CoolShift: Lightweight modeling of building cooling demand with causal machine learning

Tong Xiao
School of Mechanical Engineering
Tongji University
Shanghai, China
1910439@tongji.edu.cn

Peng Xu
School of Mechanical Engineering
Tongji University
Shanghai, China
xupeng@tongji.edu.cn

Abstract

We present **CoolShift**, a lightweight CausalML framework for counterfactual cooling-demand prediction under indoor setpoint interventions. CoolShift estimates condition-specific effects (CATE) via double machine learning with compact covariates, then composes effects into building-level counterfactuals—supporting fast "what-if" screening and aggregation to city-scale impacts. To evaluate both levels and effects, we build a quasi-random simulation corpus and the *Setpoint–Shift Benchmark* (SSB) with seen/unseen splits across heterogeneous buildings. CoolShift outperforms strong non-causal baselines (LightGBM, XGBoost, TabNet), maintaining low error on counterfactual levels (MAE $\approx$ 0.023–0.024) and accurate effects (MAE $\approx$ 0.028) on both splits, while baselines collapse on unseen effects (negative R^2). Results show that explicitly estimating causal effects, rather than differencing level predictors, is key for intervention queries and out-of-distribution generalization, enabling rapid portfolio- or city-scale DSM assessments.

1 Introduction

As major energy consumers, buildings benefit from accurate cooling-demand prediction to support demand-side management (DSM). Cooling demand depends on weather, building thermal dynamics, occupant behavior, and operational schedules. Operators can actively modulate demand by changing usage patterns, e.g., adjusting indoor air-temperature (IAT) setpoints, so models must provide reliable predictions under multiple prospective policies. Such forecasting is inherently a *what-if* problem that requires counterfactual predictions for interventions not observed in historical data. While control actions occur at the single-building level, DSM objectives are evaluated at portfolio or city scale, demanding building-level counterfactuals that transfer and aggregate to urban impacts.

First-principles models, such as building energy models (BEMs), are physically grounded but costly to calibrate and slow to scale to cities; purely data-driven models are lightweight and scalable [1-2] but offer no guarantees under interventions (i.e., out-of-distribution settings). Causal modeling offers a principled path to answer intervention queries when causal structure is learned from data [3]. Recent advances in causal machine learning (CausalML) address counterfactual prediction, yet most studies emphasize *discrete* interventions, limiting generality for continuous setpoint shifts [4–6]. Some researchers have explored CausalML on continoues treatment, but focuses on indoor air-temperature (IAT) prediction for a single building [7]. Overall, few CausalML approaches target cooling-demand modeling, especially at urban scale, and the limitations of standard data-driven predictors under interventions remain underexplored.

In this paper, we present **CoolShift**, a CausalML framework for decision-oriented cooling-demand modeling that delivers counterfactual predictions under IAT setpoint interventions (Fig. 1a). We construct a programmatic benchmark of setpoint shifts across heterogeneous buildings, a common

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: UrbanAI Workshop.

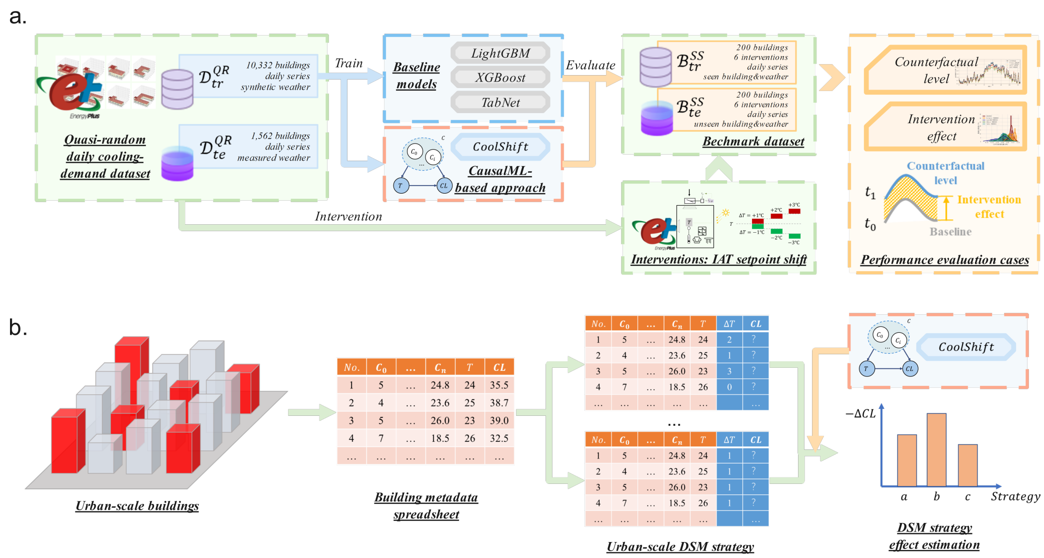

Figure 1: Framework overview. **(a)** *CoolShift*: CausalML estimator trained on quasi-random simulation to estimate the effect of IAT setpoint shifts (CATE) and produce counterfactual levels, then validated against non-causal baselines. **(b)** *Urban application (concept)*: lightweight feature inputs enable rapid, city/portfolio-scale "what-if" screening to support DSM decision-making.

DSM strategy, and compare CoolShift against XGBoost [8], LightGBM [9], and TabNet [10]. CoolShift is lightweight with a compact covariate set, enabling rapid "what-if" screening; Fig. 1b outlines a potential urban workflow for fast portfolio assessments. Controlled comparisons reveal the limits of standard predictors under interventions and the superior accuracy and robustness of CoolShift.

A summary of our contributions is given below:

- **Task & benchmark.** We formalize counterfactual cooling-demand prediction under setpoint interventions and build the *Setpoint–Shift Benchmark* (SSB) with seen/unseen splits ($\mathcal{B}^{\mathrm{SS}}_{\mathrm{tr}}$, $\mathcal{B}^{\mathrm{SS}}_{\mathrm{te}}$).
- **CoolShift.** A double–machine-learning estimator for continuous treatments using compact covariates to estimate condition-specific effects and compose counterfactual levels; designed for rapid screening.
- **Evidence.** CoolShift consistently outperforms strong non-causal baselines on both *levels* and *effects*, including unseen buildings/weather—substantiating the need for causal modeling beyond standard predictive training.

2 Method and case setup

2.1 Method

Problem formulation. We model IAT setpoint interventions as continuous shifts $T \in \mathbb{R}$ applied to an operational or physical variable of a building (e.g., indoor air-temperature setpoint or an envelope thermal parameter). Let CL denote the daily cooling demand per area (outcome), and let C denote the covariates that modulate intervention effects (weather, building morphology such as window–wall ratio, occupancy, and operational schedules). Within the potential-outcomes framework, the conditional average treatment effect (CATE) of shifting T from t_0 to t_1 for units with $C = c$ is

$$\tau_{CL}(\Delta t, c) = \mathbb{E}[\,CL(t_1) - CL(t_0)\,|\,C = c\,], \qquad \Delta t = t_1 - t_0. \tag{1}$$

This quantity captures how the same intervention magnitude can yield different cooling-demand changes under different covariate conditions c. For counterfactual prediction under a specified

Table 1: Counterfactual level under setpoint shifts on the benchmarks (overall aggregation). Left block: training-building split $\mathcal{B}_{\mathrm{tr}}^{\mathrm{SS}}$ (seen); right block: test-building split $\mathcal{B}_{\mathrm{te}}^{\mathrm{SS}}$ (unseen). Best per row within each split is **bold-underlined**; second-best is *italic-underlined*.

	$\mathcal{B}_{\mathrm{tr}}^{\mathrm{SS}}$ (seen)				$\mathcal{B}_{\mathrm{te}}^{\mathrm{SS}}$ (unseen)			
Metric	**CoolShift (ours)**	**LightGBM**	**XGBoost**	**TabNet**	**CoolShift (ours)**	**LightGBM**	**XGBoost**	**TabNet**
MAE	**0.0243**	*0.0426*	0.0442	0.0492	**0.0232**	0.0543	0.0495	*0.0416*
NMBE (%)	**-0.0535**	*-1.4136*	-2.1668	-2.3953	**-0.0885**	-3.2961	*-2.2599*	-2.9795
MAPE (%)	**8.1553**	*13.1234*	13.9430	14.9441	**8.3231**	17.5031	16.1206	*13.3911*
CV(RMSE)	**0.0956**	*0.1687*	0.1809	0.2021	**0.0976**	0.2308	0.2154	*0.1802*
R^2	**0.9891**	*0.9660*	0.9608	0.9511	**0.9884**	0.9353	0.9437	*0.9606*

Note: lower is better for MAE, MAPE, CV(RMSE); for NMBE, closer to zero is better; higher is better for R^2.

intervention, we target

$$CL(t_1, c) = CL(t_0, c) + \tau_{CL}(\Delta t, c), \tag{2}$$

which links a baseline operating point (t_0, c) to its post-intervention counterpart (t_1, c) while holding c fixed.

CausalML for effect estimation. We estimate the effect of indoor setpoint changes on daily cooling demand using double machine learning (DML) [11]. Let the setpoint T be a continuous treatment with shift $\Delta t = t_1 - t_0$, outcome CL, and covariates C (weather, morphology, occupancy, schedule; Appendix A). We target the CATE in Eq. (1) and form counterfactuals via Eq. (2). DML with cross-fitting learns $f(C) = \mathbb{E}[T \mid C]$ and $g(C) = \mathbb{E}[CL \mid C]$ (LightGBM), then regresses residuals via $CL - g(C) = \theta\,(T - f(C)) + \varepsilon$; motivated by domain knowledge, θ is linear. *Importantly,* the quasi-random simulation varies a *superset* of physics-plausible drivers; after data generation, we *down-select* the treatment T and the physics-guided, significance-tested covariate set C (Appendix A) and project the dataset onto (C, T, CL) for DML. Although T and C are not explicitly controlled during sampling, the LHS design provides broad coverage of their ranges and weakens spurious correlations, supporting approximate ignorability/overlap and stabilizing effect identification. This targets causal effects, handles continuous interventions, and scales across heterogeneous buildings.

Model development and training data. Because randomized experiments are infeasible, we train on a quasi-random simulation corpus generated via Latin hypercube sampling over physics-plausible geometry, envelope, usage, and weather (Appendix B). We denote the split from typical/synthetic weather as $\mathcal{D}_{\mathrm{tr}}^{\mathrm{QR}}$ (10,332 buildings; daily) and from measured 2015–2017 weather as $\mathcal{D}_{\mathrm{te}}^{\mathrm{QR}}$ (1,562 buildings; daily), with disjoint buildings. Our DML estimator trained on $\mathcal{D}_{\mathrm{tr}}^{\mathrm{QR}}$ is referred to as **CoolShift**.

2.2 Case setup

Baselines (non-causal). For comparison, we include strong non-causal predictors widely used on large tabular datasets: **LightGBM**, **XGBoost**, and a deep learning model **TabNet**. Each baseline learns $CL = \widehat{f}(C, T)$ on $\mathcal{D}_{\mathrm{tr}}^{\mathrm{QR}}$. Given a setpoint shift, the baseline *effect* is $\widehat{f}(c, t_1) - \widehat{f}(c, t_0)$; the *counterfactual level* is $\widehat{f}(c, t_1)$. Training protocols follow standard practice (Appendix C).

Evaluation setup. We evaluate under two benchmark splits built by simulations: $\mathcal{B}_{\mathrm{tr}}^{\mathrm{SS}}$ sampled from $\mathcal{D}_{\mathrm{tr}}^{\mathrm{QR}}$ ("seen") and $\mathcal{B}_{\mathrm{te}}^{\mathrm{SS}}$ from $\mathcal{D}_{\mathrm{te}}^{\mathrm{QR}}$ ("unseen"). Each split includes 200 buildings; per building we simulate six setpoint interventions ($-3, -2, -1, +1, +2, +3\,^{\circ}$C), yielding 1,200 building–intervention cases per split. We evaluate two targets: the *counterfactual level* $CL(t_1, c)$ and the *intervention effect* $\Delta CL = CL(t_1, c) - CL(t_0, c)$. Construction and metrics appear in Appendix D.

3 Results

Counterfactual level results. Under IAT setpoint shifts on the benchmark dataset, **CoolShift** consistently surpasses non-causal baselines across both the training-building split $\mathcal{B}_{\mathrm{tr}}^{\mathrm{SS}}$ ("seen") and the test-building split $\mathcal{B}_{\mathrm{te}}^{\mathrm{SS}}$ ("unseen"). It achieves the lowest overall errors (MAE $\approx 0.023 \sim 0.024$,

Table 2: Intervention effect estimation under setpoint shifts on the benchmarks (overall aggregation). Left: training-building split $\mathcal{B}_{\mathrm{tr}}^{\mathrm{SS}}$ (seen); right: test-building split $\mathcal{B}_{\mathrm{te}}^{\mathrm{SS}}$ (unseen). Best is **bold-underlined**; second-best is *italic-underlined*.

	$\mathcal{B}_{\mathtt{tr}}^{\mathrm{SS}}$ (seen)				$\mathcal{B}_{\mathtt{te}}^{\mathrm{SS}}$ (unseen)			
Metric	**CoolShift (ours)**	**LightGBM**	**XGBoost**	**TabNet**	**CoolShift (ours)**	**LightGBM**	**XGBoost**	**TabNet**
MAE	*0.0281*	0.0392	0.0340	**0.0187**	**0.0275**	0.1630	*0.1610*	0.1646
R^2	*0.8859*	0.7389	0.7794	**0.9422**	**0.8902**	-3.0315	*-3.0164*	-3.1754

Note: lower is better for MAE; higher is better for R^2.

MAPE $\approx 8\%$, CV(RMSE) < 0.10), near-zero bias (NMBE $\approx -0.05\% \sim -0.09\%$), and the highest $R^2 (> 0.984)$, indicating accurate counterfactual levels with minimal systematic bias and strong cross-domain stability. LightGBM and XGBoost show larger errors and pronounced negative bias with noticeable degradation from $\mathcal{B}_{\mathrm{tr}}^{\mathrm{SS}}$ to $\mathcal{B}_{\mathrm{te}}^{\mathrm{SS}}$; TabNet generalizes more steadily but remains inferior to the estimator. Table 1 summarizes *overall* metrics; building-wise aggregation and additional breakdowns appear in Appendix E.

Intervention effect results. On the benchmark dataset, **CoolShift** delivers accurate and stable effect estimates across both splits: MAE $= 0.0281$ / 0.0275 and $R^2 = 0.8859$ / 0.8902 on $\mathcal{B}_{\mathrm{tr}}^{\mathrm{SS}}$ ("seen") and $\mathcal{B}_{\mathrm{te}}^{\mathrm{SS}}$ ("unseen"), respectively (Table 2). This indicates that CoolShift captures the magnitude of setpoint–induced changes in daily cooling demand with strong generalization to unseen buildings and weather. In contrast, while TabNet appears competitive on seen buildings (MAE $= 0.0187$, $R^2 = 0.9422$), all non-causal baselines (LightGBM/XGBoost/TabNet) collapse on the unseen split with MAE > 0.16 and large negative R^2 (e.g., < -3), i.e., worse than predicting the sample mean—clear evidence that purely predictive models fail to learn a transferable causal relationship for the effect.

4 Discussions

Reasons for causal modeling beyond standard data-driven predictors. Evaluation results show that a causal formulation is necessary for intervention queries. **CoolShift** remains accurate on seen and unseen splits (low level error, near-zero bias; high effect R^2), while non-causal baselines degrade sharply on unseen effects (large MAE, negative R^2). The gap arises because standard predictors fit $\mathbb{E}[CL \mid C, T]$ and rely on plug-in differencing, which is not identified under confounding and covariate/weather shift. In contrast, **CoolShift** estimates the conditional treatment effect via double machine learning and then composes counterfactual levels, yielding transferable responses with a compact feature set and lightweight training/inference.

Potential application in urban-scale decision-making As a *potential application* (Fig. 1b), **CoolShift** enables portfolio- and city-scale "what-if" screening because it is lightweight (few features, fast inference) and yields condition-specific effects of setpoint shifts. With a basic building registry, usage schedules, and forecast weather, forming covariates C and baseline setpoints T, agencies can compute building-wise effects for small indoor setpoint changes and aggregate them by feeder, neighborhood, or district for *day-ahead* targeting and portfolio screening. Inference is milliseconds per building, enabling next-day rollups and simple scenario stress tests. Furthur deployment should include minimal guardrails (comfort/humidity limits, critical-facility exemptions) and light calibration with limited metering, with periodic retraining to handle nonstationarity. The same workflow extends beyond setpoints to envelope or ventilation interventions by swapping the treatment variable.

5 Conclusion

In this paper, we presented **CoolShift**, a lightweight CausalML framework for decision-oriented modeling of building cooling demand under setpoint interventions. By pairing a quasi-random simulation corpus with the Setpoint–Shift Benchmark (SSB), we evaluated both counterfactual levels and condition-specific effects and showed that CoolShift generalizes reliably to unseen buildings and weather, outperforming strong non-causal baselines. These findings indicate that explicitly estimating causal effects, rather than differencing level predictors, is crucial for robust "what-if" reasoning. Given its small feature footprint and fast inference, CoolShift can underpin rapid portfolio- or city-scale screening to support DSM planning. Future work will address uncertainty quantification and calibration with limited metering, extend to additional interventions (e.g., envelope/ventilation changes), and integrate comfort and grid objectives for multi-criteria decision-making at urban scale.

References

[1] Akyol, I. C. Halacli, E. G. Ucar, S. *et al.* (2025) Machine learning based prediction of long-term energy consumption and overheating under climate change impacts using urban building energy modeling. *Sustainable Cities and Society* **130**:106500. DOI:10.1016/j.scs.2025.106500.

[2] Ali, U., Bano, S., Shamsi, M. H. *et al.* (2024) Urban building energy performance prediction and retrofit analysis using data-driven machine learning approach. *Energy and Buildings* **303**:113768. DOI:10.1016/j.enbuild.2023.113768.

[3] Peters, J., Janzing, D. & Schölkopf, B. (2017) *Elements of causal inference: foundations and learning algorithms*. Cambridge, Mass: The MIT press (Adaptive computation and machine learning).

[4] Massidda, L. & Marrocu, M. (2023) Total and thermal load forecasting in residential communities through probabilistic methods and causal machine learning. *Applied Energy* **351**:121783. DOI:10.1016/j.apenergy.2023.121783.

[5] Jiang, F. & Kazmi, H. (2025) What-if: A causal machine learning approach to control-oriented modelling for building thermal dynamics. *Applied Energy* **377**:124550. DOI:10.1016/j.apenergy.2024.124550.

[6] Mun, J. & Park, C. S. (2025) A causal lens for building data: What lies beyond the measured? *Building Simulation*. DOI:10.1007/s12273-025-1314-y.

[7] Mun, J. & Park, C. S. (2025) Beyond correlation: A causality-driven model for indoor temperature control. *Energy and Buildings*, **338**:115739. DOI: 10.1016/j.enbuild.2025.115739.

[8] Chen, T. & Guestrin, C. (2016) XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*. DOI:10.1145/2939672.2939785.

[9] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W. et al. (2017) LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems (NeurIPS)*.

[10] Arik, S. O. & Pfister, T. (2019) TabNet: Attentive interpretable tabular learning. *arXiv preprint*. arXiv:1908.07442.

[11] Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W. & Robins, J. (2018) Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*. DOI:10.1111/ectj.12097.

Table 3: Covariate set C for the IAT setpoint–shift model, selected by physics knowledge and significance tests (two–sided, $p < 0.05$). Coefficients are from the linear effect component θ in the DML specification.

Category	Variable (description)	Name	p-value	Coef. in θ
Weather	Daily mean outdoor *wet-bulb* temperature	`WTemp`	$< 10^{-3}$	-0.003
Operations	Avg. outdoor air per capita	`AvgOA`	$< 10^{-3}$	-4.341
Usage	Day type	`DayType`	$< 10^{-3}$	-0.002
Form	Real window–wall ratio (overall)	`RWWR`	$< 10^{-3}$	-0.005
	Shape factor	`SF`	$< 10^{-3}$	0.063
	Longitudinal shape–factor ratio	`SFRatio`	$< 10^{-3}$	0.007
	Site coverage ratio	`SC`	$< 10^{-3}$	0.017
Envelope	Roof/wall heat–transfer coefficient	`WallU`	$< 10^{-3}$	-0.008
	Window heat–transfer coefficient	`WinU`	$< 10^{-3}$	-0.003

Table 4: Composition of the quasi-random daily cooling-demand datasets.

Subset	#Buildings	Weather source	Temporal granularity
Training	10,332	Typical/synthetic (e.g., TMY, CSWD, Meteonorm)	Daily (1 year/building)
Test	1,562	Measured (2015–2017)	Daily (1 year/building)

A Covariate selection

See Table 3.

B Quasi-random dataset generation

Variable space. We restrict the sampling domain to physics-plausible ranges informed by codes, standards, GIS surveys, and literature: (i) building form and massing (24 base geometries with abstract internal zoning and deformations), (ii) schedules/usage (type-specific with day/hour randomness and zone-level assignment), (iii) weather (17 typical/synthetic Shanghai files plus three years of measured data, 2015–2017), and (iv) other design/operation parameters (orientation, window–wall ratio, envelope U/SHGC, infiltration, indoor setpoint, outdoor-air rate, internal loads, etc.). Shape-related parameters are sampled uniformly within admissible bounds; many non-geometric parameters are sampled from normal distributions centered at code-recommended values to emulate practice while promoting independence across factors.

Sampling and simulation. Cases are drawn via Latin hypercube sampling (LHS) and auto-modeled in batch; two geometry families are used with equal counts: podium + tower and non-deformed massing (6,000 each). Simulations were executed in parallel on a 2.8 GHz 10-core CPU with 64 GB RAM (Windows 10); total wall-clock time $\approx$ 4 days.

Quality control and split. Auto-generated geometries may rarely yield invalid merges; after filtering anomalies, 11,894 runs remain. We build a *quasi-random daily cooling-demand training set* from typical/synthetic weather and a *test set* from measured 2015–2017 weather; buildings are disjoint across splits. See Table 4.

C Baseline models: training protocol and predictive performance

Training protocol. All baselines are trained on $\mathcal{D}_{\mathrm{tr}}^{\mathrm{QR}}$ using the same feature set C and treatment T; targets are daily cooling demand CL. We use a 75/25 split of $\mathcal{D}_{\mathrm{tr}}^{\mathrm{QR}}$ into internal train/validation for model selection. LightGBM and XGBoost hyperparameters are tuned via Bayesian optimization; TabNet uses early stopping (patience = 5). Model performance is listed in Table 5. This appendix only reports *predictive* (non-causal) accuracy to characterize the baselines.

Table 5: Predictive performance of baseline models on the quasi-random daily cooling-demand datasets. Metrics are computed as standard regression scores on building-day pairs.

On $\mathcal{D}_{\mathrm{tr}}^{\mathrm{QR}}$					
Model	**MAE**	**NMBE (%)**	**MAPE (%)**	**CV-(RMSE)**	$\mathbf{R^2}$
LightGBM	0.0191	-0.0000	6.3903	0.0732	0.9935
XGBoost	0.0224	-0.0002	7.5410	0.0925	0.9896
TabNet	0.0413	-0.1305	13.6753	0.1741	0.9631
On $\mathcal{D}_{\mathrm{te}}^{\mathrm{QR}}$					
Model	**MAE**	**NMBE (%)**	**MAPE (%)**	**CV-(RMSE)**	$\mathbf{R^2}$
LightGBM	0.0476	-0.3346	15.7655	0.2006	0.9483
XGBoost	0.0388	-0.4887	12.7304	0.1693	0.9632
TabNet	0.0383	-3.0684	12.6137	0.1670	0.9642

Table 6: Composition of evaluation benchmarks.

Benchmark (symbol)	**#Buildings**	**#Interventions**	**#Cases**
Training-building ($\mathcal{B}_{\mathrm{tr}}^{\mathrm{SS}}$)	200	6 (setpoint shifts)	1,200
Test-building ($\mathcal{B}_{\mathrm{te}}^{\mathrm{SS}}$)	200	6 (setpoint shifts)	1,200

D Benchmarks for evaluation

Benchmark construction. Starting from $\mathcal{D}_{\mathrm{tr}}^{\mathrm{QR}}$ and $\mathcal{D}_{\mathrm{te}}^{\mathrm{QR}}$, we sample 200 buildings for each benchmark ($\mathcal{B}_{\mathrm{tr}}^{\mathrm{SS}}$, $\mathcal{B}_{\mathrm{te}}^{\mathrm{SS}}$). For each sampled building, six indoor setpoint shifts are applied ($-3, -2, -1, +1, +2, +3\,^{\circ}$C) and simulated with EnergyPlus to obtain post-intervention cooling-demand series. Ground truth effects are computed as the difference between post- and pre-intervention simulations for the same building and day. See Table 6.

Usage protocol. All models consume identical inputs (covariates C and treatment T) and produce two outputs per case: a counterfactual level and an intervention effect. CausalML forms the counterfactual level via the baseline-plus-effect relation; baselines use direct level prediction and pre–post differencing for the effect.

Metric. For the level target, we report standard regression metrics: MAE, MAPE, CV(RMSE), NMBE, and R^2. Given that outcomes are daily series over the cooling season, we summarize errors in two complementary ways: (i) an overall aggregate across all building–day records to reflect fleet-level accuracy, and (ii) a building-wise average (first averaging over days per building, then averaging across buildings) to capture cross-building consistency and avoid dominance by buildings with longer seasons or more records. For the effect target, true values are often small (near zero), making ratio–based metrics (MAPE, CV(RMSE), NMBE) numerically unstable or misleading; therefore we use MAE and $\mathbf{R^2}$ as the core metrics.

E Additional Results

Countefactual level results. See Table 7.

Table 7: Building-wise metrics under setpoint shifts on the benckmarks (per-building average over cooling-season days, then averaged across buildings). Best is **<u>bold-underlined</u>**; second-best is *<u>italic-underlined</u>*.

Metric	**Split**	**CoolShift (ours)**	**LightGBM**	**XGBoost**	**TabNet**
MAPE (%)	$\mathcal{B}_{\mathrm{tr}}^{\mathrm{SS}}$ (seen)	**<u>8.5625</u>**	*<u>13.8377</u>*	14.8497	15.3918
	$\mathcal{B}_{\mathrm{te}}^{\mathrm{SS}}$ (unseen)	**<u>8.7988</u>**	18.4956	17.4168	*<u>13.9798</u>*
CV(RMSE)	$\mathcal{B}_{\mathrm{tr}}^{\mathrm{SS}}$ (seen)	**<u>0.0971</u>**	*<u>0.1658</u>*	0.1721	0.1928
	$\mathcal{B}_{\mathrm{te}}^{\mathrm{SS}}$ (unseen)	**<u>0.0990</u>**	0.2349	0.2176	*<u>0.1803</u>*
R^2	$\mathcal{B}_{\mathrm{tr}}^{\mathrm{SS}}$ (seen)	**<u>0.9843</u>**	*<u>0.9525</u>*	0.9464	0.9371
	$\mathcal{B}_{\mathrm{te}}^{\mathrm{SS}}$ (unseen)	**<u>0.9846</u>**	0.9152	0.9227	*<u>0.9508</u>*