Benchmarking Forecasting Models for Long-Horizon Prediction of Temperature Distribution in Smart Buildings

Vaibhav Sourirajan [1]

Abstract

Accurate long-term forecasting of temperature distribution in buildings is critical for optimizing control strategies, improving energy efficiency, and maintaining occupant comfort. In this work, we benchmark three approaches for forecasting one-week ahead temperature distributions in a smart building. We evaluate the following models: (1) SeqCast, a Seq2Seq Encoder Decoder LSTM; (2) a Transformer-based direct forecasting model trained with curriculum learning; (3) a Robust Kalman Filter, a lightweight baseline grounded in classical state-space modeling. All methods are evaluated using mean absolute error across the prediction horizon. Our results show that the Transformer model significantly outperforms both SeqCast and the Robust Kalman Filter. Our study highlights the trade-offs between model complexity, interpretability, and forecasting performance in the context of building-level time series forecasting.

1. Introduction

Predicting indoor temperature distributions and dynamics is essential for enabling predictive control in smart buildings. As buildings account for 34% of global energy demand and 37% of energy and process-related CO2 emissions (BPIE, 2024), strategies that anticipate thermal needs can lead to significant energy savings and improved occupant comfort if deployed effectively. Data-driven strategies for long-horizon prediction are gaining traction as a measure of bridging sustainability and comfort.

With the increasing deployment of sensors in commercial and residential buildings, large volumes of high-resolution time series data are now available, providing a rich temporal and spatial view of indoor climate dynamics. These data pave the way for approaches that not only forecast overall temperature trends but also capture fine-grained, localized variations. Notably, the Smart Buildings Dataset (Goldfeder et al., 2025) includes 5-minute interval readings of Google buildings spanning all of 2022, capturing several physical building parameters. Each data sample $\mathbf{x}_t$ includes the following:

1. Observation values of sensor and setpoint measurements for various rooms in the building
2. Action values representing three water and air temperature setpoints
3. A floor-plan matrix and a device layout map providing spatial context of the sensors and actuators
4. Rich metadata including sensor/device IDs, timestamps, and mappings to physical locations

The dataset is divided into temporal partitions and is designed to support tasks such as forecasting, control, and representation learning in building environments. For this work, we focus specifically on learning building dynamics of the Google Smart Building from 1 month of 5-minute interval readings (June, 2022) and performing one-week ahead temperature distribution forecasting (July 1-7, 2022). Temperature readings for three devices in the dataset across the training period can be seen in Figure 1.

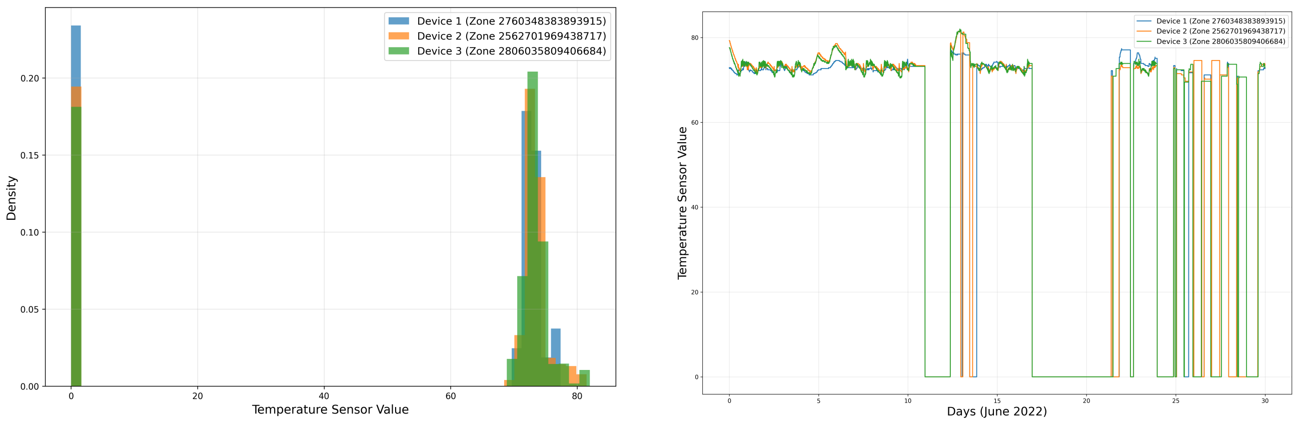
SeqCast (Sourirajan, 2025) has shown promise as an LSTM-based approach for indoor temperature forecasting. In this study, we benchmark it against two alternative approaches: a Transformer-based direct forecasting model and a Robust Kalman Filter baseline. We evaluate all models on both Mean Absolute Error (MAE) and end-to-end runtime to capture trade-offs between accuracy and computational efficiency.

2. Related Work

Previous work on modeling temperature in buildings has explored both physics-based simulations and data-driven techniques.

[1]Department of Computer Science, Columbia University, New York, USA. Correspondence to: Vaibhav Sourirajan <vs2787@columbia.edu>.

Proceedings of the 42^{nd} International Conference on Machine Learning, Vancouver, Canada. PMLR 267, 2025.



(a) Probability density graph of temperatures across 3 sensors

(b) Average temperature values of 3 sensors throughout the training period

Figure 1. Analysis of temperatures of three devices during the training period (June, 2022)

Classical control techniques such as Model Predictive Control (MPC) (Oldewurtel et al., 2012) and Kalman filtering (Alam et al., 2018) have long been used for temperature regulation and state estimation in buildings. MPC generally relies on physics-based thermal models to forecast future states and optimize control inputs, while Kalman filters use state-space formulations to estimate hidden system states from noisy observations. Both approaches require careful calibration and may struggle to generalize across buildings due to differences in geometry, construction materials, and occupancy patterns. While data-driven variants of Kalman filtering aim to learn system dynamics directly from sensor data, they remain constrained by assumptions of linearity and Gaussian noise.

In response to these challenges, the field has shifted toward machine learning techniques such as support vector regression (Lee & Baltazar, 2020), decision trees (Yu et al., 2010), and Gaussian processes (Rhee & Myoung, 2022). These methods have shown promise, but often lack scalability or fail to capture the complex spatial-temporal dependencies inherent in building systems. With the advent of deep learning, Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) models have emerged as more promising alternatives, demonstrating improved accuracy in short-term indoor temperature prediction (Elmaz et al., 2021; Shah, 2021).

Sequence-to-sequence (Seq2Seq) encoder-decoder models have also been applied to multi-step forecasting tasks like weather and building energy prediction (Kim et al., 2021). SeqCast, in particular, employs a Seq2Seq encoder-decoder architecture for long-term temperature distribution forecasting, offering stronger temporal modeling than standard autoregressive LSTMs (Sourirajan, 2025).

Transformer-based models—known for capturing long-range dependencies and enabling parallel computation—have most recently shown state-of-the-art results in long time series forecasting (Zhou et al., 2021; Wu et al., 2021) and are beginning to be applied in energy and building domains. However, their use for high-resolution temperature forecasting in buildings remains limited.

This paper benchmarks three approaches for long-term temperature distribution forecasting in smart buildings: (1) the SeqCast model as a Seq2Seq encoder-decoder LSTM baseline; (2) a Transformer-based direct forecasting model optimized with curriculum learning; and (3) a lightweight but interpretable Robust Kalman Filter grounded in classical linear state-space modeling. By comparing these methods under a common framework and evaluating both forecast accuracy and computational cost, we aim to provide a clearer picture of the trade-offs involved in modeling large-scale building environments.

3. Methodology

3.1. Problem Formulation

The objective of this task is to predict the future temperature distribution across multiple zones of a smart building given historical temperature data, historical exogenous and action variables, and future exogenous and action variables. Formally, let

$$X_{1:T} \in \mathbb{R}^{T \times D}$$

denote the historical temperature, exogenous, and action input sequence over T timesteps, and let

$$Y_{1:H} \in \mathbb{R}^{H \times D}$$

be the future temperature distributions over a prediction horizon of H steps. The model learns a mapping:

$$(X_{1:T}, E_{1:H}) \mapsto Y_{1:H}$$

where $E_{1:H}$ are exogenous and action variables of the prediction window. We focus on June, 2022 as our training window, consisting of 8640 5-minute interval measurements, and July 1-7 as our forecasting horizon, consisting of 2016 measurements.

3.2. SeqCast

Originally developed by (Sourirajan, 2025), SeqCast is an encoder-decoder model designed for forecasting of temperature distributions in the Smart Buildings Dataset. It builds on the standard sequence-to-sequence (Seq2Seq) framework using LSTM units, with an autoregressive decoder that generates future temperature trajectories based on learned temporal patterns and exogenous signals. The encoder receives as input a sequence of historical temperature distributions, exogenous variables, and action values over a fixed context window. It processes the sequence using a stacked LSTM to produce a hidden representation of the past dynamics. The final hidden and cell states of the encoder serve as the initial hidden state of the decoder, forming a learned context vector. Training is performed using teacher forcing, where the decoder receives the ground truth temperature from the previous step, rather than its own prediction, to stabilize learning. The model is trained end-to-end with a fixed context window of 1 day (288 steps) and a forecast horizon of 12 hours (144 steps), with the following weighted combination of Mean-Squared Error (MSE) and smoothness loss function:

$$\mathcal{L} = \text{MSE} + 0.1 \times \frac{1}{T-1} \sum_{t=1}^{T-1} (y_{t+1} - y_t)^2$$

During inference, the model ingests the entire training data as historical context and generates temperature distribution forecasts autoregressively, using the previously predicted temperature distribution as “ground truth”. While an effective approach in short-term forecasts, this methodology has several key limitations:

1. Autoregressive error accumulation: Small prediction errors at each step compound over long horizons, degrading performance.

2. Temporal generalization gap: The model is trained on short input/output sequences but evaluated on much longer ones, leading to a mismatch in temporal scale and a failure to fully leverage long-term historical trends.

3. Limited long-term memory: LSTMs have known limitations in retaining information over many timesteps (Kandadi & Shankarlingam, 2025), and even with stacked layers, the model may struggle to remember slow-evolving patterns or seasonal shifts.

4. Static context handling: The model is not explicitly designed to ingest or leverage large historical windows during inference, as it has only seen fixed-length (1-day) sequences during training.

We seek to address these with a transformer-based approach described below.

3.3. Transformer-Based Forecasting Architecture

To overcome the limitations of autoregressive sequence models such as SeqCast, we design a non-autoregressive Transformer-based architecture for direct, parallel prediction of temperature distributions across a one-week forecast horizon. This architecture is capable of ingesting long historical contexts and leveraging future exogenous and action variables to predict full-length output sequences without recursive decoding. It is inspired by the original Transformer architecture (Vaswani et al., 2017), adapted for the forecasting setting in smart buildings. The encoder receives the historical temperature, exogenous, and action variables, each of which are mapped to a shared latent dimension and summed. We inject positional information by adding the sinusoidal positional encoding proposed in (Vaswani et al., 2017) to the input embeddings. The decoder receives a concatenated tensor of the initial temperature distribution as well as exogenous and action variables for the full forecasting horizon. It produces an output tensor of dimension $L \times d_{\text{model}}$ where $L = 2016, d_{\text{model}} = 512$ which is then fed through a feedforward neural network to produce a sequence of temperature distributions of dimension 123 over the full forecasting horizon.

During training, we randomly sample variable-length historical sequences between 1 and 7 days, padding them to the maximum length in the batch using a custom collate_fn. To improve model stability and generalization for long-horizon forecasting, we adopt a curriculum learning strategy during training. Prior work has shown that curriculum strategies can improve convergence speed and lead to more stable predictions over long horizons (Koenecke & Gajewar, 2019) (Teutsch & Mäder, 2022). We begin training with shorter prediction lengths and gradually increase the forecast horizon ranges as training progresses. This approach allows the model to learn short-term dependencies before tackling long-horizon forecasts. We adopt the following schedule:

- Stage 0: 1-2 day forecasting horizon (15 epochs)
- Stage 1: 1-4 day forecasting horizon (20 epochs)
- Stage 2: 1-7 day forecasting horizon (25 epochs)
- Stage 3: 4-7 day forecasting horizon (30 epochs)

During inference, the model forecasts one full week using an entire month of historical data, an out-of-distribution shift from its training regime. Curriculum learning helps bridge this gap by progressively conditioning the model to handle longer-term temporal dependencies. The model is trained end-to-end using mean-squared-error loss. Unlike SeqCast, the model does not use teacher forcing and instead learns to map full historical sequences and future exogenous and action signals directly to full-length temperature distributions. Thus, the architecture is robust to error accumulation and better-suited for long-horizon forecasts.

3.4. Kalman Filters

As a model-based alternative, we implement a Robust Kalman Filter for estimating the state of the building given noisy historical measurements. The Kalman Filter is a recursive estimator that infers latent state dynamics and updates predictions by combining prior estimates with noisy observations, under the assumption of linear-Gaussian dynamics. Our implementation extends the classical formulation to high-dimensional multivariate time series with an emphasis on numerical stability and practical applicability in long-horizon forecasting scenarios. Kalman filters model the system through the following dynamics:

$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{B}\mathbf{u}_t + \mathbf{w}_t, \quad \mathbf{w}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}) \tag{1}$$

$$\mathbf{y}_t = \mathbf{C}\mathbf{x}_t + \mathbf{v}_t, \quad \mathbf{v}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{R}) \tag{2}$$

where

- $\mathbf{x}_t \in \mathbb{R}^d$ is the latent state vector representing the thermal state of the building at time t,
- $\mathbf{u}_t \in \mathbb{R}^m$ is the control input comprising both exogenous variables and action setpoints,
- $\mathbf{y}_t \in \mathbb{R}^n$ is the observed room-level temperature vector (with $n = 123$),
- $\mathbf{A}, \mathbf{B}, \mathbf{C}$ are the state transition, control input, and observation matrices respectively,
- $\mathbf{Q}, \mathbf{R}$ are the process and observation noise covariance matrices.

In our case, we choose the state dimension d to match the number of temperature sensors $n = 123$, assuming full observability and simplifying the problem to learning only the $\mathbf{A}, \mathbf{B}, \mathbf{Q}, \mathbf{R}$ matrices ($\mathbf{C}$ is trivially the identity matrix). After clipping outliers and scaling both temperature and control inputs, we construct tuples $(\mathbf{x}_t, \mathbf{u}_t, \mathbf{y}_{t+1})$ and solve the following regression:

$$\mathbf{y}_{t+1} \approx \mathbf{A}\mathbf{x}_t + \mathbf{B}\mathbf{u}_t \tag{3}$$

Table 1. Mean average error across the validation horizon (July 1-7, 2022) and total training+inference time in minutes of the SeqCast, Transformer Direct Forecasting with Curriculum Learning (Transformer CL), and Robust Kalman Filter (RKF). All experiments were conducted using an NVIDIA RTX 4090 GPU

Model	MAE	Time (min)
SeqCast	15.042	17
Transformer CL	1.606	492
RKF	58.188	1

We employ a ridge regression framework, scanning a range of regularization coefficients and selecting the model minimizing the mean squared residual while ensuring $\mathbf{A}$ has a spectral radius less than 1 to guarantee discrete-time system stability. Process and observation noise covariances $\mathbf{Q}, \mathbf{R}$ are empirically estimated from the residuals and regularized to be positive definite and well-conditioned.

After we learn the dynamics, we initialize the state estimate $\mathbf{x}_0$ to be a scaled initial temperature vector and covariance matrix $\mathbf{P}_0 = 0.01\mathbf{I}$. We forecast recursively for the entire horizon using the following equations:

$$\hat{\mathbf{x}}_{t+1} = \mathbf{A}\hat{\mathbf{x}}_t + \mathbf{B}\mathbf{u}_t \tag{4}$$

$$\mathbf{P}_{t+1} = \mathbf{A}\mathbf{P}_t\mathbf{A}^\top + \mathbf{Q} \tag{5}$$

We map the latent state back to the temperature space and perform an inverse normalization. To ensure robustness over long-horizon forecasts, predictions are clipped to reasonable bounds.

4. Results

Our results in Table 1 show the relative performance of the 3 models on the Smart Buildings Dataset over the 1-week forecasting horizon (July 1-7, 2022). The transformer model trained with curriculum learning (Transformer CL) achieved the best performance by a substantial margin, with a MAE of 1.606 compared to 15.042 for SeqCast and 58.188 for the Robust Kalman Filter (RKF). This demonstrates the superiority of non-autoregressive forecasting methods in capturing the dynamics of thermal behavior over longer horizons. However, the transformer's strong performance came with a notable computational cost, requiring almost 30x the training and inference time compared to SeqCast, underscoring a common tradeoff in such problems: larger transformer-based models offer higher performance at the cost of computational resources.

Figure 2 shows the average temperature prediction of the transformer model across the full forecasting horizon compared to the ground truth average values. The model accurately captures large-scale fluctuations in average temperature, especially sharp sensor measurement drops, indicating strong temporal generalization. However, it tends to underfit fine-grained sensor-level variations, such as those seen from July 3-5. Producing constant values during these periods of subtle shift indicates room for improvement in modeling local sensor-level transitions, perhaps through integration of spatial priors or localized attention mechanisms.

Despite its simplicity, the Kalman Filter struggles with the dimensionality and nonlinearity of the problem, though the results reported reflect a preliminary implementation of the Robust Kalman Filter (RKF). Future work may significantly improve performance through more careful tuning. In particular, incorporating more aggressive eigenvalue clipping for the transition matrix $\mathbf{A}$, adaptive regularization for the noise covariances $\mathbf{Q}, \mathbf{R}$, and advanced outlier rejection mechanisms during prediction could improve numerical stability and robustness over long horizons. SeqCast offers a reasonable compromise, reducing MAE by over 70% compared to the RKF, but its autoregressive nature leads to compounding error over longer horizons.

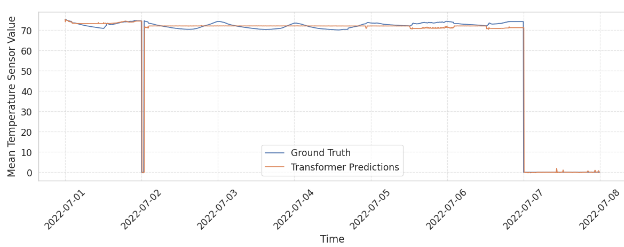


Figure 2. Average temperature prediction of transformer model trained with curriculum learning over validation forecasting window (July 1-7, 2022). We compare the predictions to the average ground truth sensor measurements during the same window.

5. Conclusion

In this work, we benchmark three distinct approaches for long-horizon forecasting of temperature distribution in smart buildings: a sequential LSTM-based Seq2Seq model (SeqCast), and a non-autoregressive transformer trained with curriculum learning, and a robust Kalman Filter. These models cover a range of paradigms, including recurrent sequence modeling, parallel direct prediction, and probabilistic modeling, respectively. As noted in Table 1, each paradigm has its unique trade-offs in accuracy and complexity.

Our findings point to the practical value of deep learning architectures, even in operational settings where forecasting accuracy can translate directly into energy savings. Future work will extend this benchmarking to incorporate spatial relationships between sensors to capture non-temporal dependencies. Integration of device layout, floorplan topology, and room adjacency could enhance forecasting performance by leveraging structural correlations between building zones. Additionally, supporting probabilistic forecasting and uncertainty estimation can further bridge the gap between forecasting and actionable energy management.

References

Alam, M., Sidhu, J., Clarke, J. A., and Elphick, R. Applying extended kalman filters to adaptive thermal modelling in homes. *Advances in Building Energy Research*, 12(1): 48–65, 2018. doi: 10.1080/17512549.2017.1358902.

BPIE, U. Global status report for buildings and construction: Beyond foundations: Mainstreaming sustainable solutions to cut emissions from the buildings sector. Report, United Nations Environment Programme (UNEP); Buildings Performance Institute Europe (BPIE); UCL Energy Institute; International Energy Agency (IEA), Global, March 2024.

Elmaz, F., Yildirim, O., Koseoglu, H., and Yildirim, S. CNN-LSTM architecture for predictive indoor temperature modeling. *Building and Environment*, 206:108327, 2021. doi: 10.1016/j.buildenv.2021.108327.

Goldfeder, J., Dean, V., Jiang, Z., Wang, X., dong, B., Lipson, H., and Sipple, J. The smart buildings control suite: A diverse open source benchmark to evaluate and scale hvac control policies for sustainability, 2025. URL `https://arxiv.org/abs/2410.03756`.

Kandadi, T. and Shankarlingam, G. Drawbacks of lstm algorithm: A case study, 2025. Available at SSRN: `https://ssrn.com/abstract=5080605`.

Kim, C. H., Kim, M., and Song, Y. J. Sequence-to-sequence deep learning model for building energy consumption prediction with dynamic simulation modeling. *Journal of Building Engineering*, 43:102577, 2021. doi: 10.1016/j.jobe.2021.102577.

Koenecke, A. and Gajewar, A. Curriculum learfning in deep neural networks for financial forecasting. In *Workshop on Mining Data for Financial Applications*. Springer International Publishing, Cham, 2019.

Lee, S. and Baltazar, J.-C. *Analysis of Support Vector Machine Regression for Building Energy Use Prediction*. Phd dissertation, Texas A&M University, 2020.

Oldewurtel, F., Sturzenegger, D., and Morari, M. Use of model predictive control and weather forecasts for energy efficient building climate control. *Energy and Buildings*, 45:15–27, 2012. doi: 10.1016/j.enbuild.2011.09.022.

Rhee, J. and Myoung, B. Objective and probabilistic long-range forecasts of summertime air temperatures in south korea based on gaussian processes. *Weather* *and Forecasting*, 37(3):329–349, 2022. doi: 10.1175/WAF-D-21-0122.1.

Shah, D. Short term temperature forecasting using lstms and cnn, 2021. Unpublished manuscript.

Sourirajan, V. SeqCast: Sequence-Based Forecasting of Temperature Distribution in Smart Buildings. In *Proceedings of the 16th ACM International Conference on Future and Sustainable Energy Systems*, 2025.

Teutsch, P. and Mäder, P. Flipped classroom: Effective teaching for time series forecasting. *arXiv preprint arXiv:2210.08959*, 2022.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

Wu, H., Xu, Y., Wang, J., Long, X., Zhao, T., Wang, Y., Tan, J., Zhang, C., Wu, J., Xue, Q., et al. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. In *Advances in Neural Information Processing Systems*, volume 34, pp. 22419–22430, 2021.

Yu, Z., Haghighat, F., Fung, A. S., and Zhou, L. A decision tree method for building energy demand modeling. *Energy and Buildings*, 42(10):1637–1646, 2010. doi: 10.1016/j.enbuild.2010.04.006.

Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., and Zhang, W. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 11106–11115, 2021.