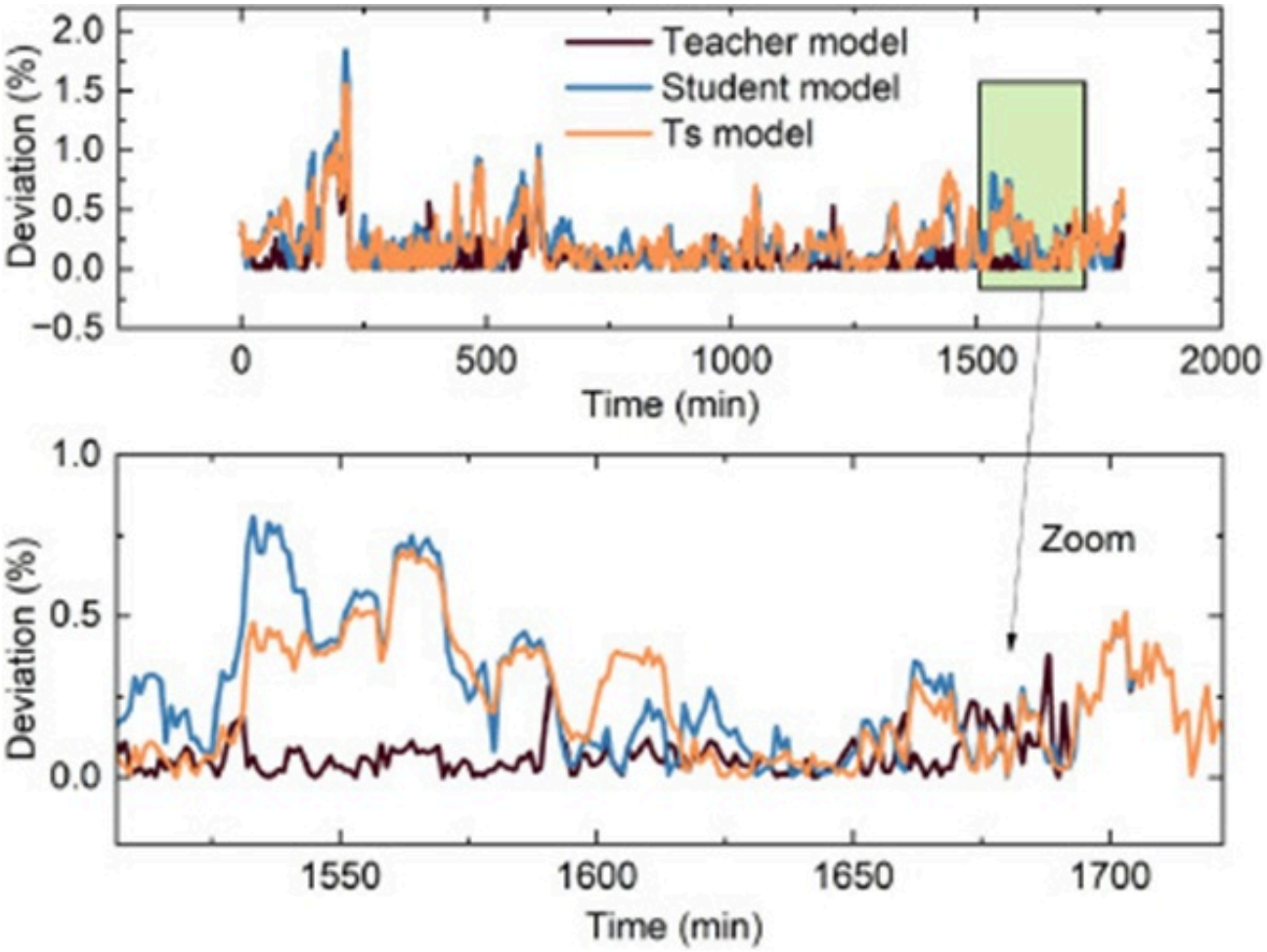
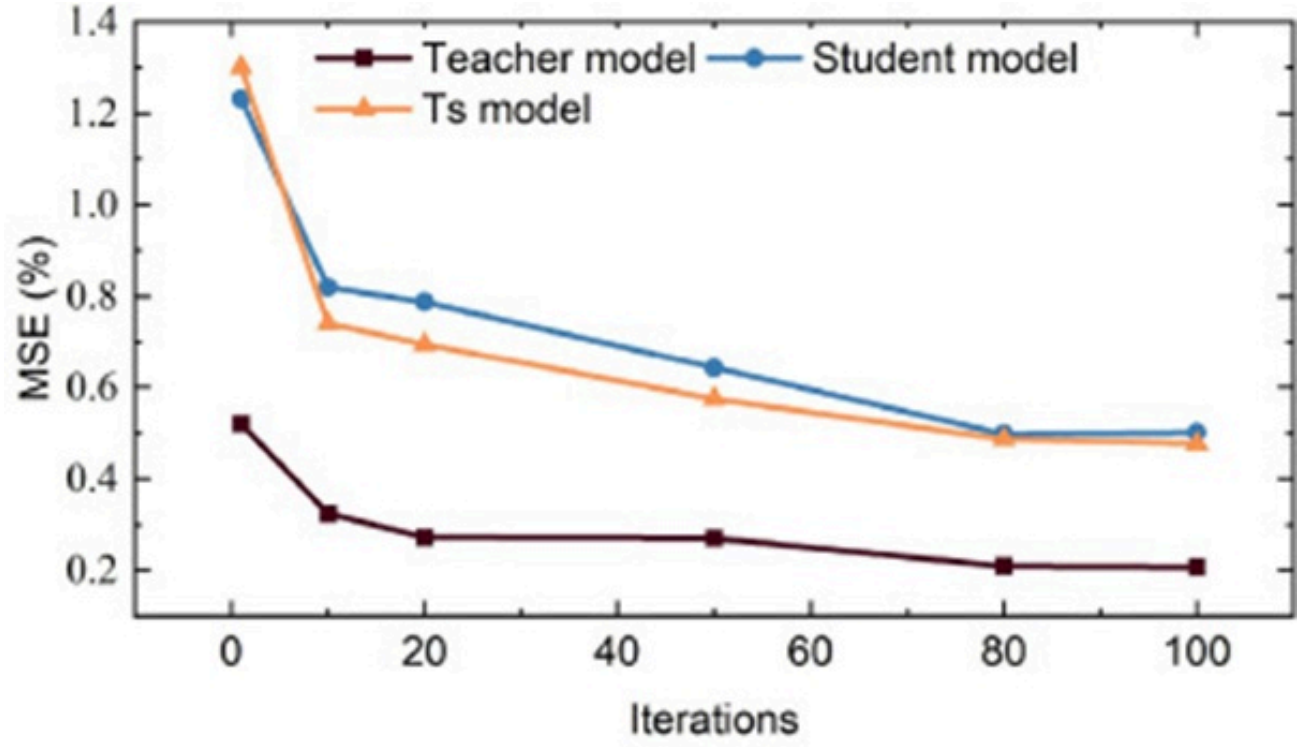


Figure 13. Online model update application effect - based on case #3.

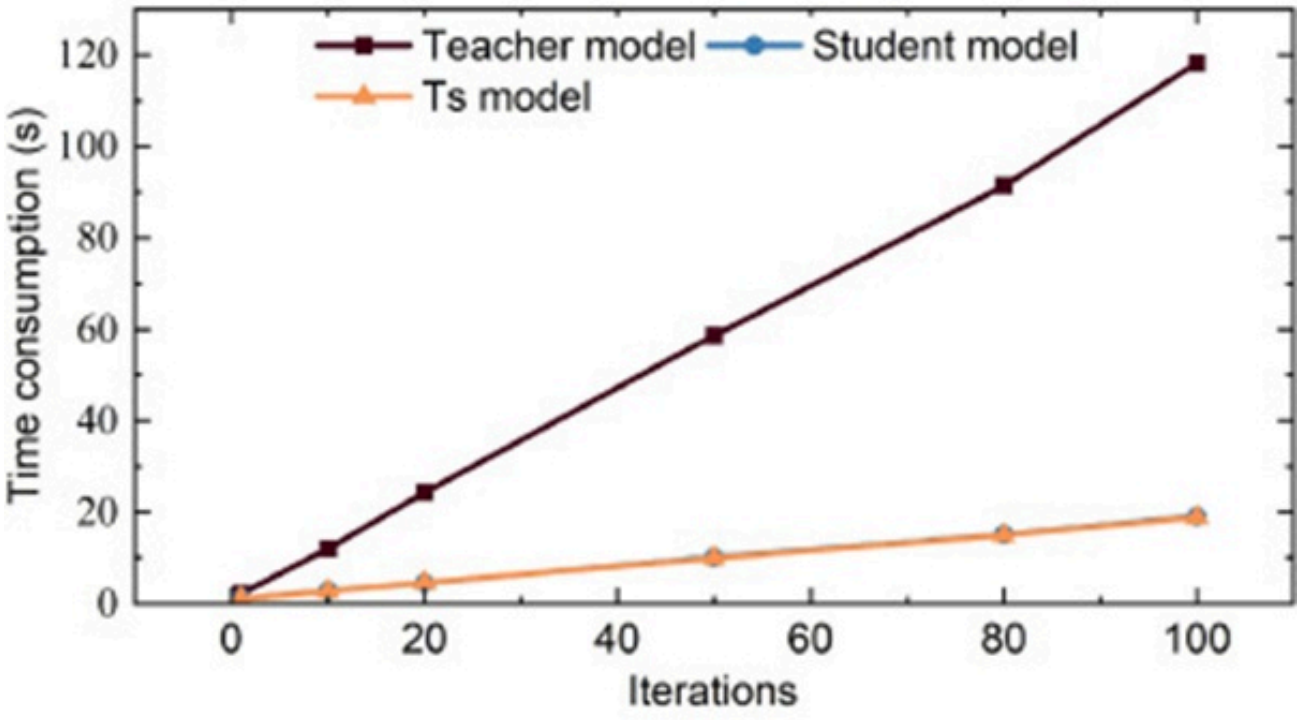
When placed alongside **Figure 12**, the ts model shows an average deviation of 0.23%, marking a drop of 1.31% from the pre-refresh level of 1.54%. In terms of time, executing ten refresh iterations on the ts model takes a scant 2.77 s, whereas the student and teacher models require 2.81 s and 12.03 s, respectively. These results underscore that model refreshment via knowledge distillation can substantially reduce the time required for online model renewal. Compared with the teacher model, the time requirement shrinks by 76.67%. Measured against the student model, the time demand contracts by 1.43%. On top of this, the model discrepancy tied to the knowledge distillation route contracts by 9.66%.

Figure 14 places side by side the outcomes of online refreshments anchored in the various models. The patterns indicate that with rising iteration counts, MSE diminishes while its downward slope flattens. Time expenditure, meanwhile, scales linearly with the iteration tally. In case #3, the MSE associated with the post-distillation model, under matched iteration counts, dips below that of the student model, and the time outlay of the post-distillation model is markedly compressed relative to the teacher model.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press



(a) MSE



(b) Time consumption

Figure 14. Results of online updates based on different models - case #3.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

Figure 15 captures how online model refreshments play out under the evolving operating conditions represented by case #4. The tendencies observed in case #4 echo those already substantiated in case #3. The knowledge-distilled model retains its superiority in both MSE and time consumption dimensions. The MSE that follows online refresh computations rooted in the ts model stands lower than what is obtained by refreshing straight from the student model. From a computational efficiency standpoint, it likewise handily outstrips refreshments grounded in the teacher model.

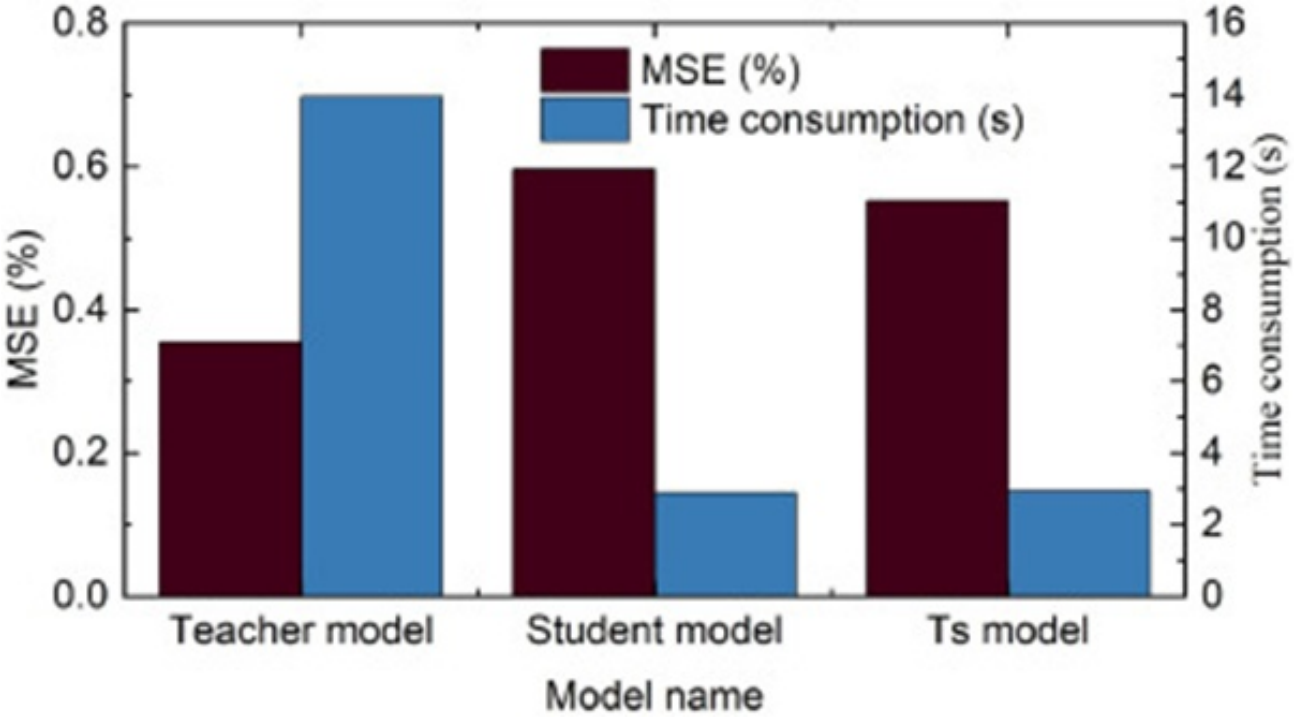


Figure 15. Results of online updates based on different models-case #4.

Universality of the method

As laid out in **Table 6**, the framework advanced in this investigation is equally deployable for refreshing the turbine representation. The refresh tactic built around the proposed methodology consistently delivers a more favorable MSE across diverse iteration budgets when contrasted with direct refreshments from the student model, with the average drop reaching 41.42%. What is more, the MSE-reducing benefit becomes increasingly evident as the iteration budget expands. When set against direct refreshments performed on the teacher model, the refreshment executed through the ts model advocated in this work yields a dramatic shortening of computation time. At an iteration budget of 10, the turbine model refreshment under this study cuts 27.35 s compared to direct refreshments on the teacher model. Pushed to an iteration budget of 2000, the savings in computation time can swell to 4722.42 s. The inference is clear: the knowledge-distillation-centered model refreshment technique proposed in this investigation carries demonstrable strengths in computational throughput.

Table 6. Updating results of the turbine model using the proposed method.

Method	Iterations							
	2000		1000		100		10	
	Time (s)	MSE (%)	Time (s)	MSE (%)	Time (s)	MSE (%)	Time (s)	MSE (%)
Ts model	14.10	0.0054	9.26	0.0086	4.85	0.029	4.61	0.053

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

Student model	46.43	0.019	19.03	0.020	5.76	0.043	5.13	0.056
Teacher model	4736.52	0.0035	13,541.89	0.0023	253.51	0.0074	31.96	2.37

Impact of model structure on prediction effect

To gain a more granular understanding of how distinct architectural decisions influence the distillation outcome, the effect of varying the number of neural network layers is probed. With the teacher model held unchanged, as rendered in **Figure 16**, a pattern emerges wherein a gradual decline in model MSE accompanies an increase in the depth of the student model. Offset against this, time consumption charts a gently rising course. A point deserving attention is that, as layer depth grows, its incremental influence on model MSE tapers toward a plateau. Hence, once dimensions such as model MSE and time demand are jointly factored in, one may deduce the existence of a favorable layer count—one that upholds model capability while accommodating the constraints of computational pace.

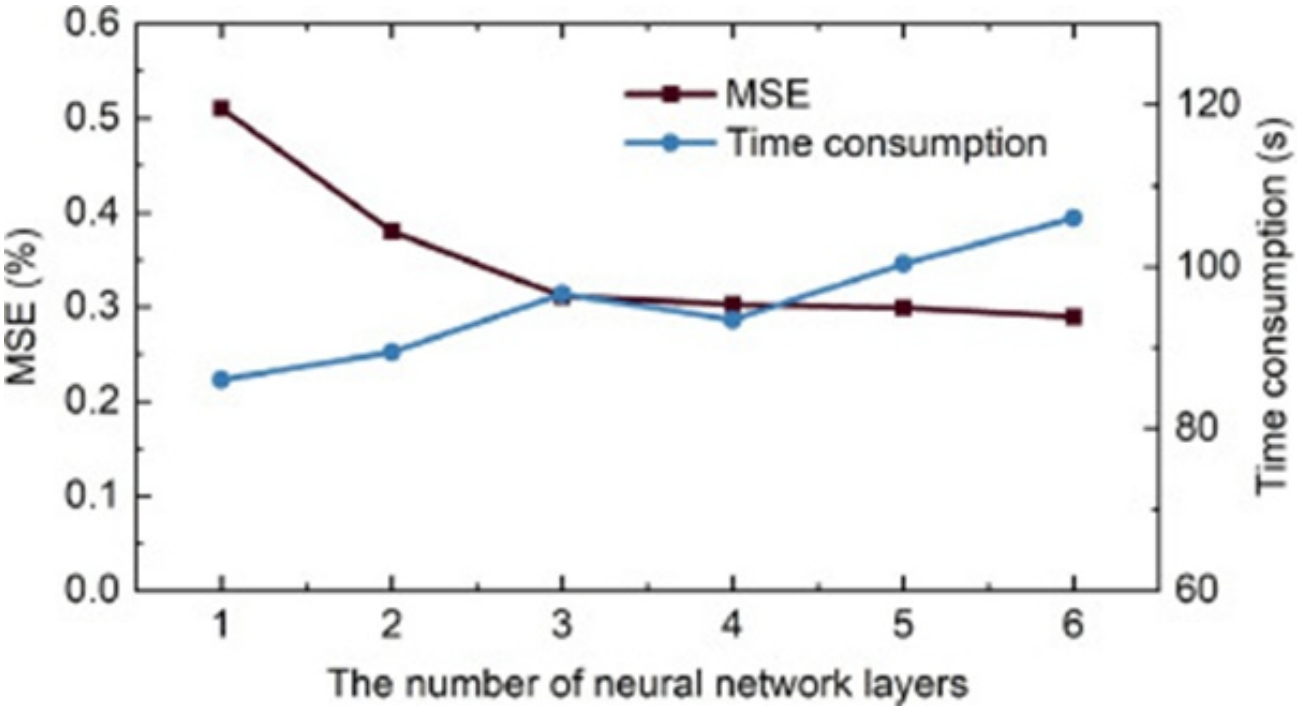


Figure 16. The influence of neural network layers on prediction performance.

Figure 17 portrays the manner in which the width of each layer, expressed as neuron count, conditions the potency of knowledge distillation. The evidence points toward a scenario where, under a fixed layer depth, augmenting the neuron count produces mild fluctuations in time consumption without engendering a substantive uptick. The increase in time consumption is smaller than the shift caused by changes in layer depth. Furthermore, for a constant layer architecture, the MSE traces a descending trajectory that progressively stabilizes as the neuron count climbs.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

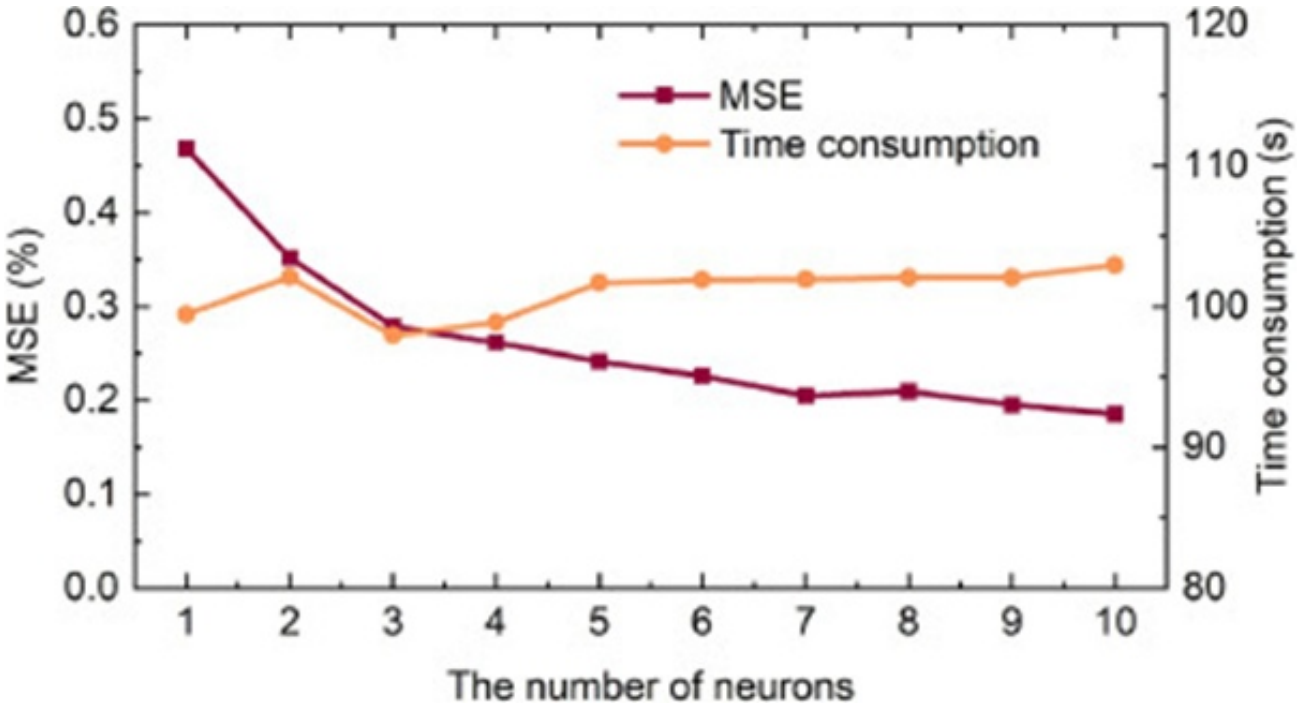


Figure 17. The influence of the number of neurons on prediction performance.

Adaptability analysis: comparison with traditional methods

To shed additional light on the adaptive capacity of the technique advanced in this paper, a side-by-side evaluation has been performed against two widely adopted variable-condition modeling paradigms, specifically the multivariate linear model and the piecewise linear model. Case #3 has been designated as the testing dataset.

A multivariate linear formulation is first assembled. The computation takes 1.29 s, and the model attains a score of 0.76. Next, a five-segment linear formulation is constructed, using main steam flow as the segmentation criterion. The forecasting behavior of the three approaches on the training partition is displayed in **Figure 18**. “Yreal” denotes the actual measured boiler efficiency. What can be observed is that, on the training partition, the forecasts produced by the proposed technique sit nearest to the actual readings, with the piecewise linearization approach coming next. In contrast, the multivariate linear forecasts exhibit the largest error spread.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

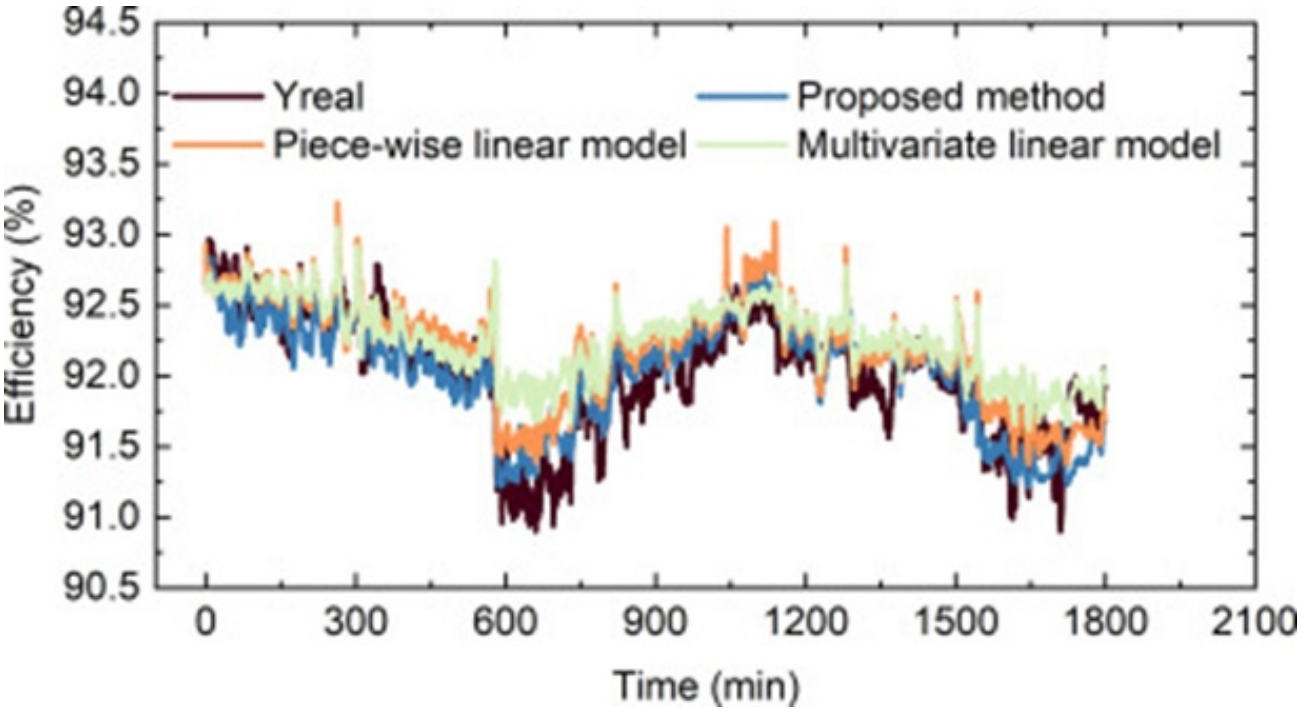


Figure 18. Prediction results compared to traditional methods - on training set.

Moving beyond this, a test of prediction under shifting conditions is executed across the three methodologies, as summarized in **Table 7**. Within case #3, the divergence yielded by piecewise linearization turns out to be more pronounced than that produced by straightforward linearization. The neural-network-based method yields an average deviation that is 12.61% smaller than the piecewise linear method and 3.49% smaller than the multivariate linear method. The outcomes on the pair of training sets bear a resemblance to one another; however, once projected onto changing operating regimes, the ability of piecewise linearization to generalize proves inferior. The principal cause lies in the difficulty piecewise linearization faces in accommodating multivariable interdependencies. The methodology outlined in this investigation delivers superior generalization relative to the conventional multivariate linear and piecewise linear approaches, all while offering forecasts that approximate those of the baseline model, yet with a leaner architecture and swifter refresh cadence.

Table 7. Comparison of condition adaptability with traditional methods.

Method	**Train**			**Case #4**			**Case #3**		
	MSE (%)	Max deviation (%)	Mean deviation (%)	MSE (%)	Max deviation (%)	Mean deviation (%)	MSE (%)	Max deviation (%)	Mean deviation (%)
Multivariate linear methods	0.43	1.25	0.27	7.88	3.42	1.46	8.80	3.37	1.60
Piecewise linear methods	0.27	0.96	0.23	7.97	3.53	1.47	11.21	3.67	1.76
The proposed method	0.17	0.98	0.18	4.62	3.26	1.04	8.35	3.37	1.54

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

Necessity analysis: impact of component model updates on IES scheduling

The shifting share of renewable energy influences the loading of IES components. Such load variability, in turn, places greater demands on the fidelity of component models. To probe how model refreshment affects overall dispatch outcomes under deep renewable penetration, this study first conducts a sensitivity analysis of renewable energy's effect on component loading. Subsequently, it examines the impact of component model inaccuracies on system-wide dispatching.

As portrayed in **Figure 19**, when the buffering effect of energy storage is left out of consideration, each 1.00 MW uptick in renewable power output leads to a marginally differing drop in coal burn across various main steam flow rates, the mean reduction amounting to 1.64 t/h. **Figure 20** reveals that, for a fixed main steam flow rate, the relationship linking the rise in renewable generation to the associated decline in coal consumption is strictly linear.

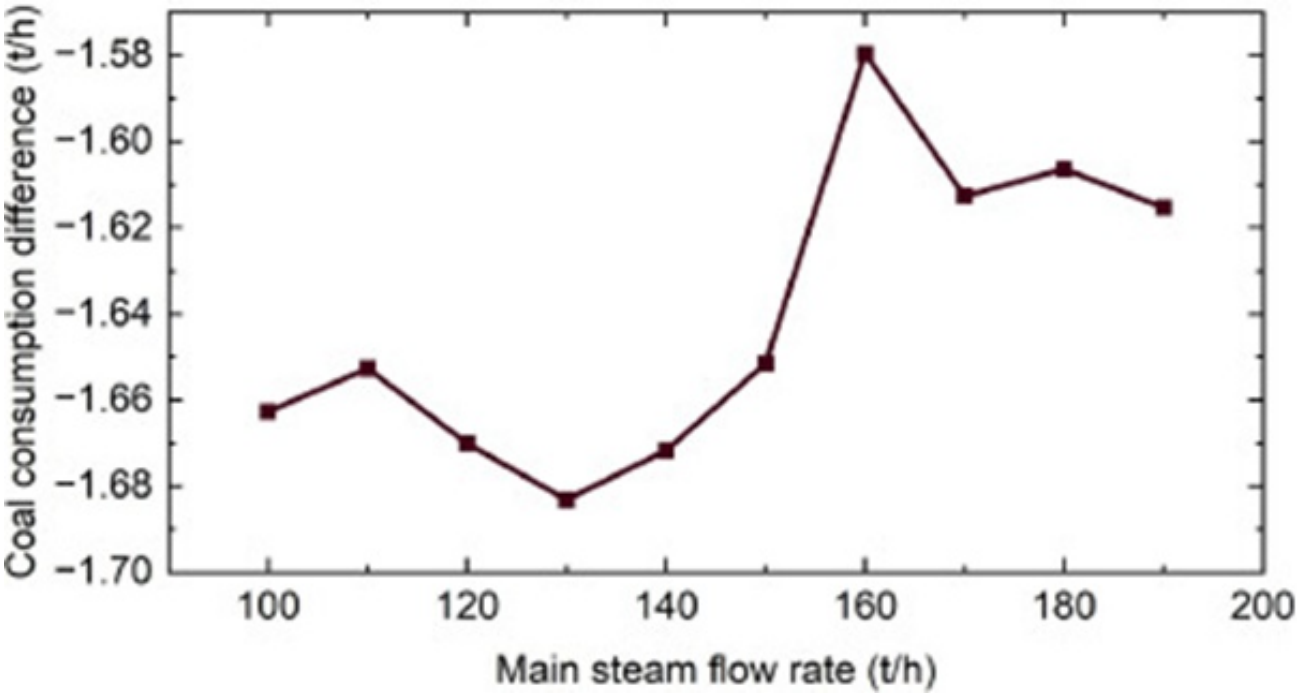


Figure 19. Impact of each MW increase in renewable energy on boiler coal consumption.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

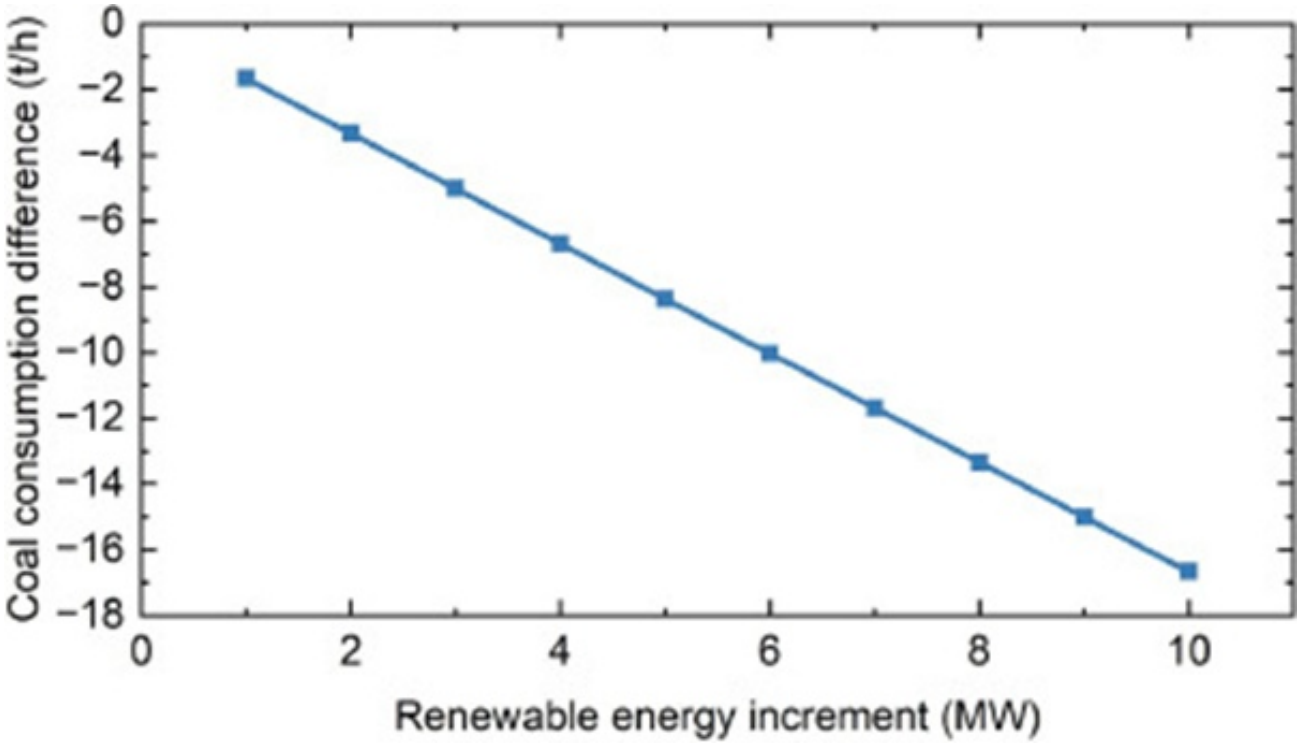


Figure 20. Impact of renewable energy increment on coal consumption (Main steam flow rate=100.00 t/h).

Delving deeper, **Figure 21** demonstrates that the coal-saving effect triggered by a unitary increase in renewable output differs depending on the prevailing main steam flow rate. Holding the renewable increment constant, a higher main steam flow rate correlates with a smaller coal consumption drop, implying higher boiler efficiency under elevated load conditions. In concrete terms, a unitary boost in renewable generation cuts coal usage by 1.62 t/h when the main steam flow rate stands at 190.00 t/h, and by 1.65 t/h when the rate is 100.00 t/h. It can also be discerned that a reduced main steam flow rate does not invariably translate into greater coal burn for an identical uptick in renewable output. Collectively, these observations indicate that assimilating differing fractions of renewable energy can have varied repercussions on boiler loading.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

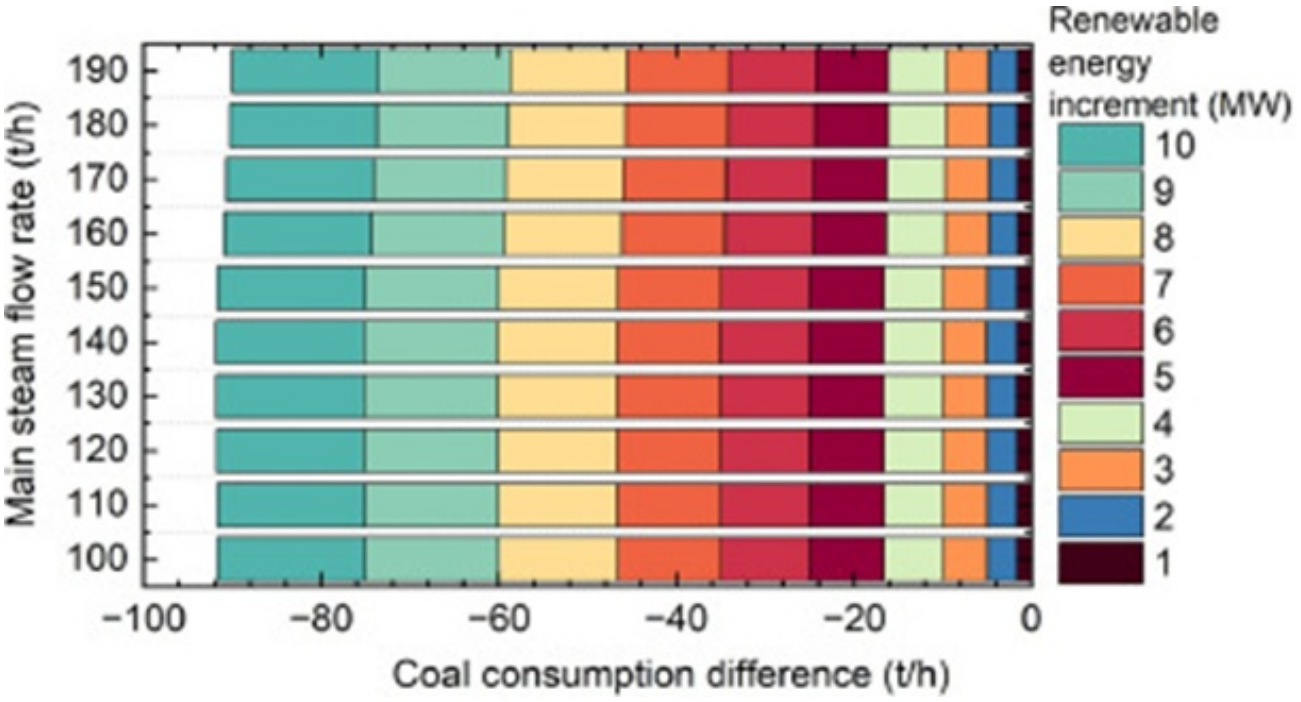


Figure 21. Impact of renewable energy increment on coal consumption (under different main steam flow rates).

To extend the investigation into how model inaccuracies ripple through IES dispatch choices and to underscore the criticality of adaptive model refreshment, a series of tests under differing MSE levels is conducted.

As laid out in **Table 8**, the figures reflect the MSE expressed in relation to boiler efficiency. An MSE of 1.00%, upon undergoing reverse denormalization computation, translates to a 0.50% shift in boiler efficiency.

Table 8. The reflection of MSE on boiler efficiency.

Id	Efficiency (%) (Reverse normalization)	RMSE	MSE (%)
1	0.049	0.01	0.01
2	0.16	0.032	0.10
3	0.50	0.10	1.00
4	1.11	0.22	5.00
5	1.57	0.32	10.00

As rendered in **Figure 22**, when the MSE takes on values of 0.01%, 0.10%, 1.00%, 5.00%, and 10.00%, the corresponding average deviations in coal consumption work out to 0.0048 t/h, 0.015 t/h, 0.048 t/h, 0.11 t/h, and 0.16 t/h, respectively. Priced against thermal coal at 950 ¥/t, an MSE of 1.00 % would translate into a daily profit discrepancy of ¥ 1094.40.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

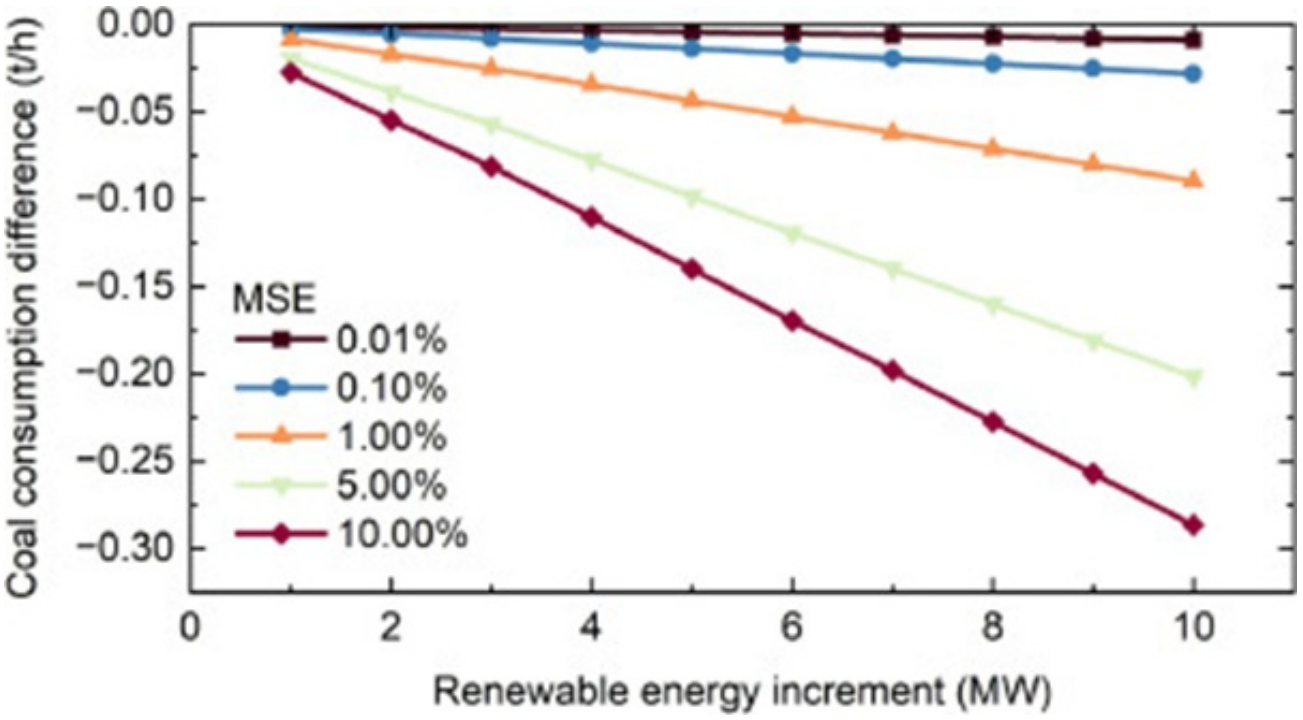


Figure 22. The impact of model MSE on coal consumption.

Extending this line of inquiry, the consequence of model deviation on coal usage is computed under the assumption of fixed power output (i.e., a constant main steam flow rate). Taking the main steam flow to be 190.00 t/h and adopting 90.00 % as the baseline boiler efficiency, **Table 9** enumerates the effect that variations in boiler efficiency within a ±4.00 % band exert on both coal consumption and financial return.

Table 9. Boiler efficiency deviation on coal consumption and operational profits.

Item	Boiler efficiency (%)				
	86.00	87.00	88.00	89.00	89.50 (MSE=1.00%)
Coal consumption (t/h)	24.39	24.11	23.84	23.57	23.44
Coal consumption difference (t/h)	1.08	0.80	0.53	0.26	0.13
Profit difference (¥/day)	24,717.47	18,325.02	12,077.86	5971.08	2950.95

Within certain markets, explicit penalty frameworks have been instituted for entities participating in ancillary service markets. To illustrate, the Zhejiang market levies a doubled assessment charge on awarded entities that fail to deliver contracted ancillary services. As a further analytical step, an evaluation anchored to fixed coal consumption is appended, computing the downstream effect of model deviation on power generation. Presuming a boiler coal burn rate of 23.00 t/h, again with 90.00% as the reference efficiency, **Table 10** details the impact of boiler efficiency swings within a 4.00% bracket on both power output and associated costs. The upshot is that a 1.00% MSE, traceable to a failure to refresh the model in a timely fashion, may give rise to a power deviation of 0.075 MW. At an electricity tariff of 300 ¥/MWh, this could generate a daily penalty expense of ¥ 1084.64.

Table 10. Boiler efficiency deviation on power and operational profits.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

Item	Boiler efficiency (%)				
Empty Cell	**86.00**	**87.00**	**88.00**	**89.00**	**89.50 (MSE=1.00 %)**
Main steam flow rate (t/h)	179.16	181.24	183.32	185.40	186.45
Power difference (MW)	0.61	0.46	0.30	0.15	0.075
Profit difference (¥/day)	8729.51	6547.14	4364.76	2182.38	1084.64

Inputs: Dataset I and corresponding output dataset O, N_T hidden layers, layer neuron settings N_i, learning rate η, momentum factor μ, activation function settings.
Outputs: Trained neural network model parameters (weights and thresholds of each neural network layer), training error.
(1) Network initialization: randomly generate weights and thresholds;
(2) Input training dataset I, O;
(3) Based on input I, calculate the output of each hidden layer neuron and the output of the output layer O′ according to Eq. (13);
(4) Calculate the error between the hidden layer and the output layer;
(5) Update connection weights and thresholds;
(6) Calculate the error function to determine whether it is less than the given learning threshold. If satisfied, the learning ends and jumps to step 7; otherwise, jumps to step 2.
(7) Retain weights and thresholds, output training error.

Algorithm 1. Construction of teacher model.

One may therefore conclude that, for IES, the adaptive refreshment of models is a matter of the highest consequence.

To encapsulate, the knowledge-distillation-anchored component modeling and refreshment framework put forward in this study carries substantial advantages over conventional approaches in both computational throughput and resilience to shifting operating conditions, and proves capable of registering alterations in the operating state of IES.

Conclusion

At present, model-driven decision-making constitutes the bedrock upon which IES conducts energy management and regulation. This investigation offers a data-driven construction and refreshment methodology to counter the erosion of component model adaptability within IES when confronted with evolving operating circumstances. Taking boiler efficiency as the illustrative case, the proposed framework can surveil shifts in operating state and execute model refreshes without undue delay. In terms of real-time responsiveness, the knowledge distillation model reduces time expenditure by 76.67% relative to direct refreshes applied to elaborate models. With respect to operating condition adaptability, when placed beside conventional multivariate linear and piecewise linear formulations, the proposed framework exhibits stronger generalization capabilities on the test partition, posting an average deviation that trails piecewise linear by 12.61% and multivariate linear by 3.49%. Viewed through the lens of refresh necessity, a 1.00 MW augmentation in renewable generation can trim coal consumption by 1.64 t/h, and a 1.00% MSE arising from delayed model refreshes can induce a power discrepancy of 0.075 MW. This insight further reinforces the significance of timely model renewal. Additionally, supplementary analysis reveals that expanding the number of neural network layers exerts a more appreciable influence on post-distillation errors.

The contributions and pivotal findings of this work can be enumerated as follows:

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

(1) A framework for component modeling and refreshment within IES is advanced, weaving together the assessment of refresh criteria, instantaneous model renewal, and model selection. The evidence demonstrates that this framework is timely, flexible, and essential.

(2) A knowledge-distillation-grounded technique is harnessed, guiding pre-existing complex component models toward the construction of a leaner surrogate, thereby supplying a new tactic for real-time model refreshment. In parallel, a PCA-rooted method for detecting component operating states is brought to bear, offering a fresh answer to the challenge of pinpointing the right moment for data model renewal.

(3) An inquiry is mounted that takes the expansion of renewable energy as its point of departure, derives the resultant shifts in boiler coal consumption, and scrutinizes whether refreshing the boiler model shapes the aggregate dispatch benefit. This furnishes a novel vantage point on the repercussions of renewable growth for boiler operation and on the role model refreshment plays in dispatch effectiveness.

Future investigations will concentrate on reducing reliance on data quality and volume, fusing mechanistic representations with data-driven techniques to pursue research at the intersection of physical principles and data, and probing modeling strategies suited to data-scarce settings. Beyond this, enhancements to the PCA-based condition detection procedure will be sought, alongside the exploration of tactics aimed at boosting surveillance efficiency across diverse operating regimes.

Nomenclature

Acronyms

MPC = Model predictive control

PCA = Principal component analysis

DRL = Deep reinforcement learning

MSE = Mean squared error

IES = Integrated energy systems

Symbols

X_0 The original data set with n observations and m variables

X_M The mean of the X_0

X_V The variance of X_0

X The normalized input data

Z The covariance matrix

λ The eigenvalues

The eigenvectors

k Features

d Accumulated contribution limit

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

P	The feature matrix
T	The score matrix
E	The residual matrix
T^2	The statistics
SPE	The statistics
X_n	The new data set
1 - α	The confidence level
F	The F distribution
T^2_{lim}	The control limit of T^2
SPE_{lim}	The control limit of SPE
c_α	The confidence value of the standard normal distribution
N_T	The number of hidden layers of the teacher model
N_i	The number of neurons of layer i
I	The input dataset
O	The output dataset
L_t	The loss value of the teacher model
c_i	The model coefficient, i equals one to four
F_1	Extraction steam flow rate [t/h]
P_{re}	The output power of renewable energy [MW]
C	The coefficient
h_s	The enthalpy value of the main steam [KJ/kg]
Q	The calorific value of the coal [KJ/kg]
W	The coefficient term
O_{pre_t}	The predicted value predicted using the teacher's model
O_r	The real value
y_1, y_2	The two variable arguments of the function MSE

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

$O_{m,j}$ The output of neuron j of the hidden layer m

σ_m The activation function of the hidden layer m

In The number of features entered for the hidden layer

$w_{i,j}$ The weight coefficient of input x_i for neuron j of hidden layer m

x_i The input i of neuron j of the hidden layer m

b_j The threshold of neuron j of hidden layer m

η The learning rate

μ The momentum factor

N_S The number of hidden layers of the student model

N_j The number of neurons of hidden layer j

L The total loss value

O_{pre_s} The predicted value predicted using the student's model

f_1 The complex structure prediction model

f_2 The simplified structure prediction model

ts model The model obtained based on knowledge distillation

β The adjustable parameter of the total loss function

f_{new} The updated model for online application

ε The acceptable boundary value of the cumulative error range

d_{t1},d_{t2} The boundary values of the cumulative duration of model updates

I_{new} The new input data

O_{new} The new output data

$P_{turbine}$ Power generation of the turbine [MW]

D The main steam flow rate [t/h]

F_2 Exhaust steam flow rate [t/h]

P_{adjust} The adjustable power within the IES [MW]

B The coal consumption of the boiler [t/h]

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

h_w The enthalpy value of the feedwater [KJ/kg]

μ The efficiency of the boiler [%]

B′,D′ B, D values after the change in renewable energy outputs[t/h]

Acknowledgements

None

Conflict of interest

None

Financial support

This work is supported by the National Key R&D Program of China (Grant No.2023YFE0108600), the National Natural Science Foundation of China (Grant No.51806190), the National Key R&D Program of China (Grant No.2022YFB3304502), and a self-directed project, State Key Laboratory of Clean Energy Utilization.

Ethics statement

None

Received: 25 Apr 2022 Revised: 30 Jun 2022 Accepted: 15 Jul 2022
Published online: 15 June 2023

Rights and permissions

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at https://creativecommons.org/licenses/by-nc-sa/4.0/.

References

1. Lu G, Wang Z, Bhatti U, Aslani A, Kumar R, Zahedi R, et al. Recent progress in carbon dioxide capture technologies: a review. Clean Energy Sci Technol. 2023;1(1):32.

2. Mishra S, Pradhan P, Jena S. Performance and combustion study of a low heat rejection engine running with biogas–diethyl ether–diesel. Waste Dispos Sustain Energy. 2023;1-12.

3. Yang W, Zhang E, Zhao J, Meng X, Liu Y, Wang H, et al. Dawn of clean energy: Enhanced heat transfer, radiative cooling, and firecracker-style controlled nuclear fusion power generation system. Clean Energy Sci Technol. 2023;1(1):61.

4. Parimi K, Kaur B, Lankapalli S, Rao V, Sharma A, Reddy P, et al. Performance, combustion, and emission characteristics of on a diesel engine fuelled with hydrogen compressed natural gas and Kusum seed biodiesel. Waste Dispos Sustain Energy. 2023;1-13.

5. National Energy Administration. China's renewable energy installations surpass coal [EB/OL]. National Energy Administration; 2023 Aug 4 [cited 2023 Aug 21]. Available from: http://www.nea.gov.cn/2023-08/04/c_1310735564.htm

6. Yan R, Wang J, Huo S, Li M, Zhang Y, Chen X, et al. Flexibility improvement and stochastic multi-scenario hybrid optimization for an integrated energy system with high-proportion renewable energy. Energy. 2023;263:125779.

7. Mancarella P. MES (multi-energy systems): an overview of concepts and evaluation models. Energy. 2014;65:1-17.

8. Wu J, Yan J, Jia H, Hatziargyriou N, Djilali N, Sun H. Integrated energy systems. Appl Energy. 2016;167:155-7.

9. Ramachandra T. RIEP: regional integrated energy plan. Renew Sustain Energy Rev. 2009;13(2):285-317.

10. Saboori H, Mohammadi M, Taghe R. Virtual power plant (VPP), definition, concept, components and types. In: Proceedings of the 2011 Asia-Pacific Power and Energy Engineering Conference; 2011; IEEE. p. 1-4.

11. Aghdam F, Mudiyanselage M, Mohammadi-Ivatloo B, Marzband M, Asadi S, Anvari-Moghaddam A, et al. Optimal scheduling of multi-energy type virtual energy storage system in reconfigurable distribution networks for congestion management. Appl Energy. 2023;333:120569.

12. Oskouei M, Mirzaei M, Mohammadi-Ivatloo B, Shafie-khah M, Catalão JPS, Zare K. A hybrid robust-stochastic approach to evaluate the profit of a multi-energy retailer in tri-layer energy markets. Energy. 2021;214:118948.

13. Mirzaei MA, Zare K, Mohammadi-Ivatloo B, Marzband M, Anvari-Moghaddam A, Asadi S. Robust network-constrained energy management of a multiple energy distribution company in the presence of multi-energy conversion and storage technologies. Sustain Cities Soc. 2021;74:103147.

14. Mobasseri A, Tostado-Véliz M, Ghadimi A, Miveh MR, Jurado F, Catalão JPS. Multi-energy microgrid optimal operation with integrated power to gas technology considering uncertainties. J Clean Prod. 2022;333:130174.

15. Li Y, Zhang H, Liang X, Huang B, Liu Y, Li X. Event-triggered-based distributed cooperative energy management for multienergy systems. IEEE Trans Industr Inform. 2018;15(4):2008-22.

16. Zheng J, Xiao W, Wu C, Li Y, Zhang H, Wang J. A gradient descent direction based-cumulants method for probabilistic energy flow analysis of individual-based integrated energy systems. Energy. 2023;265:126290.

17. Ceseña E, Mancarella P. Energy systems integration in smart districts: Robust optimisation of multi-energy flows in integrated electricity, heat and gas networks. IEEE Trans Smart Grid. 2018;10(1):1122-31.

18. Li Z, Xu Y. Temporally-coordinated optimal operation of a multi-energy microgrid under diverse uncertainties. Appl Energy. 2019;240:719-29.

19. Zhang N, Sun Q, Yang L, Li Y, Zhang H, Wang J, et al. Event-triggered distributed hybrid control scheme for the integrated energy system. IEEE Trans Industr Inform. 2021;18(2):835-46.

20. Nittaya T, Douglas P, Croiset E, Ricardez-Sandoval LA, Elkamel A. Dynamic modelling and control of MEA absorption processes for CO2 capture from power plants. Fuel. 2014;116:672-91.

21. Sani AG, Najafi H, Azimi S. Dynamic thermal modeling of the refrigerated liquified CO2 tanker in carbon capture, utilization, and storage chain: a truck transport case study. Appl Energy. 2022;326:119990.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

22. Gabrielli P, Gazzani M, Mazzotti M. Electrochemical conversion technologies for optimal design of decentralized multi-energy systems: Modeling framework and technology assessment. Appl Energy. 2018;221:557-75.

23. Ma J, Gu Y, Shen Y, Wang J, Ye L, Gao R, et al. Stability analysis of power grid connected with direct-drive wind farm containing virtual inertia based on integrated dissipation energy model. IEEE Trans Sustain Energy. 2021;12(4):2378-92.

24. Abbasimehr H, Paki R, Bahrini A. A novel XGBoost-based featurization approach to forecast renewable energy consumption with deep learning models. Sustain Comput Inform Syst. 2023;38:100863.

25. Ajbar W, Parrales A, Cruz-Jacobo U, Hernández-Pérez JA, García-Pérez J, Pérez-García J, et al. The multivariable inverse artificial neural network combined with GA and PSO to improve the performance of solar parabolic trough collector. Appl Therm Eng. 2021;189:116651.

26. Wang Z, Jing W, Ling Z, Zhang H, Liu Y, Chen X, et al. A thermal energy usage prediction method for electric thermal storage heaters based on deep learning. In: Proceedings of the 2019 IEEE 4th International Conference on Cloud Computing and Big Data Analysis (ICCCBDA); 2019; IEEE. p. 149-54.

27. Kämper A, Leenders L, Bahl B, von der Aßen N, Lampe M, Bardow A, et al. AutoMoG: automated data-driven model generation of multi-energy systems using piecewise-linear regression. Comput Chem Eng. 2021;145:107162.

28. Mansoor M, Stadler M, Zellinger M, Kupzog F, Hauer T, Kozek M. Optimal planning of thermal energy systems in a microgrid with seasonal storage and piecewise affine cost functions. Energy. 2021;215:119095.

29. Benzaama M, Rajaoarisoa L, Ajib B, Abbes M, Belarbi R, Leklou A. A data-driven methodology to predict thermal behavior of residential buildings using piecewise linear models. J Build Eng. 2020;32:101523.

30. Qin C, Zhao J, Chen L, Li Y, Zhang H, Wang J, et al. An adaptive piecewise linearized weighted directed graph for the modeling and operational optimization of integrated energy systems. Energy. 2022;244:122616.

31. Rahbar K, Xu J, Zhang R. Real-time energy storage management for renewable integration in microgrid: An off-line optimization approach. IEEE Trans Smart Grid. 2014;6(1):124-34.

32. Labidi M, Eynard J, Faugeroux O, Grieu S, Ben Youssef Y, Bournot P, et al. A new strategy based on power demand forecasting to the management of multi-energy district boilers equipped with hot water tanks. Appl Therm Eng. 2017;113:1366-80.

33. Vasilj J, Jakus D, Sarajcev P. Robust nonlinear economic MPC based management of a multi energy microgrid. IEEE Trans Energy Convers. 2020;36(2):1528-36.

34. Jiménez D, Angulo A, Street A, Muñoz-Organero M, Moreno R, Conejo AJ. A closed-loop data-driven optimization framework for the unit commitment problem: A Q-learning approach under real-time operation. Appl Energy. 2023;330:120348.

35. Ye Y, Qiu D, Wu X, Sun H, Xia S, Li Z. Model-free real-time autonomous control for a residential multi-energy system using deep reinforcement learning. IEEE Trans Smart Grid. 2020;11(4):3068-82.

36. Yang L, Li X, Sun M, Zhang Y, Wang J, Chen H, et al. Hybrid policy-based reinforcement learning of adaptive energy management for the Energy transmission-constrained island group. IEEE Trans Industr Inform. 2023; [Epub ahead of print].

37. Wen D, Aziz M. Data-driven energy management system for flexible operation of hydrogen/ammonia-based energy hub: a deep reinforcement learning approach. Energy Convers Manag. 2023;291:117323.

38. Brauns K, Scholz C, Schultz A, Seckmeyer G, Schröder A, Mette J, et al. Vertical power flow forecast with LSTMs using regular training update strategies. Energy AI. 2022;8:100143.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

39. Wang J, Ye L, Gao R, Li Y, Zhang H, Liu Y. Digital Twin for rotating machinery fault diagnosis in smart manufacturing. Int J Prod Res. 2019;57(12):3920-34.

40. Zhang H, Qi Q, Ji W, Wang J, Liu Y, Li X, et al. An update method for digital twin multi-dimension models. Robot Comput Integr Manuf. 2023;80:102481.

41. Schreiber J, Sick B. Model selection, adaptation, and combination for transfer learning in wind and photovoltaic power forecasts. Energy AI. 2023;14:100249.

42. Huang H, Li J, Zhao Q, Zhang Y, Wang H, Liu S, et al. Adaptive update method of digital twin model for feed system based on hybrid drive. Comput Integr Manuf Syst. 2023;29(6):1840.

43. Yu J, Song Y, Tang D, Dai J, Fang J, Wang S, et al. A digital twin approach based on nonparametric Bayesian network for complex system health monitoring. J Manuf Syst. 2021;58:293-304.

44. Hutter F, Kotthoff L, Vanschoren J. Automated machine learning: methods, systems, challenges. Cham: Springer Nature; 2019.

45. Masana M, Liu X, Twardowski B, Menta M, Bagdanov AD, van de Weijer J. Class-incremental learning: survey and performance evaluation on image classification. IEEE Trans Pattern Anal Mach Intell. 2022;45(5):5513-33.

46. Sarmas E, Strompolas S, Marinakis V, Doukas H, Psarras J, Kaldellis J. An incremental learning framework for photovoltaic production and load forecasting in energy microgrids. Electronics (Basel). 2022;11(23):3962.

47. Wu H, Elizaveta D, Zhadan A, Liu Y, Zhang H, Wang J, et al. Forecasting online adaptation methods for energy domain. Eng Appl Artif Intell. 2023;123:106499.

48. Hinton G, Vinyals O, Dean J. Distilling the knowledge in a neural network. arXiv:1503.02531. 2015.

49. Li Y, Hu F, Liu Y, Zhang H, Wang J, Chen X, et al. A hybrid model compression approach via knowledge distillation for predicting energy consumption in additive manufacturing. Int J Prod Res. 2023;1-23.

50. Ai M, Xie Y, Ding SX, Zhang H, Wang J, Li Y, et al. Domain knowledge distillation and supervised contrastive learning for industrial process monitoring. IEEE Trans Ind Electron. 2022;70(9):9452-62.

51. Elsamanty M, Ibrahim A, Salman W. Principal component analysis approach for detecting faults in rotary machines based on vibrational and electrical fused data. Mech Syst Signal Process. 2023;200:110559.

52. Swiercz M, Mroczkowska H. Multiway PCA for early leak detection in a pipeline system of a steam boiler-selected case studies. Sensors. 2020;20(6):1561.

53. Sakiewicz P, Piotrowski K, Ober J, Karwot J, Wierzbicki J, Rychlik M, et al. Innovative artificial neural network approach for integrated biogas–wastewater treatment system modelling: effect of plant operating parameters on process intensification. Renew Sustain Energy Rev. 2020;124:109784.

54. Zhang G, Huang X. Applications of neural network in complex system modeling. Sci Technol Rev. 2018;36(12):66-70.

55. Garcia-Alvarez D, Bregon A, Pulido B, Alonso-González CJ, Díaz I, González J. Integrating PCA and structural model decomposition to improve fault monitoring and diagnosis with varying operation points. Eng Appl Artif Intell. 2023;122:106145.

56. Gou J, Yu B, Maybank S, Chen J, Li Y, Zhang H. Knowledge distillation: A survey. Int J Comput Vis. 2021;129:1789-819.

57. Shao R, Liu Y, Zhang W, Wang J, Li Y, Chen H. A survey of knowledge distillation in deep learning. Chin J Comput. 2022;45(8):1638-73.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press

58. Kukreja H, Bharath N, Siddesh C. An introduction to artificial neural network. Int J Adv Res Innov Ideas Educ. 2016;1:27-30.

59. Wei C. Historical trend and drivers of China's CO2 emissions from 2000 to 2020. Environ Dev Sustain. 2022;1-20.

60. Lin X, Lin X, Zhong W. Predictive operation optimization of multi-energy virtual power plant considering behavior uncertainty of diverse stakeholders. Energy. 2023;128130.

© 2023 The Author(s). | Published by Institute of Smart Energy Systems Press