Disentangling the Evolving Effects of Climate Change and Technology Diffusion on Building Electricity Demand

Mai Shi[a*], Ziqi Wei[b], Michael T. Craig[a,c*]

[a] *School for Environment and Sustainability, University of Michigan, Ann Arbor, MI, USA.*

[b] *Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA.*

[c] *Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI, USA.*

Supplementary Methods

Supplementary M1 Methodology framework

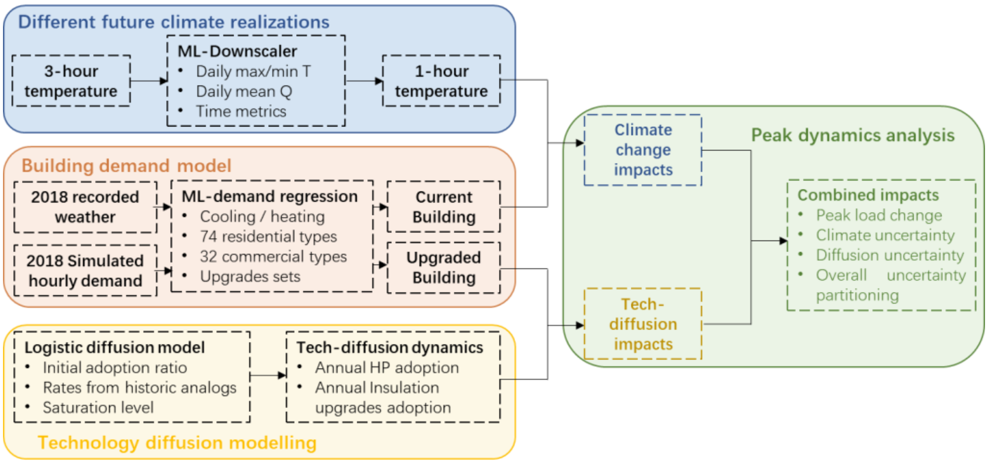

Figure S1 Methodology Framework of this study. It offers supplementary information for Methods in main text.

Supplementary M2 Supplementary information on building energy demand model

Supplementary M2.1 Introduction of input trained data

We use the individual building information and their simulated power consumption profiles in ResStock[1] and ComStock[2] database, for residential and commercial building loads training and prediction. ResStock contains simulations of 550,000 statistically representative dwelling units, covering all single-family and multifamily housing across a wide variety of climates, housing characteristics, and occupant behavior. ComStock contains simulation of 350,000 statistically representative commercial building stocks, covering a wide range of building types which roughly cover 64% total primary energy consumption of total commercial buildings. The distribution of household characteristics of buildings in ResStock and ComStock are determined by the survey datasets from Residential Energy Consumption Survey (RECS) and Commercial Building Energy Consumption Survey (CBECS) by EIA, respectively. Therefore, we here recognize the aggregated results of all sampled buildings in each city representative for the actual building demand. In this study, we specifically select 111 typical counties, covering the rage of different states, weather zones and power grid regions, where 178,587 residential and 122,006 commercial buildings are included. The map and detailed information for each county are detailed below in Figure S2 and Table S1.

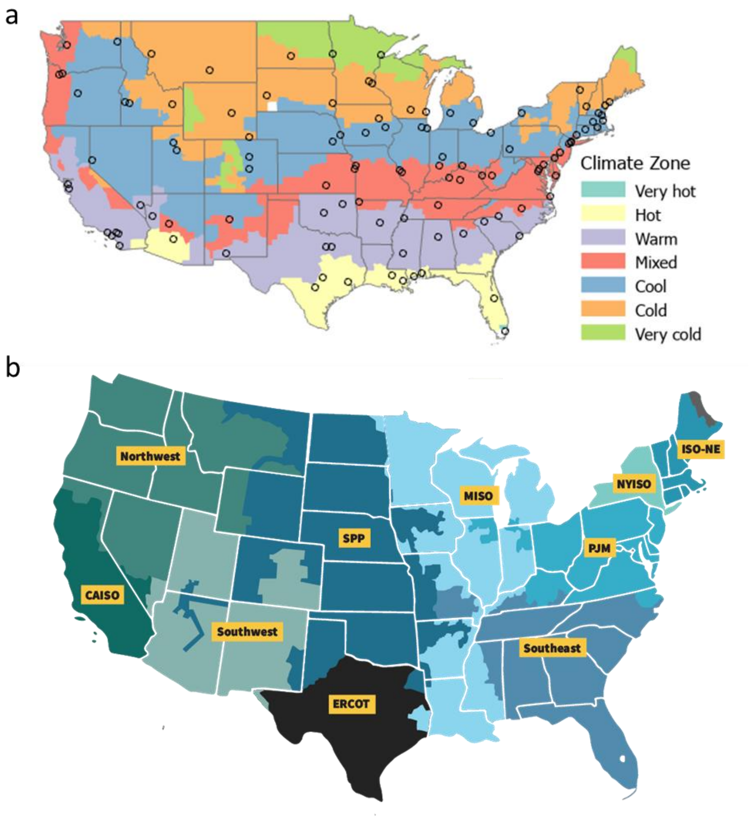

Figure S2 Map of selected counties (a) and power grid regions (b) in this study

Table S1 Detailed information of select counties

County Name	State	Grid	Weather	Res_number	Com_number
Jefferson County	AL	Southeast	3A	1259	1127
Mobile County	AL	Southeast	2A	750	671
Maricopa County	AZ	Southwest	2B	6938	3674
Mohave County	AZ	Southwest	3B	462	178
Yavapai County	AZ	Southwest	4B	465	199
Benton County	AR	MISO	4A	405	286
Pulaski County	AR	MISO	3A	748	869
Alameda County	CA	CAISO	3C	2443	1877
Los Angeles County	CA	CAISO	3B	14399	10370
Orange County	CA	CAISO	3B	4427	3159
Riverside County	CA	CAISO	3B	3386	1833
San Bernardino County	CA	CAISO	3B	2922	2267
San Diego County	CA	CAISO	3B	4898	2968

Santa Clara County	CA	CAISO	3C	2692	1852
Denver County	CO	Southwest	5B	1235	963
El Paso County	CO	Southwest	5B	1096	785
Fairfield County	CT	ISO-NE	5A	1505	992
Hartford County	CT	ISO-NE	5A	1546	1155
New Castle County	DE	PJM	4A	910	500
Sussex County	DE	PJM	4A	534	167
District of Columbia	DC	PJM	4A	1265	773
Miami-Dade County	FL	Southeast	1A	4146	2598
Orange County	FL	Southeast	2A	2100	1423
DeKalb County	GA	Southeast	3A	1269	794
Fulton County	GA	Southeast	3A	1856	1381
Ada County	ID	Northwest	5B	691	660
Bonneville County	ID	Northwest	6B	168	132
Canyon County	ID	Northwest	5B	295	266
Cook County	IL	MISO	5A	8985	4982
DuPage County	IL	MISO	5A	1478	1019
Madison County	IL	MISO	4A	488	288
Lake County	IN	MISO	5A	869	516
Marion County	IN	MISO	5A	1732	1424
Monroe County	IN	MISO	4A	247	122
Black Hawk County	IA	MISO	6A	235	131
Linn County	IA	MISO	5A	391	275
Polk County	IA	MISO	5A	787	512
Johnson County	KS	SPP	4A	962	619
Sedgwick County	KS	SPP	4A	887	773
Fayette County	KY	PJM	4A	573	412
Jefferson County	KY	PJM	4A	1411	1221
East Baton Rouge Parish	LA	MISO	2A	791	675
Orleans Parish	LA	MISO	2A	794	746
Cumberland County	ME	ISO-NE	6A	580	457
York County	ME	ISO-NE	6A	442	312
Baltimore County	MD	PJM	4A	1390	801
Montgomery County	MD	PJM	4A	1592	535
Middlesex County	MA	ISO-NE	5A	2557	1436
Worcester County	MA	ISO-NE	5A	1360	731
Kent County	MI	MISO	5A	1027	954
Wayne County	MI	MISO	5A	3367	2275
Hennepin County	MN	MISO	6A	2147	1093
Ramsey County	MN	MISO	6A	900	529
St. Louis County	MN	MISO	7A	425	251
Harrison County	MS	MISO	2A	370	212
Hinds County	MS	MISO	3A	430	393
Jackson County	MO	SPP	4A	1302	1115
St. Louis County	MO	MISO	4A	1810	1112
Missoula County	MT	Northwest	6B	212	115
Yellowstone County	MT	Northwest	6B	275	183
Douglas County	NE	SPP	5A	938	703

Lancaster County	NE	SPP	5A	517	412
Clark County	NV	Southwest	3B	3566	1783
Washoe County	NV	Northwest	5B	774	588
Hillsborough County	NH	ISO-NE	5A	690	441
Rockingham County	NH	ISO-NE	5A	531	446
Bergen County	NJ	PJM	5A	1464	1122
Essex County	NJ	PJM	4A	1297	1073
Bernalillo County	NM	Southwest	4B	1189	1059
Dona Ana County	NM	Southwest	3B	348	185
Erie County	NY	NYISO	5A	1739	920
New York County	NY	NYISO	4A	3578	1137
Mecklenburg County	NC	Southeast	3A	1732	1397
Wake County	NC	Southeast	4A	1657	1068
Burleigh County	ND	SPP	6A	163	87
Cass County	ND	SPP	7A	311	181
Cuyahoga County	OH	PJM	5A	2552	1586
Hamilton County	OH	PJM	4A	1557	1314
Oklahoma County	OK	SPP	3A	1363	1527
Tulsa County	OK	SPP	3A	1142	1148
Deschutes County	OR	Northwest	5B	342	228
Multnomah County	OR	Northwest	4C	1371	1291
Washington County	OR	Northwest	4C	907	548
Allegheny County	PA	PJM	5A	2432	2023
Philadelphia County	PA	PJM	4A	2768	1924
Providence County	RI	ISO-NE	5A	1088	908
Greenville County	SC	Southeast	3A	831	955
Horry County	SC	Southeast	3A	800	644
Minnehaha County	SD	SPP	6A	314	159
Pennington County	SD	SPP	6A	193	110
Davidson County	TN	Southeast	4A	1217	1173
Shelby County	TN	Southeast	3A	1664	1355
Bexar County	TX	ERCOT	2A	2803	2005
Dallas County	TX	ERCOT	3A	4028	3238
Harris County	TX	ERCOT	2A	6967	4862
Tarrant County	TX	ERCOT	3A	3055	2328
Travis County	TX	ERCOT	2A	1961	1421
Salt Lake County	UT	Southwest	5B	1555	1364
Utah County	UT	Southwest	5B	654	483
Chittenden County	VT	ISO-NE	6A	279	161
Windsor County	VT	ISO-NE	6A	141	36
Fairfax County	VA	PJM	4A	1697	551
Virginia Beach city	VA	PJM	4A	752	417
King County	WA	Northwest	4C	3644	2036
Spokane County	WA	Northwest	5B	856	831
Cabell County	WV	PJM	4A	192	199
Kanawha County	WV	PJM	4A	381	377
Dane County	WI	MISO	6A	920	635
Milwaukee County	WI	MISO	6A	1722	1251
Laramie County	WY	Northwest	6B	172	87

Natrona County	WY	Northwest	6B	147	91

In this study, we aggregate the cooling/heating consumption for buildings which share the same cooling, heating and insulation type in each city, and building machine-learning model especially for each type of building. Specifically, we aggregate the current stock of residential and commercial buildings into 76 and 32 types, detailed in Table S2 and Table S3.

Table S2 Aggregated type of all residential buildings

Heating	Cooling	Insulation upgraded
ASHP	ASHP	FALSE
ASHP	ASHP	TRUE
Electric Baseboard	AC	FALSE
Electric Baseboard	AC	TRUE
Electric Baseboard	None	FALSE
Electric Baseboard	None	TRUE
Electric Baseboard	Room AC	FALSE
Electric Baseboard	Room AC	TRUE
Electric Baseboard	Shared Cooling	FALSE
Electric Baseboard	Shared Cooling	TRUE
Electric Boiler	AC	FALSE
Electric Boiler	AC	TRUE
Electric Boiler	None	FALSE
Electric Boiler	None	TRUE
Electric Boiler	Room AC	FALSE
Electric Boiler	Room AC	TRUE
Electric Boiler	Shared Cooling	FALSE
Electric Boiler	Shared Cooling	TRUE
Electric Furnace	AC	FALSE
Electric Furnace	AC	TRUE
Electric Furnace	None	FALSE
Electric Furnace	None	TRUE
Electric Furnace	Room AC	FALSE
Electric Furnace	Room AC	TRUE
Electric Furnace	Shared Cooling	FALSE
Electric Furnace	Shared Cooling	TRUE
Electric Wall Furnace	AC	FALSE
Electric Wall Furnace	AC	TRUE
Electric Wall Furnace	None	FALSE
Electric Wall Furnace	None	TRUE
Electric Wall Furnace	Room AC	FALSE
Electric Wall Furnace	Room AC	TRUE
Electric Wall Furnace	Shared Cooling	FALSE
Electric Wall Furnace	Shared Cooling	TRUE
Fuel Boiler	AC	FALSE
Fuel Boiler	AC	TRUE
Fuel Boiler	None	FALSE
Fuel Boiler	None	TRUE
Fuel Boiler	Room AC	FALSE
Fuel Boiler	Room AC	TRUE

Fuel Boiler	Shared Cooling	FALSE
Fuel Boiler	Shared Cooling	TRUE
Fuel Furnace	AC	FALSE
Fuel Furnace	AC	TRUE
Fuel Furnace	None	FALSE
Fuel Furnace	None	TRUE
Fuel Furnace	Room AC	FALSE
Fuel Furnace	Room AC	TRUE
Fuel Furnace	Shared Cooling	FALSE
Fuel Furnace	Shared Cooling	TRUE
Fuel Wall/Floor Furnace	AC	FALSE
Fuel Wall/Floor Furnace	AC	TRUE
Fuel Wall/Floor Furnace	None	FALSE
Fuel Wall/Floor Furnace	None	TRUE
Fuel Wall/Floor Furnace	Room AC	FALSE
Fuel Wall/Floor Furnace	Room AC	TRUE
Fuel Wall/Floor Furnace	Shared Cooling	FALSE
Fuel Wall/Floor Furnace	Shared Cooling	TRUE
MSHP	MSHP	FALSE
MSHP	MSHP	TRUE
None	AC	FALSE
None	AC	TRUE
None	None	FALSE
None	None	TRUE
None	Room AC	FALSE
None	Room AC	TRUE
None	Shared Cooling	FALSE
None	Shared Cooling	TRUE
Shared Heating	AC	FALSE
Shared Heating	AC	TRUE
Shared Heating	None	FALSE
Shared Heating	None	TRUE
Shared Heating	Room AC	FALSE
Shared Heating	Room AC	TRUE
Shared Heating	Shared Cooling	FALSE
Shared Heating	Shared Cooling	TRUE

Table S3 Aggregated type of all commercial buildings

Heating	Cooling	Insulation upgraded
District Heating	ACC	FALSE
District Heating	DX	FALSE
District Heating	District	FALSE
District Heating	WCC	FALSE
ASHP	ASHP	FALSE
Electric Resistance	ACC	FALSE
Electric Resistance	DX	FALSE
Electric Resistance	District	FALSE
Electric Resistance	Evaporative Cooling	FALSE
Electric Resistance	None	FALSE

Electric Resistance	WCC	FALSE
Electricity Furnace	DX	FALSE
GSHP	GSHP	FALSE
WSHP	WSHP	FALSE
FuelOil Boiler	ACC	FALSE
FuelOil Boiler	DX	FALSE
FuelOil Boiler	WCC	FALSE
FuelOil Furnace	DX	FALSE
FuelOil Furnace	Evaporative Cooling	FALSE
FuelOil Furnace	None	FALSE
NaturalGas Boiler	ACC	FALSE
NaturalGas Boiler	DX	FALSE
NaturalGas Boiler	District	FALSE
NaturalGas Boiler	WCC	FALSE
NaturalGas Furnace	DX	FALSE
NaturalGas Furnace	Evaporative Cooling	FALSE
NaturalGas Furnace	None	FALSE
Propane Boiler	DX	FALSE
Propane Boiler	WCC	FALSE
Propane Furnace	DX	FALSE
Propane Furnace	Evaporative Cooling	FALSE
Propane Furnace	None	FALSE

Besides current baseline building stocks, ResStock and ComStock datasets also performs building simulation for each building under a series of different future technology uptakes. In this study, we select 8 different packages and 4 different packages in ResStock and ComStock, respectively, which are related to adoption of heat pump and insulation upgrades.

Table S4 Technology upgrade packages in ResStock and ComStock used in this study

Building	Version	Package
ResStock	2024.2	Baseline
ResStock	2024.2	Envelope Only - Light Touch Envelope
ResStock	2024.2	ENERGY STAR Air-to-Air Heat Pump With Electric Backup
ResStock	2024.2	High Efficiency Cold-Climate Air-to-Air Heat Pump With Electric Backup
ResStock	2024.2	Ultra High Efficiency Air-to-Air Heat Pump With Electric Backup
ResStock	2024.2	ENERGY STAR Air-to-Air Heat Pump With Electric Backup + Light Touch Envelope
ResStock	2024.2	High Efficiency Cold-Climate Air-to-Air Heat Pump With Electric Backup + Light Touch Envelope
ResStock	2024.2	Ultra High Efficiency Air-to-Air Heat Pump With Electric Backup + Light Touch Envelope
ComStock	2023	Baseline
ComStock	2023	Wall and Roof Insulation and New Windows
ComStock	2023	HP-RTU and ASHP-Boiler
ComStock	2023	Wall and Roof Insulation, New Windows,HP-RTU and ASHP-Boiler

Supplementary M2.2 Model setup and Parameterization

In this study, we used grid research to for the hyperparameter tuning of our MLP model. Based on the grid search on the hyperparameter in case counties, we set the MLP with: 1) (64,16) hidden layer sizes; 2) $1*10^{-4}$ Ridge regularization and 3) 'relu' activation function. To ensure the model sees enough extreme events, we performed stratified sampling by temperature: the training set was sampled to have roughly uniform coverage of low, medium, and high outdoor temperatures. Further, observations with the top 5% of loads were up-weighted in the loss function to emphasize correct prediction of peak demand.

To construct predictive models of building-level energy demand, we selected input features that capture the physical, temporal, and behavioral determinants of heating and cooling load. These include: (1) the instantaneous outdoor dry-bulb temperature, which serves as the primary climatic driver of HVAC energy use; (2) the 24-hour rolling mean temperature and its lagged derivative, to encode short-term thermal inertia and recent climatic trends; (3) sinusoidal encodings of hour-of-day (hour_sin, hour_cos) to represent diurnal load cycles; (4) interaction terms between temperature and time-of-day (temp × hour_sin, temp × hour_cos) to capture state-dependent load responses (e.g., increased afternoon cooling demand under high temperatures); and (5) categorical indicators for season and weekend status to account for broader behavioral or operational patterns. We further verified the contribution of each input feature using permutation analysis of the contribution from each variable on peak estimation, taking Dallas as an example, as shown in Figure S3. Result shows that all input features contribute to at least 5% improvement on Symmetric Mean Absolute Percentage Error of either cooling and heating hourly peak.

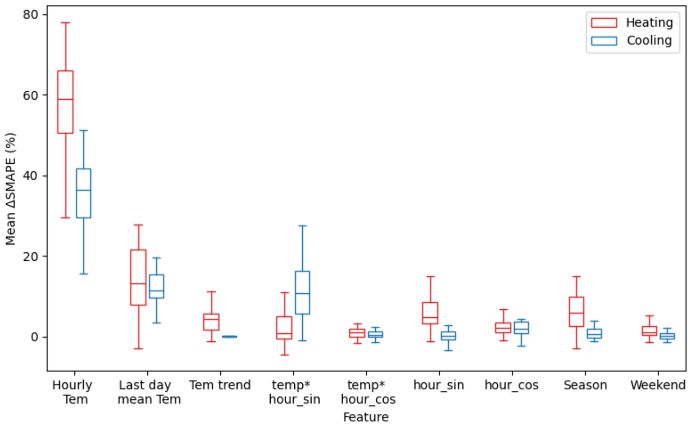

Figure S3 Boxplots of the distribution of mean ΔSMAPE for each input feature on resampled hourly peak load predictions, taking Dallas as an example. *Box plot shows the distribution among different types of buildings.*

Supplementary M2.3 Model performance

In this study, we validate our MLP-based building energy consumption model by hourly R-square for observed and predicted demand and Symmetric Mean Absolute Percentage Error (SAMPE) for observed and predicted annual peak, respectively.

Figure S4 and Figure S5 illustrate the R-square distribution result for the 7828 (103 cities * 76 types) and 3296 (103 cities * 32 types) for residential and commercial building stocks under different upgraded scenarios. Results show that the median R^2 for residential building amounts at 0.97-0.98 and 0.98 for heating and cooling, respectively. (Figure S4) Median R^2 for commercial buildings are 0.95-0.97 and 0.97-0.98 for heating and cooing across upgrade scenarios, respectively. R^2 tends to be greater than 0.90 for all of the 7828 residential models and greater than 0.80 for 3296 commercial models.

Figure S6 and Figure S7 validate the performance of our MLP model on peak demand analysis, through the

Symmetric Mean Absolute Percentage Error (SMAPE) metric. For 7828 residential models, median SMAPE across all upgrade scenarios ranges at 0.02-0.03, 0.02-0.06 and 0.01-0.02 for heating, cooling and total peak loads, respectively. For 3296 commercial loads, median SMAPE ranges at 0.03-0.04, 0.04-0.06 and 0.01-0.02. respectively.

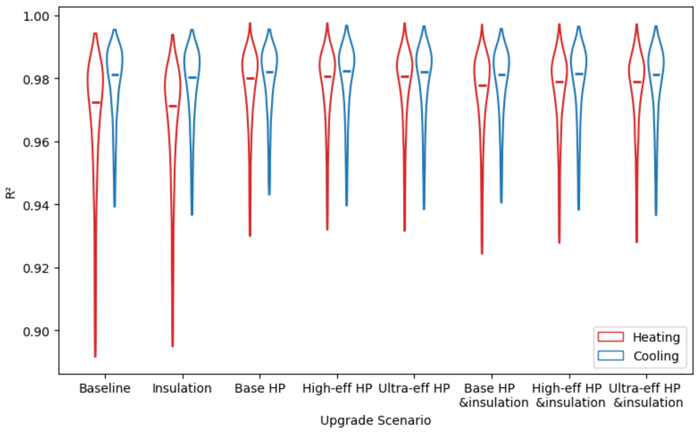

Figure S4 Violin plot of the distribution of R^2 of MLP building regression for residential buildings. *X-axis indicate different upgrade scenarios. Each violin include 103(city_number) * 76 (aggregated residential type) points for its R^2 estimation. The violin plots don't include outliers.*

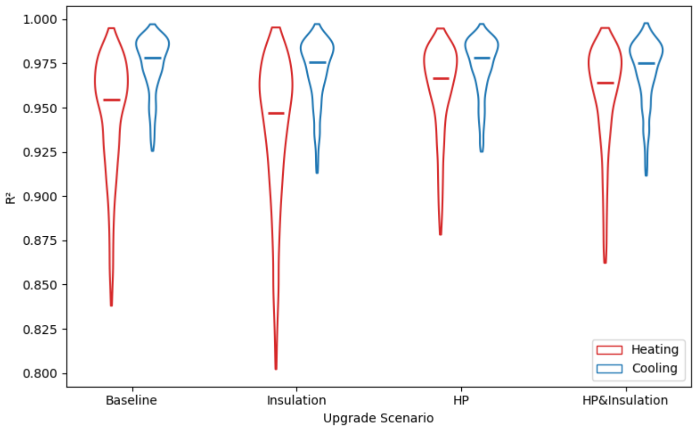

Figure S5Violin plot of the distribution of R^2 of MLP building regression for commercial buildings. *X-axis indicate different upgrade scenarios. Each violin include 103(city_number) * 32 (aggregated residential type) points for its R^2 estimation. The violin plots don't include outliers*

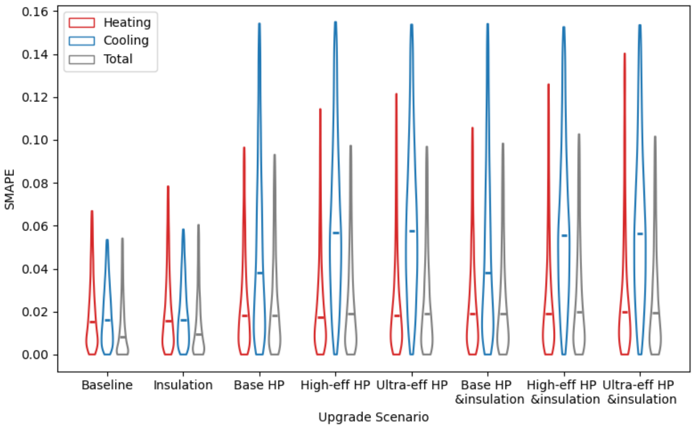

Figure S6 Violin plot of the distribution of Symmetric Mean Absolute Percentage Error (SAMPE) of MLP building regression for residential buildings. *X-axis indicate different upgrade scenarios. Each violin include 103(city_number) * 76 (aggregated residential type) points for its SMAPE estimation. Different boxes indicate the SAMPE for heating, cooling and total peak. The violin plots don't include outliers.*

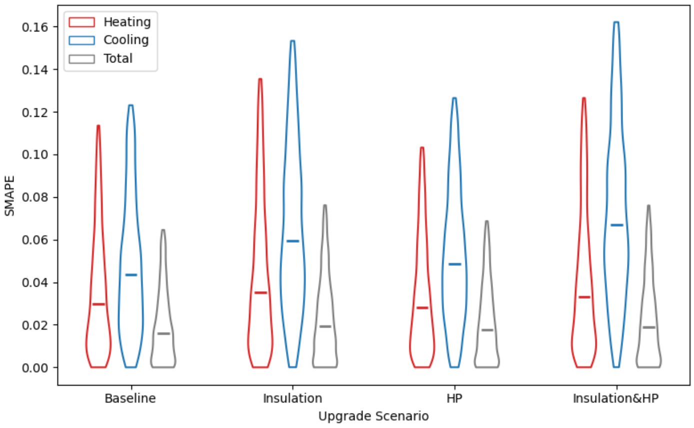

Figure S7 Violin plot of the distribution of Symmetric Mean Absolute Percentage Error (SAMPE) of MLP building regression for commercial buildings. *X-axis indicate different upgrade scenarios. Each violin include 103(city_number) * 32 (aggregated commercial type) points for its SMAPE estimation. Different boxes indicate the SAMPE for heating, cooling and total peak. The violin plots don't include outliers.*

Supplementary M3 Downscaling of the climate large ensembles

Supplementary M3.1 Framework

Our study employs a machine learning-driven statistical downscaling framework. For each location, we first train and validate a model using historical ERA5 data. The model learns to predict hourly temperature anomalies from aggregated three-hour data. Once validated, this trained model is then used to downscale the three-hour CESM2 Large Ensemble projections to produce the final hourly climate data. The flowchart is as follows, shown in Figure **S8**. Specifically, the major input-output of our climate downscaling model includes: 1) **Training Input:** Three-hour aggregated data and derived features from ERA5; 2) **Training Target:** Hourly temperature anomalies from ERA5 (relative to the 3-hour mean); 3) **Application Input:** Three-hour climate projections from a 40-member CESM2-LENS2 subset (2015–2100); 4) **Final Output:** An ensemble of hourly temperature projections for 103 U.S. counties.

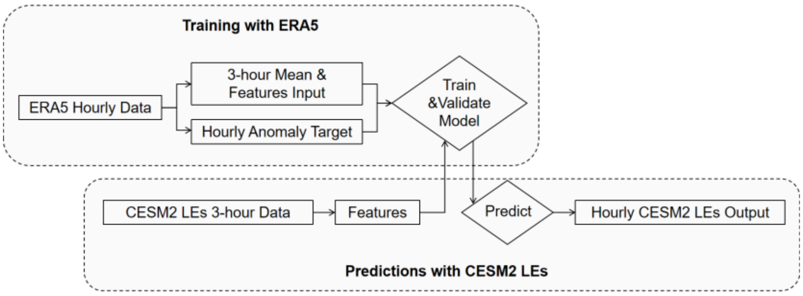

Figure S8. Flowcharts for downscaling.

Supplementary M3.2 Downscaling model setup

Supplementary M3.2.1 Datasets

CESM2-LENS2[3]: This dataset provides the low-resolution climate model output. We use a 40-member subset at three-hour resolution, derived from initializations that sample a broad spectrum of the Atlantic Meridional Overturning Circulation (AMOC) phases, ensuring a rich representation of internal climate variability under a common external forcing (SSP3-7.0).

ERA5 Reanalysis[4]: This dataset serves as the high-resolution training reference. Its hourly temporal resolution (approximate 31 km spatial resolution) from 1980 to 2024 provides the detailed ground truth required to train the model to reconstruct realistic local climate phenomena

Supplementary M3.2.2 Feature Engineering

A set of 14 informative features was engineered to predict the hourly temperature profile. The selection followed a rigorous evaluation procedure to maximize explanatory power while minimizing redundancy. We combined correlations, statistical tests (univariate F-test/ANOVA) to assess the linear dependency of each feature with the target variable, and machine learning importance metrics (from a Random Forest regressor) to rank predictors by their contribution to error reduction. This process yielded an optimal set of features, including: 1) **Baseline Descriptors**: The eight three-hour mean temperatures for the day, the daily maximum and minimum temperatures, and the daily mean specific humidity; 2)**Auxiliary Features**: Cyclical sine/cosine encodings for hour, day-of-week, and day-of-year; 3-day running mean and standard deviation of temperature; interaction terms (temperature × sinusoidal hours, humidity × sinusoidal hours); and quadratic terms for temperature and humidity.

Supplementary M3.2.3 Model Selection

We evaluated seven different algorithms spanning classical machine learning and deep learning: Random Forest (RF), XGBoost, k-Nearest Neighbors (KNN), Support Vector Regression (SVR), Long Short-Term Memory (LSTM), Multi-Layer Perceptron (MLP), and Convolutional Neural Networks (CNN).

The pilot study showed that ensemble tree methods (Random Forest and XGBoost) delivered the lowest errors and highest explained variance. Extreme-event analysis echoed the bulk metrics: these ensemble learners reproduced both the annual mean and the top-5% hourly peaks within a few percent, while other models systematically under- or over-estimated extremes. The performance metrics for all tested models are summarized in Table S5:

Table S5 The performance metrics of downscaling model selection.

Model	MAE	RMSE	R^2 Score
XGBoost	0.7	0.91	0.96
Random Forest	0.7	0.92	0.95
KNN	0.93	1.27	0.91
LSTM	1.1	1.7	0.93
SVR	1.25	1.73	0.93
MLP	1.38	1.81	0.92
CNN	1.47	1.88	0.91

Supplementary M3.2.4 Model Sketch

Based on the comparative analysis, a Random Forest (RF) model with a residual learning strategy was selected for each county. We chose RF rather than XGBoost because the extreme value distribution of the latter does not match the actual situation well.

Model Definition: The model is trained to predict the hourly temperature anomaly relative to the corresponding three-hour mean, defined as:

$$hourly\ anomaly_{h,d} = T_{h,d}^{ERA5} - T_{3hr,d}^{ERA5} \qquad \text{Eq.S1}$$

where $T_{h,d}$ is the hourly temperature and $T_{3hr,d}$ is the temperature averaged over the three-hour window containing hour h on day d.

Training and Validation: Based on the comparison across multiple models, we fit a random forest for each county location and select its specific configuration via parameter estimation within the ranges $n_{estimators} \in \{100, 200, 500\}$; $max_{depth} \in \{None, 10,20\}$; $min_{samples\ split} \in \{2,5\}$, using year-blocked 5-fold cross-validation. Data from 1980-2015 is used for training, and data from 2016-2024 is used for validation.

Supplementary M3.3 Model performance

Overall performance: Across all locations, the model exhibits uniformly low errors and minimal biases (**Table S6**), with tight, unimodal distributions for all metrics **Figure S9**). The concentration of sites near the central bins and the narrow spread indicate stable accuracy across diverse settings. Bias in both daily highs and lows remains close to zero on average, supporting good calibration. 95th-percentile diagnostics are tightly centered near zero for both hot and cold extremes, confirming well-aligned extremes across sites.

Table S6. Overall performance summary of the ERA5 t2m downscaling model across all locations.
Descriptive statistics (K) for MAE, RMSE, average daily maximum temperature bias ("hot bias"), and average daily minimum temperature bias ("cold bias"). Low central tendency and small dispersion across sites indicate uniformly small errors and minimal systematic bias. Note: MAE = mean absolute error; RMSE = root mean squared error; all temperature metrics in Kelvin.

Metric	Mean	Median	Std	Min	Max
MAE (K)	0.13	0.13	0.03	0.07	0.18
RMSE (K)	0.21	0.22	0.04	0.12	0.29

Hot bias (Avg. max error) (K)	0.13	0.14	0.04	0.06	0.19
Cold bias (Avg. min error) (K)	0.1	0.09	0.02	0.07	0.15
95th (annual max) Hot relative error (%)	0.05	-0.04	0.08	-0.33	0.12
95th (annual min) Cold relative error (%)	-0.10	-0.02	0.17	-0.59	0.09

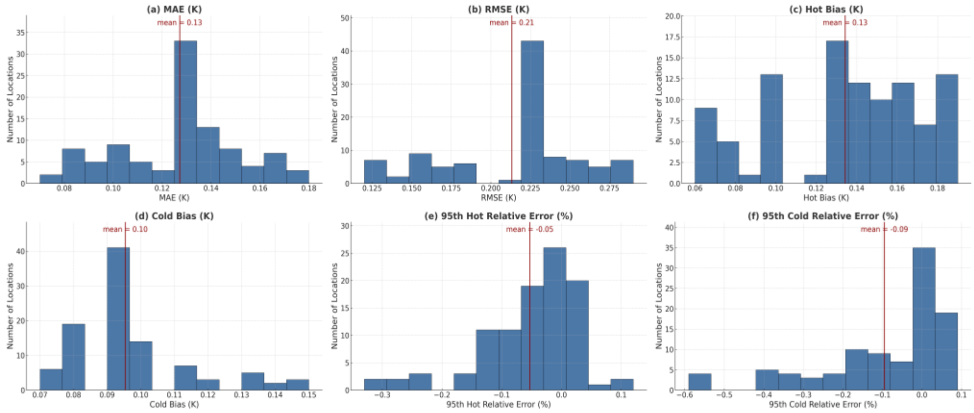

Figure S9. Distributions of error and bias metrics across all locations.
Histograms for (a) MAE, (b) RMSE, (c) hot bias, (d) cold bias, (e) 95th hot relative error (%), and (f) 95th cold relative error (%). All four metrics exhibit tight, unimodal distributions centered at low values, indicating stable accuracy and calibration across sites. Units: K for MAE/RMSE/bias; % for 95th-percentile relative errors.

Regional patterns: Grouping sites by broad climate regimes (warm vs. cold; humid/coastal-leaning vs. dry/continental) reveals best performance in warm, humid regions, where error and bias metrics—including the 95th-percentile hot/cold relative errors—are compact, with medians tightly clustered and cold-side 95th errors remaining near zero across regions (**Figure S10**). Colder and drier zones show slightly higher and more variable values, yet medians remain within a tight operational range and distributions largely overlap across zones; overall performance is consistently strong, with only modest degradation under colder, drier conditions.

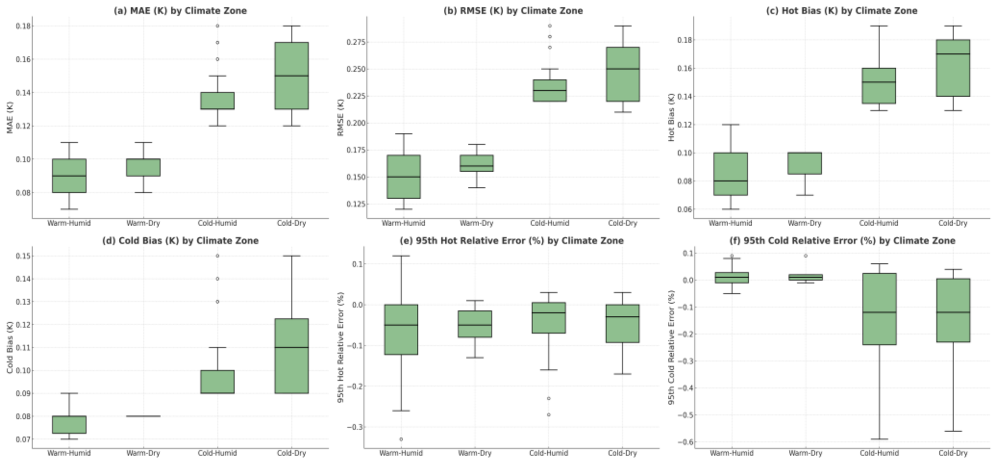

Figure S10. Performance by climate zone.

Boxplots of (a) MAE, (b) RMSE, (c) hot bias, (d) cold bias, (e) 95th hot relative error (%), and (f) 95th cold relative error (%) for four broad climate regimes (Warm–Humid, Warm–Dry, Cold–Humid, Cold–Dry), defined by latitude (warm vs. cold) and continentality (humid/coastal–leaning vs. dry/continental). Warm–Humid (and Warm–Dry) regions show the lowest medians and most compact spreads; colder/drier regions show slightly higher and more variable values, yet remain within a tight operational range. Note: Boxes depict interquartile ranges with medians; whiskers indicate variability across locations.

Locations with low MAE also display correspondingly low hot/cold biases (both mean bias and 95th-percentile relative errors), and MAE aligns closely with RMSE (**Figure S11**). The association is strongest on the warm side (near-linear with MAE) and positive but weaker on the cold side, indicating that improving mean accuracy reduces both central and tail-bias components in a self-consistent manner across space and conditions. Taken together, the metrics provide a self-consistent picture of robust model fidelity across space and conditions.

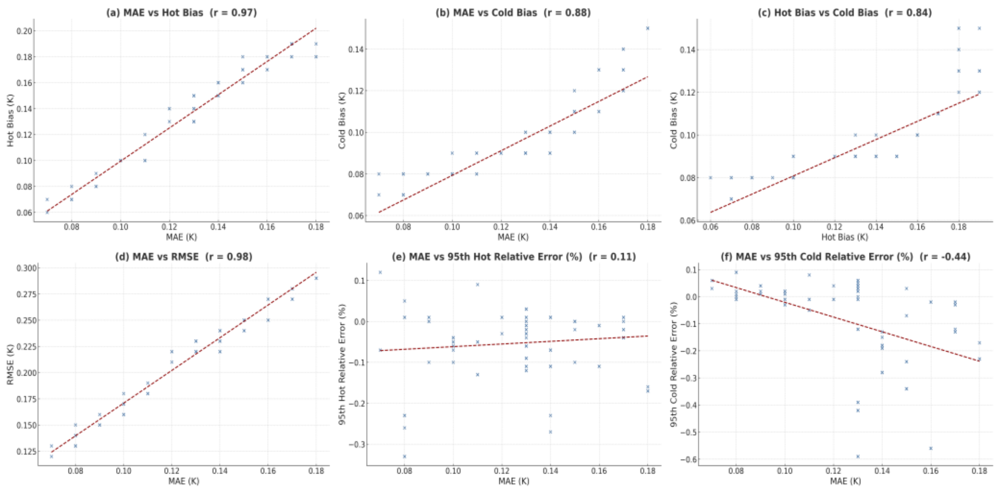

Figure S11. Relationships among metrics across locations (scatter).

(a) MAE vs hot bias; (b) MAE vs cold bias; (c) Hot bias vs cold bias; (d) MAE vs RMSE; (e) MAE vs 95th hot relative error; (f) MAE vs 95th cold relative error. MAE increases with warm-side metrics and shows a positive, weaker association on the cold side; sites with low MAE also exhibit low biases and small 95th-percentile errors. Note: Each point represents one location.

Supplementary M3.4 Characteristics of Downscaled Large Ensembles

The downscaled hourly temperature projections, $T_{h,d}^{LENS}$, are reconstructed by adding the predicted anomaly to the corresponding three-hour mean from the LENS2 dataset for each ensemble member:

$$T_{h,d}^{LENS} = T_{3hr,d}^{LENS} + predicted\ hourly\ anomaly_{h,d} \qquad \text{Eq.S2}$$

The application of this model to the 40-member CESM2-LENS2 ensemble produced a detailed projection of future hourly climate, revealing key insights into trends, extremes, and uncertainty.

The downscaled projections (**Figure S12**) show a clear and robust warming trend in annual mean temperatures across all U.S. locations. The absolute range of uncertainty for these annual means, as measured by the ensemble spread, remains relatively stable over time. A temporal comparison of temperature distributions (e.g., 2020-2030 vs. 2050-2060, shown in **Figure S13**) reveals a distinct shift toward warmer conditions, accompanied by an increase in variance, particularly in the tails of the distribution.

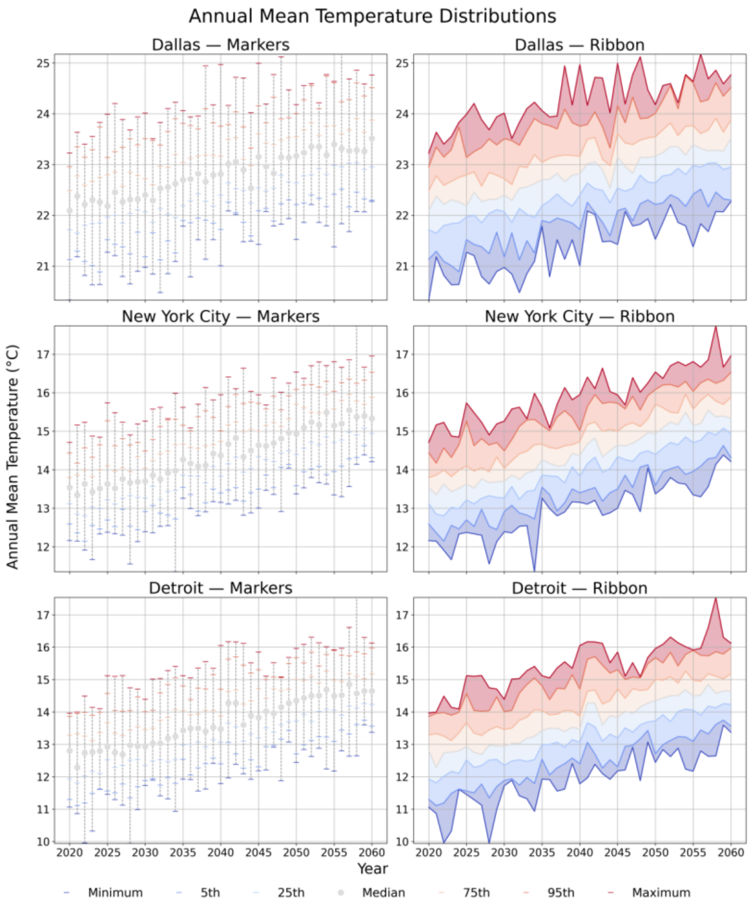

Figure S12 Annual mean temperature trends and uncertainty.

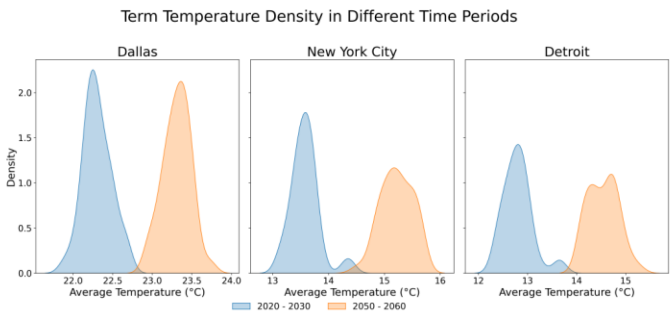

Figure S13 Temporal Comparison of Distributions.

The frequency of extreme heat events, defined relative to a historical baseline (e.g., hours exceeding the 95th percentile of the baseline period), is projected to increase dramatically, shown in Figure **S14**. Furthermore, the uncertainty in this frequency (i.e., the spread across ensemble members) also widens over time, indicating growing unpredictability in the year-to-year occurrence of extreme heat, especially in already hot climates like Dallas.

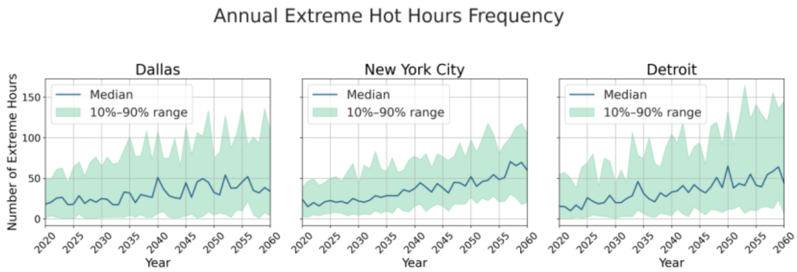

Figure S14. Occurrence of extreme heat in evolving years

As **Figure S15** shows, the spatial analysis of ensemble uncertainty reveals distinct geographical patterns: the largest uncertainty in mid-century warming is concentrated in the interior northern U.S. (e.g., Great Plains, Upper Midwest), whereas the lowest uncertainty is found in the South costal areas. This highlights that the "irreducible" uncertainty from internal climate variability is not uniform and is instead geographically dependent.

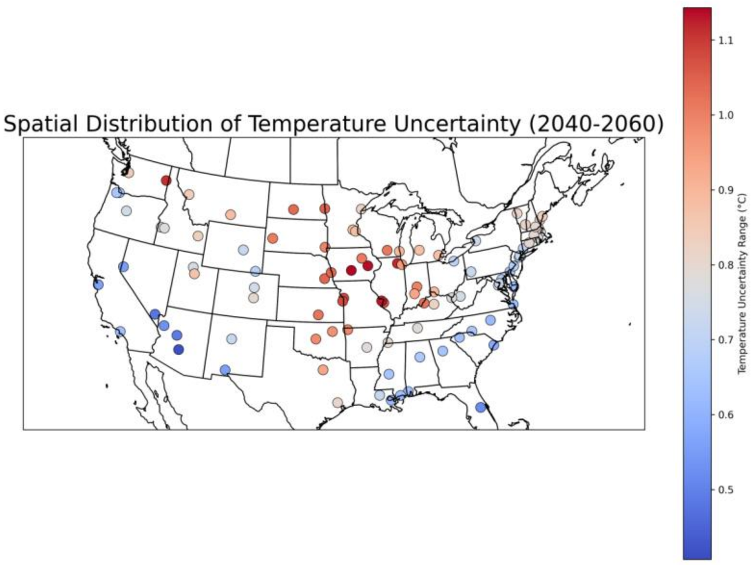

Figure S15 Spatial pattern of ensemble uncertainty.

Supplementary M4 Technology diffusion modelling

In this study, we use S-curve-based logistic model to project future technology diffusion. S-curve was first adopted in biology research to capture biological population growth under resource constraints[5]. Researchers started to adopt S-curve to project technology growth from 1960s[6]. S-curve assumes that the growth process begins after the invention stage and proceeds in three phases including formative, growth, and saturation. Logistic growth model is often used to represent the technology diffusion as an S-curve, with the use of growth rate k, inflection year t_0 and saturation level L:

$$C_t = \frac{L}{1 + e^{-k(t-t_0)}} \qquad \text{Eq.S3}$$

where C_t represents the capacity of technology at year t.

To project future growth for more mature technologies, logistic growth can be extrapolated based on existing capacity data. This method of projection is most accurate for data series that include the inflection point, or are very close to this point. However, for early-stage technologies which haven't reach the inflection point, the logistic model could be shifted to its derivative form and used to iteratively calculate future capacity from an early stage:

$$C_{t+1} = C_t(1 + k(1 - \frac{C_t}{L}) \qquad \text{Eq.S4}$$

Therefore, we use Eq.S2 to project the annual diffusion of heat pump and insulation upgrades among buildings in ResStock and ComStock in this study, through three parameters: growth rate K, initial capacity C_0 and saturation level L. We assume L as 100% in this study in the main scenario towards a fully electrified heating and cooling future with higher insulation efficiency.

In this study, we adopt technology diffusion rate k from historic technologies. Specifically, we apply SHARD methods[7] and identify 18 technologies in HATCH database[8] related to building thermal energy use, classify them into 3 categories due to their similarities to reflect lower to higher tech-diffusion uncertainties. We then fit their annual growth with logistic model and calculate their k value. The specific information of each historic technology analog is shown in Table S5 below.

Table S7 Historic analog and their technology diffusion rates

Number	Technology name	Categoty	Diffusion rate k	District	Data point
1	Laundry Dryers	1	0.071	Global	87
2	Washing Machines	1	0.087	Global	89
3	Home Air Conditioning	1	0.100	USA	55
4	Refrigerators	1	0.113	Global	92
5	Dishwashers	1	0.154	USA	40
6	Passenger Cars	2	0.042	USA	103
7	Bicycles	2	0.058	Global	147
8	Stove	2	0.079	USA	108
9	Telephones	2	0.090	USA	117
10	Microcomputers	2	0.184	USA	33
11	Colour TV	2	0.261	USA	40
12	Compact Fluorescent Light Bulbs	2	0.271	World	14
13	Vacuums	3	0.020	USA	56
14	Electricity	3	0.038	Global	37
15	Running Water	3	0.041	USA	27
16	Solar Thermal Energy	3	0.290	USA	22
17	Solar Photovoltaic	3	0.337	USA	22
18	Cellphones	3	0.363	USA	25

For the initial percentage of heat pump and insulation upgrades, we calculate it by each county based on the baseline results in ResStock and ComStock, respectively. The initial residential heat pump, residential insulation upgrades and commercial heat pump adopters are illustrated in Figure S16-S18. As insulation upgrades are applicable for 100% ComStock buildings, we assume an initial starting point at 1% for insulation upgrades adoption among commercial buildings.

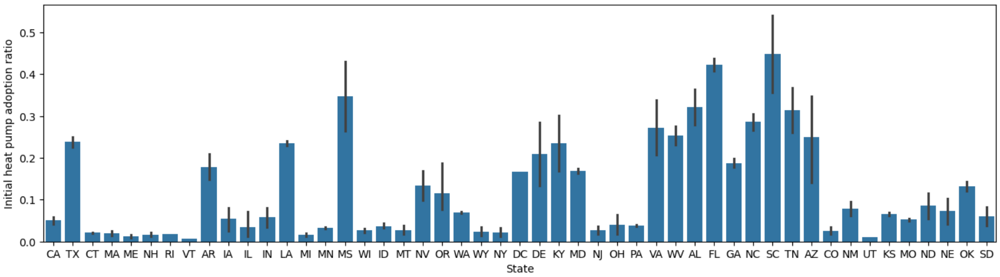

Figure S16 Ratio of initial heat pump adopters across different states in residential buildings.

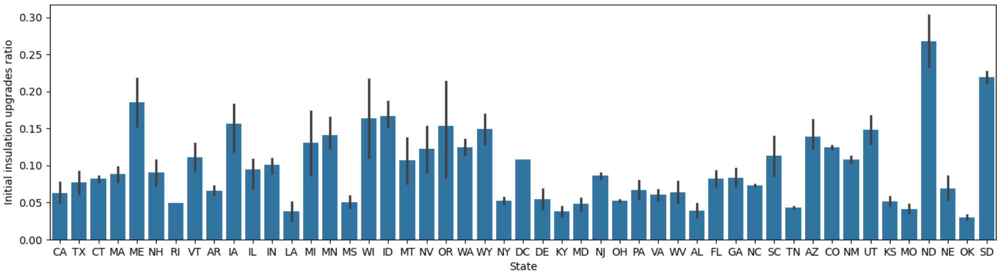

Figure S17 Ratio of initial insulation upgraded residential buildings across different states

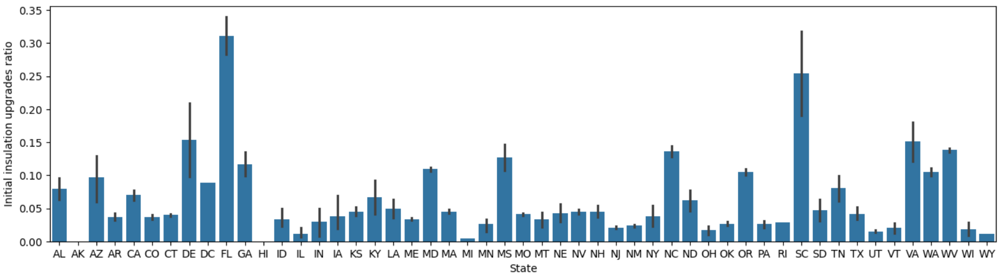

Figure S18 Ratio of initial heat pump adopters across different states in commercial buildings.

Based on the diffusion rate k, city-specific initial capacity C_0 and target saturation as 100%, we model the annual diffusion dynamics of the initial-, new- and non-adopters for heat pump and insulation upgrades for residential and commercial buildings, respectively. Figure S19 and Figure S20 illustrate an example in Chicago and Dallas, for residential and commercial buildings, respectively. The adoption dynamics for residential and commercial buildings in other cities are accessible at (*Data share link*)

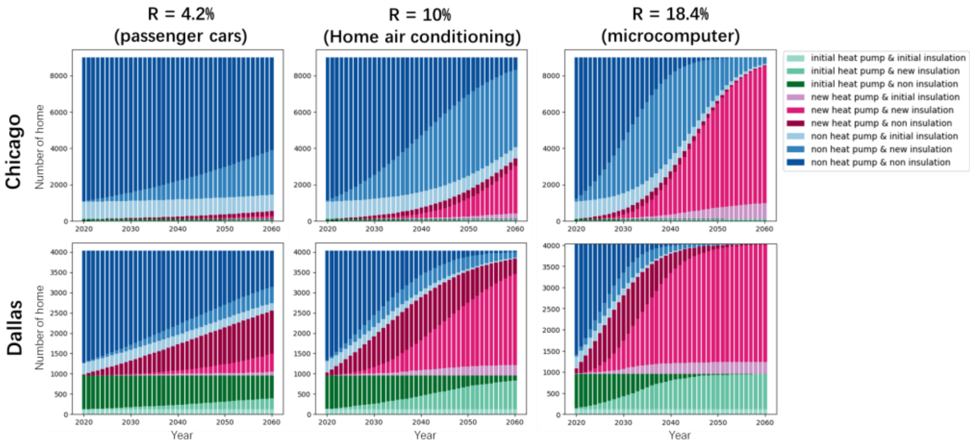

Figure S19 Dynamics of heat pump and insulation upgrades diffusion among residential homes under different diffusion rate k, taking Chicago and Dallas as examples.

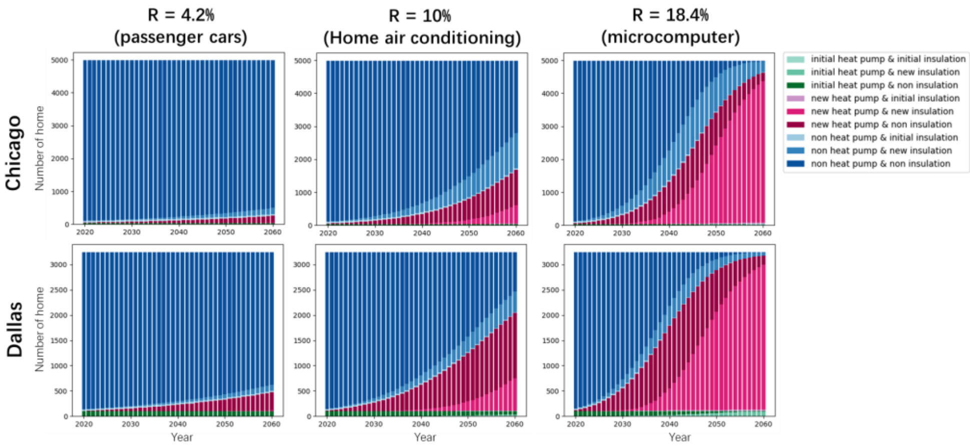

Figure S20 Dynamics of heat pump and insulation upgrades diffusion among commercial buildings under different diffusion rate k, taking Chicago and Dallas as examples.

Supplementary M5 Loads aggregation across cities

Here, we aggregate the city-scale simulated results in ResStock and ComStock into hourly demand of total building sector, through weighting them based on state-level statistical power consumption data in Residential Energy Consumption Survey (RECS) and Commercial Building Energy Consumption survey (CBECS), as shown in EqS5-S7:

$$Res_{city,i} = ResStock_{city,i} \times \frac{RECS_{state,2020}}{ResStock_{state,2020}} \qquad \text{Eq.S5}$$

$$Com_{city,i} = ComStock_{city,i} \times \frac{CBECS_{state,2018}}{ComStock_{state,2018}} \qquad \text{Eq.S6}$$

$$Total_{grid,i} = \sum_{city \in grid} (Res_{city,i} + Com_{city,i}) \qquad \text{Eq.S7}$$

where $ResStock_{city,i}$, $ComStock_{city,i}$ indicate the model prediction results of residential and building electricity consumption in year i in each city, respectively; $Res_{city,i}$, $Com_{city,i}$ indicate the weighted results of residential and building electricity consumption in year i in each city, respectively; $RECS_{state,2020}$, $CBECS_{state,2018}$ indicate the statistical results of state-level electricity consumption in each state, recorded in 2020 RECS and 2018 CBECS, respectively; $ResStock_{state,2020}$ and $ComStock_{state,2018}$ indicate the simulated total state-level building consumption in 2020 and 2018, respectively, as a comparison with statistical data.

Supplementary Results

Supplementary R1 Supplementary results on peak demand dynamics

Supplementary R1.1 Peak demand uncertainty distribution

Table S8 Thiel-Sen metrics of peak loads growth in each power grid region across 2020-2060 (as Supplementary for Figure 1 in the main text)

Region	Peak TS-slope per 40 year		
	Climate change only	Diffusion only	Combined effects
ISO-NE	3.5%	80.5%	67.7%
NYISO	2.8%	56.5%	40.3%
MISO	4.8%	65.3%	17.6%
PJM	4.6%	99.8%	55.4%
Northwest	-11.0%	67.6%	47.2%
SPP	4.7%	81.9%	26.1%
CAISO	2.0%	0.5%	2.1%
Southwest	4.9%	-4.3%	0.6%
Southeast	4.8	-2.0%	-4.9%
ERCOT	2.6%	-9.5%	-6.9%

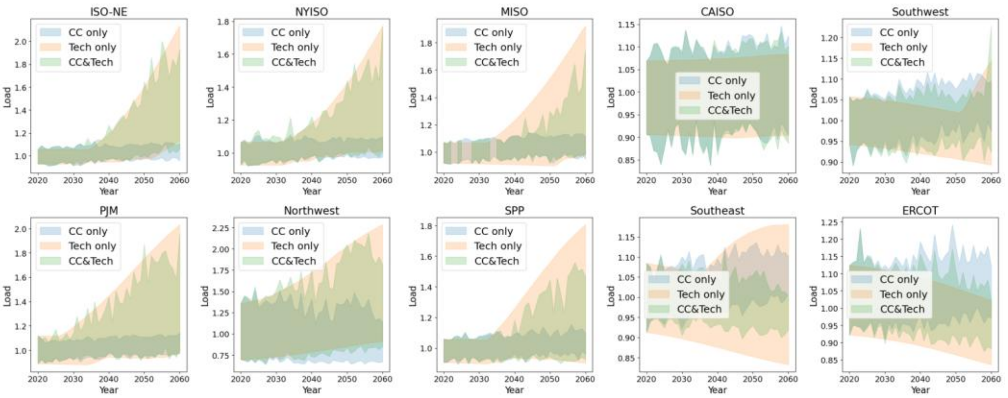

Figure S21. Climate uncertainty of future peak dynamics under climate change, tech-diffusion and combined impacts. *Y-axis indicate the ratio of annual peak to the median peak in 2020. Diffusion rate is adopted as 0.100 here, consistent with the main scenario in the main text.*

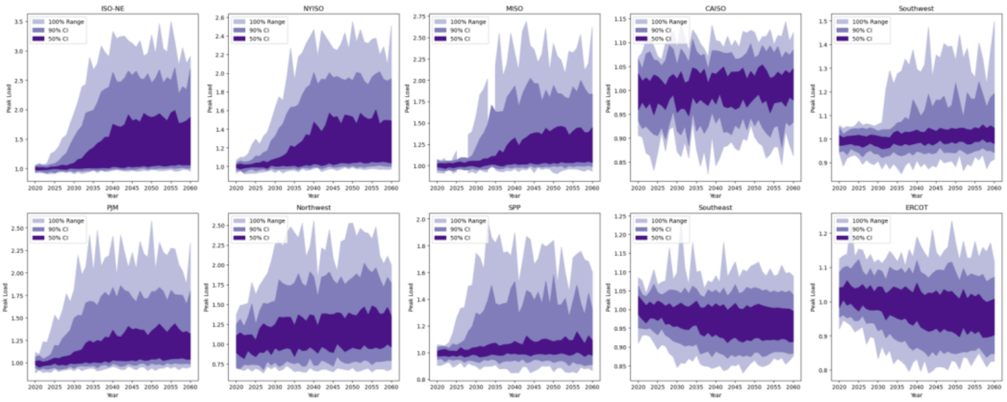

Figure S22 Uncertainty of peak load dynamics across 720 (18 diffusion rates * 40 climate scenario) ensembles. *Y-axis indicate the ratio of annual peak to the median peak in 2020. The different dashed regions indicate 50% confidence interval, 90% confidence interval and full range, respectively.*

Supplementary R1.2 Supplementary on cooling and heating peak demand dynamics

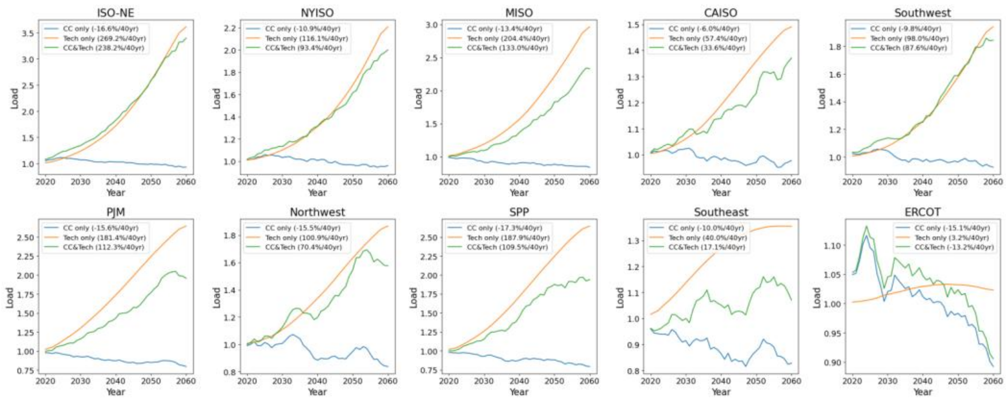

Figure S23 Dynamics of the 95th extreme heating peak under climate change, tech-diffusion and combined impacts. *Y-axis indicate 5-year rolling average for the ratio of annual 95th peak across climates to the 95th peak in 2020. Diffusion rate is adopted as 0.100 here, consistent with the main scenario in the main text.*

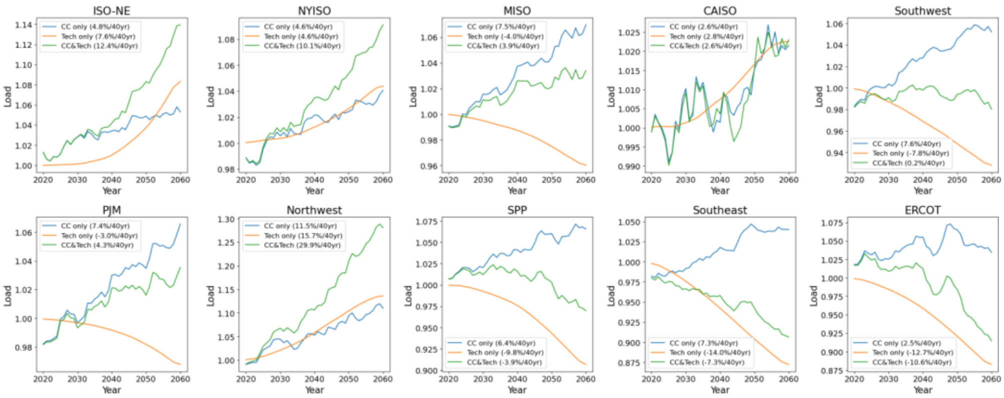

Figure S24 Dynamics of the 95th extreme cooling peak under climate change, tech-diffusion and combined impacts. *Y-axis indicate 5-year rolling average for the ratio of annual 95th peak across climates to the 95th peak in 2020. Diffusion rate is adopted as 0.100 here, consistent with the main scenario in the main text.*

Supplementary R1.3 Supplementary on residential and commercial peak demand dynamics

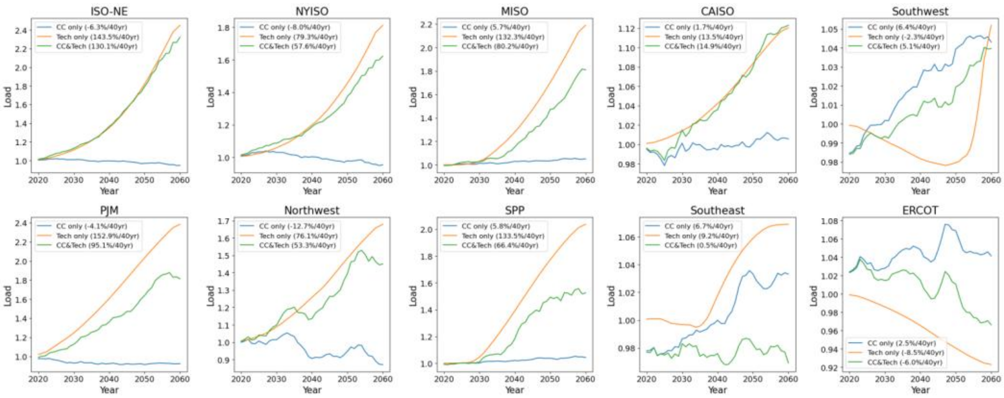

Figure S25 Dynamics of the 95th extreme peak for residential buildings under climate change, tech-diffusion and combined impacts. *Y-axis indicate 5-year rolling average for the ratio of annual 95th peak across climates to the 95th peak in 2020. Diffusion rate is adopted as 0.100 here, consistent with the main scenario in the main text*

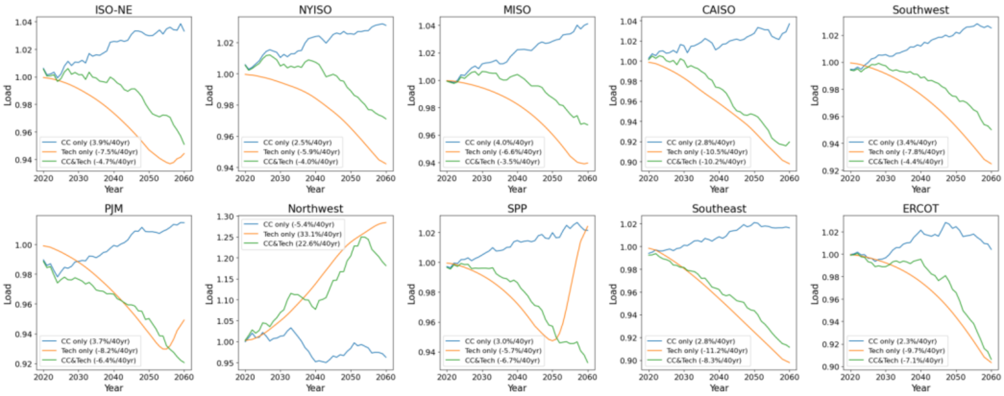

Figure S26 Dynamics of the 95th extreme peak for commercial buildings under climate change, tech-diffusion and combined impacts. *Y-axis indicate 5-year rolling average for the ratio of annual 95th peak across climates to the 95th peak in 2020. Diffusion rate is adopted as 0.100 here, consistent with the main scenario in the main text*

Supplementary R1.4 Supplementary on city-scale peak demand dynamics

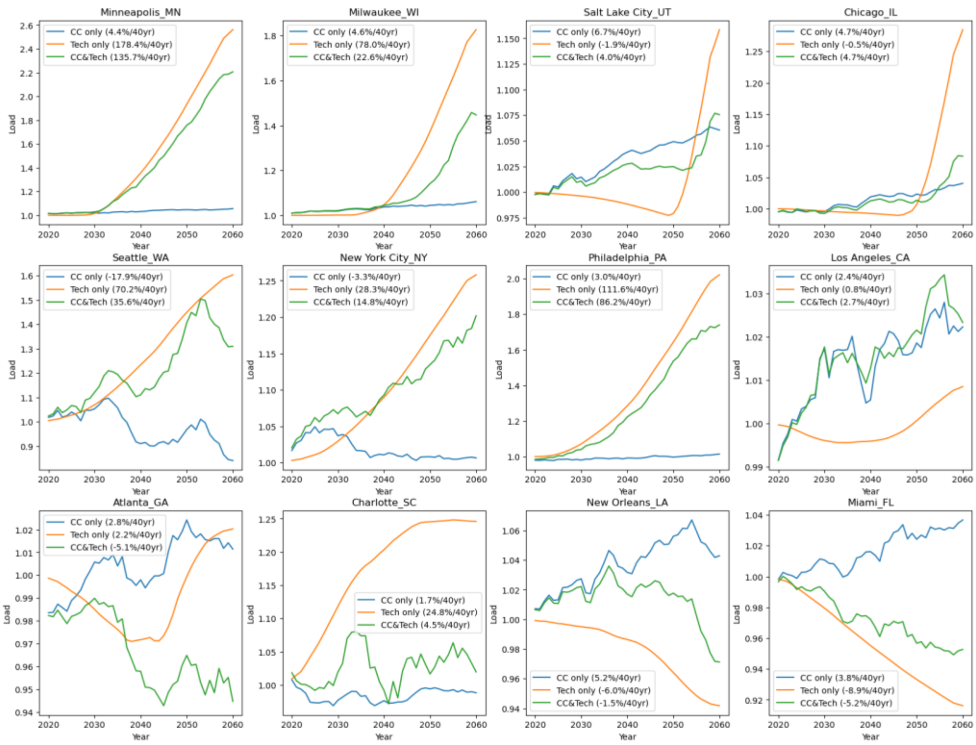

Figure S27 Dynamics of the 95th extreme peak under climate change, tech-diffusion and combined impacts in 12 typical cities across continental U.S. *Y-axis indicate 5-year rolling average for the ratio of annual 95th peak across climates to the 95th peak in 2020. Diffusion rate is adopted as 0.100 here, consistent with the main scenario in the main text*

Supplementary R1.5 Supplementary on peak demand uncertainty analysis

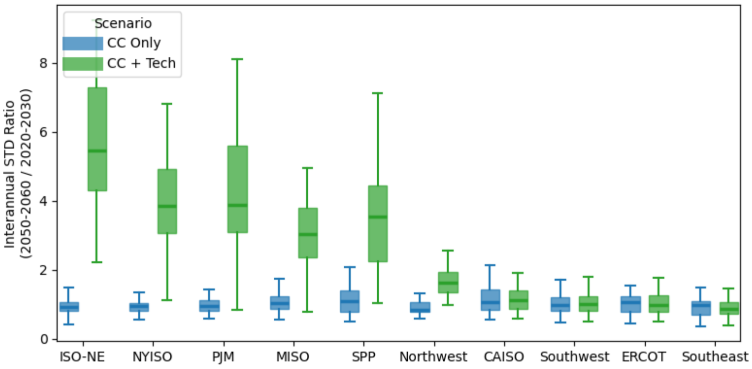

Figure S28 Ratio between the 2050s and 2020s interannual peak loads standard deviation under climate only and combined climate & tech-diffusion impacts.

The relationship of MPL between technology diffusion only and the combined effects scenarios are also largely dynamic across different diffusion rates. In cold regions, MLP mitigated by climate warming (different between MLP in combined effects and tech-diffusion only scenarios) first increase then decrease along with the growth of tech-diffusion rates, ranging between 0-15% (Figure S29a). In warm regions (except Southeast), climate warming impacts tend to increase the MPL growth than tech-diffusion only scenario at a relatively constant ratio among all diffusion rates.(Figure S29b)

批注 [MC1]: If move 4c & 4d to SI, I would move this text too. It's not adding much I'd say.

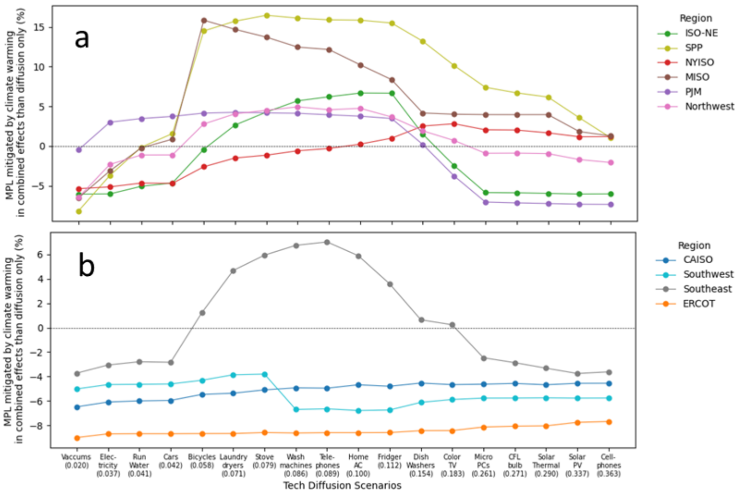

Figure S29 MPL mitigated by future warming climates in cold (c) and hot (d) regions, plotted as the MPL different (%) between tech-diffusion with climate change (combined effects scenarios) and without climate change (tech-diffusion only) scenarios.

Supplementary R2 Sensitivity analysis

Supplementary R2.1 Sensitivity on residential heat pump efficiency

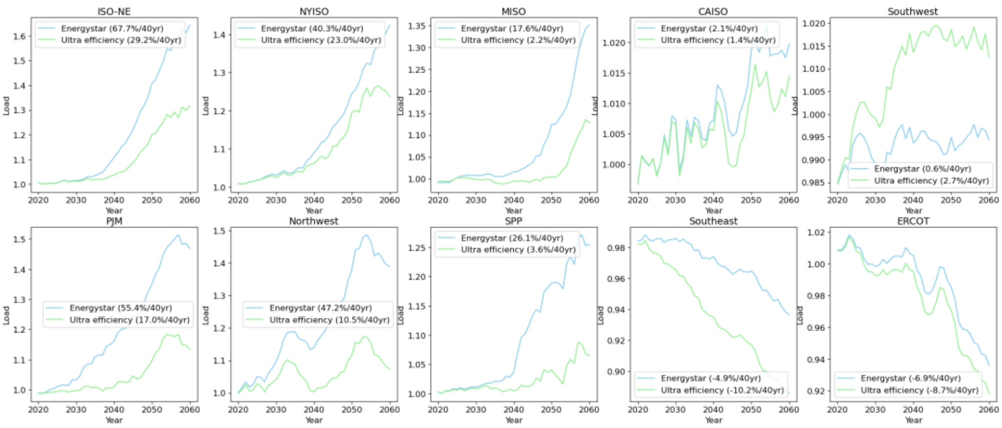

Figure S30 Residential heat pump efficiency on the dynamics of the 95th extreme peak demand under combined climate and diffusion impacts. *Y-axis indicate 5-year rolling average for the ratio of annual 95th peak across climates to the 95th peak in 2020. Diffusion rate is adopted as 0.100 here, consistent with the main scenario in the main text.*

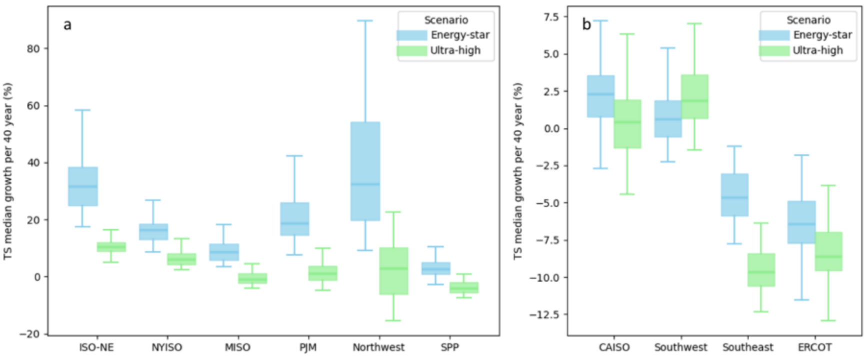

Figure S31 Residential heat pump efficiency on TS peak growth rate across 40 climate realizations. *Diffusion rate is adopted as 0.100 here, consistent with the main scenario in the main text.*

Supplementary R2.2 Sensitivity on heat pump adoption priority

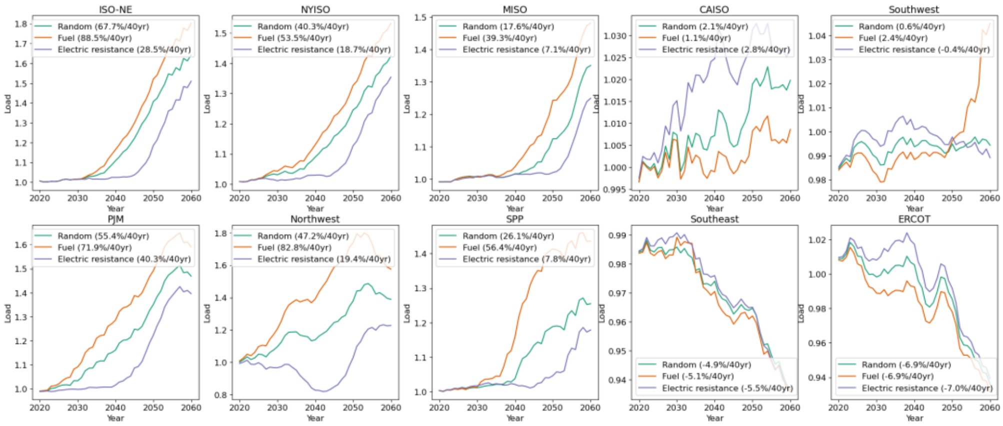

Figure S32 Heat pump adoption priority on the dynamics of the 95th extreme peak demand under combined climate and diffusion impacts. *Y-axis indicate 5-year rolling average for the ratio of annual 95th peak across climates to the 95th peak in 2020. Diffusion rate is adopted as 0.100 here, consistent with the main scenario in the main text.*

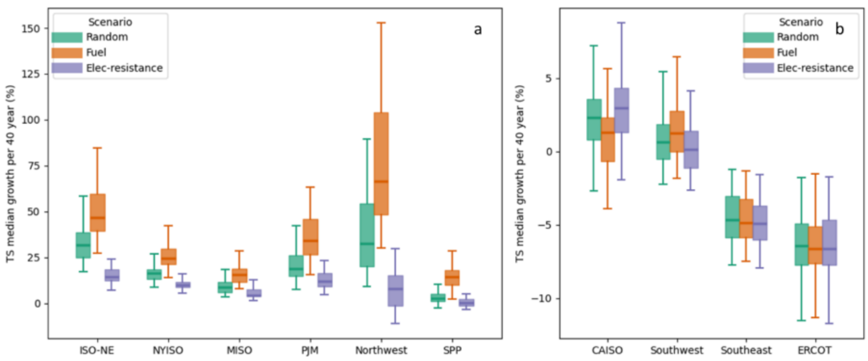

Figure S33 Heat pump adoption priority on TS peak growth rate across 40 climate realizations. *Diffusion rate is adopted as 0.100 here, consistent with the main scenario in the main text.*

Supplementary R2.3 Sensitivity on saturation level

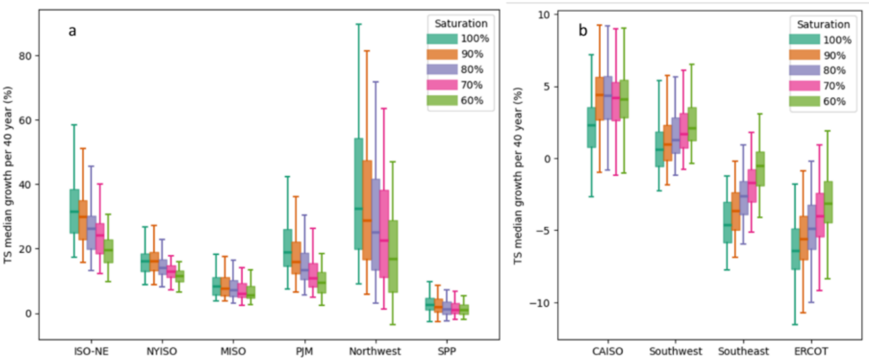

Figure S34 Saturation level of heat pump and insulation upgrades adoption on TS peak growth rate across 40 climate realizations. *Diffusion rate is adopted as 0.100 here, consistent with the main scenario in the main text.*