

Research Article

A New Model for Building Energy Modeling and Management Using Predictive Analytics: Partitioned Hierarchical Multitask Regression (PHMR)

Shuluo Ning¹ and Hyunsoo Yoon²

¹Amazon Web Service, 410 Terry Ave N, Seattle, WA 98109-5210, USA ²Department of Industrial Engineering, Yonsei University, Seoul 03722, Republic of Korea

Correspondence should be addressed to Hyunsoo Yoon; hs.yoon@yonsei.ac.kr

Received 3 August 2023; Revised 12 February 2024; Accepted 13 February 2024; Published 11 March 2024

Academic Editor: Faming Wang

Copyright © 2024 Shuluo Ning and Hyunsoo Yoon. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Buildings are major consumers of energy, accounting for a significant proportion of total energy use worldwide. This substantial energy consumption not only leads to increased operational costs but also contributes to environmental concerns such as greenhouse gas emissions. In the United States, building energy consumption accounts for about 40% of total energy use, highlighting the importance of efficient energy management. Therefore, accurate prediction of energy usage in buildings is crucial. However, accurate prediction of building energy consumption remains a challenge due to the intricate interaction of indoor and outdoor variables. This study introduces the Partitioned Hierarchical Multitask Regression (PHMR), an innovative model integrating recursive partition regression (RPR) with multitask learning (hierML). PHMR adeptly predicts building energy consumption by integrating both indoor factors, such as building design and operational variables, and outdoor environmental influences. Rigorous simulation studies illustrate PHMR's efficacy. It outperforms traditional single-predictor regression models, achieving a 32.88% to 41.80% higher prediction accuracy, especially in scenarios with limited training data. This highlights PHMR's robustness and adaptability. The practical application of PHMR in managing a modular house's Heating, Ventilation, and Air Conditioning (HVAC) system in Spain resulted in a 37% improvement in prediction accuracy. This significant efficiency gain is evidenced by a high Pearson correlation coefficient (0.8) between PHMR's predictions and actual energy consumption. PHMR not only offers precise predictions for energy consumption but also facilitates operational cost reductions, thereby enhancing sustainability in building energy management. Its application in a real-world setting demonstrates the model's potential as a valuable tool for architects, engineers, and facility managers in designing and maintaining energy-efficient buildings. The model's integration of comprehensive data analysis with domain-specific knowledge positions it as a crucial asset in advancing sustainable energy practices in the building sector.

1. Introduction

In an era marked by escalating energy demands, the imperative to enhance building energy efficiency has become increasingly critical. The Department of Energy's 2023 report underscores this urgency, highlighting that buildings account for a substantial portion of total energy use in the United States [1]. This situation presents not just a challenge in resource management but also raises significant environmental and economic concerns. The integration of wireless sensors and IoT technologies has paved new avenues for understanding and controlling building energy usage. When combined with predictive analytics, these technologies show considerable promise in forecasting energy consumption and enabling the fine-tuning of building parameters to minimize waste.

Despite these technological advancements, the development of effective predictive models for building energy consumption is fraught with challenges, primarily due to the dynamic and multifaceted nature of energy usage. While recent studies, such as those by Himeur et al. [2, 3], Deng et al. [4], Elnour et al. [5], and Han et al. [6], have illuminated the potential of AI and big data in revolutionizing building energy management, they also reveal significant gaps in current predictive modeling approaches. These studies highlight the need for models that can effectively navigate the complex interplay between various indoor and outdoor variables affecting energy consumption.

In response to this identified research gap, our study introduces the Partitioned Hierarchical Multitask Regression (PHMR) model. The PHMR model is designed to address the complexities of building energy consumption by integrating recursive partitioning regression (RPR) with multitask learning (hierML). This approach not only enhances the accuracy of energy consumption predictions but also facilitates the nuanced control and adaptation of building parameters. Our model represents a significant advancement from existing RPR models Chan et al. [7], Chaudhuri et al. [8], Landwehr et al. [9], Loh [10], Loh [11], Zeileis et al. [12], Rusch and Zeileis [13], filling a critical gap in the literature and offering a more sophisticated tool for energy management.

The key contributions of this study are manifold. We present the model formulation of PHMR, which bridges recursive partitioning on outdoor variables with hierarchical multitask learning to enhance prediction and control of building energy consumption. The transformation of the hierML algorithm into a convex optimization problem is detailed, ensuring optimal and computationally efficient solutions. Furthermore, the practical application of PHMR in managing a modular house's HVAC system in Spain demonstrates the model's utility in reducing unnecessary energy consumption, aligning with the sustainable goals outlined in recent research by Borràs et al. [14] and Mehdizadeh Khorrami et al. [15].

This paper is organized to detail the PHMR model formulation, the algorithm used for model estimation, its comparison against other methods through simulation studies, and its practical application in building energy prediction and management.

2. Model Formulation for PHMR

The PHMR model consists of two components: (1) a treegrowing process that utilizes the outdoor variables for recursive partitioning and (2) hierML performed at each step of the tree-growing process. In this section (Section 2), we will discuss the hierML model at a fixed step (*s*-th step) of the tree-growing process. The other steps use hierML in a similar manner. In the following section (Section 3), we will present the algorithm that is used to estimate the model parameters for hierML, which is integrated with recursive partitioning to grow a tree.

The *s*-th step of the tree-growing process yields a tree T_s , consisting of internal and leaf nodes. All nodes of T_s , denoted as \mathbf{V}_s , include the leaf nodes $\mathbf{V}_{l,s}$, which do not have any children in the tree. At each internal node of \mathbf{V} , a \mathbf{Z} variable is used to partition the samples into left and right branches. The leaf nodes correspond to different subdivisions of the space defined by the outdoor variables, and for each leaf node, there exists a regression model linking the response with indoor variables: $\mathbf{y}_{v_i} = \mathbf{X}_{v_i} \mathbf{\beta}_{v_i} + \mathbf{\varepsilon}_{v_i}$ and $v_l \in$

 $\mathbf{V}_{l,s}$. Figure 1 provides an example of T_s with leaf nodes $\mathbf{V}_s = \{3, 5, 6, 7\}$.

In previous recursive partitioning regression (RPR) models, including our SPR [16, 17], each leaf node's regression model was fitted independently. However, in our proposed hierML model, we incorporate the hierarchical multilevel similarity structure of the leaf nodes to fit the models jointly. According to the principle of recursive partitioning, models at lower levels of leaf nodes should be more similar to each other. For instance, in Figure 1, the model at leaf node 6 should be more similar to the model at 7 than to that at 5 and that at 3 (and in fact, less similar to 3 than to 5). This is because the recursive partitioning process performs easier splits at earlier stages, resulting in more dissimilar branches and easier splits. As the split becomes later, it becomes harder because the resulting branches become more and more similar.

To jointly estimate regression models for the leaf nodes while incorporating the hierarchical multilevel similarity structure, a penalized formulation is proposed as follows:

$$\widehat{\mathbf{B}} = \underset{\mathbf{B}}{\operatorname{argmin}} \sum_{\nu_l} \left(\mathbf{y}_{\nu_l} - \mathbf{X}_{\nu_l} \boldsymbol{\beta}_{\nu_l} \right)^T \left(\mathbf{y}_{\nu_l} - \mathbf{X}_{\nu_l} \boldsymbol{\beta}_{\nu_l} \right) + \lambda \sum_j \sum_{\nu} w_{\nu} \left\| \boldsymbol{\beta}_{\mathbf{G}_{\nu}}^j \right\|_2,$$
(1)

where $\mathbf{B} = \{\boldsymbol{\beta}_{v_l}\}_{v_l \in \mathbf{V}_l}$.

The first part of equation (1) is the least-square error loss. For simplicity, the subscript "s" has been removed from all the notations in this section since we are focusing on a single step in the tree-growing process. The second term in equation (1) is the hierML penalty, and there are some additional notations that need to be defined before explaining the penalty clearly. For each node in the tree, denoted as $v \in \mathbf{V}$, \mathbf{G}_v represents the set of leaf nodes that grow from node v. Taking T_s in Figure 1 as an example, there are seven nodes in the tree. The \mathbf{G}_v for each node is $\mathbf{G}_1 = \{3, 5, 6, 7\}$, $\mathbf{G}_2 = \{5, 6, 7\}$, $\mathbf{G}_3 = \{3\}$, $\mathbf{G}_4 = \{6, 7\}$, $\mathbf{G}_5 = \{5\}$, $\mathbf{G}_6 = \{6\}$, and $\mathbf{G}_7 = \{7\}$. Let $\boldsymbol{\beta}_{\mathbf{G}_v}^j$ contain the set of regression coefficients corresponding to the *j*-th predictor (i.e., *j*-th indoor variable) in \mathbf{G}_v . For example, $\boldsymbol{\beta}_{\mathbf{G}_1}^j = \{\beta_3^j, \beta_5^j, \beta_6^j, \beta_7^j\}$, $\boldsymbol{\beta}_{\mathbf{G}_2}^j = \{\beta_5^j,$ $\boldsymbol{\beta}_6^j, \boldsymbol{\beta}_7^j\}$, and $\boldsymbol{\beta}_{\mathbf{G}_3}^j = \{\beta_3^j\}$.

To jointly estimate the leaf node-wise regression models while incorporating the hierarchical multilevel similarity structure, we propose a penalized formulation inspired by multitask learning using L_1/L_2 regularization Obozinski et al. [18]. Specifically, the set of leaf nodes growing from a given node ν is denoted as \mathbf{G}_{ν} . By considering the \mathbf{G}_{ν} as distinct tasks, we apply a weighting strategy to L_2 -norm on $\boldsymbol{\beta}_{\mathbf{G}_1}^j$, denoted as $w_{\nu} \|\boldsymbol{\beta}_{\mathbf{G}_{\nu}}^j\|_2$, where w_{ν} is a weight to be discussed later. They then put an L_1 -norm outside the weighted L_2 -norm, i.e., $\sum_{\nu} w_{\nu} \|\boldsymbol{\beta}_{\mathbf{G}_{\nu}}^j\|_2$, to enable the selection of regression coefficients contained in each $\boldsymbol{\beta}_{\mathbf{G}_{\nu}}^j$ as a group. Additionally, they put another L_1 -norm further outside to enable the selection of regression coefficients corresponding to the *j*-th predictor as a group, i.e., $\sum_j \sum_{\nu} w_{\nu} \|\boldsymbol{\beta}_{\mathbf{G}_{\nu}}^j\|_2$. This completes the explanation of the hierML penalty, which



FIGURE 1: To simplify the presentation, an instance of the tree growth process and its corresponding symbols are displayed, showcasing only two consecutive steps.

is the second term in the formulation. λ in the formulation is a tuning parameter that balances the least-square loss and the proposed penalty [17].

In addition, we will explain how to select the weight w_{ν} for each node ν . When the tree splits into a left and a right branch at each internal node, the regression models at the two branches should exhibit some similarities because they share the same internal node. However, the models should not be identical, or else the internal node would not have been split. For instance, at the lowest internal node, $\nu = 4$ in Figure 1, the tree splits into nodes 6 and 7 as the two branches. The regression coefficients of the *j*-th predictor in the models of nodes 6 and 7, β_6^j and β_7^j , should be comparable but not precisely identical.

$$W_{j}(4) \triangleq g_{4}\sqrt{\left(\beta_{6}^{j}\right)^{2} + \left(\beta_{7}^{j}\right)^{2}} + s_{4}\left(\left|\beta_{6}^{j}\right| + \left|\beta_{7}^{j}\right|\right), \quad (2)$$

where $\sqrt{(\beta_6^j)^2 + (\beta_7^j)^2}$ encourages the two coefficients to be selected jointly to account for their similarity, while $|\beta_6^j|$ and $|\beta_7^j|$ encourage selection separately to account for their difference. g_4 and s_4 are the corresponding weights. Using the definition of $\beta_{G_4}^j$, (2) can be written as

$$W_{j}(4) = g_{4} \left\| \boldsymbol{\beta}_{\mathbf{G}_{4}}^{j} \right\|_{2} + s_{4} \left(\left| \boldsymbol{\beta}_{6}^{j} \right| + \left| \boldsymbol{\beta}_{7}^{j} \right| \right).$$
(3)

Furthermore, we can move up to the next internal node, i.e., v = 2. At node 2, the tree splits into a subtree rooted from node 4 as the left branch and node 5 as the right branch. Following a similar idea to (3), we can write down the penalization on regression coefficients that simultaneously account for the similarity and difference between the two branches, i.e.,

$$W_{j}(2) = g_{2} \left\| \boldsymbol{\beta}_{\mathbf{G}_{2}}^{j} \right\|_{2} + s_{2} \left(W_{j}(4) + \left| \boldsymbol{\beta}_{5}^{j} \right| \right).$$
(4)

In a similar way, the penalization associated with the internal node v = 1 is as follows:

$$W_{j}(1) = g_{1} \left\| \boldsymbol{\beta}_{\mathbf{G}_{1}}^{j} \right\|_{2} + s_{1} \left(W_{j}(2) + \left| \boldsymbol{\beta}_{3}^{j} \right| \right).$$
(5)

To generalize the above scheme, we can write the definition of $W_i(v)$ as

$$W_{j}(v) = \begin{cases} g_{\nu} \left\| \boldsymbol{\beta}_{\mathbf{G}_{\nu}}^{j} \right\|_{2} + s_{\nu} \sum_{c \in children(\nu)} W_{j}(c) & \text{if } \nu \text{ is an internal node} \\ \left| \boldsymbol{\beta}_{\nu}^{j} \right| & \text{if } \nu \text{ is a leaf node} \end{cases}$$
(6)

with $g_v + s_v = 1$ for identifiability consideration. Using this definition, we can write the hierML penalty in (1) as

$$\sum_{j} \sum_{\nu} w_{\nu} \left\| \boldsymbol{\beta}_{\mathbf{G}_{\nu}}^{j} \right\|_{2} = \sum_{j} W_{j}(\nu_{\text{root}}).$$
⁽⁷⁾

By some algebraic operations, it can be demonstrated that w_v is associated with g_v and s_v in the following manner, where v_{root} is the tree's root node:

$$w_{v} = \begin{cases} g_{v} \prod_{m \in \operatorname{Ancestors}(v)} s_{m} & \text{if } v \text{ is an internal node} \\ \prod_{m \in \operatorname{Ancestors}(v)} s_{m} & \text{if } v \text{ is a leaf node} \end{cases} \end{cases}.$$
(8)

This completes the discussion on designing the weight w_{ν} for the proposed hierML penalty. For better illustration, take T_s in Figure 1 as an example. The right-hand side of (7) can be shown to be:

$$W_{j}(1) = g_{1} \left\| \boldsymbol{\beta}_{G_{1}}^{j} \right\|_{2}^{} + s_{1} \left(W_{j}(2) + \left| \boldsymbol{\beta}_{3}^{j} \right| \right) \\ = g_{1} \left\| \boldsymbol{\beta}_{G_{1}}^{j} \right\|_{2}^{} + s_{1} \left[g_{2} \left\| \boldsymbol{\beta}_{G_{2}}^{j} \right\|_{2}^{} + s_{2} \left(W_{j}(4) + \left| \boldsymbol{\beta}_{5}^{j} \right| \right) + \left| \boldsymbol{\beta}_{3}^{j} \right| \right] \\ = g_{1} \left\| \boldsymbol{\beta}_{G_{1}}^{j} \right\|_{2}^{} + s_{1} \left\{ g_{2} \left\| \boldsymbol{\beta}_{G_{2}}^{j} \right\|_{2}^{} + s_{2} \left[g_{4} \left\| \boldsymbol{\beta}_{G_{4}}^{j} \right\|_{2}^{} \right. \\ \left. + s_{4} \left(\left| \boldsymbol{\beta}_{6}^{j} \right| + \left| \boldsymbol{\beta}_{7}^{j} \right| \right) + \left| \boldsymbol{\beta}_{5}^{j} \right| \right] + \left| \boldsymbol{\beta}_{3}^{j} \right| \right\} \\ = g_{1} \left\| \boldsymbol{\beta}_{G_{1}}^{j} \right\|_{2}^{} + g_{2} s_{1} \left\| \boldsymbol{\beta}_{G_{2}}^{j} \right\|_{2}^{} + g_{4} s_{1} s_{2} \left\| \boldsymbol{\beta}_{G_{4}}^{j} \right\|_{2}^{} + s_{1} s_{2} s_{4} \left| \boldsymbol{\beta}_{6}^{j} \right| \\ \left. + s_{1} s_{2} s_{4} \left| \boldsymbol{\beta}_{7}^{j} \right| + s_{1} s_{2} \left| \boldsymbol{\beta}_{5}^{j} \right| + s_{1} \left| \boldsymbol{\beta}_{3}^{j} \right|. \end{cases}$$

$$\tag{9}$$

Comparing the above equation with the left-hand side of (7), we can get $w_1 = g_1$, $w_2 = g_2 s_1$, $w_4 = s_1 s_2 g_4$, $w_6 = s_1 s_2 s_4$,

 $w_7 = s_1 s_2 s_4$, $w_5 = s_1 s_2$, and $w_3 = s_1$. It is easy to verify that these weights comply with the formula in (8).

It is important to note that a coefficient can be penalized in multiple groups using the proposed hierML penalty. For example, β_6^j is penalized in four groups according to (9): once by itself as $|\beta_6^j|$ and three other times in groups \mathbf{G}_1 , \mathbf{G}_2 , and \mathbf{G}_4 , respectively. These groups have a nested structure, where $\{6\}$ is a subset of \mathbf{G}_4 , \mathbf{G}_4 is a subset of \mathbf{G}_2 , and \mathbf{G}_2 is a subset of \mathbf{G}_1 . This is a general property of hierML, where a regression coefficient of each leaf node is penalized in multiple nested groups with each group corresponding to an ancestor of the leaf node. However, the weighting scheme proposed ensures that the weights corresponding to the multiple nested groups for each leaf node sum up to one, which balances the penalization of regression coefficients in all leaf nodes [17]. This property is presented mathematically in Proposition 1.

Proposition 1. For any leaf node v_l , let $Path(v_l)$ be a set of nodes including all the ancestors of v_l and itself. For each $v \in Path(v_l)$, let w_v be the weight associated with node v in the hierML penalty in (1). Then,

$$\sum_{v \in Path(v_l)} w_v = 1.$$
 (10)

Detailed proofs are available in a separate supplementary document (Supplementary Material, Proposition 1 (available here)) for clarity.

In summary, our proposed hierML model is defined by the optimization in (1) with the weight w_{ν} given in (8), which depends on the choice of g_v and s_v . To determine s_v , we suggest the following approach: recall that s_{ν} represents the idea that the regression coefficients at each branch of node v should be different, while q_v assumes they should be similar. Therefore, a larger value of s_{ν} implies that we want the coefficients to be estimated more independently of each other. The proposed approach for choosing the weight s_v is to make it proportional to the distance between node v and the bottom level of the tree. This is based on the idea that the farther away a node is from the bottom of the tree, or the closer it is to the root node, the more different the regression coefficients at each branch of v should be. This is because recursive partitioning generally produces easier splits at earlier stages of the partitioning process, resulting in more dissimilar branches. As the split occurs at later stages, the resulting branches become more similar, making the split more difficult. By making s_{ν} proportional to the distance between node v and the bottom level of the tree, the proposed approach takes into account this principle of recursive partitioning.

Once the s_v for all the nodes are specified, w_v can be obtained using (8). Next, we use the example in Figure 2 to explain how the weight for node 6, w_6 , is obtained. Because node 6 is a leaf node, we need to specify s_u for all $u \in$ Ancestors(6) which include nodes 1, 2, and 4. Node 1 is three levels up from the bottom level, so $s_1 = 3$. Likewise, we can get $s_2 = 2$ and $s_4 = 1$. Normalize these weights to warrant the equality in (10) to hold, we can get $s_1 = 1$, $s_2 = 0.67$,

and $s_4 = 0.33$. Then, $w_6 = 0.22$ using (8). The weights corresponding to other nodes can be obtained in a similar way. Eventually, we can get the hierML penalty term in (1) written as

$$\sum_{j} \sum_{\nu} w_{\nu} \left\| \boldsymbol{\beta}_{\mathbf{G}_{\nu}}^{j} \right\|_{2} = 0 \cdot \left\| \boldsymbol{\beta}_{\mathbf{G}_{1}}^{j} \right\|_{2} + 0.33 \left\| \boldsymbol{\beta}_{\mathbf{G}_{2}}^{j} \right\|_{2} + 0.45 \left\| \boldsymbol{\beta}_{\mathbf{G}_{4}}^{j} \right\|_{2} + 0.22 \left| \boldsymbol{\beta}_{6}^{j} \right| + 0.22 \left| \boldsymbol{\beta}_{7}^{j} \right| + 0.67 \left| \boldsymbol{\beta}_{5}^{j} \right| + \left| \boldsymbol{\beta}_{3}^{j} \right|.$$
(11)

3. Algorithm for Model Estimation of PHMR

To solve the optimization problem in (1), our first step is to convert it into an alternative convex optimization following a similar idea proposed by Bach [19] for solving group lasso, i.e.,

$$\widehat{\mathbf{B}} = \underset{\mathbf{B}}{\operatorname{argmin}} \sum_{\nu_l} \left(\mathbf{y}_{\nu_l} - \mathbf{X}_{\nu_l} \boldsymbol{\beta}_{\nu_l} \right)^T \left(\mathbf{y}_{\nu_l} - \mathbf{X}_{\nu_l} \boldsymbol{\beta}_{\nu_l} \right) + \lambda \left(\sum_j \sum_{\nu} w_{\nu} \left\| \boldsymbol{\beta}_{\mathbf{G}_{\nu}}^j \right\|_2 \right)^2.$$
(12)

The optimization in (12) is convex but with a nonsmooth penalty, which is difficult to solve directly. We relax the penalty term as follows:

$$\left(\sum_{j}\sum_{\nu}w_{\nu}\left\|\boldsymbol{\beta}_{\mathbf{G}_{\nu}}^{j}\right\|_{2}\right)^{2}\leq\sum_{j}\sum_{\nu}d_{j,\nu}^{-1}\left(w_{\nu}\left\|\boldsymbol{\beta}_{\mathbf{G}_{\nu}}^{j}\right\|_{2}\right)^{2},\qquad(13)$$

where $\sum_{j}\sum_{\nu}d_{j,\nu} = 1$ and $d_{j,\nu} \ge 0$.

Proposition 2. The equality of (13) holds, i.e., $(\sum_{j}\sum_{v}w_{v} \|\beta_{G_{v}}^{j}\|_{2})^{2} = \sum_{j}\sum_{v}d_{j,v}^{-1}(w_{v}\|\beta_{G_{v}}^{j}\|_{2})^{2}$, if and only if $d_{j,v} = w_{v}$ $\|\beta_{G_{v}}^{j}\|_{2}/\sum_{j}\sum_{v}w_{v}\|\beta_{G_{v}}^{j}\|_{2}$.

The findings of this study align with the theory presented in Supplementary Proposition 1 (Supplementary Material, Proposition 1). Detailed proofs are available in a separate supplementary document (Supplementary Material, Proposition 2) for clarity [17].

Using the results in Proposition 2, we can write the optimization in (12) into an equivalent format with a smooth penalty term, i.e.,

$$(\widehat{\mathbf{B}}, \widehat{\mathbf{D}}) = \underset{\mathbf{B}, \mathbf{D}}{\operatorname{argmin}} \sum_{\nu_l} \left(\mathbf{y}_{\nu_l} - \mathbf{X}_{\nu_l} \boldsymbol{\beta}_{\nu_l} \right)^T \left(\mathbf{y}_{\nu_l} - \mathbf{X}_{\nu_l} \boldsymbol{\beta}_{\nu_l} \right) + \lambda \sum_j \sum_{\nu} d_j^{-1} \left(w_{\nu} \left\| \boldsymbol{\beta}_{\mathbf{G}_{\nu}}^j \right\|_2 \right)^2,$$
(14)

subject to $\sum_{j}\sum_{\nu}d_{j,\nu} = 1$ and $d_{j,\nu} \ge 0$.



FIGURE 2: (a) True tree structure partitioned by Z_1 and Z_2 . (b) Pattern of regression coefficients within each leaf node of the tree.

 $\mathbf{D} = \text{Diag}(\sum_{\nu} d_{1,\nu}, \dots, \sum_{\nu} d_{j,\nu}, \dots, \sum_{\nu} d_{J,\nu}).$ (14) can be solved using an iterative algorithm that alternates between solving **B** and **D**. Given **D**, **B** can be solved analytically, i.e.,

$$\boldsymbol{\beta}_{\nu_l} = \left(\mathbf{X}_{\nu_l}^{\mathrm{T}} \mathbf{X}_{\nu_l} + \lambda \mathbf{D} \right)^{-1} \mathbf{X}_{\nu_l}^{\mathrm{T}} \mathbf{y}_{\nu_l}.$$
 (15)

Given **B**, $d_{i\nu}$ can be obtained using the result in Proposition 2. The tuning parameter λ is selected based on a line search from 0 to Λ , where 0 is the smallest possible value and Λ is the largest value that no further improvement can be made by increasing this value. The optimal λ is chosen by minimizing the mean squared prediction error (MSPE) on the validation set or via a cross-validation scheme. We summarize the above estimation procedure for the hierML model in Algorithm 1. Note that Algorithm 1 provides the model estimation method for hierML at each step of the recursive partitioning process for growing the tree. Next, we present the steps involved in the entire process, which compose the algorithm for constructing the PHMR model. Input to the algorithm includes a training set and a validation set on the indoor variables X, outdoor variable Z, and response variable Y. At each step, the algorithm needs to select an outdoor variable, Z_i , to split the samples belonging to a leaf node v_s into a left and a right branch (a.k.a. the left

child and right child node), $\{Z_j \le z_j\}$ and $\{Z_j > z_j\}$, respectively. z_j is the splitting point. To choose the optimal outdoor variable and the associated splitting point, our algorithm goes through each outdoor variable included in the dataset and each candidate splitting point and chooses the ones to be such that the empirical risk reduction evaluated on the validation set is the largest. The empirical risk reduction is computed as follows:

$$\Delta \widehat{R}_{\text{val}}^{j} = \widehat{R}_{\text{val}} \left(\widehat{\boldsymbol{\beta}}_{tr}^{\{v_s\}} \right) - \widehat{R}_{\text{val}} \left(\widehat{\boldsymbol{\beta}}_{tr}^{\{Z_j \leq z_j\}} \right) - \widehat{R}_{\text{val}} \left(\widehat{\boldsymbol{\beta}}_{tr}^{\{Z_j > z_j\}} \right),$$
(16)

where $\widehat{\boldsymbol{\beta}}_{tr}^{\{v_s\}}$ contains regression coefficients for leaf node v_s estimated using the training set, which has been obtained in the previous step. $\widehat{\boldsymbol{\beta}}_{tr}^{\{Z_j \leq z_j\}}$ and $\widehat{\boldsymbol{\beta}}_{tr}^{\{Z_j > z_j\}}$ contain coefficients of the regression models at two child nodes of v_s , respectively. The child nodes are obtained by using Z_j to split v_s according to the splitting point z_j . $\widehat{\boldsymbol{\beta}}_{tr}^{\{Z_j \leq z_j\}}$ and $\widehat{\boldsymbol{\beta}}_{tr}^{\{Z_j > z_j\}}$ are estimated by the hierML model or two lasso models for computational ease. A commonly used empirical risk function $\widehat{R}_{val}(\bullet)$ is the sum of squared prediction errors over all the samples in the validation set. Suppose that the outdoor variable Z_{j^*} and associated

Input: A set of *L* leaf nodes, v_l 's, and their associated data (training and validation) of indoor variables \mathbf{X}_{v_l} and response variable \mathbf{y}_{v_l} . A set of weights w_v for each node v of the tree **Initialize:** Initial value of **D** as $\mathbf{D}^{(0)}$ and start from iteration t = 1. Iterate for all the values of λ : for $\lambda \in (0, \Lambda]$ do while $t < t_{max}$ and $\|\hat{\mathbf{B}}^{(t+1)} - \hat{\mathbf{B}}^{(t)}\|_F > \epsilon$ do Update $\hat{\mathbf{\beta}}_{v_1}^{(t+1)}$ by solving $\hat{\mathbf{\beta}}_{v_1}^{(t+1)} = (\mathbf{X}_{v_l}^T \mathbf{X}_{v_l} + \lambda \mathbf{D}^{(t)})^{-1} \mathbf{X}_{v_l}^T \mathbf{y}_{v_l}$ Update $\hat{\mathbf{D}}^{(t+1)}$ using Proposition 2 end while Calculate $MSPE_{val}(\lambda)$ on validation set using $\hat{\mathbf{B}}(\lambda)$ end for Output $\hat{\mathbf{B}}(\lambda)$ with the smallest $MSPE_{val}(\lambda)$ Output: Regression coefficients' estimation for each leaf node $\hat{\mathbf{B}} = (\mathbf{\beta}_{v_1}, \dots, \mathbf{\beta}_{v_L})$.

ALGORITHM 1: Algorithm for hierML model estimation.

splitting point z_{j^*} are found to be leading to the highest empirical risk reduction, then our algorithm will split the leaf node using (Z_{j^*}, z_{j^*}) . This creates two new leaf nodes corresponding to $\{Z_{j^*} \le z_{j^*}\}$ and $\{Z_{j^*} > z_{j^*}\}$, respectively. Algorithm 1 will be used to refit the hierML model for all the leaf nodes. This completes the partitioning at one leaf node. The partitioning will be recursively performed on each newly generated leaf node until no reduction in the empirical risk function can be found and the algorithm stops [17]. We summarize the above-described recursive partitioning tree-growing process in Algorithm **2**.

4. Simulation Studies

The efficacy of the Partitioned Hierarchical Multitask Regression (PHMR) model was rigorously evaluated through a series of simulation studies designed to reflect various complexities encountered in real-world data. This section details the data generation process, compares PHMR against existing methods, and provides an in-depth analysis of the results.

4.1. Data Generation. Data for the simulation studies were generated to mimic a real-world scenario with one indoor variable (*I*), five outdoor variables (Z_1 to Z_5), and a response variable (*Y*). Z_1 and Z_2 were the true partitioning variables, dividing the outdoor variable space into distinct subdivisions, while Z_3 to Z_5 were included as noise. The 75 input variables followed a multivariate normal distribution with a correlation structure in $\Sigma_{75\times75}$ set to $\sigma_{ij} = 0.5^{|i-j|}$, $i, j = 1, \cdots$, 75. This setup aimed to test the robustness and adaptability of PHMR in a controlled yet complex environment.

PHMR's performance was compared against Single Partition Regression (SPR), model-based recursive partitioning (MOB), and Generalized, Unbiased, Interaction Detection and Estimation (GUIDE) [16, 17]. The selection of these methods provided a diverse range of comparison points, from traditional regression approaches to more modern partitioning techniques. The tuning parameters for all models were optimized based on the mean squared prediction error (MSPE) on a validation set, ensuring a fair and consistent evaluation framework.

4.2. *Result Analysis and Model Performance.* The mean squared prediction error (MSPE) of each method on the test set is reported in Table 1.

The results show that PHMR outperforms the other methods significantly, particularly in smaller training sizes. This superior predictive accuracy and efficiency are attributed to its hierarchical structure, facilitating information sharing across nodes and enhancing prediction capability.

The recovery rate of the true tree structure, detailed in Table 2 and visualized in Figure 2(a), further demonstrates PHMR's proficiency in revealing and representing complex data relationships. This was attributed to its hierarchical structure, which allows for information sharing across different nodes and enhances the overall prediction capability. The recovery rate of the true tree structure further demonstrated PHMR's ability to uncover and represent complex data relationships. A nuanced examination of the MSPE and Pearson correlation within each leaf node, as shown in Table 3, illustrates the model's adaptability across varying data segments. Nodes closer to the root exhibited better prediction performances, with smaller MSPEs and larger Pearson correlations, than those nearer the bottom of the tree, like nodes 8 and 9. However, even in these challenging nodes, PHMR achieved high Pearson correlations between the true and predicted responses on test data, underscoring its robustness.

In 100 independent simulations, we assessed the reconstruction frequency of the actual tree structure (Figure 2(a)) by SPR and PHMR, with outcomes and MSPE of fullyreconstructed runs detailed in Table 2. PHMR notably outperformed SPR in replicating the ground-truth structure, primarily because SPR often prematurely terminates node splitting. Further analysis of PHMR's performance across outdoor variable subdivisions (nodes 2, 4, 6, 8, and 9) showed that

Input : A training set D_{tr} and a validation set D_{val} on the indoor variable X , outdoor variable Z		
Fit a lasso model at the current node u using D to obtain the current empirical risk		
The above hold at the current hold v_s using D_{tr} to obtain the current empirical lisk		
$\hat{R}_{val}(\boldsymbol{\beta}_{trs}^{(r)})$		
Assume that there are <i>Q</i> outdoor variables,		
for $j = 1$ to Q do		
Split the data of node v_s into left and right child node, $\{Z_i \leq z_i\}$ and $\{Z_i > z_i\}$, respectively;		
Fit two lasso models in both child nodes and obtain the coefficients of the regression		
models, $\widehat{\boldsymbol{\beta}}_{tr}^{\{Z_j \leq z_j\}}$ and $\widehat{\boldsymbol{\beta}}_{tr}^{\{Z_j > z_j\}}$;		
Calculate empirical risk reduction $\Delta \widehat{R}_{val}^{j} = \widehat{R}_{val}(\widehat{\beta}_{tr}^{\{v_s\}}) - \widehat{R}_{val}(\widehat{\beta}_{tr}^{\{Z_j \le z_j\}}) - \widehat{R}_{val}(\widehat{\beta}_{tr}^{\{Z_j \le z_j\}})$.		
end for		
if $(\Delta \widehat{R}_{val}^{j} < 0 \text{ for all } j)$ then		
Stop		
else		
Choose outdoor variable Z_{j^*} and associated splitting point z_{j^*} leading to the		
largest empirical risk reduction $\Delta \hat{R}^{i}_{ad}$;		
Split the data of node ν_i into two new leaf nodes, $\{Z_{i^*} \leq z_{i^*}\}$ and $\{Z_{i^*} > z_{i^*}\}$,		
respectively:		
Apply Algorithm 1 on all leaf nodes to obtain hierML estimation:		
Re-calculate empirical risk for each leaf node on $D_{\rm ext}$ by using new estimates		
$\boldsymbol{\beta}_{val}$ from hierML:		
$\hat{\mathbf{r}}_{ij}$ =		
Select the real node with largest $\mathbf{R}_{val}(\mathbf{p}_{v_l})$.		
end if		
Output: A set of leaf nodes and fitted regression models in each leaf node.		

ALGORITHM 2: Algorithm for constructing the PHMR model.

 TABLE 1: MSPE on test data for four methods under two different training sample sizes.

	Sample size = 2000	Sample size = 1000
GUIDE	39.32	45.20
MOB	1.80	10.61
SPR	0.73	3.11
Proposed method (PHMR)	0.49	1.81

TABLE 2: Recovery rate of the true tree structure and MSPE of fully recovered run sizes.

	# of full recovery out of 100 runs		MSPE	
	Sample size: 2000	Sample size: 1000	Sample size: 2000	Sample size: 1000
SPR	69	57	0.54	1.99
Proposed method (PHMR)	94	81	0.45	0.90

nodes nearer the root had better predictions with lower MSPE and higher Pearson correlation than those at the bottom, as documented in Table 3. Despite smaller sample sizes in lower nodes, PHMR achieved high Pearson correlations, indicating its robust predictive capability across varied data segments. 4.3. Implications. The simulation studies underscore the robustness and practical applicability of PHMR in predicting complex phenomena. Its ability to accurately capture and represent underlying data structures makes it a valuable tool for various domains. Moreover, the insights gained from these studies provide a solid foundation for further research, suggesting avenues such as extending the hierML model to accommodate nonlinear relationships and developing more efficient algorithms for tree growth in the PHMR approach. These simulation studies provide a comprehensive understanding of PHMR's capabilities and advantages. The model's superior performance, coupled with its methodological sophistication, positions it as a promising approach for tackling intricate predictive tasks in real-world scenarios.

5. Application of PHMR in Predictive HVAC Management

5.1. Dataset Description and Experimental Settings. This study utilized a dataset from a modular house in Madrid, Spain, collected every 15 minutes over 42 days, resulting in 4,137 samples to predict indoor temperature. The dataset was collected from the SML system, a prototype dwelling equipped with cutting-edge energy-saving features (Bache and Lichman [20] and Zamora-Martinez et al. [21]). Figure 3 provides a comprehensive depiction of the sensors and actuators employed in the study. This dataset included six indoor variables like CO_2 and relative humidity, and six outdoor variables

	MSPE		Pearson's correlation (ρ)	
	Sample size: 2000	Sample size: 1000	Sample size: 2000	Sample size: 1000
Node 2	0.25 (0.01)	0.26 (0.02)	1.00 (0.00)	1.00 (0.00)
Node 4	0.26 (0.02)	0.26 (0.02)	1.00 (0.00)	1.00 (0.00)
Node 6	0.26 (0.03)	0.30 (0.24)	1.00 (0.00)	1.00 (0.00)
Node 8	2.08 (2.86)	5.00 (3.94)	0.99 (0.02)	0.96 (0.03)
Node 9	2.18 (2.76)	5.60 (3.92)	0.99 (0.02)	0.96 (0.03)

TABLE 3: MSPE and Pearson's correlation on testing data for each leaf node (mean (std) over 100 simulation runs) size.



FIGURE 3: SML system, a prototype house with advanced energy-saving technologies and actuator maps [21].

such as outdoor temperature and sun irradiance, detailed in Table 4. The data were split into a training set (50%), a validation set (25%), and a test set (25%). PHMR was applied to this dataset, creating a significant tree structure (depicted in Figure 4) that highlighted the importance of variables such as Sun_light, Sun_irradiance, and Outdoor_temp in predicting indoor temperature variations [17]. This section elaborates on the dataset's specifics, the variable importance as discovered by PHMR, and the model's implementation details.

5.2. Model Performance and Key Findings. The model's performance was evaluated against Single Partition Regression (SPR), with PHMR demonstrating a 37% improvement in prediction accuracy, evidenced by a lower mean squared prediction error (MSPE) of 4.57 compared to SPR's 7.27. This performance is depicted in a scatter plot (Figure 5) illustrating the relationship between predicted and actual temperatures. Key findings from this application include the crucial role of sunlight-related variables in temperature prediction and the model's ability to fit different outdoor temperature bins, enabling the HVAC system to adjust its activation/deactivation strategy efficiently. This subsection details the performance metrics used, the comparative analysis with other models, and the critical insights gained from applying PHMR to the real-world dataset.

5.3. Practical Implications. The application of PHMR in predictive HVAC management has underscored its substantial

	Variable abbreviation (unit)	Physical meaning
	CO ₂ _dining (ppm by volume)	CO ₂ in the kitchen area
	CO ₂ _living (ppm by volume)	CO ₂ in the living room area
In do on vanishlas	RH_dining (%)	Relative humidity in the kitchen area
Indoor variables	RH_living (%)	Relative humidity in the living room area
	Light_dining (Lux)	Lighting in the kitchen area
	Light_living (Lux)	Lighting in the living room area
	Outdoor_temp (°C)	Outdoor temperature
	Outdoor_humidity(%)	Outdoor relative humidity
Outdoor variables	Rain	The proportion of the last 15 minutes where rain was detected (a value in range (0, 1))
	Wind speed (m/s)	Wind speed
	Sun_light (Lux)	Sunlight
	Sun_irradiance (W/m ²)	Sun irradiance
Response variable	Indoor temperature (°C)	Indoor temperature

TABLE 4: Abbreviation and physical meanings of indoor, outdoor, and response variables in the application case study.



FIGURE 4: Tree structure found by PHMR.



FIGURE 5: Predicted vs. true response variable (indoor temperature) on the test set by PHMR with Pearson's correlation = 0.8.

potential for enhancing building energy efficiency. By providing accurate indoor temperature forecasts, PHMR allows HVAC systems to operate more strategically, significantly reducing energy consumption while maintaining comfort levels. This achievement not only highlights the practical utility of the model but also its contribution to advancing sustainable energy practices in building management. The robustness and adaptability of PHMR, as demonstrated by its superior performance over conventional models like SPR, suggest its wide applicability and potential to revolutionize various predictive tasks within the energy management domain.

In light of the discussions on the PHMR model's capability, it is pertinent to highlight its adaptability and precision in processing diverse datasets. The model's design allows it to seamlessly accommodate a broad spectrum of variables, including those indicative of occupancy levels, such as CO₂ concentrations. While the current case study's findings showed CO₂ levels in dining and living areas as not significantly influencing the model's predictions relative to other variables, this should not be construed as diminishing the potential relevance of CO₂ in different contexts or datasets. The PHMR model's robust architecture is well-equipped to handle controlled and uncontrolled variables, demonstrating its applicability across a wide range of building types and environmental conditions. This ensures the model's utility in capturing the nuanced dynamics of indoor environments, reaffirming its broad applicability and value in predictive HVAC management and beyond.

As we look to the future, this study opens new avenues for research and development. Enhancing PHMR to accommodate nonlinear relationships could lead to even more precise predictions in complex scenarios, and optimizing the algorithms for tree growth and model selection could improve its computational efficiency and scalability. Furthermore, integrating PHMR with other building management systems could provide a comprehensive approach to intelligent building operations. Collaborative efforts with industry practitioners will be crucial in validating the model's effectiveness and exploring its full potential. Ultimately, this research not only validates PHMR's effectiveness in a real-world scenario but also sets the stage for its broader application in creating energy-efficient and intelligent building systems.

6. Conclusion

This study introduced the PHMR model, a significant advancement in predictive modeling, tailored to optimize building energy management. PHMR's innovative integration of recursive partitioning for outdoor variables with a hierarchical machine learning (hierML) model for indoor variables at each partitioning node has demonstrated its strength. Notably, the model's ability to incorporate multilevel hierarchical similarity structures into the joint model fitting process has led to improved prediction accuracy and robust performance, outperforming traditional methods in both simulated and real-world datasets. The successful application of PHMR in accurately predicting building energy usage underscores its potential as a powerful tool in the field, opening new avenues for intelligent, data-driven decision-making in building management systems.

6.1. Limitations and Future Directions. Despite its promising capabilities, PHMR comes with limitations that future research should aim to address. The current model relies on linear relationships within the hierarchical structure, potentially limiting its ability to capture more intricate, nonlinear interactions prevalent in complex data. Additionally, the scalability and computational efficiency, particularly concerning larger datasets and real-time applications, require enhancement to make PHMR more accessible and practical for broader usage. Future developments might include extending the hierML model to encompass nonlinear relationships, thereby broadening the model's applicability and accuracy. Enhancements in tree growth algorithms and model selection processes are also critical to improving PHMR's computational efficiency and scalability. These improvements are not just enhancements; they are necessary steps to evolve PHMR into a more comprehensive tool for various predictive modeling applications.

6.2. Broader Applications beyond Building Energy Prediction. Beyond building energy management, the PHMR model holds the potential for impactful applications across diverse industries. In healthcare, leveraging PHMR could lead to breakthroughs in predictive diagnostics and personalized treatment plans, providing a more nuanced understanding of patient data. The financial sector could benefit from the model's predictive accuracy in areas like risk assessment and market trend analysis, making more informed and strategic decisions. Manufacturing and supply chain operations can harness PHMR to predict maintenance needs and optimize production processes, thereby enhancing efficiency and reducing operational costs. Environmental sciences could utilize the model for more accurate climate modeling and effective pollution control strategies, contributing to conservation and sustainability efforts. These various applications not only demonstrate PHMR's versatility but also emphasize its potential to drive significant advancements in numerous fields, making it a valuable tool for researchers and practitioners alike.

Nomenclature

Acronyms

- PHMR: Partitioned Hierarchical Multitask Regression—a predictive modeling approach used for building energy prediction
- HVAC: Heating, Ventilation, and Air Conditioning—a technology of indoor environmental comfort
- hierML: Hierarchical machine learning—a machine learning approach that considers hierarchical structures within data
- RPR: Recursive partition regression—a regression method that involves partitioning data recursively
- SPR: Single Partition Regression—a regression method involving a single partition of data
- MOB: Model-based recursive partitioning—a statistical method for recursive partitioning using model-based criteria
- GUIDE: Generalized, Unbiased, Interaction Detection and Estimation—a method for detecting and estimating interactions in nonlinear models
- MSPE: Mean squared prediction error—a measure of prediction accuracy in statistical models.

Parameters and Variables

- *I*: Indoor variable—represents an indoor measurement or condition in the study.
- Z_1 to Z_5 : Outdoor variables—variables representing outdoor conditions, where Z_1 and Z_2 are specifically the true partitioning variables
- *Y*: Response variable—typically represents building energy consumption or indoor temperature in the study
- β_{v_l} : Regression coefficients—coefficients for the leaf node v_l in the regression model
- $\boldsymbol{\varepsilon}_{v_l}$: Error term—represents the error term for the leaf node v_l in the model
- λ : Regularization parameter—used in the PHMR model to control the complexity of the model
- σ_{ij} : Correlation matrix elements—elements of the correlation matrix for the input variables.

Mathematical Notations

- $\Sigma_{75 \times 75}$: Correlation matrix—a 75 × 75 matrix representing the correlations between the 75 input variables
- *B*: Coefficient vector—a vector of regression coefficients across all leaf nodes in the PHMR model
- G_{v} : Set of leaf nodes—represents the set of leaf nodes that grow from node v in the tree structure of the model.

Data Availability

The SML2010 data used to support the findings of this study have been deposited in the UCI repository. The hyperlink is available as follows: https://archive.ics.uci.edu/dataset/274/ sml2010.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was supported by the Yonsei University Research Fund of 2022-22-0295.

Supplementary Materials

Accompanying this manuscript is a supplementary document that contains in-depth mathematical proofs for "Proof of Proposition 1" and "Proof of Proposition 2" essential for a thorough understanding of the foundational principles behind the "Partitioned Hierarchical Multitask Regression (PHMR)" model, as introduced in our study. (Supplementary Materials)

References

- US Department of Energy, "About building technologies office. Office of Energy Efficiency & Renewable Energy," n.d., https://www.energy.gov/eere/buildings/about-buildingtechnologies-office.
- [2] Y. Himeur, K. Ghanem, A. Alsalemi, F. Bensaali, and A. Amira, "Artificial intelligence based anomaly detection of energy consumption in buildings: a review, current trends and new perspectives," *Applied Energy*, vol. 287, article 116601, 2021.
- [3] Y. Himeur, M. Elnour, F. Fadli et al., "AI-big data analytics for building automation and management systems: a survey, actual challenges and future perspectives," *Artificial Intelli*gence Review, vol. 56, no. 6, pp. 4929–5021, 2023.
- [4] Z. Deng, Y. Chen, J. Yang, and F. Causone, "AutoBPS: a tool for urban building energy modeling to support energy efficiency improvement at city-scale," *Energy and Buildings*, vol. 282, article 112794, 2023.
- [5] M. Elnour, Y. Himeur, F. Fadli et al., "Neural network-based model predictive control system for optimizing building automation and management systems of sports facilities," *Applied Energy*, vol. 318, article 119153, 2022.
- [6] J. M. Han, S. Lim, A. Malkawi et al., "Data-informed building energy management (DiBEM) towards ultra-low energy buildings," *Energy and Buildings*, vol. 281, article 112761, 2023.
- [7] K. Y. Chan and W. Y. Loh, "LOTUS: an algorithm for building accurate and comprehensible logistic regression trees," *Journal* of Computational and Graphical Statistics, vol. 13, no. 4, pp. 826–852, 2004.
- [8] P. Chaudhuri and W. Y. Loh, "Nonparametric estimation of conditional quantiles using quantile regression trees," *Bernoulli*, vol. 8, no. 5, pp. 561–576, 2002.
- [9] N. Landwehr, M. Hall, and E. Frank, "Logistic model trees," *Machine Learning*, vol. 59, no. 1-2, pp. 161–205, 2005.

- [10] W. Y. Loh, "Regression tress with unbiased variable selection and interaction detection," *Statistica Sinica*, vol. 12, no. 2, pp. 361–386, 2002.
- [11] W. Y. Loh, "Regression by Parts: Fitting Visually Interpretable Models with GUIDE," in *Handbook of Data Visualization*. *Springer Handbooks Comp.Statistics*, Springer, Berlin, Heidelberg, 2008.
- [12] A. Zeileis, T. Hothorn, and K. Hornik, "Model-based recursive partitioning," *Journal of Computational and Graphical Statistics*, vol. 17, no. 2, pp. 492–514, 2008.
- [13] T. Rusch and A. Zeileis, "Gaining insight with recursive partitioning of generalized linear models," *Journal of Statistical Computation and Simulation*, vol. 83, no. 7, pp. 1301–1315, 2013.
- [14] I. M. Borràs, D. Neves, and R. Gomes, "Using urban building energy modeling data to assess energy communities' potential," *Energy and Buildings*, vol. 282, article 112791, 2023.
- [15] B. Mehdizadeh Khorrami, A. Soleimani, A. Pinnarelli, G. Brusco, and P. Vizza, "Forecasting heating and cooling loads in residential buildings using machine learning: a comparative study of techniques and influential indicators," *Asian Journal of Civil Engineering*, vol. 25, pp. 1163–1177, 2024.
- [16] S. Ning, E. Byon, T. Wu, and J. Li, "A sparse partitionedregression model for nonlinear system-environment interactions," *IISE Transactions*, vol. 49, no. 8, pp. 814–826, 2017.
- [17] S. Ning, Structure-regularized partition-regression models for nonlinear system-environment interactions, [Ph.D. thesis], Arizona State University, 2018.
- [18] G. Obozinski, B. Taskar, and M. I. Jordan, "Joint covariate selection and joint subspace selection for multiple classification problems," *Statistics and Computing*, vol. 20, no. 2, pp. 231–252, 2010.
- [19] F. R. Bach, "Consistency of the group lasso and multiple kernel learning," *Journal of Machine Learning Research*, vol. 9, no. 6, pp. 1179–1225, 2008.
- [20] K. Bache and M. Lichman, "UCI Machine Learning Repository," 2013, http://archive.ics.uci.edu/ml.
- [21] F. Zamora-Martinez, P. Romeu, P. Botella-Rocamora, and J. Pardo, "On-line learning of indoor temperature forecasting models towards energy efficiency," *Energy and Buildings*, vol. 83, pp. 162–172, 2014.