

scientific data



OPEN

DATA DESCRIPTOR

A Large-Scale Residential Load Dataset in a Southern Province of China

Bo Li^{1,2,6}, Ruotao Yu^{1,6}, Kaiye Gan¹, Guangchun Ruan³, Shangwei Liu^{1,4}, Mingxia Yang¹, Daiyu Xie⁵, Wei Dai¹✉ & Haiwang Zhong²✉

Granular, localized data are essential for generating actionable insights that facilitate the transition to a net-zero energy system, particularly in underdeveloped regions. Understanding residential electricity consumption—especially in response to extreme weather events such as heatwaves and tropical storms—is critical for enhancing grid resilience and optimizing energy management strategies. However, such data are often scarce. This study introduces a comprehensive dataset comprising hourly transformer-level residential electricity load data collected between 2022 and 2023 from 23 residential communities across 10 cities in Guangxi Province, China. The dataset is augmented with meteorological data, including temperature, humidity, and records of extreme weather events. Additionally, calendar-related data (e.g., holidays) are included to facilitate the analysis of consumption patterns. The paper provides a detailed overview of the methodologies employed for data collection, preprocessing, and analysis, with a particular emphasis on how extreme weather influences electricity demand in residential areas. This dataset is anticipated to support future research on energy consumption, climate change adaptation, and grid resilience.

Background & Summary

The increasing extreme weather events attributed to climate change significantly impact energy consumption in urban residential. Specifically, during periods of both low and high ambient temperatures, building heating, ventilation, and air conditioning (HVAC) are keys to maintain thermal comfort^{1,2}. The heightened reliance on HVAC systems leads to increased peak electricity demand across various regions, such as the U.S.³, European², Mexico⁴, China⁵, *et al.* Consequently, accurate forecasting of electricity load under diverse weather conditions is essential for effective power system operation and planning, encompassing aspects such as unit commitment, economic dispatch, demand-side management, and transmission line cogeneration management. Given that power grids are designed to meet maximum load, the availability of high-quality electricity demand data is critical for analyzing electricity consumption patterns and informing the deployment of peak generation capacity.

Currently, numerous load datasets are accessible from various countries, including Australia⁶, Austria^{7,8}, Canada⁹, China^{10,11}, German^{7,12,13}, Irish¹⁴, Italy⁸, Japan^{15,16}, Korean¹⁷, Portugal^{18,19}, Spain²⁰, UK^{21–29}, Uruguay³⁰, and USA^{31–33}. For instance, the ECD-UY³⁰ dataset provides comprehensive electricity consumption records for residential households in Uruguay, featuring the total household consumption with sample intervals ranging from one to fifteen minutes. Another dataset is collected from an industrial park with multiple types of buildings in Suzhou, China¹¹, spanning six years with 5-minite intervals. The Spanish dataset provides valuable insights into electricity consumption patterns of Spanish household²⁰, comprising 25,559 raw hourly time series data collected over three years. Additionally, the EWELD dataset represents a large-scale compilation of electricity consumption data from 386 industrial and commercial users in southern China, recorded over six years at fifteen-minute intervals¹⁰. Furthermore, the IoT-based data collected in a single-family apartment in China³⁴ includes information on occupant behavior, thermal environment, and appliance electricity use. These datasets

¹School of Electrical Engineering, Guangxi University, Nanning, 530004, China. ²State Key Laboratory of Power System Operation and Control, Department of Electrical Engineering, Tsinghua University, Beijing, 100190, China.

³Laboratory for Information & Decision Systems, Massachusetts Institute of Technology, Boston, MA, 02139, USA.

⁴Princeton School of Public and International Affairs, Princeton University, Princeton, NJ, 08540, USA. ⁵Power Dispatching and Control Center, Guangxi Grid Co., Ltd, Nanning, 530000, China. ⁶These authors contributed equally: Bo Li, Ruotao Yu. ✉e-mail: weidai@gxu.edu.cn; zhonghw@mail.tsinghua.edu.cn

Dataset	Country	Duration	Resolution	Description	Weather	Extreme event	Source
SGSC	Australia	2010–2014	30 min	78,720 Houses	—	—	6
ADRES		2009–2010	2 second	30 Houses	—	—	7
GREEND	Austria	2013	1 second	8 Houses	—	—	8
AMPds2	Canada	2012–2014	1 min	1 Houses	Yes	—	9
EWELD	China	2015–2022	15 min	386 enterprises	Yes	Yes	10
—		2016–2021	5 min	5,600 buildings	Yes	—	11
NOVAREF	German	2013–2016	2 second	12 Houses	—	—	7
DEDDIAG		2016–2020	1 second	15 Houses	—	—	13
WPuQ		2018–2020	10 second	38 Houses	Yes	—	48
CoSSMic		2014–2019	1 min	11 households	—	—	12
ISSDA	Irish	2009–2010	30 min	5000 homes and businesses	—	—	14
—	Japan	2012–2020	30 min	Over 7000 buildings	—	—	15,16
ENERTALK	Korean	2016–2017	15 Hz	22 Houses	—	—	17
FIKElectricity	Portugal	2019	1 second	3 Restaurant Kitchens	—	—	18
Elergone		2011–2014	15 min	370 clients	—	—	19
—	Spain	2014–2022	1 hour	25,559 customers	—	—	20
LCL	UK	2011–2014	30 min	5567 Houses	—	—	27
NESEMP		2010–2012	5 min	215 Houses	Yes	—	25
EDRP		2007–2010	30 min	14000 Houses	—	—	22
IDEAL		2018	1 second	255 Houses	Yes	—	24
METER		2016–2019	1 min	529 individuals	—	—	23
REFIT		2013–2015	8 second	20 Houses	—	—	21
SAVE		2016–2019	15 min	Over 5000 homes	—	—	28
SERL		2021	30 min	13000 Houses	Yes	—	26
UK-DALE		2013–2015	1 second	5 Houses	—	—	29
ECD-UY	Uruguay	2019–2020	15 min	110,953 customers	—	—	30
MFRED	USA	2019	10 second	390 apartments	—	—	33
EULP		2018	15 min	277 buildings	Yes	—	31,32
Our dataset	China	2022–2023	1 hour	80,000 households	Yes	Yes	—

Table 1. Selection of public load demand datasets.

are instrumental for analyzing energy consumption patterns, conducting energy disaggregation studies, and designing energy billing plans.

Table 1 presents a summary of the existing load demand datasets. It should be noted that most of these datasets are outdated with limited coverage of recent five years. Furthermore, many of these datasets feature a restricted number of supply points, including those from sources such as AMPds2, FIKElectricity, UK-DALE, GREEND, CoSSMic, NOVAREF, DEDDIAG, REFIT, ENERTALK, ADRES, WPuQ, NESEMP, IDEAL, EULP, Elergone, EWELD, and MFRED. The sample size of these datasets is typically fewer than 500 houses, and represent the usage patterns of a small group of residents, which is insufficient for reliable demand analysis. Additionally, climate change and extreme weather events exert consideration influence on energy systems, impacting both energy demand and supply resilience. However, only a limited number of datasets, such as those from AMPds2, REFIT, WPuQ, NESEMP, IDEAL, EULP, EWELD, SERL, and SGSC, provide climate data, and even when such data is available. It predominantly consists of temperature information. Furthermore, there is a notable lack of load datasets pertaining to China and residential areas, as well as a lack of publicly accessible data on load forecasting in the context of extreme weather events. Most of the existing datasets originate from English-speaking countries that share similar cultural attributes, which may introduce biases in the interpretations and conclusions drawn from the data, particularly in the domain of human behavior analysis.

Comprehensive datasets of with above features are seldom employed for research purposes and are often difficult to obtain. Furthermore, certain available load datasets are exclusively applicable to single-family households and do not extend to building-level analysis within residential communities. Most public load datasets exhibit the limitations of small sample sizes, which has become increasingly evident in recent years. Additionally, there is a significant lack of load datasets pertaining to China and its residential areas as well as publicly available datasets on load analysis during extreme weather events.

We are committed to establishing an open load dataset that incorporates extreme weather information to fill the existing gaps. This study focuses on electricity load of residential communities, which contributes to regional diversity in accessible energy datasets. In specific, the dataset includes extreme weather events such as tropical storms and extremely high temperatures, collected from Jan. 2022 to Nov. 2023 from 23 residential communities in ten cities of Guangxi, covering more than 80,000 family households. Despite being limited in scope, the data offer valuable insights that can significantly contribute to regional decision-making and planning.

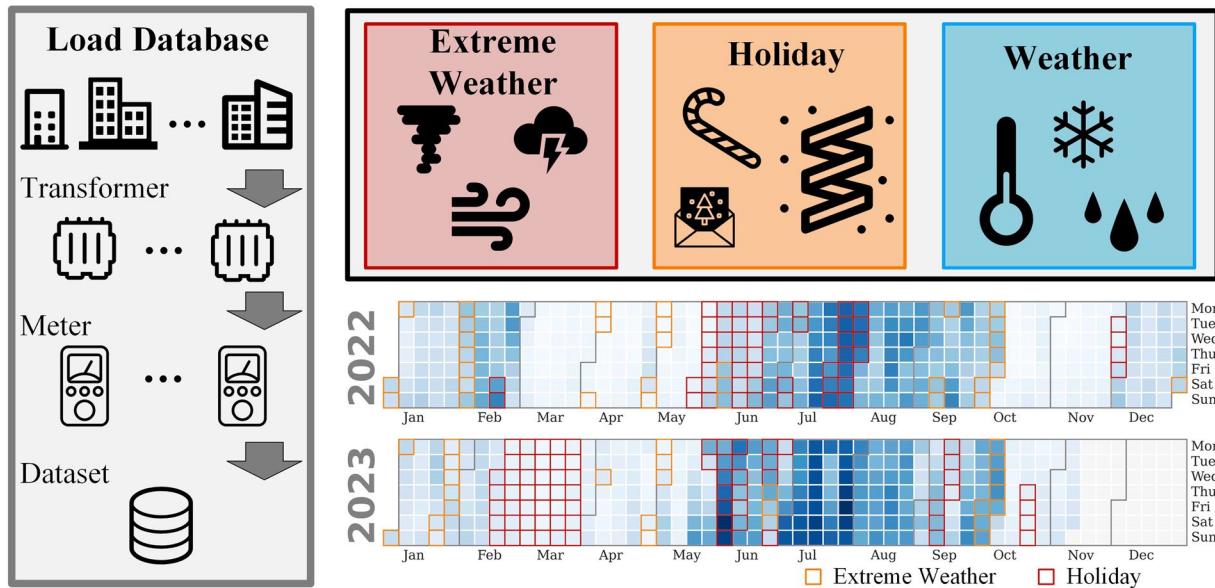


Fig. 1 Illustration of the dataset construction.

The proposed dataset is featured with the following unique characteristics.

- It exhibits a high degree of diversity, encompassing electricity consumption data from 23 residential communities across 10 cities, which collectively represent over 80,000 family households in Guangxi, China. This dataset comprises aggregated load profiles derived from electricity transformer meters and emphasizes consumption patterns at the community level. The variety of sources enhances the dataset's comprehensiveness, facilitating electricity load forecasting under a range of extreme conditions through data-driven methodologies.
- The dataset encompasses a substantial collection period from 2022 to 2023, with an hourly sampling frequency. This high sampling rate accurately captures contemporary trends in energy consumption, providing a precise representation of current societal energy usage patterns. It enables high-fidelity analysis such as real-time forecasting into electricity demand during this period. The high temporal resolution, characterized by hourly measurements, allows for an in-depth examination of electrical consumption behaviors.
- The dataset integrates weather data, as well as extreme weather events. It encompasses 12 distinct definitions of extreme weather and over 13 types of weather-related information. This comprehensive dataset facilitates quantitative analyses of climate change impacts on electricity consumption, such as peak demand. Examining load variations in response to weather fluctuations, deepen the current understanding of the interplay between external environmental factors and power consumption.

Methods

Data acquisition and collection. The dataset comprises hourly electricity demand data from 423 electricity transformer meters, which serve a total of 23 residential communities encompassing approximately 80,000 family households. Figure 1 illustrates the methodology employed in constructing the dataset. Table 2 presents essential information regarding the communities, including transformer capacity, the number of transformers, and the number of households. It is noted that the dataset is aggregated based on electricity transformers with rated capacities ranging from 400 kVA to 5,000 kVA. Variations in the number of transformers are attributed to the population size of the respective residential community. The consumption patterns observed across different transformer meters within the same community exhibit highly comparable trends, leading to the aggregation of each community into a singular load profile. The data spans from January 2022 to November 2023 and encompasses a wide array of information relevant to electricity demand, geographical location, holidays, climatic conditions, and occurrences of extreme weather.

The electricity consumption data utilized in this study were collected and validated by the local electrical department, ensuring high accuracy and reliability. The data were obtained through a hierarchical metering system that monitors electricity usage at multiple levels, including individual households, residential buildings, and transformers. At the household level, electricity meters record the energy consumption of individual residences, while each residential building is equipped with a meter to aggregate the consumption of all its households. At the community scale, transformer meters capture the total electricity demand of all buildings connected to a specific transformer.

The dataset used in this study comprises exclusively transformer-level data, providing aggregated load profiles for residential communities. These transformer meters, specifically the Holley DTSD546 Smart Meter, are equipped with advanced monitoring technology capable of recording hourly electricity consumption

City	Capacity of transformer/kVA	Number of transformers	Households
C0-Baise	800/3200	37	5,112
C1-Liuzhou	500/630/800	36	5,082
C2-Guilin	800/1000	18	5,139
C3-Beihai	500/630/800/1000	4	878
C4-Guigang	800/1000/1250	46	2,510
C5-Nanning	500/630/800/1000	67	5,122
C6-Fangchenggang	6080/9190	90	46,837
C7-Qinzhou	800/2000/2800	13	3,185
C8-Hechi	800/1100/1200/2300/3100	79	8,456
C9-Wuzhou	400/500/800	33	5,988
Total	—	423	88,309

Table 2. Statistics of cities.

(<http://www.holleymeter.com/prodel.html?id=117>). The Holley DTSD546 is a multi-function, three-phase electronic meter designed for high precision in energy measurement. It features capabilities such as real-time data collection, remote monitoring, and comprehensive diagnostics to detect issues such as abnormal voltage, current imbalances, and meter tampering. The meter's high-resolution data capture ensures detailed insights into energy usage trends, making it ideal for load analysis and energy planning. The electrical department conducts routine validation and quality assurance to ensure the integrity of the data. This includes flagging and addressing anomalies such as irregular consumption patterns or meter errors during the preprocessing stage, thereby maintaining the dataset's reliability and accuracy.

The location-specific climate data come from the National Oceanic and Atmospheric Administration (NOAA) and provides essential information, including daily averages, maximum and minimum temperatures, humidity levels, mean wind speed, maximum sustained wind speed, gusts, precipitation amounts, and mean station pressure. This climate data serves as a crucial component for understanding the relationship between electricity consumption patterns and meteorological conditions. The solar radiation data used in this study were obtained from the ERA5 reanalysis dataset, provided by the European Centre for Medium-Range Weather Forecasts (ECMWF)³⁵. This dataset, specifically titled "Reanalysis Era5 Single Levels," contains hourly surface net solar radiation data for the years 2022 and 2023. The data was collected for a specific geographical area, encompassing latitudes from 20°N to 27°N and longitudes from 104°E to 113°E, which cover the geographic scope of the dataset. The data was retrieved in GRIB format and compressed into a ZIP file to facilitate efficient handling and analysis.

The criteria for defining extreme weather events are outlined in Table 3, which includes classifications such as heat index caution, heat index extreme caution, heat index danger, heat index extreme danger, wind chill, wind chill frostbite danger, wind chill great frostbite danger, tropical storm, severe tropical storm, typhoon, strong typhoon, and super typhoon. The extreme weather data are incorporated into the dataset using two distinct methodologies. First, we gathered data on extreme weather events from local media sources, which identified the "Top Ten" extreme weather events for 2022 and 2023 in the region. This approach ensures that significant local weather occurrences are well represented. Second, a statistical approach was employed, using NOAA weather data to compute relative humidity, heat index, and wind chill metrics, which are important for understanding extreme temperature conditions. Additionally, tropical storms and typhoons were categorized based on the Beaufort wind force scale, with gust-related threats excluded due to the absence of gust data at most monitoring stations. Table 3 outlines the criteria for defining extreme weather events. For example, a day is classified as a high-temperature day if the temperature exceeds the upper limits of the 95% confidence interval established within the dataset.

Table 4 provides a summary of the statistics related to electricity consumption data from 23 residential communities across 10 cities. It includes metrics such as mean values, standard deviation, skewness, and percentiles. Notably, significant disparities exist in the mean electricity consumption among the various communities, which can be attributed to differences in household sizes. All weather data are recorded daily and encompass mean and maximum values. Table 5 summarizes the statistics of the weather data from the same 23 residential communities, detailing parameters such as temperature, wind speed, dew point, and station pressure.

Data processing. *Data censoring.* To ensure data privacy and prevent the reverse engineering of transformer identities, we implemented a data censoring technique. In the original dataset, transformers were numbered according to their respective cities and communities in a logical order. Subsequently, each transformer across all cities was randomized and renumbered. As a result, transformers sharing the same city and community numbers indicate that they belong to the same community within the same city; however, the specific transformer numbers and corresponding community names have been censored. All code associated with the randomization process, as well as the original data prior to censoring, has been completely removed to safeguard privacy. This method guarantees that sensitive information regarding individual transformers cannot be traced back to specific locations.

Data cleaning. To ensure the quality of our dataset, we identified and removed outliers in the electricity consumption data, specifically instances where the maximum power consumption exceeded the rated capacity of

Extreme Weather	Definition
High Temperature	Temperature above the upper bound of the 95% confidence interval of maximum temperatures
Low Temperature	Temperature below the lower bound of the 95% confidence interval of minimum temperatures
High Humidity	Relative humidity above the upper bound of the 95% confidence interval of relative humidities
Heat Index Caution	Heat index higher than 27 °C and lower than 32 °C
Heat Index Extreme Caution	Heat index higher than 32 °C and lower than 41 °C
Heat Index Danger	Heat index higher than 41 °C and lower than 54 °C
Heat Index Extreme Danger	Heat index higher than 54 °C
Wind Level 5	Beauford wind scale level 5
Wind Level 6	Beauford wind scale level 6
Wind Level 7	Beauford wind scale level 7
Wind Level 8	Beauford wind scale level 8
Wind Chill Very Cold	Wind chill larger than -35 °C and lower than -25 °C
Wind Chill Frostbite Danger	Wind chill larger than -60 °C and lower than -35 °C
Wind Chill Great Frostbite Danger	Wind chill lower than -60 °C
Tropical Storm	Wind speed larger than 39 mph and smaller than 54 mph
Severe Tropical Storm	Wind speed larger than 54 mph and smaller than 73 mph
Typhoon	Wind speed larger than 73 mph and smaller than 93 mph
Strong Typhoon	Wind speed larger than 93 mph and smaller than 114 mph
Super Typhoon	Wind speed larger than 114 mph

Table 3. Extreme weather definition.

City	Mean	Standard Deviation	Skew	Kurtosis
C0	1209.03	716.42	0.75	-0.15
C1	1509.82	751.51	1.05	1.06
C2	1903.33	906.11	0.87	0.37
C3	91.08	36.49	0.93	0.55
C4	2386.18	1108.14	0.84	0.46
C5	4895.87	2290.43	0.8	0.2
C6	4010.95	1701.94	0.76	0.3
C7	917.57	429.02	0.92	0.66
C8	1791.28	855.26	0.85	0.33
C9	2141.14	1080.25	1.00	0.79

Table 4. Statistics of electricity consumption data.

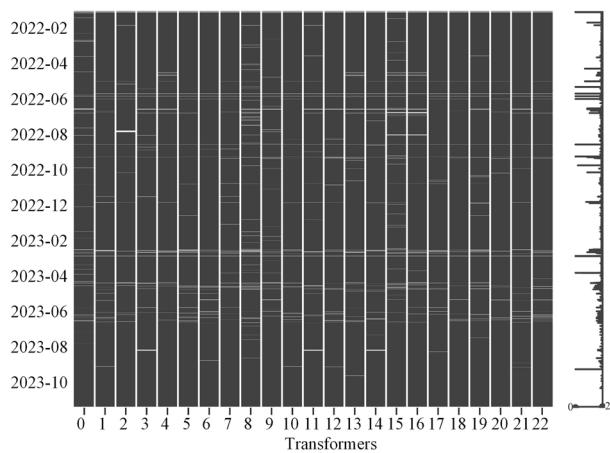
the transformer. These cases were deemed erroneous and excluded from the dataset. For the weather data, outliers were identified through statistical analysis to confirm that measurements—such as temperature, humidity, and precipitation—fell within reasonable ranges for the region. Any data points that fell outside these expected ranges were also considered erroneous and removed to maintain the integrity of the dataset.

Extreme weather events. The extreme weather events were compiled using two distinct methodologies: local identification and a statistical approach. The local identification method involved gathering data on the “Top Ten” extreme weather events for 2022 and 2023 from local media sources, thereby identifying the most significant weather occurrences in the region. The statistical approach utilized a script to retrieve historical weather data from NOAA. From this data, we calculated relative humidity, heat index, and wind chill metrics to characterize extreme temperature conditions. Additionally, tropical storms and typhoons were categorized based on the Beaufort wind force scale. Due to the lack of gust data at most monitoring stations, gust-related threats were excluded from the extreme weather categorization.

Imputation of missing data. The dataset includes several categories, such as electricity consumption data, weather data, extreme weather indicators, and holiday data. Figure 2 presents statistics on the missing values for each meter in Community 0 of City 0. Among the 200 transformer meters, 180 exhibit data completeness exceeding 95%. It is considered acceptable to exclude the entire dataset of transformer meters if the proportion of missing values exceeds 30%.

To address missing values, we employed the forward-backward imputation method for both the electricity consumption and weather datasets. This approach was chosen to ensure that missing values were imputed with context from both preceding and subsequent time periods. For the electricity consumption data, we implemented forward-backward imputation by averaging the data from the previous and following weeks. This

City	Mean value				Extreme value		
	Temperature/°C	Dew point/°C	Station Pressure/kPa	Wind Speed/(m/s)	Maximum Temperature/°C	Minimum Temperature /°C	Maximum sustained wind speed/ (m/s)
C0	22.05	16.88	85.36	1.59	34.78	6.61	6.02
C1	20.10	15.1	97.77	3.71	31.11	1.83	11.99
C2	20.23	14.44	72.94	2.12	31.44	2.72	11.01
C3	23.12	19.45	3.80	2.75	31.00	6.11	9.00
C4	22.01	16.98	60.96	3.1	31.39	5.61	18.01
C5	22.12	17.57	77.34	2.39	31.83	5.67	13.02
C6	22.32	19.28	22.64	2.99	31.22	5.44	9.00
C7	22.32	19.28	22.64	2.99	31.22	5.44	9.00
C8	21.17	15.35	97.99	1.59	32.28	4.67	6.02
C9	21.93	17.37	52.7	1.94	32.17	4.94	7.97

Table 5. Statistics of weather data.**Fig. 2** Missing data distribution of the community 0 in City 0 from 2022 to 2023. The sparkline on the right provides a concise summary of the data completeness, highlighting the rows with the highest and lowest number of missing values in the dataset. The y-axis displays datetime, while the x-axis depicts transformers. The white spaces denote the absence of data.

method is effective for filling gaps in data where consumption patterns are expected to be similar within a given week. The forward-backward method guarantees that the imputed values are balanced and reflect both prior and subsequent consumption trends. Similarly, for the weather data, the forward-backward imputation method was applied, with missing values filled by averaging the values from the day before and the day after the missing data point. This approach is particularly effective in situations where weather conditions follow daily cycles, as any missing data would likely resemble the values of the surrounding days. Three commonly used standard imputation methods were employed for comparison, as illustrated in Fig. 3. The linear interpolation method tends to yield suboptimal results when the time interval exceeds 24 hours; however, it performs more favorably than other methods for very short time intervals. It is crucial to recognize that missing values may convey significant information, such as meter failures due to extreme weather conditions. Furthermore, the peak demand within the load profile can be approximated using the mean value, as peak demand is integral to power grid dispatch strategies. Consequently, the raw data is preserved in the repository for future research endeavors.

Data Records

The dataset is available on Figshare and can be imported into any software that is compatible with the SQLite3 database format³⁶. A detailed analysis workflow is provided in the Code Availability section. It is important to note that the initial dataset comprises only raw data, while subsequent analyses are conducted on processed data. The dataset comprises seven distinct tables: raw transformer data, transformer metadata, weather data, weather station metadata, extreme weather data sourced from the internet, calculated extreme weather data, and festival data. The timestamps are recorded in UTC + 8 format (YYYY-MM-DD hh:mm:ss). The temporal range of the data extends from January 2022 to November 2023, with measurements taken at one-hour intervals.

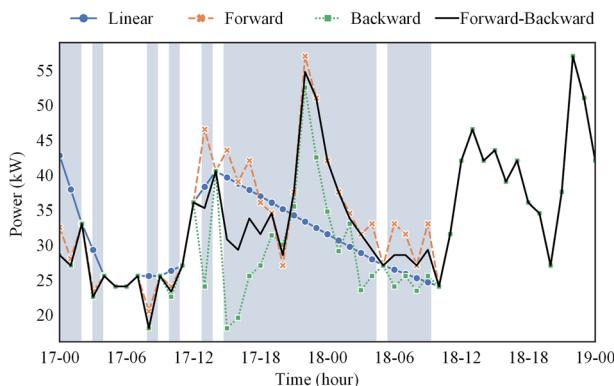


Fig. 3 Comparison of various imputation methods for addressing missing data. The selected forward-backward imputation method is denoted by the black line, while the shaded background indicates the presence of missing data. Three days in June 2022 are selected. Additional imputation methods, including linear interpolation, forward imputation, and backward imputation, are also evaluated for comparison purposes. Forward imputation addresses missing values by utilizing data from the preceding week, whereas backward imputation employs data from the subsequent week to fill in gaps. The forward-backward imputation method integrates both approaches by averaging the data from the week prior and the week following the missing values.

Technical Validation

This section evaluates the quality and reliability of the dataset using systematic validation approaches. We concentrate on verifying the key attributes of the data, including temporal trends, geographical patterns, and the influence of meteorological factors on electricity consumption. The validation process guarantees the dataset's robustness for future applications and research.

Temporal validation of electricity consumption patterns. We first validate the electricity consumption patterns across various time scales—annual, weekly, and daily—to ensure consistency with established trends. We examine the seasonal and daily consumption profiles of 23 residential communities, emphasizing the variation in electricity demand throughout different periods of the year. For instance, Fig. 4 illustrates the annual electricity consumption profiles of 10 cities, demonstrating a clear increase in demand during the summer months, which aligns with typical seasonal trends. Additionally, Fig. 5 presents the weekly and daily electricity consumption profiles, identifying peak consumption times, including midday and evening periods. The daily profiles reveal distinct peaks during midday and evening hours, driven by residential activities such as cooling, cooking, and lighting, with a noticeable dip during early morning hours when household activity is minimal. Weekly patterns indicate higher consumption on weekdays compared to weekends, reflecting standard work-week schedules. Monthly profiles reveal significant seasonal variations, with increased demand in the summer months due to air conditioning use and decreased demand in cooler months, influenced by weather conditions. These findings are validated against historical consumption patterns in the region to confirm their accuracy and reliability.

The Diversity Factor (Fig. 6) is evaluated as a measure of variability in electricity demand over time. This factor emphasizes the fluctuations in demand across different seasons and hourly intervals. The diversity factor is higher in summer and winter, which corresponds to increased variability in electricity consumption during extreme temperature conditions. Notably, the diversity factor is particularly pronounced during the late afternoon and early evening hours (hours 17–22), confirming the dataset's capacity to capture demand fluctuations effectively.

Geographical validation. We evaluate the geographical variability in electricity consumption across various cities, selected for their diverse socio-economic conditions and geographic locations. The six chosen cities in Guangxi, each characterized by distinct climates and economic structures, are validated for their representative attributes. Figure 5 illustrates the differences in electricity consumption profiles among the cities, including the influence of varying work schedules on consumption patterns. These geographical variations are supported by external data sources, such as demographic and economic reports, ensuring that the selected cities accurately represent broader regional trends.

Validation of meteorological influence on electricity consumption. Moreover, we examined the effects of meteorological variables on electricity consumption patterns, utilizing C0 and C2 as case studies. C0 refers to the coastal city of BH, which has an average temperature of 22.05 degrees Celsius. Coastal cities like BH are particularly vulnerable to marine-related disasters, such as typhoons. In contrast, C2 is located in the northern Guangxi province, an area prone to cold waves and rainstorms. Figure 7 illustrates the relationship between electricity consumption and various meteorological indicators for C0, including mean temperature and humidity. The correlation is assessed using the Pearson correlation coefficient, revealing a strong association between temperature and humidity with electricity consumption, suggesting that these factors exert a significant positive influence on power usage, in alignment with the prior knowledge.

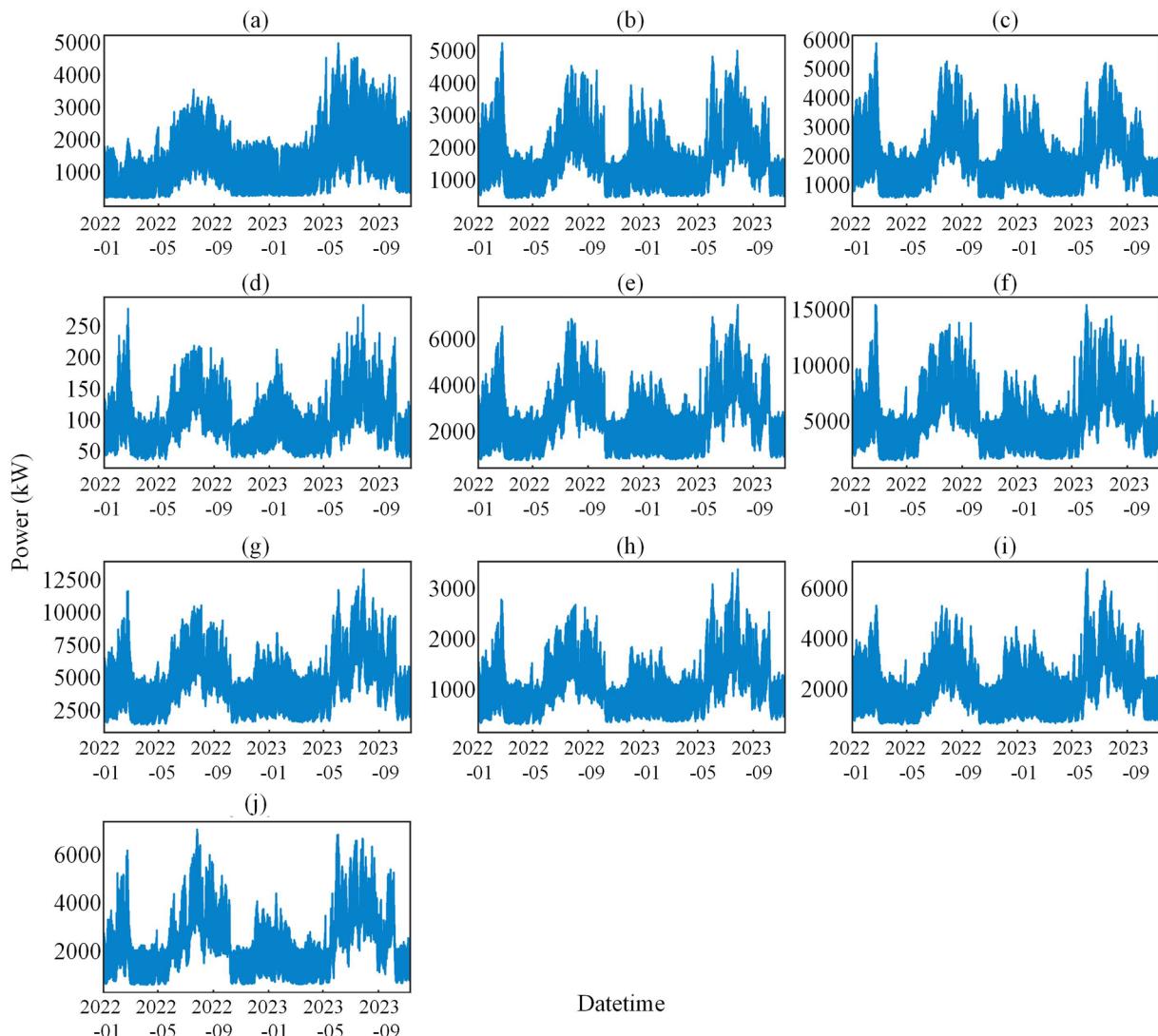


Fig. 4 Annual electricity consumption profiles of 10 cities from 2022 to 2023. (a–j) represent city 0 to city 9.

The lower left section of Fig. 7(a) illustrates the daily fluctuations in electricity consumption in relation to temperature. The data indicates a U-shaped relationship between temperature and electricity consumption. Specifically, the curve remains relatively flat within the comfortable temperature range of 18–25 °C, while it rises sharply when temperatures exceed 25 °C and increases more gradually when temperatures fall below 18 °C.

Validation of extreme weather event impact. Table 6 provides a summary of the proportion of extreme weather events recorded annually at the provincial level. In 2023, drought conditions were the most prevalent extreme weather event, comprising 9.9% of all recorded weather occurrences. As global temperatures continue to rise, climate change is likely to have significant implications for peak electricity demand during the summer moments, necessitating an increase in generation capacity. Figure 8 presents the effects of extreme weather on electricity consumption patterns in two selected cities, indicating that different types of extreme weather exert varying impacts on electricity demand. Figure 9 compares electricity demand profiles on typical days versus extreme weather days, which include conditions such as extreme heat indices, high and low temperatures, elevated humidity, level 5 winds, and precipitation exceeding 100 mm.

Usage Notes

HVAC energy consumption is significantly influenced by various factors, including outdoor temperature, relative humidity, solar radiation, building design, indoor setpoints, and user behavior. In addition to the outdoor temperature, solar radiation serves as a crucial determinant that affects the thermal comfort experienced within a building³⁷. The architectural characteristics of a building, such as insulation quality, window placement, and material selection, play a vital role in determining its thermal efficiency. For instance, buildings equipped with high-performance glazing and strategically designed shading devices can achieve a reduction in solar heat gain of approximately 30% to 50%, resulting in considerable energy savings for HVAC operations³⁸. Moreover, indoor

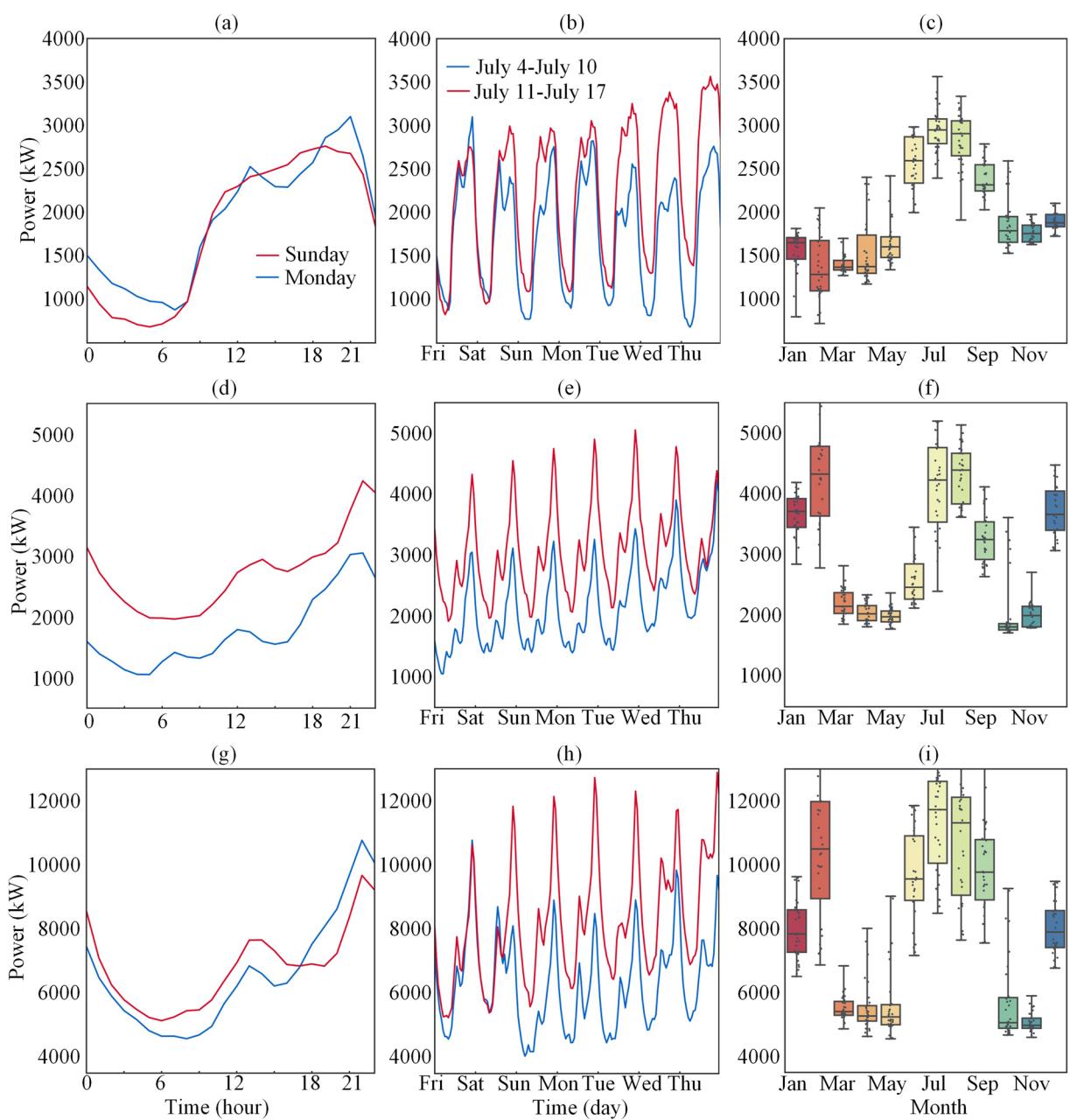


Fig. 5 Daily electricity consumption, weekly electricity consumption, and daily peak load distribution of selected three cities. The first row, second row, and third row are city 0, city 2, and city 5, respectively.

setpoints are critical, as adjustments to thermostat settings can lead to significant energy savings. Specifically, adjusting the setpoint can reduce the overall energy consumption of HVAC systems by over 7%³⁹. User behaviors also significantly impact electricity consumption, particularly when meteorological variables fall within certain thresholds. For example, extreme heat conditions often lead to increased air conditioner usage⁴⁰. These factors should be considered in future studies to enhance the accuracy and applicability of energy demand modeling. A summary of relevant datasets related to these factors is presented in Supplementary Table 1.

The dataset is derived from a limited number of cities, each possessing unique geographical and socio-economic characteristics. While this may present challenges for direct generalization to other cities, the dataset still offers a valuable foundation for decision-making and planning. By analyzing load patterns and their responses to meteorological variations, this dataset can serve as a baseline for extrapolating insights to cities with similar climates and population dynamics. The cities exhibit distinct relationships concerning average temperature, reflecting the climatic variability across Guangxi. According to the Statistics Bureau of China (<https://data.stats.gov.cn/>), in 2023, Guangxi's GDP per capita is comparatively lower, at approximately \$8,317, positioning it 29th among all provinces. Nanning, the provincial capital, has a population of approximately 8.89 million and a GDP per capita of around \$9,474. Nanning experiences a subtropical climate, with average temperatures ranging

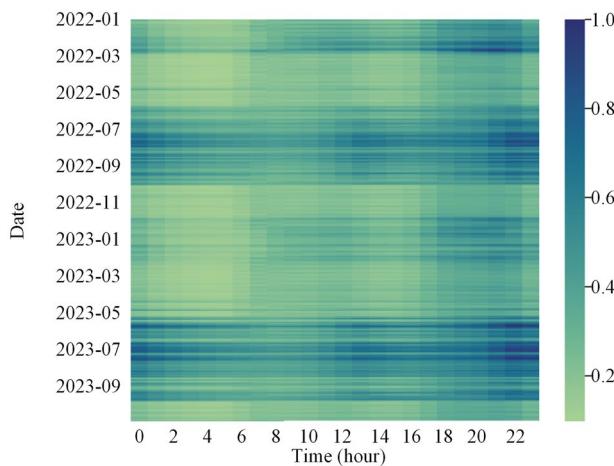


Fig. 6 The diversity factor of electricity demand from January 2022 to November 2023. The temporal resolution is one hour.

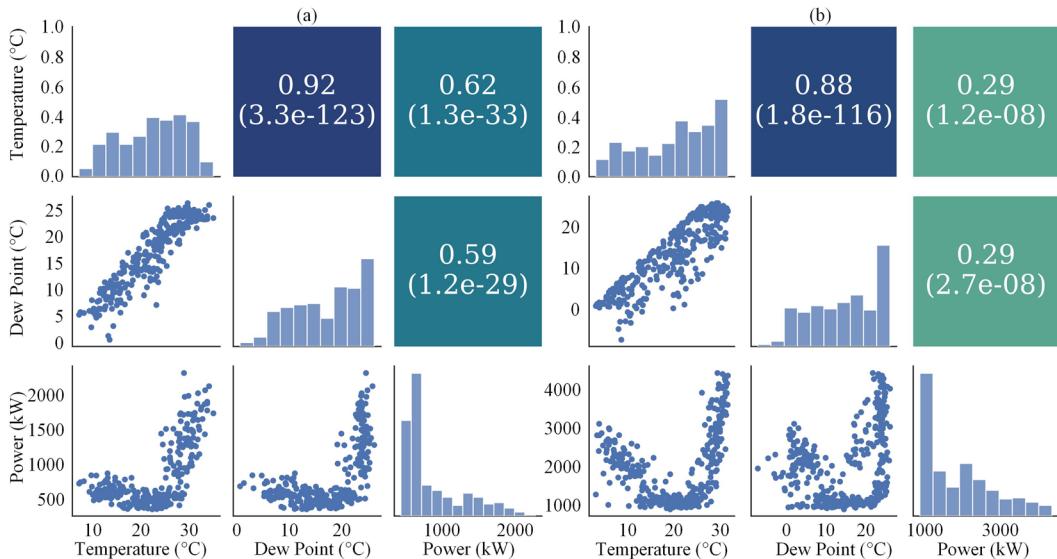


Fig. 7 The relationship between electricity consumption and meteorological indicators for City 0 and City 2 throughout the data collection period. **(a)** City 0. **(b)** City 2. The upper triangles present the Pearson correlation coefficients among the various factors, while the lower triangles illustrate the corresponding diagrams. The correlation coefficient is represented by the Pearson coefficient, and the second row indicates the p-values associated with these correlations.

Year	Cold wave	Dragon-boat rain	Excessive flooding	Hot weather	Typhoon	Drought	Extreme rainfall	Rainstorm	Severe convective weather
2022	1%	8.8%	0.6%	3.8%	1.1%	0	0	0	0
2023	0	2.2%	0	2.2%	3%	9.9%	1.1%	0.6%	1.1%

Table 6. The proportion of extreme weather events collected from the internet of Guangxi Province in 2022 and 2023.

from 10°C to 18°C in January to 26°C to 34°C in July. Guilin, located in the northeastern region of Guangxi, has a population of about 4.95 million and a GDP per capita of approximately \$7,851. Guilin's climate is characterized by mild winters and warm, humid summers, with average temperatures ranging from 6 to 15°C in January to 26°C to 34°C in July. In the western part of Guangxi, Baise has a population of 3.57 million and a GDP per capita of roughly \$7,980. Baise has a tropical monsoon climate, with average temperatures ranging from 10 to 20°C in January to 25°C to 35°C in July.

Moreover, this dataset serves as a resource for developing time-of-use electricity pricing strategies, providing a data-driven approach to structuring tariffs that align with consumption behaviors. By analyzing peak

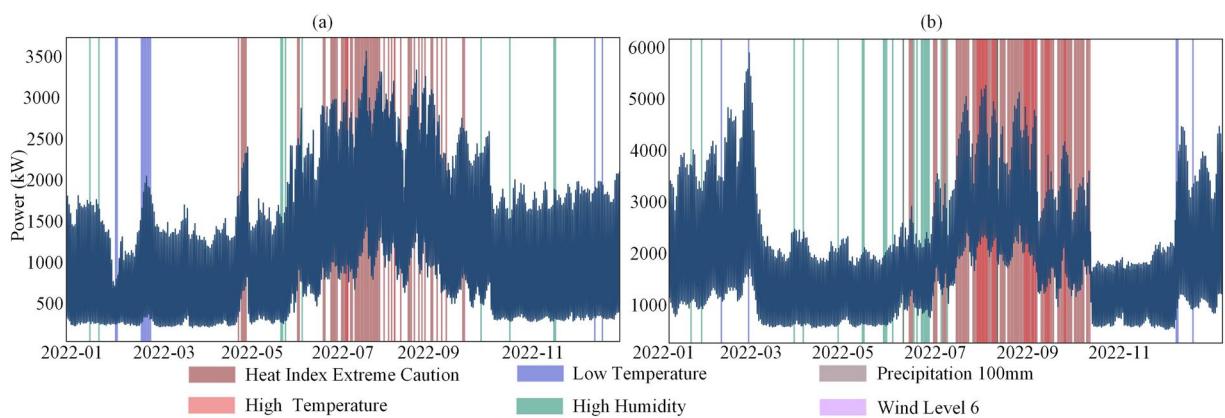


Fig. 8 The impacts of extreme weather on electricity consumption for selected cities, C0 and C2. (a) City 0. (b) City 2. Extreme weather events are assessed based on the criteria outlined in Table 3. The colored background serves to highlight these weather occurrences.

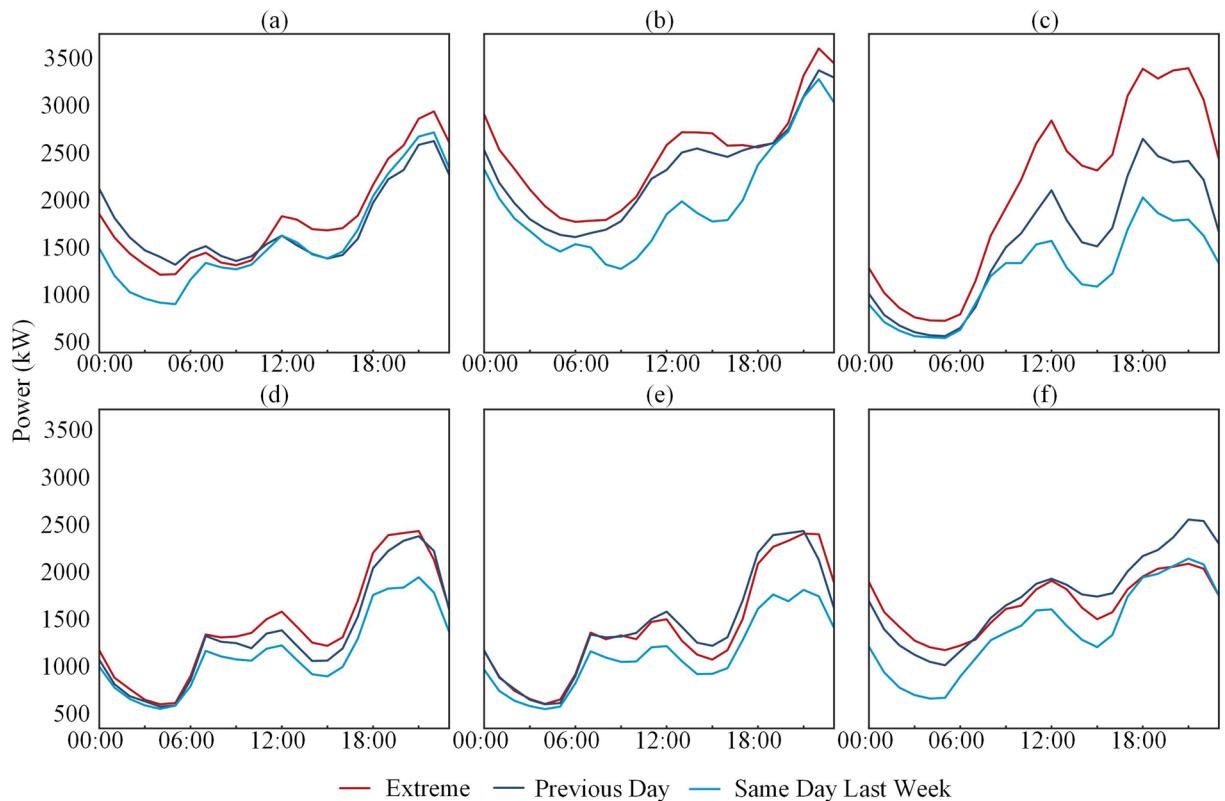


Fig. 9 The comparison of electricity demand profiles between normal days and extreme days during extreme weather for C2. (a) Heat index extreme caution. (b) High temperature. (c) Low temperature. (d) High humidity. (e) Level 5 wind. (f) Precipitation reached 100 mm.

and off-peak usage patterns, policymakers can design pricing schemes that encourage consumers to shift their energy consumption to periods of lower demand. This approach alleviates stress on the grid during peak times and promotes more efficient energy use⁴¹.

While prospective climate scenarios may differ from the conditions observed during the data collection phase, this dataset serves as a vital resource for long-term strategic planning by providing baseline information on energy consumption trends. Historical data offers a foundational reference for analyzing the potential evolution of energy consumption patterns in response to rising temperatures. For instance, Matteo *et al.* demonstrated the use of historical data in modeling the relationship between average temperature and energy demand across Italy⁴². Furthermore, the integration of historical datasets with advanced climate models enables researchers to

predict energy demand under various climate change scenarios, thereby strengthening the planning for resilient energy systems⁴³.

Climate change is characterized by long-term alterations in average weather patterns, where immediate changes may not be readily observable. The data presented in this study remain instrumental as a foundational reference for comprehending current HVAC energy consumption trends, which can be leveraged to track future variations. The Intergovernmental Panel on Climate Change (IPCC) projects that if current warming trends persist, future temperatures could rise by 1.5 °C above pre-industrial levels between 2030 and 2052, which would have significant implications for energy consumption⁴⁴. Our dataset, gathered over a period exceeding one year, predominantly captures typical conditions, thereby providing a solid basis for evaluating future shifts in energy usage attributable to rising temperatures. Additionally, the dataset encompasses instances of extreme weather, including both high and low temperature events, which can serve as benchmarks for assessing the effects of such occurrences on energy demand. Additionally, the dataset includes instances of extreme weather conditions, such as high and low temperatures, which provide essential insights into their potential impacts on load demand. A study examining the relationship between electricity usage and temperature across various regions of China revealed consistent patterns over time⁴⁰. Although climate change unfolds gradually, the frequency of extreme weather events, including heatwaves and cold spells, is increasing annually. Such events have a pronounced effect on residential electricity consumption, as evidenced by recent extreme weather occurrences in different regions. Consequently, this dataset remains instrumental in understanding user behavior and electricity consumption patterns over an extended timeframe, thereby supporting adaptive planning for future climate scenarios.

Achieving net-zero energy buildings, structures that generate as much energy as they consume, requires the effective integration of distributed energy resources (DERs), such as solar photovoltaic (PV) systems and energy storage solutions. Net-zero buildings not only contribute to sustainability but also reduce reliance on grid electricity. The critical challenge in this integration is addressing the temporal mismatch between energy generation and consumption. Solar PV systems, for example, produce peak energy during midday, while residential electricity demand often peaks in the late afternoon and evening. This discrepancy underscores the necessity for efficient energy storage solutions to balance supply and demand⁴⁵. Accurate modeling of energy consumption patterns allows for the optimal sizing and placement of renewable energy systems, ensuring they effectively meet the energy requirements of buildings. Research has demonstrated that analyzing temporal variations in energy usage is crucial for integrating energy storage systems, enabling the storage of excess energy during low-demand periods and its utilization during peak demand times⁴⁶.

This dataset offers valuable insights into consumption patterns and load demands, serving as a key tool for planning distributed energy systems aimed at achieving net-zero energy goals. It facilitates the design of cost-effective energy systems by enabling precise modeling of energy needs, which is essential for sustainable and resilient building designs. In summary, leveraging detailed consumption data and optimizing the integration of distributed energy resources and storage systems are fundamental steps toward realizing net-zero energy buildings. These efforts are vital for creating a sustainable and resilient built environment, aligning with global initiatives to reduce carbon emissions and mitigate climate change⁴⁷.

Code availability

The code implementation was done with Python 3.11. The source code including data preprocessing, data analysis, and data visualization is available on GitHub (<https://github.com/yuruotao/District-power>) and is licensed under the MIT license. The repository includes a README.md file with detailed instructions and all functions are well commented.

Received: 8 August 2024; Accepted: 6 March 2025;

Published online: 18 March 2025

References

1. Petri, Y. & Caldeira, K. Impacts of global warming on residential heating and cooling degree-days in the United States. *Scientific reports* **5**, 12427 (2015).
2. Wenz, L., Levermann, A. & Auffhammer, M. North–south polarization of European electricity consumption under future warming. *Proceedings of the National Academy of Sciences* **114**, E7910–E7918 (2017).
3. Auffhammer, M., Baylis, P. & Hausman, C. H. Climate change is projected to have severe impacts on the frequency and intensity of peak electricity demand across the United States. *Proceedings of the National Academy of Sciences* **114**, 1886–1891 (2017).
4. Davis, L. W. & Gertler, P. J. Contribution of air conditioning adoption to future energy use under global warming. *Proceedings of the National Academy of Sciences* **112**, 5962–5967 (2015).
5. Cao, J. *et al.* Spatially heterogeneous effect of temperature on electricity consumption in Shenzhen, China. *Building and Environment* **241**, 110468 (2023).
6. in *Smart-Grid Smart-City Customer Trial Data* (2022).
7. Anvari, M. *et al.* Data-driven load profiles and the dynamics of residential electricity consumption. *Nature communications* **13**, 4593 (2022).
8. Monacchi, A., Egarter, D., Elmenreich, W., D'Alessandro, S. & Tonello, A. M. in *2014 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, 511–516 (IEEE).
9. Makonin, S., Ellert, B., Bajić, I. V. & Popowich, F. Electricity, water, and natural gas consumption of a residential house in Canada from 2012 to 2014. *Scientific data* **3**, 1–12 (2016).
10. Liu, G. *et al.* EWLD: A Large-Scale Industrial and Commercial Load Dataset in Extreme Weather Events. *Scientific Data* **10**, 615 (2023).
11. Zhou, K., Hu, D., Hu, R. & Zhou, J. High-resolution electric power load data of an industrial park with multiple types of buildings in China. *Scientific Data* **10**, 870 (2023).
12. *Data Platform – Open Power System Data*, <https://open-power-system-data.org/> (2020).
13. Wenninger, M., Maier, A. & Schmidt, J. DEDDIAG, a domestic electricity demand dataset of individual appliances in Germany. *Scientific Data* **8**, 176 (2021).
14. Regulation, C. f. E. (Irish Social Science Data Archive Dublin, Ireland, 2012).

15. in *Sustainable Open Innovation Initiative*.
16. Komatsu, H. & Kimura, O. Customer segmentation based on smart meter data analytics: Behavioral similarities with manual categorization for building types. *Energy and Buildings* **283**, 112831 (2023).
17. Shin, C. et al. The ENERTALK dataset, 15 Hz electricity consumption data from 22 houses in Korea. *Scientific data* **6**, 193 (2019).
18. Pereira, L. et al. FIKElectricity: A Electricity Consumption Dataset from Three Restaurant Kitchens in Portugal. *Scientific Data* **10**, 779 (2023).
19. Trindade, A. in *Uci machine learning repository-electricity load diagrams 2011–2014 dataset* (2016).
20. Quesada, C., Astigarraga, L., Merveille, C. & Borges, C. E. An electricity smart meter dataset of Spanish households: insights into consumption patterns. *Scientific Data* **11**, 59 (2024).
21. Murray, D., Stankovic, L. & Stankovic, V. An electrical load measurements dataset of United Kingdom households from a two-year longitudinal study. *Scientific data* **4**, 1–12 (2017).
22. in *Energy demand research project: Early smart meter trials, 2007–2010* (2018).
23. in *ETER: UK Household Electricity and Activity Survey, 2016–2019: Secure Access* (ed Diakonova Grunewald P., M.) (2020).
24. Pullinger, M. et al. The IDEAL household energy dataset, electricity, gas, contextual sensor data and survey data for 255 UK homes. *Scientific Data* **8**, 146 (2021).
25. Craig, T., Polhill, J. G., Dent, I., Galan-Diaz, C. & Heslop, S. The north east scotland energy monitoring project: exploring relationships between household occupants and energy usage. *Energy and Buildings* **75**, 493–503 (2014).
26. Webborn, E. et al. The SERL Observatory Dataset: Longitudinal smart meter electricity and gas data, survey, EPC and climate data for over 13,000 households in Great Britain. *Energies* **14**, 6934 (2021).
27. DataStore, L. in *Smartmeter energy consumption data in London households* (ed UK Power Networks) (2013).
28. Rushby, T., Anderson, B., James, P. & Bahaj, A. Solent Achieving Value from Efficiency (SAVE) Data, 2017–2018 (2020).
29. Kelly, J. & Knottenbelt, W. The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes. *Scientific data* **2**, 1–14 (2015).
30. Chavat, J., Nesmachnow, S., Graneri, J. & Alvez, G. ECD-UY, detailed household electricity consumption dataset of Uruguay. *Scientific Data* **9**, 21 (2022).
31. Wilson, E. J. et al. End-use load profiles for the US building stock: Methodology and results of model calibration, validation, and uncertainty quantification. (National Renewable Energy Lab.(NREL), Golden, CO (United States) 2022).
32. Wang, C. et al. Impacts of climate change, population growth, and power sector decarbonization on urban building energy use. *Nature Communications* **14**, 1–16 (2023).
33. Meinrenken, C. J. et al. MFRED, 10 second interval real and reactive power for groups of 390 US apartments of varying size and vintage. *Scientific Data* **7**, 375 (2020).
34. Wang, C., Li, X., Sun, W., An, J. & Gao, S. Occupant behavior, thermal environment, and appliance electricity use of a single-family apartment in china. *Scientific Data* **11**, 65 (2024).
35. Bell, B. et al. The ERA5 global reanalysis: Preliminary extension to 1950. *Quarterly Journal of the Royal Meteorological Society* **147**, 4186–4227 (2021).
36. Li, B. et al. A Large-The provincial shapefile and whole database. *figshare* <https://doi.org/10.6084/m9.figshare.26333452.v1> (2024).
37. La Gennusa, M., Nucara, A., Pietrafesa, M. & Rizzo, G. A model for managing and evaluating solar radiation for indoor thermal comfort. *Solar Energy* **81**, 594–606 (2007).
38. Tzempelikos, A. & Athienitis, A. K. The impact of shading design and control on building cooling and lighting demand. *Solar energy* **81**, 369–382 (2007).
39. Kusiak, A., Li, M. & Tang, F. Modeling and optimization of HVAC energy consumption. *Applied Energy* **87**, 3092–3102 (2010).
40. Li, Y., Pizer, W. A. & Wu, L. Climate change and residential electricity consumption in the Yangtze River Delta, China. *Proceedings of the National Academy of Sciences* **116**, 472–477 (2019).
41. Nicolson, M. L., Fell, M. J. & Huebner, G. M. Consumer demand for time of use electricity tariffs: A systematized review of the empirical evidence. *Renewable and Sustainable Energy Reviews* **97**, 276–289 (2018).
42. De Felice, M., Alessandri, A. & Catalano, F. Seasonal climate forecasts for medium-term electricity demand forecasting. *Applied Energy* **137**, 435–444 (2015).
43. Morais, L. B. S. et al. Short-term load forecasting using neural networks and global climate models: An application to a large-scale electrical power system. *Applied Energy* **348**, 121439 (2023).
44. Allan, R. P. et al. in *Climate change 2021: The physical science basis. Contribution of working group I to the sixth assessment report of the intergovernmental panel on climate change 3–32* (Cambridge University Press, 2023).
45. Lund, H., Möller, B., Mathiesen, B. V. & Dyrelund, A. The role of district heating in future renewable energy systems. *Energy* **35**, 1381–1390 (2010).
46. Wörman, A., Pechlivanidis, I., Mewes, D., Riml, J. & Bertacchi Uvo, C. Spatiotemporal management of solar, wind and hydropower across continental Europe. *Communications Engineering* **3**, 3 (2024).
47. O'Shaughnessy, E., Cutler, D., Ardani, K. & Margolis, R. Solar plus: Optimization of distributed solar PV through battery storage and dispatchable load in residential buildings. *Applied Energy* **213**, 11–21 (2018).
48. Schlemminger, M., Ohrdes, T., Schneider, E. & Knoop, M. Dataset on electrical single-family house and heat pump load profiles in Germany. *Scientific data* **9**, 56 (2022).

Acknowledgements

This work was supported in part by the Specific Research Project of Guangxi for Research Bases and Talents (2022AC21257), and the National Natural Science Foundation of China (52122706).

Author contributions

Conceptualization and methodology, B.L., R.Y.; visualization and writing, B.L., R.Y., M.Y.; formal analysis, H.Z., G.R.; investigation, W.D., S.L., G.R., D.X.; funding acquisition, W.D., H.Z.; supervision, H.Z. All authors have read and agreed to the published version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41597-025-04766-7>.

Correspondence and requests for materials should be addressed to W.D. or H.Z.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025