Supplemental information for: Data-driven Characterization Of Cooling Needs In A Portfolio Of Co-located Commercial Buildings

Aqsa Naeem, Sally M. Benson, Jacques A. de Chalendar

December 13, 2023

1 Review of related literature for residential and commercial buildings

In the residential sector, regression studies coupling electric meter data with surveys have shown that along with weather, several factors such as internal gains, a building's physical characteristics, occupancy, and occupant behavior affect its energy consumption [1, 2, 3, 4]. Residential meter data have been used for load clustering, customer segmentation, and building archetype identification, not only for electricity [4, 5], but also for heating [6].

Commercial buildings have not received as much attention as residential buildings and that attention has almost exclusively been limited to electricity meter data [7]. Temperature is almost ubiquitously included as a predictor in regression models for commercial building electricity consumption [8]. A recent study suggests that humidity has a critical role in building's cooling loads and subsequent greenhouse emissions [9]. A study of chilled water and steam consumption in commercial buildings on two university campuses in the northeastern U.S. found that outdoor temperature explained steam consumption but not chilled water consumption [10].

Some approaches have used whole-building electricity meter data to improve simulation models by post-processing their output for institutional planning [11] and retrofit strategies [12].

In addition to weather, several studies have demonstrated that the Heating, Ventilation and Air Conditioning (HVAC) control system and control set points have a strong impact on energy consumption. Occupancy-driven zonal control was shown to reduce annual energy consumption by 7-15% compared to a fixed schedule in a recent simulation study [13]. Another study laid emphasis on finding alternative means of humidity control to ensure thermal comfort in hot-humid climate, rather than the conventional practice of cooling the air below its dew point and reheating it to the desired indoor temperature [14].

Increasingly available on-the-ground measurements such as the ones in this paper bring tremendous added value to simulation approaches for analyzing energy consumption and efficiency in buildings [15, 16]. They could greatly enhance the current capabilities of modeling tools [12, 11].

2 Data

In this section, we provide additional information on the cooling consumption data at the individual building and portfolio levels. Figure S1 shows an example building that follows a weekly operating schedule.

The portfolio-level cooling load is shown in Figure S2 and 2(B). The stacked plot in Figure S2 shows the 7-day rolling mean of buildings' cooling consumption in TJ over the five-year period. The Healthcare and Laboratory buildings make up the major proportion of the total cooling load. Figure 2(B) shows the percentage change in cooling load intensity of buildings with respect to a change in mean daily outside air temperature.

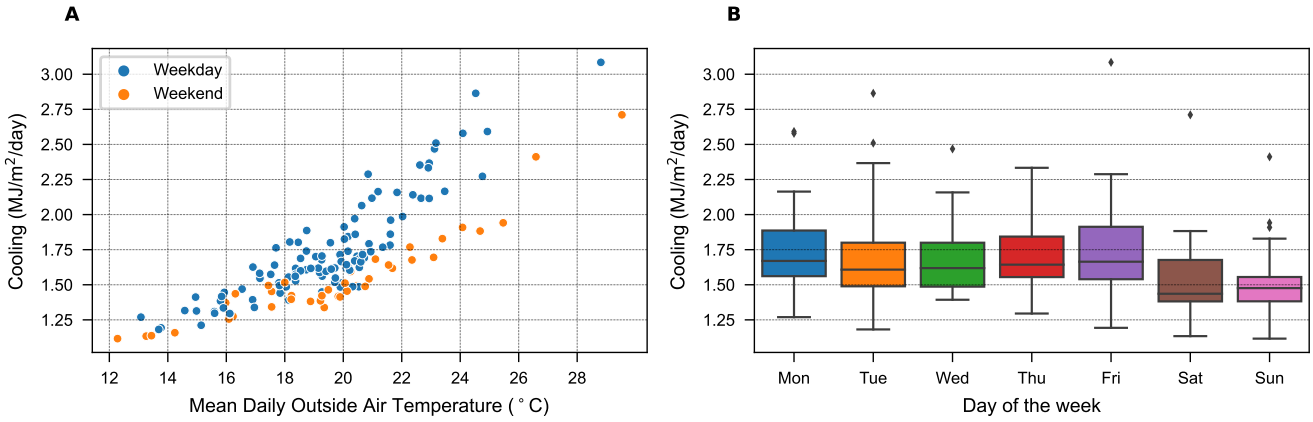

Figure S1: An example building that follows a weekly operating schedule. (A) Daily cooling consumption is linearly related to the mean daily outside air temperature (OAT). (B) Daily cooling consumption pattern over the course of a week.

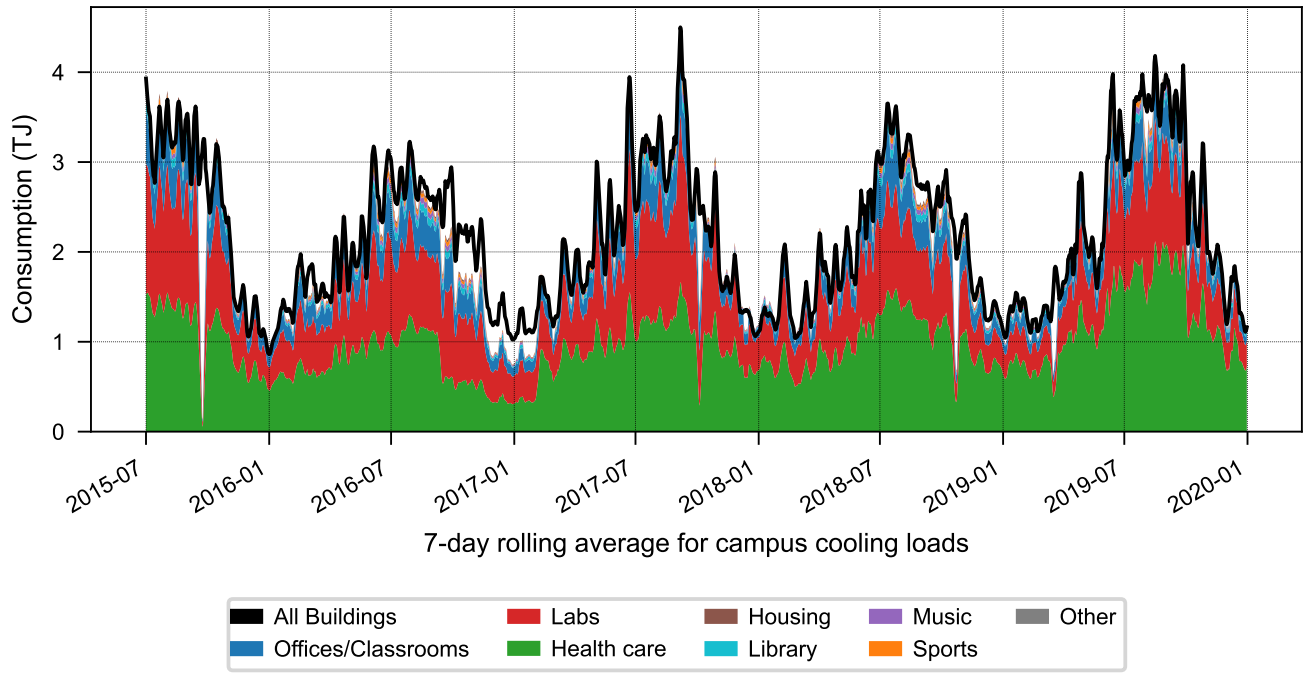

Figure S2: 7-day rolling mean cooling consumption of buildings in the portfolio.

3 Feature exploration

Figure S3 provides summary data on the considered building portfolio's climate. The heatmaps show the mean daily variations in different weather variables recorded over the course of five summers. The mean daily OAT is lower during the shoulder months (May/September) while July and August experience a greater number of hot days. Average daily solar radiation is below 225 W/m^2 throughout the summer of 2019 and drops even further by the end of September. The wind speeds remain calm, except in May of 2016 and 2017 and September 2019, when some days experience light to gentle breeze. On most days during the peak summer months (Jun-Aug), the percentage of RH remains close to the acceptable thermal comfort level (below 65%). Further, Tdew remains below 18°C (64.4°F), which means that the air holds little amount of moisture.

Figure S4 shows the distributions of mean daily OAT profiles during the summer season recorded in each year. Over 95% of mean daily temperatures vary over the range of 14 to 25°C, with an average value of 19°C. Figure S4-A shows the empirical cumulative distribution function of mean daily OAT for the summer season (May - September) during the five-year period of interest. The years 2017 and 2019 experienced some heat waves with an average daily OAT rising above 25°C (77°F), while 2015 had the largest number of moderately 'cold' days with average daily OAT of less than 15°C (59°F) during the summer. In the remaining years, OAT remained below 20°C (68°F) in about 8 out of 10 days. Figure S4-B through Figure S4-F shows the hourly profiles of OAT colored by mean daily OAT. The highest temperature recorded in 2017 is 42.8°C (109°F), while the mean daily temperature during that day remained above 26.7°C (80°F).

4 Feature Curation

We use a three-step feature curation process to process the features and obtain feature sets in each year that could help explain the buildings' cooling loads (illustrated in Figure S5). First, we remove features with small variation, as indicated by computing the standard deviation of data. Second, we repeatedly compute pairwise correlations to eliminate features that exhibit an absolute correlation coefficient of more than 0.5 with another feature. For instance, in 2015, the correlation between OAT and RH is 0.65, while Tdew and RH exhibit a correlation coefficient of 0.77. In this case, only RH is eliminated. In contrast, in 2017, Tdew and wind exhibit larger than 0.5 correlation with OAT, so both features are eliminated. In the third step, we regress each feature on the remaining features and compute the corresponding Variance Inflation Factor (VIF). Any feature that has a VIF of more than 10 is removed from the set. This additional step eliminates the possibility of having two or more features in the feature set of each building that together could explain a high proportion of the variance in the remaining predictor. However, in our dataset, VIF of each feature after step (2) remains significantly low, so none of the remaining features is eliminated.

5 Model selection criterion

The one standard-error rule can be used to compare models with different features and choose the simplest model that has comparable predictive performance as other models. We apply this rule to select the most parsimonious model for all buildings each year as follows. For a k-fold CV, we first calculate the MSE (denoted by ϵ) and standard error ($\bar{se}$) averaged over k folds. The cutoff value is computed as $\bar{\epsilon}_j + \frac{\bar{\sigma}_j}{\sqrt{k}}$, where $\bar{\epsilon}_j$ is the minimum average MSE of all the models and its corresponding standard deviation is represented by $\bar{\sigma}_j$. The most parsimonious model among all the models that have lower CV-MSE than this cutoff value, is the one with the lowest error. For instance, for the portfolio-level models in 2017, M3 gets chosen out of 4 possibilities, which are tabulated in Table S1.

We also considered the use of R^2_{adj} in determining the optimal model for individual buildings. It was found that selecting the models based on R^2_{adj} does not alter the estimated responses of buildings at different temperatures (see

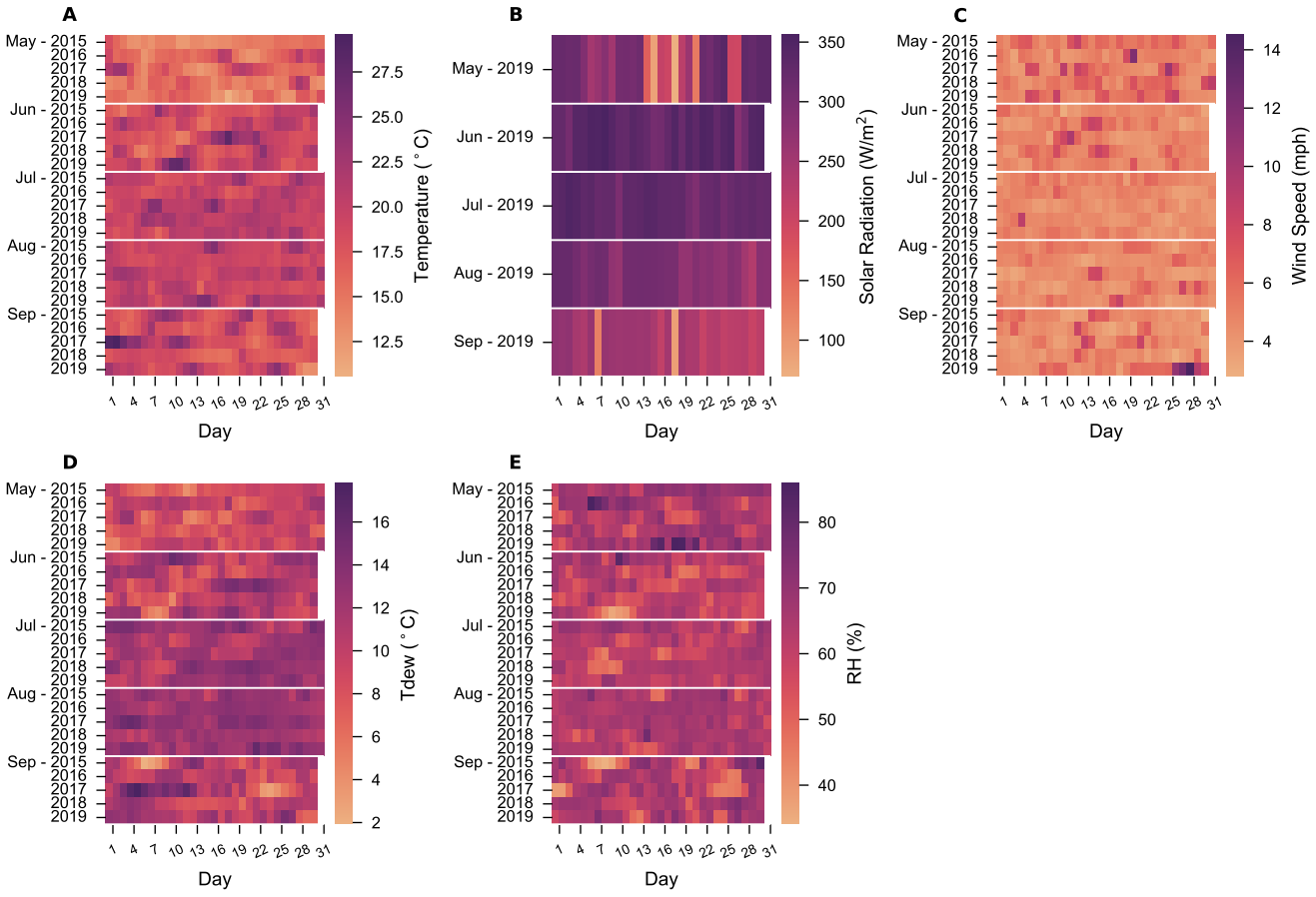

Figure S3: Warm-summer Mediterranean climate as illustrated by the mean daily variations in weather characteristics across years during the summer months (May - September). The x-axis represents the day of the month, while the y-axis represents the month and year. Tdew and RH represent dew point of temperature and relative humidity respectively. (A) Outside air temperature: Darker colors represent higher temperatures. (B) Solar radiation. Shoulder months receive a lesser amount of radiation. (C) Wind speed: Light to gentle breeze on some days in the region. (D) Dew point of temperature: Air holds little amount of moisture. (E) Relative humidity stays around 65% during the peak summer months (June-August). Note that the data for solar radiation prior to 2019 is missing and thus, has been excluded.

Figure S6). However, the model selected using the one standard-error rule requires fewer features to characterize cooling consumption with a comparable performance.

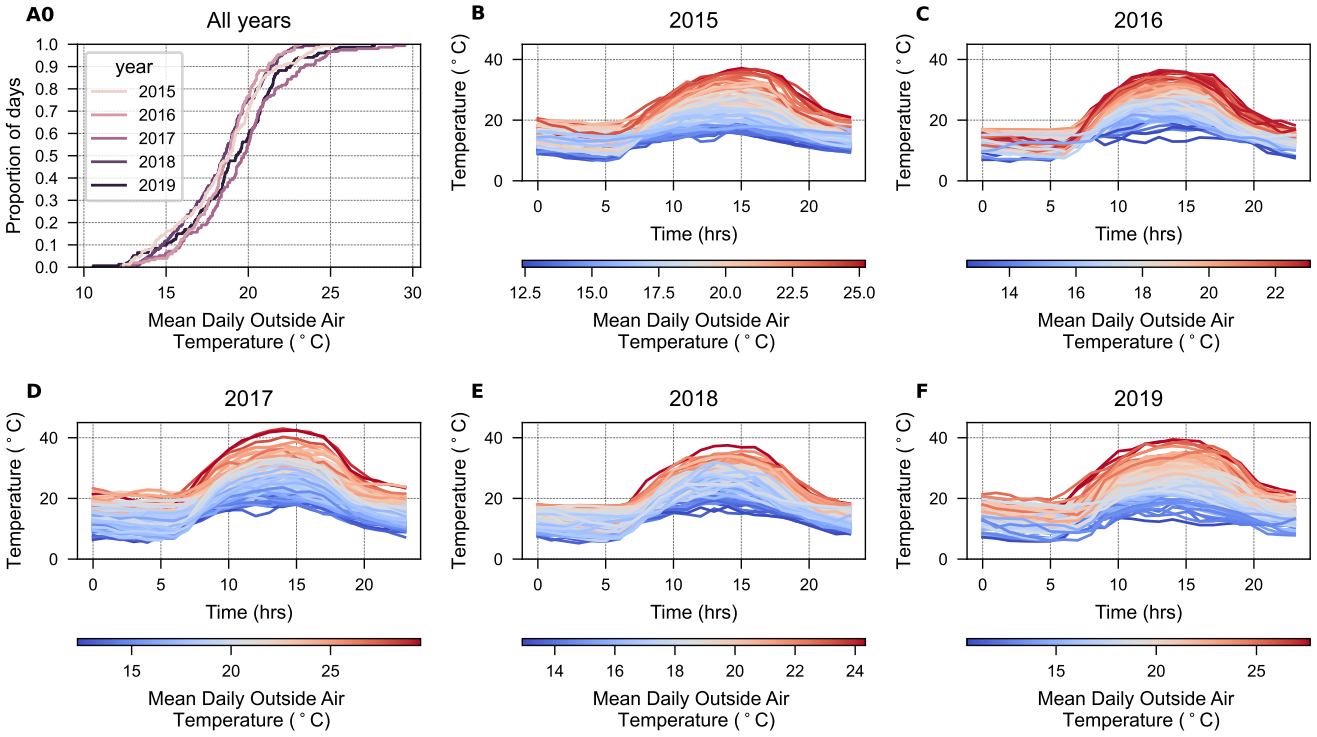

Figure S4: Distributions of outside air temperature. (A) Empirical cumulative distribution function for outside air temperature in different years. (B) - (F) show hourly temperature profiles colored by mean daily temperature in 2015 through 2019.

6 Model cross validation

We choose three CV error metrics, CV-R^2, CV-RMSE, and CV-MAPE to quantify the performance of selected models. Using the CV-based model selection criterion outlined above, we choose the best parsimonious model M^n_{opt} for n^{th} building, where n $\in$ N. The average CV error metrics over the training and testing data are plotted against the fraction of modeled load in Figure S7. The fraction of modeled load is obtained using the expression given in Section . The training CV-R^2 stays above the 0.7 threshold for about 92% of the modeled load, while for the testing CV-R^2, this percentage of load drops to 65%. The CV-RMSE and CV-MAPE remain comparable for the training and testing periods.

7 Model Evaluation Metric

The performance is evaluated using the R^2 statistic. The Akaike and Bayesian Information Criteria (AIC and BIC) were also considered as performance criteria during model selection but did not significantly affect our results (see Table S1). The table shows the portfolio-level model selection and evaluation error metrics in 2017. Notice that even though both M2 and M3 have the CV R^2_{adj} value of 0.94, the most parsimonious model is M3. Evaluating the

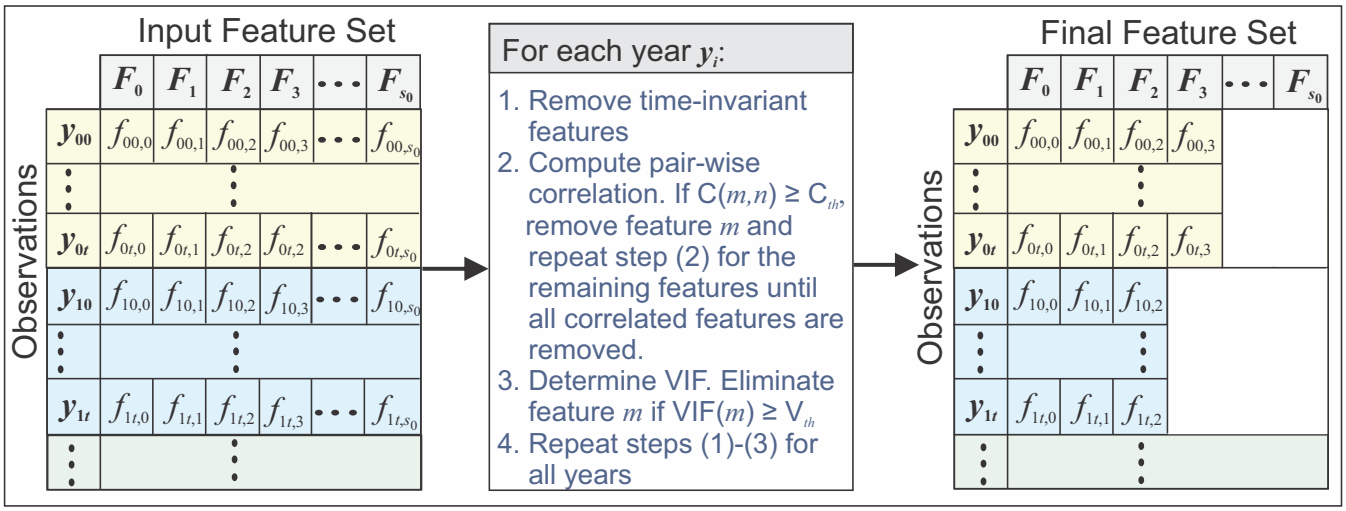

Figure S5: Feature curation process as referenced in Figure 6.

Table S1: Model Selection and Evaluation. Using the Cross-Validation (CV) Mean-Squared Errors (CV-MSE) to determine the most parsimonious model for all buildings with low CV error and retraining the models on the whole dataset to evaluate them using R^2, Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) metrics. Note that Nobs stands for number of observations, while M0 through M3 represent different models.

Model	Model Evaluation				Model Cross-Validation		Features
	Nobs	R^2	AIC	BIC	CV-R^2_{adj}	CV-MSE (test)	
M0	153	0.91	-353	-347	0.91	0.0897	[temperature]
M1	153	0.92	-372	-363	0.92	0.0675	[temperature, RH]
M2	153	0.94	-407	-399	0.94	0.0648	[temperature, Weekend]
M3	153	0.95	-423	-411	0.94	0.0521	[temperature, RH, Weekend]

Table S2: Performance of buildings at the portfolio-scale.

Level	2015	2016	2017	2018	2019
All Buildings	0.94	0.96	0.95	0.94	0.90
Offices/Classrooms	0.92	0.94	0.89	0.87	0.75
Housing	0.78	0.81	0.78	0.77	0.74
Labs	0.94	0.94	0.92	0.89	0.90
Music	0.68	0.93	0.92	0.84	0.79
Library	0.88	0.81	0.80	0.79	0.76
Sports	0.75	0.89	0.70	0.79	0.81
Health care	0.88	0.33	0.91	0.90	0.86
Other	-	-	-	0.78	0.76

performance of this model on the summer-long data, R^2 turns out to be 0.95.

Using the features in M^n_{opt}, we retrain the selected model on the entire summer data of each year. The performance of the retrained model is evaluated using R^2 in Figure S8 for all years, which remains comparable over the years at the portfolio-level. Table S2 compares the results at the portfolio-scale. The buildings that house laboratories ("Labs") exhibit the highest R^2 in all years, while the proportion of load explained in "Housing' buildings is the lowest. The buildings in the "Other" category were constructed in 2018, which is why the previous years have no data. It is worth mentioning that the predictive performance of log-linear models remain on par with linear models (see Figure S9), but for the ease of interpreting the model coefficients, we used log-linear formulation in this work.

8 Feature sensitivity test

We use the model estimates to determine how sensitive the cooling loads are to variations in each feature. Defining the base load intensity at 18°C (64.4°F) and average values of other features on a weekday (weekend indicator = 1) as the base case, we record the variations in cooling intensity by varying each feature, one at a time, from its minimum to maximum limit while keeping other features fixed at their mean values.

Figure S10 shows the corresponding tornado chart for the top 42 buildings with the highest R^2 in 2017. The x-axis shows the daily cooling consumption, while the y-axis shows different features. The base case is denoted by the dotted line. The width of the bars indicate the influence of each feature on the cooling load intensity, that is, the bigger the width, the higher the sensitivity of each feature. The max-min variations for each feature besides the weekend indicator are shown in blue and red colors respectively. The weekend indicator is a binary variable which is why we only show the change in cooling consumption on the weekend with respect to the base case (weekday). As shown, temperature is the most important feature for all buildings. For large-sized buildings, RH seems to be the next important feature, while for small-sized buildings, Weekend indicator is the second most important feature. Further, notice that some buildings are only affected by temperature and other features seem to have no influence on the cooling load intensity.

We plot the tornado chart at the category-level of buildings in Figure S11. At the portfolio-level, besides temperature, the cooling load is sensitive to variations in RH, weekend indicator, and solar radiation. The remaining three features do not make any significant impact on the cooling load. Comparing different categories, some buildings are affected more by temperature variations than others. After temperature, RH is the next important feature for all categories except Healthcare and Offices and Classrooms buildings, the latter of which are impacted more by the weekly operating schedules.

9 Feature importance in other years

Figure S12 shows the significance of sequentially adding features to the baseline model (that uses just outside air temperature to predict the cooling load) in all years. Figure S12(A) illustrates the number of buildings that choose a given feature in each year. Temperature gets picked by buildings all the time (note that there were fewer buildings in 2015 than in 2019). Weekend indicator is the next important feature that is selected by more than half of the buildings, however, its importance varies across the years. The data for solar radiation is not available in 2018, which is why it never gets picked up by any building. In contrast, the data for RH is available in other years but it gets excluded from the feature set due to having a high (negative) correlation with temperature. The remaining plots (Figure S12(B) through Figure S12(F)) plots the impact on the model performance of subsequently adding features to the baseline model in each year.

10 Variation in estimated coefficients

Figure S13 identifies the effect that building size has on the estimated model coefficients. As shown in Figure S13, there is no correlation between the building size and the estimated coefficients. Additionally, most small-sized Offices and Classrooms buildings have smaller base load intensities, while the Laboratories and Healthcare buildings are not only bigger in size, but they also consume more cooling at 18°C. Figure S14, Figure S15, and Table S3 supplement the discussion in Section .

Table S3: Quantifying the statistical significance of variation in estimated base load intensity and temperature sensitivity of buildings within a certain category over years.

Type	Temperature Sensitivity				Base Load Intensity			
	Year-1	Year-2	t-statistic	p-value	Year-1	Year-2	t-statistic	p-value
Offices/Classrooms	2015	2016	-3.8555	0.0004	2015	2016	2.0462	.0477
	2016	2017	0.7264	0.4721	2016	2017	-0.6475	0.5212
	2017	2018	-0.6827	0.4989	2017	2018	0.2303	0.8191
	2018	2019	1.7329	0.0912	2018	2019	-0.9440	0.3512
Labs	2015	2016	-4.0312	0.0003	2015	2016	3.7746	0.0006
	2016	2017	2.6208	0.0130	2016	2017	-0.8651	0.3930
	2017	2018	-0.6967	0.4907	2017	2018	1.1162	0.2722
	2018	2019	-1.8529	0.0726	2018	2019	0.7192	0.4769
Health care	2015	2016	-5.8742	0.0004	2015	2016	3.0699	0.0154
	2016	2017	1.9616	0.0854	2016	2017	1.2164	0.2585
	2017	2018	-2.2187	0.0573	2017	2018	2.0117	0.0791
	2018	2019	1.1029	0.3021	2018	2019	-1.7088	0.1259

11 Performance monitoring and anomaly detection

In addition to the implications discussed in Section , our chilled-water use characterization can also be used to detect anomalies in data and to identify the buildings that vary their consumption behavior over years. To demonstrate this, we test the model in year t that has been trained on data from $t-1$. Figure S17 shows that the models perform poorly in some buildings likely due to differences in the availability of data for buildings that were performing well in the previous year. To determine the underlying cause of performance deterioration, we plot the training R^2 value for all buildings in Figure S16(A). The buildings are sorted in order of decreasing training R^2 that has been averaged over five years. Note that the white space indicates that the data is either missing or insufficient to train the model. Figure S16(B) plots the R^2 of models that are trained in t year and tested on data for $t+1$ year. As displayed, some buildings exhibit poor testing R^2, even though they exhibit a good training R^2 in t and $t+1$ years. As an example consider three buildings shown in Figure S17.

Figure S17(A) shows the measured cooling consumption in years 2017 and 2018, while Figure S17(B) shows the cooling consumed in 2018 and 2019. The blue and orange colored regression lines represents the modeled load during the training and testing periods. The building B-57 is a Laboratories building and has a good training and testing R^2 in both years. The remaining two buildings exhibit a good training R^2 but have a poor testing R^2 due to changes in consumption pattern and anomalous data. For instance, B-127 is a Healthcare building that dropped its mean daily consumption in 2018 by about 21% but then returned to its usual consumption level in 2019. As a result, the models that are trained on $t-1$ year result in a poor fit on t-year data. In contrast, B-40 uses temperature and RH to train the model in 2017, while in subsequent years, the building starts to follow weekly schedule and the absence of weekend indicator from the model results in poor testing R^2.

References

[1] A. Kavousian, R. Rajagopal, and M. Fischer, "Determinants of residential electricity consumption: Using smart meter data to examine the effect of climate, building characteristics, appliance stock, and occupants' behavior," *Energy*, vol. 55, pp. 184–194, 2013.

[2] A. Nilsson, M. Wester, D. Lazarevic, and N. Brandt, "Smart homes, home energy management systems and real-time feedback: Lessons for influencing household energy consumption from a swedish field study," *Energy and Buildings*, vol. 179, pp. 15–25, 2018.

[3] C. E. Kontokosta and R. K. Jain, "Modeling the determinants of large-scale building water use: Implications for data-driven urban sustainability policy," *Sustainable Cities and Society*, vol. 18, pp. 44–55, 2015.

[4] A. Albert and R. Rajagopal, "Smart meter driven segmentation: What your consumption says about you," *IEEE Transactions on power systems*, vol. 28, no. 4, pp. 4019–4030, 2013.

[5] A. Satre-Meloy, M. Diakonova, and P. Grünewald, "Cluster analysis and prediction of residential peak demand profiles using occupant activity data," *Applied Energy*, vol. 260, p. 114246, 2020.

[6] J. P. Real, J. K. Møller, R. Li, and H. Madsen, "A data-driven framework for characterising building archetypes: A mixed effects modelling approach," *Energy*, p. 124278, 2022.

[7] C. Miller, "More buildings make more generalizable models—benchmarking prediction methods on open electrical meter data," *Machine Learning and Knowledge Extraction*, vol. 1, no. 3, pp. 974–993, 2019.

[8] X. Li, Y. Zhou, S. Yu, G. Jia, H. Li, and W. Li, "Urban heat island impacts on building energy consumption: A review of approaches and findings," *Energy*, vol. 174, pp. 407–419, 2019.

[9] J. Woods, N. James, E. Kozubal, E. Bonnema, K. Brief, L. Voeller, and J. Rivest, "Humidity's impact on greenhouse gas emissions from air conditioning," *Joule*, vol. 6, no. 4, pp. 726–741, 2022.

[10] M. Heidarinejad, J. G. Cedeño-Laurent, J. R. Wentz, N. M. Rekstad, J. D. Spengler, and J. Srebric, "Actual building energy use patterns and their implications for predictive modeling," *Energy Conversion and Management*, vol. 144, pp. 164–180, 2017.

[11] D. L. Villa, "Institutional heat wave analysis by building energy modeling fleet and meter data," *Energy and Buildings*, vol. 237, p. 110774, 2021.

[12] A. Nutkiewicz, Z. Yang, and R. K. Jain, "Data-driven urban energy simulation (due-s): A framework for integrating engineering simulation and machine learning methods in a multi-scale urban energy modeling workflow," *Applied energy*, vol. 225, pp. 1176–1189, 2018.

[13] S. D. Nazemi, E. Zaidan, and M. A. Jafari, "The impact of occupancy-driven models on cooling systems in commercial buildings," *Energies*, vol. 14, no. 6, 2021.

[14] S. Garimella, K. Lockyear, D. Pharis, O. El Chawa, M. T. Hughes, and G. Kini, "Realistic pathways to decarbonization of building energy systems," *Joule*, vol. 6, no. 5, pp. 956–971, 2022.

[15] C. Miller, A. Kathirgamanathan, B. Picchetti, P. Arjunan, J. Y. Park, Z. Nagy, P. Raftery, B. W. Hobson, Z. Shi, and F. Meggers, "The building data genome project 2, energy meter data from the ashrae great energy predictor iii competition," *Scientific data*, vol. 7, no. 1, pp. 1–13, 2020.

[16] N. Luo, Z. Wang, D. Blum, C. Weyandt, N. Bourassa, M. A. Piette, and T. Hong, "A three-year dataset supporting research on building energy management and occupancy analytics," *Scientific Data*, vol. 9, no. 1, pp. 1–15, 2022.

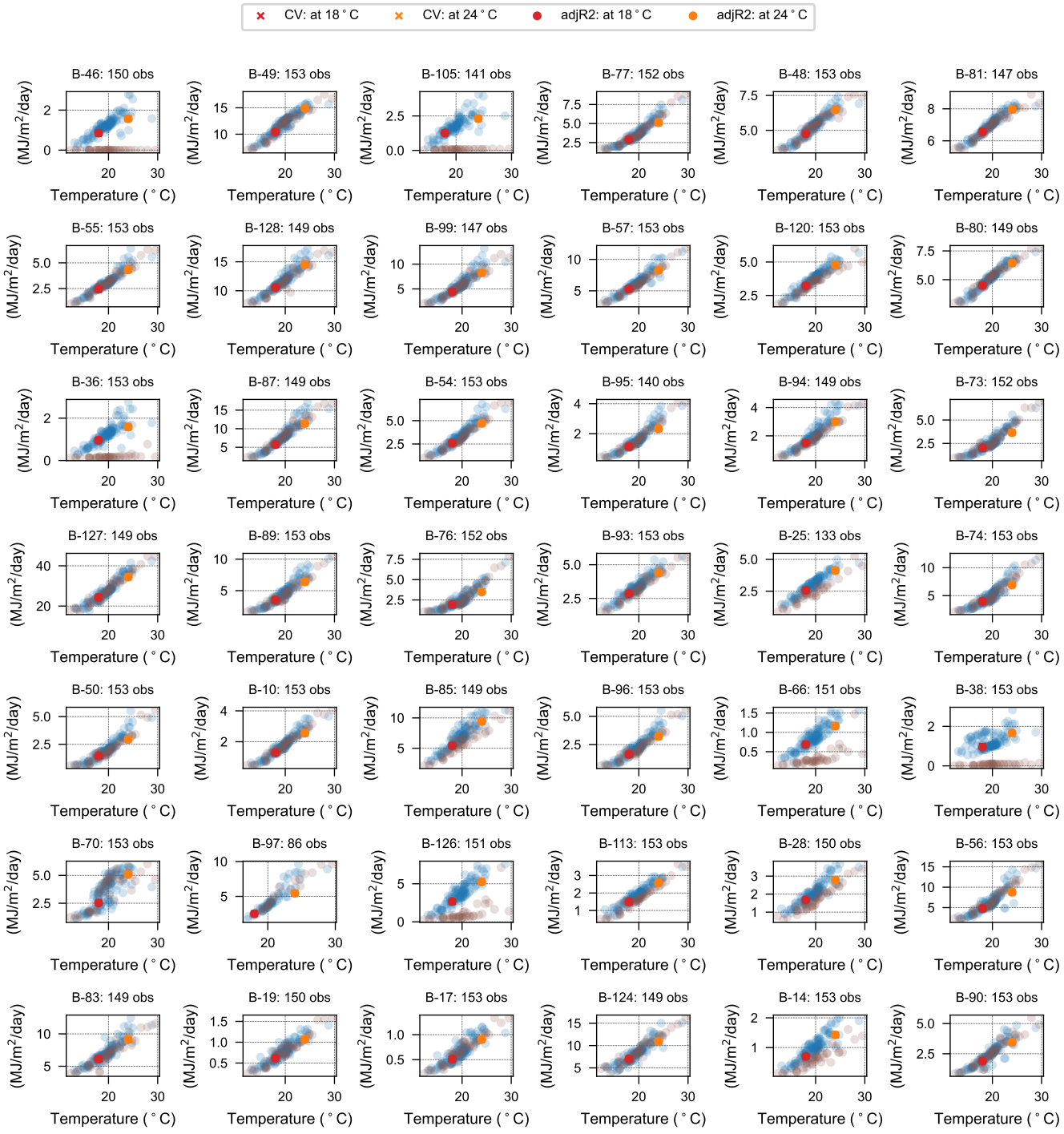

Figure S6: Impact of the model selection criterion on estimated base load intensity. Note that the base load intensities in MJ/m^2/day are plotted on y-axis and mean daily outside air temperatures in °C are plotted on x-axis.

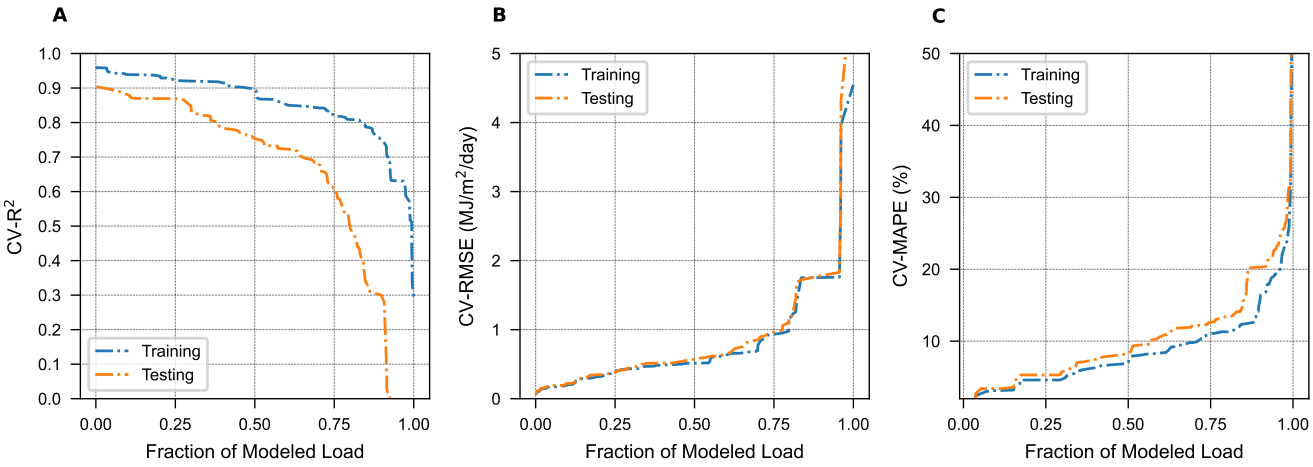

Figure S7: Cross validation (CV) errors. (A) Cross validation coefficient of determination (R^2) against fraction of modeled load. (B) Cross validation Root Mean Squared Error (RMSE) against fraction of modeled load. (C) Cross validation Mean Absolute Percentage Error (MAPE) against fraction of modeled load.

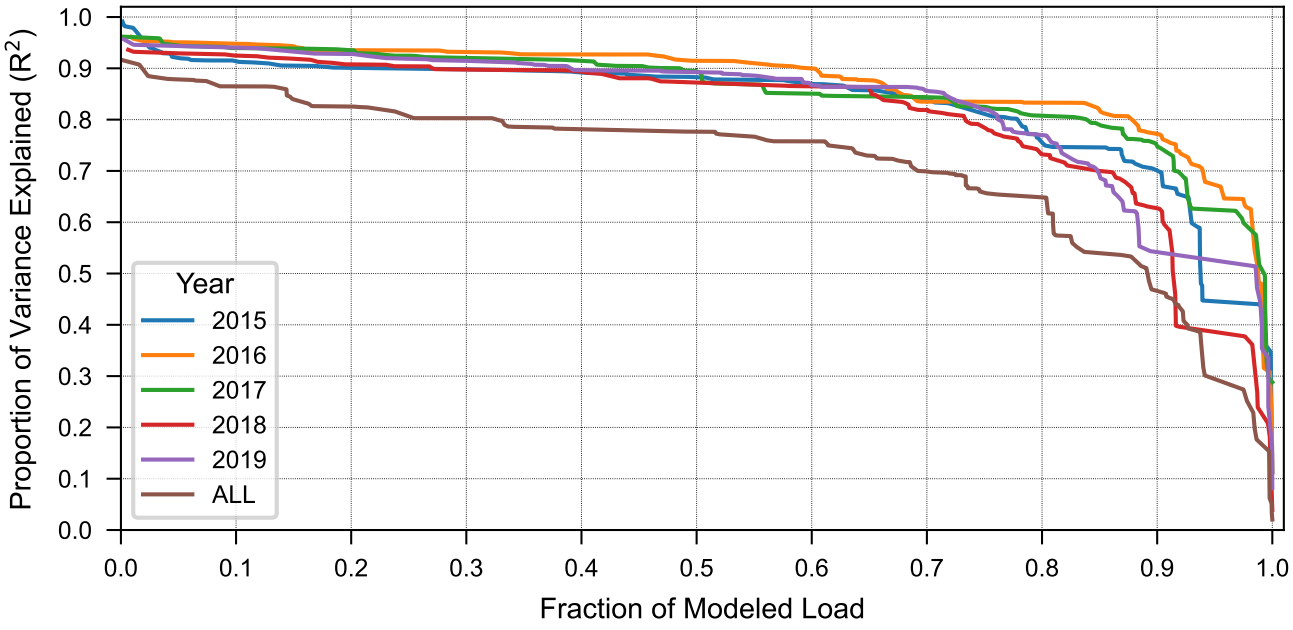

Figure S8: Performance of models in different years. 'ALL' represents the model that has been trained using all five years of data.

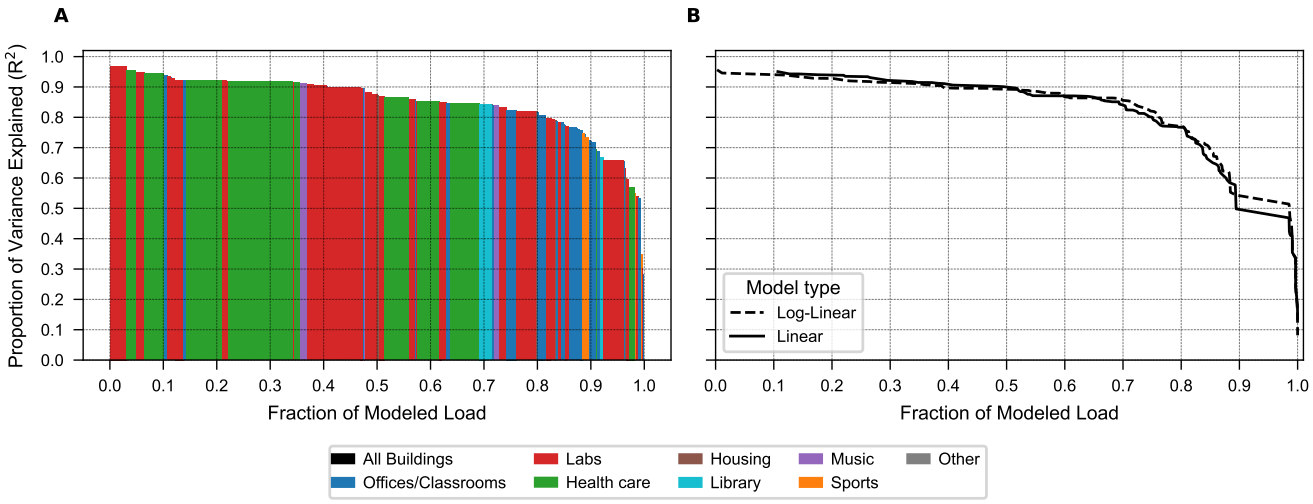

Figure S9: Comparison of Linear and Log-linear Models. (A) Linear Model: Proportion of variance explained as described by the coefficient of determination (R^2), plotted against the fraction of modeled load for year 2017. (B) Comparing the proportion of variance explained by Linear and Log-linear models.

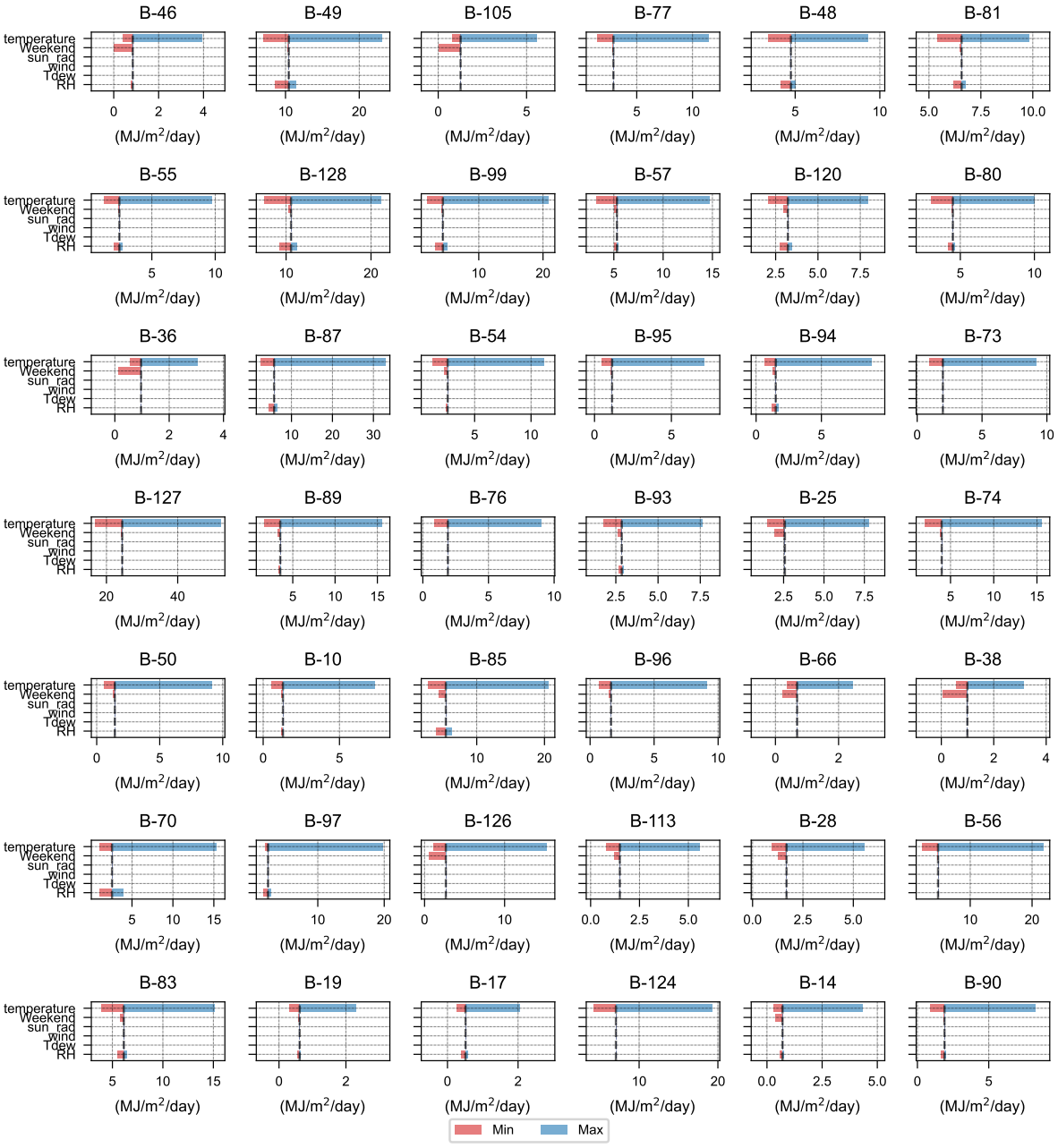

Figure S10: Feature sensitivity test for the top 42 buildings with the highest R^2 in 2017. The x-axis on each subplot shows the modeled daily cooling consumption, while the y-axis shows different features. The vertical dashed line represents the base case, which is define as the base load intensity at 18°C and average values of other features on a weekday. The max-min range corresponds to the maximum and minimum limits of each feature.

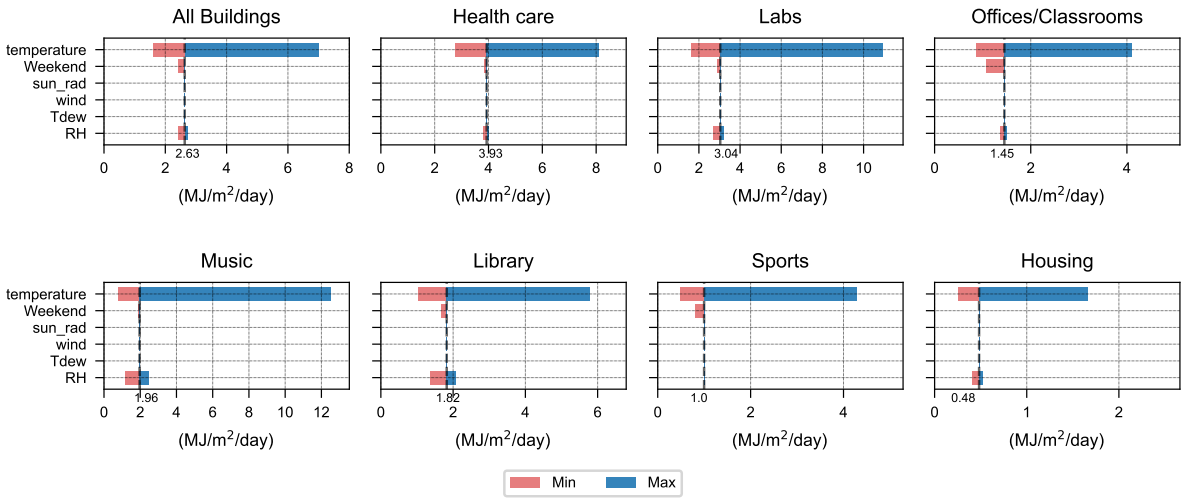

Figure S11: Feature sensitivity test of buildings at the portfolio-scale and building category-level aggregates in 2017. The x-axis on each subplot shows the daily cooling consumption, while the y-axis shows different features. The vertical dashed line represents the base case, which is define as the base load intensity at 18°C and average values of other features on a weekday. The max-min range corresponds to the maximum and minimum limits of each feature.

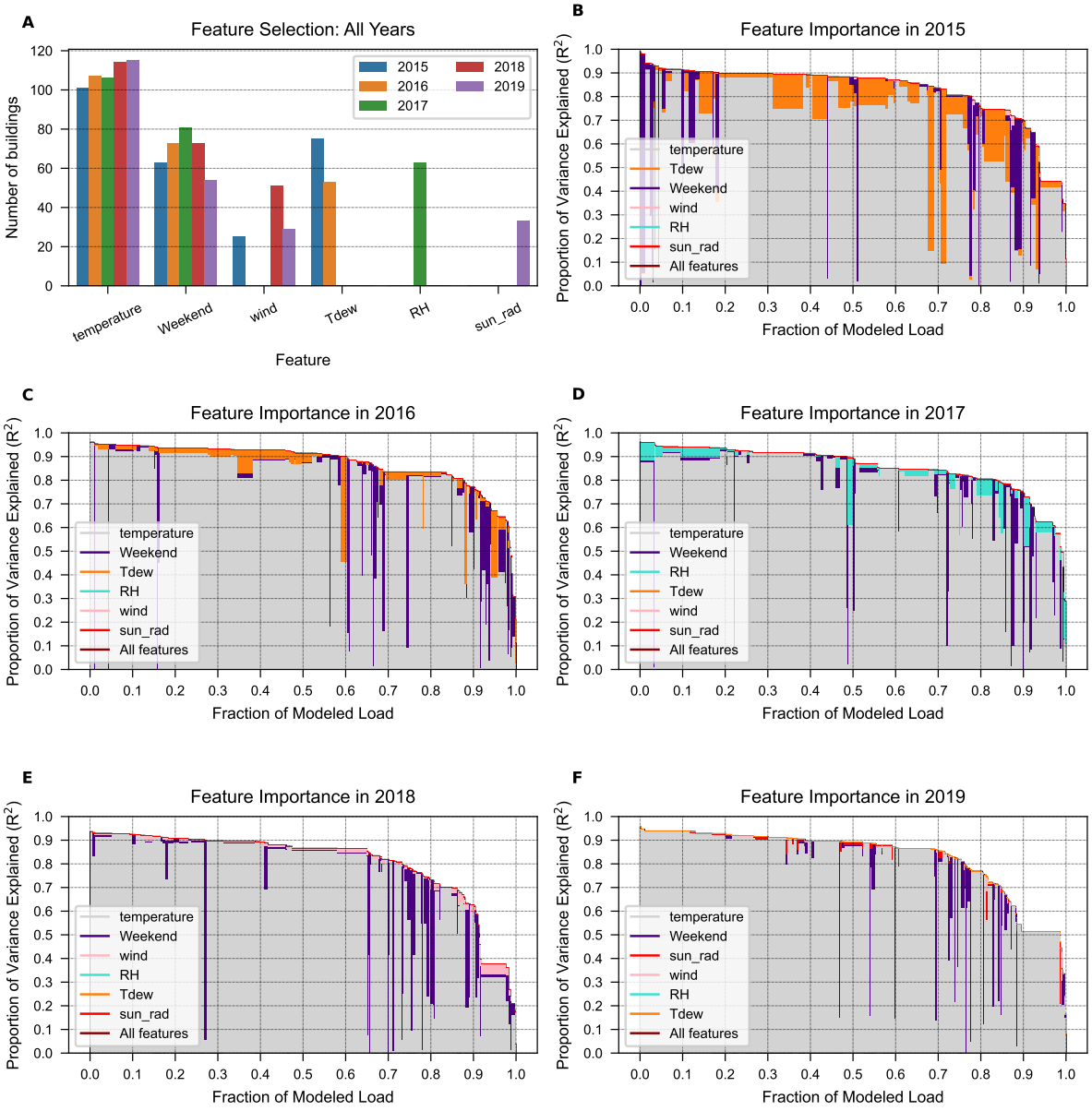

Figure S12: Significance of features across all years. (A) A bar plot showing the number of buildings that select a given feature in all years. (B) - (F) show the improvement in the model performance when features are sequentially added to the baseline model in 2015 through 2019.

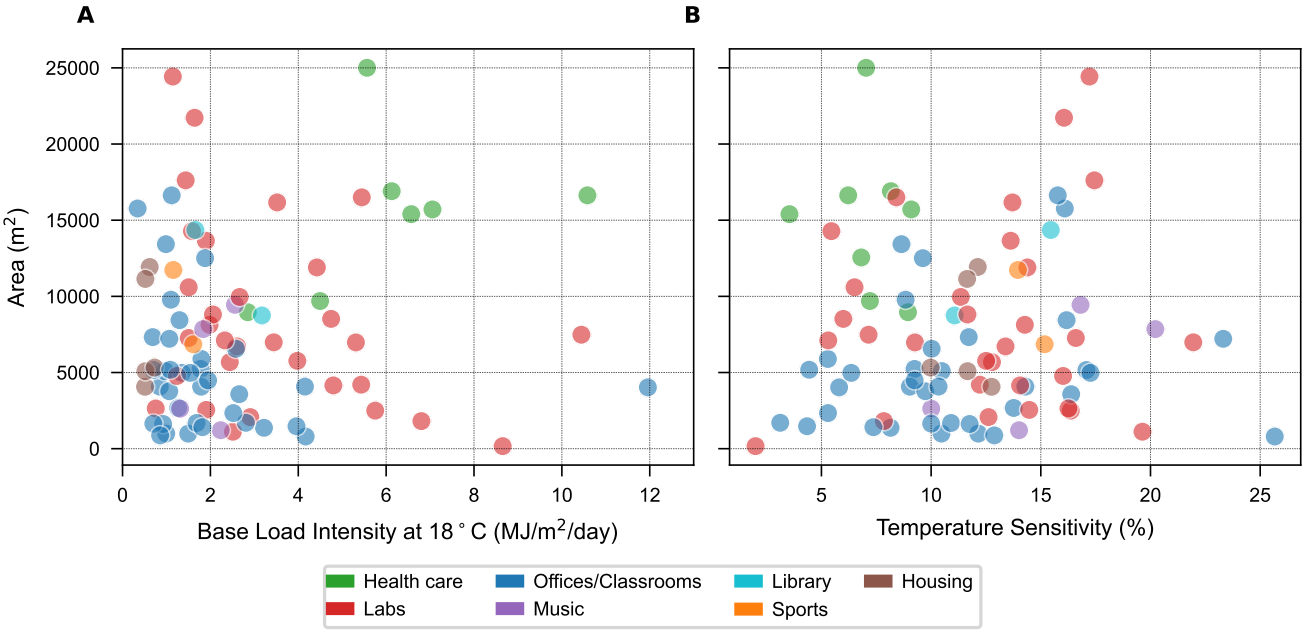

Figure S13: Relationship between the building size and estimated coefficients. The plots show area of buildings against their (A) Base load intensities. (B) Temperature Sensitivities.

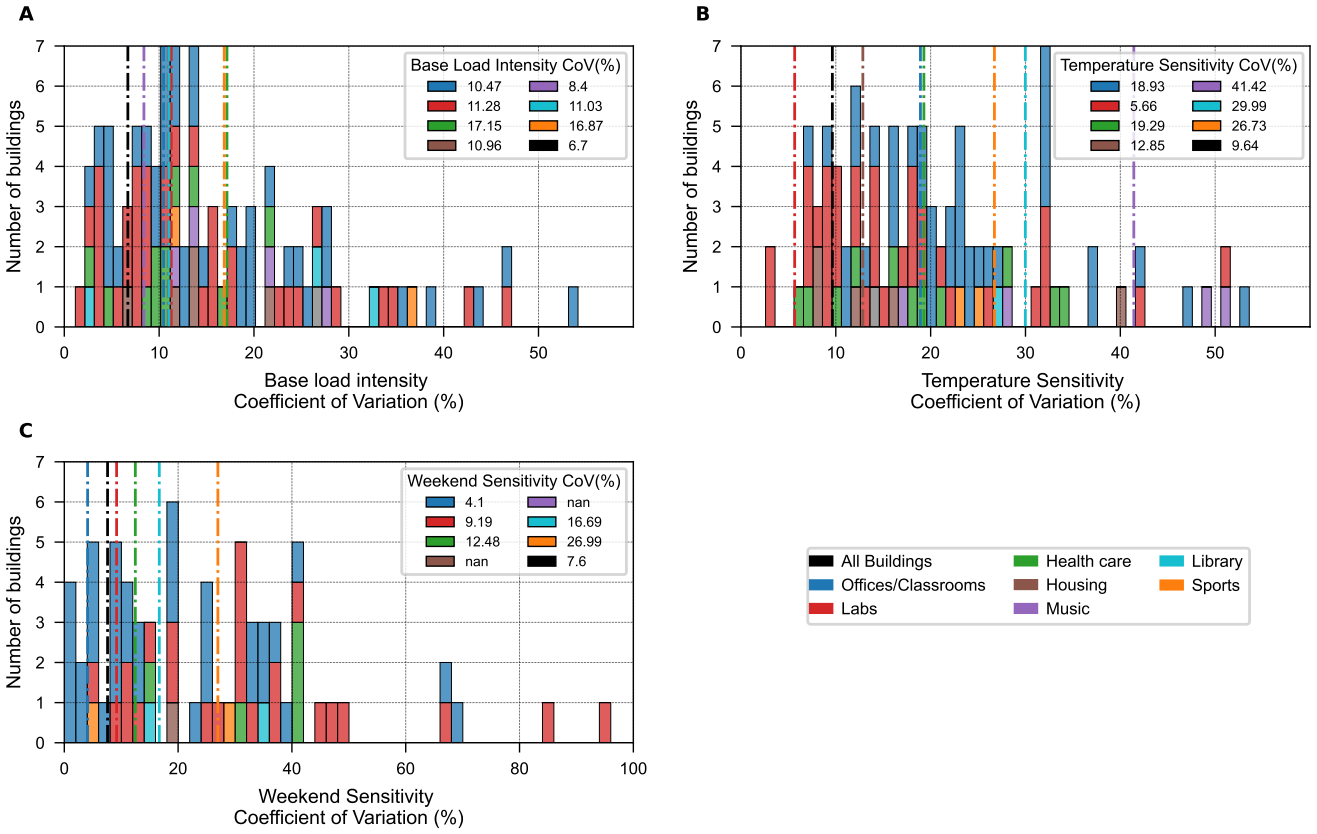

Figure S14: Year-to-year variability of model coefficient estimates. Histograms of Coefficient of Variation (CoV) computed for model coefficients across different years (2015 through 2019). The vertical dashed lines denote CoV of buildings in different categories. (A) Base load intensity. (B) Temperature sensitivity. (C) Weekend sensitivity.

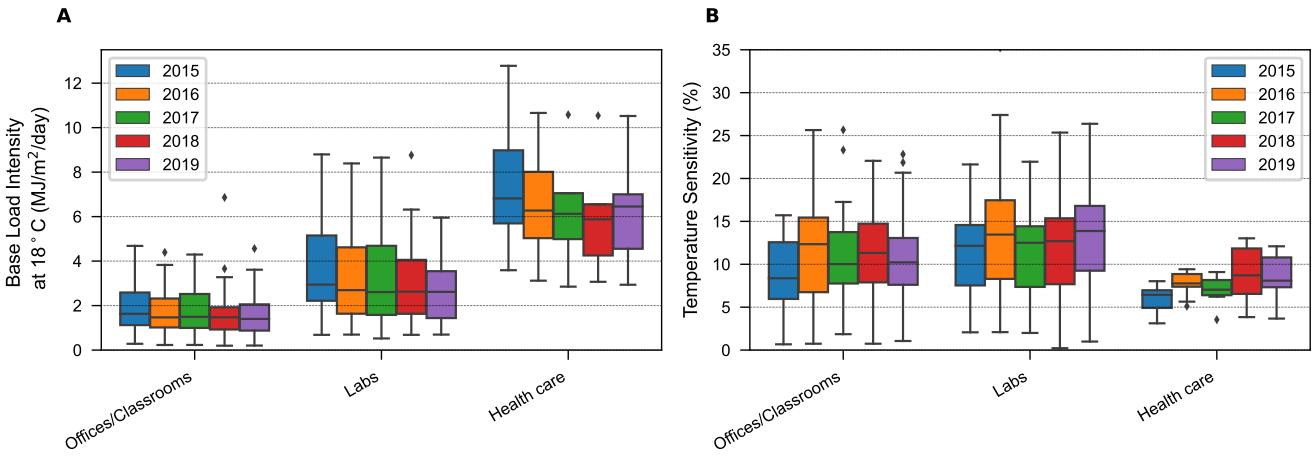

Figure S15: Spread of estimated model coefficients of buildings in different categories that get trained in all years. Boxplots show the variation in model coefficients of buildings in different years. (A) Base load intensity. (B) Temperature sensitivity.

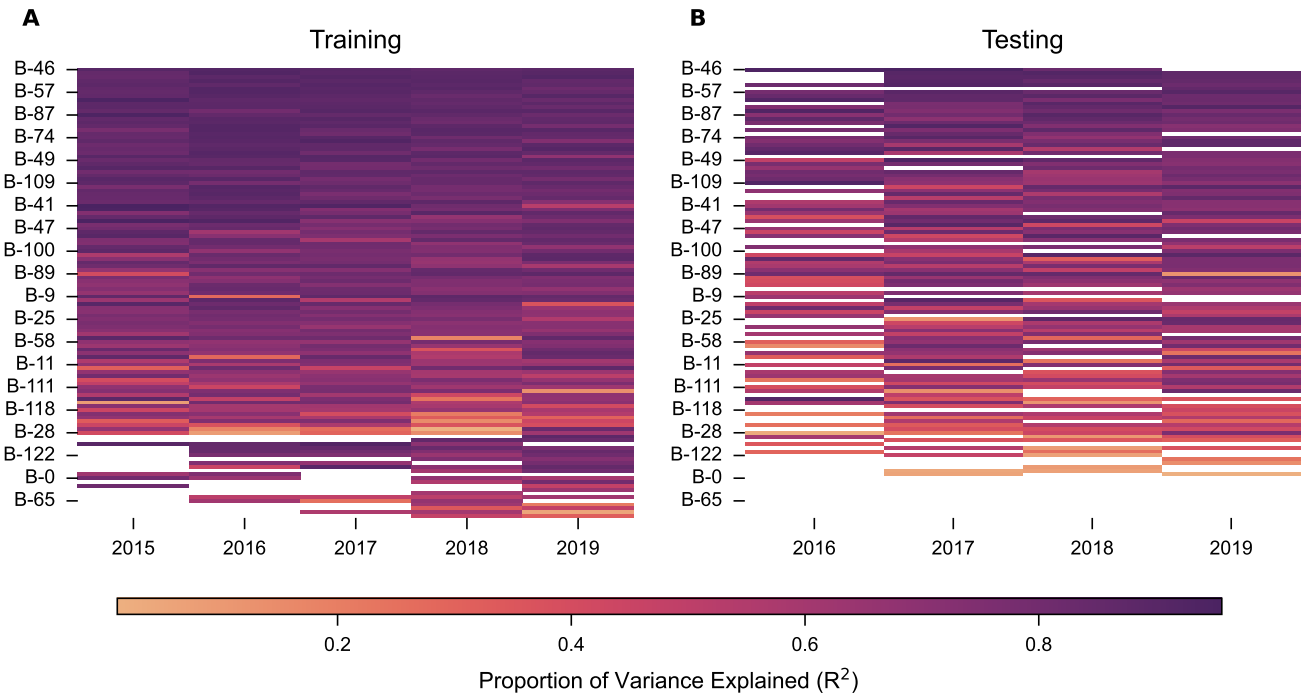

Figure S16: Performance of models in different years. Note that each row indicates a different building, but the labels are shown only for every fifth building to avoid overlapping, while columns represent different years. The color is indicative of how well the model explains the variance in load for that year. (A) Performance of models trained on data for t year. (B) Performance of models in (A) tested on data for $t + 1$ year.

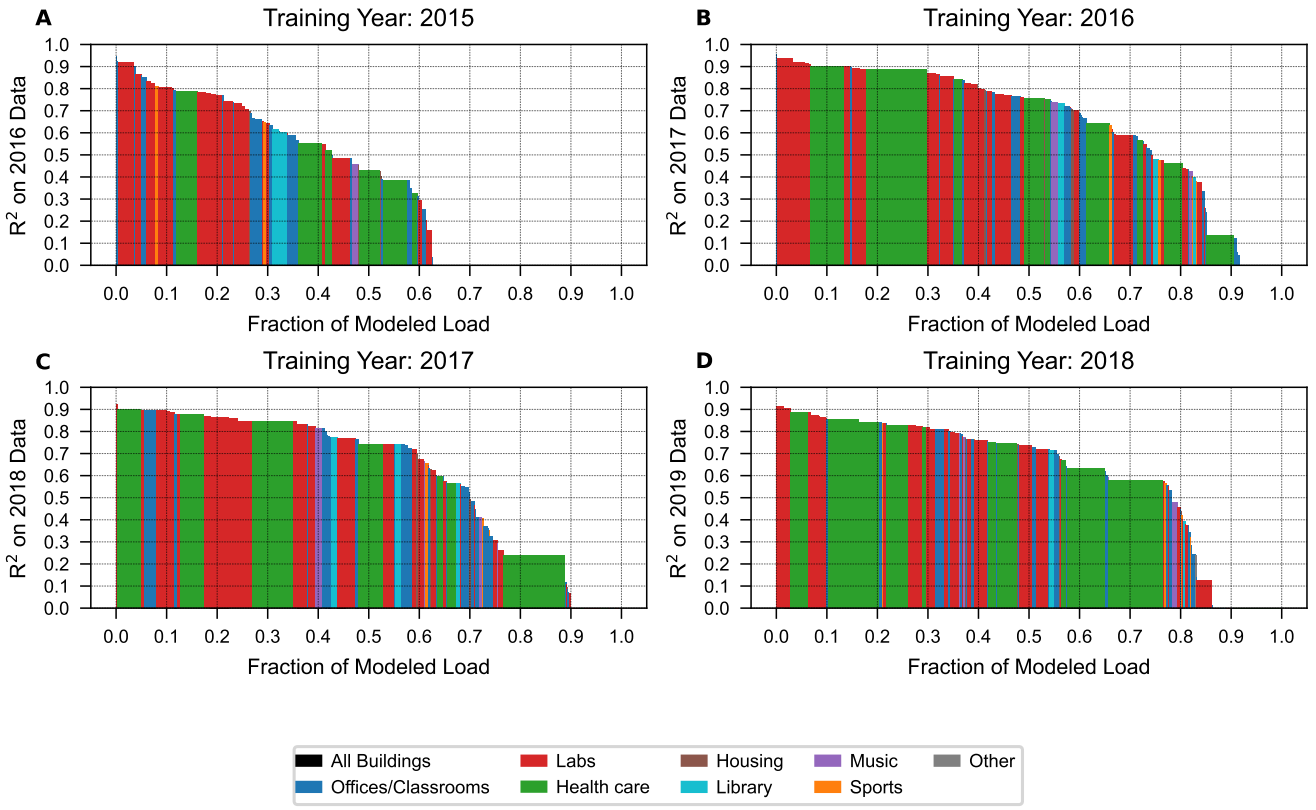

Figure S17: Performance of models trained in $t-1$ year and tested on data for t year. (A) Test year: 2016. (B) Test year: 2017. (C) Test year: 2018. (D) Test year: 2019. Note that the tail end of these charts is not plotted because the respective models have a negative test R^2.

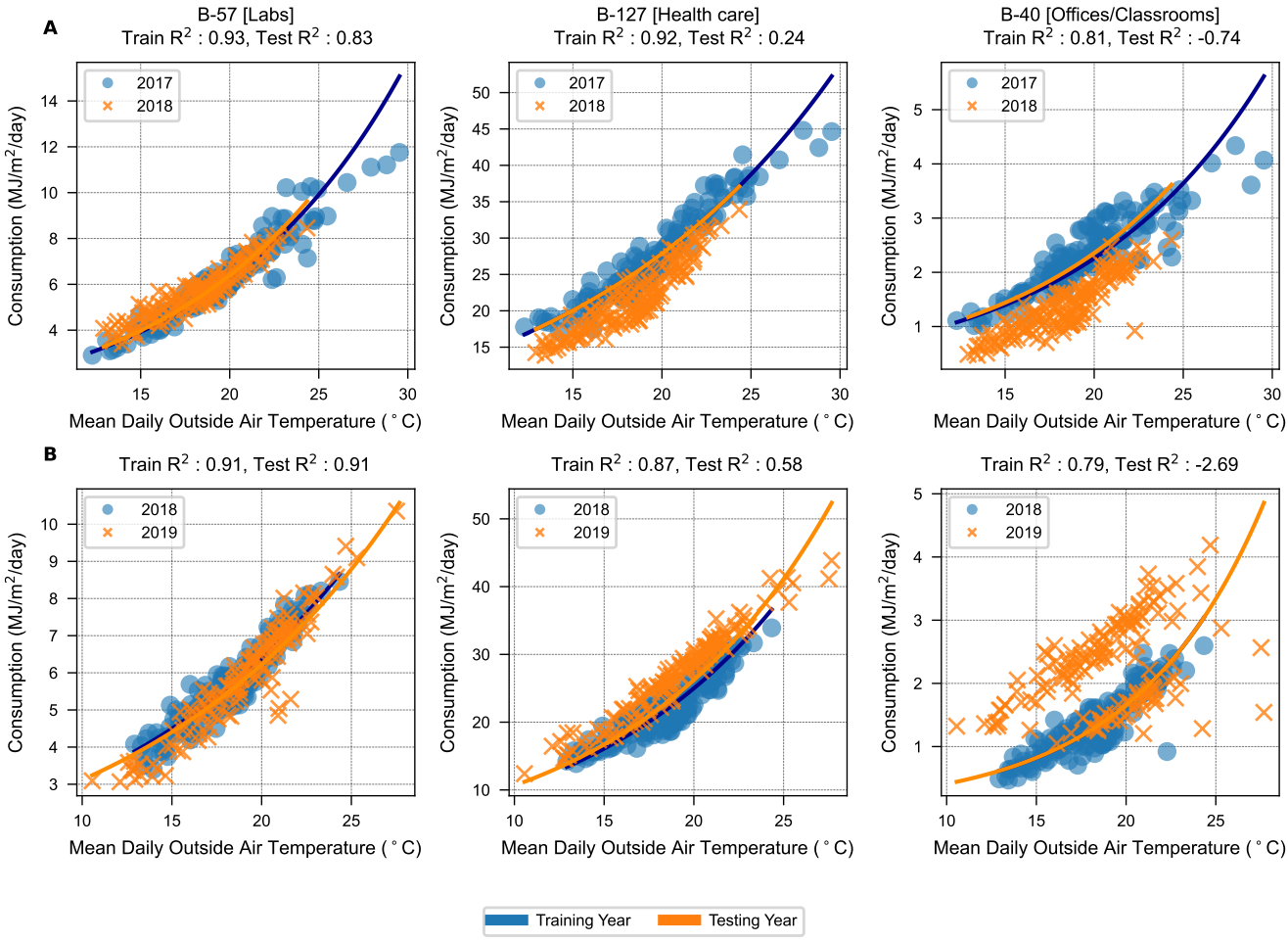

Figure S18: Measured cooling consumption of three buildings against the mean daily outside air temperature in 2017 through 2019. (A) Models trained in 2017 and tested on data measured in 2018. (B) Models trained in 2018 and tested on data measured in 2019.