IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 7, 2019, accepted March 5, 2019, date of publication March 8, 2019, date of current version March 26, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2903830

Machine Learning Approach for Graphical Model-Based Analysis of Energy-Aware Growth Control in Plant Factories

YU FUJIMOTO[1], (Member, IEEE), SAYA MURAKAMI[2], NANAE KANEKO[3], HIDEKI FUCHIKAMI[4], TOSHIROU HATTORI[4], AND YASUHIRO HAYASHI[2], (Member, IEEE)

[1]Advanced Collaborative Research Organization for Smart Society, Waseda University, Tokyo 169-8555, Japan
[2]Department of Electrical Engineering and Bioscience, Waseda University, Tokyo 169-8555, Japan
[3]Department of Advanced Science and Engineering, Waseda University, Tokyo 169-8555, Japan
[4]Mayekawa Mfg. Co., Ltd., Ibaraki 302-0118, Japan

Corresponding author: Yu Fujimoto (y.fujimoto@aoni.waseda.jp)

ABSTRACT In recent decades, there has been a gradual penetration of plant factories achieving semiautomated crop cultivation. However, efficient energy utilization, as well as quality control of crops, are very important factors with regard to sustainable operation. Operating parameters, such as room temperature, affect not only the quality of crops but also the electric power required to realize the target operation while being influenced by the environment outside the plant. Therefore, a methodology is needed to analyze and interpret the relationships among these manipulated variables, exogenous variables, crop quality, and the amount of required electric power. Constructing a directed acyclic graph composed of regression models is an attractive approach for such analysis; however, the relationships can possibly be nonlinear, so the direct application of existing analytic approaches will not be appropriate. In this paper, we propose a methodology for relationship analysis among variables based on the directed acyclic graphs while identifying the linearity/nonlinearity in their relationships. In general, the construction of such a graphical model has computational issues, especially when the number of variables is large, and the risk of overfitting. The proposed method utilizes the idea of sparse regularization, which has been actively discussed in the field of machine learning, for realizing the automatic identification of linearity/nonlinearity between variables and screening redundant candidate structures; this approach relaxes the computational complexity issue and controls the risk of overfitting. As a case study, the proposed method is applied to a dataset collected from a real-world cultivation system in a plant factory to discuss its usefulness.

INDEX TERMS Analysis of plant data, directed graphical model, energy-aware plant growth control, identification of linearity/nonlinearity, overlap group lasso, plant factory, sparse partially linear model.

I. INTRODUCTION

In Japan, the production of plants in plant factories has attracted considerable attention in recent years due to a decrease in the number of agricultural workers and a decrease in the food self-sufficiency rate. The quality of the product largely depends on the growing environment (e.g. temperature, humidity, CO2 concentration, nutrient components, etc.); therefore, automatic big-data-driven plant growth control, which involves monitoring the plant growing environments, analyzing the acquired data, and using it for tuning the control parameters for maintaining production quality, has attracted significant attention [1]–[3]. However, despite the fact that data utilization is being carried out for such production control, nearly half of the domestic plant factories are forced to operate in deficit [4]. One of the reasons is that the electric energy cost due to the power consumption of the air conditioning system and the cultivation system for controlling the plant growing environments and realizing automatic production is relatively large. This fact suggests that balancing the quality of products and the energy cost required for growth is still a big problem in current plant factories.

Recently, there have been attempts to analyze and grasp the relationships among the control parameters of the growing environment, the quality of products, and the energy cost required for control [5], [6]. However, there has not been much discussion on the methodology to unifiedly discuss

The associate editor coordinating the review of this manuscript and approving it for publication was Ricardo De Castro.

and analyze the relationships among multiple control parameters, monitored environmental data, and energy consumption required for system control.

The construction of a directed graphical model based on data will provide a useful approach for organizing the relationships among various monitoring information and discussing the impact of control results under the exogenous factors. In particular, the construction of a directed graphical model based on traditional regression models [7], [8] is an attractive approach with an excellent interpretability for grasping and discussing the implicit relationships existing in the datasets collected in plant factories.

For example, the additive Bayesian Network [9], [10] is an analysis tool based on the directed graphical model that uses a class of generalized linear regression models; the model has been used for analyzing the relationships among the growth environment, quality of products, and the energy consumption required for the environmental control considering the effect of the weather conditions [11]. However, the class of the linear model may be inappropriate for describing the relationships existing in multivariate data handled at a plant factory; hence, a methodology is required for explicitly describing the nonlinear relationships. Therefore, a framework is required for constructing such an analytical model while actively identifying the linearity/nonlinearity of the relationship between variables.

In this paper, a strategy for constructing a directed acyclic graphical model based on partially linear additive models (PLAMs) is proposed; the proposed method selects a limited number of variables, directly explains each variable, and identifies the linear/nonlinear relationships among the variables based on the given data set. The proposed scheme consists of two key ideas that have recently developed in the machine learning community; i.e., regularization schemes for finding the linear/nonlinear relationships that are used to estimate sparse PLAMs [12], [13] and an enumeration scheme for finding the next-best alternative sparse regression models [14]. In our scheme, the search space for obtaining a plausible directed acyclic graphical representation is expected to be reduced effectively by introducing the sparseness assumption. The major contributions of this paper are as follows: (i) the concept of a directed graphical model that can represent the nonlinear additive relationship between variables is proposed, (ii) a method of selecting the linear/nonlinear relationships among variables is implemented based on a natural expansion of the regularized sparse regression scheme, (iii) a computationally efficient method for finding the plausible directed acyclic graphical structure corresponding to a set of PLAMs is discussed, (iv) the proposed scheme is applied to the dataset collected from a real-world plant factory.

The rest of the paper is organized as follows. In Section II, the current situation of the growth control in plant factories is briefly introduced. Traditional approaches for directed graphical model-based analysis are also described in this section. In Section III, the basic idea of a directed graphical model composed of partially linear additive models and its difficulties in implementation are described. Then, we propose a novel directed graphical model for the analysis of the partially linear additive statistical structure under a kind of sparseness assumption in Section IV; a computationally efficient and data-driven model estimation procedure is also described in this section. In Section V, the proposed approach is applied to a dataset collected from a real-world plant factory. Finally, our concluding remarks are provided in Section VI.

II. DATA-CENTRIC DECISION MAKING IN PLANT FACTORIES

A. GROWTH CONTROL IN PLANT FACTORIES

A plant factory is an idea that realizes sustainable and high-quality production of crops by applying systematic control for the cultivation process from an engineering viewpoint. In general, its main objective is the realization of optimal control of the environmental conditions by using the information processing results of the measured physical quantities of the plants. In particular, the measurement and information processing of plant growth states used in this scheme is called the *speaking plant approach* [15]. The control of the production environment based on information processing [1], [2], [16] utilizing the measurement of crop conditions [17], [18] has been discussed well from the viewpoint of automated crop production [3], [19] since the late 1970s. Various applications of artificial intelligence technology and machine learning approaches have been studied to realize data-driven operation and provide decision making support to the plant factories; e.g., [20] proposed a classification algorithm for the detection of fungal diseases in crops, and [21] introduced the essence of fuzzy theory into the operation scheduling of crop production (see [22] for a comprehensive review of machine learning approaches applied to the agricultural domain).

Meanwhile, as the plant factories began to be put into practical use, the importance of grasping the trade-off between energy consumption and control results began to be recognized from the viewpoints of improving profitability and reducing environmental impact [23]; therefore, energy-aware growth control has become a crucial topic [5], [6], [11], [24]. Fig. 1 shows the schematic relationship between the variables used in a plant factory. The analysis of real-world operation data plays a key role in grasping such a trade-off. This study focuses on an approach for analyzing the relationships among controllable variables and target variables, e.g., crop quality and power consumption, under the given exogenous variables, for decision making in the energy-aware operation in plant factories.

B. GRAPHICAL STRUCTURE USING REGRESSION MODELS FOR OPERATION ANALYSIS

Graphical models [25] are useful tools for describing and interpreting the complex statistical relationships existing in real-world datasets. In particular, regression-based graphical representation of statistical models for the description of dependencies among variables, e.g. traditional path analy-

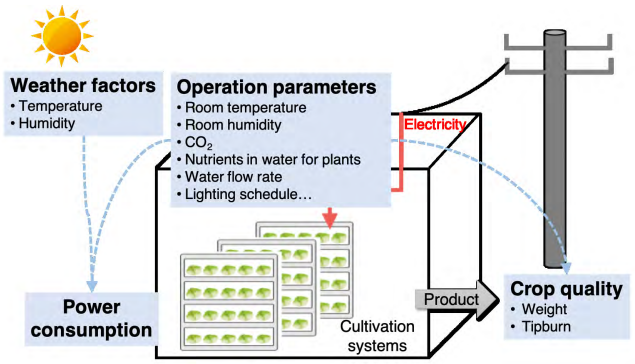


FIGURE 1. Schematic relationship among exogenous variables (weather factors), controllable variables (operation parameters), and target variables (power consumption and crop quality).

sis [7] and structural equation model [8], have been frequently used in the context of data analysis. Such graphical representations have been originally developed for describing linear relationships, and occasionally, they have been applied to real-world datasets without sufficient verification of the validity of the linearity assumption among variables, although real-world data might not satisfy this assumption. For this reason, several graphical modeling approaches based on expanded regression methods have been discussed for grasping the nonlinear variable relationships. For example, the structural equation model has mainly evolved to describe the nonlinear relationships between variables by handling the latent variables in the model [26]. Some works have also proposed the use of interaction or quadratic effects for the representation of nonlinear relationships [27]–[31]. Recently, another regression-based graphical modeling scheme based on generalized linear models [32], [33] was proposed to describe a relatively flexible statistical structure [10], [11], [34]. However, frameworks that can deal with highly nonlinear relationships have hardly been discussed. In particular, the utilization of nonlinear regression makes it difficult to identify the essentially linear relationships among variables; this will complicate the data-centric decision making unnecessarily.

In this study, we focus on an attractive expanded regression approach based on a partially linear additive model (PLAM) [13], [35]–[37]; this model belongs to a class of additive models [38], which is one of the practical *white box models*, and is superior in terms of the representability of nonlinear effects and interpretability of the estimation result by focusing on the additive contribution of each variable. A practical difficulty in conducting analyses based on a PLAM is the explicit identification of linear/nonlinear relationships between variables; however, the idea of regularization for obtaining a sparse representation [39], which has been intensively discussed in the field of machine learning in recent years, has incubated fascinating variable selection [40], [41] and linearity/nonlinearity selection [12], [13], [42] schemes. In the next section, we discuss the construction approach of the graphical model based on PLAMs.

III. BASIC IDEA OF GRAPHICAL MODEL BASED ON PARTIALLY LINEAR ADDITIVE MODELS

A. GRAPHICAL MODEL BASED ON PARTIALLY LINEAR ADDITIVE MODELS

Let $\boldsymbol{v} = (v^1, \ldots, v^P) \in \mathbb{R}^P$ be the variables[1] and $\mathcal{S} = \{1, \ldots, P\}$ be the index set of these variables. In addition, let $\boldsymbol{v}^{-p} = \{v^i; i \in \mathcal{S}^p\}$ be the candidate set of explanatory variables for v^p, where $\mathcal{S}^p \subseteq (\mathcal{S} \setminus \{p\})$ is the candidate index set of explanatory variables for the representation of v^p. We discuss a class of graphical model based on the following additive models, which represent linear and nonlinear relationships flexibly.

Definition 1 (PLAM; Partially Linear Additive Model): The following regression formulation,

$$\begin{aligned} v^p &\simeq f^p(\boldsymbol{v}^{-p}) \\ &= \beta_p^0 + \sum_{i \in \mathcal{L}^p} \beta_p^i v^i + \sum_{i \in \mathcal{N}^p} \phi_p^i(v^i), \end{aligned} \tag{1}$$

is called a partially linear additive model (PLAM) [35], [36], where β_p^i is the coefficient parameter, $\phi_p^i(\cdot)$ is a nonlinear function, and $\mathcal{L}^p, \mathcal{N}^p \subseteq \mathcal{S}^p$ are variable index subsets containing linear/nonlinear relationships with the variable v^p, respectively; assume that $(\forall p \in \mathcal{S})\ \mathcal{L}^p \cap \mathcal{N}^p = \emptyset$ holds.

In this study, we introduce the following set of cubic spline bases with $(K-1)$ knots, i.e.

$$\begin{aligned} \boldsymbol{\psi}_p^i(v^i) &= \left[\psi_{p,k}^i(v^i); k = 1, \ldots, K\right] \\ &= \left[v^i, (v^i - v^i_{(1)})_+^3, \ldots, (v^i - v^i_{(K-1)})_+^3\right], \end{aligned} \tag{2}$$

for expressing the nonlinear function ϕ_p^i where $v^i_{(1)} \ldots, v^i_{(K-1)}$ are knots of the spline chosen from quantiles in the sample set and

$$(z)_+ = \begin{cases} 0 & (z < 0) \\ z & (z \geq 0). \end{cases} \tag{3}$$

Therefore, Eq. (1) is alternatively given as follows:

$$\begin{aligned} &f^p(\boldsymbol{v}^{-p}; \boldsymbol{\theta}^p, \mathcal{L}^p, \mathcal{N}^p) \\ &\quad = \beta_p^0 + \sum_{i \in \mathcal{L}^p} \beta_p^i v^i + \sum_{i \in \mathcal{N}^p} \underbrace{\sum_k \tau_{p,k}^i \psi_{p,k}^i(v^i)}_{\phi_p^i(v^i)}, \end{aligned} \tag{4}$$

where $\tau_{p,k}^i$ is the coefficient for the basis $\psi_{p,k}^i$ and $\boldsymbol{\theta}^p = \{\beta_p^0, \{\beta_p^i\}, \boldsymbol{\tau}_p^i = [\tau_{p,k}^i]\}$ is the whole parameter set of the model. If $\mathcal{L}^p$ and $\mathcal{N}^p$ are given, one can estimate the parameter $\boldsymbol{\theta}^p$ of PLAM, whose form is given by Eq. (4), as follows:

$$\hat{\boldsymbol{\theta}}^p = argmin_{\boldsymbol{\theta}^p} G^p(\boldsymbol{\theta}^p; \mathcal{L}^p, \mathcal{N}^p) \tag{5}$$

$$= \arg\min_{\boldsymbol{\theta}^p} \sum_n \{v_n^p - f^p(\boldsymbol{v}_n^{-p}; \boldsymbol{\theta}^p, \mathcal{L}^p, \mathcal{N}^p)\}^2. \tag{6}$$

[1]Throughout this paper, we assume that v^p is standardized to have zero mean and unit variance.

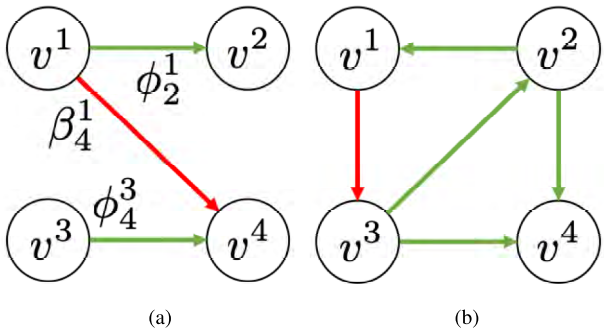


FIGURE 2. Examples of partially linear additive relationships represented by a directed graph. The red arrows indicate the linear relationships and the green arrows indicate the nonlinear relationships. (a): Example of PLAS, i.e., the partially linear additive relationships represented by the DAG. The graph implies that $v^1 \simeq \beta_1^0$; $v^2 \simeq \beta_2^0 + \phi_2^1(v^1)$; $v^3 \simeq \beta_3^0$; and $v^4 \simeq \beta_4^0 + \beta_4^1 x^1 + \phi_4^3(v^3)$. The SCCs are $\{\mathcal{S}_c^m(A)\} = \{\{1\}, \{2\}, \{3\}, \{4\}\}$. (b): Example of a directed graph containing a cyclic relationship. The SCCs are $\{\mathcal{S}_c^m(A)\} = \{\{1, 2, 3\}, \{4\}\}$.

Our objective is to find this type of partially linear additive relationship among variables based on the given data set $\{\boldsymbol{v}_n = (v_n^1, \ldots, v_n^P)\}$ in order to construct a directed graphical model for the representation of variable dependencies; in particular, we assume that the derived structure is a directed acyclic graph (DAG) as well as the ordinary Bayesian network [43]. Fig. 2 shows examples of the graphical representation of the statistical relationship using PLAMs. Here, variables $\boldsymbol{v}$ are represented as P nodes, and the partially linear additive relationships among these nodes are represented by two types of edges representing the linear/nonlinear relationships directed from $\{v^i; i \in \mathcal{L}_p\}$ and $\{v^j; j \in \mathcal{N}_p\}$ to x^p, respectively. In this study, we call the graphical representation associated with the DAG shown in Fig. 2(a) the partially linear additive structure (PLAS), and aim to derive this representation based on the given dataset.

Let $\mathcal{A}^p = \mathcal{L}^p \cup \mathcal{N}^p$ be the explanatory variable index set of the variable v^p. Assume that the set $\mathcal{A}^p$ indicates the edges directed from v_i $(\forall i \in \mathcal{A}^p)$ to v_p in the graphical representation. We obtain what we shall call the *adjacency matrix* $A = [a_{pi}] \in \{0, 1\}^{P \times P}$, where its component a_{pi} is defined as follows,

$$a_{pi} = \begin{cases} 1 & (i \in \mathcal{A}^p) \\ 0 & (i \notin \mathcal{A}^p). \end{cases} \tag{7}$$

The directed graph representation corresponding to this adjacency matrix A is discussed in this paper.

Now, we introduce the following idea for this graph representation.

Definition 2 (SCC; Strongly Connected Component:) A strongly connected component (SCC) of a directed graph is a maximal node subset $\mathcal{S}_c \subset \mathcal{S}$ such that its component nodes are mutually reachable.

Note that the SCC represents the node set of a maximal cyclic subgraph, and if there is no node $i \in \mathcal{S}^p$ such that it is mutually reachable from p, the corresponding SCC is given as $\mathcal{S}_c = \{p\}$. We denote the set of SCCs in the directed graph corresponding to the adjacency matrix A by $\{\mathcal{S}_c^m(A); m = 1, \ldots, M\}$, where M is the number of SCCs.

By using the idea of SCCs, we focus on the following property.

Theorem 1 (SCCs of DAG): The graphical representation of the set of PLAMs is a DAG iff the cardinality of all the SCCs corresponding to A in the directed graph is equal to 1.

Proof: The definition $\mathcal{A}^p = \mathcal{L}^p \cup \mathcal{N}^p$ suggests that the adjacency matrix A represents the corresponding set of PLAMs. If $|\mathcal{S}_c^m(A)| = 1$ $(\forall m)$ holds, then the graph has no cyclic paths (e.g. see [44]). □

Now, our task, i.e. finding the DAG representation based on PLAMs for the derivation of the PLAS, is formulated by considering the property given in Theorem 1 as follows:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} G(\boldsymbol{\theta}) \quad \text{s.t. } |\mathcal{S}_c^m(A)| = 1 \ (\forall m), \tag{8}$$

where $G(\boldsymbol{\theta})$ is the total objective function evaluating the whole structure, i.e.

$$G(\boldsymbol{\theta}) = \sum_{p \in \mathcal{S}} G^p(\boldsymbol{\theta}^p), \tag{9}$$

where G^p is defined in Eq. (5).

B. NAIVE APPROACH AND ITS ISSUES

Lewis and McCormick proposed a method so-called the additive Bayesian network [34] to derive directed acyclic statistical relationships based on the generalized linear regression model. In their framework, the following two-step strategy was implemented: 1) enumerate generalized linear models for possible combination of explanatory variables for all the variables, and 2) search the best model set that constructs a DAG. A naive approach for realization of the framework discussed in this paper can be constructed by extending this scheme as follows:

- **Enumerate** PLAMs for possible combination of explanatory variables for all the variables.
- **Search** the best model set that constructs a DAG.

In the enumeration step, we assume that the possible combination $\mathcal{R}^p = \{(\mathcal{L}^p, \mathcal{N}^p)\}$ is given in advance for all the variables; if there is no prior knowledge about the validity of the explanatory variables,[2] one may adopt

$$\mathcal{R}^p = \{(\mathcal{L}^p, \mathcal{N}^p) \mid \mathcal{L}^p, \mathcal{N}^p \subseteq \mathcal{S}^p, \mathcal{L}^p \cap \mathcal{N}^p = \emptyset\}. \tag{10}$$

Then, the PLAM is estimated for each $(\mathcal{L}^p, \mathcal{N}^p) \in \mathcal{R}^p$ by optimizing the problem given in Eq. (6). Algorithm 1 shows a naive implementation of this enumeration step; this algorithm returns a set of estimation results and their corresponding objective values, $\mathcal{Q}^p$, for all the possible combinations

[2]Note that the idea of heuristic constraints, so-called *ban list* [34], for the reduction of DAG search space, which explicitly describes the undesirable edge sets in the graph structure in advance, can be implemented by defining the candidate index set, $\mathcal{S}^p$, with a limited number of explanatory variables for each x^p.

Algorithm 1 Naive Enumeration of Possible PLAMs for Variable v^p

Require: dataset $\{\boldsymbol{v}_n\}$, possible sets of linear/nonlinear relationships $\mathcal{R}^p = \{(\mathcal{L}^p, \mathcal{N}^p)\}$.
1: **for** $(\mathcal{L}^p, \mathcal{N}^p) \in \mathcal{R}^p$ **do**
2: Obtain $\hat{\boldsymbol{\theta}}^p$ according to Eq. (6) under $\mathcal{L}^p$ and $\mathcal{N}^p$.
3: Store a pair $\mathcal{Q}^p(\mathcal{L}^p, \mathcal{N}^p) \leftarrow (\hat{\boldsymbol{\theta}}^p, G^p(\hat{\boldsymbol{\theta}}^p))$.
4: **end for**
Ensure: $\mathcal{Q}^p = \{\mathcal{Q}^p(\mathcal{L}^p, \mathcal{N}^p); (\mathcal{L}^p, \mathcal{N}^p) \in \mathcal{R}^p\}$.

Algorithm 2 Hill Climbing Approach for Finding PLAS

Require: $\mathcal{Q} = \{\mathcal{Q}^p; p \in \mathcal{S}\}, \{\mathcal{R}^p\}$.
1: **function** SelectTuple($\mathcal{L}^p, \mathcal{N}^p$) ▷ Select the corresponding PLAM for given relation.
2: $(\hat{\boldsymbol{\theta}}^p, G^p) \leftarrow \mathcal{Q}^p(\mathcal{L}^p, \mathcal{N}^p)$.
3: **return** $(\mathcal{L}^p \cup \mathcal{N}^p, \hat{\boldsymbol{\theta}}^p, G^p)$.
4: **end function**

5: **for** $p \in \mathcal{S}$ **do** ▷ Find initial graph structure.
6: Randomly select $(\bar{\mathcal{L}}^p, \bar{\mathcal{N}}^p) \in \mathcal{R}^p$.
7: $(\bar{\mathcal{A}}^p, \bar{\boldsymbol{\theta}}^p, \bar{G}^p) \leftarrow$ SelectTuple$(\bar{\mathcal{L}}^p, \bar{\mathcal{N}}^p)$.
8: **end for**
9: **while** $|\mathcal{S}_c^m(\bar{A})| \neq 1 (\exists m)$ **do** ▷ Find initial DAG.
10: Randomly select $p \in \mathcal{S}$.
11: $(\check{\mathcal{L}}^p, \check{\mathcal{N}}^p) \leftarrow (\bar{\mathcal{L}}^p, \bar{\mathcal{N}}^p)$, $\check{\mathcal{A}}^p \leftarrow \bar{\mathcal{A}}^p$.
12: Randomly select $(\bar{\mathcal{L}}^p, \bar{\mathcal{N}}^p) \in \{(\mathcal{L}^p, \mathcal{N}^p) \in \mathcal{R}^p \mid (\mathcal{L}^p \cup \mathcal{N}^p) \subset \check{\mathcal{A}}^p\}$
13: $(\bar{\mathcal{A}}^p, \bar{\boldsymbol{\theta}}^p, \bar{G}^p) \leftarrow$ SelectTuple$(\bar{\mathcal{L}}^p, \bar{\mathcal{N}}^p)$
14: **end while**
15: $\underline{G} \leftarrow \infty$
16: **while** $\sum_p \bar{G}^p < \underline{G}$ **do**
17: $\{(\check{\mathcal{L}}^p, \check{\mathcal{N}}^p)\} \leftarrow \{(\bar{\mathcal{L}}^p, \bar{\mathcal{N}}^p)\}$, $\{\check{\mathcal{A}}^p\} \leftarrow \{\bar{\mathcal{A}}^p\}$. ▷ Back up the current best.
18: **for** $p \in \mathcal{S}$ **do**
19: **for** $(\check{\mathcal{L}}^p, \check{\mathcal{N}}^p) \in \{(\mathcal{L}^p, \mathcal{N}^p) \in \mathcal{R}^p \mid |\bar{\mathcal{A}}^p \cup (\mathcal{L}^p \cup \mathcal{N}^p)| - |\bar{\mathcal{A}}^p \cap (\mathcal{L}^p \cup \mathcal{N}^p)| = 1\}$ **do** ▷ Remove/add an edge.
20: $(\check{\mathcal{A}}^p, \check{\boldsymbol{\theta}}^p, \check{G}^p) \leftarrow$SelectTuple$(\check{\mathcal{L}}^p, \check{\mathcal{N}}^p)$
21: **if** $|\mathcal{S}_c^m(\check{A})| = 1 (\forall m)$ and $\sum_p \check{G}^p < \underline{G}$ **then**
22: $(\dot{\mathcal{L}}^p, \dot{\mathcal{N}}^p) \leftarrow (\check{\mathcal{L}}^p, \check{\mathcal{N}}^p)$, $\underline{G} \leftarrow \sum_p \check{G}^p$. ▷ Store the current best DAG.
23: **end if**
24: **end for**
25: **end for**
26: $(\bar{\mathcal{L}}^p, \bar{\mathcal{N}}^p) \leftarrow (\dot{\mathcal{L}}^p, \dot{\mathcal{N}}^p)$.
27: **end while**
Ensure: $\{(\bar{\mathcal{L}}^p, \bar{\mathcal{N}}^p); p \in \mathcal{S}\}$

$(\mathcal{L}^p, \mathcal{N}^p) \in \mathcal{R}^p$. Note that the enumeration procedure given by Algorithm 1 can be applied for all $p \in \mathcal{S}$ in parallel.

In the DAG structure search step, we can naively adopt the idea of the hill climbing method [45], which has also been adopted in the structure search step of the additive Bayesian network. This approach finds a local minimum of the total objective function by iteratively replacing the local PLAMs in a greedy manner; the edges, which provides the greatest decrease in the objective function, are sequentially added/removed while keeping the DAG constraints. Algorithm 2 shows a procedure for finding the PLAS represented by the DAG. At lines 5-14, an initial DAG structure is randomly assigned using the enumerated candidates $\{\mathcal{Q}^p\}$. At lines 18-25, the next best DAG structure is greedily adopted by focusing on each substructure $\check{\mathcal{A}}^p$. Note that Algorithm 2 derives one of the local minima; therefore, the best result from multiple trials of the various initial states is adopted from the viewpoint of the total objective function. Multiple trials of Algorithm 2 for selecting the best result can also proceed in parallel. It should also be stressed that the SCCs for checking the DAG constraint in this procedure can be found in the linear computational time by applying the depth-first search algorithm [46], [47].

The above-mentioned naive approach enables the construction of a directed graphical model based on a set of PLAMs; however, the following two points become issues: 1) the size

of the search space in the DAG search step, and 2) structural overfit for the given data set.

The computational time required for the DAG search step in the naive approach basically depends on the number of combinations of candidate PLAMs to be evaluated. If no prior knowledge is applied to create a set of candidates $\mathcal{R}^p$, the number of possible candidates derived with Eq. (10) becomes $|\mathcal{R}^p| = 3^{P-1}$ for each v^p. Therefore, in this case, the structure search of the graphical model based on PLAMs inherently requires searching a solution in the huge space of $\prod_p |\mathcal{R}^p| = 3^{P(P-1)}$. Since the naive approach with a very large search space intuitively produces a large number of local solutions, the problem of obtaining a relatively good solution becomes very difficult.

The optimization problem shown in Eq. (6) corresponds to the maximum likelihood estimation of the PLAM parameters used for representing the target v^p under the assumption of Gaussian error distribution. Therefore, the minimization problem Eq. (8), which focuses on the structure constructed by such PLAMs, is also derived as a kind of maximum likelihood estimation under the constraint of being a DAG. Note that in general, such maximum likelihood estimation tends to derive a model structure overfitted to the learning data. For example, the idea of majority consensus network [48] has been proposed for alleviating the overfitting problem in the random-restart hill-climbing algorithm [34]. This approach adopts a network composed of the directed edges existing in a majority of networks derived from the various initial DAGs by the hill climbing approach so as to reduce the overfitting. However, such a majority consensus network does not generally guarantee to be a DAG and significantly depends on the random initial conditions. Therefore, an efficient approach is required to control the structural overfitness to the given data while satisfying the DAG constraints.

IV. GRAPHICAL MODEL BASED ON PARTIALLY LINEAR ADDITIVE MODELS UNDER SPARSENESS ASSUMPTION

The naive approach introduced in the previous section is a natural extension of the existing implementation; however, this approach has issues of computational cost caused by the huge search space and overfitness caused by the design of the objective function. In this section, we discuss a novel scheme for finding the PLAS for construction of a directed graphical model by introducing the sparseness assumption in the explanatory variables.

Consider that the following form of additive model $f^p(\cdot)$ is constructed for the explanation of v^p using the variables indicated by $\mathcal{S}^p$,

$$\begin{aligned} v^p &\simeq f^p(\boldsymbol{v}^{-p}; \boldsymbol{\theta}^p, \mathcal{S}^p) \\ &= \beta_p^0 + \sum_{i \in \mathcal{S}^p} \underbrace{\left\{ (\xi_p^i + \tau_{p,1}^i) v^i + \sum_{k=2}^{K} \tau_{p,k}^i \psi_{p,k}^i(v^i) \right\}}_{\phi_p^i(v^i)}, \end{aligned} \tag{11}$$

where $\boldsymbol{\theta}^p = \{\beta_p^0, \xi_p^i, \boldsymbol{\tau}_p^i = [\tau_{p,k}^i]\}$ is a parameter set, ξ_p^i and $\tau_{p,1}^i$ are coefficients for representing the linear trend on v^i, $\psi_{p,k}^i(\cdot)$ is a basis function, and $\tau_{p,k}^i$ is the coefficient of the basis $\psi_{p,k}^i$ for representing the nonlinearity. Note that the reparameterized term of the additive model shown in Eq. (11) reduces to a linear term with $\beta_p^i = \xi_p^i$ when $(\forall j)\ \tau_{p,j}^i = 0$ holds. We also note that when $\tau_{p,k}^i \neq 0$ holds for any $k \geq 2$, the corresponding $\phi_p^i(v^i)$ expresses the nonlinear term.

To evaluate a model given by Eq. (11) using data set $\{\boldsymbol{v}_n\}$, we focus on the minimization problem of the following objective function:

$$G_\lambda^p(\boldsymbol{\theta}^p; \mathcal{S}^p) = \sum_n \left\{ v_n^p - f^p(\boldsymbol{v}_n^{-p}; \boldsymbol{\theta}^p, \mathcal{S}^p) \right\}^2 + \lambda \Omega(\boldsymbol{\theta}^p), \tag{12}$$

where λ is a positive regularization parameter and $\Omega(\boldsymbol{\theta}^p)$ is a regularization term. We particularly focus on the learning task for finding appropriate linear/nonlinear relationships explaining a specific variable v^p while searching the sparse variable representation by using the data set $\{\boldsymbol{v}_n\}$; therefore, we apply the following type of regularization:

$$\Omega(\boldsymbol{\theta}^p; \mathcal{S}^p) = \sum_{i \in \mathcal{S}^p} \left(|\xi_p^i| + \|\boldsymbol{\tau}_p^i\|_2 \right). \tag{13}$$

Note that the regularizer, Eq. (13), is motivated by the technique so-called *overlap group lasso* [49], and it tends to reduce the coefficient ξ_p^i and/or the coefficient vector $\boldsymbol{\tau}_p^i$ to become zeros. Therefore, the minimizer of Eq. (12) with the regularizer given by Eq. (13) enables the selection of one of the following states:

- v^i is linearly related to v^p when $\|\boldsymbol{\tau}_p^i\|_2 = 0$ and $\xi_p^i \neq 0$,
- v^i is nonlinearly related to v^p when $\tau_{p,j}^i \neq 0$ for any $k \geq 2$,
- v^i is unrequired for explanation of v^p when $|\xi_p^i| = 0$ and $\|\boldsymbol{\tau}_p^i\|_2 = 0$.

This property derives a sparse PLAM for the target variable v^p. The parameters $\boldsymbol{\theta}^p$ in the additive model can be estimated by using the group decent algorithm [50].

Here, we focus on the construction of a DAG structure containing two types of directed edges for representing the linear/nonlinear relationships of a set of PLAMs. Assume that the parameter $\boldsymbol{\theta}^p$ in the model is obtained for each variable v^p; our requirement for a set of models $\{f^p(\cdot)\}$ is that the corresponding directed graphical representation is acyclic. Let $\boldsymbol{\theta} = \{\boldsymbol{\theta}^p; p = 1, \ldots, P\}$ be the parameter set of PLAMs for P variables and

$$\mathcal{A}_{\boldsymbol{\theta}}^p = \{ i \mid |\xi_p^i| \neq 0 \text{ or } \|\boldsymbol{\tau}_p^i\|_2 \neq 0 \} \tag{14}$$

be the variable index set derived from the given parameter $\boldsymbol{\theta}^p$. We construct the adjacency matrix $A_{\boldsymbol{\theta}}$ of the graphical representation by using the estimation results $\{\mathcal{A}_{\boldsymbol{\theta}}^p\}$ in accordance with Eq. (7). By introducing the regularizer shown in Eq. (12), the whole optimization problem, Eq. (8), is slightly modified as follows:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} G_\lambda(\boldsymbol{\theta}; \{\mathcal{S}^p\}) \quad \text{s.t. } |\mathcal{S}_c^m(A_{\boldsymbol{\theta}})| = 1\ (\forall m), \tag{15}$$

Algorithm 3 Enumerate PLAMs Under Overlap Group Lasso Constraint for Variable v^p

Require: $\{\boldsymbol{v}_n\}, \mathcal{S}^p, \lambda$.
Set target $\mathcal{C} \leftarrow \mathcal{S}^p$ and discovered links $\mathcal{D} \leftarrow \emptyset$.
$\hat{\boldsymbol{\theta}}^p \leftarrow \arg\min_{\boldsymbol{\theta}^p} G^p_\lambda(\boldsymbol{\theta}^p; \mathcal{C})$. ▷ Estimate sparse PLAM under possible regressors.
Initialize min-heap $\mathcal{T}^p \leftarrow \{(\hat{\boldsymbol{\theta}}^p, \mathcal{C}, \mathcal{D})\}$ with key $G^p_\lambda(\hat{\boldsymbol{\theta}}^p; \mathcal{C})$.
for $i = 1, 2, \ldots$ **do**
  Extract $(\hat{\boldsymbol{\theta}}^p, \mathcal{C}, \mathcal{D})$ from heap $\mathcal{T}^p$. ▷ The current best PLAM is popped from the heap.
  Store a pair $(\hat{\boldsymbol{\theta}}^{p[i]}, G^{p[i]}_\lambda) \leftarrow (\hat{\boldsymbol{\theta}}^p, G^p_\lambda(\hat{\boldsymbol{\theta}}^p; \mathcal{C}))$. ▷ The PLAM is stored as the i-th best.
  for $d \in \mathcal{A}^p_{\hat{\boldsymbol{\theta}}^{p[i]}}$ and $d \notin \mathcal{D}$ **do** ▷ Select a regressor with nonzero coefficients not involved in $\mathcal{D}$.
    $\hat{\boldsymbol{\theta}}^p \leftarrow \arg\min_{\boldsymbol{\theta}^p} G^p_\lambda(\boldsymbol{\theta}^p; \mathcal{C} \setminus \{d\})$.
    Insert $(\hat{\boldsymbol{\theta}}^p, \mathcal{C} \setminus \{d\}, \mathcal{D})$ to $\mathcal{T}^p$ with key $G^p_\lambda(\hat{\boldsymbol{\theta}}^p; \mathcal{C} \setminus \{d\})$.
    $\mathcal{D} \leftarrow \mathcal{D} \cap \{d\}$.
  end for
end for
Ensure: $\mathcal{Q}^p_\lambda = \{\mathcal{Q}^{p[i]} = (\hat{\boldsymbol{\theta}}^{p[i]}, G^{p[i]}_\lambda)\}$.

Algorithm 4 Hill Climbing Approach for Finding PLAS Based on Sparse PLAMs

Require: $\{\{(\boldsymbol{\theta}^{p[i]}, G^{p[i]}_\lambda)\} = \mathcal{Q}^p_\lambda; p \in \mathcal{S}\}$.
Set $\bar{\mathcal{A}}^p \leftarrow \cup_i \mathcal{A}^p_{\hat{\boldsymbol{\theta}}^{p[i]}}$ $(\forall p)$ and construct adjacency matrix $\bar{A}$ with $\{\bar{\mathcal{A}}^p\}$. ▷ Consider all possible links in candidates.
Set $\bar{\boldsymbol{\theta}}^p \leftarrow \hat{\boldsymbol{\theta}}^{p[1]}$ $(\forall p)$ and $\bar{\mathcal{S}} \leftarrow \emptyset$.
for $m = 1, \ldots, |\{\mathcal{S}^m_c(\bar{A})\}|$ **do** ▷ Identify possible nodes constructing cyclic subgraphs.
  if $\mathcal{S}^m_c(\bar{A}) > 1$ **then**
    $\bar{\mathcal{S}} \leftarrow \bar{\mathcal{S}} \cup \mathcal{S}^m_c(\bar{A})$.
    Randomly select $(\bar{\boldsymbol{\theta}}^p, \bar{G}^p_\lambda) \in \mathcal{Q}^p_\lambda$ $(\forall p \in \mathcal{S}^m_c(A_{\hat{\boldsymbol{\theta}}}))$.
  end if
end for
while $|\mathcal{S}^m_c(A_{\bar{\boldsymbol{\theta}}})| > 1 (\exists m)$ **do** ▷ Find initial DAG.
  Randomly select $p \in \bar{\mathcal{S}}$.
  Randomly select $(\bar{\boldsymbol{\theta}}^p, \bar{G}^p_\lambda) \in \mathcal{Q}^p_\lambda$.
end while
$\underline{G} \leftarrow \infty$
while $\sum_p \bar{G}^p_\lambda < \underline{G}$ **do**
  $\breve{G}^p_\lambda \leftarrow \bar{G}^p_\lambda$ $(\forall p)$
  for $p \in \bar{\mathcal{S}}$ **do**
    for $(\breve{\boldsymbol{\theta}}^p, \breve{G}^p_\lambda) \in \{\mathcal{Q}^{p[i]}_\lambda \mid \left||\mathcal{A}^p_{\theta^{p[i]}}| - |\mathcal{A}^p_{\bar{\boldsymbol{\theta}}}|\right| = 1, \mathcal{Q}^{p[i]}_\lambda \in \mathcal{Q}^p_\lambda\}$ **do** ▷ Remove/add an edge.
      if $|\mathcal{S}^m_c(A_{\breve{\boldsymbol{\theta}}})| = 1 (\forall m)$ and $\sum_p \breve{G}^p_\lambda < \underline{G}$ **then**
        $\dot{\boldsymbol{\theta}}^p \leftarrow \breve{\boldsymbol{\theta}}^p$, and $\underline{G} \leftarrow \sum_p \breve{G}^p_\lambda$. ▷ Store the current best DAG.
      end if
    end for
  end for
  $\bar{\boldsymbol{\theta}} \leftarrow \dot{\boldsymbol{\theta}}$.
end while
Ensure: $\{\bar{\boldsymbol{\theta}}^p; p \in \mathcal{S}\}$

where

$$G_\lambda(\boldsymbol{\theta}; \{\mathcal{S}^p\}) = \sum_{p \in \mathcal{S}} G^p_\lambda(\boldsymbol{\theta}^p; \mathcal{S}^p). \tag{16}$$

Now, to obtain the optimizer of Eq. (15), we adopt the same enumerate and search strategy as the naive approach, but in a relatively efficient way.

In the enumerate step, we utilize the idea inspired from the enumerate lasso [14]; this algorithm efficiently enumerates good regression models with various subsets of regressors in ascending order with the lasso regularization based on their objective values. In our context, we focus on the following proposition.

Proposition 1 (Different Support Solutions to Overlap Group Lasso): Given $\mathcal{S}^p$ and $\bar{\mathcal{S}}^p \subseteq \mathcal{S}^p$, the objective function of the overlap group lasso model, Eq. (12)*, has the following property,*

$$\min_{\boldsymbol{\theta}^p} G^p_\lambda(\boldsymbol{\theta}^p; \mathcal{A}^p_{\hat{\boldsymbol{\theta}}^p}) = \min_{\boldsymbol{\theta}^p} G^p_\lambda(\boldsymbol{\theta}^p; \bar{\mathcal{S}}^p) = \min_{\boldsymbol{\theta}^p} G^p_\lambda(\boldsymbol{\theta}^p; \mathcal{S}^p), \tag{17}$$

if $\bar{\mathcal{S}}^p \supseteq \mathcal{A}^p_{\hat{\boldsymbol{\theta}}^p}$ holds, where

$$\hat{\boldsymbol{\theta}}^p = \mathrm{argmin}_{\boldsymbol{\theta}^p} G^p_\lambda(\boldsymbol{\theta}^p; \mathcal{S}^p). \tag{18}$$

Proof: The explanatory variables set $\mathcal{A}^p_{\hat{\boldsymbol{\theta}}^p}$ defined in Eq. (14) suggests that $\hat{\xi}^i_p = 0$ and $\|\boldsymbol{\tau}^i_p\|_2 = 0$ for $i \in (\mathcal{S}^p \setminus \mathcal{A}^p_{\hat{\boldsymbol{\theta}}^p})$. Therefore, the explanatory variables subset $\mathcal{A}^p_{\hat{\boldsymbol{\theta}}^p}$ used in the first formulation of Eq. (17) does not violate the optimality of $\hat{\boldsymbol{\theta}}^p$. Similarly, we can see that the second formulation of Eq. (17) also does not violate the optimality of $\hat{\boldsymbol{\theta}}^p$ under the condition that $\mathcal{A}^p_{\hat{\boldsymbol{\theta}}^p} \subseteq \bar{\mathcal{S}}^p \subseteq \mathcal{S}^p$ holds. □

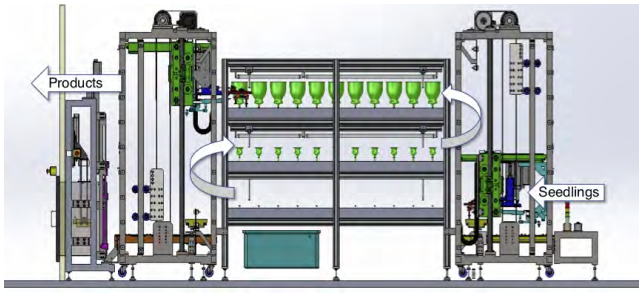
Proposition 1 suggests that if $\bar{\mathcal{S}}^p$ satisfies $\mathcal{A}^p_{\hat{\boldsymbol{\theta}}^p} \subseteq \bar{\mathcal{S}}^p \subseteq \mathcal{S}^p$, $\hat{\boldsymbol{\theta}}^p$ is also a minimizer of $G^p_\lambda(\boldsymbol{\theta}^p; \bar{\mathcal{S}}^p)$. This property significantly reduces the computational cost for the parameter estimation by avoiding redundant enumeration under the given λ. Algorithm 3 shows the procedure for enumerating possible PLAMs under the given $\mathcal{S}^p$ and λ. Note that $\mathcal{T}^p$ at line 3 indicates the heap of triplets composed of the estimated parameter, candidate subset of explanatory variables, and the objective value, which are sorted in ascending order of objective value G^p_λ; the condition at line 7 is given to avoid redundant search. Algorithm 3 returns a sorted list of a limited number of good models for v^p from the viewpoint of the individual objective function G^p_λ. The enumeration procedure shown in Algorithm 3 can also be applied for all $p \in \mathcal{S}$ in parallel.

Algorithm 4 describes a hill climbing approach for selecting a plausible DAG based on a set of PLAMs derived using Algorithm 3 under the sparseness assumption. Basically, the procedure follows the naive approach given in Algorithm 2. However, the number of targeting structures can be drastically reduced from the naive approach by enumerating the component PLAMs in Algorithm 3; the number of structural candidates to be searched is $\prod_p |\mathcal{Q}^p_\lambda| \ll \prod_p |\mathcal{R}^p_\lambda|$. In order to further reduce the search range of the combination of component PLAMs, the following proposition is used.

Proposition 2 (Screening of Acyclic Subgraphs): Let $\{\{(\boldsymbol{\theta}^{p[i]}, G^{p[i]}_\lambda)\} = \mathcal{Q}^p_\lambda; p \in \mathcal{S}\}$ be a set of pairs composed of a parameter set of PLAM, $\boldsymbol{\theta}^{p[i]}$, and its corresponding objective function, $G^{p[i]}_\lambda$, where $[i]$ indicates the i-th best model from the viewpoint of the objective function value in the set, $\bar{\mathcal{A}}^p = \cup_i \mathcal{A}^p_{\hat{\boldsymbol{\theta}}^{p[i]}}$, be the possible parent nodes for v^p, and $\bar{A}$ be the adjacency matrix constructed with $\{\bar{\mathcal{A}}^p\}$. The directed graph based on $\bar{A}$ suggests that $\boldsymbol{\theta}^{p[1]}$ is a subset of the minimizer of Eq. (15) *if $|\mathcal{S}^m_c(\bar{A})| = 1$ holds for $\mathcal{S}^m_c(\bar{A}) \ni p$.*

Proof: The SCCs of the directed graph based on $\bar{A}$ with a single node v^p indicates that the node v^p does not violate the DAG constraint regardless of which candidate model combination is adopted. This suggests that the PLAM parameter $\boldsymbol{\theta}^{p[1]}$ is a subset of the minimizer given in Eq. (15) since $G^{p[1]}_\lambda \leq G^{p[i]}_\lambda$ holds for any $i > 1$. □

Proposition 2 suggests that we can screen nodes that can adopt the best component PLAM in terms of the objective function before the search process; the search space of the optimization problem composed of combinations of the model candidates derived by Algorithm 3 can be further reduced by this procedure. We also note that multiple trials of Algorithm 4 for selecting the best whole structure can be conducted in parallel.



(a)

(b)

FIGURE 3. Cultivation system used in the case study. (a) Schematic image. (b) Appearance of the system.

V. CASE STUDY: MODELING BASED ON REAL-WORLD DATA

In this section, a graphical model is constructed based on the dataset collected from a real-world plant factory by using the proposed approach.

A. DERIVATION OF PARTIALLY LINEAR ADDITIVE GRAPH STRUCTURE

In this case study, we focus on a dataset intermittently collected at a pilot cultivation system in a plant factory for the period from April 2015 to July 2018; the dataset consists of

TABLE 1. Variables for model construction.

Variable†	Type*	Controllable	Description
v^1	Q	-	Weight of product
v^2	Q	-	# leaves w. tipburns
v^3	S	-	# days for raising seedlings
v^4	S	-	# days for harvesting
v^5	S	-	Amount of nitrate nitrogen in nutrient solution
v^6	S	-	Amount of ammonia nitrogen in nutrient solution
v^7	S	-	Amount of phosphoric acid in nutrient solution
v^8	S	-	Amount of potassium in nutrient solution
v^9, v^{10}	S	-	EC of nutrient circulation tank
v^{11}, v^{12}	S	-	pH of nutrient circulation tank
v^{13}, v^{14}	G	-	Temperature of nutrient solution
v^{15}, v^{16}	G	✓	Water level of nutrient solution tank
v^{17}	G	✓	Moles of lifetime photons
v^{18}	G	✓	Operation state of pure clearizer
v^{19}, v^{20}	G	✓	Water amount of nutrient solution
v^{21}, v^{22}	G	✓	Indoor CO2
v^{23}, v^{24}	G	✓	Temperature of cultivation room
v^{25}, v^{26}	G	✓	Humidity of cultivation room
v^{27}, v^{28}	G	✓	CO2 flow rate
v^{29}, v^{30}	G	✓	Temperature setting
v^{31}, v^{32}	X	-	Outside temperature
v^{33}, v^{34}	X	-	Outside humidity
v^{35}, v^{36}	P	-	Power consumption at internal control panel 200V
v^{37}, v^{38}	P	-	Power consumption of servo motor
v^{39}, v^{40}	P	-	Power consumption at external control panel 200V
v^{41}, v^{42}	P	-	Power consumption of chiller unit
v^{43}, v^{44}	P	-	Power consumption at AC100V
v^{45}	G	✓	Types of fluorescent lamps

† For duplicate variables, the first and second variables represent the average of the light and dark periods, respectively.
∗ Type is categorized as follows: crop quality (Q), scheduled environmental variables (S), growing environmental variables (G), exogenous variables (X), and power consumption (P).

FIGURE 4. Tipburn of frillice lettuce.

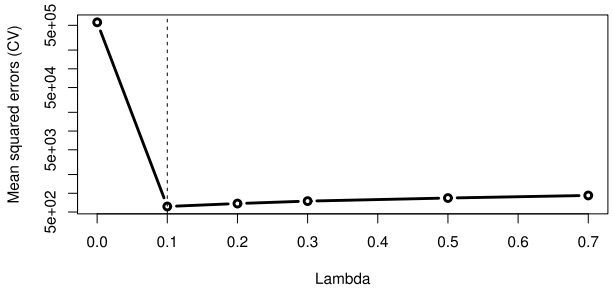


FIGURE 5. Result of cross validation. The cross validation error is minimized at $\lambda = 0.1$.

records of 2,047 frillice lettuces and the average environmental data during their cultivation period. Fig. 3(a) shows the schematic image of the target system and Fig. 3(b) shows its appearance [51], [52]. The pilot system is capable of harvesting five heads of frillice lettuce a day. The system operates various controllable parameters for growing the plants. The daily operation schedule consists of a 16-hour light period and an 8-hour dark period. The average growth environment and growth result for each crop were recorded. Table 1 shows the recorded data used in the evaluation. Typical variables indicating the quality of the target crops include the weight of the crops and the number of leaves in which tipburn has occured (see Fig. 4).

In our proposed procedure, firstly, the sparseness parameter λ was tuned according to the three-fold cross validation from the viewpoint of the total sum of squared errors to the validation data subset. The number of bases was set as $K = 6$ in Eq. (2). Algorithm 4 was performed from 1,000 different initial states. Fig. 5 shows the result of the cross validation. The average of the mean squared errors was minimized at $\lambda = 0.1$.

Fig. 6 shows the graphical model derived with the proposed method by using all the datasets under the selected parameter,

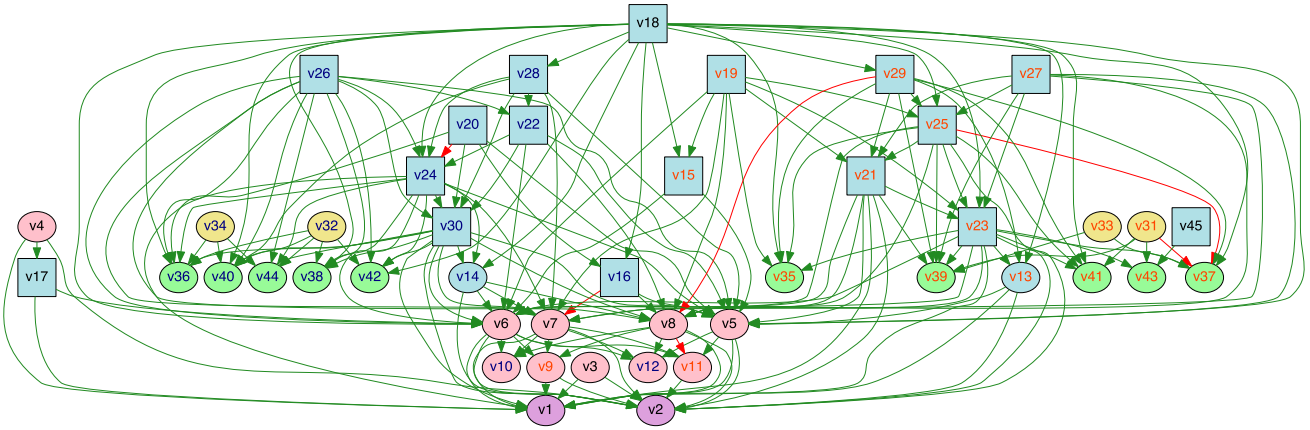


FIGURE 6. Graph structure derived by using the proposed method. Square nodes show the controllable variables. The colors of the nodes show the types of variables: purple for representing crop quality; pink for scheduled environmental variables; blue for growing environmental parameters; yellow for exogenous variables; and green for power consumption. Variable names written in red/blue show the variables of the light/dark periods.

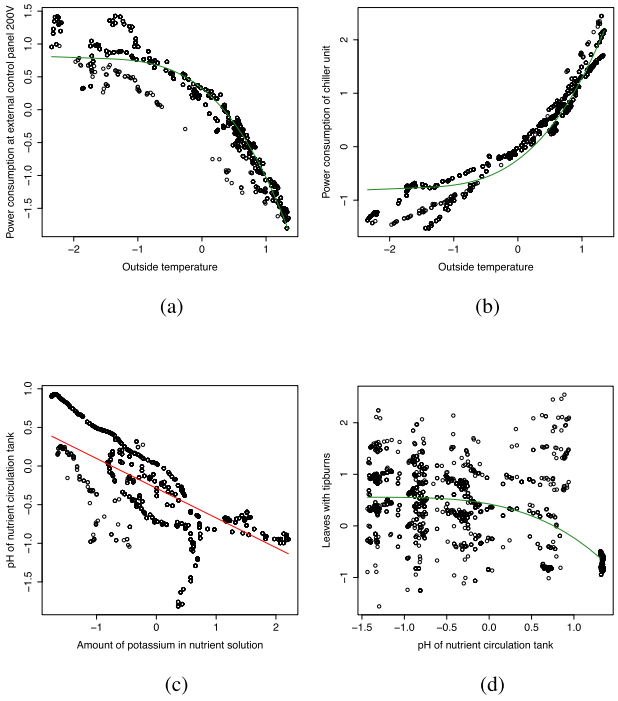


FIGURE 7. Example of the derived relationships between variables: (a) outside temperature versus power consumption at external control panel (200[V]), (b) outside temperature versus power consumption of chiller unit, (c) amount of potassium in nutrient solution versus pH of nutrient circulation tank, and (d) pH of nutrient circulation tank versus the number of leaves with tipburns.

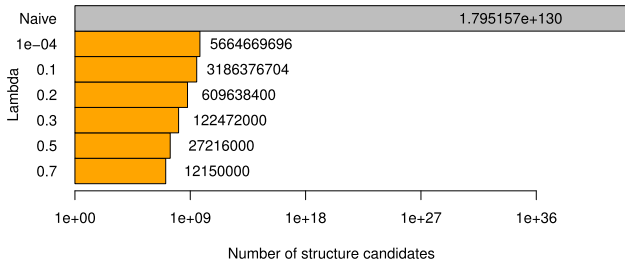


FIGURE 8. The number of candidates for DAG search.

i.e. $\lambda = 0.1$. The arrows in the results indicate the direct relationships of the expected change in the other variables when a variable is controlled. For example, the arrows from v^{32} to v^{40} and v^{42} indicate the relationships between the outside air temperature and the power consumptions. Fig. 7(a) shows the concrete relationship between the outside air temperature and the power consumption at the external control panel (200V); note that the variables shown in Fig. 7 are standardized and plotted by subtracting the expected value of the influence of other explanatories. Since the power consumption of the plug heater accounts for most of the electric power monitored at the external control panel, this result supports an intuitive interpretation that more energy is required when the outside temperature is low. Fig. 7(b) shows another example of the relationship between the outside air temperature and the power consumption of the chiller unit; since the chiller unit is used to lower the room temperature, the higher the outside temperature, the more the power consumption that is required. The results derived by the proposed approach naturally extracts these types of nonlinear relationships. Fig. 7(c) shows an example of the relationship between the pH of the nutrient circulation tank and the amount of potassium in the nutrient; a linear relationship was selected as a result of the partially linear model construction based on the overlap group lasso. Fig. 7(d) shows the relationship between the pH of the nutrient circulation tank and the number of leaves with tipburn; the result suggests a nonlinear relationship by which the number of tipburns tend to decrease when the pH is relatively high. We expect that the analyses performed using such procedures can provide valuable knowledge for realizing energy-aware crop quality control.

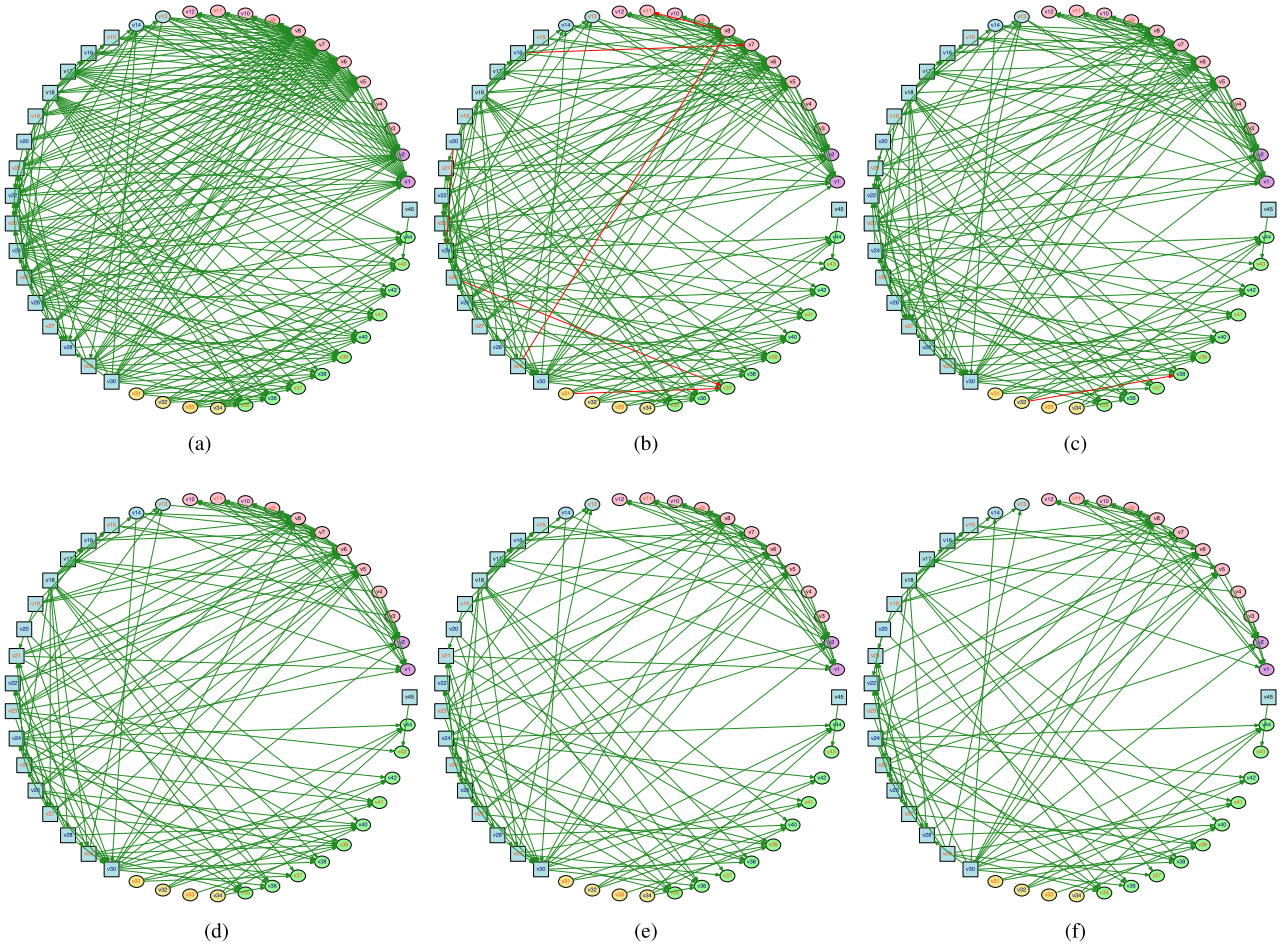


FIGURE 9. Graph structures under various λ. (a) λ = 0.0001 (#links = 234). (b) λ = 0.1 (#links = 177). (c) λ = 0.2 (#links = 162). (d) λ = 0.3 (#links = 136). (e) λ = 0.5 (#links = 113). (f) λ = 0.7 (#links = 101).

B. DISCUSSION

Now, we briefly discuss about the computational difficulty and the model consistency of the proposed approach under various λ.

Fig. 8 shows the size of the search space required for finding the optimal DAG for the given dataset. In the case of the naive approach, the size of the search space, i.e. $\prod_p |\mathcal{R}^p|$, is very large in this problem. However, in the proposed approach, the size of the search space $\prod_p |\mathcal{Q}_\lambda^p|$ is significantly reduced under the sparseness assumption by setting the λ value. The result indicates that the proposed approach realizes efficient reduction of the space for searching the appropriate DAG structure with PLAMs as its components.

Fig. 9 shows the derived graph structures under various λ. The results show that the number of links decreases and becomes a sparse structure as the value of λ increases. This property suggests that the mechanism for controlling the over-fitting in the graphical modeling works well as expected by adjusting the number of links between nodes. Note that the links existing under situations where λ is large also exist even under situations where λ is small in most of the substructures in the derived graphs; these arrows imply a relationship between variables that is determined to be consistently related regardless of the value of λ, suggesting that there is an explicit relationship. Fig. 10 shows the subgraphs under various λ in the corresponding graph structure shown in Fig. 9, focusing on the node subset $\{v^{13}, v^{21}, v^{23}, v^{25}, v^{27}, v^{29}\}$. For example, variables v^{25}, v^{27}, v^{29} are used as explanatories for v^{21} under any λ value; the result implies that these three variables are essentially important for explaining v^{21}. The directed links consistently existing under different λ support the fact that there are very clear relationships among these variables. Meanwhile, focusing on v^{25} and v^{29}, the direction of the arrow between these variables changes depending on λ. As shown in this example, the result has some sub-structure whose link presence and direction changes discontinuously with changes in the value of λ. In the graph structure derived in the previous subsection, the model is derived according to a policy that selects a whole structure with low expected mean squared error by adopting cross validation.

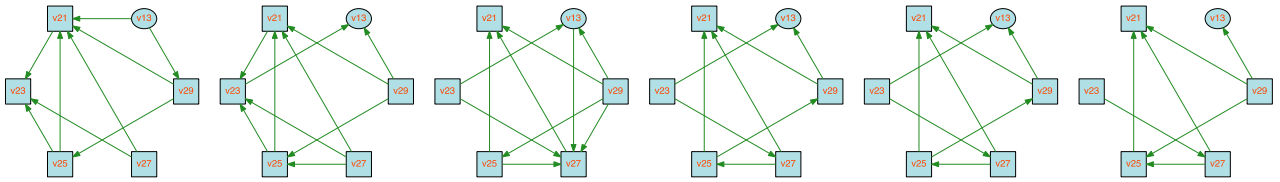


FIGURE 10. Subgraph structures focusing on $\{v^{13}, v^{21}, v^{23}, v^{25}, v^{27}, v^{29}\}$ under various λ. (a) $\lambda = 0.0001$. (b) $\lambda = 0.1$. (c) $\lambda = 0.2$. (d) $\lambda = 0.3$. (e) $\lambda = 0.5$. (f) $\lambda = 0.7$.

However, discontinuous and inconsistent changes in the substructures corresponding to changes in the value of λ can lead to degradation of the interpretability in the analysis; hence, this is an open issue in our proposed approach.

VI. CONCLUDING REMARKS

In this paper, we introduced an idea of directed graphical modeling based on sparse partially linear additive models and proposed approaches to find a sparse partially linear additive structure for the construction of directed acyclic graphical representations. We provided the results of a case study using a multivariate dataset collected at a real-world plant factory and showed that the proposed scheme is effective for finding strong linear/nonlinear relationships among variables while efficiently reducing the search space for obtaining the plausible directed acyclic graph. The proposed analytical model will provide an interpretable estimate of the expected operational effect of plant factories under the changes in exogenous factors such as weather conditions; e.g., the proposed approach can be used in realization of data-driven decision making for operations in plant factories that minimize the expected energy cost while keeping the expected crop quality [11].

This study proposed an approach for analyzing the energy-aware growth control in plant factories; however, the proposed approach is expected to be applicable to various analyses based on the statistical relationship described by the directed acyclic graphical structure between various variables. In this paper, we particularly focused on the class of ordinary (partially linear) additive models; however, but our scheme can be easily extended to a broader class of generalized additive models by introducing arbitrary link functions. We expect that our proposed scheme provides a novel approach for describing statistical relationships while identifying linearity/nonlinearity. The proposed approach will be useful when flexible and interpretable statistical structures among variables are required while focusing on the generalization performance. In this paper, we adopted a hill climbing approach for obtaining relatively good solutions within a finite time, which works well under a large number of variables; however, the realization of efficient and scalable exact search is an interesting future work. We will continue discussing an open issue regarding the lack of structural consistency against changes in the value of λ, which was mentioned during the discussion of our case study, as our future work. We also plan to release the scripts for data-driven construction of the proposed model near future.

REFERENCES

[1] T. Morimoto, K. Hatou, and Y. Hashimoto, "Speaking plant approach for environment control based on artificial intelligence," in *Proc. 12th CIGR World Congr.*, 1994, pp. 190–197.

[2] T. Morimoto and Y. Hashimoto, "AI approaches to identification and control of total plant production systems," *Control Eng. Pract.*, vol. 8, no. 5, pp. 555–567, 2000.

[3] Y. Hashimoto, H. Murase, T. Morimoto, and T. Torii, "Intelligent systems for agriculture in Japan," *IEEE Control Syst. Mag.*, vol. 21, no. 5, pp. 71–85, Oct. 2001.

[4] Japan Greenhouse Horticultural Association. (2018). *Large Scale Facilities Gardening/Plant Factory—Actual Situation Survey Case Studies (in Japanese)*. [Online]. Available: http://www.jgha.com/jisedai/h29/report/29bessatsu1.pdf

[5] D. Kolokotsa, G. Saridakis, K. Dalamagkidis, S. Dolianitis, and I. Kaliakatsos, "Development of an intelligent indoor environment and energy management system for greenhouses," *Energy Convers. Manage.*, vol. 51, no. 1, pp. 155–168, 2010.

[6] T. Yano *et al.*, "Energy-saving night temperature regime for satsuma mandarins (citrus unshiu marc.) Grown in a plastic house with heating. I. Effect of temperature and water condition on fruit growth and quality," *Environ. Control Biol.*, vol. 52, no. 3, pp. 161–166, 2014.

[7] S. Wright, "The method of path coefficients," *Ann. Math. Statist.*, vol. 5, no. 3, pp. 161–215, 1934.

[8] K. G. Jöreskog, "A general method for estimating a linear structural equation system," in *Structural Equation Models in the Social Sciences*. New York, NY, USA: Academic, 1973, pp. 85–112.

[9] A. Ludwig, P. Berthiaume, P. Boerlin, S. Gow, and D. Léger, and F. I. Lewis, "Identifying associations in *Escherichia coli* antimicrobial resistance patterns using additive Bayesian networks," *Preventive Veterinary Med.*, vol. 110, no. 1, pp. 64–75, 2013. doi: 10.1016/j.prevetmed.2013.02.005.

[10] M. Pittavino *et al.*, "Comparison between generalized linear modelling and additive Bayesian network; Identification of factors associated with the incidence of antibodies against *Leptospira interrogans sv Pomona* in meat workers in New Zealand," *Acta Tropica*, vol. 173, pp. 191–199, Sep. 2017.

[11] S. Murakami, Y. Fujimoto, Y. Hayashi, H. Fuchikami, and T. Hattori, "Energy cost minimization in plant factories considering weather factors using additive Bayesian networks," *J. Int. Council Elect. Eng.*, vol. 8, no. 1, pp. 127–134, 2018.

[12] A. Chouldechova and T. Hastie. (2015). "Generalized additive model selection." [Online]. Available: https://arxiv.org/abs/1506.03850

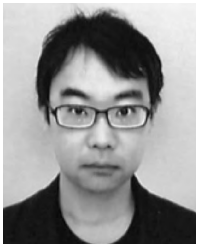
[13] Y. Lou, J. Bien, R. Caruana, and J. Gehrke, "Sparse partially linear additive models," *J. Comput. Graph. Statist.*, vol. 25, no. 4, pp. 1126–1140, 2016.

[14] S. Hara and T. Maehara, "Enumerate lasso solutions for feature selection," in *Proc. AAAI*, 2017, pp. 1985–1991.

[15] A. J. U. ten Cate, "Computer control of greenhouse climates," in *Proc. Symp. Potential Productiv. Protected Cultivation*, vol. 87, 1978, pp. 265–272.

[16] H. Nishina, Y. Hashimoto, and A. Matsumoto, "Identification and control of environment in greenhouse systems," in *Proc. II IFAC/ISHS Workshop, Math. Control Appl. Agricult. Horticulture*, vol. 406, 1995, pp. 29–37.

[17] Y. Hashimoto, "Computer control of short term plant growth by monitoring leaf temperature," in *Proc. Symp. Comput. Greenhouse Climate Control*, vol. 106, 1980, pp. 139–146.
[18] Y. Hashimoto, "Recent strategies of optimal growth regulation by the speaking plant concept," in *Proc. Int. Symp. Growth Yield Control Vegetable Prod.*, vol. 260, 1989, pp. 115–121.
[19] Y. Hashimoto, G. P. Bot, W. Day, H.-J. Tantau, and H. Nonami, Eds., *The Computerized Greenhouse -Automatic Control Application in Plant Production*. New York, NY, USA: Academic, 1993.
[20] J. D. Pujari, R. Yakkundimath, and A. S. Byadgi, "Neuro-kNN classification system for detecting fungal disease on vegetable crops using local binary patterns," *Agricult. Eng. Int., CIGR J.*, vol. 16, no. 4, pp. 299–308, 2014.
[21] N. Monjezi, M. Sheikhdavoodi, H. Zakidizaji, A. Marzban, and M. Shomeili, "Operations scheduling of sugarcane production using fuzzy GERT method (Part II: Preserve operations, harvesting and ratooning)," *Agricult. Eng. Int., CIGR J.*, vol. 18, no. 3, pp. 343–349, 2016.
[22] K. G. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis, "Machine learning in agriculture: A review," *Sensors*, vol. 18, no. 8, pp. 1–29, 2018.
[23] K. Omasa, "A perspective on smart greenhouse- factory production systems, environment and energy measures, and intelligent informatization," *J. Sci. High Technol. Agricult.*, vol. 26, no. 2, pp. 89–97, 2014.
[24] P. Qasemi-Kordkheili, M. A. Asoodar, and N. Kazemi, "Application of a non-parametric method to analyze energy consumption for orange production," *Agricult. Eng. Int., CIGR J.*, vol. 16, no. 1, pp. 157–166, 2014.
[25] J. Pearl, *Probabilistic Reasoning in Intelligent Systems*. San Mateo, CA, USA: Morgan Kaufmann, 1988.
[26] J. C. Westland, *Structural Equation Models: From Paths to Networks*. Cham, Switzerland: Springer, 2015.
[27] M. M. Wall and Y. Amemiya, "Title: Nonlinear structural equation modeling as a statistical method," *Handbook of Latent Variable and Related Models*, S.-Y. Lee, Ed. Amsterdam, The Netherlands: North-Holland, 2007, pp. 321–343.
[28] H. W. Marsh *et al.*, "Doubly-latent models of school contextual effects: Integrating multilevel and structural equation approaches to control measurement and sampling error," *Multivariate Behav. Res.*, vol. 44, no. 6, pp. 764–802, 2009.
[29] K. Schermelleh-Engel, C. S. Werner, A. G. Klein, and H. Moosbrugger, "Nonlinear structural equation modeling: Is partial least squares an alternative?" *AStA Adv. Stat. Anal.*, vol. 94, no. 2, pp. 167–184, 2010.
[30] W. L. Leite and Y. Zuo, "Modeling latent interactions at level 2 in multilevel structural equation models: An evaluation of mean-centered and residual-centered unconstrained approaches," *Struct. Equation Model., Multidisciplinary J.*, vol. 18, no. 3, pp. 449–464, 2011.
[31] K. Schermelleh-Engel, M. Kerwer, and A. G. Klein, "Evaluation of model fit in nonlinear multilevel structural equation modeling," *Frontiers Psychol.*, vol. 5, pp. 1–11, Mar. 2014.
[32] J. A. Nelder and R. W. Wedderburn, "Generalized linear models," *J. Roy. Stat. Soc., A (Gen.)*, vol. 135, no. 3, pp. 370–384, 1972.
[33] A. J. Dobson and A. G. Barnett, *An Introduction to Generalized Linear Models*, 3rd ed. London, U.K.: Chapman & Hall, 2008.
[34] F. I. Lewis and B. J. J. McCormick, "Revealing the complexity of health determinants in resource-poor settings," *Amer. J. Epidemiology*, vol. 176, no. 11, pp. 1051–1059, 2012.
[35] Q. Li, "Efficient estimation of additive partially linear models," *Int. Econ. Rev.*, vol. 41, no. 4, pp. 1073–1092, 2000.
[36] W. Härdle and H. Liang, "Partially linear models," in *Statistical Methods for Biostatistics and Related Fields*, 2007, pp. 87–103.
[37] T. Zhao, G. Cheng, and H. Liu, "A partially linear framework for massive heterogeneous data," *Ann. Statist.*, vol. 44, no. 4, pp. 1400–1437, 2016.
[38] A. Buja, T. Hastie, and R. Tibshirani, "Linear smoothers and additive models," *Ann. Statist.*, vol. 17, no. 2, pp. 453–510, 1989.
[39] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Statist. Soc., B (Methodol.)*, vol. 58, no. 1, pp. 267–288, 1996.
[40] P. Ravikumar, J. Lafferty, H. Liu, and L. Wasserman, "Sparse additive models," *J. Roy. Stat. Soc., B, Stat. Methodol.*, vol. 71, no. 5, pp. 1009–1030, 2009.
[41] J. Huang, J. L. Horowitz, and F. Wei, "Variable selection in nonparametric additive models," *Ann. Statist.*, vol. 38, no. 4, pp. 2282–2313, 2010.
[42] H. H. Zhang, G. Cheng, and Y. Liu, " Linear or nonlinear? Automatic structure discovery for partially linear models," *J. Amer. Stat. Assoc.*, vol. 106, no. 495, pp. 1099–1112, 2011.
[43] J. Pearl, "Bayesian networks: A model of self-activated memory for evidential reasoning," in *Proc. 7th Conf. Cognit. Sci. Soc.*, 1985, pp. 329–334.
[44] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, 2nd ed. Cambridge, MA, USA: MIT Press, 2001.
[45] D. Heckerman, D. Geiger, and D. M. Chickering, "Learning Bayesian networks: The combination of knowledge and statistical data," *Mach. Learn.*, vol. 20, no. 3, pp. 197–243, Sep. 1995.
[46] R. Tarjan, "Depth-first search and linear graph algorithms," *SIAM J. Comput.*, vol. 1, no. 2, pp. 146–160, 1972.
[47] B. Korte and J. Vygen, *Combinatorial Optimization: Theory and Algorithms*, 4th ed. Berlin, Germany: Springer, 2008.
[48] A. F. Y. Poon, F. I. Lewis, S. L. K. Pond, and S. D. W. Frost, "Evolutionary interactions between N-linked glycosylation sites in the HIV-1 envelope," *PLoS Comput. Biol.*, vol. 3, no. 1, pp. 110–119, 2007.
[49] L. Jacob, G. Obozinski, and J.-P. Vert, "Group lasso with overlap and graph lasso," in *Proc. ICML*, 2009, pp. 433–440.
[50] P. Breheny and J. Huang, "Group descent algorithms for nonconvex penalized linear and logistic regression models with grouped predictors," *Statist. Comput.*, vol. 25, no. 2, pp. 173–187, 2015.
[51] N. Sakamoto, N. Kashihara, and T. Kamimura, "Plant cultivation apparatus," Japan Patent 63 712A, Apr. 6, 2017.
[52] N. Sakamoto, N. Kashihara, and T. Kamimura, "Plant cultivation appratus and plant factory," Japan Patent 63 713A, Apr. 6, 2017.



YU FUJIMOTO (M'17) received the Ph.D. degree in engineering from Waseda University, Tokyo, Japan, in 2007, where he is currently an Associate Professor with the Advanced Collaborative Research Organization for Smart Society. His primary areas of interest are machine learning and statistical data analysis. His current research interests include data mining in energy domains, especially with regard to controlling the power in smart grids, and statistical prediction of the power fluctuations due to the introduction of a large number of renewable energy sources. He is a member of the Information Processing Society of Japan.

SAYA MURAKAMI received the B.Eng. degree in electrical engineering and bioscience from Waseda University, Japan, in 2017, where she is currently pursuing the degree with the Department of Electrical Engineering and Bioscience. Her research interest includes energy management of plant factories. She is a Student Member of the Institute of Electrical Engineers in Japan (IEEJ).

NANAE KANEKO received the B.Eng. degree in electrical engineering and bioscience from Waseda University, Japan, in 2018, where she is currently pursuing the degree with the Department of Advanced Science and Engineering. Her research interest includes forecast of electricity power demand.

HIDEKI FUCHIKAMI received the M.E. degree from Saga University, in 2004. He has been with Mayekawa Mfg. Co., Ltd., Ibaraki, Japan, since 2004. His current research interest includes industrial heat pump systems.

TOSHIROU HATTORI received the M.E. degree from Waseda University, Japan, in 1979. He joined Mayekawa Mfg. Co., Ltd., in 1979. He became the Director of the R&D Center, in 2009, and is currently the Executive Director. His research interests include chemical engineering and product development.

YASUHIRO HAYASHI (M'91) received the B.Eng., M.Eng., and D.Eng. degrees from Waseda University, Japan, in 1989, 1991, and 1994, respectively. In 1994, he became a Research Associate with Ibaraki University, Mito, Japan. In 2000, he became an Associate Professor with the Department of Electrical and Electronics Engineering, Fukui University, Fukui, Japan. He has been with Waseda University as a Professor of the Department of Electrical Engineering and Bioscience, since 2009, and as the Director of the Research Institute of Advanced Network Technology, since 2010. Since 2014, he has been the Dean of the Advanced Collaborative Research Organization for Smart Society, Waseda University. His current research interests include the optimization of distribution system operation and forecasting, and operation, planning, and control concerned with renewable energy sources and demand response. He is a member of the Institute of Electrical Engineers of Japan and a Regular Member of CIGRE SC C6 (Distribution Systems and Dispersed Generation).