https://doi.org/10.5194/essd-2025-709
Preprint. Discussion started: 28 December 2025

Open Access Earth System Science Data Discussions

A global base temperature dataset for building energy demand modeling

Xiujuan He[1], Jiyong Eom[2], Sha Yu[3], Shu Liu[1], Wenru Xu[4], Yuyu Zhou[1,5]

[1]Department of Geography, The University of Hong Kong, Hong Kong, PR China
[2]School of Business & Technology Management, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea
[3]Center for Global Sustainability, School of Public Policy, University of Maryland, College Park, MD, USA
[4]CAS Key Laboratory of Forest Ecology and Management, Institute of Applied Ecology, Chinese Academy of Science, Shenyang 110016, China
[5]Institute for Climate and Carbon Neutrality, The University of Hong Kong, Hong Kong, PR China

Correspondence to: Yuyu Zhou (yuyuzhou@hku.hk)

Abstract. Accurate building energy demand modeling is critical to decarbonizing regional energy systems. The cooling and heating degree-day models are widely used due to their simplicity and low data requirements; however, the lack of accurate base temperature data limits their performance. In particular, the scarcity of high temporal resolution building energy demand data constrains regional-scale base temperature estimation through conventional methods such as the energy signature method and the performance line method. To address this limitation, this study develops a global regional-scale base temperature dataset based on the BiLSTM neural network framework with an attention mechanism. The dataset includes both cooling base temperature (*Tcool*) and heating base temperature (*Theat*) for each region, defined at a spatial scale equivalent to a U.S. state or a Chinese province. The BiLSTM framework demonstrates strong performance, with RMSE values of 1.39℃ for training and 1.33℃ for testing, and Pearson correlation coefficients of 0.84 for *Tcool* and 0.70 for *Theat*. Predicted results show that global *Tcool* ranges from 19-25℃ and *Theat* from 14-18℃, consistent with physical principles. External validations using 16 independent datasets demonstrate that the predicted base temperatures significantly improve the accuracy of building energy demand modeling, reducing RMSE by 10.01% for cooling and 10.02% for heating, compared to official or empirical base temperatures. This dataset supplements sparse observational base temperature data and enhances the accuracy of building energy demand modeling, contributing to low-carbon energy system planning, broader climate impact assessment and weather-related financial applications. The proposed global *Tbase* dataset can be acquired from https://doi.org/10.6084/m9.figshare.30646376.v2 (He et al., 2025).

1 Introduction

Reductions in energy demand and carbon emissions in the building sector play a crucial role in achieving the low-carbon energy system transition and the Paris Agreement climate goals (Camarasa et al., 2022). As reported, the building sector consumes 36% of global end-use energy and contributes approximately 39% of energy and process-related CO_2 emissions in 2018 (UN, 2019). Space cooling and heating dominate energy use within the building sector, representing roughly 40% of total building energy consumption globally as of 2018 (UN, 2019). In the future, with continued climate change, population and economic growth, energy demand in the building

https://doi.org/10.5194/essd-2025-709
Preprint. Discussion started: 28 December 2025

Open Access Earth System Science Data Discussions

sector will increase further, putting enormous pressure on the global energy supply system. In response, reducing energy demand is considered a reliable coping strategy.

The spatial variability of building energy demand requires differentiated solutions to achieve energy demand reduction. Ambient temperature is the most significant factor affecting energy demand in buildings (Deroubaix et al., 2021; Wenz et al., 2017). An increase in ambient temperature will reduce heating demand but increase cooling demand across different regions (Abajian et al., 2025; Wang et al., 2023; Filahi et al., 2025; Clarke et al., 2018; Zhou et al., 2013). Simulation analysis reveals that identical building configurations exhibit distinct energy demand patterns across different climate zones (Mahdavinejad et al., 2024; Yu et al., 2014). Among the influencing factors, occupant behaviours in response to temperature changes play a particularly important role in shaping building energy demand. In Seoul, some occupants preferred behavioural adaptations—such as adjusting clothing insulation or activity level—in response to temperature changes, while others relied more directly on air-conditioning (Bae et al., 2025) or opted to visit temperature-controlled stores (Yoo et al., 2024). Occupant behaviour also interacts with socioeconomic factors to shape building energy demand patterns. In India, income levels significantly determine cooling service accessibility among urban poor populations, but there is considerable heterogeneity in their perceived thermal comfort (Gupta et al., 2024). In some developing nations, the increased availability of cooling equipment is anticipated to drive increased space cooling demand as emerging markets expand and affluence levels rise (IEA, 2024; Clarke et al., 2018). Given the spatial variability of factors such as weather, occupant behaviour, and socioeconomic level, a comprehensive understanding of building energy demand response to local conditions is critical (Staffell et al., 2023). This understanding facilitates the reduction of uncertainties in building energy demand modeling and contributes to the design of regional energy systems adapted to future climate conditions.

The Degree-Days method is one of the most commonly used approaches for modeling heating and cooling energy demand in buildings. Degree-day models exhibit substantial advantages in terms of data availability and geographic generalization when compared to physical and statistical models (Prataviera et al., 2021; Ahmad et al., 2018). This is due to the fact that degree-day models rely exclusively on ambient and base temperatures. The base temperatures define thresholds for heating (below the heating base temperature) or cooling (above the cooling base temperature) in buildings (De Rosa et al., 2015; Martinopoulos et al., 2019). Degree days are calculated as the difference between the average daily temperature and this base temperature, with values below the threshold being the heating degree day (HDD) and those above the threshold being the cooling degree day (CDD). Base temperature values are influenced by local climatic conditions, occupant behavior, building types, and thermal properties (Ramon et al., 2020; Hao et al., 2022), which suggests significant spatial heterogeneity in base temperature selection. In the United States, 18.3°C is the most common base temperature. Above 18.3°C, an increase in cooling energy demand is considered, and below 18.3°C, an increase in heating demand is considered (Petri & Caldeira, 2015). In the UK, the HDD and CDD are calculated using 15.5°C and 22°C as the base temperatures, respectively (UK Climate Projections, 2014). In China, 18°C and 24°C are widely used as the base temperatures for the degree-day model (Yang et al., 2024; Li et al., 2018). Conventionally, recommended base temperatures have served as key parameters in modeling heating and cooling energy demand across buildings, regions, and countries (Camarasa et al., 2022; Deroubaix et al., 2021). However, the application of uniform base temperatures across broad geographical areas inadequately represents the sensitivity of building energy demand to local climatic variations, raising concerns about the accuracy and uncertainty of degree-day modeling. Even

https://doi.org/10.5194/essd-2025-709
Preprint. Discussion started: 28 December 2025

minor uncertainties in base temperatures can lead to substantial variations in CDDs. To illustrate, a 1°C change in the base temperature can result in approximately 20 additional or fewer CDDs per month in San Luis Obispo (Woods & Fuller, 2014). Inaccurate base temperatures also trigger a nonlinear relationship between building energy demand and degree-days (Day et al., 2003). Moreover, the implementation of inappropriate base temperatures in building sustainability policies may result in energy inefficiencies or compromised indoor thermal comfort, undermining effective building operation and management.

Accurate base temperature selection is key to optimizing the reliability of degree-day models. Methods such as the energy signature approach and performance line analysis are commonly used to determine optimal base temperatures for specific regions, using building energy consumption data in conjunction with local meteorological data (Bhatnagar et al., 2018; Staffell et al., 2023; Anjomshoaa & Salmanzadeh, 2017; Woods & Fuller, 2014). The energy signature method refers to correlating energy consumption with weather data to obtain a non-linear relationship between them (U-curve or L-curve) and using the inflection point of the curve as the base temperature (Anjomshoaa & Salmanzadeh, 2017; Meng et al., 2018). The performance line method determines the base temperature by plotting the best-fit straight line of energy consumption relative to the monthly HDD or CDD scatterplot data (Bhatnagar et al., 2018). Both approaches employ actual building energy consumption data that reflects the combined influence of local weather conditions, ambient temperature fluctuations, and occupancy behaviour patterns, thus enabling the identification of optimal base temperatures. For example, Bhatnagar et al. (2018) applied both the energy signature method and the performance line method to determine base temperatures across 60 Indian cities, reporting median base temperatures of 18.3°C for cooling and 17.4°C for heating. In Kerman, Iran, the L-curve relationship between average daily temperature and natural gas consumption is derived using the energy signature method, revealing base temperatures of 15.42 and 21.18°C for the region (Anjomshoaa & Salmanzadeh, 2017). Furthermore, Staffell et al. (2023) evaluated indoor temperature thresholds for heating and cooling across several countries and regions based on the energy signature method, thereby enhancing the reliability of building energy modeling. These studies have paved the way for understanding the energy requirements for achieving universal thermal comfort across different zones.

However, the reliance on building energy consumption data with high-temporal resolution makes the determination of accurate base temperatures costly and impractical for many regions, thereby limiting the precision of building energy demand modeling at broader spatial scales. Particularly at the global scale, most studies have assessed temperature-dependent energy demand at the grid, city, or national level using a harmonized base temperature (e.g., Deroubaix et al., 2021; Scoccimarro et al., 2023; Falchetta et al., 2024). These studies potentially limit their value for informing global energy system transitions, as successful building energy transitions require a nuanced understanding of building energy demand necessary to ensure universal thermal comfort (Kikstra et al., 2021). These challenges highlight the critical need for a practical methodology that can predict base temperatures at a global scale using limited data—a prerequisite for accurate global energy demand modeling. Deep learning presents a promising solution by linking base temperatures to local weather patterns and socioeconomic factors through the analysis of high-temporal-resolution building energy consumption datasets. This approach enables the estimation of base temperatures from observable inputs, without the need for detailed physical knowledge of individual building energy systems.

In summary, this research aims to use a statistical model and a deep learning model with limited building energy data, along with readily available regional weather and socioeconomic data, to map base temperatures at

https://doi.org/10.5194/essd-2025-709
Preprint. Discussion started: 28 December 2025

provincial resolution and scale this up to create continuous global maps. The statistical model is a segmented linear regression model based on the energy signature method. Given the strong spatiotemporal dependencies in meteorological data, this study adopts a Long Short-Term Memory Model (LSTM) enhanced with an Attention Mechanism (AMM) as its core framework. The LSTM is capable of capturing temporal correlations between variables, making it effective for analyzing weather data (Yu et al., 2011). To ensure the robustness of the approach, an integrated validation system is also established to verify the reliability of provincial-scale base temperatures for global building energy modeling. The resulting dataset offers important support for regional energy demand modeling and energy system planning.

2 Data and methods

2.1 Data

Energy demand data

The energy demand data encompasses multi-timescale time series electricity or natural gas demand data from 172 regions worldwide (see Table A1). These data are utilized to estimate region-specific base temperatures (*Tbase*), which serve as critical input data for subsequent analyses. Two types of energy demand are considered: building-level and region-level. Building-level energy demand data predominantly consists of residential demand across various temporal resolutions (second, minute, and hourly intervals), which are aggregated to a daily scale to better reflect seasonal patterns linked to local temperatures and weather conditions. Due to the scarcity of available residential data, a small number of public building energy demand datasets are also included to expand coverage. Region-level energy demand data, available at daily and monthly temporal scales, can also capture seasonal variations when the time series is sufficiently long, as demonstrated by Wenz et al. (2017), Huang & Gurney (2016), and Staffell et al. (2023). All energy demand datasets are spatially matched to the finest available administrative boundaries using location metadata, thereby enabling integrated analysis with local climatic conditions.

Meteorological variables

Meteorological data are primary drivers of temporal variations in building energy demand. This study utilizes five key variables from the ERA5-Land Reanalysis data (monthly averages from *1950 to present*): 2-meter air temperature, dew point temperature, 10-meter wind speed (u and v components), and surface downward solar radiation. The 2-meter air temperature reflects diurnal variations and seasonal trends across a wide range of environments and is strongly correlated with building energy demand (Li et al., 2018; Mishra et al., 2024). Dew point temperature, in relation to air temperature, captures atmospheric humidity variations. In winter, higher humidity enhances thermal comfort at lower temperatures, while in summer, it intensifies perceived heat stress and elevates cooling energy requirements, as occupants tend to set lower setpoint temperatures to achieve comfort (Pawlak et al., 2021; Choi et al., 2017; Mansouri et al., 2023). Wind speed affects the heat transfer process through the building envelope (Cholewa et al., 2021). Higher wind speeds typically increase heating energy demand during winter through enhanced heat losses and may reduce cooling loads in summer due to improved natural ventilation (Hirth et al., 2024). However, the impact of wind speed varies by region, with some areas showing little sensitivity (e.g., Chambers & Oreszczyn, 2019). Solar radiation provides beneficial passive heat gains during winter, thereby

https://doi.org/10.5194/essd-2025-709
Preprint. Discussion started: 28 December 2025

reducing building heating needs, but can exacerbate overheating in summer, increasing cooling demand (Huang & Kang, 2021; Zhao et al., 2024).

The spatiotemporal heterogeneity of meteorological variables, combined with the varying sensitivity of building energy demand to different meteorological conditions, adds complexity to the accurate estimation of *Tbase*. Among these variables, air temperature is the most dominant factor affecting building energy demand, while humidity, wind speed, and solar radiation function as correction terms that modify temperature effects. Their physical interdependence with temperature amplifies the nonlinear dynamics of building energy demand and contributes to the uncertainty in determining *Tbase*.

Static features

Static features include socioeconomic variables, along with the geographic coordinates (longitude and latitude) of each region, as shown in Table 1. Population density is critical in shaping regional microclimate through urban densification, which in turn affects thermal comfort. In densely built urban environments, reduced wind speeds in narrow canyons limit heat circulation and removal, thereby worsening thermal conditions (Perera et al., 2023). This degraded thermal environment forces residents to rely more on mechanical cooling during the summer months. In addition, per capita heating energy demand varies significantly across population density ranges (Zarco-Periñán et al., 2021), which reflects the heterogeneous impact of population density on energy demand. Regional affluence, characterized by economic density (GDP per unit area), also strongly influences residential energy access and consumption patterns (Rode et al., 2021; Abajian et al., 2025; van Ruijven et al., 2019). Typically, in low-income areas, the adoption of air conditioning is often delayed due to financial constraints (Duan et al., 2023), resulting in higher summer base temperatures. Conversely, in high-income areas, cooling systems are typically activated at lower ambient temperatures. This disparity is particularly pronounced in rural-urban comparisons, where urban populations tend to exhibit lower temperature thresholds for initiating electricity demand (Shi et al., 2025).

Geographic coordinates (longitude and latitude) serve as fundamental determinants of regional climate variability and influence local populations' adaptive behavioural responses to thermal conditions. For instance, residents in high-latitude regions tend to be more acclimated to cold climates and typically initiate heating systems at lower temperatures than those in lower-latitude counterparts (Yilmaz & Canan, 2025).

2.2 Model development

This study developed a novel framework to spatially generate an accurate dataset of base temperatures at the regional scale by integrating a statistical model and a Bidirectional Long Short-Term Memory (BiLSTM)-based neural network framework (Fig.1-a). First, the target data for the training set is prepared by a statistical model. The BiLSTM model is then trained to obtain optimal parameters, incorporating both stratified sampling and an attention mechanism. Stratified sampling ensures a more representative division of the dataset into training and test sets, while the attention mechanism enhances prediction accuracy by automatically focusing on the most relevant time steps for *Tbase* estimation. The final step is the generation and assessment of *Tbase* values at the regional scale.

2.2.1 Target base temperature data preparation

https://doi.org/10.5194/essd-2025-709
Preprint. Discussion started: 28 December 2025

Two types of base temperature, the cooling base temperature (*Tcool*) and the heating base temperature (*Theat*), are used as training targets. These targets are derived from the nonlinear relationship between actual energy demand and outdoor temperature across different regions. To adequately capture seasonal changes in energy use, the time series of daily energy demand should be at least one year long, while monthly energy demand data should be at least 5 years long. Furthermore, location metadata is essential for each energy demand dataset, as regional climate-energy relationships are critical for accurately estimating spatial variations in *Tbase*. Detailed information on the datasets used in this research is provided in Table A1.

Regression approaches are used to estimate *Tbase* values by modeling the relationship between energy demand and outdoor temperature. This relationship can be represented using a variety of specifications, including linear, polynomial, and exponential models (Fung et al., 2006; Ihara et al., 2008; Li, 2018). Although complex nonlinear models can provide precise model fits, they often pose challenges such as overfitting and limited interpretability of parameters, which limits their applicability across diverse regions (Staffell et al., 2023). To address these limitations, this study adopts a segmented linear regression model, grounded in the energy signature method (Bhatnagar et al., 2018), to identify the critical turning points in the energy demand-temperature relationship, namely the cooling and heating base temperatures (*Tcool* and *Theat*). The segmented linear regression model is expressed as follows:

$$E_{ti} = \begin{cases} a + b_1(Theat - T_{ti}), & T_{ti} < Theat \\ a, & Theat \le T_{ti} \le Tcool \\ a + b_2(T_{ti} - Tcool), & T_{ti} > Tcool \end{cases} \quad (1)$$

where E_{ti} represents energy demand at location i on day (or month) t, and T_{ti} denotes the corresponding average outdoor temperature. *Theat* and *Tcool* represent the heating and cooling base temperatures at location i, respectively. Coefficients b_1 and b_2 denote the slopes of heating and cooling demand. Specifically, $b_1 < 0$ indicates that heating demand decreases as outdoor temperature increases, while $b_2 > 0$ indicates that cooling demand increases with rising temperature. When outdoor temperature T_{ti} falls between *Theat* and *Tcool*, the building does not require active heating nor cooling, and the energy demand reflects the baseline load independent of temperature variations. When $T_{ti} < Theat$, heating is required to maintain indoor thermal comfort, and when $T_{ti} > Tcool$, cooling is necessary.

The segmented linear regression model identifies *Tbase* values through a three-step process. (1) Determining the number of base temperatures: for each energy-temperature dataset, both two-segment linear regression and quadratic regression are performed. The number of *Tbase* values is determined by comparing their R^2 values. When the quadratic regression yields a higher R^2, the energy-temperature relationship is likely U- or V-shaped, indicating the presence of both heating and cooling base temperatures. Otherwise, only a single *Tbase* is identified: a positive slope in the non-horizontal segment corresponds to *Tcool*, while a negative slope indicates *Theat*. (2) Calculating base temperatures: for dataset with a single base temperature, *Tbase* corresponds to the breakpoint of the fitted line. For datasets with two base temperatures, the minimum temperature point (Tmin) from quadratic regression is used as a reference. Intervals of (Tmin-6 °C, Tmin) and (Tmin, Tmin+8 °C) are examined at 0.05 °C increments to search for two breakpoints maximizing the R^2 of a three-segment linear regression, and the resulting breakpoints are identified as *Theat* and *Tcool*, respectively. Studies have shown that people usually feel cold when the ambient temperature is 4°C below the optimal temperature and feel hot when it is 5°C above the optimal temperature (Krause et al., 2023). This perception of cold and heat directly determines whether the cooling or heating system needs to be started. Tmin represents the ambient temperature with the lowest energy consumption

https://doi.org/10.5194/essd-2025-709
Preprint. Discussion started: 28 December 2025

and can be regarded as the optimal temperature. Considering that environmental factors such as differentiated humidity will affect the actual perception of cold and heat, it is necessary to expand the search range on the 4-5°C threshold; at the same time, people have more active cooling measures (ventilation, reducing clothing, shading, etc.) under hot conditions, while the response measures under cold conditions are relatively limited, so the search range on the hot side is appropriately expanded. Based on the above considerations, the search range is finally set to (Tmin-6 °C, Tmin+8 °C) to fully capture the temperature threshold that causes significant changes in energy consumption. (3) Producing the target dataset: detected *Tbase* values are retained only if the segmented regression achieves an R^2 above 0.5, ensuring an adequate model fit, and if the resulting *Theat* and *Tcool* values are physically plausible (i.e., *Theat* < *Tcool*, and both values fall within typical comfort temperature ranges of 10-30°C). Ultimately, base temperature values from 131 out of 172 geographical units are included in the final labeled dataset, as shown in Fig.1-b.c.d and Table A2.

Building-level energy demand datasets require careful pre-processing prior to segmented linear regression to ensure reliable analysis. For each dataset, three pre-processing steps are performed: (1) The energy demand time series for each building is visualized to exclude those without cyclical variation patterns, as such buildings typically exhibit temperature-independent demand (e.g., warehouses) (Staffell et al., 2023); (2) Daily energy demand of all buildings within each geographical unit are aggregated and averaged to represent regional building energy demand; (3) Using two-dimensional energy-temperature data, a density-based clustering method is used to eliminate abnormal daily records, which are often the results of data errors or abnormal demand patterns due to holidays. For region10-level datasets, only outlier detection is performed, as their monthly temporal resolution naturally smooths out short-term fluctuations such as holiday effects.

2.2.2 Model training

The model has been trained using *Tbase* values and corresponding feature data from 131 geographic units. Monthly-scale meteorological data are chosen as dynamic features because their temporal smoothing effectively filters out high-frequency noise while retaining medium- and long-term trends. These retained signals are more strongly correlated with energy demand and *Tbase*, enabling the model to capture more stable physical relationships and enhance spatial generalization.

Weather data exhibit complex time-series characteristics, including cyclical patterns, seasonal variations, and long-term trends, which are closely linked to variations in occupant energy demand for cooling and heating across different regions. To address the temporal complexity of meteorological data and their nonlinear relationships with energy demand, this study constructs a neural network framework based on a Bidirectional Long Short-Term Memory (BiLSTM) architecture to model the relationship between time-series weather data and *Tbase* (Fig.1-a).

LSTM networks represent an advancement over traditional recurrent neural networks (RNNs), effectively addressing the fundamental challenge of preserving information across extended sequences—also known as long-term memory capacity, which makes them particularly well suited for processing time-series data (Hochreiter & Schmidhuber, 1997). Each LSTM cell integrates three gates (input, forget, and output) and a storage unit (Zhang et al., 2024). The input gate regulates how much new information is added to the memory. The storage unit preserves historical information across time steps, enabling the network to maintain long-term dependencies. The forget gate determines which information is discarded, while the output gate determines which information is passed on to the next hidden state (Kim et al., 2024b; Chen et al., 2023). By controlling whether to retain or discard

https://doi.org/10.5194/essd-2025-709
Preprint. Discussion started: 28 December 2025

information, LSTM networks mitigate the vanishing and exploding gradient problems that commonly affect traditional RNNs during training.

The BiLSTM model extends the standard LSTM by processing input sequences in both forward and backward directions (Palazzoli et al., 2025). This bidirectional architecture captures temporal dependencies more comprehensively, enabling the model to learn complex patterns from both past and future contexts in historical weather data. The forward processing captures the temporal evolution of weather patterns, while the reverse processing identifies important contextual relationships that may be missed in a unidirectional processing. By leveraging this full temporal context, the BiLSTM model yields more accurate estimates of *Tbase* values.

In the BiLSTM framework, monthly meteorological data at the regional level and annual socioeconomic data by region from 2000 to 2020, as well as latitude and longitude, are used for model training. For dynamic features, sequence normalization is applied within each geographic unit to preserve the periodicity and seasonality inherent in meteorological variables. For static features, global standardization across all units is performed to highlight absolute differences between geographic units. Furthermore, latitude and longitude are encoded to capture rich location information. A sequence length of 24 months is chosen, which provides a balance between capturing patterns of climate variability and maintaining a sufficient number of training samples. To address the issue of sparse and unevenly distributed samples, K-means clustering is applied using geographic coordinates (latitude and longitude) to classify the 131 geographic units into three clusters. Training and test sets is then constructed proportionally within each cluster, ensuring stratified sampling and preserving the geographical representativeness of the training set—thus enhancing the model's capacity for spatial generalization. The final train-test split follows an 8:2 ratio. To present data leakage, all sequences from each geographic unit are assigned exclusively to either the training or test set. Model training adopts an early stopping strategy, halting if no improvement is observed over 20 consecutive epochs, with a maximum of 50 epochs.

The mean squared error (*MSE*) is used as the loss function, which is particularly sensitive to large errors and thus encourages the model to minimize substantial prediction biases. This study designs an adaptive *MSE* to handle missing target variables in certain regions, such as *Tcool* being absent in cold regions or *Theat* being unavailable in hot regions. The loss function automatically identifies and skips missing values through a dynamic masking mechanism, ensuring that only valid labels are used for error calculation. This design enables the model to fully utilize the label data, improving the efficiency of data utilization and the stability of model training. For each sample in an epoch, the loss function is defined as:

$$L_i = \frac{1}{|M_i|}\sum_{j\in M_i}\left(Tbase_{i,j} - \widehat{Tbase}_{i,j}\right)^2 \quad (2)$$

where $Tbase_{i,j}$ is the target value j of sample i, where $j \in \{0,1\}$ corresponding to *Tcool* and *Theat*, respectively. The $\widehat{Tbase}_{i,j}$ is the predictive value of $Tbase_{i,j}$. *Mi* represents the set of indexes of non-missing labels in sample *i*, and |*Mi*| indicates the number of non-missing labels. Given that each region contains at least one valid label, |*Mi*| $\in \{1,2\}$. The total loss per epoch is:

$$L_{batch} = \frac{1}{N}\sum_{i=1}^{N} L_i \quad (3)$$

where *N* is the number of samples.

In addition, the correlation coefficient are further considered to evaluate the overall performance of the prediction model. The formula for *r* is as follows:

https://doi.org/10.5194/essd-2025-709
Preprint. Discussion started: 28 December 2025

$$r = \frac{\sum_{m=1}^{K}(Tbase_{obse,m} - \overline{Tbase})(Tbase_{pred,m} - \widehat{Tbase})}{\sqrt{\sum_{m=1}^{K}(Tbase_{obse,m} - \overline{Tbase})^2}\sqrt{\sum_{m=1}^{K}(Tbase_{pred,m} - \widehat{Tbase})^2}} \quad (4)$$

where K represents the number of geographical units, $\overline{Tbase}$ represents the mean of the observations. $Tbase_{obse,m}$ is the observation value of the m^{th} geographic unit; $Tbase_{pred,m}$ is the predicted value of the m^{th} geographic unit; $\widehat{Tbase}$ is the the mean of the predicted values.

The hyperparameters for the BiLSTM model are determined through combinatorial optimization, with model performance evaluated using the MSE as the selection criterion. The hyperparameter search space included hidden state dimensions (128, 256), dropout rates (0.2, 0.3), and learning rates (0.001, 0.002, 0.0005). Based on experimental comparisons, the optimal configuration was selected as a dropout rate of 0.2, a learning rate of 0.002, and a hidden layer size of 256.

2.2.3 Global data generation

By utilizing provincial-level monthly meteorological data, annual socioeconomic data aligned with the training set, and geographic coordinates (Table 1), the above estimation procedure for *Tbase* can be extended to the global scale. Global regional administrative boundaries are obtained from GDAM (https://gadm.org/data.html), with second-level boundaries selected to correspond to the U.S. state or Chinese province scales. Considering the validity of the population mask, a total of 3385 spatial units with valid feature data are obtained globally. Theoretically, each provincial unit corresponds to two base temperatures: *Tcool* and *Theat*. This study adopts a unified modeling strategy that does not predefine the ranges or distribution patterns of *Tbase* based on traditional climate zoning. This approach allows the model to learn the distribution pattern of dynamic features from the data, avoiding fixed climate zoning that might obscure climate change evolution, thus revealing *Tbase* in a more objective way.

3 Results and discussion

3.1 Model performance

The performance of the BiLSTM model is assessed using *MSE* and *r* by comparing actual values with predicted values. The results showed that the MSE of the training set is 1.94 ℃², while the test set achieved an MSE of 1.78 ℃², with corresponding RMSE values of 1.39 ℃ and 1.33 ℃, respectively, indicating good model generalization without significant overfitting. Target-specific analysis revealed differences in predictive performance: *Tcool* achieved a strong correlation (r = 0.84) while *Theat* showed a moderate correlation (r = 0.70). The lower performance in predicting *Theat* compared to *Tcool* reflects the greater complexity of heating demand patterns, as evidenced by Kim et al. (2024a), who found that *Theat* showed significantly higher variability than *Tcool* when differentiating buildings by construction year and floor area in South Korea.

3.2 Global *Tbase* analysis

Global *Tbase* mapping reveals regionally differentiated temperature-dependent energy demand responses. Globally, *Tcool* ranges from 19 to 25 °C, with a mean of 22 °C and a median of 21.94 °C, while *Theat* ranges

https://doi.org/10.5194/essd-2025-709
Preprint. Discussion started: 28 December 2025

from 14 to 18 °C, with a mean of 16 °C and a median of 15.87 °C, which is consistent with physical logic (Fig. 2-a, b).

The *Tcool* and *Theat* present a generally consistent distribution pattern, characterized by lower *Tbase* values in high-latitude regions and higher values in low-latitude regions (Fig. 2-c). North America and Europe, which are predominantly composed of developed countries, have the lowest median *Tcool* and Theat. Higher economic levels enable residents to utilize cooling services at lower outdoor temperatures to meet greater comfort demands (Cong et al., 2022), well-insulated buildings require heating only when outdoor temperatures drop to lower thresholds. Africa and South America—located in the middle and low latitudes—show higher *Tbase* values. A higher *Tcool* means historical adaptation of residents to warmer climates, that is, they can achieve a certain degree of comfort in hotter conditions (Djongyang & Tchinda, 2010; Aljawabra & Nikolopoulou, 2018). A higher *Theat* is closely related to the living conditions in these regions, which are characterized by high population density and low economic output. In many areas with widespread informal settlements, UN-Habitat reports that as of 2018, 237.84 million people lived in slums in Sub-Saharan Africa (UN-Habitat, 2020). The rapid heat loss caused by poor building performance in these regions is due to the prevalence of substandard heating and cooling systems (Bah, 2018), as well as the low thermal inertia of housing facilities (Correia & Cristina Delgado, 2025; Pohl et al., 2021). However, Libya, Gabon, Algeria, Botswana, and South Africa in the African region have relatively low Theat, indicating that residents there benefit from relatively liveable housing conditions supported by housing renovation policies (Benzaama et al., 2021; Santos et al., 2022). Asia spans the longest latitudinal range and exhibits the largest inter-regional economic disparities, resulting in the widest distribution of *Tbase* values (*Tcool*: 19.57 - 24.99℃, *Theat*: 14.18 - 18.05℃). Southern China and India have higher *Tbase* values. A base temperature test for India based on the energy signature method showed that the median *Theat* in many regions is 17.4℃ (Bhatnagar et al., 2018), which closely aligns with the median *Theat* of 17.05℃ found in this study. Oceania's *Tbase* is more dispersed, which is due to the smallest number of regions included. Differentiated *Tbase* will provide strong support for energy-saving planning, building renovation, and climate-adaptive design across various regions around the world.

3.3 Technical validation

To further demonstrate the effectiveness of the predicted *Tbase*, a comprehensive validation framework is employed, as shown in Fig.3. In this framework, energy demand datasets from 16 regions are employed to perform external validations of *Tbase* (Table A3). These datasets cover major continents globally, with relatively uniform distribution to ensure the comprehensiveness and representativeness of the validation results.

Three validation approaches are employed to assess whether predicted *Tbase* improves the accuracy of temperature-dependent energy demand modelling, including direct validation, indirect validation and sensitivity analysis (Fig.3). Direct validation applies the segmented linear regression method described earlier to determine *Tbase* for each dataset, and then compares these results with the predicted values to validate the accuracy and reliability of the prediction results in practical applications. Indirect validation uses both the predicted *Tbase* and reference values from literature or official sources to forecast building energy demand through degree-day modeling. For each energy dataset, the following indirect validation procedure is adopted to evaluate the accuracy of the predicted *Tbase*. First, based on the corresponding temperature data, two sets of degree days are calculated: (1) HDD and CDD calculated using the predicted *Tbase* from this study; (2) HDD and CDD calculated using

https://doi.org/10.5194/essd-2025-709
Preprint. Discussion started: 28 December 2025

reference *Tbase* values from literature or official sources. Subsequently, statistical analyses are performed on the above two sets of degree days with actual energy consumption data, including: (1) correlation analysis, calculating Pearson correlation coefficients to quantify the strength of linear relationships between degree days and energy consumption; (2) regression analysis, establishing degree day-energy consumption regression models and calculating root mean square error (RMSE) to evaluate prediction accuracy. Finally, the modeling performance of the predicted *Tbase* is evaluated through comparative analysis. If the HDD and CDD calculated based on the predicted *Tbase* exhibit higher correlation coefficients with actual energy consumption and the corresponding regression models have lower RMSE, this indicates superior accuracy. For sensitivity analysis, the time series correlation coefficient (TCC) is used to analyse how changes in *Tbase* affect the correlation between energy demand and HDD or CDD, thereby evaluating the performance of the predicted *Tbase*.

For the 16 energy demand datasets at the regional and building levels, the validation results are summarized in Table 2. Both direct and indirect validation results show that the predicted *Tbase* performs well across most datasets. Direct validation yields average errors of 0.68℃ for *Tcool* and 1.08℃ for *Theat* across the 16 datasets. Indirect validation demonstrates that the predicted *Tbase* significantly improves the accuracy of building energy demand modeling in two key aspects. First, the TCC between energy demand and HDD/CDD calculated using the predicted *Tbase* is greater than or equal to that obtained using reference *Tbase* values. This indicates that predicted *Tbase* maintains or enhances the explanatory power of temperature-dependent energy demand variations. On average, the predicted *Tbase* improves TCC across the validation datasets from 0.67 to 0.81. Second, the predicted *Tbase* significantly reduces the RMSE of both cooling and heating energy demand modeling. Specifically, the RMSE of cooling and heating energy demand modeling decreases by an average of 10.01% and 10.02%, respectively.

The sensitivity analysis evaluates the proximity of the TCC value derived from the predicted *Tbase* to the maximum achievable TCC value. A TCC value approaching the maximum indicates higher accuracy of the predicted *Tbase* and enhanced performance in temperature-dependent energy demand modeling. Fig. A1-A14 demonstrate that across all datasets, the TCC values calculated using the predicted *Tbase* approximate their respective maxima, confirming the effectiveness of the predicted *Tbase* parameters in HDD/CDD-based modeling. Although three datasets show prediction errors greater than 2°C, the indirect validation results indicate that predicted *Tbase* still contributes to enhanced energy demand modeling accuracy, as evidenced by consistent RMSE reductions in all instances. Overall, the predicted *Tbase* improves the ability to explain and model temperature-dependent energy demand variations.

3.3.1 Building-level validation

In order to present the validation process more clearly, the detailed results of a building-level dataset and a regional-level dataset are presented below. Fig.4 shows the validation results of a residential building in South Australia, Australia. The residential building is powered entirely by photovoltaics, and the dataset contains hour-scale load data (Sharma et al., 2019). Direct validation shows absolute prediction errors of 0.3°C between the predicted *Tcool* (21.4°C) and the actual value (21.10°C), and 0.6°C between the predicted *Theat* (15.5°C) and the actual value (16.10°C) (Fig. 4-a). Fig. 4-b demonstrates that the TCCs corresponding to *Tcool* and *Theat* approach their respective maximum values across the tested base temperature range, implying that the predicted *Tbase* effectively captures temperature-dependent energy demand variations. Following previous studies, 14°C and 24°C

https://doi.org/10.5194/essd-2025-709
Preprint. Discussion started: 28 December 2025

are adopted as reference base temperatures for this dataset (Livada et al., 2021), and indirect validation is performed to compare the reliability of predicted versus reference *Tbase* values in residential energy demand modeling. The TCC calculated using predicted *Tbase* (15.5°C/21.4°C) reaches 0.91, exceeding the TCC based on the reference *Tbase* (14°C/24°C) of 0.88, representing a correlation improvement of 0.03 (Fig. 4-c). For heating energy demand modeling, RMSE decreases from 1.82 kWh to 1.69 kWh, achieving a relative improvement of 7.14% (Fig. 4-d, f). For cooling energy modeling, RMSE is reduced from 1.84 kWh to 1.45 kWh, corresponding to a relative improvement of 21.20% (Fig. 4-e, g). These results indicate that energy demand modeling based on predicted *Tbase* is more reliable, although consistent with the research of Staffell et al. (2023), regional *Tbase* have higher accuracy in cooling energy demand modeling.

3.3.2 Region-level validation

For the regional-level energy demand dataset, Fig. 5 presents the validation results of daily energy demand modeling in Tacoma, USA. To extract residential energy demand from the raw dataset, the residential sector's share of annual end-use energy demand from the U.S. Energy Information Administration (EIA) (https://www.eia.gov/) has been applied to daily total energy consumption data to derive daily residential energy demand. The direct validation results suggest that this area has significant heating demand, with an error of only 0.15°C between the predicted *Theat* (14.5°C) and the actual value (14.35°C) (Fig. 5-a). Fig. 5-b shows that the TCC corresponding to *Theat* is very close to its maximum value, demonstrating that the predicted *Tbase* effectively captures residential energy demand variations. The official value of 18°C is adopted as the reference *Tbase* for residential energy demand modeling in this dataset. The TCC calculated using the predicted *Tbase* (14.5 °C) reaches 0.94, exceeding the TCC based on the reference *Tbase* (18°C) of 0.91, representing a correlation improvement of 0.03 (Fig. 5-c). The RMSE is reduced from 172.65 MWh to 150.46 MWh, corresponding to a relative improvement of 12.85% (Fig. 5-d, e). These results demonstrate that the predicted base temperatures have good applicability in regional-scale energy demand modeling.

3.4 Potential application of the *Tbase* dataset

Efforts to enhance degree-day models persist in order to improve efficiency and accuracy in building energy modeling. Numerous studies have calculated region-specific *Tbase* across Europe, Africa, India, and globally (Wenz et al., 2017; Bhatnagar et al., 2018; Staffell et al., 2023), significantly improving the accuracy of regional building energy demand modeling. However, the heavy reliance on actual building energy demand datasets constrains the extension of *Tbase* estimation to broader detailed spatial scales. This study overcomes this critical bottleneck through a machine learning approach.

The *Tbase* dataset developed in this study has significant potential for use in a wide range of areas related to degree-day modeling. Firstly, applying *Tbase* to heating and cooling energy demand modeling on a consistent global scope can provide scientific support for differentiated energy system planning and climate adaptation strategies in building design. Secondly, *Tbase* can be used for a wide range of heat exposure assessments, contributing to studies on the health dimensions of climate change impacts and providing data to support strategies for heatwave warning and the protection of vulnerable populations. Finally, *Tbase* holds significant value in large-scale model coupling. Integration of the *Tbase* datasets into various Earth System Models (e.g., CESM) and Integrated Assessment Models (e.g., GCAM, IMAGE) can effectively reduce long-term modeling uncertainties.

https://doi.org/10.5194/essd-2025-709
Preprint. Discussion started: 28 December 2025

In turn, incorporating these more accurate representations of energy demand modeling results into energy-economic models (e.g., TIMES, MARKAL) can provide a reliable modeling framework for designing decarbonization pathways. Beyond energy and health applications, the *Tbase* dataset can support weather-related financial products, such as temperature-based derivatives and parametric insurance, which rely on accurate HDD/CDD calculations to help weather-sensitive industries hedge against climate-related financial risks (e.g., Berhane et al., 2021).

3.5 Limitations and uncertainties of the *Tbase* dataset

Limitations and uncertainties related to the research design and model should be highlighted. Regrading the limitations, firstly, the dataset's reliance on electricity and natural gas consumption statistics inadequately captures heating demand from alternative energy sources. Especially in energy-poor regions, the decentralized nature of heating makes it difficult to account for the use of polluting sources, such as coal and diesel (Burguillo & Juez-Martel, 2024; Fleiter et al., 2016). Also, district heat does not require high temporal resolution monitoring to ensure system stability, and the high cost of heat meters limits their adoption—unlike smart meters used for electricity (Kozarcanin et al., 2019). This limitation in data acquisition makes it difficult to accurately characterize heating demand patterns across time and temperature. Secondly, given the unavailability of extensive data, building characteristics such as type, age, floor area, insulation performance, number of floors, and other attributes are not differentiated in this study. The *Tbase* dataset, therefore, provides generalized energy demand estimates rather than building-type-specific predictions. This limitation particularly affects demand-responsive heating demand modeling, since the energy flexibility potential of a building is mainly related to the level of insulation in the envelope during the heating season, that is, well-insulated buildings maintain indoor temperatures longer, even if heating is temporarily reduced or stopped. In contrast, during the cooling season, energy flexibility is mainly influenced by the timing of active demand response events (Vivian et al., 2020). This also explains why the prediction accuracy for *Theat* in this study is lower than that for *Tcool*. In addition, the dynamic changes in *Tbase* were not taken into account due to the limitations of acquiring extensive long-term energy consumption data.

The effects of data unavailability described above will be passed on to the model, leading to uncertainty in the *Tbase* results. Therefore, this study focuses on analyzing how the model limitations affect the reliability of *Tbase*, which can be decomposed into aleatoric (data noise) and epistemic (model knowledge) uncertainty (Xu et al., 2025; He et al., 2025). The epistemic uncertainty of the *Tbase* dataset was estimated by 30 data sampling analyses using MC Dropout for meteorological, socioeconomic, and geographic background data collected for 3,385 spatial units. The MC Dropout employs a neural disconnection mechanism based on randomness to evaluate network sensitivity to model variations (Xu et al., 2025). This approach is widely used for uncertainty quantification due to its computational simplicity and efficiency. Uncertainty was quantified by calculating the standard deviation (std) across the results of multiple analyses, where larger values indicate greater uncertainty in *Tbase* estimates due to model variability or limited data. The results are shown in Fig. 6. The uncertainty ranges for *Theat* and *Tcool* are [0.56, 1.61] and [0.82, 2.26] ℃, respectively. Although *Theat* shows lower model accuracy, its target distribution is more concentrated (the interquartile range, IQR, is narrower, see Fig.1-d), resulting in a smaller uncertainty interval. By focusing on the maximum uncertainty and normalizing by their respective mean values (22°C for *Tcool*, 16°C for *Theat*), both variables exhibit comparable relative uncertainties of approximately 10%, indicating consistent uncertainty quantification across different targets. The spatial distribution of std shows

https://doi.org/10.5194/essd-2025-709
Preprint. Discussion started: 28 December 2025

that the uncertainty is at a medium-low level in most regions, such as North America, Europe, northern Asia, and Oceania. The training samples in Africa, South America, and western Asia are sparse, and the std is relatively high. In the future, further increasing the sample size in sparsely sampled regions will reduce the epistemic uncertainty of the model.

4 Data availability

The proposed global *Tbase* dataset can be acquired from https://doi.org/10.6084/m9.figshare.30646376.v2 (He et al., 2025).

5 Conclusion

This study developed a global regional-scale *Tbase* dataset, including both *Tcool* and *Theat*, by integrating comprehensive energy demand, meteorological, and socio-economic data using a BiLSTM framework with an attention mechanism. The *Tbase* values varied across regions worldwide, with *Tcool* ranging from 19 to 25°C (mean: 22°C; median: 21.94°C), and *Theat* ranging from 14 to 18°C (mean: 16°C; median: 15.87°C). Compared to the use of a fixed *Tbase*, these heterogeneous values better capture local climate characteristics, socioeconomic conditions, and energy use behaviours. The *Tbase* in this study is primarily applicable to general building energy demand modelling for assessing energy requirements to achieve universal thermal comfort.

The model construction incorporates several strategies to minimize the impact of small samples. First, stratified sampling based on geographic clustering ensures representative distribution between the training and test sets. Second, joint training of *Tcool* and *Theat* using an adaptive MSE-based loss function maximizes data utilization. Third, the attention mechanism captures critical temporal patterns to enhance training efficiency. Based on 131 regions (with 80% used for training and 20% for testing), internal validation shows RMSE values of 1.39℃ (training) and 1.33℃ (testing), with correlation coefficients of 0.84 for *Tcool* and 0.70 for *Theat*, demonstrating good generalization performance.

External validation using 16 independent building energy demand datasets from various global regions (completely separate from the 131 regions used in the model construction phase) demonstrates that predicted *Tbase* significantly improves the accuracy of energy demand modeling. The average prediction errors for *Tcool* and *Theat* are 0.68℃ and 1.08℃, respectively. Compared with official or empirical *Tbase* values, the predicted *Tbase* increases the average correlation between energy demand and CDD/HDD from 0.67 to 0.81 and reduces RMSE by 10.01% and 10.02% for cooling and heating energy demand modeling, respectively. Uncertainty analysis revealed epistemic uncertainties of 0.56-2.26°C, with higher values sparsely distributed, supporting the dataset's overall reliability. These results underscore the model’s ability to effectively capture regional heterogeneity in temperature-dependent energy demand responses.

The resulting dataset supports a wide range of regional-scale applications and increases research productivity across multiple domains. Most significantly, it improves the accuracy of degree-day energy demand models, enabling the exploration of how adjustments to temperature thresholds affect energy requirements. Additionally, the dataset facilitates climate change impact assessments, such as accurate heat exposure analysis. Finally, coupling the *Tbase* dataset with Earth system models and integrated assessment models has the potential to reduce uncertainty in climate-energy interactions, thereby enhancing long-term energy modeling performance. Overall,

https://doi.org/10.5194/essd-2025-709
Preprint. Discussion started: 28 December 2025

this spatially continuous *Tbase* dataset helps overcome key limitations in building energy modeling and opens new research opportunities in climate-responsive energy planning.

Author contributions

Y. Zhou contributed to the conceptualization, project supervision and administration, funding acquisition, and review. X. He contributed to the data analysis, methodology, writing and dataset submission. J. Eom contributed to the data analysis, review and editing. S. Yu contributed to investigation, supervision, review and editing. S. Liu contributed to the data curation and methodology. W. Xu contributed to review and editing.

Competing interests

The contact author has declared that none of the authors has any competing interests.

Acknowledgements

This work was supported by the University of Hong Kong HKU-100 Scholars Fund, the HKU Social Sciences Internal Seed Grant Scheme, and Research Grants Council-Strategic Topics Grant STG2/P-705/24-R. The authors would like to express their sincere gratitude to Dr. Steven J. Smith (Center for Global Sustainability, University of Maryland) for his valuable comments and suggestions, which greatly improved the quality of this manuscript.

https://doi.org/10.5194/essd-2025-709
Preprint. Discussion started: 28 December 2025

Table 1 Description of meteorological variables and static features

types	variables	unit	temporal resolution	source	description
dynamic	2m temperature	K	monthly	ERA5-Land (Muñoz-Sabater et al., 2021)	Calculated temperature and its lag features and rolling average
	2m dewpoint temperature	K			Calculated relative humidity
	surface solar radiation downwards	J m^{-2}			Calculated total energy received per unit area of surface
	10m u-component of wind	m s^{-1}			Calculated wind speed
	10m v-component of wind	m s^{-1}			
static	population density	persons/km^2	yearly	Lebakula et al., 2024	Population per unit area
	GDP	Int. $ 2017		Kummu et al., 2025	GDP per unit area
	latitude	-	-	GADM (https://gadm.org/data.html)	Latitude and longitude of the centre point
	longitude	-			

https://doi.org/10.5194/essd-2025-709
Preprint. Discussion started: 28 December 2025

Table 2 Summaries of validation

Location	predicted value		direct validation				indirect validation					
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
Austin, USA	20.4	14.8	19.8	0.6	14.1	0.7	-0.29	0.81	6.83%	29.99%	23.8	23.8①
Kitakyushu, Japan	20	14.4	19.12	0.88	13.97	0.43	0.04	0.76	12.27%	20.41%	26	18②
Lower Saxony, Germany	-	14.5	-	-	15.3	0.8	0.93	0.93	-	1.98%	-	15.5③
Trondheim, Norway	-	14.8	-	-	12.3	2.5	0.97	0.97	-	1.47%	-	15.5③
Agder, Norway	-	14.8	-	-	12.6	2.2	0.95	0.95	-	1.13%	-	15.5③
Northerneast, Mexico	20.1	-	19.8	0.3	-	-	0.73	0.75	2.60%	-	18④	-
South Australia, Australia	21.4	15.5	21.1	0.3	16.1	0.6	0.88	0.91	21.13%	6.92%	24	14⑤
Hawke's Bay, New Zealand	-	16.2	-	-	-	-	0.8	0.81	-	1.46%	-	14⑤
Itakyry, Paraguay	22.1	-	21.4	0.7	-	-	0.63	0.66	3.99%	-	24⑥	-
Presidente Franco, Paraguay	22.1	-	20.6	1.5	-	-	0.81	0.87	18.18%	-	24⑥	-
Acaray, Paraguay	22.1	-	20.7	1.4	-	-	0.69	0.75	8.91%	-	24⑥	-
Philadelphia, USA	20.2	14.7	20.1	0.1	12.8	1.9	0.62	0.66	15.90%	4.99%	18	18⑦
Tacoma, USA	-	14.5	-	-	14.4	0.1	0.91	0.94	-	12.85%	-	18⑦
Victoria, Australia	21	15.2	20.8	0.2	15.7	0.5	0.58	0.6	0.75%	2.46%	24	14⑦
Delhi, India	23.1	16.7	22.3	0.8	-	-	0.75	0.86	17.67%	26.51%	26	16⑧
Abidjan, Côte d'Ivoire	22.8	-	-	-	-	-	0.73	0.74	1.89%	-	26⑨	-
Mean	-	-	-	0.68	-	1.08	0.67	0.81	10.01%	10.02%	-	-

Note 1: I: predicted *Tcool* from BiLSTM framework; II: predicted *Theat* from BiLSTM framework; III: *Tcool* from segmented linear regression method; IV: Absolute error of *Tcool*; V: *Theat* from segmented linear regression method; VI: Absolute error of *Theat*; VII: TCC based on reference *Tbase*; VIII: TCC based on predicted *Tbase*; IX: decreased RMSE for CDD-energy demand modeling by predicted *Tcool*; X: decreased RMSE for HDD-energy demand modeling by predicted *Theat*; XI: reference *Tcool*; XII: reference *Theat*.

Note 2: ①: Feng et al., 2021; ②: Yuan et al., 2024; ③: Filahi et al., 2024; ④: Corrales-Suastegui et al., 2021; ⑤: Livada et al., 2021; ⑥: Kennard et al., 2022; ⑦: ASHRAE, 2009; ⑧: Borah et al., 2015. ⑨: Dicko et al., 2024.

https://doi.org/10.5194/essd-2025-709
Preprint. Discussion started: 28 December 2025

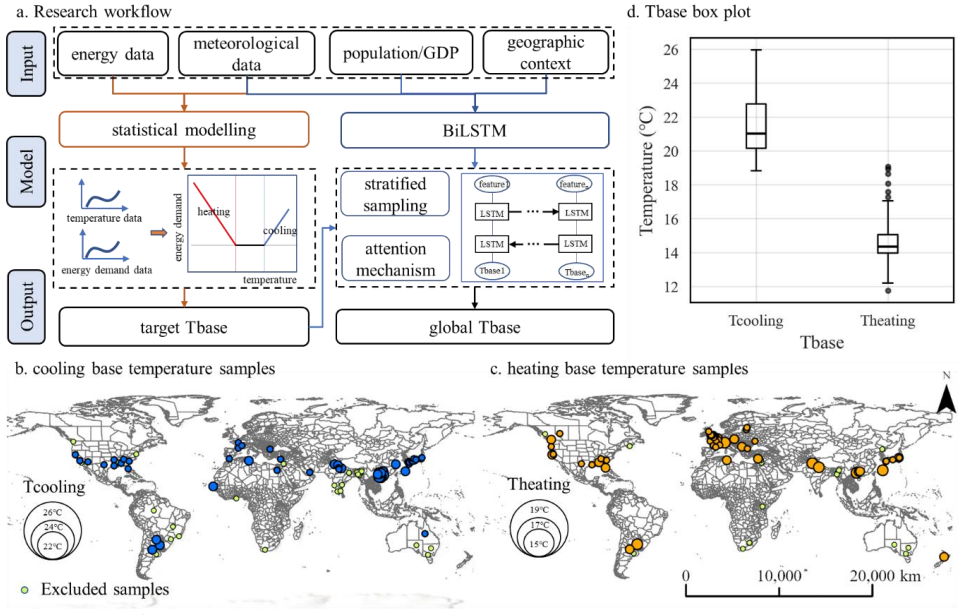


Figure 1: Workflow and Data Preparation.

(a: Research workflow based on a segmented linear regression model and BiLSTM; b: cooling base temperature samples; c: heating base temperature samples; d: Box plot of base temperature samples.)

https://doi.org/10.5194/essd-2025-709
Preprint. Discussion started: 28 December 2025

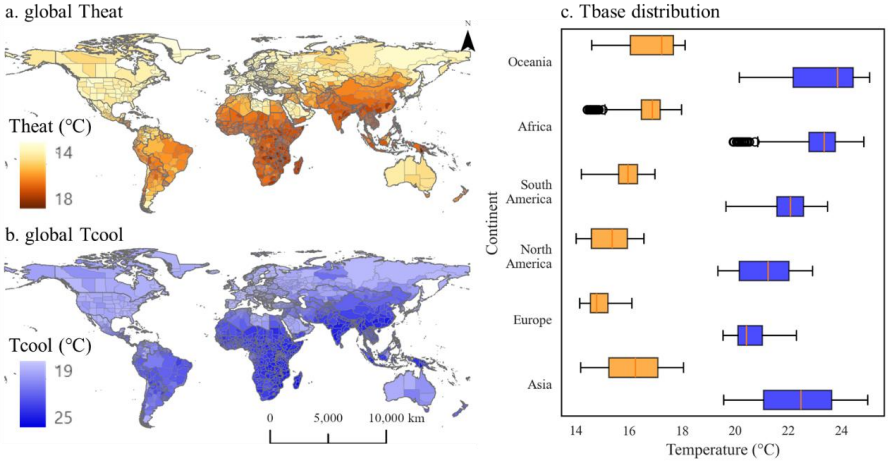


Figure 2: Global *Tbase* mapping based on the results of BiLSTM framework.

https://doi.org/10.5194/essd-2025-709
Preprint. Discussion started: 28 December 2025

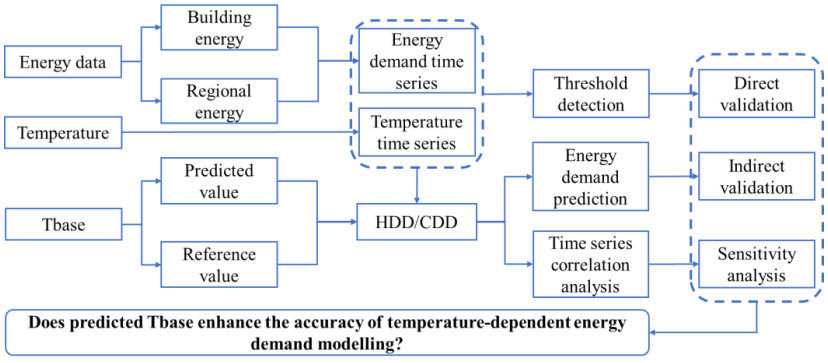


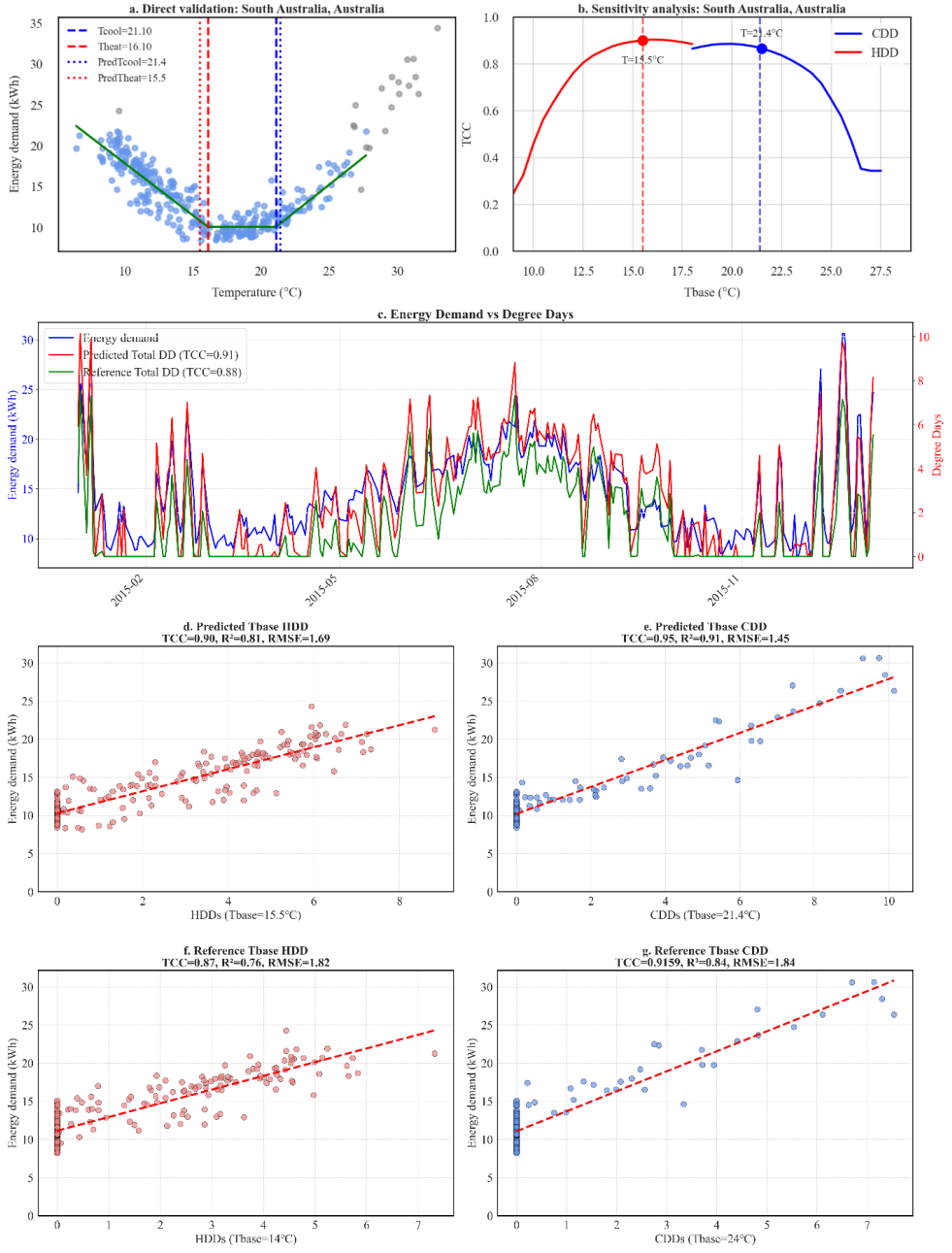
Figure 3: Workflow for technical validation.

650

655

660

https://doi.org/10.5194/essd-2025-709
Preprint. Discussion started: 28 December 2025



665

Figure 4: Visualization of CDD and HDD versus energy demand at different base temperatures of a building in South Australia, Australia.

https://doi.org/10.5194/essd-2025-709
Preprint. Discussion started: 28 December 2025

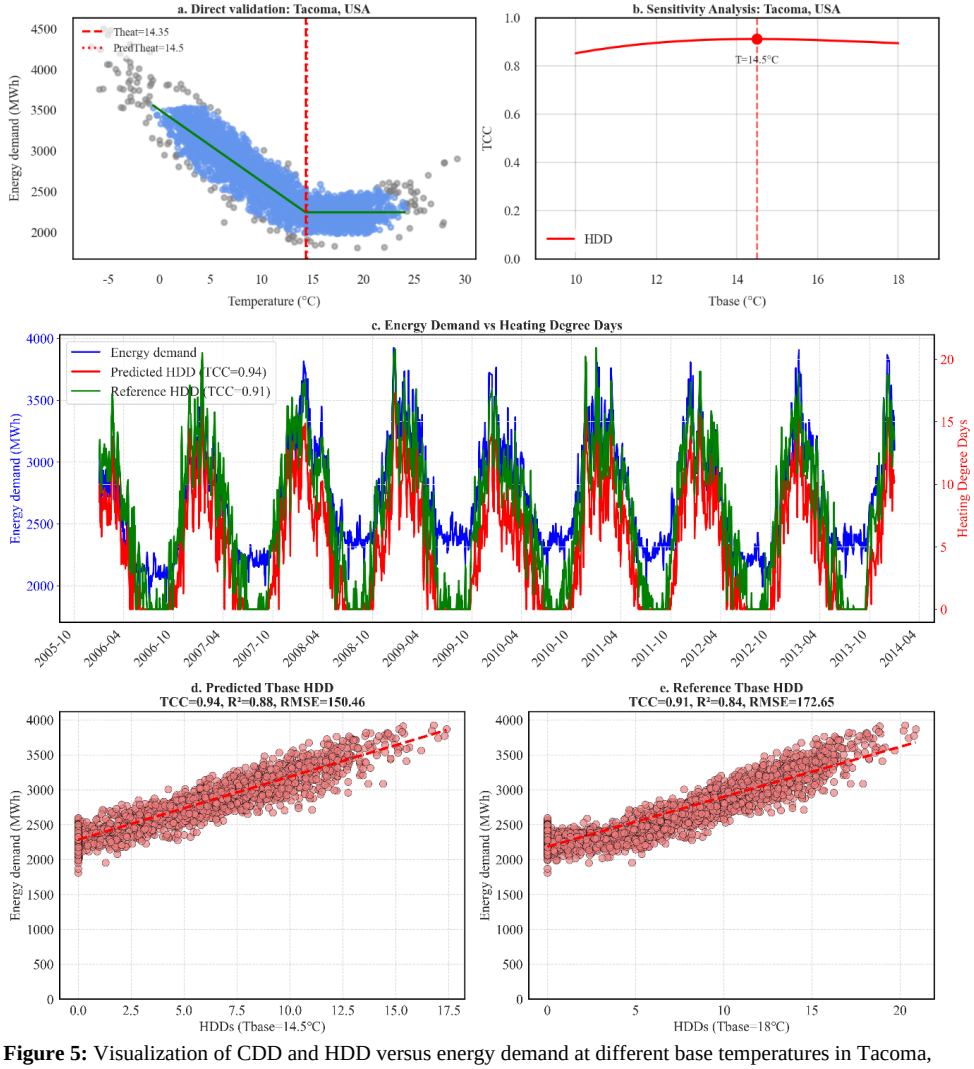


Figure 5: Visualization of CDD and HDD versus energy demand at different base temperatures in Tacoma, USA.

https://doi.org/10.5194/essd-2025-709
Preprint. Discussion started: 28 December 2025

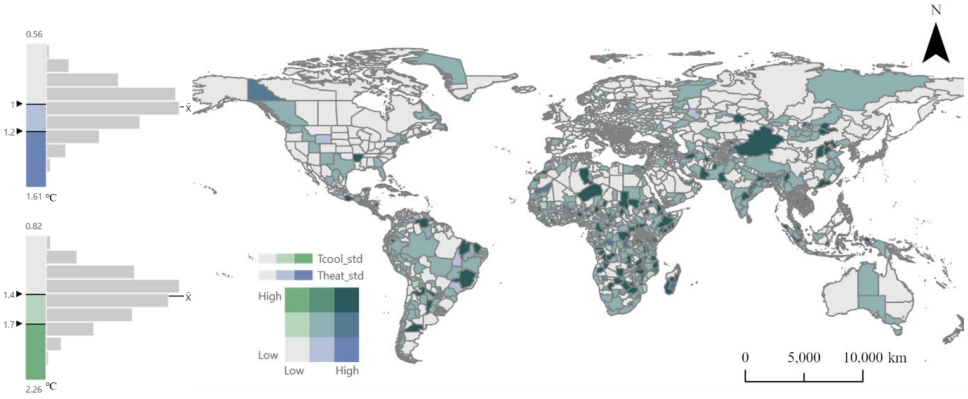


Figure 6: *Tbase* uncertainty analysis based on the limitations in the model.

https://doi.org/10.5194/essd-2025-709
Preprint. Discussion started: 28 December 2025

Appendix: Appendix A

Table A1 Details of the energy datasets used for calculating the target Tbase.

Country	Datasets	Year period	Time resolution	Gas / Electricity	Spatial units	Type	Source
Argentina	Base Demanda Diaria	2017-2025	monthly	Electricity	3	energy	CAMMESA, 2025
Australia	AEMO	2012-2020	daily	Electricity	4	energy	AEMO, 2025
Brazil	Empresa de Pesquisa Energética - EPE	2004-2024	monthly	Electricity	4	energy	EPE, 2024
Bulgaria	EMBER	2015-2025	monthly	Electricity	1	energy	Ember, 2025
Canada	HUE	2012-2020	hourly	Electricity	1	building energy	Makonin, 2019
Canada	-	2000-2007	hourly	Electricity	1	energy	Canadian Centre for Energy Information, n.d.
Canada	-	2008-2024	monthly	Electricity	3	energy	Canadian Centre for Energy Information, n.d.
China	residential electricity load data	2022-2023	hourly	Electricity	10	building energy	Li et al., 2025
China	High-resolution electric power load data	2016-2021	hourly	Electricity	1	building energy	Zhou et al., 2023
Cyprus	EMBER	2017-2024	monthly	Electricity	1	energy	Ember, 2025
Czech Republic	EMBER	2015-2025	monthly	Electricity	1	energy	Ember, 2025
Egypt	EMBER	2016-2024	monthly	Electricity	1	energy	Ember, 2025
Estonia	EMBER	2015-2025	monthly	Electricity	1	energy	Ember, 2025
France	Rte-France	2014-2024	monthly	Electricity	10	energy	RTE, n.d.
Germany	DEKN	2015-2019	minute	Electricity	1	building energy	Open Power System Data, 2020
India	Monthly energy demand dataset	2007-2011	monthly	Electricity	13	energy	Central Electricity Authority, 2024
India	Ecohen4/Energy	2011-2013	hourly	Electricity	2	energy	Cohen et al., 2015
India	Smart Meter Data: Mathura & Bareilly	2019-2021	minute	Electricity	2	building energy	Agrawal et al., 2021
Japan	Statistics-METI	2016-2024	monthly	Electricity	21	energy	Agency for Natural Resources and Energy, n.d.
Japan	Ecohen4/Energy	2006-2014	hourly	Electricity	1	energy	Cohen et al., 2015
Latvia	EMBER	2015-2025	monthly	Electricity	1	energy	Ember, 2025
Morocco	Power Consumption Prediction project	2017	minute	Electricity	1	energy	Salam et al., 2018
New Zealand	New Zealand Green Grid	2014-2016	minute	Electricity	2	building energy	B. Anderson et al., 2018
Pakistan	PRECON	2018-2019	minute	Electricity	1	building energy	Nadeem & Arshad, 2019
Senegal	Ecohen4/Energy	2011-2014	hourly	Electricity	1	energy	Cohen et al., 2015
Serbia	EMBER	2016-2025	monthly	Electricity	1	energy	Ember, 2025
Switzerland	Swiss Smart Meter Data	2019-2020	minute	Electricity	1	building energy	Afroz, 2023
the United Arab Emirates	SEWA Energy Demand data	2020-2021	daily	Electricity	1	energy	Rastogi, 2024
Tunisia	EMBER	2020-2024	monthly	Electricity	1	energy	Ember, 2025
UK	Energy Demand Research Project	2007-2010	minute	both	64	building energy	AECOM Ltd., 2014
Ukraine	EMBER	2014-2022	monthly	Electricity	1	energy	Ember, 2025
USA	Ecohen4/Energy	2003-2018	hourly	Electricity	5	energy	Cohen et al., 2015
USA	EMBER	2001-2024	monthly	Electricity	7	energy	Ember, 2025
USA	Building Data Genome Project 2	2016-2017	hourly	Electricity	2	building energy	C. Miller, 2020
USA	Residential power usage-Houston	2016-2020	hourly	Electricity	1	building energy	Srinuti, 2020

https://doi.org/10.5194/essd-2025-709
Preprint. Discussion started: 28 December 2025

Open Access Earth System Science Data Discussions

Table A2 Tcool and Theat values used for model training.

RegionID	Region	Tcool	Theat	RegionID	Region	Tcool	Theat
1	Chattanooga, USA	19.15	11.75	45	Melton, UK	-	14.19
2	Eugene, USA	-	13.55	46	North West Leicestershire, UK	-	14.48
3	Los Angeles, USA	20.06	-	47	Oadby and Wigston, UK	-	14.65
4	Sacramento, USA	19.85	13.85	48	North Kesteven, UK	-	14.25
5	Maricopa, USA	21.02	-	49	South Kesteven, UK	-	14.44
6	Alameda, USA	-	16.43	50	Daventry, UK	-	14.09
7	Alabama, USA	20.46	15.46	51	East Northamptonshire, UK	-	14.32
8	Arkansas, USA	21.08	-	52	Kettering, UK	-	14.17
9	Florida, USA	21.21	16.21	53	Northampton, UK	-	14.07
10	Louisiana, USA	20.18	14.73	54	South Northamptonshire, UK	-	14.46
11	South Carolina, USA	21.14	13.34	55	Wellingborough, UK	-	14.33
12	Texas, USA	20.28	14.45	56	Ashfield, UK	-	14.63
13	Vancouver, Canada	-	15.60	57	Bassetlaw, UK	-	14.78
14	Alberta, Canada	-	12.20	58	Broxtowe, UK	-	14.70
15	Auvergne-Rhône-Alpes, France	19.09	13.14	59	Gedling, UK	-	14.45
16	Bretagne, France	-	14.41	60	Mansfield, UK	-	14.02
17	Centre-Val de Loire, France	19.77	14.72	61	Newark and Sherwood, UK	-	14.37
18	Hauts-de-France, France	-	15.42	62	Rushcliffe, UK	-	14.20
19	Normandie, France	-	14.85	63	East Staffordshire, UK	-	13.77
20	Nouvelle-Aquitaine, France	-	13.95	64	Lichfield, UK	-	14.29
21	Occitanie, France	19.53	12.78	65	Staffordshire Moorlands, UK	-	13.99
22	Pays de la Loire, France	-	14.96	66	Elmbridge, UK	-	14.33
23	Île-de-France, France	-	15.21	67	Guildford, UK	-	14.46
24	Lucerne, Switzerland	-	17.05	68	Reigate and Banstead, UK	-	15.12
25	Tetouan,Morocco	20.31	-	69	Nuneaton and Bedworth, UK	-	14.17
26	Dakar,Senegal	22.25	-	70	Rugby, UK	-	13.88
27	Sharjah,United Arab Emirates	20.15	-	71	Warwick, UK	-	14.07
28	Chandigarh, India	22.08	-	72	Arun, UK	-	15.47
29	Delhi, India	21.40	-	73	Mid Sussex, UK	-	13.80
30	Lahore, Pakistan	23.22	18.07	74	Redditch, UK	-	14.35
31	Bareilly, India	22.42	17.27	75	Wyre Forest, UK	-	13.66
32	Mathura, India	19.45	-	76	Birmingham, UK	-	14.14
33	Tokyo, Japan	21.42	15.97	77	Coventry, UK	-	14.25
34	Kyoto, Japan	20.20	-	78	Dudley, UK	-	14.22
35	Saga, Japan	20.86	13.01	79	Sandwell, UK	-	13.90
36	Chiba, Japan	21.02	13.47	80	Solihull, UK	-	14.48
37	Wakayama, Japan	20.28	13.28	81	Walsall, UK	-	13.86
38	Osaka, Japan	20.79	-	82	Wolverhampton, UK	-	15.01
39	Nara, Japan	19.28	-	83	Bromley, UK	-	14.77
40	Tokushima, Japan	20.21	-	84	Croydon, UK	-	14.96
41	Ehime, Japan	20.11	-	85	South Lanarkshire, UK	-	13.62
42	Okinawa, Japan	23.64	18.64	86	Suzhou, China	23.82	-
43	Shiga, Japan	19.91	-	87	Guilin, China	25.18	16.68
44	Kumamoto, Japan	20.14	-	88	Liuzhou, China	23.84	16.79

https://doi.org/10.5194/essd-2025-709
Preprint. Discussion started: 28 December 2025

89	Fukuoka, Japan	20.95	-	111	Hechi, China	24.03	18.93
90	Ibaraki, Japan	19.21	13.81	112	Baise, China	22.91	-
91	Nagasaki, Japan	21.41	13.41	113	Wuzhou, China	24.86	-
92	Kagawa, Japan	20.69	-	114	Guigang, China	25.12	-
93	Kochi, Japan	20.02	-	115	Nanning, China	24.76	-
94	Tottori, Japan	19.33	-	116	Qinzhou, China	24.77	-
95	Kagoshima, Japan	21.85	-	117	Fangchengang, China	24.22	19.07
96	Derby, UK	-	14.02	118	Beihai, China	25.96	-
97	Rutland, UK	-	13.99	119	CENTRO, Argentina	22.87	15.02
98	Nottingham, UK	-	14.27	120	LITORAL, Argentina	23.28	17.58
99	Stoke-on-Trent, UK	-	13.65	121	NORESTE, Argentina	22.66	-
100	Medway, UK	-	15.28	122	Queensland, Australia	21.16	-
101	Amber Valley, UK	-	14.09	123	Bulgaria	18.82	13.62
102	Bolsover, UK	-	13.74	124	Cyprus	20.38	15.38
103	Erewash, UK	-	13.98	125	Czechia	-	15.58
104	South Derbyshire, UK	-	14.55	126	Egypt	21.69	-
105	Wealden, UK	-	15.43	127	Estonia	-	14.24
106	Blaby, UK	-	13.74	128	Latvia	-	14.76
107	Charnwood, UK	-	14.16	129	Serbia	-	15.56
108	Harborough, UK	-	14.44	130	Tunisia	22.23	16.68
109	Hinckley and Bosworth, UK	-	14.06	131	Ukraine	-	13.88
110	New Plymouth, New Zealand	-	15.24				

https://doi.org/10.5194/essd-2025-709
Preprint. Discussion started: 28 December 2025

Table A3 Details of 17 energy demand datasets used for external validation.

Location	Year period	Time resolution	Gas / Electricity	Type	Source
Austin, USA	2000-2020	monthly	Electricity	building energy	U.S. EIA, 2020
North eastern Mexico	2022	daily	Gas / Electricity	building energy	Aguirre-Fraire et al., 2024
Trondheim, Norway	2019-2021	hourly	Electricity	building energy	Hofmann et al., 2023
Agder, Norway	2019-2022	hourly	Electricity	building energy	Hofmann et al., 2023
Dhaka, Bangladesh	2018-2019, 2021-2023	hourly	Electricity	building energy	Salehin et al., 2024
Estación Itakyry, Paraguay	2017-2020	hourly	Electricity	energy	Velázquez et al., 2022
Presidente Franco, Paraguay	2017-2020	hourly	Electricity	energy	Velázquez et al., 2022
Acaray, Paraguay	2017-2020	hourly	Electricity	energy	Velázquez et al., 2022
South Australia, Australia	2015	hourly	Electricity	building energy	Sharma et al., 2019
Kitakyushu, Japan	2015-2018	hourly	Gas/Electricity	building energy	Liao et al., 2024
Lower Saxony, Germany	2018-2021	hourly	Electricity	building energy	Schlemminger et al., 2022
Hawke's Bay, New Zealand	2014-2016	minute	Electricity	building energy	B. Anderson et al., 2018
Philadelphia, USA	2009-2011	daily	Electricity	energy	Cohen et al., 2015
Tacoma, USA	2006-2013	daily	Electricity	energy	Cohen et al., 2015
Victoria, Australia	2000-2007	daily	Electricity	energy	AEMO, 2025
Delhi, India	2011-2013	daily	Electricity	energy	Cohen et al., 2015
Abidjan, Côte d'Ivoire	2012-2013	daily	Electricity	energy	Cohen et al., 2015

740

745

750

https://doi.org/10.5194/essd-2025-709
Preprint. Discussion started: 28 December 2025

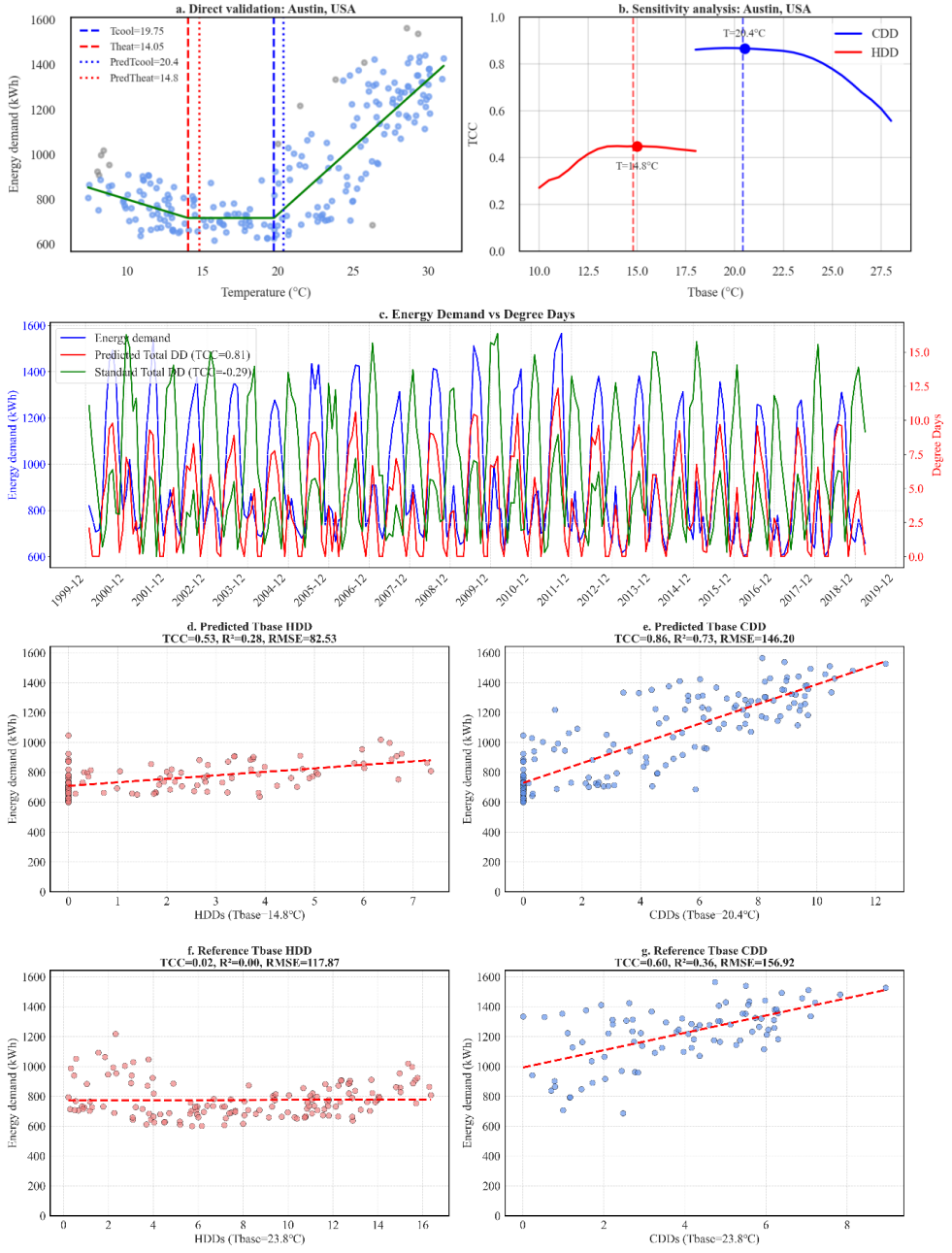


Figure A1: Visualization of HDD and CDD versus energy consumption at different base temperatures in Austin, USA.

https://doi.org/10.5194/essd-2025-709
Preprint. Discussion started: 28 December 2025

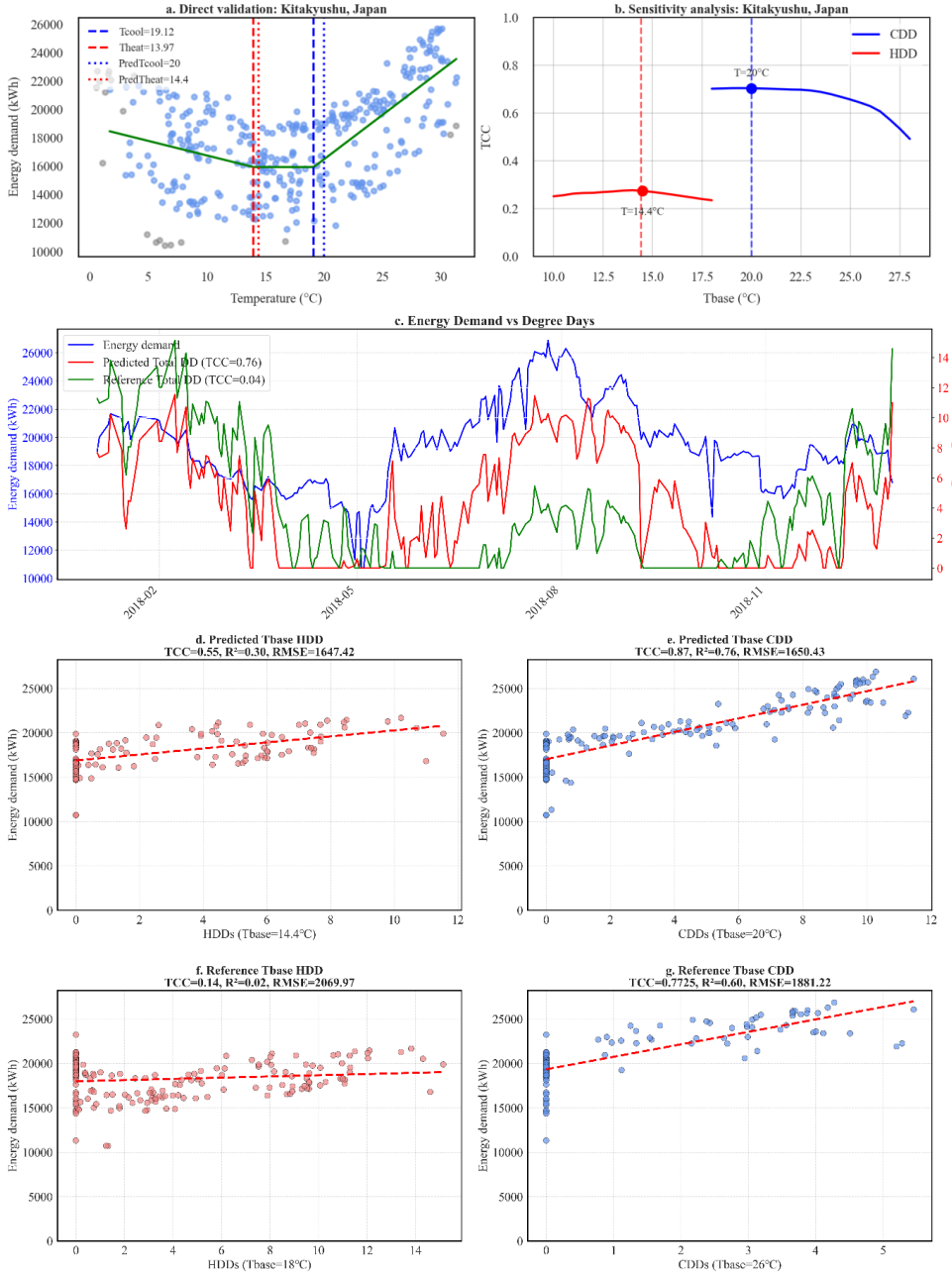


Figure A2: Visualization of HDD and CDD versus energy consumption at different base temperatures in Kitakyushu, Japan.

https://doi.org/10.5194/essd-2025-709
Preprint. Discussion started: 28 December 2025

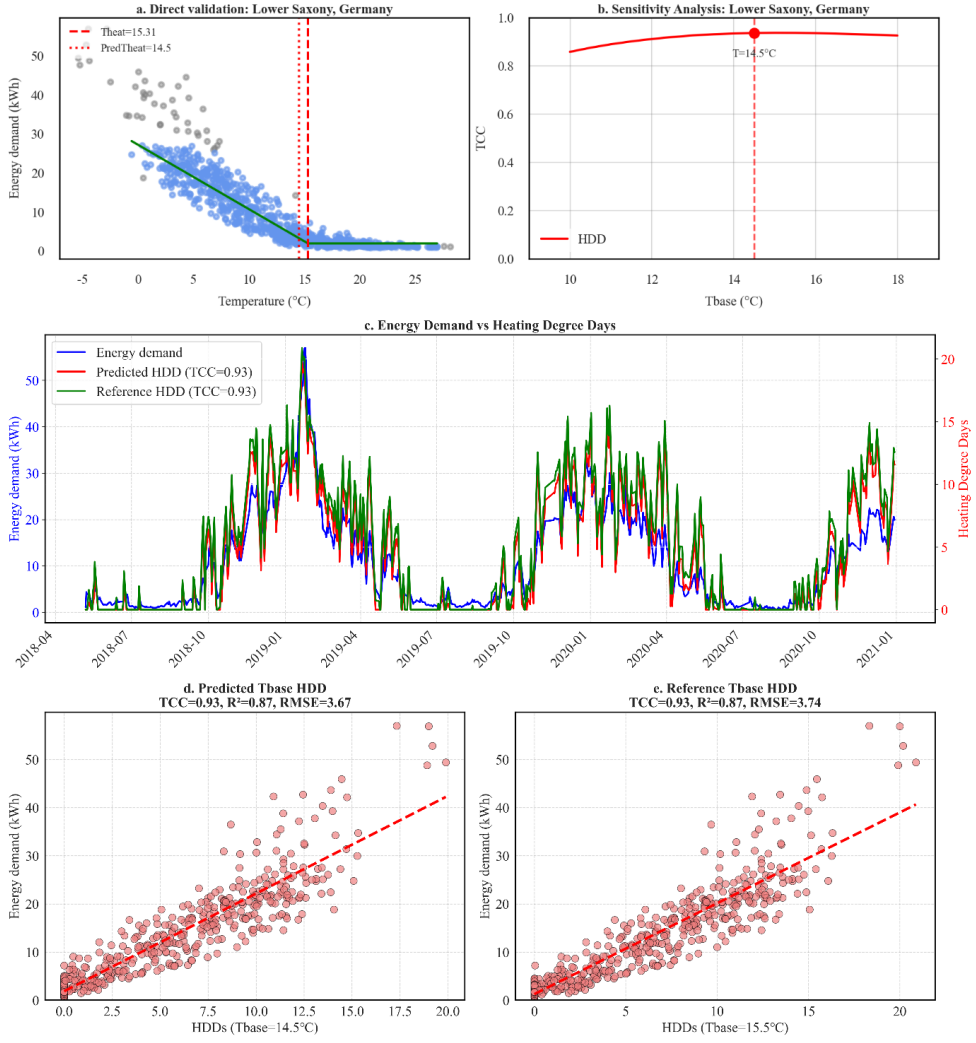


Figure A3: Visualization of HDD versus energy consumption at different base temperatures in Lower Saxony, Germany.

https://doi.org/10.5194/essd-2025-709
Preprint. Discussion started: 28 December 2025

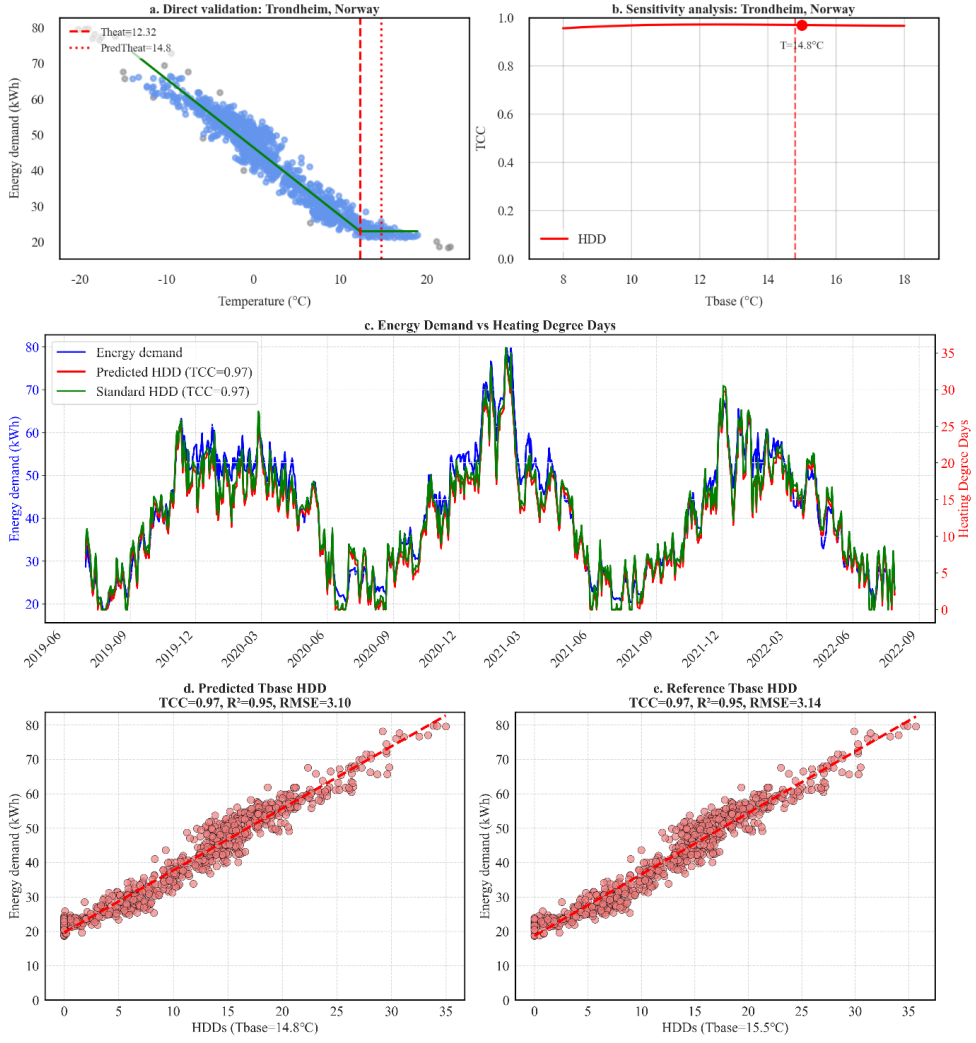


Figure A4: Visualization of HDD and CDD versus energy consumption at different base temperatures in Trondheim, Norway.

https://doi.org/10.5194/essd-2025-709
Preprint. Discussion started: 28 December 2025

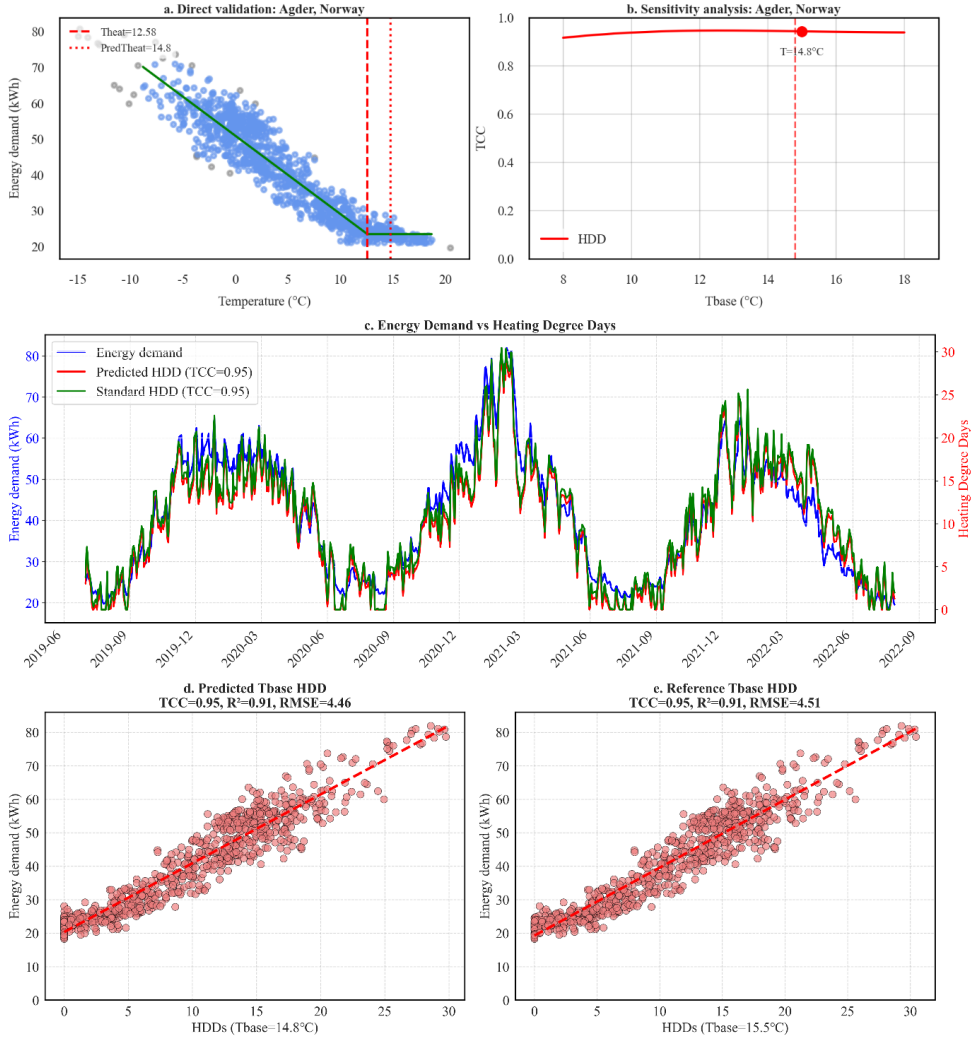


Figure A5: Visualization of HDD versus energy consumption at different base temperatures in Agder, Norway.

https://doi.org/10.5194/essd-2025-709
Preprint. Discussion started: 28 December 2025

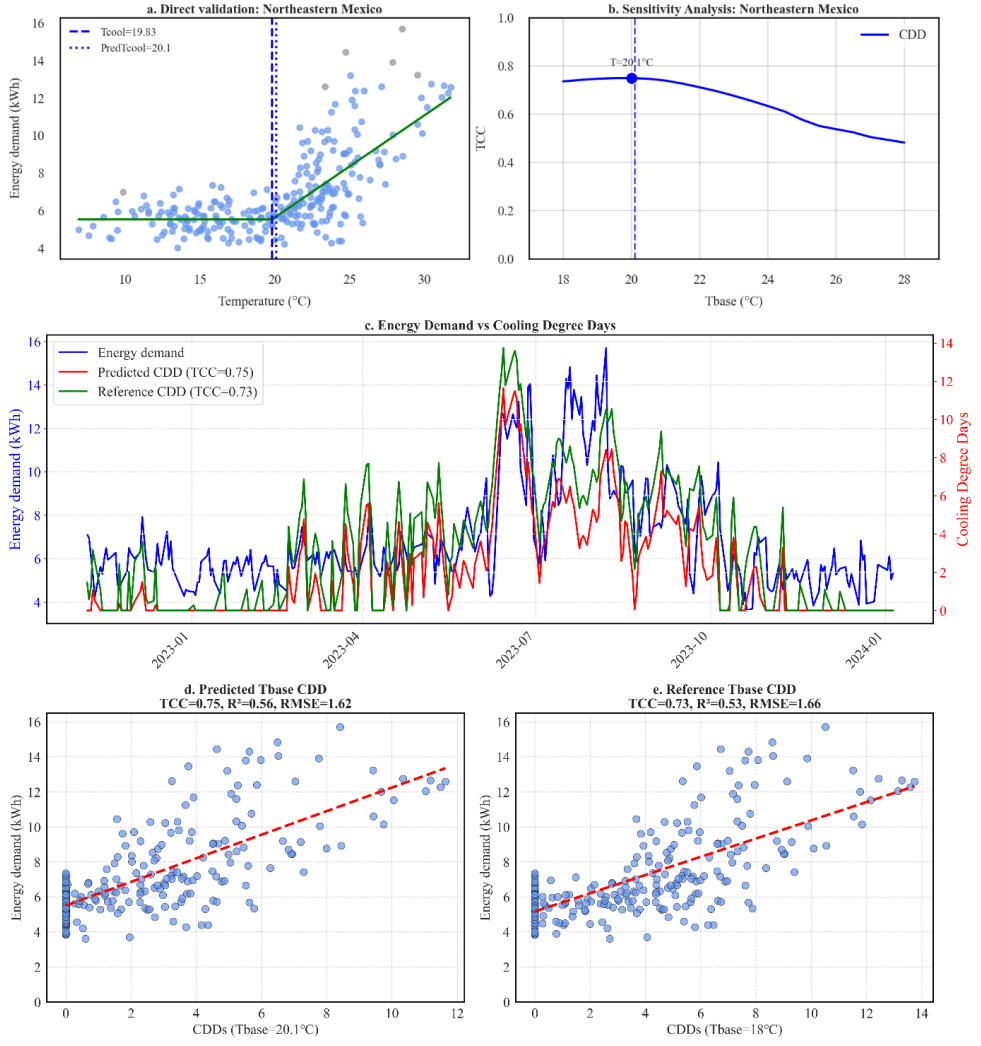


Figure A6: Visualization of HDD and CDD versus energy consumption at different base temperatures in North eastern Mexico.

https://doi.org/10.5194/essd-2025-709
Preprint. Discussion started: 28 December 2025

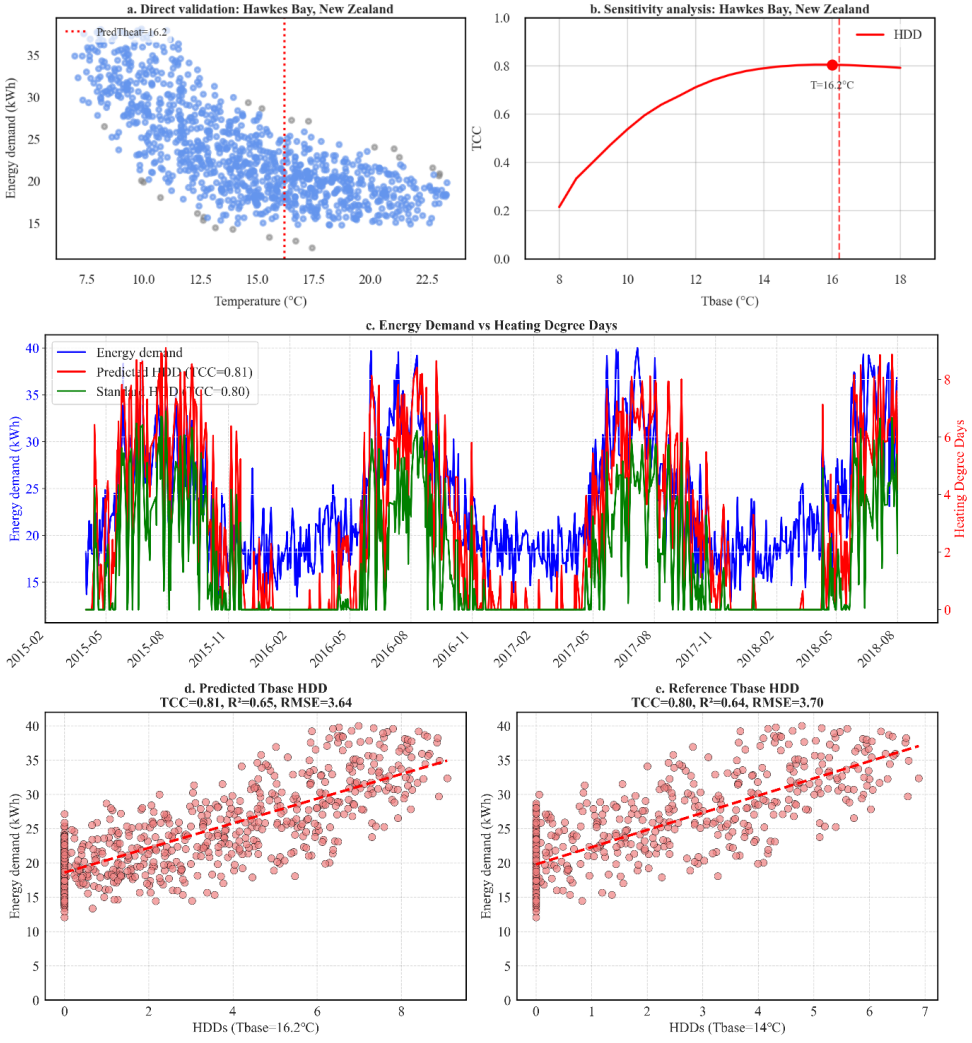


Figure A7: Visualization of HDD versus energy consumption at different base temperatures in Hawkes Bay, New Zealand.

https://doi.org/10.5194/essd-2025-709
Preprint. Discussion started: 28 December 2025

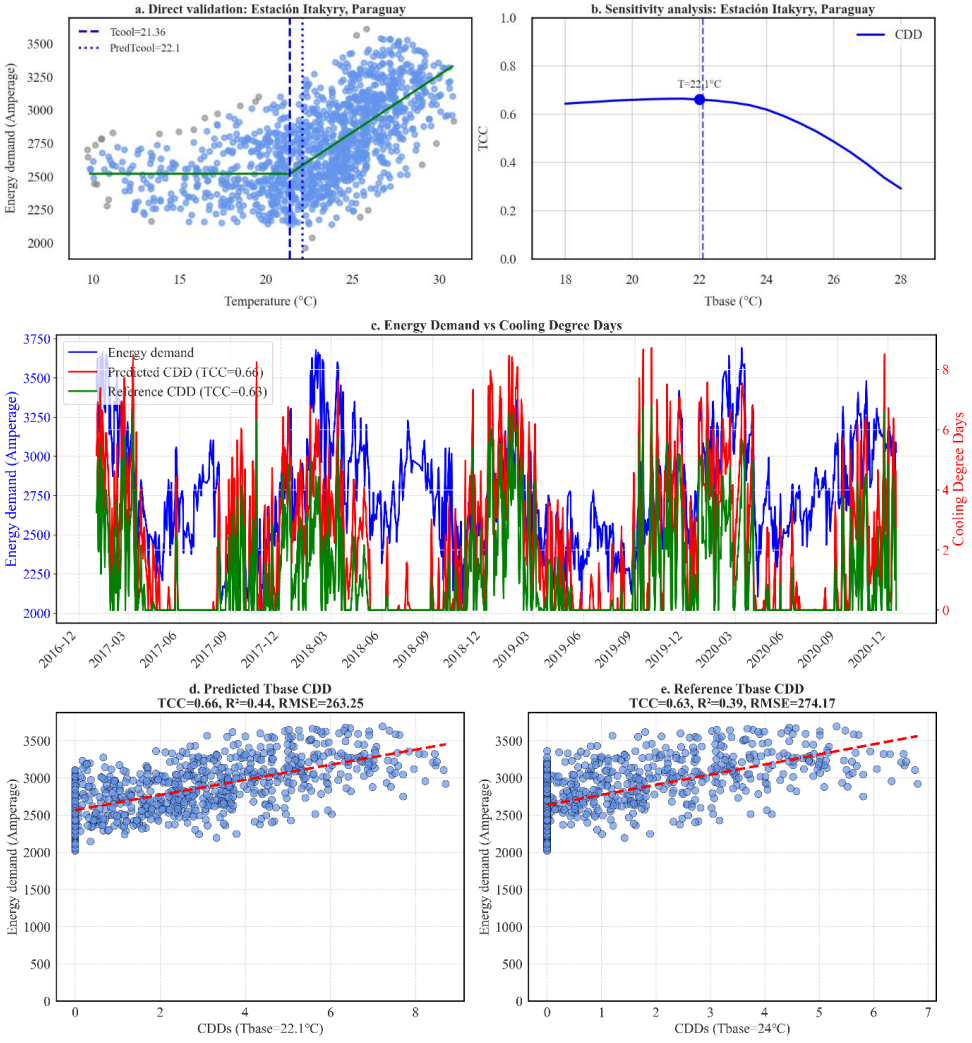


Figure A8: Visualization of HDD versus energy consumption at different base temperatures in Itakyry, Paraguay.

https://doi.org/10.5194/essd-2025-709
Preprint. Discussion started: 28 December 2025

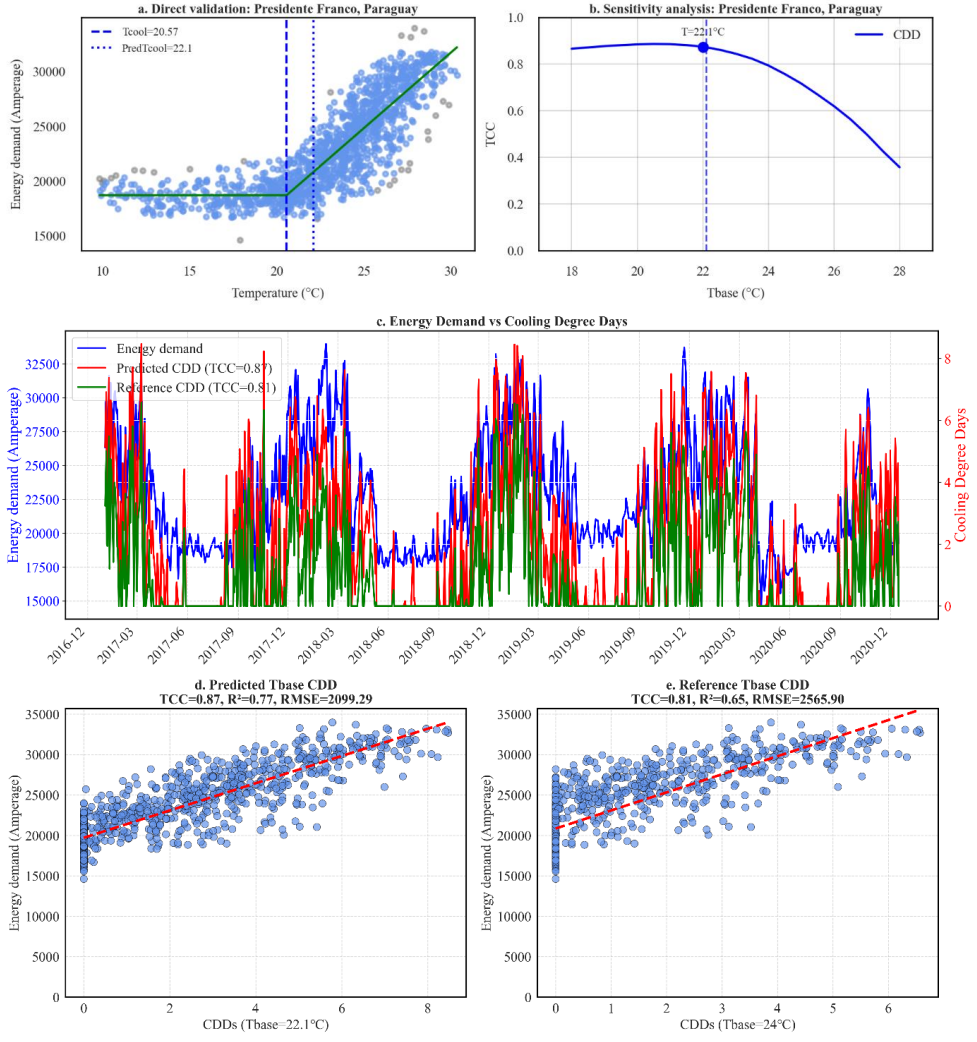


Figure A9: Visualization of CDD versus energy consumption at different base temperatures in Presidente Franco, Paraguay.

https://doi.org/10.5194/essd-2025-709
Preprint. Discussion started: 28 December 2025

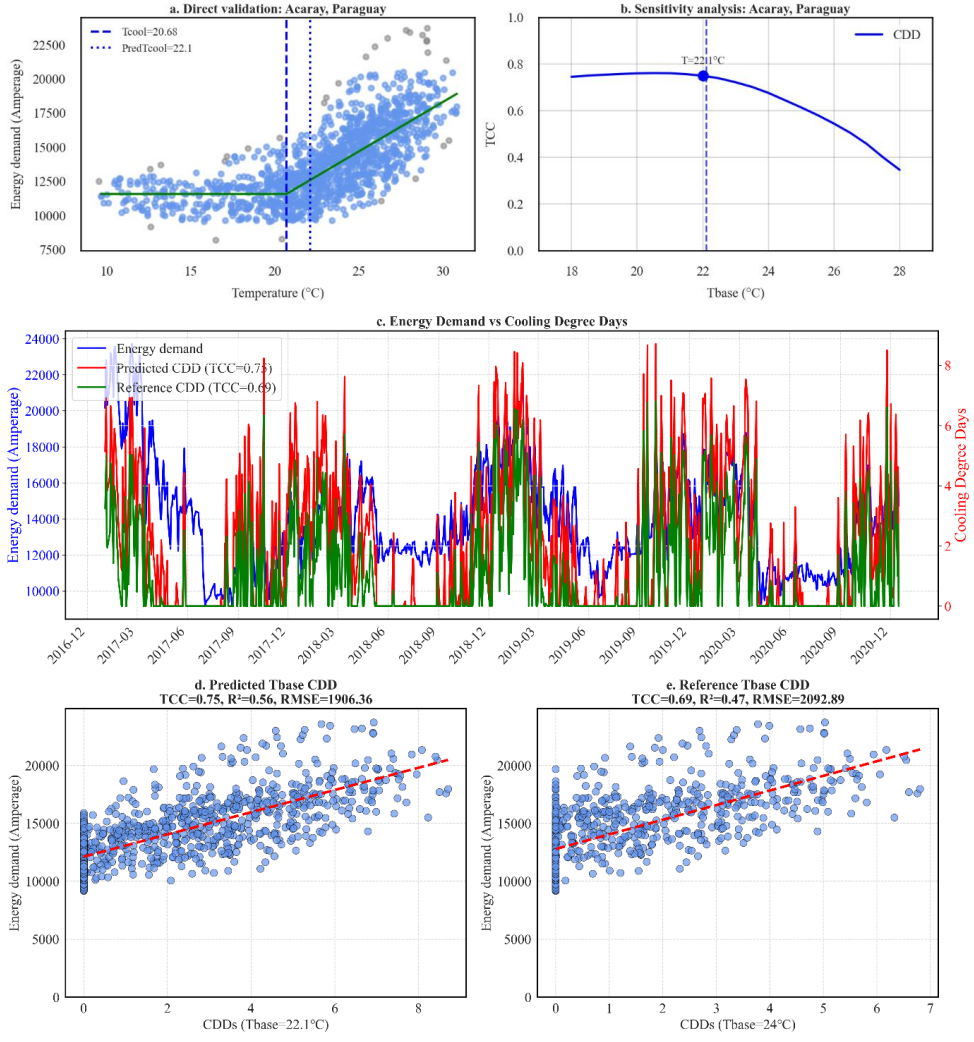


Figure A10: Visualization of CDD versus energy consumption at different base temperatures in Acaray, Paraguay.

https://doi.org/10.5194/essd-2025-709
Preprint. Discussion started: 28 December 2025

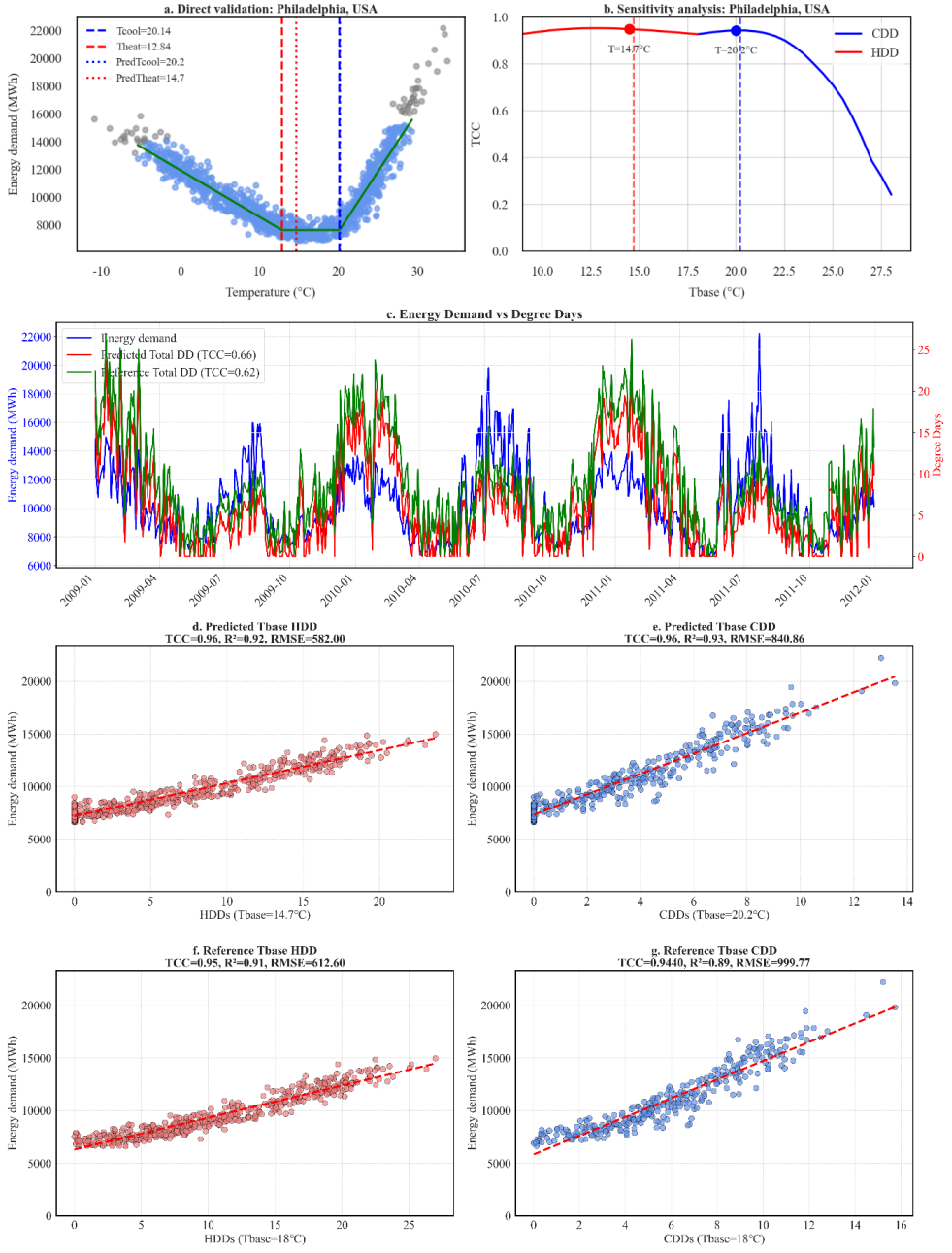


Figure A11: Visualization of HDD and CDD versus energy consumption at different base temperatures in Philadelphia, USA.

https://doi.org/10.5194/essd-2025-709
Preprint. Discussion started: 28 December 2025

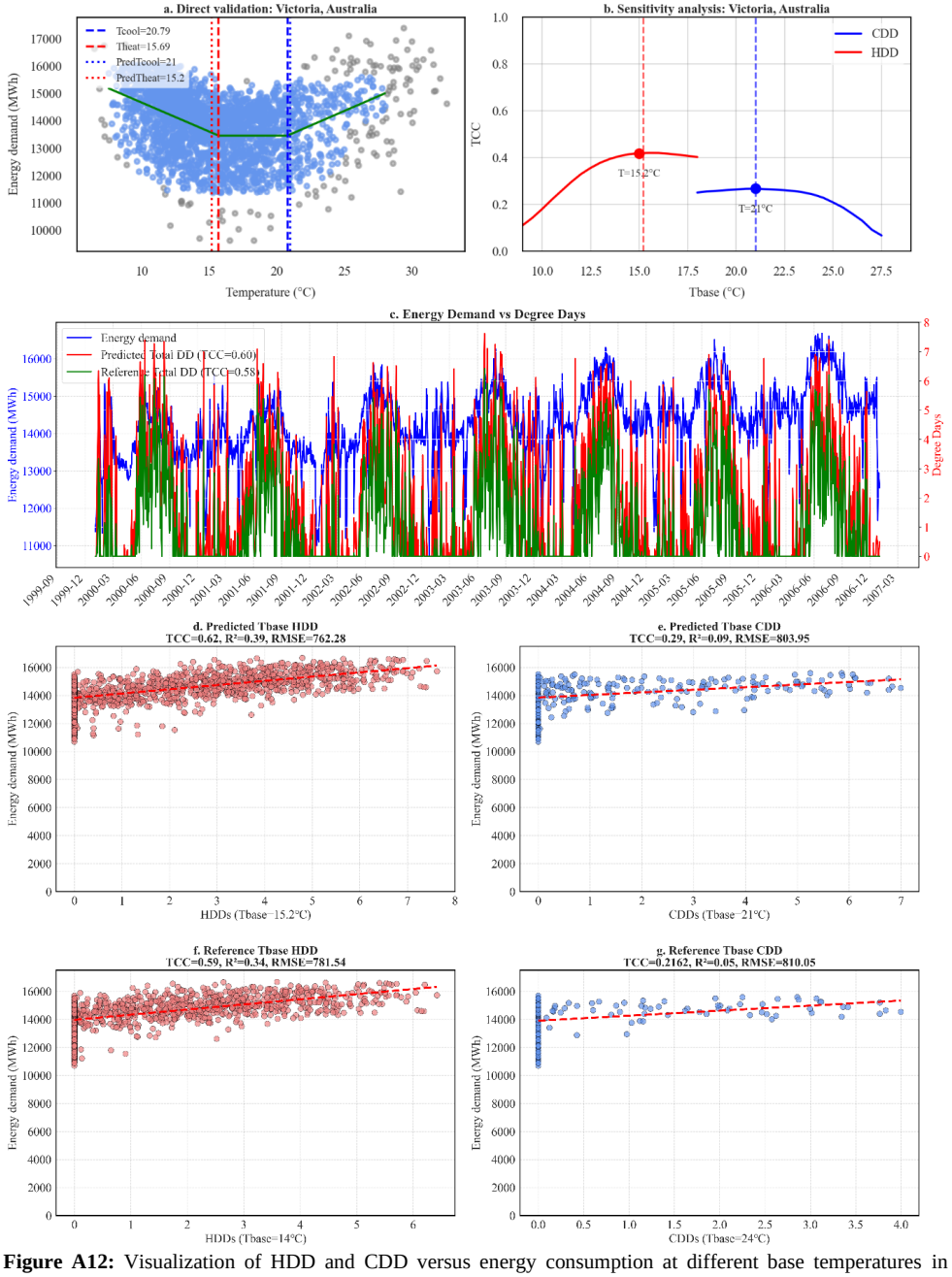


Figure A12: Visualization of HDD and CDD versus energy consumption at different base temperatures in Victoria, Australia.

https://doi.org/10.5194/essd-2025-709
Preprint. Discussion started: 28 December 2025

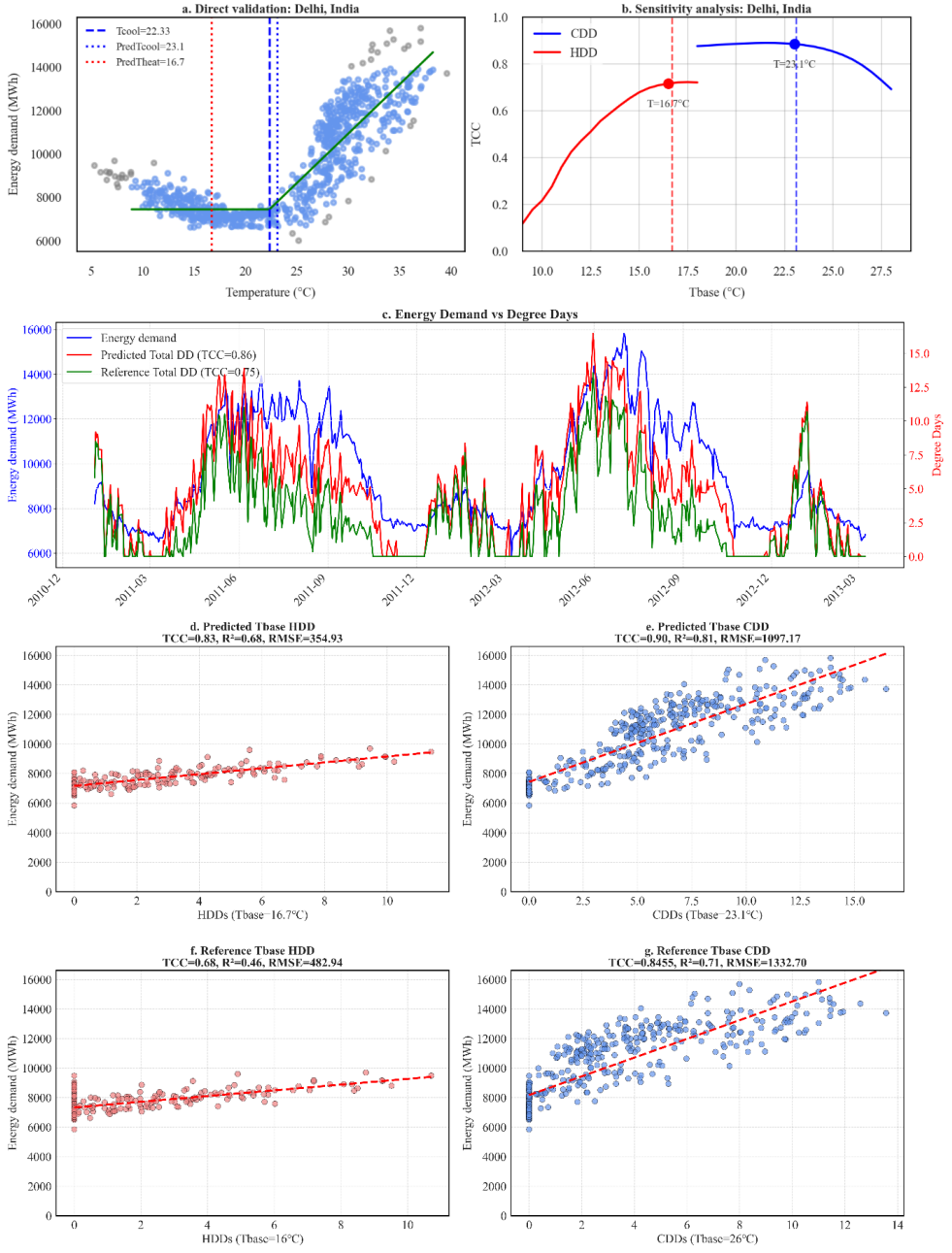


Figure A13: Visualization of HDD and CDD versus energy consumption at different base temperatures in Delhi, India.

https://doi.org/10.5194/essd-2025-709
Preprint. Discussion started: 28 December 2025

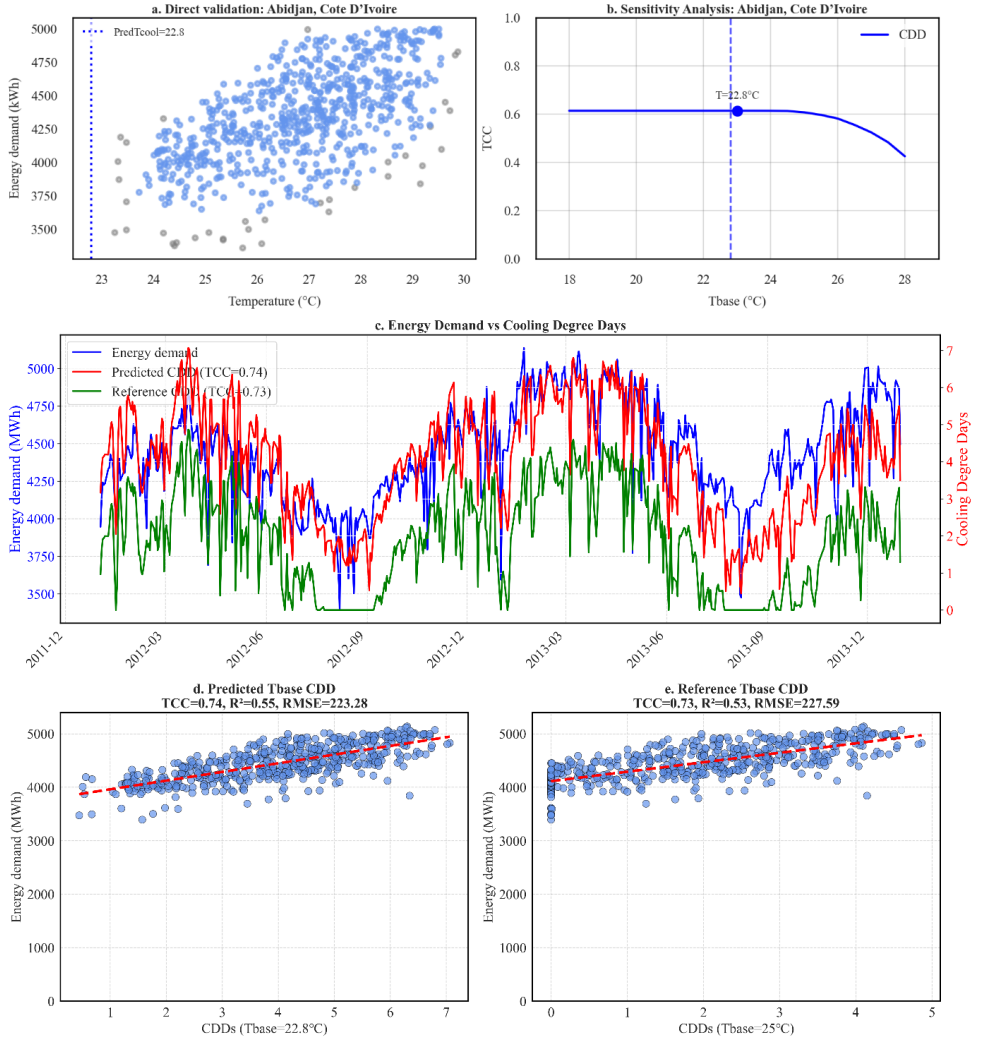


Figure A14: Visualization of CDD versus energy consumption at different base temperatures in Abidjan, Cote D' Ivoire.

https://doi.org/10.5194/essd-2025-709
Preprint. Discussion started: 28 December 2025

Reference

Abajian, A. C., Carleton, T., Meng, K. C., & Deschênes, O. (2025). Quantifying the global climate feedback from energy- based adaptation. Nature Communications, 16(1). https://doi.org/10.1038/s41467-025-59201-7

AECOM Building Engineering, Energy Demand Research Project: Early Smart Meter Trials, 2007-2010, (2014).

Afroz. (2023). Swiss Smart Meter Data [Data set]. Kaggle. https://www.kaggle.com/datasets/pythonafroz/swiss-smart-meter-data

Agency for Natural Resources and Energy. (n.d.). Electric power survey statistics: Historical data [Data set]. Ministry of Economy, Trade and Industry. Retrieved November 12, 2025, from https://www.enecho.meti.go.jp/statistics/electric_power/ep002/results_archive.html

Agrawal, S., Mani, S., Ganesan, K., & Jain, A. (2021). High frequency smart meter data from two districts in India (Mathura and Bareilly) (Version 2) [Data set]. Harvard Dataverse. https://doi.org/10.7910/DVN/GOCHJH

Aguirre-Fraire, B., Beltrán, J., & Soto-Mendoza, V. (2024). A comprehensive dataset integrating household energy consumption and weather conditions in a north-eastern Mexican urban city. Data in Brief, 54, 110452.

Ahmad, T., Chen, H., Guo, Y., & Wang, J. (2018). A comprehensive overview on the data driven and large scale based approaches for forecasting of building energy demand: A review. In Energy and Buildings (Vol. 165, pp. 301–320). Elsevier Ltd. https://doi.org/10.1016/j.enbuild.2018.01.017

Aljawabra, F., & Nikolopoulou, M. (2018). Thermal comfort in urban spaces: a cross-cultural study in the hot arid climate. International Journal of Biometeorology, 62(10), 1901–1909. https://doi.org/10.1007/s00484-018-1592-5

Anjomshoaa, A., & Salmanzadeh, M. (2017). Estimation of the changeover times and degree-days balance point temperatures of a city using energy signatures. Sustainable Cities and Society, 35, 538–543. https://doi.org/10.1016/j.scs.2017.08.028

Australian Energy Market Operator (AEMO). (2025). Electricity price & demand [dataset]. Retrieved from https://aemo.com.au/

B. Anderson, D. Eyers, R. Ford, D. Giraldo Ocampo, R. Peniamina, J. Stephenson, K. Suomalainen, L. Wilcocks, M. Jack, New Zealand GREEN Grid Household Electricity Demand Study 2014-2018, (2018). https://doi.org/http://doi.org/10.5255/UKDA-SN 853334.

Bae, C., Lee, H., & Chun, C. (2025). Beyond the thermostat: A behavioral study of residential cooling practices. Building and Environment, 282. https://doi.org/10.1016/j.buildenv.2025.113250

Bah, E.-H. M., Faye, I., & Geh, Z. F. (n.d.). HOUSING MARKET DYNAMICS IN AFRICA.

Benzaama, M. H., Rajaoarisoa, L. H., Lekhal, M. C., Menhoudj, S., & Mokhtari, A. M. (2021). Thermal inertia and energy efficiency assessment of Direct Solar Floor system using a switching-linear model. Applied Energy, 300. https://doi.org/10.1016/j.apenergy.2021.117363

Berhane, T., Shibabaw, A., Awgichew, G., & Walelgn, A. (2021). Pricing of weather derivatives based on temperature by obtaining market risk factor from historical data. Modeling Earth Systems and Environment, 7(2), 871-884.

Bhatnagar, M., Mathur, J., & Garg, V. (2018). Determining base temperature for heating and cooling degree-days for India. Journal of Building Engineering, 18, 270–280. https://doi.org/10.1016/j.jobe.2018.03.020

https://doi.org/10.5194/essd-2025-709
Preprint. Discussion started: 28 December 2025

Borah, P., Singh, M. K., & Mahapatra, S. (2015). Estimation of degree-days for different climatic zones of North-East India. Sustainable Cities and Society, 14(1), 70–81. https://doi.org/10.1016/j.scs.2014.08.001

Burguillo, M., & Juez-Martel, P. (2024). Just energy heating transitions: lessons from characteristics of households using different heating sources. Energy Efficiency, 17(6). https://doi.org/10.1007/s12053-024-10241-w

C. Miller, A. Kathirgamanathan, B. Picchetti, P. Arjunan, J.Y. Park, Z. Nagy, P. Raftery, B.W. Hobson, Z. Shi, F. Meggers, The Building Data Genome Project 2, energy meter data from the ASHRAE Great Energy Predictor III competition, (2020).

Camarasa, C., Mata, É., Navarro, J. P. J., Reyna, J., Bezerra, P., Angelkorte, G. B., Feng, W., Filippidou, F., Forthuber, S., Harris, C., Sandberg, N. H., Ignatiadou, S., Kranzl, L., Langevin, J., Liu, X., Müller, A., Soria, R., Villamar, D., Dias, G. P., … Yaramenka, K. (2022). A global comparison of building decarbonization scenarios by 2050 towards 1.5–2 °C targets. Nature Communications, 13(1). https://doi.org/10.1038/s41467-022-29890-5

CAMMESA. (2025). Base Demanda Diaria 2017-2025 [Dataset]. Compañía Administradora del Mercado Mayorista Eléctrico. https://cammesaweb.cammesa.com/2025/11/07/covid-19-comportamiento-de-la-demanda-de-energia-electrica-en-el-mem

Canadian Centre for Energy Information. (n.d.). High-Frequency Electricity Data: Visualization Tool (Beta) [Data set]. Canadian Centre for Energy Information, in collaboration with Natural Resources Canada and the Canada Energy Regulator. https://energy-information.canada.ca/en/resources/high-frequency-electricity-data

Central Electricity Authority. (2024). Power supply position [Data set]. Open Government Data Platform India. Retrieved November 12, 2025, from https://www.data.gov.in/catalog/power-supply-position

Chambers, J. D., & Oreszczyn, T. (2019). Deconstruct: A scalable method of as-built heat power loss coefficient inference for UK dwellings using smart meter data. Energy and Buildings, 183, 443–453. https://doi.org/10.1016/j.enbuild.2018.11.016

Chen, J., Chen, J., Zhang, X. J., Peng, P., & Risi, C. (2023). A century and a half precipitation oxygen isoscape for China generated using data fusion and bias correction. Scientific Data, 10(1). https://doi.org/10.1038/s41597-023-02095-1

Choi, S., Pellen, A., & Masson, V. (2017). How does daylight saving time affect electricity demand? An answer using aggregate data from a natural experiment in Western Australia. Energy Economics, 66, 247–260. https://doi.org/10.1016/j.eneco.2017.06.018

Cholewa, T., Siuta-Olcha, A., Smolarz, A., Muryjas, P., Wolszczak, P., Anasiewicz, R., & Balaras, C. A. (2021). A simple building energy model in form of an equivalent outdoor temperature. Energy and Buildings, 236. https://doi.org/10.1016/j.enbuild.2021.110766

Clarke, L., Eom, J., Marten, E. H., Horowitz, R., Kyle, P., Link, R., ... & Zhou, Y. (2018). Effects of long-term climate change on global building energy expenditures. Energy Economics, 72, 667-677.

Cohen, E., Modi, V., Torbey, H., Piccirelli, M. & Yu-Tian. Global trends in urban energy use. GitHub Repository (2015). https://github.com/Ecohen4/Energy

Cong, S., Nock, D., Qiu, Y. L., & Xing, B. (2022). Unveiling hidden energy poverty using the energy equity gap. Nature Communications, 13(1). https://doi.org/10.1038/s41467-022-30146-5

https://doi.org/10.5194/essd-2025-709
Preprint. Discussion started: 28 December 2025

Corrales-Suastegui, A., Ruiz-Alvarez, O., Torres-Alavez, J. A., & Pavia, E. G. (2021). Analysis of cooling and heating degree days over mexico in present and future climate. Atmosphere, 12(9). https://doi.org/10.3390/atmos12091131

Correia, E., & Delgado Henriques, C. (2025). Indoor thermal comfort in informal settlements: a case study of Mafalala, Maputo. African Geographical Review. https://doi.org/10.1080/19376812.2025.2520775

Day, A. R., Knight, I., Dunn, G., & Gaddas, R. (2003). Improved methods for evaluating base temperature for use in building energy performance lines. Building Services Engineering Research and Technology, 24(4), 221-228.

De Rosa, M., Bianco, V., Scarpa, F., & Tagliafico, L. A. (2015). Historical trends and current state of heating and cooling degree days in Italy. Energy Conversion and Management, 90, 323–335. https://doi.org/10.1016/j.enconman.2014.11.022

Deroubaix, A., Labuhn, I., Camredon, M., Gaubert, B., Monerie, P. A., Popp, M., Ramarohetra, J., Ruprich-Robert, Y., Silvers, L. G., & Siour, G. (2021). Large uncertainties in trends of energy demand for heating and cooling under climate change. Nature Communications, 12(1). https://doi.org/10.1038/s41467-021-25504-8

Dicko, K., Umaru, E. T., Sanogo, S., Okhimamhe, A. A., & Loewner, R. (2024). A Trend Analysis of Changes in Cooling Degree Days in West Africa Under Global Warming. Atmosphere, 15(11). https://doi.org/10.3390/atmos15111376

Djongyang, N., & Tchinda, R. (2010). An investigation into thermal comfort and residential thermal environment in an intertropical sub-Saharan Africa region: Field study report during the Harmattan season in Cameroon. Energy Conversion and Management, 51(7), 1391–1397. https://doi.org/10.1016/j.enconman.2010.01.011

Duan, H., Ming, X., Zhang, X. B., Sterner, T., & Wang, S. (2023). China's adaptive response to climate change through air-conditioning. IScience, 26(3). https://doi.org/10.1016/j.isci.2023.106178

Ember. (n.d.). Monthly electricity data [Data set]. Retrieved November 12, 2025, from https://ember-energy.org/data/monthly-electricity-data/

Empresa de Pesquisa Energética (EPE). Monthly Electricity Consumption by Class (regions and subsystems). Rio de Janeiro: EPE, [2024]. https://www.epe.gov.br/en/open-data/publications/monthly-electricity-consumption

Energy Agency, I. (2024). World Energy Outlook 2024. www.iea.org/terms

Falchetta, G., Cian, E. De, Pavanello, F., & Wing, I. S. (2024). Inequalities in global residential cooling energy use to 2050. Nature Communications, 15(1). https://doi.org/10.1038/s41467-024-52028-8

Feng, Y., Duan, Q., Chen, X., Yakkali, S. S., & Wang, J. (2021). Space cooling energy usage prediction based on utility data for residential buildings using machine learning methods. Applied Energy, 291. https://doi.org/10.1016/j.apenergy.2021.116814

Filahi, H., Drobinski, P., Oueslati, B., Claudel, S., & Laurent, M. H. (2025). Rising power system strains despite decreasing heating demand in a warming climate. Scientific Reports, 15(1). https://doi.org/10.1038/s41598-025-96174-5

Filahi, H., Omrani, H., Claudel, S., & Drobinski, P. (2024). Temporal fragmentation of the energy demand in Europe: Impact of climate change on the maneuverability of energy system. Climate Services, 34. https://doi.org/10.1016/j.cliser.2024.100469

https://doi.org/10.5194/essd-2025-709
Preprint. Discussion started: 28 December 2025

Fung, W. Y., Lam, K. S., Hung, W. T., Pang, S. W., & Lee, Y. L. (2006). Impact of urban temperature on energy consumption of Hong Kong. In Energy (Vol. 31, Issue 14, pp. 2623–2637). Elsevier Ltd. https://doi.org/10.1016/j.energy.2005.12.009

Gupta, V., Sarkar, A., Saxena, A., Jana, A., & Bandyopadhyay, S. (2024). Appliance utility in low-income housing: On the relation between socio-economic characteristics, thermal comfort and appliance usage pattern. Energy and Buildings, 323. https://doi.org/10.1016/j.enbuild.2024.114849

Hao, Z., Zhang, X., Xie, J., Yin, K., & Liu, J. (2022). Balance point temperature and heating degree-days in different climate conditions for building energy efficiency applications. Building and Environment, 216. https://doi.org/10.1016/j.buildenv.2022.109013

He, J., Zhang, S., Yu, M., Liang, Q., Cao, M., Xu, H., ... & Liu, J. (2025). Predicting indoor PM2. 5 levels in shared office using LSTM method. Journal of Building Engineering, 104, 112407.

He, X., Eom, J., Yu, S., Liu, S., Xu, W., Zhou, Y. (2025). A global base temperature dataset for building energy demand modeling. figshare. Dataset. https://doi.org/10.6084/m9.figshare.30646376.v2

Hirth, L., Khanna, T. M., & Ruhnau, O. (2024). How aggregate electricity demand responds to hourly wholesale price fluctuations. Energy Economics, 135. https://doi.org/10.1016/j.eneco.2024.107652

Hochreiter, S., & Urgen Schmidhuber, J. ¨. (n.d.). Long Short-Term Memory. http://direct.mit.edu/neco/article-pdf/9/8/1735/813796/neco.1997.9.8.1735.pdf

Hofmann, M., Bjarghov, S., & Nessa, S. (2023). Norwegian hourly residential electricity demand data with consumer characteristics during the European energy crisis. Data in Brief, 51, 109687.

Huang, J., & Gurney, K. R. (2016). The variation of climate change impact on building energy consumption to building type and spatiotemporal scale. Energy, 111, 137–153. https://doi.org/10.1016/j.energy.2016.05.118

Huang, L., & Kang, J. (2021). Thermal comfort in winter incorporating solar radiation effects at high altitudes and performance of improved passive solar design—Case of Lhasa. Building Simulation, 14(6), 1633–1650. https://doi.org/10.1007/s12273-020-0743-x

Ihara, T., Genchi, Y., Sato, T., Yamaguchi, K., & Endo, Y. (2008). City-block-scale sensitivity of electricity consumption to air temperature and air humidity in business districts of Tokyo, Japan. Energy, 33(11), 1634–1645. https://doi.org/10.1016/j.energy.2008.06.005

International Energy Agency (IEA). (2024). World energy outlook 2024. https://www.iea.org/reports/world-energy-outlook-2024

Islam, A. R. M. T., Ahmed, I., & Rahman, M. S. (2020). Trends in cooling and heating degree-days overtimes in Bangladesh? An investigation of the possible causes of changes. Natural Hazards, 101(3), 879–909. https://doi.org/10.1007/s11069-020-03900-5

Kennard, H., Oreszczyn, T., Mistry, M., & Hamilton, I. (2022). Population-weighted degree-days: The global shift between heating and cooling. Energy and Buildings, 271. https://doi.org/10.1016/j.enbuild.2022.112315

Kikstra, J. S., Mastrucci, A., Min, J., Riahi, K., & Rao, N. D. (2021). Decent living gaps and energy needs around the world. Environmental Research Letters, 16(9). https://doi.org/10.1088/1748-9326/ac1c27

Kim, H. G., Kim, S. S., & Ahn, H. (2024a). Data-driven analysis of heating and cooling base temperatures for buildings: Case studies in South Korea. Energy and Buildings, 322, 114709.

https://doi.org/10.5194/essd-2025-709
Preprint. Discussion started: 28 December 2025

Kim, S., Nathaniel, J., Hou, Z., Zheng, T., & Gentine, P. (2024b). Spatiotemporal upscaling of sparse air-sea pCO2 data via physics-informed transfer learning. Scientific Data , 11(1). https://doi.org/10.1038/s41597-024-03959-w

Kozarcanin, S., Andresen, G. B., & Staffell, I. (2019). Estimating country-specific space heating threshold temperatures from national gas and electricity consumption data. Energy and Buildings, 199, 368–380. https://doi.org/10.1016/j.enbuild.2019.07.013

Krause, J. S., Brandt, G., Schmidt, U., & Schunk, D. (2023). Don't sweat it: Ambient temperature does not affect social behavior and perception. Journal of Economic Psychology, 99, 102657.

Kummu, M., Kosonen, M., & Masoumzadeh Sayyar, S. (2025). Downscaled gridded global dataset for gross domestic product (GDP) per capita PPP over 1990–2022. Scientific Data, 12(1), 178.

Lebakula, V., Epting, J., Moehl, J., Stipek, C., Adams, D., Reith, A., Kaufman, J., Gonzales, J., Reynolds, B., Basford, S., Martin, A., Buck, W., Faxon, A., Cunningham, A., Roy, A., Barbose, Z., Massaro, J., Walters, S., Woody, C., … Urban, M. (2024). LandScan Silver Edition [Data set]. Oak Ridge National Laboratory. https://doi.org/10.48690/1531770

Li, B., Yu, R., Gan, K., Ruan, G., Liu, S., Yang, M., ... & Zhong, H. (2025). A Large-Scale Residential Load Dataset in a Southern Province of China. Scientific Data, 12(1), 1-14.

Li, J., Yang, L., & Long, H. (2018). Climatic impacts on energy consumption: Intensive and extensive margins. Energy Economics, 71, 332–343. https://doi.org/10.1016/j.eneco.2018.03.010

Li, X. X. (2018). Linking residential electricity consumption and outdoor climate in a tropical city. Energy, 157, 734–743. https://doi.org/10.1016/j.energy.2018.05.192

Liao, W., Jin, X., Ran, Y., Xiao, F., Gao, W., & Li, Y. (2024). A twenty-year dataset of hourly energy generation and consumption from district campus building energy systems. Scientific data, 11(1), 1400.

Livada, I., Pyrgou, A., Haddad, S., Sadeghi, M., & Santamouris, M. (2021). Recent climatic trends and analysis of monthly heating and cooling degree hours in Sydney. Climate, 9(7). https://doi.org/10.3390/cli9070114

Mahdavinejad, M., Bazazzadeh, H., Mehrvarz, F., Berardi, U., Nasr, T., Pourbagher, S., & Hoseinzadeh, S. (2024). The impact of facade geometry on visual comfort and energy consumption in an office building in different climates. Energy Reports, 11, 1–17. https://doi.org/10.1016/j.egyr.2023.11.021

Makonin, S. (2019). HUE: The hourly usage of energy dataset for buildings in British Columbia. Data in brief, 23, 103744.

Mansouri, A., Abolmasoumi, A. H., & Ghadimi, A. A. (2023). Weather sensitive short term load forecasting using dynamic mode decomposition with control. Electric Power Systems Research, 221. https://doi.org/10.1016/j.epsr.2023.109387

Martinopoulos, G., Alexandru, A., & Papakostas, K. T. (2019). Mapping temperature variation and degree-days in metropolitan areas with publicly available sensors. Urban Climate, 28. https://doi.org/10.1016/j.uclim.2019.100464

Meng, Q., Mourshed, M., & Wei, S. (2018). Going beyond the Mean: Distributional Degree-Day Base Temperatures for Building Energy Analytics Using Change Point Quantile Regression. IEEE Access, 6, 39532–39540. https://doi.org/10.1109/ACCESS.2018.2852478

https://doi.org/10.5194/essd-2025-709
Preprint. Discussion started: 28 December 2025

Mishra, A., Lone, H. R., & Mishra, A. (2024). DECODE: Data-driven energy consumption prediction leveraging historical data and environmental factors in buildings. Energy and Buildings, 307. https://doi.org/10.1016/j.enbuild.2024.113950

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., ... & Thépaut, J. N. (2021). ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. Earth system science data, 13(9), 4349-4383.

Nadeem, A. & Arshad, N. Precon: Pakistan residential electricity consumption dataset. 52–57, https://doi-org.eproxy.lib.hku.hk/10.1145/3307772.3328317 (2019).

Open Power System Data. Data package household data. https://data.open-power-system-data.org/household_data/2020-04-15/ (2020).

Palazzoli, I., Ceola, S., & Gentine, P. (2025). GRAiCE: reconstructing terrestrial water storage anomalies with recurrent neural networks. Scientific Data , 12(1). https://doi.org/10.1038/s41597-025-04403-3

Pawlak, J., Faghih Imani, A., & Sivakumar, A. (2021). How do household activities drive electricity demand? Applying activity-based modelling in the context of the United Kingdom. Energy Research and Social Science, 82. https://doi.org/10.1016/j.erss.2021.102318

Perera, A. T. D., Javanroodi, K., Mauree, D., Nik, V. M., Florio, P., Hong, T., & Chen, D. (2023). Challenges resulting from urban density and climate change for the EU energy transition. Nature Energy, 8(4), 397–412. https://doi.org/10.1038/s41560-023-01232-9

Petri, Y., & Caldeira, K. (2015). Impacts of global warming on residential heating and cooling degree-days in the United States. Scientific reports, 5(1), 12427.

Pohl, B., Dos Santos, S., Bai, G. M., Compaoré, Y., Dianou, K., Diallo-Dudek, J., Soura, A., & Janicot, S. (2021). Indoor temperature variability in the Sahel: a pilot study in Ouagadougou, Burkina Faso. Theoretical and Applied Climatology, 146(3–4), 1403–1420. https://doi.org/10.1007/s00704-021-03800-z

Prataviera, E., Romano, P., Carnieletto, L., Pirotti, F., Vivian, J., & Zarrella, A. (2021). EUReCA: An open-source urban building energy modelling tool for the efficient evaluation of cities energy demand. Renewable Energy, 173, 544–560. https://doi.org/10.1016/j.renene.2021.03.144

Ramon, D., Allacker, K., De Troyer, F., Wouters, H., & van Lipzig, N. P. M. (2020). Future heating and cooling degree days for Belgium under a high-end climate change scenario. Energy and Buildings, 216. https://doi.org/10.1016/j.enbuild.2020.109935

Rastogi, K. (2024). SEWA Energy Demand Forecasting [Data set]. Kaggle. https://www.kaggle.com/datasets/kaustubhrastogi17/sewa-energy-demand-forecasting

Rode, A., Carleton, T., Delgado, M., Greenstone, M., Houser, T., Hsiang, S., Hultgren, A., Jina, A., Kopp, R. E., McCusker, K. E., Nath, I., Rising, J., & Yuan, J. (2021). Estimating a social cost of carbon for global energy consumption. Nature, 598(7880), 308–314. https://doi.org/10.1038/s41586-021-03883-8

RTE. (n.d.). Regional electricity data [Data set]. Retrieved November 12, 2025, from https://analysesetdonnees.rte-france.com/en/regions/overview-region

Salam, A. R., & El Hibaoui, A. (2018). Comparison of machine learning algorithms for the power consumption prediction: Case study of Tetouan city. In 2018 6th International Renewable and Sustainable Energy Conference (IRSEC) (pp. 1–5). IEEE.

https://doi.org/10.5194/essd-2025-709
Preprint. Discussion started: 28 December 2025

Salehin, I., Noman, S. M., & Hasan, M. M. (2024). Electricity energy dataset "BanE-16": Analysis of peak energy demand with environmental variables for machine learning forecasting. Data in Brief, 52, 109967.

Santos, M. M., Ferreira, A. V., & Lanzinha, J. C. G. (2022). Passive Solar Systems for the Promotion of Thermal Comfort in African Countries: A Review. In Energies (Vol. 15, Issue 23). MDPI. https://doi.org/10.3390/en15239167

Schlemminger, M., Ohrdes, T., Schneider, E., & Knoop, M. (2022). Dataset on electrical single-family house and heat pump load profiles in Germany. Scientific data, 9(1), 56.

Scoccimarro, E., Cattaneo, O., Gualdi, S., Mattion, F., Bizeul, A., Risquez, A. M., & Quadrelli, R. (2023). Country-level energy demand for cooling has increased over the past two decades. Communications Earth and Environment, 4(1). https://doi.org/10.1038/s43247-023-00878-3

Sharma, V., Haque, M. H., & Aziz, S. M. (2019). PV generation and load profile data of net zero energy homes in South Australia. Data in brief, 25, 104235.

Shi, H., Wang, B., Deng, N., Xu, S., & Wang, Z. (2025). Examining energy poverty under high temperatures in China. Renewable and Sustainable Energy Reviews, 215. https://doi.org/10.1016/j.rser.2025.115575

Srinuti. (2020). Residential power usage 3years data - Timeseries [Data set]. Kaggle. https://www.kaggle.com/datasets/srinuti/residential-power-usage-3years-data-timeseries

Staffell, I., Pfenninger, S., & Johnson, N. (2023). A global model of hourly space heating and cooling demand at multiple spatial scales. Nature Energy, 8(12), 1328–1344. https://doi.org/10.1038/s41560-023-01341-5

UK Climate Projections Projections. (2014). How have cooling degree days (CDD) and heating degree days (HDD) been calculated in UKCP09? (http://ukclimateprojections.metoffice.gov.uk/22715)

United Nations (UN). (2019). Environment Programme & Global Alliance for Buildings and Construction: 2019 Global Status Report for Buildings and Construction.

United Nations Human Settlements Programme (UN-Habitat). (2020). World cities report 2020: The value of sustainable urbanization.

U.S. Energy Information Administration. (2020). Residential Average Monthly kWh and Bills [Data set]. Data.gov. https://catalog.data.gov/dataset/residential-average-monthly-kwh-and-bills

van Ruijven, B. J., De Cian, E., & Sue Wing, I. (2019). Amplification of future energy demand growth due to climate change. Nature Communications, 10(1). https://doi.org/10.1038/s41467-019-10399-3

Velázquez, G., Morales, F., García-Torres, M., Gómez-Vela, F., Divina, F., Noguera, J. L. V., ... & Román, J. C. M. (2022). Distribution level electric current consumption and meteorological data set of the east region of Paraguay. Data in Brief, 40, 107699.

Vivian, J., Chiodarelli, U., Emmi, G., & Zarrella, A. (2020). A sensitivity analysis on the heating and cooling energy flexibility of residential buildings. Sustainable Cities and Society, 52. https://doi.org/10.1016/j.scs.2019.101815

Wang, C., Song, J., Shi, D., Reyna, J. L., Horsey, H., Feron, S., Zhou, Y., Ouyang, Z., Li, Y., & Jackson, R. B. (2023). Impacts of climate change, population growth, and power sector decarbonization on urban building energy use. Nature Communications, 14(1). https://doi.org/10.1038/s41467-023-41458-5

Wenz, L., Levermann, A., & Auffhammer, M. (2017). North–south polarization of european electricity consumption under future warming. Proceedings of the National Academy of Sciences, 114(38), 201704339.

https://doi.org/10.5194/essd-2025-709
Preprint. Discussion started: 28 December 2025

Woods, J., & Fuller, C. (2014). Estimating base temperatures in econometric models that include degree days. Energy Economics, 45, 166–171. https://doi.org/10.1016/j.eneco.2014.06.006

Xu, X., Hu, Y., Atamturktur, S., Chen, L., & Wang, J. (2025). Systematic review on uncertainty quantification in machine learning-based building energy modeling. Renewable and Sustainable Energy Reviews, 218, 115817.

Yang, S., Jahanger, A., & Awan, A. (2024). Temperature variation and urban electricity consumption in China: Implications for demand management and planning. Utilities Policy, 90. https://doi.org/10.1016/j.jup.2024.101782

Yilmaz, S., & Canan, F. (2025). Establishing the thresholds for outdoor thermal comfort in cold climate urban areas. Energy and Buildings, 342. https://doi.org/10.1016/j.enbuild.2025.115861

Yoo, J., Eom, J., & Zhou, Y. (2024). Thermal comfort and retail sales: A big data analysis of extreme temperature's impact on brick-and-mortar stores. Journal of Retailing and Consumer Services, 77, 103699.

Yu, S., Eom, J., Zhou, Y., Evans, M., & Clarke, L. (2014). Scenarios of building energy demand for China with a detailed regional representation. Energy, 67, 284-297.

Yu, Z., Fung, B. C. M., Haghighat, F., Yoshino, H., & Morofsky, E. (2011). A systematic procedure to study the influence of occupant behavior on building energy consumption. Energy and Buildings, 43(6), 1409–1417. https://doi.org/10.1016/j.enbuild.2011.02.002

Yuan, J., Jiao, Z., Xiao, X., Emura, K., & Farnham, C. (2024). Impact of future climate change on energy consumption in residential buildings: A case study for representative cities in Japan. Energy Reports, 11, 1675–1692. https://doi.org/10.1016/j.egyr.2024.01.042

Zarco-Periñán, P. J., Zarco-Soto, I. M., & Zarco-Soto, F. J. (2021). Influence of the population density of cities on energy consumption of their households. Sustainability (Switzerland), 13(14). https://doi.org/10.3390/su13147542

Zhang, W., Liu, D., Tian, H., Pan, N., Yang, R., Tang, W., Yang, J., Lu, F., Dayananda, B., Mei, H., Wang, S., & Shi, H. (2024). Parsimonious estimation of hourly surface ozone concentration across China during 2015–2020. Scientific Data, 11(1). https://doi.org/10.1038/s41597-024-03302-3

Zhao, Y., Ding, X., Wu, Z., Yin, S., Fan, Y., & Ge, J. (2024). Impact of urban form on building energy consumption in different climate zones of China. Energy and Buildings, 320. https://doi.org/10.1016/j.enbuild.2024.114579

Zhou, K., Hu, D., Hu, R., & Zhou, J. (2023). High-resolution electric power load data of an industrial park with multiple types of buildings in China. Scientific Data, 10(1), 870.

Zhou, Y., Eom, J., & Clarke, L. (2013). The effect of global climate change, population distribution, and climate mitigation on building energy use in the US and China. Climatic Change, 119(3), 979-992.