ThermoStill: Distilling Time Series Foundation Model into Thermal Dynamics Model for HVAC Model Predictive Control

Rui Liang
The Hong Kong Polytechnic University
Hong Kong, Hong Kong
maxwell-rui.liang@connect.polyu.hk

Yang Deng*
The Hong Kong Polytechnic University
Hong Kong, Hong Kong
yang2.deng@connect.polyu.hk

Yaohui Liu
The Hong Kong Polytechnic University
Hong Kong, Hong Kong
yaohui.liu@connect.polyu.hk

Dafang Zhao
The University of Osaka
Osaka, Japan
zhao.dafang@ist.osaka-u.ac.jp

Ozan Baris Mulayim
Carnegie Mellon University
Pittsburgh, USA
omulayim@andrew.cmu.edu

Ittetsu Taniguchi
The University of Osaka
Osaka, Japan
i-tanigu@ist.osaka-u.ac.jp

Dan Wang
The Hong Kong University of Science and Technology
Hong Kong, Hong Kong
wangdan@ust.hk

Abstract

Model predictive control (MPC) is a widely adopted paradigm for building HVAC operation optimization, and its effectiveness critically depends on an accurate thermal dynamics model. Although numerous modeling approaches have been proposed to improve prediction performance, they remain fundamentally constrained by limited knowledge coverage. The main reason is that learned dynamics are often bounded by the scope of available data, making it difficult to support diverse operating conditions. Recent advances in large pre-trained foundation models offer a new perspective by providing ready-made knowledge learned from large-scale corpora. In this work, we revisit thermal dynamics modeling from the viewpoint of extracting useful knowledge from Time Series Foundation Models (TSFMs) to support HVAC control.

To this end, we propose ThermoStill, a novel approach to leverage Knowledge Distillation (KD) to transfer rich temporal representations from existing pre-trained TSFMs into a physically consistent and tractable Resistance-Capacitance (RC) model. Recognizing that TSFMs possess divergent inductive biases and distinct pre-training priors, ThermoStill employs a multi-teacher framework to harness and integrate their complementary knowledge. Evaluated on hundreds of real buildings, ThermoStill achieves an accuracy improvement of 23.9% on thermal dynamics prediction. For HVAC control, we evaluate ThermoStill on the BOPTEST framework and conduct a total of 720 hours of closed-loop control in a real meeting room under normal occupancy and operation conditions. The results show that ThermoStill saves 29.2% energy consumption over baseline controllers while maintaining acceptable occupant comfort. Our code is available at the repository: https://github.com/Dylan0211/ThermoStill-release.

*Corresponding author.

E-Energy '26, Banff, AB, Canada

ACM ISBN 979-8-4007-2011-6/26/06
https://doi.org/10.1145/3744255.3811745

CCS Concepts

• **Applied computing → Physical sciences and engineering**.

Keywords

Energy Efficiency, HVAC Controls, Reinforcement Learning, Knowledge Distillation, Time Series Foundation Model

ACM Reference Format:
Rui Liang, Yang Deng, Yaohui Liu, Dafang Zhao, Ozan Baris Mulayim, Ittetsu Taniguchi, and Dan Wang. 2026. ThermoStill: Distilling Time Series Foundation Model into Thermal Dynamics Model for HVAC Model Predictive Control. In *The 17th ACM International Conference on Future and Sustainable Energy Systems (E-Energy '26), June 22–25, 2026, Banff, AB, Canada.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3744255.3811745

1 Introduction

The worldwide building sector annually accounts for 30% of global final energy consumption, and 45% of the building energy consumption is directly related to heating, ventilation, and air conditioning (HVAC) systems, according to the International Energy Agency (IEA) report [23]. Optimizing HVAC operations is not merely an economic concern but a critical imperative for global decarbonization [13, 27, 62]. To address this, a variety of control paradigms have been studied to enhance HVAC efficiency, ranging from rule-based (e.g., PIDs) to sophisticated black-box methods. Among these, Model Predictive Control (MPC) is a widely applied framework due to its solid theoretical foundation and dynamic optimization capabilities [10]. MPC has been extensively validated to optimally manage the trade-off between energy cost and comfort in both simulations and real-world deployments [6, 19, 79].

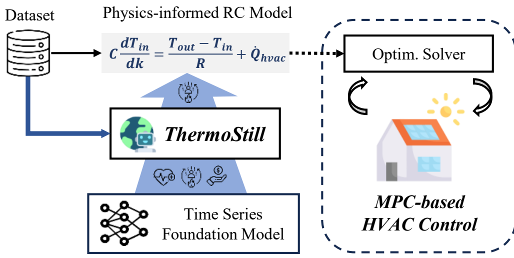

Figure 1: The proposed ThermoStill distills knowledge from TSFMs into a physics-informed RC model.

At the core of MPC lies the system dynamics model, i.e., the building thermal dynamics model in the HVAC context. The performance of MPC critically depends on the predictive accuracy of this model, as it directly guides control optimization [45]. The model is typically constructed using target building data, either through physics-based formulations or data-driven methods, and they are often limited by the expressive capacity of the model or by the amount and diversity of available training data. To address this, prior studies have explored incorporating external knowledge into building thermal modeling, for example, by leveraging information from other building environments through meta-learning or transfer learning techniques [7, 29, 53, 65]. Intrinsically, these approaches aim to extract reusable thermal patterns from auxiliary data sources and adapt them to a target building. While these approaches have shown some effectiveness, they face inherent limitations: They rely on task-specific or customized source models that require non-trivial effort to explicitly design and train; moreover, the scope of transferable knowledge is constrained by the diversity and scale of the available source data.

In contrast, recent advances in large pre-trained Foundation Models (FMs) offer an alternative way to introduce external knowledge into modeling pipelines [4]. As a subcategory of FM for the time series domain, Time Series Foundation Model (TSFM) aligns naturally with building thermal dynamics. This alignment arises not only from the temporal dependencies in thermal states (e.g., seasonal and diurnal patterns), but also from the fact that mainstream TSFMs are pretrained on datasets including energy or building related time series [1, 9, 57]. In practice, current TSFMs such as Chronos [1] and TimesFM [9] (from Amazon and Google) have demonstrated remarkable zero-shot forecasting capabilities across diverse downstream applications [18, 28, 37, 43]. This motivated recent studies of applying TSFMs to the field of energy management, such as building energy consumption [25, 26, 35] and solar irradiance [30] forecasting. Beyond forecasting, recent works also explored the use of TSFMs in control-oriented tasks, such as network congestion control [37] and traffic flow optimization [5].

In this work, we systematically investigate applying TSFMs to contribute to HVAC MPC, specifically, the thermal dynamics modeling part. While recent works have explored TSFMs in thermal dynamics-related tasks, existing efforts largely remain at the level of forecasting [48, 49] and do not explicitly address the requirements of control-oriented applications. Incorporating TSFMs into MPC raises two fundamental challenges. First, MPC requires tractable models to ensure reliable optimization, yet TSFMs are inherently black-box and highly nonlinear. Second, TSFMs do not explicitly encode physical constraints, which may lead to predictions that violate basic thermodynamic principles. To address these challenges, we leverage Knowledge Distillation (KD) [20], a popular technique for transferring knowledge from a complex teacher model to a compact student model, commonly applied to deploy large models in real-world systems [42]. Based on this framework, our idea is to distill a TSFM into a physics-informed Resistance-Capacitance (RC) model (Figure 1), where TSFMs serve as teachers that provide rich temporal representations, and the RC model serves as a physically consistent and tractable student. Furthermore, to effectively leverage the complementary strengths of multiple TSFMs, we employ a practical Multi-Teacher Knowledge Distillation (MTKD) framework in our context, and propose ThermoStill. Rather than relying on a fixed teacher configuration, ThermoStill adaptively combines knowledge from different TSFMs, enabling the student model to extract the most relevant temporal information for a given building.

We evaluate ThermoStill through two aspects: i) performance of thermal dynamics modeling, and ii) control performance. For i), ThermoStill is evaluated on hundreds of real buildings. Results indicate that ThermoStill enhances thermal states prediction accuracy of the RC model by 23.9% and outperforms knowledge distillation baseline methods by up to 19.0%. For ii), we conduct both a simulation study using BOPTEST and a month-long real-world deployment. The experiments demonstrate that ThermoStill effectively saves 29.2% energy consumption over baseline controllers while maintaining acceptable occupant comfort. The contributions of this paper are summarized as follows:

- To the best of our knowledge, we conduct the first study on applying TSFMs for MPC-based building HVAC control. We identify two critical challenges of TSFM integration and leverage the KD framework to effectively address them.
- We undertake empirical studies to investigate how KD performs when distilling TSFMs into physics-informed RC models. The insufficiency of knowledge from a single TSFM is observed, thus motivating us to propose ThermoStill, which adaptively distills knowledge from multiple TSFMs.
- We validate the performance of ThermoStill through both a simulation study and a one-month real-world deployment, demonstrating superior energy efficiency and improved thermal comfort in practical settings.

2 Background and Related Work

In this section, we introduce background and related work on key concepts of this study. Related works are summarized in Table 1.

2.1 MPC-based HVAC control and Thermal Dynamics Model

MPC is a constrained control strategy that calculates the optimal control inputs by minimizing a given objective function over a finite prediction horizon [10]. A general MPC formulation[1] can be

[1]We omit constraints on the optimization problem here for a simplified illustration. Detailed MPC formulations are presented in experiments.

Table 1: Summary of studies on thermal dynamics modeling. Our work fills the gap of leveraging general temporal knowledge to obtain thermal dynamics models for control.

Modeling with Target Building Data	Modeling with External Knowledge	
Physics-based: [16, 59], Grey-box: [12, 46], Data-driven: [64, 77, 78]	Cross-building knowledge	Meta-learning: [65], Transfer-learning: [7, 29, 53]
	General temporal knowledge (TSFM)	[48, 49] (for forecasting), THERMOSTILL (for control)

represented as follows:

$$\min_{u_1,\ldots,u_H} \quad \sum_{k=1}^{H} J(x_k, u_k) \tag{1a}$$

$$\text{s.t.} \quad x_{k+1} = Ax_k + B_u u_k + B_d d_k, \forall k \tag{1b}$$

Here, x_k, u_k, and d_k respectively denote the thermal states (e.g., indoor temperature), control inputs (e.g., HVAC power), and exogenous disturbances (e.g., outdoor temperature, solar irradiance) at time step k, respectively. $J(\cdot)$ is the objective function for MPC and H denotes the optimization horizon. A key component within the MPC formulation is the thermal dynamics model presented in Eq.(1b). This state-space representation predicts future thermal states, thereby guiding the optimization of HVAC control. A, B_u, and B_d denote building-specific coefficients.

Table 1 summarizes existing approaches to thermal dynamics modeling. Traditional methods rely primarily on information derived directly from the target building, such as historical data or physical specifications. Depending on the specific technique, these are categorized into physics-based [16, 59], grey-box [12, 46], or data-driven [64, 77, 78] methods. To further enhance modeling accuracy, recent research has sought to incorporate external knowledge as complementary information. One prominent direction involves transferring knowledge from other buildings through meta-learning [65] and transfer learning approaches [7, 29, 53]. Alternatively, recognizing that thermal dynamics inherently exhibit strong temporal dependencies, recent studies [48, 49] have utilized general-purpose TSFMs to encode generalized temporal knowledge for accurate forecasting. Motivated by these advancements, this paper aims to move beyond pure forecasting and investigate the applicability of TSFMs in closed-loop HVAC control.

2.2 Time Series Foundation Model

Recent progress in self-supervised pre-training has led to the development of large-scale, general-purpose models for time series analysis, known as Time Series Foundation Models (TSFMs) [36]. A TSFM is a deep learning model pre-trained on large-scale multi-source datasets, allowing it to be applied directly to various downstream tasks in a zero-shot setting. Formally, given a pre-trained TSFM f^t with fixed parameters and a historical input time series of length L with C variables, denoted as $\mathbf{X}_{k-L:k} \in \mathbb{R}^{L\times C}$, the TSFM predicts $\hat{\mathbf{Y}}_{k+1:k+H} = f^t(\mathbf{X}_{k-L:k})$ for the future H steps, which accurately approximates the ground truth $\mathbf{Y}_{k+1:k+H} \in \mathbb{R}^{H\times C}$ without task-specific training.

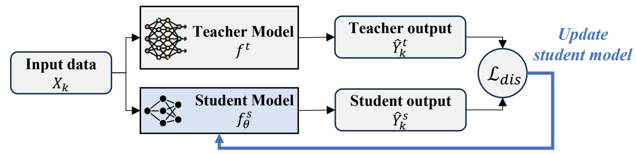

Figure 2: A general knowledge distillation framework.

Consequently, TSFMs have demonstrated remarkable performance across diverse fields, including healthcare [22, 28, 50], finance [43, 52, 80], traffic [5, 37], and energy [11, 18, 44, 60]. In the context of building energy management, recent studies have utilized TSFMs to forecast both endogenous system variables, such as energy consumption [25, 26, 35] and indoor thermal states [48, 49], as well as exogenous disturbances like outdoor temperature and solar irradiance [30]. These findings indicate that TSFMs can model complex building dynamics with superior accuracy and reduced data dependency. Furthermore, in line with known scaling laws [69], TSFMs are expected to achieve continuous performance gains as more datasets become available for pre-training. This growing potential motivates us to look beyond pure forecasting and explore a deeper integration of TSFMs within the building energy management pipeline. Specifically, we investigate integration of TSFMs into control frameworks (e.g., MPC) to effectively translate their generalized temporal knowledge into optimized control actions.

2.3 Knowledge Distillation

Knowledge Distillation (KD) is a paradigm for transferring knowledge from a highly capable yet complex teacher model to a compact student model. The primary objective is to obtain a lightweight surrogate that approximates the teacher's predictive capability, thereby enabling efficient deployment. A standard KD process is shown in Figure 2. Formally, given a transfer set $\mathcal{D}_{\text{train}} = \{(\mathbf{X}_k, \mathbf{Y}_k)\}_{k=1}^{N}$, a pre-trained teacher model f^t, and a student model f_θ^s parameterized by θ, the teacher prediction $\hat{\mathbf{Y}}_k^t = f^t(\mathbf{X}_k)$ is considered as a supervision signal for the student prediction $\hat{\mathbf{Y}}_k^s = f_\theta^s(\mathbf{X}_k)$. The student parameter θ is then optimized by minimizing the objective:

$$\mathcal{L}_{\text{KD}}(\theta) = \mathcal{L}_{gt} + \lambda \cdot \mathcal{L}_{dis}, \tag{2}$$

where $\mathcal{L}_{gt} = \|\mathbf{Y}_k - \hat{\mathbf{Y}}_k^s\|_2^2$ is the ground-truth loss and $\mathcal{L}_{dis} = \|\hat{\mathbf{Y}}_k^t - \hat{\mathbf{Y}}_k^s\|_2^2$ is the distillation loss. Here, $\mathcal{L}_{dis}$ drives the transfer of teacher knowledge to improve the student's predictive capability, whereas $\mathcal{L}_{gt}$ anchors the student to the ground truth, mitigating the risk of inheriting teacher-induced biases. λ is a weighting factor to control the influence of teacher.

Recently, KD has become a prevalent technique for deploying FMs, primarily due to its dual advantages of model compression and knowledge transfer [42]. First, KD enables the deployment of computationally intensive FMs in resource-constrained environments, such as IoT systems [33] or latency-sensitive applications like real-time control [39]. Second, it facilitates the transfer of generalized knowledge from FMs to task-specific models in domains such as climate [17] and healthcare [34, 41], thereby enhancing their performance. These capabilities motivate us to adopt a KD framework to integrate TSFMs into MPC-based HVAC control.

3 Preliminary Study and proposed approach

Although prior studies [48, 49] have successfully applied TSFMs as thermal dynamics models for indoor temperature prediction, we identify two fundamental challenges of integrating them into MPC-based HVAC control. First, MPC requires an explicit and tractable thermal dynamics model for efficient and reliable optimization, while TSFMs are black-box and highly nonlinear, making their direct embedding into the MPC optimization loop impractical. Second, TSFMs do not inherently encode physical laws. Consequently, they may generate predictions that violate basic thermodynamic principles, undermining the reliability required for safe control.

We consider Knowledge Distillation (KD) as a potential solution to these challenges. Specifically, we propose to distill knowledge from a TSFM (teacher) into a physics-informed Resistance-Capacitance (RC) model (student), yielding two major advantages for MPC. First, the RC model serves as a tractable surrogate, enabling direct integration into MPC optimization. Second, the thermodynamic formulation of the RC model naturally enforces physical consistency and extrapolation to unseen data regimes, thereby enhancing control reliability. Building on this idea, we first conduct a preliminary study to empirically investigate applicability of the KD framework and then introduce the proposed approach.

3.1 Preliminary Study

We perform two experiments to study the applicability of KD. In the first experiment, we investigate whether TSFMs offer reasonably accurate predictions on building thermal dynamics to serve as high quality teachers. In the second experiment, we study whether knowledge distilled from TSFMs can improve the student RC model. The experimental setup is presented below.

Setup. Following prior measurement studies [47, 49], we use a subset of the public *ecobee* [40] dataset, which contains hourly thermal data from 841 real-world buildings. For this preliminary study, we specifically selected 74 buildings with no missing values. Each building provides 28 days of training data and 4 days of testing data during the cooling season. Two product-level TSFMs, i.e., Chronos [1] and TimesFM [9], and a first-order RC model[2] are utilized for experiments. All models perform a rolling one-step-ahead forecasting using a sliding 168-sample historical window.

3.1.1 *Performance of TSFMs in thermal dynamics modeling.*

To verify whether TSFMs can provide high-quality supervision signals, we compare them in a zero-shot setting against a customized RC model trained specifically on the target building data across 74 buildings. Remarkably, as shown in Figure 3, TSFMs significantly outperform the customized RC model in overall accuracy and accurately capture complex diurnal temperature patterns. These results validate the superior predictive capability of TSFMs, justifying their role as teachers in our distillation framework.

3.1.2 *Performance of KD with TSFMs.*

Given the strong predictive power of TSFMs, we next investigate whether a distilled RC model can outperform a customized RC model solely trained on the data without KD. We employ the KD objective in Eq.(2) to distill a student RC model from each teacher TSFM on the training set and evaluate their performance on the testing set. We show results on ten selected buildings in Figure 4 for detailed analysis. Here, each scatter represents one KD experiment result, where the x-axis denotes teacher's accuracy (calculated as $1/\text{MSE}$), while the y-axis indicates student's MSE improvement relative to the customized RC model. Specifically, we observe that all distilled students show positive performance improvement (i.e., $y > 0\%$), confirming that knowledge in TSFMs consistently enhances RC model accuracy.

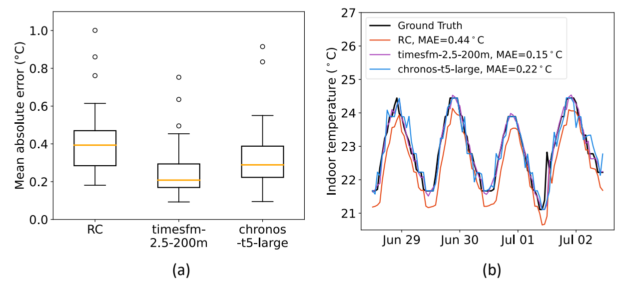

Figure 3: Performance of TSFMs and a customized RC model. Figure (a) is a box plot of overall accuracy on 74 tested buildings. Figure (b) shows a four-day example prediction.

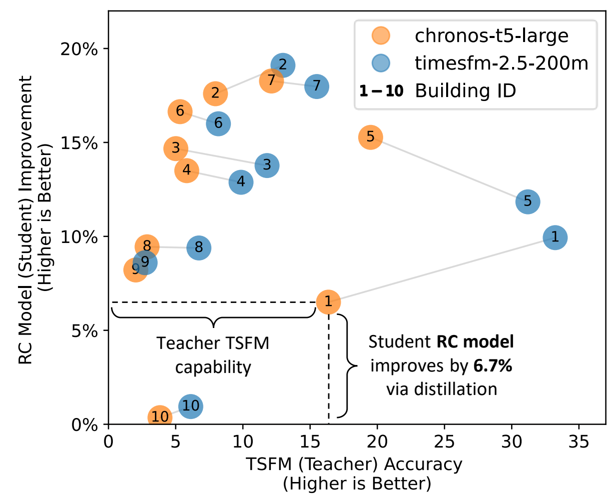

Figure 4: Performance of two TSFMs (teacher) and their distilled RC model (student) on ten buildings. Each scatter is one KD experiment result, where x-axis and y-axis respectively denotes the performance of teacher and student (higher is better). The number denotes building id and the color denotes teacher TSFM.

3.1.3 *Key insights.*

We derive two insights from experiment results in Figure 4. First, teacher superiority is building-dependent. No single TSFM consistently dominates across all buildings. For instance, while Chronos yields a superior student on Building 5, TimesFM distills a better student on Building 1. Second, and more

[2]The first-order RC model adopted can be represented as: $C\frac{dT_{in}}{dk} = \frac{T_{out}-T_{in}}{R} - \dot{Q}_{hvac}$, where T_{in} denotes the indoor temperature, T_{out} the outdoor temperature, and $\dot{Q}_{hvac}$ the controlled cooling input. R and C represent the thermal resistance and capacitance, which are learnable parameters $\theta = \{R, C\}$. k denotes the time step. While the current RC model is simplified by omitting solar irradiance, this variable will be used in the comprehensive experiments later.

importantly, teacher accuracy is not a guarantee of distillation effectiveness. An example is on Building 5, where TimesFM achieves higher predictive accuracy than Chronos but produces a strictly inferior student. These findings exemplify a commonly observed but non-trivial challenge of optimal teacher selection in KD. To address this, adopting a Multi-Teacher Knowledge Distillation (MTKD) framework has become a standard practice in the field [14, 42]. By aggregating and controlling knowledge of multiple teachers throughout the distillation process, MTKD aims at leveraging their complementary strengths to achieve substantial accuracy improvement on the distilled student. Following the mainstream paradigm that formulates MTKD as an RL problem due to its intrinsic sequential decision-making nature, we present a standard formulation of RL-based MTKD in the next section.

3.2 Proposed Approach: Multi-Teacher Knowledge Distillation via RL

In the setting of MTKD, multiple teachers are aggregated to collaboratively distill a student. Specifically, the influence of each teacher on the distillation process is controlled by an associated teacher weight [67]. We formulate the MTKD framework for distilling multiple TSFMs into an RC model as follows.

3.2.1 ***Multi-Teacher Knowledge Distillation***. Given a transfer set $\mathcal{D}_{\text{train}} = \{(\mathbf{X}_k, \mathbf{Y}_k)\}_{k=1}^{N}$. The input sequence $\mathbf{X}_k$ consists of the historical thermal state x, control inputs u, and disturbances d over a look-back window of length L, denoted as $\mathbf{X}_k = [x_t, u_t, d_t]_{t=k-L}^{k}$. The corresponding target $\mathbf{Y}_k = x_{k+1}$ is the one-step-ahead thermal state[3]. With M pre-trained TSFMs $\{f^j\}_{j=1}^{M}$ and a physics-informed RC model f_θ^s, the student f_θ^s is optimized via minimizing the following MTKD objective:

$$\mathcal{L}_{\text{MTKD}}(\theta) = \mathcal{L}_{gt} + \lambda \sum_{j=1}^{M} w_k^j \mathcal{L}_{dis}^j. \tag{3}$$

As compared to the traditional KD objective in Eq.(2), the only difference is that $\mathcal{L}_{\text{MTKD}}$ includes a weighted sum of distillation loss from each teacher TSFM, i.e., $\mathcal{L}_{dis}^j = \|\hat{\mathbf{Y}}_k^s - \hat{\mathbf{Y}}_k^j\|_2^2$, where $\hat{\mathbf{Y}}_k^j$ is the j-th teacher prediction. Here, $[w_k^1, \ldots, w_k^M]$ is a vector of teacher weights to control each distillation loss on the training sample $(\mathbf{X}_k, \mathbf{Y}_k)$. The goal is to optimize teacher weights to maximize distillation effectiveness.

3.2.2 ***Motivation for RL-based MTKD***. Existing MTKD methods mainly differ in how teacher contributions are estimated during distillation. A first family adopts fixed or heuristic weighting strategies, such as ensemble-style averaging, confidence-aware weighting, or weights computed from predefined reliability scores [63, 68, 74]. These methods are simple and effective when teacher usefulness can be approximated by teacher-side or sample-level indicators, e.g., confidence, prediction error, uncertainty, or apparent reliability. However, this assumption does not hold reliably in our setting, as validated in our preliminary study in §3.1. The indirect relation between teacher accuracy and distillation usefulness therefore motivates a learnable weighting mechanism.

[3]We set $L > 1$ to accommodate the TSFM's requirement for a relatively long temporal context, whereas the RC model only uses the last step in the input data for prediction.

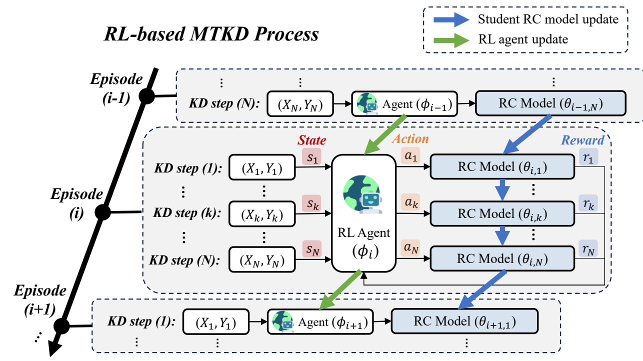

Figure 5: Paradigm of RL-based MTKD. Parameters of the RL agent (ϕ) and student RC model (θ) are jointly optimized.

A second family therefore employs learnable adaptive weighting methods, such as meta-weight networks or gating networks, to generate input-dependent teacher weights from observable training signals, such as the current sample, teacher performance, or teacher–student discrepancies [38, 55, 75]. These methods improve over fixed heuristics by learning teacher combinations in a differentiable and data-dependent manner. Nevertheless, many such methods remain largely myopic, as teacher weights are typically optimized according to current samples or immediate training feedback. In practice, the effect of a weighting decision may only become evident after subsequent student updates. A typical example is curriculum-like distillation [32]: a student at an early training stage may first benefit from simpler or more consistent supervision, even if the corresponding teacher is not the most accurate one, and later exploit more accurate but harder-to-fit teachers as training proceeds. Therefore, preferred teacher weights should depend not only on the current sample or teacher quality, but also on the student's learning stage and the expected effect of each weighting decision on subsequent training.

Based on the above discussion, teacher contributions should be optimized throughout the distillation process, rather than determined by one-time teacher selection or a static teacher combination. This naturally leads to a sequential decision-making formulation, where each weighting action affects both the current distillation loss and the subsequent state of the student model. Following prior RL-based MTKD studies [66, 67, 71, 72], we use RL to optimize teacher weights throughout the distillation process. Specifically, the RL agent interacts with student training and learns a weighting policy that accounts for both immediate feedback and accumulated effects over the distillation trajectory. The detailed problem formulation is presented in the following.

3.2.3 ***Problem Statement***. Figure 5 illustrates a standard paradigm of RL-based MTKD, where an RL agent π_ϕ parameterized by ϕ is introduced to interact with the distillation process and learn the optimal teacher-weighting policy to enhance distillation gains. At each KD step k, the agent π_ϕ observes state s_k, which encodes information from the current training sample $(\mathbf{X}_k, \mathbf{Y}_k)$, and outputs an action $a_k = [w_k^1, \ldots, w_k^M]$. The student f_θ^s is then updated via $\mathcal{L}_{\text{MTKD}}$ and its prediction accuracy is used as the reward r_k. After a whole episode of training samples, the agent π_ϕ will be updated to maximize the expected return:

$$J(\phi) = \mathbb{E}_{\pi_\phi}\left[\sum_{k=1}^{N} \gamma^{k-1} r_k\right]. \tag{4}$$

Here, the discount factor $\gamma = 1$ since student's performance on each data sample is considered of equal importance. Ideally, after multiple training episodes, the performance of the distilled RC model f_θ^s will be significantly enhanced, driven by the ability of π_ϕ in continuously providing optimal combinations of teacher knowledge.

4 Methodology

In this paper, we propose ThermoStill, a novel RL-based MTKD framework to address the problem of distilling knowledge from multiple TSFMs into a student RC model for MPC-based HVAC control. The overall framework is shown in Figure 6.

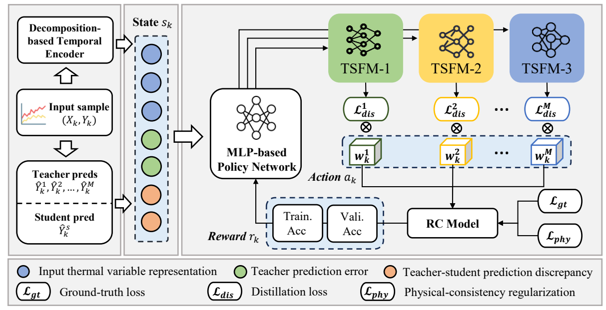

Figure 6: Overview of the ThermoStill approach.

4.1 RL Formulation

We first present the RL formulation in ThermoStill by introducing the instantiation of the state, action, and reward.

4.1.1 ***State.*** To enable the RL agent to make judicious decisions, the state s_k is designed to encompass the following information:

(1) *Input thermal variable representation.* The input sequence $\mathbf{X}_k$ implicitly encodes the current operating context of the thermal dynamics. Motivated by our preliminary findings that teacher efficacy varies across different data patterns, we employ a temporal encoder $E(\cdot)$ to extract a compact embedding $e_k = E(\mathbf{X}_k)$ to augment the state vector. The architecture of $E(\cdot)$ will be introduced later.

(2) *Teacher prediction error.* The capability of teacher TSFMs serve as important information for the agent to adjust their contributions. Therefore, the prediction error of each teacher TSFM over the current sample, i.e., $\varepsilon_k^j = \|\mathbf{Y}_k - \hat{\mathbf{Y}}_k^j\|_2^2$ is included in the state vector.

(3) *Teacher-student prediction discrepancy.* We also introduce the prediction discrepancy between the student and each teacher, i.e., $\delta_k^j = \|\hat{\mathbf{Y}}_k^s - \hat{\mathbf{Y}}_k^j\|_2^2$, into the state vector. The rationale is to discourage assigning high weights to teachers with drastically divergent predictions, as such large gaps might indicate a misalignment or excessive difficulty that could lead to negative transfer.

Combining the above components, the state vector s_k is defined in Eq. (5), where Concat$(\cdot)$ denotes the concatenation operator.

$$s_k = \text{Concat}(e_k, \varepsilon_k^1, \dots, \varepsilon_k^M, \delta_k^1, \dots, \delta_k^M) \tag{5}$$

4.1.2 ***Action.*** The action a_k is defined as a continuous weight vector $[w_k^1, \dots, w_k^M]$, where each element $w_k^j \in [0, 1]$ represents the weight assigned to the j-th teacher. By dynamically adjusting these weights, the agent can selectively leverage informative teachers while suppressing less relevant ones, thereby facilitating effective and adaptive knowledge transfer.

4.1.3 ***Reward.*** The reward function r_k is designed to quantify the effectiveness of the action a_k in improving the thermal dynamics modeling accuracy of the student RC model. It consists of two components. First, we compute the student model's prediction error on the current sample, i.e., $\varepsilon_k = \|\mathbf{Y}_k - \hat{\mathbf{Y}}_k^s\|_2^2$, to reflect the direct improvement. Second, to encourage generalization and mitigate the risk of overfitting, we evaluate the student model on a held-out validation set $\mathcal{D}_{\text{val}}$, whose samples are excluded from training. The average prediction error on this set, i.e., $\varepsilon^{\text{val}} = \mathbb{E}_{k' \in \mathcal{D}_{\text{val}}}[\|\mathbf{Y}_{k'} - \hat{\mathbf{Y}}_{k'}^s\|_2^2]$, reflects the student's generalization capability. We use k' to denote the sample index in the validation set. The final reward function is represented in Eq.(6), where η controls the trade-off between fitting the current sample and generalizing to unseen data. Here, the reward is defined by the negative loss, since a lower loss value indicates a better student.

$$r_k = -\left[\eta \varepsilon_k + (1 - \eta)\varepsilon^{\text{val}}\right] \tag{6}$$

4.2 Architectural Design

In this section, we present the architectural design for networks in ThermoStill, which includes a decomposition-based temporal encoder and an MLP-based policy network.

4.2.1 ***Decomposition-based Temporal Encoder.*** We propose a lightweight and effective decomposition-based encoder to extract latent temporal features from the input $\mathbf{X}_k$. Specifically, $\mathbf{X}_k$ is first decomposed into a trend component $\mathbf{X}_k^{\text{tre}}$ and a seasonality component $\mathbf{X}_k^{\text{sea}}$, each of which is projected via linear layers separately into latent spaces and then combined to form the final temporal feature. In the context of thermal systems, this decomposition allows the encoder to capture slow-varying temperature changes caused by building thermal inertia and recurring fluctuations induced by periodic heating/cooling cycles, thereby extracting structured temporal information. The process can be represented as:

$$\mathbf{X}_k^{\text{tre}} = \text{AvgPool}(\mathbf{X}_k),\ \mathbf{X}_k^{\text{sea}} = \mathbf{X}_k - \mathbf{X}_k^{\text{tre}}, \tag{7a}$$

$$e_k = W^{\text{tre}}\mathbf{X}_k^{\text{tre}} + W^{\text{sea}}\mathbf{X}_k^{\text{sea}}. \tag{7b}$$

Here, we follow a mainstream decomposition strategy using average pooling [51, 73]. W^{tre} and W^{sea} are learnable weight matrices for linear transformations of the trend and seasonality components.

4.2.2 ***MLP-based Policy Network.*** The policy network π_ϕ is implemented as a multi-layer perceptron (MLP) that maps the current state s_k to the action distribution. Given that the teacher weights in a_k are bounded in $[0, 1]$, we model the policy using Beta distributions, which naturally enforces the value constraint and allows the agent to flexibly adjust the contribution of each teacher individually.

Algorithm 1: ThermoStill Training Procedure.

Input: The transfer dataset $\mathcal{D}_{\text{train}}$, Pre-trained TSFMs $\{f^j\}_{j=1}^{M}$, and policy network π_ϕ.
Output: The trained student RC model f_θ^s.
/* **Pre-training of Student Model** */
Pre-train the student model f_θ^s by $\mathcal{L}_{gt}$ on all training samples $(\mathrm{X}_k, \mathrm{Y}_k) \in \mathcal{D}_{\text{train}}$;
/* **Joint Training of Student Model & Policy Network** */
for $i = 1, ..., \#Episodes$ **do**
  Initialize set $\mathcal{H}$ to store episode history ;
  for *each data sample* $(\mathrm{X}_k, \mathrm{Y}_k) \in \mathcal{D}_{\text{train}}$ **do**
    Construct state vector s_k for sample $(\mathrm{X}_k, \mathrm{Y}_k)$ via Eq.(5);
    Generate teacher weights a_k via Eq.(8,9);
    Compute $\mathcal{L}_{\text{MTKD}}$ via Eq.(3) with generated weights a_k and teacher TSFMs $\{f^j\}_{j=1}^{M}$;
    Update student RC model f_θ^s using $\mathcal{L}_{\text{MTKD}}$;
    Compute reward r_k via Eq. (6);
    Save (s_k, a_k, r_k) to the episode history $\mathcal{H}$;
  for *each* $(s_k, a_k, r_k) \in \mathcal{H}$ **do**
    Update policy network π_ϕ via PPO;

Formally, the network consists of two fully-connected layers:

$$h = \text{ReLU}(W^{\text{in}} s_k + b^{\text{in}}),$$
$$\alpha_k, \beta_k = \text{Softplus}(W^{\text{out}} h + b^{\text{out}}), \tag{8}$$

where W and b denote the learnable weight matrices and bias vectors. Note that we apply the Softplus activation in the output layer to ensure that the distribution parameters α_k and β_k remain strictly positive. The weight for each teacher is sampled from its corresponding Beta distribution:

$$w_k^j \sim \text{Beta}(\alpha_k^j, \beta_k^j), \quad j = 1, \ldots, M. \tag{9}$$

4.3 Model Training

In this section, we first describe a physical-consistency regularization to ensure the validity of physical parameters in the student RC model during distillation. Then, we introduce the RL training algorithm together with the overall training procedure.

4.3.1 ***Physical-consistency Regularization.*** The learnable parameters in RC models represent intrinsic physical properties like thermal resistance and capacitance. Direct optimization often yields physically implausible values inapplicable for practical HVAC control [2]. While we constrain these parameters within realistic bounds following prior studies [47, 54, 61], a critical structural constraint arises from discretization of RC models for forecasting. Specifically, to ensure a valid discrete-time state transition with a time interval Δk, the building's thermal time constant τ ($\tau = RC$ for a first-order RC model) must significantly exceed the sampling interval, i.e., $\tau \gg \Delta k$ [24]. To enforce this condition, we incorporate a regularization term $\mathcal{L}_{\text{phy}} = (\Delta k/\tau)^2$ into the MTKD objective in Eq.(3). Notably, the gradient magnitude scales as $|\nabla_\tau| \propto \tau^{-3}$, ensuring the regularization activates only when τ approaches Δk and remains negligible when $\tau \gg \Delta k$.

4.3.2 ***RL Training via PPO.*** The RL agent is trained via interacting with the distillation process of the student RC model, which naturally falls into the setting of online RL. To ensure stable policy learning, we adopt Proximal Policy Optimization (PPO) [56] to optimize the RL agent. PPO is a policy gradient method that constrains policy updates within a trust region through a clipped surrogate objective, effectively preventing overly aggressive policy changes.

4.3.3 ***Overall Training Process.*** Algorithm 1 illustrates the overall ThermoStill training procedure, which comprises a pre-training phase and a joint-training phase. Specifically, we first pre-train the student RC model f_θ^s using the ground-truth loss $\mathcal{L}_{gt}$ without KD for a few epochs (i.e., 20 epochs). The rationale behind this design is that the policy network requires a valid reward signal for optimization, while an untrained student model's feedback could be completely useless and even harmful for policy learning. Therefore, pre-training gives the student model a reasonable initialization with basic knowledge on the thermal dynamics modeling task, allowing the policy network to correctly assess the performance of diverse combinations of knowledge from TSFMs. Then, in the joint training phase of the policy network π_ϕ and RC model f_θ^s, for each episode, the policy network generates teacher weights, and the student RC model f_θ^s is trained via distillation. At the end of each episode, the collected (state, action, reward) triplet is utilized to optimize the agent π_ϕ via a standard PPO algorithm.

5 Evaluation on thermal dynamics modeling

In this section, we present the evaluation of ThermoStill with the following research questions:

RQ1: How does ThermoStill perform when compared with existing MTKD techniques?

RQ2: How do the key components in KD framework, i.e., student, teacher, transfer set, affect the performance of ThermoStill?

RQ3: How does each designed component in ThermoStill contribute to its performance?

5.1 Experiment Setup

Datasets. To evaluate the thermal dynamics modeling accuracy of ThermoStill, we adopt a publicly available processed subset of the large-scale *ecobee* [40] dataset, which has been utilized for benchmarking TSFMs in prior measurement studies [48, 49]. This subset comprises hourly thermal data from 841 real-world buildings across four U.S. states: California, Texas, Illinois, and New York in the year of 2017. Each building is further subsampled to have 28 days of training data and 4 days of testing data from cooling seasons.

Teacher and student models. We adopt three representative TSFMs as teacher models for this evaluation. Specifically, we select two product-level TSFMs, i.e., TimesFM [9] developed by Google and Chronos [1] developed by Amazon, as well as the largest publicly available TSFM to date, TimeMoE [57] [4]. We adopt a first-order RC model as the student for most experiments in this evaluation.

Baselines. We compare ThermoStill against four MTKD baselines: (1) U-Ensemble [68] assigns equal weights to all candidate teacher models. (2) W-Ensemble [63] is a weighted ensemble method

[4]The detailed model versions used in this study are TimesFM (google/timesfm-2.5-200m-pytorch), TimeMoE (Maple728/TimeMoE-200M), and Chronos (amazon/chronos-t5-large, 710M).

Table 2: Overall performance comparison of ThermoStill and MTKD baselines. The *ecobee* dataset is divided into four subsets based on the geographical location (state) of buildings for detailed analysis. Performance of RC models trained with ThermoStill and baselines are compared in the *Student+KD* category, where the best results are highlighted in bold.

Category	Method	California			Illinois			New York			Texas		
		MAE	MSE	CV-RMSE	MAE	MSE	CV-RMSE	MAE	MSE	CV-RMSE	MAE	MSE	CV-RMSE
Teacher	TimesFM	0.2395	0.1341	0.1135	0.2291	0.1250	0.1079	0.2330	0.1448	0.1121	0.2375	0.1338	0.1109
	TimeMoE	0.2424	0.1349	0.1144	0.2314	0.1248	0.1081	0.2394	0.1498	0.1138	0.2383	0.1327	0.1099
	Chronos	0.3183	0.2237	0.1502	0.3028	0.2106	0.1441	0.3087	0.2441	0.1506	0.3054	0.2077	0.1410
Student	RC Model	0.4754	0.3566	0.1959	0.3199	0.1908	0.1389	0.3281	0.2207	0.1450	0.3503	0.2434	0.1539
Student + KD	+ U-Ensemble	0.4132	0.2891	0.1764	0.3004	0.1777	0.1318	0.3149	0.2124	0.1394	0.3581	0.2705	0.1571
	+ W-Ensemble	0.4113	0.2872	0.1740	0.2972	0.1743	0.1306	0.3081	0.2057	0.1370	0.3278	0.2256	0.1454
	+ MMKD	0.4297	0.3121	0.1829	0.2978	0.1727	0.1307	0.3088	0.2045	0.1379	0.3198	0.2137	0.1424
	+ MTKD-RL	0.4253	0.3021	0.1790	0.2957	0.1711	0.1303	0.3084	0.2043	0.1378	0.3193	0.2131	0.1421
	+ ThermoStill	**0.3524**	**0.2318**	**0.1536**	**0.2683**	**0.1550**	**0.1211**	**0.2768**	**0.1861**	**0.1278**	**0.2967**	**0.1968**	**0.1338**

which assigns a fixed but different weight to each teacher model. (3) MMKD [75] proposes a meta-weight network to learn the optimal teacher weights at a batch-level. (4) MTKD-RL [67] employs an RL agent to adaptively decide teacher weights based on both teacher performance and teacher-student gaps.

Implementation details. For each building in the *ecobee* dataset, we consider the first 28-day data as the transfer set for RC model distillation. The follow-up 4-day data is considered as the test set for RC model evaluation. This setup follows existing measurement studies for TSFMs [48, 49]. For ThermoStill, the last 4 days in the transfer set is kept as the validation set. The length of input data L is set to 168. We adopt mean absolute error (MAE), mean squared error (MSE), and coefficient of variation of the root mean squared error (CV-RMSE) as metrics to evaluate the thermal dynamics modeling accuracy. All methods are trained using an Adam optimizer with maximum 200 epochs. The weight for the distillation loss λ is set to 0.5. The hyperparameter for reward function η is set to 0.8.

5.2 Main Results

To answer **RQ1**, we comprehensively evaluate ThermoStill on the *ecobee* dataset and compare it against state-of-the-art MTKD baselines. Table 2 summarizes the performance across buildings in four U.S. states. The "Teacher" category reports the zero-shot forecasting error of the TSFMs, while the "Student" category reflects the baseline RC model trained solely on data without distillation. Overall, ThermoStill significantly enhances the baseline RC model, reducing average prediction error by 19.0%, 23.9%, and 15.4% in terms of MAE, MSE, and CV-RMSE. Furthermore, ThermoStill consistently outperforms all four baselines, achieving 19.0%, 13.8%, 14.8%, and 13.6% MSE improvement respectively. These results highlight the effectiveness of ThermoStill in dynamically aggregating appropriate knowledge from TSFMs for enhanced student RC model performance. Furthermore, ThermoStill exhibits robust resilience against negative transfer, whereas methods assigning fixed teacher weights (e.g., U-Ensemble) remain vulnerable to inferior teacher predictions. For instance, in the New York subset, the teacher *Chronos* performs worse than even the naive student, causing U-Ensemble to provide a limited MSE improvement of 3.7%. In contrast, ThermoStill effectively filters this noise to achieve a superior improvement of 15.7%, validating the proposed RL agent's capability to distinguish and reject unreliable supervision signals.

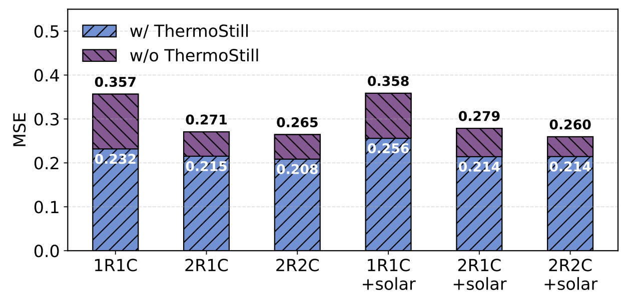

Figure 7: Performance of different RC models trained with and without ThermoStill.

5.3 Detailed Analysis

To answer **RQ2**, we further study the effectiveness of ThermoStill by varying three key components of the KD framework: 1) the architecture of the student RC model; 2) the selection of teacher TSFMs; and 3) the impact of different training set sizes.

Different students. To assess the effectiveness of ThermoStill across distinct physical representations, we extend our evaluation to two additional RC architectures with extra physical parameters, i.e., 2R1C and 2R2C (detailed in Appendix A). Furthermore, since the *ecobee* dataset includes ambient temperature as the sole exogenous disturbance, we augment the feature space by integrating auxiliary solar irradiation data from the National Solar Irradiation Dataset[5]. This setup yields three additional RC model variants that can account for solar heat gain. The results on buildings in California are in Figure 7. ThermoStill delivers substantial performance gains, achieving an average MSE improvement of 24.4% across all six RC models compared to baselines trained without ThermoStill. This confirms the effectiveness of ThermoStill across diverse student architectures. Notably, the performance enhancement is more pronounced in simpler RC architectures compared to their complex counterparts, ranging from 17.7% to 35.0%. We attribute this to the limited capacity of lower-order models to capture complex thermal

[5]The dataset is available at: https://developer.nrel.gov/docs/solar/nsrdb/.

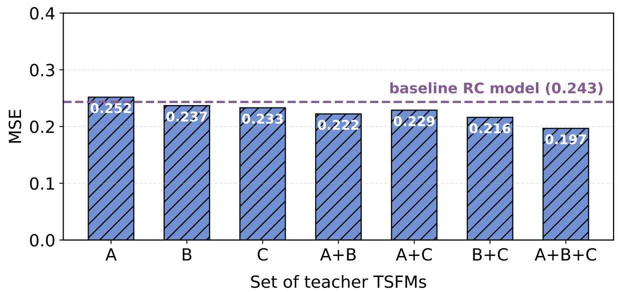

Figure 8: Performance of RC models trained with ThermoStill under different sets of teacher TSFMs. A, B, and C respectively denote Chronos, TimesFM, and TimeMoE. The dashed line indicates the average MSE of the baseline RC model trained without distillation.

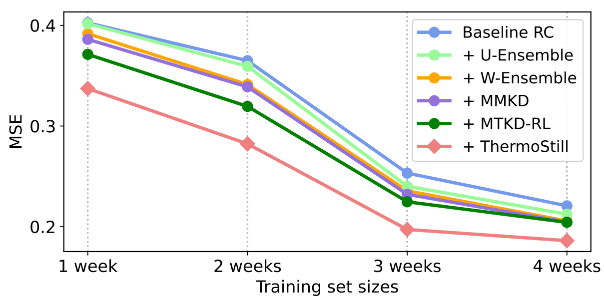

Figure 9: Performance of ThermoStill and other MTKD baselines under different training set sizes.

dynamics from noisy raw data. By providing smoothed supervision signals, TSFMs compensate for these structural limitations, guiding the simpler students toward a more optimal solution.

Different teachers. To examine the performance of ThermoStill with varying sets of teacher TSFMs, we designate the three evaluated TSFMs as Chronos (A), TimesFM (B), and TimeMoE (C), creating seven distinct teacher configurations ranging from single-model setups to the full combination, as illustrated in Figure 8 on Texas buildings. In single-teacher settings, ThermoStill functions as an adaptive weighting mechanism, dynamically adjusting the supervision signal of the single teacher for different samples. The horizontal dashed line represents the average MSE of the baseline RC model trained without distillation. Overall, the forecasting error exhibits a consistent downward trend as the teacher set expands, achieving average MSE improvement of 1.1%, 8.7%, and 19.1% for combinations of one to three TSFMs as compared to the baseline RC model. This suggests that diverse TSFMs encode complementary knowledge, the aggregation of which yields performance superior to that of any single model. Notably, the RC model distilled solely from Chronos performs slightly worse than the baseline, further underscoring the insufficiency and potential bias of relying on a single TSFM for different buildings.

Different transfer set sizes. We next study the effectiveness of ThermoStill and baselines under limited training data, which constitutes a practical constraint in real-world building energy systems. Specifically, we vary the training set size from one week to four weeks and evaluate on buildings in New York. Note that for the one-week setting, the input data length L is adjusted from 168 to 24 to ensure that there are enough data for obtaining predictions from TSFMs. As shown in Figure 9, ThermoStill consistently outperforms all baselines, delivering average MSE improvements of 13.6%, 17.9%, 16.7%, and 11.1% for each tested training set size. Notably, we observe that ThermoStill is able to surpass baselines that utilize more training data. For instance, ThermoStill trained on three-week data outperforms all baselines on the full four-week dataset. These results indicate that ThermoStill provides impressive data-efficiency and a consistent performance enhancement regardless of the specific data setup.

Table 3: Ablation study on key designs of ThermoStill. The best results are highlighted in bold.

Variant	MAE	MSE	CV-RMSE
w/o temporal encoder	0.2815	0.1654	0.1278
w/o validation reward	0.2734	0.1663	0.1247
w/o physical regularization	0.2845	0.1670	0.1335
w/o student pre-training	0.2886	0.1691	0.1367
ThermoStill	**0.2683**	**0.1550**	**0.1211**

5.4 Ablation Study

To answer **RQ3**, we conduct an ablation study in Table 3 to validate the contributions of key designs in ThermoStill. By isolating these designs, we propose four variants of ThermoStill, each purposefully modified to test the significance of specific components:

- **w/o temporal encoder** replaces the decomposition-based temporal encoder with a fully connected layer.
- **w/o validation reward** removes the reward term of forecasting accuracy on a held-out validation set.
- **w/o physical regularization** excludes the physical-consistency regularization on learnable parameters in RC models.
- **w/o student pre-training** removes the pre-training phase of the student RC model, directly performing joint training.

The results on buildings in Illinois indicate that each design component contributes to the superior performance of ThermoStill. Notably, omitting student pre-training leads to the most significant performance drop, increasing the MSE by 8.3%. This highlights the necessity of a well-initialized student model to provide stable reward signals for the optimization of the RL agent.

6 Control experiment 1: Simulation Study

To evaluate the efficacy of ThermoStill, we conduct closed-loop experiments using the Building Optimization Performance Tests (BOPTEST) simulation framework [3]. Specifically, the "BESTEST Hydronic Heat Pump" test case is selected. We first operate the BOPTEST environment using its default PI controller for a duration of 45 days (Jan 1 to Feb 14) at 30-minute intervals. The operational data collected during this period is split into a 30-day transfer set and a 15-day validation set. Using this training data, we develop two

Table 4: Control performance for all controllers. The results of RBC are considered as the baseline for energy savings.

ω	Controller	Energy (kWh)↓	Energy Savings(%)↑	Thermal Discomfort(Kh)↓
-	RBC	1179.09	–	22.01
5	MPC-RC	892.51	24.30	31.07
	MPC-ThermoStill	812.92	31.05	16.47
10	MPC-RC	876.69	25.64	22.07
	MPC-ThermoStill	836.96	29.01	9.99
20	MPC-RC	864.80	26.65	18.45
	MPC-ThermoStill	852.53	27.69	9.38

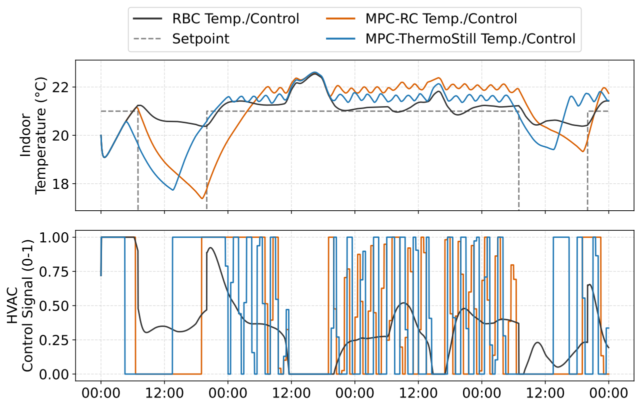

Figure 10: Comparison of indoor temperature and HVAC control signal for RBC, MPC-RC, and MPC-ThermoStill over the first four days of the simulation period ($\omega = 5$).

distinct RC models: one distilled via ThermoStill, and a baseline RC model trained solely on the data. The training process is repeated five times and the best model is selected for deployment due to stochasticity in ThermoStill. Subsequently, the two RC models are separately integrated into an MPC framework and deployed into the BOPTEST environment for a follow-up 30-day simulation (Feb 15 to Mar 16). The control objective is to minimize energy consumption and thermal discomfort, i.e., $J = J_{energy} + \omega \cdot J_{comfort}$, where ω denotes a weight factor for the thermal comfort objective. The MPC operates with a horizon of $H = 24$ steps and a control interval of 30 minutes. Details on the RC model and MPC formulations are presented in Appendix B.1.

We compare the performance of three controllers: MPC with the RC model distilled via ThermoStill (MPC-ThermoStill), MPC with the baseline RC model (MPC-RC), and the default rule-based PI controller (RBC) from BOPTEST. The experiments are conducted under three settings with varying weights $\omega \in \{5, 10, 20\}$ assigned to the thermal comfort objective. This allows us to study the behavior of controllers under different optimization priorities. The one-month simulation results are summarized in Table 4. Here, thermal discomfort is a default BOPTEST performance indicator, defined as the cumulative temperature deviation from comfort thresholds (in $K \cdot h$). Overall, both MPC controllers achieve significant energy savings compared to RBC, with average energy savings of 25.5% and

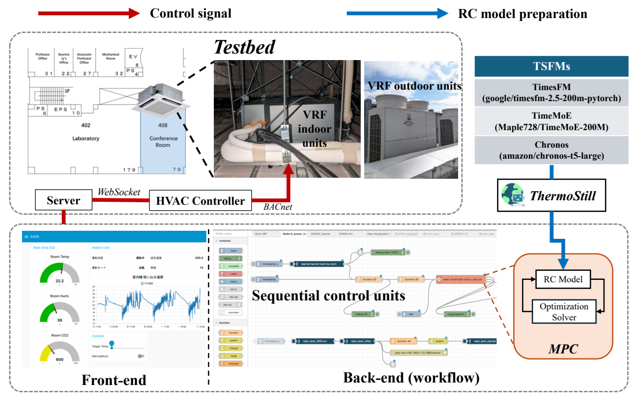

Figure 11: Deployed HVAC control system integrated with TSFMs and ThermoStill.

29.2% delivered by MPC-RC and MPC-ThermoStill, respectively. Beyond energy efficiency, MPC-ThermoStill effectively reduces an average thermal discomfort of 49.9% as compared to MPC-RC across all experimental settings. These results confirm that the superiority of ThermoStill in accurate thermal dynamics modeling can translate to an effective HVAC control strategy.

Figure 10 illustrates the detailed control behaviors during the first four days of the simulation period. Since MPC-RC and MPC-ThermoStill share the same MPC formulation, they exhibit generally similar control patterns. However, the more accurate modeling provided by ThermoStill results in more precise timing for control actions. For instance, prior to the start of the occupancy period, MPC-ThermoStill initiates preheating earlier than MPC-RC. This proactive action successfully maintains the zone temperature within the comfortable range, whereas MPC-RC results in temperatures falling significantly below the setpoint. Similarly, before the occupied period ends, MPC-ThermoStill turns off the heat pump earlier than MPC-RC. This better matches the shift to the unoccupied period, saving energy without affecting comfort.

7 Control experiment 2: Real-World Deployment

This section presents an end-to-end deployment of ThermoStill in a meeting room of a research building at the University of Osaka, Japan, from Dec. 18, 2025, to Jan. 18, 2026. Before deployment, we collect the operational data generated with the default rule-based PI controller from Nov 10, 2025 to Dec 10, 2025 as the transfer set to train an RC model with ThermoStill. The data from Dec 11, 2025 to Dec 17, 2025 are used as the validation set which we use to choose the best RC model under five training runs with ThermoStill. Then, the RC model is integrated into an MPC framework which operates with a horizon of $H = 24$ steps and a control interval of one hour. The control objective is to minimize energy consumption while maintaining the room temperature within a comfort bound ($22 \pm 2°C$) during working hours (from 8:00 a.m. to 8:00 p.m.). A diagram of the deployed HVAC control system integrated with ThermoStill is depicted in Figure 11. Details on the RC model and MPC formulations are presented in Appendix B.2.

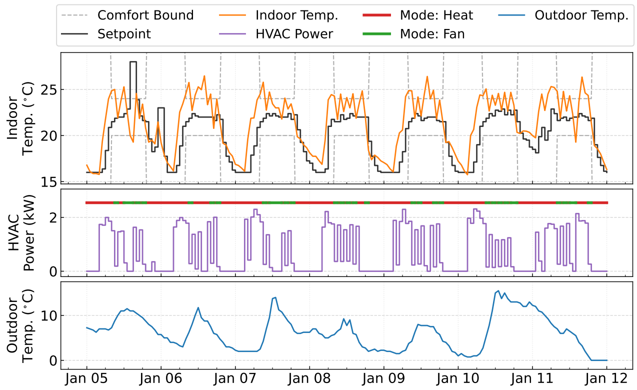

Figure 12: Deployment results of ThermoStill with an MPC framework in a meeting room.

We present the one-week deployment results in Figure 12 and provide the full one-month results in Appendix C. Overall, ThermoStill demonstrates effective control behavior. Each day, the HVAC system initiates preheating around 6:00 a.m. to gradually warm the room prior to occupancy. During working hours, the controller maintains the temperature setpoint near 22°C, keeping the room temperature within the comfort range, and switches off early to conserve energy. These results demonstrate that ThermoStill enables the MPC controller to closely adhere to the occupancy schedule, achieving energy savings without sacrificing comfort. While the general operation is stable, we observe some anomalous behaviors. For instance, two distinct spikes in the temperature setpoint occurred on Jan 5. Operational logs confirm that these were caused by manual overrides from occupants. Additionally, the room temperature frequently hovers near or exceeds the upper comfort bound. This discrepancy arises because the data collection sensor is mounted at the ceiling level. Consequently, the recorded temperature is typically lower than the temperature actually perceived by the occupants, leading to slight overheating.

8 Discussion

Integration of RL and MPC in HVAC control. Many prior works have explored RL for direct HVAC control, e.g., outputting real-time setpoints. However, real-world deployment is often hindered by safety concerns, limited interpretability, and high data dependency [8, 70]. Conversely, MPC is widely adopted in industrial control due to receding-horizon optimization and explicit constraint handling. Integrating RL into MPC thus offers a synergistic perspective: RL optimizes learnable components in the model-based pipeline, while MPC handles structured decision-making and constraint satisfaction. For instance, differentiable-MPC-based HVAC controllers such as Ibex-RL [46] use RL to adapt model-based components, including physics-informed dynamics and cost-related terms. Similarly, ThermoStill introduces RL agent to distill thermal dynamics model from pre-trained TSFMs for MPC. Thus, RL neither directly generates HVAC actions nor replaces MPC. The final HVAC control actions are still computed by MPC under explicit operational and comfort constraints.

TSFMs with covariate handling. While incorporating covariates inherently aligns with the multivariate nature of thermal dynamics, we argue that the success of KD depends primarily on the knowledge encapsulated in the teacher's outputs, rather than the features they utilize as input. Existing benchmark [49] reveals that covariate-aware TSFMs often yield comparable or even inferior accuracy in unseen domains compared to their univariate counterparts. This underperformance is likely attributed to feature misalignment between general-purpose pre-training data and building-specific control signals, which introduces noise rather than valid information. Motivated by these findings, we configure TSFMs to perform univariate forecasting in this study to ensure the delivery of high-quality supervision signals.

FM fusion. Model fusion [31] is an emerging paradigm that merges the parameters or predictions of multiple deep learning models into a single, more robust model. While FMs have demonstrated capabilities superior to those of traditional models, no single FM is universally optimal. Distinct FMs possess unique strengths due to variations in architecture, objectives, and pre-training data. This has motivated a growing body of research on FM fusion to enhance general performance [15], improve cross-task adaptability [21], and mitigate computational overhead [58]. The MTKD framework adopted in ThermoStill serves as a concrete implementation of this fusion concept. Consequently, this work represents a pioneering exploration into the fusion of TSFMs, specifically tailored for building energy management applications.

9 Conclusion

In this paper, we presented ThermoStill, a novel approach to apply general-purpose TSFMs for MPC-based HVAC control. Specifically, ThermoStill utilizes an RL-based multi-teacher knowledge distillation framework to aggregate and transfer knowledge from multiple TSFMs into a physics-informed RC model for MPC integration. Extensive evaluations, including modeling benchmarks across hundreds of buildings, standardized BOPTEST simulations, and a month-long real-world deployment, demonstrate that ThermoStill significantly enhances both modeling accuracy and energy efficiency in practical MPC-based HVAC control.

Future work. Future research may expand in two promising directions. First, methodologically, integrating ThermoStill with decision-focused learning techniques, such as differentiable MPC, warrants further exploration. Aligning the distillation objective with downstream control performance could yield models optimized for decision-making rather than pure prediction. Second, regarding application scope, extending ThermoStill to other control frameworks, such as model-based RL, or to energy-intensive environments like data centers, would help validate the broader generalizability of this foundation-model-driven control paradigm.

Acknowledgments

Dan Wang's work is supported in part by RGC GRF 15201322, 15230624, 15239925, CRF C5020-25GF, ITC ITS/052/23MX, and a start up fund of HKUST. This work is also supported by JSPS, Japan KAKENHI Grant Number 22H03697, 24K20901, and DAIKIN Industries, Ltd.

References

[1] Abdul Fatir Ansari, Lorenzo Stella, Caner Turkmen, Xiyuan Zhang, Pedro Mercado, Huibin Shen, Oleksandr Shchur, Syama Sundar Rangapuram, Sebastian Pineda Arango, Shubham Kapoor, et al. 2024. Chronos: Learning the language of time series. *arXiv preprint arXiv:2403.07815* (2024).

[2] Peder Bacher and Henrik Madsen. 2011. Identifying suitable models for the heat dynamics of buildings. *Energy and Buildings* 43, 7 (2011), 1511–1522. doi:10.1016/j.enbuild.2011.02.005

[3] David Blum, Javier Arroyo, Sen Huang, Ján Drgoňa, Filip Jorissen, Harald Taxt Walnum, Yan Chen, Kyle Benne, Draguna Vrabie, Michael Wetter, et al. 2021. Building optimization testing framework (BOPTEST) for simulation-based benchmarking of control strategies in buildings. *Journal of Building Performance Simulation* 14, 5 (2021), 586–610.

[4] Rishi Bommasani. 2021. On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258* (2021).

[5] Petros Brimos, Evangelos Kalampokis, and Konstantinos Tarabanis. 2024. Pretrained Time Series Foundation Models for Traffic Flow Prediction Using Open Government Data: The case of Attica traffic OGD. In *Proceedings of the 28th Pan-Hellenic Conference on Progress in Computing and Informatics*. 424–430.

[6] Felix Bünning et al. 2022. Physics-informed linear regression is competitive with two Machine Learning methods in residential building MPC. *Applied Energy* 310 (2022), 118491.

[7] Gaurav Chaudhary, Hicham Johra, Laurent Georges, and Bjørn Austbø. 2025. Transfer learning in building dynamics prediction. *Energy and Buildings* 330 (2025), 115384.

[8] Xin Chen et al. 2022. Reinforcement Learning for Selective Key Applications in Power Systems: Recent Advances and Future Challenges. *IEEE Transactions on Smart Grid* 13, 4 (2022), 2935–2958. doi:10.1109/TSG.2022.3154718

[9] Abhimanyu Das, Weihao Kong, Rajat Sen, and Yichen Zhou. 2024. A decoder-only foundation model for time-series forecasting. In *Forty-first International Conference on Machine Learning*.

[10] Ján Drgoňa, Javier Arroyo, Iago Cupeiro Figueroa, David Blum, Krzysztof Arendt, Donghun Kim, Enric Perarnau Ollé, Juraj Oravec, Michael Wetter, Draguna L Vrabie, et al. 2020. All you need to know about model predictive control for buildings. *Annual reviews in control* 50 (2020), 190–232.

[11] Yuwei Fan, Tao Song, Chenlong Feng, Chao Liu, and Dongxiang Jiang. 2025. Wind power prediction using foundation large time series models enhanced by time series prompt in exogenous and tuning forms. *Applied Energy* 400 (2025), 126535.

[12] Pietro Favaro, Jean-François Toubeau, François Vallée, and Yury Dvorkin. 2025. Decision-Focused Learning for Complex System Identification: HVAC Management System Application. In *Proceedings of the 16th ACM International Conference on Future and Sustainable Energy Systems*. 347–358.

[13] Srinivas Garimella, Kristian Lockyear, David Pharis, Omar El Chawa, Matthew T Hughes, and Girish Kini. 2022. Realistic pathways to decarbonization of building energy systems. *Joule* 6, 5 (2022), 956–971.

[14] Jianping Gou, Baosheng Yu, Stephen J Maybank, and Dacheng Tao. 2021. Knowledge distillation: A survey. *International journal of computer vision* 129, 6 (2021), 1789–1819.

[15] Yanggan Gu, Yuanyi Wang, Zhaoyi Yan, Yiming Zhang, Qi Zhou, Fei Wu, and Hongxia Yang. 2025. InfiFPO: Implicit model fusion via preference optimization in large language models. *arXiv preprint arXiv:2505.13878* (2025).

[16] Dimitri Guyot, Florine Giraud, Florian Simon, David Corgier, Christophe Marvillet, and Brice Tremeac. 2020. Building energy model calibration: A detailed case study using sub-hourly measured data. *Energy and Buildings* 223 (2020), 110189.

[17] Fang He, Jiaqi Fan, Yang Deng, Xiaoyang Zhang, Ka Tai Lau, and Dan Wang. 2025. Weather foundation model enhanced decentralized photovoltaic power forecasting through spatio-temporal knowledge distillation. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*. 9665–9673.

[18] Basu Hela, Praveen Prasad Handigol, and Pandarasamy Arjunan. 2025. Are Time Series Foundation models good for Energy Anomaly Detection?. In *Proceedings of the 16th ACM International Conference on Future and Sustainable Energy Systems*. 656–665.

[19] Seon-Young Heo, Jin-Hong Kim, Sunghyun Kim, Young Sub Kim, and Cheol Soo Park. 2025. Graphical federated MPC for real-life cooling towers and chillers using two simple thermal and energy efficiencies. *Energy and Buildings* (2025), 116769.

[20] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. 2015. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531* (2015).

[21] Chengsong Huang, Qian Liu, Bill Yuchen Lin, Tianyu Pang, Chao Du, and Min Lin. 2023. Lorahub: Efficient cross-task generalization via dynamic lora composition. *arXiv preprint arXiv:2307.13269* (2023).

[22] Nan Huang, Haishuai Wang, Zihuai He, Marinka Zitnik, and Xiang Zhang. 2024. Repurposing foundation model for generalizable medical time series classification. *arXiv preprint arXiv:2410.03794* (2024).

[23] IEA. 2022. Global status report for buildings and construction. https://www.iea.org/energy-system/buildings.

[24] Camille John, Charalampos Vallianos, José Candanedo, and Andreas Athienitis. 2018. Estimating time constants for over 10,000 residential buildings in North America: towards a statistical characterization of thermal dynamics. (2018).

[25] Anuj Kumar, Harish Kumar Saravanan, Shivam Dwivedi, and Pandarasamy Arjunan. 2025. Mix-BEATS: Mixer-enhanced Basis Expansion Analysis for Load Forecasting. In *Proceedings of the 16th ACM International Conference on Future and Sustainable Energy Systems*. 169–176.

[26] Anuj Kumar, Harish Kumar Saravanan, Shivam Dwivedi, and Pandarasamy Arjunan. 2025. MixForecast: Mixer-Enhanced Foundation Model for Load Forecasting. In *Proceedings of the 2nd International Workshop on Foundation Models for Cyber-Physical Systems & Internet of Things*. 25–30.

[27] Benjamin D Leibowicz, Christopher M Lanham, Max T Brozynski, José R Vázquez-Canteli, Nicolás Castillo Castejón, and Zoltan Nagy. 2018. Optimal decarbonization pathways for urban residential building energy services. *Applied energy* 230 (2018), 1311–1325.

[28] Hao Li, Bowen Deng, Chang Xu, Zhiyuan Feng, Viktor Schlegel, Yu-Hao Huang, Yizheng Sun, Jingyuan Sun, Kailai Yang, Yiyao Yu, et al. 2025. MIRA: Medical Time Series Foundation Model for Real-World Health Data. *arXiv preprint arXiv:2506.07584* (2025).

[29] Han Li, Giuseppe Pinto, Marco Savino Piscitelli, Alfonso Capozzoli, and Tianzhen Hong. 2024. Building thermal dynamics modeling with deep transfer learning using a large residential smart thermostat dataset. *Engineering Applications of Artificial Intelligence* 130 (2024), 107701.

[30] Jiatong Li, Bokai Ji, Guangxia Li, Peilin Zhao, Liu Liu, and Zhigang Ren. 2025. Tackling Non-Stationarity in HVAC Control with TimeGPT-Enhanced Deep Reinforcement Learning. In *Proceedings of the 12th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*. 384–388.

[31] Weishi Li, Yong Peng, Miao Zhang, Liang Ding, Han Hu, and Li Shen. 2023. Deep model fusion: A survey. *arXiv preprint arXiv:2309.15698* (2023).

[32] Zheng Li, Xiang Li, Lingfeng Yang, Borui Zhao, Renjie Song, Lei Luo, Jun Li, and Jian Yang. 2023. Curriculum temperature for knowledge distillation. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 37. 1504–1512.

[33] Zuguang Li, Shaohua Wu, Wen Wu, Qiaohua Ling, Yaping Sun, and Hui Wang. 2025. Split Knowledge Distillation for Large Models in IoT: Challenges and Solutions. *IEEE Internet of Things Magazine* (2025).

[34] Pengchen Liang, Haishan Huang, Bin Pu, Jianguo Chen, Xiang Hua, Jing Zhang, Weibo Ma, Zhuangzhuang Chen, Yiwei Li, and Qing Chang. 2025. Task-Specific Knowledge Distillation from the Vision Foundation Model for Enhanced Medical Image Segmentation. *arXiv preprint arXiv:2503.06976* (2025).

[35] Rui LIANG, Yang Deng, xiedonghua, and Dan Wang. 2025. Enabling Time-series Foundation Model for Building Energy Forecasting via Contrastive Curriculum Learning. In *ICML 2025 CO-BUILD Workshop on Computational Optimization of Buildings*. https://openreview.net/forum?id=7TgKHQeUsL

[36] Yuxuan Liang, Haomin Wen, Yuqi Nie, Yushan Jiang, Ming Jin, Dongjin Song, Shirui Pan, and Qingsong Wen. 2024. Foundation models for time series analysis: A tutorial and survey. In *Proceedings of the 30th ACM SIGKDD conference on knowledge discovery and data mining*. 6555–6565.

[37] Fan Liu, Behrooz Farkiani, and Patrick Crowley. 2025. Time-Series Foundation Models for ISP Traffic Forecasting. *arXiv preprint arXiv:2511.17529* (2025).

[38] Yuang Liu, Wei Zhang, and Jun Wang. 2020. Adaptive multi-teacher multi-level knowledge distillation. *Neurocomputing* 415 (2020), 106–113.

[39] Junwen Lu, Hanxiao Song, and Moudong Zhang. 2024. Optimization of Industrial Robot Navigation Algorithm Based on Pre-trained Large Model and Knowledge Distillation. In *Proceeding of the 2024 5th International Conference on Computer Science and Management Technology*. 699–705.

[40] Na Luo and Tianzhen Hong. 2022. Ecobee Donate Your Data 1,000 homes in 2017. (3 2022). doi:10.25584/ecobee/1854924

[41] Jiabo Ma, Zhengrui Guo, Fengtao Zhou, Yihui Wang, Yingxue Xu, Jinbang Li, Fang Yan, Yu Cai, Zhengjie Zhu, Cheng Jin, et al. 2024. Towards a generalizable pathology foundation model via unified knowledge distillation. *arXiv preprint arXiv:2407.18449* (2024).

[42] Amir M Mansourian, Rozhan Ahmadi, Masoud Ghafouri, Amir Mohammad Babaei, Elaheh Badali Golezani, Zeynab Yasamani Ghamchi, Vida Ramezanian, Alireza Taherian, Kimia Dinashi, Amirali Miri, et al. 2025. A Comprehensive Survey on Knowledge Distillation. *arXiv preprint arXiv:2503.12067* (2025).

[43] Ben A Marconi. 2025. Time series foundation models for multivariate financial time series forecasting. *arXiv preprint arXiv:2507.07296* (2025).

[44] Marcel Meyer, David Zapata, Sascha Kaltenpoth, and Oliver Müller. 2024. Benchmarking time series foundation models for short-term household electricity load forecasting. *arXiv preprint arXiv:2410.09487* (2024).

[45] Panagiotis Michailidis, Iakovos Michailidis, Federico Minelli, Hasan Huseyin Coban, and Elias Kosmatopoulos. 2025. Model Predictive Control for Smart Buildings: Applications and Innovations in Energy Management. *Buildings* 15, 18 (2025), 3298.

[46] Ozan Baris Mulayim and Mario Bergés. 2025. Ibex-RL: Interpretable and Scalable Control via Physics-Informed Reinforcement Learning. In *Proceedings of the 12th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*. 192–202.

[47] Ozan Baris Mulayim and Mario Bergés. 2024. Leveraging Grey Box Models for Enhanced Energy Flexibility in Centralized and Decentralized Single-Zone Multi-Node Systems. In *2024 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*. 264–270.

[48] Ozan Baris Mulayim, Pengrui Quan, Liying Han, Xiaomin Ouyang, Dezhi Hong, Mario Bergés, and Mani Srivastava. 2024. Are Time Series Foundation Models Ready to Revolutionize Predictive Building Analytics?. In *Proceedings of the 11th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*. 169–173.

[49] Ozan Baris Mulayim, Pengrui Quan, Liying Han, Xiaomin Ouyang, Dezhi Hong, Mario Bergés, and Mani Srivastava. 2025. Can Time-Series Foundation Models Perform Building Energy Management Tasks? *arXiv preprint arXiv:2506.11250* (2025).

[50] Gernot Pucher, Amin Dada, Felix Nensa, Martin Schuler, Christian Reinhardt, Jens Kleesiek, and Christopher M Sauer. 2025. Evaluating zero-shot foundation models for time series forecasting in clinical settings: A simulation study with electronic health records. *Studies in health technology and informatics* 329 (2025), 820–824.

[51] Xiangfei Qiu, Xingjian Wu, Yan Lin, Chenjuan Guo, Jilin Hu, and Bin Yang. 2025. Duet: Dual clustering enhanced multivariate time series forecasting. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 1*. 1185–1196.

[52] Eghbal Rahimikia, Hao Ni, and Weiguan Wang. 2025. Re (Visiting) Time Series Foundation Models in Finance. *arXiv preprint arXiv:2511.18578* (2025).

[53] Fabian Raisch, Thomas Krug, Christoph Goebel, and Benjamin Tischler. 2025. GenTL: A General Transfer Learning Model for Building Thermal Dynamics. In *Proceedings of the 16th ACM International Conference on Future and Sustainable Energy Systems*. 322–333.

[54] Simon Rouchier, Mickaël Rabouille, and Pierre Oberlé. 2018. Calibration of simplified building energy models for parameter estimation and forecasting: Stochastic versus deterministic modelling. *Building and Environment* 134 (2018), 181–190.

[55] Mert Bülent Sarıyıldız, Philippe Weinzaepfel, Thomas Lucas, Diane Larlus, and Yannis Kalantidis. 2024. UNIC: Universal classification models via multi-teacher distillation. In *European Conference on Computer Vision*. Springer, 353–371.

[56] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. 2017. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347* (2017).

[57] Xiaoming Shi, Shiyu Wang, Yuqi Nie, Dianqi Li, Zhou Ye, Qingsong Wen, and Ming Jin. 2024. Time-moe: Billion-scale time series foundation models with mixture of experts. *arXiv preprint arXiv:2409.16040* (2024).

[58] Yang Shu, Zhi Kou, Zhangjie Cao, Jianmin Wang, and Mingsheng Long. 2021. Zoo-tuning: Adaptive transfer from a zoo of models. In *International Conference on Machine Learning*. PMLR, 9626–9637.

[59] Rui Tang and Shengwei Wang. 2019. Model predictive control for thermal energy storage and thermal comfort optimization of building demand response in smart grids. *Applied Energy* 242 (2019), 873–882.

[60] Shihao Tu, Yupeng Zhang, Jing Zhang, Zhendong Fu, Yin Zhang, and Yang Yang. 2024. Powerpm: Foundation model for power systems. *Advances in Neural Information Processing Systems* 37 (2024), 115233–115260.

[61] Charalampos Vallianos, Andreas Athienitis, and Benoit Delcroix. 2022. Automatic generation of multi-zone RC models using smart thermostat data from homes. *Energy and Buildings* 277 (2022), 112571.

[62] Chenghao Wang, Jiyun Song, Dachuan Shi, Janet L Reyna, Henry Horsey, Sarah Feron, Yuyu Zhou, Zutao Ouyang, Ying Li, and Robert B Jackson. 2023. Impacts of climate change, population growth, and power sector decarbonization on urban building energy use. *Nature Communications* 14, 1 (2023), 6434.

[63] Meng-Chieh Wu, Ching-Te Chiu, and Kun-Hsuan Wu. 2019. Multi-teacher knowledge distillation for compressed video action recognition on deep neural networks. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2202–2206.

[64] Tianqi Xiao and Fengqi You. 2023. Building thermal modeling and model predictive control with physically consistent deep learning for decarbonization and energy optimization. *Applied Energy* 342 (2023), 121165.

[65] Jiajia Xie, Han Li, and Tianzhen Hong. 2024. A lifelong meta-learning approach for learning deep grey-box representative thermal dynamics models for residential buildings. *Energy and Buildings* 318 (2024), 114408.

[66] Qing Xu, Keyu Wu, Min Wu, Kezhi Mao, Xiaoli Li, and Zhenghua Chen. 2023. Reinforced knowledge distillation for time series regression. *IEEE Transactions on Artificial Intelligence* 5, 6 (2023), 3184–3194.

[67] Chuanguang Yang, Xinqiang Yu, Han Yang, Zhulin An, Chengqing Yu, Libo Huang, and Yongjun Xu. 2025. Multi-teacher knowledge distillation with reinforcement learning for visual recognition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 39. 9148–9156.

[68] Ze Yang, Linjun Shou, Ming Gong, Wutao Lin, and Daxin Jiang. 2020. Model compression with two-stage multi-teacher knowledge distillation for web question answering system. In *Proceedings of the 13th international conference on web search and data mining*. 690–698.

[69] Qingren Yao, Chao-Han Huck Yang, Renhe Jiang, Yuxuan Liang, Ming Jin, and Shirui Pan. 2024. Towards neural scaling laws for time series foundation models. *arXiv preprint arXiv:2410.12360* (2024).

[70] Peipei Yu, Hongcai Zhang, Yonghua Song, Zhenyi Wang, Huiyu Dong, and Liang Ji. 2025. Safe reinforcement learning for power system control: A review. *Renewable and Sustainable Energy Reviews* 223 (2025), 116022.

[71] Zeqin Yu, Jiangqun Ni, Jian Zhang, Haoyi Deng, and Yuzhen Lin. 2025. Reinforced Multi-teacher Knowledge Distillation for Efficient General Image Forgery Detection and Localization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 39. 995–1003.

[72] Fei Yuan, Linjun Shou, Jian Pei, Wutao Lin, Ming Gong, Yan Fu, and Daxin Jiang. 2021. Reinforced multi-teacher selection for knowledge distillation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 14284–14291.

[73] Ailing Zeng, Muxi Chen, Lei Zhang, and Qiang Xu. 2023. Are transformers effective for time series forecasting?. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 37. 11121–11128.

[74] Hailin Zhang, Defang Chen, and Can Wang. 2022. Confidence-aware multi-teacher knowledge distillation. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 4498–4502.

[75] Hailin Zhang, Defang Chen, and Can Wang. 2023. Adaptive multi-teacher knowledge distillation with meta-learning. In *2023 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 1943–1948.

[76] Dafang Zhao, Daichi Watari, Yuki Ozawa, Ittetsu Taniguchi, Toshihiro Suzuki, Yoshiyuki Shimoda, and Takao Onoye. 2022. A thermal comfort and peak power demand aware vrf heating/cooling management framework: Simulation and on-site experiment. *Journal of Information Processing* 30 (2022), 476–485.

[77] Dafang Zhao, Daichi Watari, Yuki Ozawa, Ittetsu Taniguchi, Toshihiro Suzuki, Yoshiyuki Shimoda, and Takao Onoye. 2023. Data-driven online energy management framework for HVAC systems: An experimental study. *Applied Energy* 352 (2023), 121921.

[78] Wanfu Zheng, Dan Wang, and Zhe Wang. 2024. Economic model predictive control for building HVAC system: A comparative analysis of model-based and data-driven approaches using the BOPTEST Framework. *Applied Energy* 374 (2024), 123969.

[79] Jie Zhu, Jide Niu, Sicheng Zhan, Zhe Tian, Adrian Chong, Huilong Wang, and Haizhu Zhou. 2025. A learning-based model predictive control method for unlocking the potential of building energy flexibility. *Energy and Buildings* 330 (2025), 115299.

[80] Zhuohang Zhu, Haodong Chen, Qiang Qu, and Vera Chung. 2025. FinCast: A Foundation Model for Financial Time-Series Forecasting. In *Proceedings of the 34th ACM International Conference on Information and Knowledge Management*. 4539–4549.

A RC Model Formulation

We present the formulations of three RC models adopted in the main experiment on the *ecobee* dataset. Let k denote continuous time. The variables include: T_{in} for indoor temperature, T_{out} for outdoor temperature, $\dot{Q}_{hvac}$ for controlled cooling input, and $\dot{Q}_{sol}$ for exogenous solar irradiance gain.

1R1C model. The 1R1C model represents the simplest thermal dynamics:

$$C\frac{dT_{in}}{dk} = \frac{T_{out} - T_{in}}{R} + \dot{Q}_{sol} - \dot{Q}_{hvac}, \tag{10}$$

where R and C denote the thermal resistance and capacitance, respectively. The solar gain is $\dot{Q}_{sol} = I_{sol}A_{eff}$, where I_{sol} is the global horizontal irradiance and A_{eff} is the effective solar aperture. The learnable parameters are $\theta = \{R, C, A_{eff}\}$.

2R1C model. The 2R1C model introduces an additional thermal mass to account for internal heat storage:

$$C\frac{dT_{in}}{dk} = \frac{T_m - T_{in}}{R_m} + \frac{T_{out} - T_{in}}{R_{out}} + \dot{Q}_{sol} - \dot{Q}_{hvac}, \tag{11}$$

where T_m represents the effective thermal mass temperature. R_m is the resistance between the zone air and internal mass and R_{out} is the resistance to the ambient environment. Solar gain is $\dot{Q}_{sol} = I_{sol}A_{eff}$. The learnable parameters are $\theta = \{R_m, R_{out}, C, A_{eff}\}$.

2R2C model. The 2R2C model describes the coupled dynamics of the indoor air and the building envelope:

$$C_i\frac{dT_{in}}{dk} = \frac{T_e - T_{in}}{R_i} + \dot{Q}_{sol,i} - \dot{Q}_{hvac}, \tag{12a}$$

$$C_e\frac{dT_e}{dk} = \frac{T_{in} - T_e}{R_i} + \frac{T_{out} - T_e}{R_e} + \dot{Q}_{sol,e}, \tag{12b}$$

where T_e is the envelope temperature. R_i and R_e represent the indoor-to-envelope and envelope-to-ambient resistances. C_i and C_e are the respective thermal capacitances. Solar gains are respectively defined as $\dot{Q}_{sol,i} = I_{sol}A_i$ and $\dot{Q}_{sol,e} = I_{sol}A_e$. The learnable parameters are $\theta = \{R_i, R_e, C_i, C_e, A_i, A_e\}$.

B Implementation Details on Control Experiments

B.1 Simulation Study

RC model. In this simulation experiment, we adopt a 2R2C model which is slightly different from the one we used for experiment on the *ecobee* dataset in Eq.(12a). The RC model is given by:

$$\begin{aligned} C_i\frac{dT_{in}}{dk} &= \frac{T_e - T_{in}}{R_i} + A_i\dot{Q}_{sol} + K\dot{Q}_{int} + \dot{Q}_{hvac}, \\ C_e\frac{dT_e}{dk} &= \frac{T_{in} - T_e}{R_i} + \frac{T_{out} - T_e}{R_e} + A_e\dot{Q}_{sol}. \end{aligned} \tag{13a}$$

Here, $\dot{Q}_{hvac}$ denotes the HVAC heat input to be controlled via adjusting the heat pump modulating signal u (ranging from 0 to 1). T_{out} is the ambient temperature. $\dot{Q}_{sol}$ is the solar irradiance; $\dot{Q}_{int}$ is the internal heat gain. The learnable parameters are $\theta = \{R_i, R_e, C_i, C_e, A_i, A_e, K\}$. The additional parameter K denotes a scaling factor to estimate total internal heat gains from $\dot{Q}_{int}$.

MPC Formulation. The MPC optimization problem is formulated as:

$$\min_{u_{1:H}} \quad \sum_{k=1}^{H}\Big(P_{hvac,k}\cdot\Delta k + \omega\cdot\delta_k\Big) \tag{14a}$$

$$\text{s.t.} \quad \hat{T}_{in,0} = T_{in,0} \tag{14b}$$

$$\hat{T}_{in,k+1} = f(\hat{T}_{in,k}, T_{out,k}, I_{sol,k}, \dot{Q}_{int,k}, \dot{Q}_{hvac,k}), \forall k \tag{14c}$$

$$T^{low}_{in,k} - \delta_k \le \hat{T}_{in,k} \le T^{high}_{in,k} + \delta_k, \forall k \tag{14d}$$

$$0 \le u_k \le 1, \forall k \tag{14e}$$

$$\delta_k \ge 0, \forall k. \tag{14f}$$

Here, H denotes the prediction horizon. The objective function aims to minimize the combined cost of energy consumption and weighted thermal discomfort over the prediction horizon. The first term, $P_{hvac,k}\cdot\Delta k$, calculates the HVAC energy usage at time step k, where $P_{hvac,k}$ is the power of the HVAC system. The second term, $\omega\cdot\delta_k$, penalizes the deviation from thermal comfort, where ω is a weighting parameter used to balance the trade-off between energy efficiency and occupant comfort. δ_k is an auxiliary variable introduced to ensure a non-empty feasible set for the optimization problem, which indicates temperature deviations for a preset thermal comfort threshold.

The constraints are defined as follows: Eq. (14b) specifies the initialization of the indoor temperature state. Eq. (14c) characterizes the system dynamics, predicting the evolution of the indoor temperature $\hat{T}_{in}$ based on the RC model defined in Eq.(13a). Eq. (14d) defines the thermal comfort bounds, maintaining the indoor temperature within the range $[T^{low}_{in,k}, T^{high}_{in,k}]$. Eq. (14e) restricts the control input u_k (i.e., modulation signal) within the feasible range of 0 to 1. Finally, Eq. (14f) ensures that the comfort violation term δ_k remains non-negative.

B.2 Real-world Deployment

RC model. In this real-world deployment, we adopt a simple 1R1C model which takes the same form as Eq.(10) but in a heating mode.

MPC Formulation. To maintain a comfortable indoor environment, the indoor temperature is regulated within a specific comfort range:

$$T_{lower} - s_k \le T_{in} \le T_{upper} + s_k, \quad \forall k, \tag{15}$$

$$0 \le s_k, \quad \forall k, \tag{16}$$

where T_{lower} and T_{upper} denote the lower and upper bounds for the indoor temperature. This temperature boundary is calculated by a linear regression Predicted Mean Vote (PMV) thermal comfort model [76]. Slack variable s_k is also a decision variable, taking a positive value when it exceeds the lower/upper bound. The thermal comfort term $J_{comfort}$ is denoted by following expression:

$$J_{comfort} = \frac{\sum_{k=1}^{H} O_k\cdot(T^k_{in} - T_{ref})^2}{\text{worst-case}}, \tag{17}$$

where O_k is the room's occupant information and $O_k = 1$ represents the room is occupied at time k for planning period H, and otherwise $O_k = 0$, and T_{ref} is the room's target temperature. Term $J_{comfort}$ is normalized by dividing it with the worst-case value for thermal comfort as,

$$\sum_{k=1}^{H} O_k \cdot \left(\max\left(T_{upper} - T_{ref}, T_{ref} - T_{lower}\right)\right)^2, \tag{18}$$

where $\max(\cdot,\cdot)$ selects the larger value. During actual HVAC operation, the HVAC unable to operate at a low power load. Therefore, we introduce following constraint on HVAC power consumption D_k for time step k as:

$$D_{lowest} \cdot \theta_k \leq D_k \leq D_{rated} \cdot \theta_k, \quad \forall k, \tag{19}$$

where D_{lowest} denoted the lowest possible power consumption when HVAC is operating, and D_{rated} is the HVAC rated power. To represent HVAC operation state, binary variable θ is introduced to represent HVAC ON/OFF status, where $\theta = 1$ represents HVAC is ON, and $\theta = 0$ represents HVAC is OFF. Same as the thermal comfort term, the power consumption term $J_{consume}$ is normalized by dividing it with the worst-case value of $D_{rated} \cdot H$. The $J_{consume}$ is defined as:

$$J_{consume} = \frac{\sum_{k=1}^{H} D_k}{D_{rated} \cdot H} \tag{20}$$

The proposed formulation includes two objectives, minimization of total energy consumption, maximization of thermal comfort. The MPC solves the following optimization problem every k step:

$$\begin{aligned}
&\text{minimize} && \omega \cdot J_{\text{consume}} + (1-\omega) \cdot J_{\text{comfort}} \\
& && + P_e \cdot \sum_{k=0}^{H-1} s_k, \\
&\text{subject to} && (10), (15) - (20), \\
&\text{input} && O_k,\ \forall k, \\
&\text{decision variables} && \{D_k, s_k\},\ \forall k,
\end{aligned} \tag{21}$$

where ω is a weight parameter that adjusts the importance between energy consumption $J_{consume}$ and thermal comfort $J_{comfort}$, which ranges from 0 to 1. In this experiment, we set $\omega = 0.1$ where thermal comfort is the main consideration, and P_e is a penalty coefficient which is set as 10,000.

C Full Real-world Deployment Result

Figure 13 presents the complete one-month real-world deployment of ThermoStill within an MPC framework in a meeting room, spanning from December 18 to January 18. The results illustrate the evolution of indoor temperature, HVAC power consumption, and outdoor conditions. Notably, the fluctuations observed during the initial days (Dec 18 to Dec 20) are attributed to the on-site commissioning and calibration phase. Following this brief adaptation period, ThermoStill achieved stable and efficient control, consistently maintaining the indoor temperature within the desired comfort bounds.

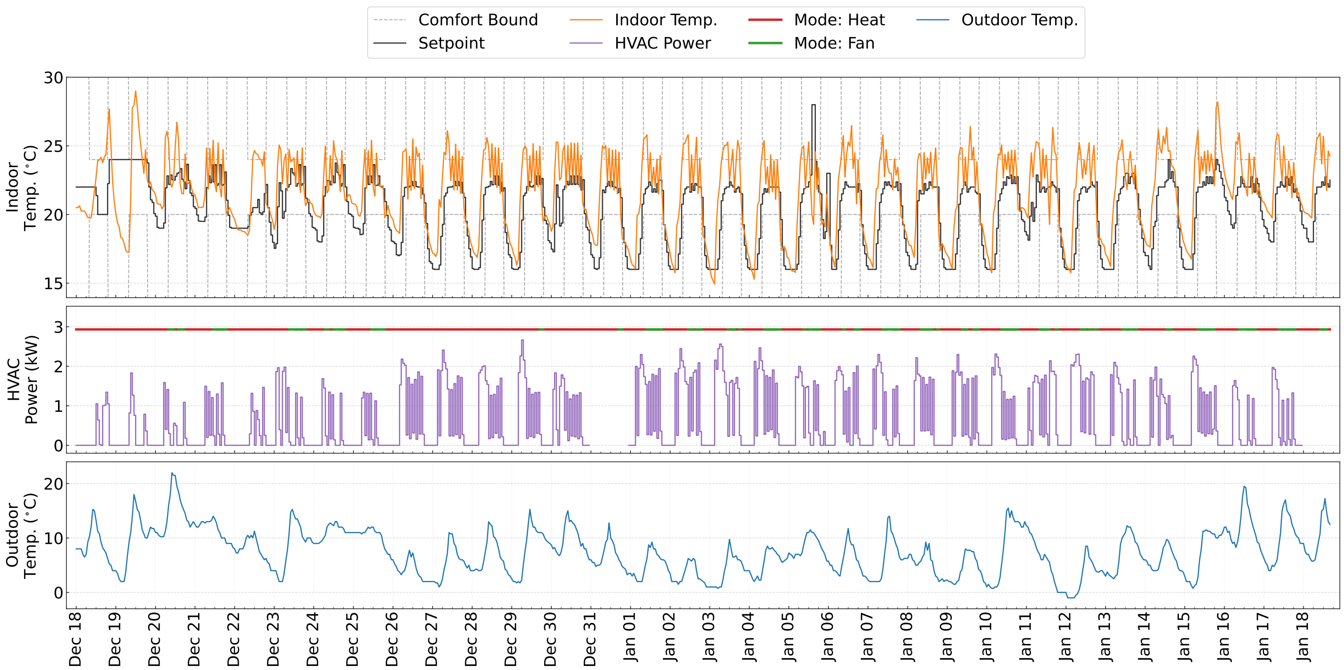

Figure 13: The complete one-month deployment results of ThermoStill with an MPC framework in a meeting room.