ECoalVis: Visual Analysis of Control Strategies in Coal-fired Power Plants

Shuhan Liu, Di Weng, Yuan Tian, Zikun Deng, Haoran Xu, Xiangyu Zhu, Honglei Yin, Xianyuan Zhan, Yingcai Wu

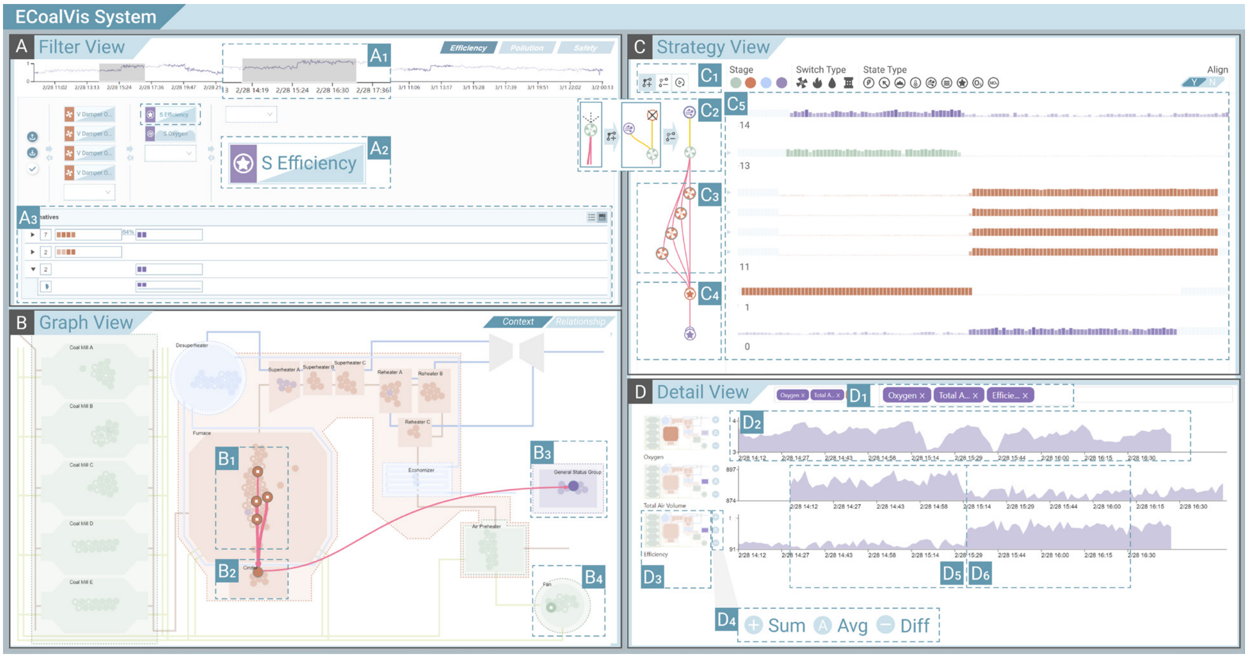

Fig. 1. The user interface of ECoalVis. A) The filter view reveals the time series of key sensors and allows users to fuzzy query control strategies. B) The graph view reveals the spatial propagation of control strategy impact across components, units and sensors. C) The strategy view depicts the temporal cascading of control strategy impact, visualizing the topology of the strategy and the time-lag-aligned time series. D) The detail view allows users to search for sensors and perform time series operations to find insights from the raw data.

Abstract— Improving the efficiency of coal-fired power plants has numerous benefits. The control strategy is one of the major factors affecting such efficiency. However, due to the complex and dynamic environment inside the power plants, it is hard to extract and evaluate control strategies and their cascading impact across massive sensors. Existing manual and data-driven approaches cannot well support the analysis of control strategies because these approaches are time-consuming and do not scale with the complexity of the power plant systems. Three challenges were identified: a) interactive extraction of control strategies from large-scale dynamic sensor data, b) intuitive visual representation of cascading impact among the sensors in a complex power plant system, and c) time-lag-aware analysis of the impact of control strategies on electricity generation efficiency. By collaborating with energy domain experts, we addressed these challenges with ECoalVis, a novel interactive system for experts to visually analyze the control strategies of coal-fired power plants extracted from historical sensor data. The effectiveness of the proposed system is evaluated with two usage scenarios on a real-world historical dataset and received positive feedback from experts.

Index Terms—Power plant visual analytics, energy data visualization, spatiotemporal visualization, smart factory

1 Introduction

Coal is one of the world's largest energy sources, contributing over one-third of the electricity generated in 2019 [45]. However, generating electricity from fossil fuels, such as coal, has serious implications for the environment [41] and human health [12]. To alleviate such implications before clean energy is deployed, the efficiency of the existing coal-fired power plants needs to be improved, such that less coal is consumed and pollutant emissions are reduced [64].

The control strategy is one of the major factors affecting the efficiency of coal-fired power plants [22, 67]. To regulate a coal-fired power plant, experts largely follow certain control strategies to achieve desired goals, like increasing burning rates [25]. A control strategy comprises a sequence of control valve adjustments and state changes, which are monitored by thousands of sensors distributed across a power plant. For example, a control strategy can be formulated as follows: turn up the burner, and then increase the opening of the water valves after temperature rise is observed. However, due to the complex and

- *S. Liu, Y. Tian, Z. Deng, and Y. Wu are with the State Key Lab of CAD&CG, Zhejiang University, Hangzhou, China. E-mail: {shliu, yuantian, zikun_rain, ycwu}@zju.edu.cn.*
- *D. Weng is with Microsoft Research Asia, Beijing, China. E-mail: diweng@microsoft.com. D. Weng is the corresponding author.*
- *H. Xu and H. Yin are with JD iCity, JD Technology, Beijing, China and JD Intelligent Cities Research, Beijing, China. E-mail: {ryanxhr, yinhonglei93}@gmail.com.*
- *X. Zhu and X. Zhan are with Institute for AI Industry Research (AIR), Tsinghua University, Beijing, China. E-mail: {zackxiangyu, zhanxianyuan}@gmail.com.*

Manuscript received 31 March 2022; revised 1 July 2022; accepted 8 August 2022. Date of publication 03 October 2022; date of current version 2 December 2022. This article has supplementary downloadable material available at https://doi.org/10.1109/TVCG.2022.3209430, provided by the authors. Digital Object Identifier no. 10.1109/TVCG.2022.3209430

dynamic environment in the power plants, it is hard to determine the effect such a control strategy has in such an environment.

Traditionally, control strategies are evaluated manually by experts with specialized monitoring and tuning software [11, 47], where the collected sensor data is organized into several dashboards to visualize the real-time states of the power plant. However, not only the trial-and-error approach is too risky for operating power plants on the fly, but also the latent propagation of the impact of a control strategy is hard to capture manually due to the scale of the system. With the development of simulation techniques, many data-driven approaches [17,48,54] were proposed to predict the outcome of control strategies by modeling and simulating a power plant. However, most of these approaches are based on limited datasets and strong assumptions. For example, the power plant system should be in a steady state. Moreover, these approaches mainly focus on physical simulation rather than interactive reasoning. For example, they cannot be used to backtrack efficiency fluctuations and identify the responsible control strategies.

The lack of a comprehensive approach for evaluating the control strategies of coal-fired power plants motivates us to explore the potential solutions driven by visual analytics. Based on massive historical sensor data, a visual analytics solution can assist experts in both forward (determine the impact of control strategies) and backward (identify the control strategies responsible for abnormal states) analyses, facilitating the understanding of the complex interactions between different control strategies and the power plant systems. However, developing such a solution poses the following three challenges:

Interactive extraction of control strategies from large-scale dynamic sensor data. The first step towards the effective analysis of control strategies is to identify and extract these strategies from the historical data. However, such a task can be challenging since a coal-fired power plant comprises numerous units and thousands of sensors, generating large-scale time series data constantly. An efficient search of control strategies and their impact should be implemented to accelerate the analysis procedure. Moreover, the diversity and dynamics in control strategies require the interactive integration of domain knowledge, such that experts can quickly find the strategies of their interest.

Intuitive visual representation of the cascading impact among the sensors in a complex power plant system. The impact of control strategies must be intuitively visualized to assist experts in analyzing how such impact propagates across the power plant system and eventually affects the power generation efficiency. However, a power plant system has numerous components and multiple levels of hierarchies. Due to the complexity and scale of the system structure, it is difficult to reveal the cascading impact of control strategies among components, units, and sensors in a systematic and comprehensible way. Specifically, producing a hierarchical and semantic graph layout with low visual clutter to facilitate understanding is both important and challenging.

Time-lag-aware analysis of the impact of control strategies on electricity generation efficiency. Time lags can be observed when the impact of control strategies propagates from one sensor to another. Such time lags are crucial to the analysis of control strategies because the order and timing of control valve adjustments heavily depend on these time lags. However, capturing and analyzing these time lags is challenging since the environment inside a power plant is highly dynamic. Moreover, the temporal analysis of these time lags should be tightly integrated with the spatial analysis of cascading impact, revealing how the manipulations of control valves are coordinated together to eventually affect the power generation efficiency.

To address these challenges, we propose ECoalVis, a novel interactive system for experts to visually analyze the control strategies of coal-fired power plants. For the first challenge, we designed an interactive query interface that allows users to describe the relationships among multiple time series and proposed a new method that fuzzy searches and extracts the desired control strategies based on the user input. For the second challenge, we employed a dual-mode hierarchical graph visualization, allowing users to switch between context- and relationship-oriented layouts and inspect the cascading impact of control strategies at different levels of hierarchy. For the third challenge, we proposed a spatiotemporal control strategy view that integrates the propagation of the impact among sensors with time-lag-aligned time-series visualization, and an iterative approach was adopted to empower users to verify and correct the extracted strategies based on the time lags. The contributions of this study are summarized as follows:

- We formulated a framework of two types of analysis and characterized the user requirements for the comprehensive analysis of the control strategies of coal-fired power plants.
- We designed an interactive approach that assists users in specifying time series relationships and retrieving the desired control strategies with a fuzzy-matching model.
- We developed ECoalVis, a novel visual analytics system that integrates a set of tailored visualizations to effectively help users query, analyze and evaluate control strategies.
- We evaluated ECoalVis with two usage scenarios on historical data and received positive feedback from domain experts.

2 Related Work

This section presents relevant studies on analyzing control strategies of coal-fired power plants, visual analytics for industries, and event pattern recognition techniques.

2.1 Control Strategies Analysis of Coal-fired Power Plants

There have been many relevant methods for analyzing, evaluating, and optimizing control strategies of coal-fired power plants. These methods can be categorized into two types: *experience-driven* and *data-driven*.

Many *experience-driven* approaches rely on specialized software to continuously collect data from coal-fired power plant sensors and display the values in simple dashboard panels (e.g., Fei et al. [20]). When experts analyze control strategies, they will first check them through some important states, such as Li et al. [35] check energy efficiency and Zheng et al. [69] check gas emission. They search for these state values from multiple sensor panels and then evaluate them based on their experience (e.g., Fan et al. [19]).

With the development of data analysis techniques, many *data-driven* approaches have been applied to the control strategy analysis and evaluation of coal-fired power plants. One type of data-driven technique is modeling coal-fired power plants using expert software (e.g., Li et al. [36]). Some of the modeling studies concentrate on evaluating the subunits of coal-fired power plants. For example, Yin et al. [65] model the coal mill and Chandrasekhara et al. [6] model the boiler. The other focus on the whole power plant and evaluate it globally. For example, Njoku et al. [40] adopt the multi-criteria evaluation. Besides, many machine learning algorithms, such as ANN-GA [50], Bayesian network [68], Reinforcement Learning [22], and Multiple Linear Regression [48], are also applied to model coal-fired power plants.

However, these methods do not combine human intelligence well with machine computing power. Experience-driven methods are poorly scalable and time-consuming. Data-driven methods use limited data instead of actual data and model coal-fired power plants under strict constraints, such as running in a steady state.

2.2 Visual Analytics for Industries

Existing visual analytics systems help users better understand industrial data and cover a wide range of industries [14, 60, 70]. The energy industry is closely related to our work, so we investigated many visual analytics systems designed for it. Höllt et al. [28,29] and CasCADe [30] model specific industrial scenarios in 3D because contextual information is very important for the energy industry, but this also makes the aforementioned systems only applicable to the corresponding scenarios. We also investigated other energy industrial visual analytics systems for high-dimensional hierarchical time series data. Xiao et al. [62] propose a visual analysis method for the optimization of transformer substations. Maljovec et al. [38] simulate and optimize the nuclear reactor process with visual analytics. However, to the best of our knowledge, no existing visual analytics systems are designed for high-dimensional hierarchical time-series data that can support time-lag analysis. This is why specially designed visual analytics systems are required to analyze the complex dynamics of coal-fired power plants.

Additionally, as the time-lag-aware analysis is vital for coal-fired power plants, we surveyed visualization techniques designed for temporal features. Line charts are frequently used to depict linear time (e.g., PlanningVis [49] and Zhou et al. [71]). For periodic time, Wu et al. [61] use calendar charts, and ViDX [63] uses a radial graph. Besides, the

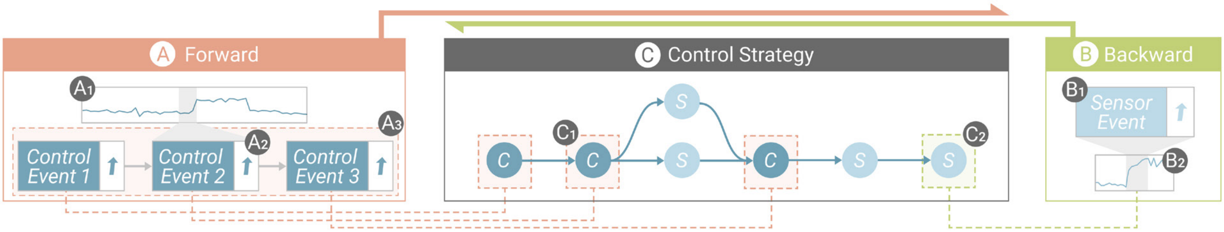

Fig. 2. Concept explanation and task abstraction. A) In the *forward analysis*, the user specifies a partial control strategy (A3), where each control event (A2) is discretely obtained from the time series (A1). B) In the *backward analysis*, the user observes anomaly (B2) from the sensor state time series and looks for the responsible control strategy (C). C) A complete control strategy. Obtained from different known conditions of the forward and backward analysis, for example, C1 corresponds to A2, and C2 corresponds to B2.

Gantt graph [31] is efficient for displaying time intervals. Since time series in the coal-fired power plants are all linear, we employed the line chart design. There are also numerous techniques for revealing the relationships between time series. ViDX [63] and OrderMonitor [51] adopt the marey graph. The parallel coordinates plot [44] can be used similarly. However, these techniques are incapable of presenting temporal trends and time lags. VisCas [13] assists in analyzing the cascading propagation of spatiotemporal events but does not consider time lags. Compass [16] is designed to analyze causality and time lags but only helps with ego-networks relationships. Therefore, we designed the graph and strategy views to simultaneously visualize the relationships, propagations, and time lags of time series in coal-fired power plants.

2.3 Event Pattern Recognition Techniques

According to the survey of Guo et al. [26], event pattern recognition techniques are categorized based on four dimensions: data scales, automated sequence analysis, visual representations, and interactions. The method used by ECoalVis belongs to the category *Filter or Query* under the dimension of the interaction. There are many *Filter or Query* methods, but none can solve the problem of ECoalVis.

Many auto-mining methods like Frequence [44] extract the frequent event patterns, but in coal-fired power plants, there may be some critical but infrequent strategies. There are also query-based methods. Sequence Synopsis [9] supports filtering similar events. Fails et al. [18] propose a visual system to query event sequences but cannot help users specify the relationships between time series.

Some query approaches assist users in specifying the relationships between events. Krause et al. [32] propose a method to query the branching event sequences. MAQUI [33] allows mining for recursive event sequences. However, they only match consecutive event sequences and cannot perform fuzzy matching. Activitree [52] uses a graph-based design to query related events, but the complex relationships between sensors will cause many crossings. The work of Lee et al. [34] is also inspirational but cannot calculate the time lag.

Therefore, we designed an interactive method to support specifying time-series relationships and fuzzy matching.

3 Data and Task Abstraction

This section presents the background of our study, the description of relevant data and concepts, and the summarized user requirements.

3.1 Background

To characterize the workflow of analyzing control strategies of coal-fired power plants, we collaborated closely with four domain experts, EA, EB, EC, and ED, in the past year. EA and EB are researchers from an intelligent city research team, and they have decades of experience in developing data-driven approaches for the energy sector. EC has worked as a senior engineer at a coal-fired power plant for more than three years and is extremely knowledgeable about power plant operation. We also invited ED, a Ph.D. candidate in energy science, to join the collaboration, such that we could leverage her expertise to better understand the rationales behind the diverse control strategies.

Through the collaboration, we learned that different control strategies have different effects on the states of a coal-fired power plant. These effects generally form cascading relationships, that is, one effect occurs due to the executed control strategy and/or other relevant effects, and these effects may eventually lead to changes in efficiency. However, due to the complexity of the power plant environment, linking causes with effects or vice versa in such an environment is a challenging yet valuable task. On the one hand, the engineers at the power plant need to know what *effects* a control strategy will have on the power plant states to reduce decision errors and facilitate operating confidence. On the other hand, the engineers need to determine the *causes* of the abnormal power plant states like sudden efficiency drops, such that the engineers can revert incorrect control strategies and take mitigation actions. Therefore, we summarized two types of analyses based on the direction of the analysis between causes and effects:

Forward analysis (*causes* → *effects*). The experts need to determine the cascading impact of diverse control strategies across different sensors and eventually on the efficiency of the power plant. To perform such a task, they want to see what happens if some control valves are manipulated under certain states. Moreover, subsequent control actions can be discovered with this type of analysis, answering questions like "*what should be adjusted next after fan A is increased to accept more oxygen?*" Therefore, the proposed system should help the experts formulate their queries into a partial control strategy, where the relationships among time series changes can be described with chain reactions. Based on such a partial control strategy, the system should efficiently find all matching control strategies and present the impact of these strategies for further analysis.

Backward analysis (*effects* → *causes*). Inappropriate control strategies may lead to the power plant's anomalous behaviors, which are reflected in the sensor time series. Hence, the experts would like to find anomalies in several critical sensors and identify the control strategies responsible for these anomalies. Questions can be asked like "*The log shows that the negative furnace pressure was abnormal; why is this happening, and which control strategies have problems?*" The insights obtained from the backward analysis can help the experts discover which actions in the control strategies lead to the anomalies and revise the control strategies to avoid similar anomalies in future operations. To support this type of analysis, the proposed system needs to enable the experts to inspect and select a time range of interest, and the responsible control strategy that comprises a series of control actions and state changes should be extracted and visualized.

In both directions of analysis, the experts are concerned about the paths of impact propagation among the sensors and the time lags of each propagation, such that control strategies can be constructed or refined based on this information. For example, after the burner is turned up, the experts would like to know how long they should wait to increase water cooling. Hence, time lags should be visualized along with key control actions and state changes to facilitate informed decision-making.

3.2 Data and Concepts

A historical dataset collected from a coal-fired power plant was used to conduct this study. The dataset comprises two types of data: a) **Structure data** describes the hierarchical structure of a coal-fired power plant, which comprises three levels: *components*, *units*, and *sensors*. These components can be further divided into three stages, *coal pul-*

verization, *burning*, and *steam circulation*, based on the workflow of generating electricity from coal. b) **Sensor data** comprises a set of time series collected from sensors in a coal-fired power plant. The sensors not only record the states (such as temperatures) inside the power plant, but also track the control feedback (such as valve openings). The time series for each sensor is a sequence of samples 1 minute apart.

Sensor events can be extracted from the sensor data (Sect. 4.1.1). Each sensor event is characterized by the corresponding sensor and one of the following temporal patterns, *rising*, *falling* or *stable*. We also categorize these events by the types of sensor, namely, the control feedback (**control events**) or state sensors (**state events**).

We further formalize a **control strategy** (Fig. 2C) as a directed acyclic graph, where each node is a sensor event, and each edge implies the propagation of impact. Additionally, each edge is associated with a time duration, indicating the time lags in the propagation.

3.3 Requirements Analysis

We followed the nine-stage design study methodology framework [46] to iteratively *discover* the user requirements by reviewing the related literature, obtain domain insights from expert interviews, and discussing the design ideas with the experts. Finally, we conclude five user requirements as follows. In particular, R1 and R2 are for the forward and backward analyses, respectively, and R3-5 are for the evaluation of the control strategies in both analyses.

R1: Extract the impact of control strategies with time series queries. In the forward analysis, the experts need to find the impact of certain control strategies. Hence, the proposed system should support the intuitive query with partial control strategies that describe the time-series changes in the sensors of concern (rising, falling, or stable) and the temporal relationships between these changes (the order of propagation). Thereafter, the matching control strategies and their impact should be efficiently extracted from the historical data, and the similarities and differences among these strategies should be visualized to help the experts select a strategy of interest for further analysis.

R2: Identify responsible control strategies for anomalies in important sensors. In the backward analysis, the experts need to identify anomalies (e.g., sudden rises or drops) of important sensors, such as the efficiency and pollutant emissions, which should be visualized in the proposed system. To find the causes of the anomalies, the experts should be able to select the period of time that comprises an anomaly, and the system should efficiently search through the historical data and extract a responsible control strategy constituted by multiple cascading sensor events, including the selected anomaly. Further analysis of this control strategy can help determine the cause of this anomaly.

R3: Explore the spatial propagation of control strategy impact. Visualizing the spatial propagation of impact is critical to the analysis of control strategies, where the experts can discover the involved components, units and sensors and identify the propagation paths. We initially adopt a *relationship-oriented* layout, visualizing the propagation on a distorted schema of the power plant to reflect the strengths of the relationships among the involved sensors. The experts appreciate the intuitiveness of this design, but they also request to add a *context-oriented* layout without schema distortion, so they can interpret the distribution of the involved sensors faster in a more familiar context.

R4: Obtain the temporal cascading of control strategy impact. Analyzing how long it takes to propagate impact from one sensor to another in a control strategy can help the experts understand the strategy's execution flow, such that more control actions can be incorporated into the strategy with confidence. The proposed system should be able to infer such time lags and visualize them along with the topology of the control strategy. This visualization can also reveal that some sensors may be missing from the extracted control strategy based on large time lags. Hence, the system should allow the experts to add or remove sensors from the topology to iteratively guide the extraction model.

R5: Inspect the details of the sensor time series. The experts request to see raw sensor data when analyzing control strategies. For example, they may search for a specific sensor to confirm whether or not it has changed during the execution of a control strategy. A query interface and time-series visualization should be presented to support such analyses. Moreover, the system should also enable the experts to perform various operations on time series, such as aggregating the sensors of the same type by sum or average, or computing the difference between the sensors on left and right sides. This utilizes the experts in discovering insights from the in-depth analysis of raw sensor data.

4 ECoalVis

To meet the user requirements summarized in Sect. 3.3, we proposed ECoalVis, a novel interactive system for experts to visually analyze the control strategies of coal-fired power plants. During the design procedure of ECoalVis, all the domain experts were also tightly integrated in the discussion of design alternatives. An overview of the system is shown in Fig. 3. ECoalVis comprises three modules, namely, data storage, backend, and frontend. The data storage module (Fig. 3A) maintains the structure and sensor data and builds spatial and temporal indexes for accelerating data queries. The backend module (Fig. 3B) can extract control strategies from the sensor time series with two types of query corresponding to the forward and backward analyses, and iterative refinement of control strategies is also supported in this module. The frontend module (Fig. 3C) consists of four views, namely, the filter, graph, strategy, and detail views. These four views combined facilitate the intuitive query and in-depth analysis of the control strategies.

This section presents the detailed implementation of ECoalVis. We first introduce the algorithm in the backend that extracts the desired control strategies, and then describe the visual design of the four views in the frontend to show how they can support the comprehensive analysis.

4.1 Time-Lag-Aware Extraction of Control Strategies

To extract control strategies from the sensor data, we propose a time-lag-aware model to answer two types of query, the forward and backward queries, which provide support for the forward (**R1**) and backward (**R2**) analyses, respectively. For the forward queries, the input is a partial strategy that comprises a sequence of sensor events arranged according to their order of occurrence. As the complete control strategy is complex, there will be errors or omissions in the user's query input. The model should fuzzy match the given partial strategy with the sensor events and find the matched strategies and their cascading impact. For the backward queries, the input is a time range of an important sensor (such as the efficiency) that contains a fluctuation. The model need to expand from the given sensor to find the correlated sensor events, obtaining a complete control strategy that is responsible for the fluctuation. We discuss how the model handles these two types of query in the following sections.

4.1.1 Forward Queries

The forward queries help users obtain possible control strategies and their cascading impact by specifying partial strategies (**R1**). A partial control strategy U consists of a sequence of L sensor event groups $U = \{U_1, U_2, ..., U_L\}$. Each sensor event group U_i comprises a set of sensor events $U_{i,j}$, each identified by a sensor ID and a time series pattern. Since the direct specification of a time series pattern is difficult, we allow users to constraining the temporal trend of a sensor event by choosing an option from *rising*, *falling* and *stable*. Moreover, to maintain the flexibility of the queries, the model performs fuzzy matching based on the given partial strategy, allowing the absence of some sensor events in the matched strategies, such that diverse results can be obtained. This is inspired by string matching methods usually used in search engines, where "fuzzy search means the process of finding strings that approximately match a given string" [58]. We regard a partial strategy as a subsequence and a complete control strategy as a string. The goal of the queries is to find a set of the control strategies that fuzzy match the given partial one, along with the time lags of impact propagation and a matching score for each matched strategy. To answer the queries, the model employs a three-fold approach:

Discretizing time series. First, we discretize sensor time series into events sequences (Fig. 3D). The increases and decreases in the time series will lead to peaks. There are relationships between the start/end time point and the extremum. For example, as shown in Fig. 3F, we circled the first minimum in the decreasing series. It corresponds to the end time point in the event. As the ternary search [10] efficiently finds extremum, we employ it to perform such discretization. However, since the ternary search only works for single peaks, we set a sliding window to address this limitation. Besides, we also fine-tuned a threshold to

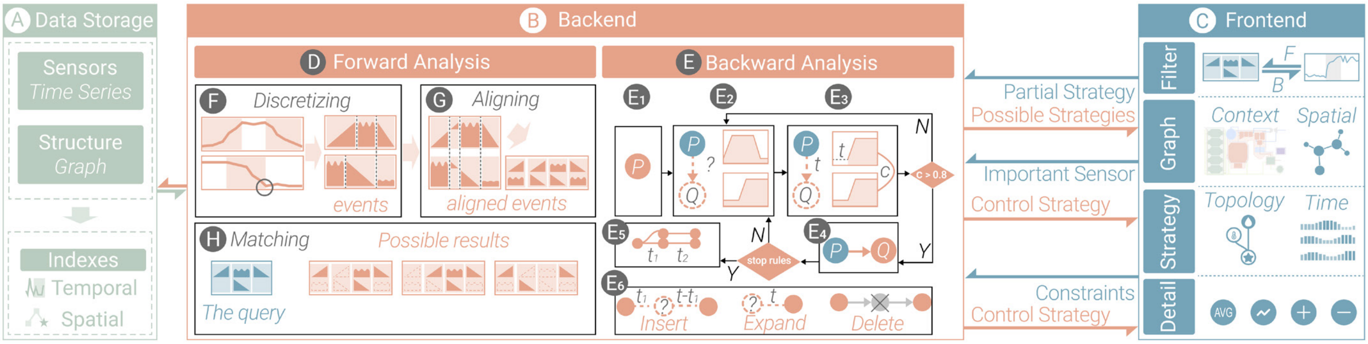

Fig. 3. The system overview of ECoalVis. ECoalVis comprises three parts, namely, data storage (A), backend (B), frontend (C). A) Data are processed and indexed in the data storage. B) The backend comprises the algorithms for the forward (B1) and backward analyses (B2). C) The frontend combines the filter, graph, strategy, and detail views to support extracting and analyzing control strategies.

reduce the impact of time series jitters. After this step, a sequence of sensor events is obtained from the time series of each sensor.

Aligning sensor events. The sensor events extracted in the first step are not aligned because the discretization is performed individually for each sensor. Hence, in this step, we split the extracted sensor events to align the start and end time of these events as illustrated in Fig. 3G. The resulting time series can be represented with a sequence of sensor event groups $S = \{S_1, S_2, ..., S_T\}$, where T is the number of events in each time series, and S_i is a sensor event group where all events in the group can be seen as occurred simultaneously.

Matching partial strategy. In this step, we match the given partial control strategy U with the extracted sequence of sensor event groups S (Fig. 3H). Such matching can be formulated as the longest common subsequence problem, and thus the dynamic programming approach [3, 27] can be employed to solve this problem. To adapt this approach for fuzzy matching, we assume two groups S_i and U_j match if there is at least one identical event in both groups. We also encode the temporal trend of a sensor event with a bit set of length 3 (each bit corresponds to a time series pattern in rising, falling and stable), such that fast matching can be achieved by intersecting two sensor events, i.e., $S_{i,k} \wedge U_{j,k} \neq 0$ indicates a positive match for sensor k. Hence, we can modify the recursive equation in the dynamic programming approach as follows:

$$F_{i,j} = \begin{cases} F_{i-1,j-1} + 1, & \exists_k \; S_{i,k} \wedge U_{j,k} \neq 0, \\ max(F_{i-1,j}, F_{i,j-1}), & \text{otherwise}, \end{cases}$$

where the state matrix $F_{i,j}$ semantically indicates the maximum number of the matching groups before U_i and S_j. We also maintain the matching sensor events $G_{i,j}$ between the groups S_i and U_j, such that we can backtrack the dynamic programming procedure to produce the sequences of the matching events as the output of this step. Furthermore, we compute a match score for each resulting sequence based on the intuition that matching more groups and more events is preferred. We denote the numbers of matching groups and events for a resulting sequence as F and G, and the match score is calculated as $F + G/N$, where N is the total number of sensor events.

After the sequences of the matching sensor events are determined, we expand these sequences into control strategies by launching the breadth-first search procedure similar to the backward queries from each sensor event towards both directions (ahead and behind), obtaining a complete control strategy and its cascading impact for each sequence.

4.1.2 Backward Queries

The backward queries help users search for the control strategies responsible for the anomalies the users observed (**R2**). Our approach is based on the cross-correlation analysis studies [43]. If the impact of sensor A propagates to sensor B after t time steps, then we will see that the time series of A is highly correlates with that of B after A and B are aligned by the time lag t (Fig. 3E3). Therefore, we employ a time-lag-aware breadth-first search approach to extract control strategies.

For the backward queries, users will specify a time range on a sensor time series that contains the anomaly. We place this sensor as the last one in the extracted control strategy and attempt to expand from this sensor to complete the control strategy. The detailed steps of the breadth-first search are described as follows:

1. The specified sensor is pushed into the search queue (Fig. 3E1).
2. The first sensor P in the search queue is taken (Fig. 3E2).
3. For each sensor Q in the reverse order of the power generation workflow, we offset the time series of Q by t minutes for each t between 1 and 15 minutes, which is the predefined upper limit for the time lags, and compute the Pearson coefficients c between the time series of P and Q during the specified time period. The Pearson coefficient [42] is a number between -1 and 1, indicating the linear correlation (Fig. 3E3).
4. If c exceeds 0.8, which is a threshold obtained by trial-and-error, we will assume that P and Q are correlated and add Q into the search queue and the control strategy (Fig. 3E4).
5. The search terminates if the search queue is empty or the duration of the control strategy has exceeded 2 hours; otherwise, the search will return to step 2 and continue (Fig. 3E5).

After the search completes, the extracted control strategy will comprise all sensors that are relevant to the changes in the specified sensor. The time complexity of the search is $O(N^2M)$, where N is the number of sensors, and M is the length of time range. Such a procedure is also applicable to the iterative editing of the control strategies (Fig. 3E6). Specifically, for expanding a leaf sensor in a control strategy, we can start the search from the specified sensor; and for inserting additional sensors between two connected sensors, we can start the search from the latter sensor and terminate once the former sensor is reached.

4.2 Filter View

The first step of the analysis is to filter interesting control strategies. As users have different query requirements, we design two types of interaction in the filter view to accommodate the forward and backward analyses, respectively. Therefore, users can quickly identify the desired control strategies. For the forward analysis (**R1**), we design an input panel to help specify user-desired patterns and a ranking panel to select a control strategy. For the backward analysis (**R2**), a line plot is used to visualize the time series of important sensors and help users select a time range that comprises anomalies.

Input panel. User-desired patterns are actually a partial control strategy according to Sect. 3. A partial control strategy is a multi-sensor event sequence. As every event has many properties, we design an event glyph (Fig. 4E) with user-concerned details [4, 66]. The glyph comprises four parts: a) the trend shapes (Fig. 4E3) depict the changes, including rising, falling, and stable; b) the sensor type is encoded by semantic icons (Fig. 4E1) to facilitate identifiability; c) the color hues encode four discrete stages (Fig. 4E1); d) the annotation before the sensor name (Fig. 4E2) distinguishes control (V) and state events (S). The color hues and icons are consistent throughout the system.

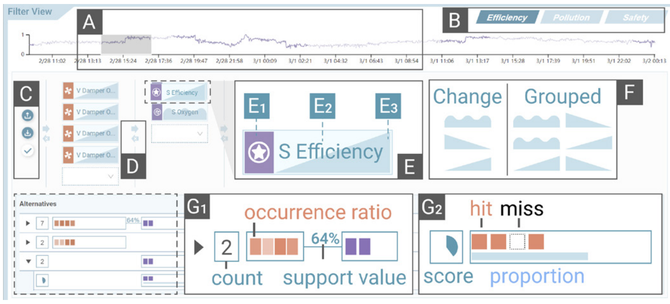

Fig. 4. The design of filter view. A-B) Grouped line charts visualize the time series of important sensors. C-D) Menu buttons help edit the input panel. E) The event glyph visualizes the stage, type, name, and trend of a sensor. F) The different grouped shapes represent possible trends of a sensor. G) A table-based ranking view compares and evaluates the alternative control strategies in the list mode and group mode.

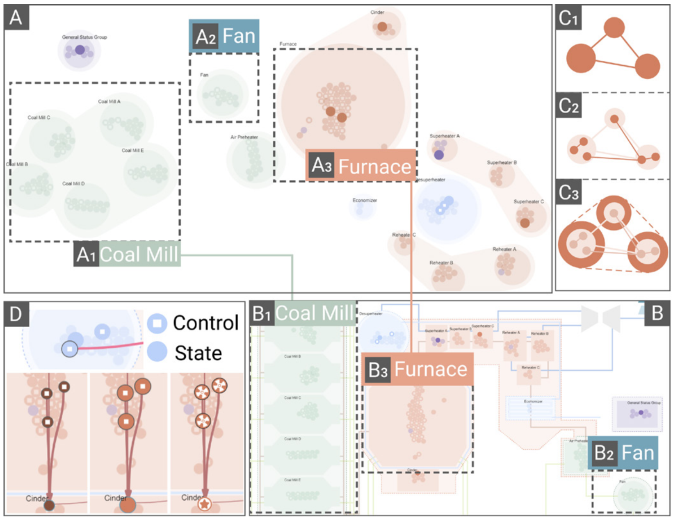

Fig. 5. The visual design of the graph view. A) This reveals the relationship-oriented view. B) This reveals the context-oriented view. C) It illustrates the three steps of the hierarchical force-directed layout algorithm. D) It introduces the design of sensors, units, and components.

To help users specify the relationship between events, we design a multi-column view. According to the discussion with the experts, we find that users focus on co-occurrence and precedence. Therefore, we split the input panel into multiple columns. Events in the same and different columns correspond to co-occurrence and precedence, respectively. To reduce the interaction burden, we offer three additional changes via grouping (not rising, not falling, and unstable as shown in Fig. 4F) and allow users to import and export files (Fig. 4C).

Ranking panel (Fig. 4G). The ranking panel has two modes: *list* and *group* mode. The *list* mode aims to help compare and select possible control strategies with user-desired patterns (Fig. 4G2). Inspired by LineUp [24], we design a table-based view for easy ranking and comparison. For sequence-level comparison, each row shows an alternative control strategy, and the pie chart reveals the matching score. For column-level comparison, each column contains events matching the above input panel, and the light blue bar below encodes the proportion of matches in this column. For event-level comparison, the colored rectangles in the column display match details. The *group* mode aims to visualize summaries of different patterns. A summary row (Fig. 4G1) is added to the top of each group of alternative control strategies. At the front is the count of alternatives in this group. The opacity of the colored rectangle indicates the occurrence rate of the corresponding event. Besides, the support value of the link is shown in Fig. 4G1. We calculate the support value using the formula $SV = N/M$. In the formula, N means the times this control strategy happens, and M means the number of all related control strategies. The support value indicates how frequently a control strategy occurs.

Line chart (Fig. 4A). A line chart depicts the time series of important sensors such as power generation efficiency, NOx emissions, and furnace pressure. Users can switch among important sensors via the tabs (Fig. 4B). We allow users to set a threshold for the first derivative because they are concerned with the time range in which there is a significant change. Opacity encoding helps highlight the critical areas that change fast. Brushing a time range on the line chart will invoke a backward query, which extracts the responsible strategy.

Justification. For the input panel, there are many event pattern recognition techniques. We discuss them in Sect. 2 in detail. Besides, we tried graph-based design initially, as it is efficient for showing relationships [8]. We designed event glyphs and used them as nodes. However, specifying complex time series relationships with the node-link diagram can be rather cumbersome and may lead to many link crossings. Therefore, we simplified the design into a multi-column view. For the event glyph, we encode stages with color hues, as it is the second effective channel to encode category [7, 39]. We encode sensor types by icons because there are so many of them. For the ranking panel, we are inspired by LineUp [24] and the comparison panel design from Wongsuphasawat et al. [59]. The table view helps users compare alternatives on column- and event-level quickly. As the compact technique is always used to group similar patterns, we improve the table view through it to make the aggregation pattern clear [1].

4.3 Graph View

The graph view adopts a dual-mode design to provide both spatial and contextual information about the coal-fired power plant (**R3**). Analyzing the relationship between sensors in control strategies helps gain insights into the power plants, so we design a relationship-oriented view. This view highlights the spatial propagation and correlation among the components, units, and sensors. In addition, depicting the complex hierarchical structure and actual spatial location is also very important, so we design a context-oriented view.

In both modes, we employ hierarchical encodings for different levels. At the sensor level, the hollow circles encode the control feedback sensors, while the solid ones encode the state sensors (Fig. 5D). At the unit and component levels, we adopt colored shapes with different opacity. Moreover, sensors that have changed significantly during the user-selected time range will be enlarged. After clicking the sensor, its impact propagation links will appear.

Relationship-oriented visualization (Fig. 5B). Initially, we found mismatches between actual spatial location and data relation. For example, as shown in (Fig. 5B), the fan (Fig. 5B2) located far from the coal mill (Fig. 5B1) and furnace (Fig. 5B2) significantly impacts them through the wind. In this situation, the actual position will lead to ambiguity, and users may feel that the correlation between the fan and the coal mill is weak. Therefore, inspired by the multi-level force-directed graph algorithm [53], we adopt a node-link diagram-based design. However, it will take a long time to compute the layout because of the large scale of sensors. Therefore, the layout algorithm is improved into a hierarchical one to minimize the search space. First (Fig. 5C1), we compute the force-directed graph layout for the units and then fix their position. Next (Fig. 5C2), the sensor-level computation is constrained by the units, so the search space quickly converges. Then (Fig. 5C3), to distinguish components, the convex hull calculated from common tangents of units displays component-level layout.

Context-oriented visualization (Fig. 5A). When users focus on the context, such as workflow and actual location, the relationship-oriented layout is not intuitive enough [15,55]. This is because domain engineers are more familiar with abstract workflow diagrams (Fig. 5B) than node-link diagrams. Therefore, we simplify the power plant structure as an abstract workflow diagram to provide a context overview.

Justification. The graph view visualizes relationships and propagation graphs, so we considered two most commonly used techniques: matrix-based and node-link designs. The matrix design is rejected because it is non-intuitive to track the propagation paths. As for the node-link design, though it reduces ambiguities, the experts are unfamil-

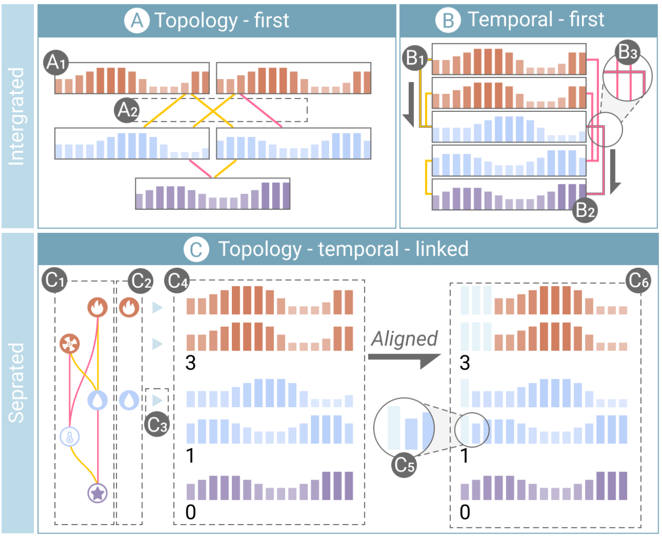

Fig. 6. Design choices for strategy view. A) A topology-first intergrated diagram. B) A temporal-first intergrated diagram. C) Our current design of topology-temporal-linked layout.

iar with it. Hence, we adopted a context-oriented diagram familiar to experts to help them locate the critical sensors. Also, after comparing color-, size-, and texture-based design [23], we employed the size-texture-mixed one, which highlighted sensors most clearly (Fig. 5D).

4.4 Strategy View

The strategy view is designed for analyzing the temporal cascading impact of control strategies (**R4**). There are two aspects to this analysis task. First, in terms of temporal dimension, the user is concerned about the exact time lag of the impact from sensor A to sensor B. In addition, the users also want to see detailed time series. Second, in terms of topological dimension, the user is concerned about the cascading impact between sensors, especially the correlation and propagating direction.

Temporal visualization. A bar-based view is used to display discretely sampled time series data (Fig. 6C4). Not only the height of each bar reveals the value, but to further highlight the trend, we also set the opacity of the bar accordingly. The benefit of this encoding is that it clearly depicts the time lag steps between sensors (Fig. 6C5). In addition, the sensors with the same delay are grouped together to facilitate exploring patterns, such as synergistic and antagonistic adjustment. Moreover, this view can be aligned to see overall trends and display propagation more clearly. We use triangles on the left to mark control sensors (Fig. 6C3), while the others are state sensors.

Topological visualization. The control strategy is shown as a node-link diagram (Fig. 6C1), where nodes encode sensors and links encode the cascading impact between sensors. The thickness of the edge encodes the correlation strength, and the color hue encodes whether the correlation is positive (yellow) or negative (pink). We place the extra column of semantic icons next to user-specified sensors (Fig. 6C2).

Interaction. On the one hand, with many control strategies taking place simultaneously in complex coal-fired power plants, inevitably, *false correlations* will occasionally occur. On the other hand, some sensors might be missing because of a large time lag. Hence, we design editing interactions to assist users in obtaining the cascading impact of the temporal control strategy. There are three types of editing interactions, namely, insert, expand, and delete. Insert is for exploring the missing sensors between two levels. Delete is for removing the uninteresting or wrong sensors. Expand is for obtaining correlated sensors with a large time lag. Users can click on the top button group to quickly edit the control strategy (Fig. 1C1).

Justification. There are also two integrated design choices for the strategy view, one is topology-first, and the other is temporal-first. The topology-first design is based on a directed graph (Fig. 6A), where each node in the graph is a time series chart that helps users inspect details (Fig. 6A1). The edge of the directed graph reflects the correlation (Fig. 6A2). The thickness of the edge encodes the strength, the color hue of the edge encodes the positive and negative (cyan is positive, red is negative), and the length of the edge reveals the time lag [2]. The topology-first design clearly shows the cascading impact but still has two disadvantages. First, although the time curve can reflect the relative length of the time lag, it is still hard to compare. Second, it is hard to evaluate it absolutely instead of relatively, as they are not aligned.

The temporal-first design is based on the bar-based chart (Fig. 6B). Each row corresponds to the time-series data (Fig. 6B1), and the links on the left side of the data are negatively correlated while positively on the right side. The thickness also indicates the strength of the correlation. This design reflects the time lags clearly and is easily aligned for inspection. However, the links have many crossings (Fig. 6B3), which clutter the cascading propagation paths. Besides, it is difficult to follow the propagation. For example, in (Fig. 6B), if the user wants to track the propagation from B1→B2, he needs to first follow the left link to find the middle sensor and then follow the right link to find B2. Therefore, we finally adopt the visual design of topology-temporal-separated design to consider both advantages.

4.5 Detail View

To support efficient inspection of data details (**R5**), we provide a detail view (Fig. 1D) to show raw data for all sensors within a specific time range. Users can add desired sensors in the search box on the top (Fig. 1D1). Then from top to bottom, the data details of each sensor are shown one by one. The structure diagram on the left (Fig. 1D3) shows the spatial position of the sensor in the coal-fired power plant, providing context information. The line chart on the right (Fig. 1D2) shows the time-lag-aligned time series during the selected time range. In addition, users also need to view the aggregation patterns of multiple sensors. We provide three aggregation interactions: sum, difference, and average (Fig. 1D4). In this way, users can inspect the data details that may help them confirm the conclusions drawn in the previous views.

Justification. The sensor data in coal-fired power plants is high-dimensional temporal data. Liu et al. [37] and Wei et al. [57] surveyed commonly used views for high-dimensional data and sensor data, including axis-based, glyphs, pixel-oriented, hierarchy-based [5], scatter plots [56], and animation designs. The line plot is the most efficient for showing trends in all of the methods, so we use it.

5 Implementation

We employed the common front-end (React, Redux, and Typescript) and back-end (Flask, Python, and Numpy) libraries to develop ECoalVis. The coal-fired power plant data is stored in many excel files, which are exported from a meta dataset. We open source ECoalVis on Github. The data configuration and guidance are available in the repository. https://github.com/ECoalVis/ECoalVis.

6 Evaluation

In this section, we present two usage scenarios explored by us and reproduced and evaluated by the experts using ECoalVis on real-world historical datasets in the forward analysis and the backward analysis, respectively. Then we interviewed the experts, and their feedback demonstrated the effectiveness and usability of ECoalVis.

Dataset. The dataset contains 8 components, 17 units, and 203 sensors. It collects data from February 28, 2019, to May 8, 2019. As the sensor records data every minute, the length of each sensor's time series is 100080. The time lag's order of magnitude is 10, and the average length of events is 11.17 minutes.

6.1 Backward: Find Causes for Efficiency Increase

In the backward analysis, the users usually observe some anomalies and want to find a responsible control strategy.

Step 1: Filter the anomaly. After loading the dataset, the user first explored the line plot in the filter view (Fig. 1A). The purple line reveals the power generation efficiency, one of the most important sensors. The user observed that the efficiency increased between 2/28 14:12 to 2/28 16:46 and then reached a stable state (Fig. 1A1). The user selected the interval to explore the pattern.

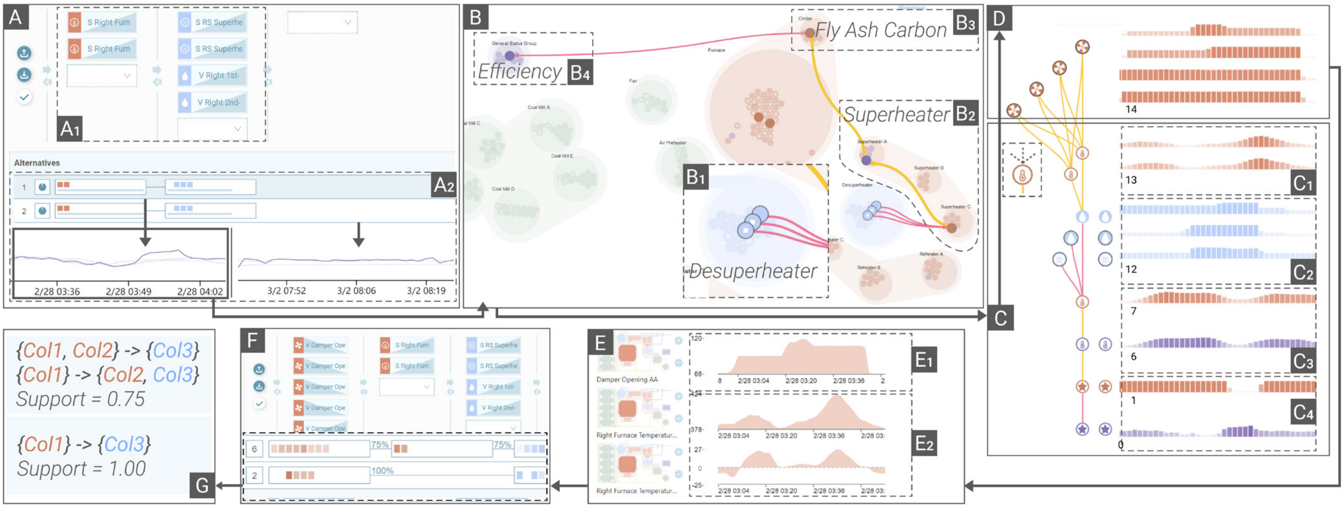

Fig. 7. The analysis process of obtaining effects of unbalanced burning. A) shows the user how to specify desired patterns and choose the best match control strategy. B) displays the spatial propagation and reveals a synergistic adjustment in the desuperheater. C) depicts the cascading impact of unbalanced burning with time lags. D) is the reasoning result of unbalanced burning. E) shows detailed evidence to support the deduction in (D). F) and G) help to verify the conclusion.

Step 2: Explore the spatial propagation impact. After the back-end algorithm extracted the control strategy responsible for this anomaly, the user started analyzing it. The user was attracted by highlighted circles (Fig. 1B) in the graph view and clicked a valve in the furnace to explore its impact propagation. The user found that the impact propagated from furnace to cinder and finally to the efficiency (B1 →B2 →B3 in Fig. 1). The user also found that there was a highlighted sensor in the fan (Fig. 1B4). As the fan was only in the wind circulation, the user guessed that this anomaly might be related to it closely. However, the user was unsure how and why the fan was adjusted, so the user continued exploring more details.

Step 3: Obtain the temporal cascading impact. The user checked the details in the strategy view. The user first edited the control strategy to understand why the fan was adjusted (Fig. 1C2). In the editing process, the user first expanded the fan node but got an unexpected sensor. So, the user deleted it and then checked the time-lag-aligned temporal view (Fig. 1C5). The user found that the fan decreased one minute after reducing the total air volume. Then, after another two minutes, the dampers in the furnace increased. Finally, the fly ash carbon decreased and the efficiency increased (C2 →C3 →C4 in Fig. 1). Based on domain knowledge, the user guessed that the reason why the efficiency increased might be reducing excess air in the furnace. The user needed to inspect more evidence.

Step 4: Inspection and check. The user further inspected some related sensors in the detail view, including oxygen, total air volume, and negative furnace pressure. The user clicked the align button and found the oxygen was stable (Fig. 1D2) while the other changed according to the efficiency (D5 →D6 in Fig. 1). Therefore, the user confirmed that reducing excess air while keeping oxygen could increase the efficiency. Finally, the user imported events one by one into the filter view (Fig. 1A3) and found the support value was 64%. In the end, the user confirmed that the control strategy frequently occurred in history.

6.2 Forward: Obtain Effects of Unbalanced Burning

In the forward analysis, the users usually want to explore the possible cascading impact after adjusting some valves.

Step 1: Specify desired patterns. The user observed the burning in the furnace was not balanced and geared to the right from the log history. According to his expertise, the user should adjust the water valve on the right side to alleviate this situation. However, the user knew little about the possible cascading effects and how long it took to affect the efficiency. Therefore, the user edited the filter view to specify the control strategy. The user first added the increasing event of the right furnace temperature and then added the opening wider event of the right-side water valves (Fig. 7A1). After the possible control strategies were shown, the user browsed the alternatives from top to bottom as they were sorted decreasingly by scores. The user selected the first one because the user found a more dramatic change of efficiency in it than that in the second one (Fig. 7A2).

Step 2: Explore the propagation. The user explored the spatial propagation of the selected control strategy. To focus more on the relationship and gain insights into the correlation between sensors and components, the user used the relationship-oriented mode of the graph view (Fig. 7B). The user clicked the highlighted circles in the desuperheater to see the impact of water valve adjustment. The colored links revealed that the desuperheater was negatively correlated with the superheater (B1→B2). Then, the positive correlation spread further to fly ash carbon (B2→B3) and finally affected efficiency (B3 →B4) in Fig. 7B. Though the user knew the correlation between each pair of sensors, the user still wondered about the exact time lags.

Step 3: Obtain the cascading impact. The user started to examine the temporal cascading impact. The user first aligned the time series data of all sensors and observed cascading patterns (C1 →C2 →C3 →C4 in Fig. 7). After the unbalanced burning appeared, the right-sided water valve was raised within 1 minute. After another 5 minutes, the relevant temperature was affected to decrease gradually, and finally, the efficiency can be recovered from decreasing. The user confirmed that the timely adjustment of the right-sided water valves could avoid the loss diffusion of abnormal unbalanced burning with these patterns.

Step 4: Reason the control strategy and verify. However, the user still wanted to know why unbalanced burning was caused and whether all unbalanced burning could be treated with such a remedy. Therefore, the user expanded the last level (Fig. 7C1) and gained the full control strategy (Fig. 7D). The user found that four dampers were opened wider before the unbalanced burning occurred, so the user saw their sum in the detail view (Fig. 7E1). Besides, the user also inspected many other sensors to verify the conclusion, such as the temperature difference between the right and left side (Fig. 7E2). After that, the user was sure that the dampers caused the unbalanced burning.

Step5: Evaluate the remedy control strategy. Next, the user evaluated the control strategy of the *Abnormal Dampers →Right Side Unbalanced Burning →Right Water Valves Increasing* into the filter view and switched to the group mode (Fig. 7F) to see aggregation information. In the end, since the support value reached more than 75% (Fig. 7G), the user believed that timely adjustment of the water valves could effectively reduce the negative impact of unbalanced burning.

6.3 Expert Interview

We conducted one-on-one structured interviews with the aforementioned four experts (EA, EB, EC, and ED). We first introduced the basic functions of ECoalVis. Second, the experts reproduced two usage scenarios according to prepared script. Third, they followed the think-aloud protocol [21] and analyzed the control strategies of interest with ECoalVis. Finally, we collected qualitative feedback from the experts on the effectiveness, designs and interactions, and other suggestions.

Effectiveness. All experts spoke highly of ECoalVis and confirmed that this system could be useful in analyzing the control strategies of coal-fired power plants. EC mentioned that in the past, the control strategies of power plants were always evaluated and judged based on individual experience, which cost a long time to inspect multiple sensor values. Besides, engineers could only qualitatively analyze sensor correlation with traditional software, resulting in an insufficient understanding of control strategies. ECoalVis clearly presented the correlation between sensors and the time lags, facilitating further insights and optimization of control strategies. EA told us that the time-lag analysis was precious because it shed insight on how the programming logic of the PID inside the power plants could be optimized. EB thought ECoalVis could assist engineers in making informed adjustments: "*This system can help engineers understand the subsequent impact and make changes in time.*" ED suggested that ECoalVis had great potential in teaching new engineers, helping them learn diverse control strategies.

Designs and interactions. Overall, the experts agreed that ECoalVis was easy to use, and the interactions were convenient and smooth. EA liked the design and interaction of the filter view very much because it could help him "quickly explore the impact of any combination of control adjustments". EB said the results retrieved in the filter view were well organized. EB and EC both praised the responsiveness of the system as they had a good experience in querying with the filter view. ED mentioned that the graph view could quickly help her analyze the impact of various units and sensors and their relationships. She told us that "relationship-oriented graph layout well captures some features like collaborative adjustments." EB, EC, and ED all agreed that the time-lag-aware design of the strategy view was very intuitive and efficient. ED emphasized that the correspondence between the impact propagation of control strategies and the resulting efficiency could be established with this design, so "In the future, ECoalVis will greatly help us build economic power plants."

Improvement. Experts also made constructive suggestions for our system. EB recommended that adding bookmarks to the filter view could help users record query history for future verification. EC suggested that users shall be able to jump between the same sensors in the graph and strategy views to further associate the spatial information of the sensors with the propagation topology. ED commented that the sensors in the graph view could be grouped at a finer granularity, especially for larger units like furnaces. It would be easier to understand if the locations of the sensors were similar to those in the real world. We have optimized the system based on these suggestions.

7 Discussion

In this section, we discuss the implications, lessons learned, and the limitations and future work of ECoalVis.

Implications. In this paper, we propose ECoalVis, a visual analytics system for analyzing control strategies of coal-fired power plants. We discuss the implications of ECoalVis from the following aspects.

Analytical framework. We identified two types of analyses based on the directions between causes and effects and further derived five requirements. Although these analyses were summarized from the domain-specific observations, we argue that such a framework can be adapted to other correlation analysis applications where either causes or effects are unknown and need to be determined. These applications can benefit from such a framework in characterizing similar requirements.

Techniques. We designed an interactive query interface that integrated a tailored model to search for multiple time series based on their relationships. The interface supports users to specify the relationships among time series and present the query results with interpretable confidence. Such an interface can also be applied to other scenarios where the relationships among time series are the focus of the analysis, like querying the diffusion patterns of air pollution based on sensor data.

Applicability. We presented two usage scenarios on the real-world coal-fired power plant data. These usage scenarios preliminarily show the potential of ECoalVis in helping users find relevant control strategies based on the preferred analysis approach and gaining unprecedented insights into the spatial and temporal characteristics of these strategies. The findings delivered by these scenarios were confirmed by the experts and may guide the future operation and optimization of the power plant.

Lessons learned. We present two design lessons during the development of ECoalVis. First, usability and flexibility should be carefully balanced while developing domain-specific visual analytics systems. During the design of the filter view, we initially adopted a graph-based query interface. However, the experts struggled with this query interface, finding it hard to understand because they had never authored such graphs. Therefore, we simplified the query interface by adopting a sequence-based design, where users can arrange sensor events linearly based on the order of occurrence. Despite the loss of certain flexibility, the experts well perceive this design. Second, familiarity matters in the presentation of spatial information. Initially, we wanted to emphasize the impact propagation process in the graph view, thereby designing a hierarchical force-directed layout based on the relationship strengths among the sensors. Despite the clear trend of propagation captured by the layout, the experts failed to locate components and units since the underlying schema of the power plant was distorted. Hence, we developed another mode in the graph view to preserve the schematic positions and the shapes of the components and units.

Limitations and future work. Three limitations are observed in the proposed system. First, the filter view does not offer automatic guidance on what can be specified as the query constraints. Inexperienced users may find it hard to initiate the analysis if they do not know which sensors to query with. A possible mixed-initiative solution is to integrate an anomaly detection model that highlights interesting control strategies, which requires further studies on identifying abnormal strategies. Hence, we resorted to the on-demand query approach and left the mixed-initiative ones as a part of future work. Second, the schema of the power plant in the graph view is not presented in 3D. The experts told us in the interview that a 3D schema may better support them in the analysis because it creates a sense of presence that helps them navigate. However, integrating such a schema is a challenging task, requiring a realistic 3D model of the power plant and a new graph layout to facilitate the immersive propagation analysis in the 3D environment, so we decide to leave it as a future direction. Third, the evaluation of ECoalVis was limited to a small number of expert users. We would like to deploy ECoalVis in the production and collect more use cases and feedback in the future to further validate its effectiveness.

In the future, we will also connect ECoalVis to the streaming data from power plants and adapt it for in-situ analyses, such that control strategies can be analyzed and optimized in real-time. This opens up new avenues for analyzing control strategies, and decision-making modules can be included to facilitate informed selection. We will also try to simplify ECoalVis by adding guidance and integrating machine learning models that detect anomalies automatically.

8 Conclusion

This study proposes ECoalVis, a novel interactive system for experts to visually analyze the control strategies of coal-fired power plants. To address three identified challenges, we collaborated with domain experts and summarized five requirements based on two types of analyses. We developed ECoalVis to suit the experts' needs, integrating a series of well-designed visualization with efficient models to support the in-depth analysis of control strategies. The effectiveness of the system was evaluated with two usage scenarios on a real-world dataset along with expert interviews. In the future, we plan to address the limitations observed in the proposed approach and adapt the system for in-situ analyses with streaming data.

Acknowledgments

The work was supported by NSFC (62072400) and the Collaborative Innovation Center of Artificial Intelligence by MOE and Zhejiang Provincial Government (ZJU). We also thank the anonymous reviewers for their helpful comments.

REFERENCES

[1] N. Andrienko, G. Andrienko, S. Miksch, H. Schumann, and S. Wrobel. A theoretical model for pattern discovery in visual analytics. *Visual Informatics*, 5(1):23–42, 2021. doi: 10.1016/j.visinf.2020.12.002

[2] B. Bach, C. Shi, N. Heulot, T. M. Madhyastha, T. J. Grabowski, and P. Dragicevic. Time curves: Folding time to visualize patterns of temporal evolution in data. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):559–568, 2015. doi: 10.1109/TVCG.2015.2467851

[3] L. Bergroth, H. Hakonen, and T. Raita. A survey of longest common subsequence algorithms. In *Proceedings of Seventh International Symposium on String Processing and Information Retrieval*, pp. 39–48, 2000. doi: 10.1109/SPIRE.2000.878178

[4] M. Brehmer, R. Kosara, and C. Hull. Generative design inspiration for glyphs with diatoms. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):389–399, 2022. doi: 10.1109/TVCG.2021.3114792

[5] M. Burch, H. van de Wetering, and N. Klaassen. The power of interactively linked hierarchy visualizations. *Journal of Visualization*, pp. 1–17, 2022. doi: 10.1007/s12650-021-00818-3

[6] S. Chandrasekharan, R. C. Panda, B. N. Swaminathan, and A. Panda. Operational control of an integrated drum boiler of a coal fired thermal power plant. *Energy*, 159:977–987, 2018. doi: 10.1016/j.energy.2018.06.157

[7] S. Chen, J. Zhang, Y. Zhao, P. L. Rosin, Y. Lai, and L. Gao. A review of image and video colorization: From analogies to deep learning. *Visual Informatics*, 2022. doi: 10.1016/j.visinf.2022.05.003

[8] X. Chen, X. Tang, Z. Luo, and J. Zhang. Evaluating user cognition of network diagrams. *Visual Informatics*, 5(4):26–33, 2021. doi: 10.1016/j.visinf.2021.12.004

[9] Y. Chen, P. Xu, and L. Ren. Sequence synopsis: Optimize visual summary of temporal event data. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):45–55, 2018. doi: 10.1109/TVCG.2017.2745083

[10] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. *Introduction to algorithms*. 2022.

[11] Cotmac Inc. Automation of 30MW Power Plant using PCS 7. `https://cotmac.io/industries/energy/power-plant-automation/`, 2018. [Online; accessed 19-Mar-2022].

[12] M. Cropper, R. Cui, S. Guttikunda, N. Hultman, P. Jawahar, Y. Park, X. Yao, and X.-P. Song. The mortality impacts of current and planned coal-fired power plants in india. *Proceedings of the National Academy of Sciences*, 118(5):e2017936118, 2021. doi: 10.1073/pnas.2017936118

[13] Z. Deng, D. Weng, Y. Liang, J. Bao, Y. Zheng, T. Schreck, M. Xu, and Y. Wu. Visual cascade analytics of large-scale spatiotemporal data. *IEEE Transactions on Visualization and Computer Graphics*, 28(6):2486–2499, 2022. doi: 10.1109/TVCG.2021.3071387

[14] Z. Deng, D. Weng, S. Liu, Y. Tian, M. Xu, and Y. Wu. A survey of urban visual analytics: advances and future directions. *Computational Visual Media*, 9:to appear, 2023. doi: 10.1007/s41095-022-0275-7

[15] Z. Deng, D. Weng, and Y. Wu. You are experienced: Interactive tour planning with crowdsourcing tour data from web. *Journal of Visualization*, 26(to appear), 2023.

[16] Z. Deng, D. Weng, X. Xie, J. Bao, Y. Zheng, M. Xu, W. Chen, and Y. Wu. Compass: Towards better causal analysis of urban time series. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):1051–1061, 2022. doi: 10.1109/TVCG.2021.3114875

[17] M. Draganescu, S. Guo, J. Wojcik, J. Wang, X. Liu, G. Hou, Y. Xue, and Q. Gao. Generalized predictive control for superheated steam temperature regulation in a supercritical coal-fired power plant. *CSEE Journal of Power and Energy Systems*, 1(1):69–77, 2015. doi: 10.17775/CSEEJPES.2015.00009

[18] J. A. Fails, A. Karlson, L. Shahamat, and B. Shneiderman. A visual interface for multivariate temporal data: Finding patterns of events across multiple histories. In *2006 IEEE Symposium On Visual Analytics Science And Technology*, pp. 167–174, 2006. doi: 10.1109/VAST.2006.261421

[19] X. Fan and H. Zheng. Design cems for flue gas from thermal power plant. In *Proceedings of Asia-Pacific Power and Energy Engineering Conference*, pp. 1–4, 2009. doi: 10.1109/APPEEC.2009.4918261

[20] X. Fei, H. Wei, and M. Yong. Real time supervise system of flue gas desulfurization in coal fired power plants. In *Proceedings of 3rd IEEE Conference on Industrial Electronics and Applications*, pp. 747–751, 2008. doi: 10.1109/ICIEA.2008.4582615

[21] M. E. Fonteyn, B. Kuipers, and S. J. Grobe. A description of think aloud method and protocol analysis. *Qualitative health research*, 3(4):430–441, 1993. doi: 10.1177/104973239300300403

[22] J. Fu, H. Xiao, H. Wang, and J. Zhou. Control strategy for denitrification efficiency of coal-fired power plant based on deep reinforcement learning. *IEEE Access*, 8:65127–65136, 2020. doi: 10.1109/ACCESS.2020.2985233

[23] L. Giovannangeli, R. Bourqui, R. Giot, and D. Auber. Color and shape efficiency for outlier detection from automated to user evaluation. *Visual Informatics*, 6(2):25–40, 2022. doi: 10.1016/j.visinf.2022.03.001

[24] S. Gratzl, A. Lex, N. Gehlenborg, H. Pfister, and M. Streit. Lineup: Visual analysis of multi-attribute rankings. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2277–2286, 2013. doi: 10.1109/TVCG.2013.173

[25] H. Gu, H. Zhu, Y. Cui, F. Si, R. Xue, H. Xi, and J. Zhang. Optimized scheme in coal-fired boiler combustion based on information entropy and modified k-prototypes algorithm. *Results in Physics*, 9:1262–1274, 2018. doi: 10.1016/j.rinp.2018.04.045

[26] Y. Guo, S. Guo, Z. Jin, S. Kaul, D. Gotz, and N. Cao. A survey on visual analysis of event sequence data. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–1, 2021. doi: 10.1109/TVCG.2021.3100413

[27] M. Held and R. M. Karp. A dynamic programming approach to sequencing problems. *Journal of the Society for Industrial and Applied mathematics*, 10(1):196–210, 1962. doi: 10.1137/0110015

[28] T. Höllt, J. Beyer, F. Gschwantner, P. Muigg, H. Doleisch, G. Heinemann, and M. Hadwiger. Interactive seismic interpretation with piecewise global energy minimization. In *Proceedings of IEEE Pacific Visualization Symposium*, pp. 59–66, 2011. doi: 10.1109/PACIFICVIS.2011.5742373

[29] T. Höllt, A. Magdy, G. Chen, G. Gopalakrishnan, I. Hoteit, C. D. Hansen, and M. Hadwiger. Visual analysis of uncertainties in ocean forecasts for planning and operation of off-shore structures. In *Proceedings of IEEE Pacific Visualization Symposium*, pp. 185–192, 2013. doi: 10.1109/PacificVis.2013.6596144

[30] P. Ivson, D. Nascimento, W. C. Filho, and S. D. J. Barbosa. Cascade: A novel 4d visualization system for virtual construction planning. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):687–697, 2018. doi: 10.1109/TVCG.2017.2745105

[31] J. Jo, J. Huh, J. Park, B. Kim, and J. Seo. Livegantt: Interactively visualizing a large manufacturing schedule. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2329–2338, 2014. doi: 10.1109/TVCG.2014.2346454

[32] J. Krause, A. Perer, and H. Stavropoulos. Supporting iterative cohort construction with visual temporal queries. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):91–100, 2016. doi: 10.1109/TVCG.2015.2467622

[33] P.-M. Law, Z. Liu, S. Malik, and R. C. Basole. Maqui: Interweaving queries and pattern mining for recursive event sequence exploration. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):396–406, 2019. doi: 10.1109/TVCG.2018.2864886

[34] T.-Y. Lee and H.-W. Shen. Visualization and exploration of temporal trend relationships in multivariate time-varying data. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):1359–1366, 2009. doi: 10.1109/TVCG.2009.200

[35] S. Li, H. Liu, W.-J. Cai, Y.-C. Soh, and L.-H. Xie. A new coordinated control strategy for boiler-turbine system of coal-fired power plant. *IEEE Transactions on control systems technology*, 13(6):943–954, 2005. doi: 10.1109/TCST.2005.854319

[36] Y. Li and R. Li. Simulation and optimization of the power station coal-fired logistics system based on witness simulation software. In *Proceedings of Workshop on Power Electronics and Intelligent Transportation System*, pp. 394–398, 2008. doi: 10.1109/PEITS.2008.103

[37] S. Liu, D. Maljovec, B. Wang, P.-T. Bremer, and V. Pascucci. Visualizing high-dimensional data: Advances in the past decade. *IEEE Transactions on Visualization and Computer Graphics*, 23(3):1249–1268, 2017. doi: 10.1109/TVCG.2016.2640960

[38] D. Maljovec, B. Wang, P. Rosen, A. Alfonsi, G. Pastore, C. Rabiti, and V. Pascucci. Rethinking sensitivity analysis of nuclear simulations with topology. In *Proceedings of IEEE Pacific Visualization Symposium*, pp. 64–71, 2016. doi: 10.1109/PACIFICVIS.2016.7465252

[39] T. Munzner. *Visualization analysis and design*. CRC press, 2014.

[40] I. H. Njoku, C. O. Oko, J. C. Ofodu, and O. E. Diemuodeke. Optimal thermal power plant selection for a tropical region using multi-criteria decision analysis. *Applied Thermal Engineering*, 179:115706, 2020. doi: 10.1016/j.applthermaleng.2020.115706

[41] A. L. Page, A. A. Elseewi, and I. R. Straughan. Physical and chemical

properties of fly ash from coal-fired power plants with reference to environmental impacts. In *Proceedings of Residue Reviews*, pp. 83–120. New York, NY, 1979. doi: 10.1007/978-1-4612-6185-8_2
[42] K. Pearson. Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin philosophical magazine and journal of science*, 2(11):559–572, 1901. doi: 10.1080/14786440109462720
[43] B. Podobnik and H. E. Stanley. Detrended cross-correlation analysis: a new method for analyzing two nonstationary time series. *Physical review letters*, 100(8):084102, 2008. doi: 10.1103/PhysRevLett.100.084102
[44] S. Radoš, R. Splechtna, K. Matković, M. źuras, E. Gröller, and H. Hauser. Towards quantitative visual analytics with structured brushing and linked statistics. *Computer Graphics Forum*, 35(3):251–260, 2016. doi: 10.1111/cgf.12901
[45] H. Ritchie and M. Roser. Electricity Mix. `https://ourworldindata.org/electricity-mix`, 2020. [Online; accessed 11-Mar-2022].
[46] M. Sedlmair, M. D. Meyer, and T. Munzner. Design study methodology: Reflections from the trenches and the stacks. *IEEE Trans. Vis. Comput. Graph.*, 18(12):2431–2440, 2012. doi: 10.1109/TVCG.2012.213
[47] S. Sengupta. The World Needs to Quit Coal. Why Is It So Hard? `https://www.nytimes.com/2018/11/24/climate/coal-global-warming.html`, 2018. [Online; accessed 19-Mar-2022].
[48] C. Sun, J. Chen, S. Cao, X. Gao, G. Xia, C. Qi, and X. Wu. A dynamic control strategy of district heating substations based on online prediction and indoor temperature feedback. *Energy*, 235:121228, 2021. doi: 10.1016/j.energy.2021.121228
[49] D. Sun, R. Huang, Y. Chen, Y. Wang, J. Zeng, M. Yuan, T.-C. Pong, and H. Qu. Planningvis: A visual analytics approach to production planning in smart factories. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):579–589, 2019. doi: 10.1109/TVCG.2019.2934275
[50] M. Suresh, K. Reddy, and A. K. Kolar. Ann-ga based optimization of a high ash coal-fired supercritical power plant. *Applied Energy*, 88(12):4867–4873, 2011. doi: 10.1016/j.apenergy.2011.06.029
[51] J. Tang, Y. Zhou, T. Tang, D. Weng, B. Xie, L. Yu, H. Zhang, and Y. Wu. A visualization approach for monitoring order processing in e-commerce warehouse. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):857–867, 2021. doi: 10.1109/TVCG.2021.3114878
[52] K. Vrotsou, J. Johansson, and M. Cooper. Activitree: Interactive visual exploration of sequences in event-based data using graph similarity. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):945–952, 2009. doi: 10.1109/TVCG.2009.117
[53] C. Walshaw. A multilevel algorithm for force-directed graph drawing. In *Proceedings of International Symposium on Graph Drawing*, pp. 171–182, 2000. doi: 10.1007/3-540-44541-2_17
[54] D. Wang, X. Wu, and J. Shen. An efficient robust predictive control of main steam temperature of coal-fired power plant. *Energies*, 13(15):3775, 2020. doi: 10.3390/en13153775
[55] H. Wang, Y. Ni, L. Sun, Y. Chen, T. Xu, X. Chen, W. Su, and Z. Zhou. Hierarchical visualization of geographical areal data with spatial attribute association. *Visual Informatics*, 5(3):82–91, 2021. doi: doi.org/10.1016/j.visinf.2021.09.001
[56] J. Wang, X. Cai, J. Su, Y. Liao, and Y. Wu. What makes a scatterplot hard to comprehend: data size and pattern salience matter. *Journal of Visualization*, 25(1):59–75, 2022. doi: 10.1007/s12650-021-00778-8
[57] D. Wei, C. Li, H. Shao, Z. Tan, Z. Lin, X. Dong, and X. Yuan. Sensoraware: visual analysis of both static and mobile sensor information. *Journal of Visualization*, 24(3):597–613, 2021. doi: 10.1007/s12650-020-00717-z
[58] What Is. The definition of fuzzy search. `https://whatis.techtarget.com/definition/fuzzy-search/`, 2005. [Online; accessed 30-Jun-2022].
[59] K. Wongsuphasawat and B. Shneiderman. Finding comparable temporal categorical records: A similarity measure with an interactive visualization. In *2009 IEEE Symposium on Visual Analytics Science and Technology*, pp. 27–34, 2009. doi: 10.1109/VAST.2009.5332595
[60] A. Wu, D. Deng, F. Cheng, Y. Wu, S. Liu, and H. Qu. In defence of visual analytics systems: Replies to critics. *IEEE Transactions on Visualization and Computer Graphics*, To Appear.
[61] W. Wu, Y. Zheng, K. Chen, X. Wang, and N. Cao. A visual analytics approach for equipment condition monitoring in smart factories of process industry. In *Proceedings of IEEE Pacific Visualization Symposium*, pp. 140–149, 2018. doi: 10.1109/PacificVis.2018.00026
[62] Y. Xiao, K. Zheng, S. Lonapalawong, W. Lu, Z. Chen, B. Qian, T. Zhang, X. Wang, and W. Chen. Ecovis: visual analysis of industrial-level spatio-temporal correlations in electricity consumption. *Frontiers of Computer Science*, 16(2):1–11, 2022. doi: 10.1007/s11704-020-0088-8
[63] P. Xu, H. Mei, L. Ren, and W. Chen. Vidx: Visual diagnostics of assembly line performance in smart factories. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):291–300, 2016. doi: 10.1109/TVCG.2016.2598664
[64] Y. Xu, C.-J. Yang, and X. Xuan. Engineering and optimization approaches to enhance the thermal efficiency of coal electricity generation in China. *Energy Policy*, 60:356–363, 2013. doi: 10.1016/j.enpol.2013.05.047
[65] J. Yin, M. Liu, Y. Zhao, C. Wang, and J. Yan. Dynamic performance and control strategy modification for coal-fired power unit under coal quality variation. *Energy*, 223:120077, 2021. doi: 10.1016/j.energy.2021.120077
[66] L. Ying, T. Tangl, Y. Luo, L. Shen, X. Xie, L. Yu, and Y. Wu. Glyphcreator: Towards example-based automatic generation of circular glyphs. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):400–410, 2022. doi: 10.1109/TVCG.2021.3114877
[67] X. Zhan, H. Xu, Y. Zhang, X. Zhu, H. Yin, and Y. Zheng. DeepThermal: Combustion optimization for thermal power generating units using offline reinforcement learning. In *Proceedings of AAAI Conference*, 2022. doi: 10.48550/arXiv.2102.11492
[68] X. Zhao, B. Peng, E. Elahi, C. Zheng, and A. Wan. Optimization of Chinese coal-fired power plants for cleaner production using bayesian network. *Journal of Cleaner Production*, 273:122837, 2020. doi: 10.1016/j.jclepro.2020.122837
[69] H. Zheng and G. Tang. Developing data acquisition and handling system for continuous emission monitoring system from coal-fired power plant. In *Proceedings of Chinese Control and Decision Conference*, pp. 3616–3619, 2008. doi: 10.1109/CCDC.2008.4598004
[70] F. Zhou, X. Lin, C. Liu, Y. Zhao, P. Xu, L. Ren, T. Xue, and L. Ren. A survey of visualization for smart manufacturing. *Journal of Visualization*, 22(2):419–435, 2019. doi: 10.1007/s12650-018-0530-2
[71] F. Zhou, X. Lin, X. Luo, Y. Zhao, Y. Chen, N. Chen, and W. Gui. Visually enhanced situation awareness for complex manufacturing facility monitoring in smart factories. *Journal of Visual Languages & Computing*, 44:58–69, 2018. doi: 10.1016/j.jvlc.2017.11.004