www.nature.com/scientificdata

scientific data

OPEN

DATA DESCRIPTOR

Monthly electricity consumption data at 1 km × 1 km grid for 280 cities in China from 2012 to 2019

Xiaoqin Yan[1,2], Zhou Huang[1,2] ✉, Shuliang Ren[1,2], Ganmin Yin[1,2] & Junnan Qi[1,2]

High spatio-temporal resolution estimates of electricity consumption are essential for formulating effective energy transition strategies. However, the data availability is limited by complex spatio-temporal heterogeneity and insufficient multi-source feature fusion. To address these issues, this study introduces an innovative downscaling method that combines multi-source data with machine learning and spatial interpolation techniques. The method's accuracy showed significant improvements, with determination coefficients (R^2) increasing by 30.1% and 33.4% over the baseline model in two evaluation datasets. With this advanced model, we estimated monthly electricity consumption across 1 km x 1 km grid for 280 Chinese cities from 2012 to 2019. Our dataset is highly consistent with officially released electricity consumption of different industries (Pearson correlation coefficients within 0.83 - 0.91). Moreover, our data can reflect the electricity consumption patterns of different urban land uses compared to other datasets. This study bridges a significant gap in fine-grained electricity consumption data, providing a robust foundation for the development of sustainable energy policies.

Background & Summary

To accelerate the transition to a carbon-neutral world powered by emerging energy technologies by 2050, the world is required to achieve net-zero emissions[1]. The electricity sector, attributing 40% of worldwide carbon dioxide (CO_2) emissions, emerges as a critical focal point for decarbonization[2]. Consequently, nations have implemented policies to transform electricity systems and reduce conventional energy use, thus promoting sustainable development[3–5]. China is the world's largest consumer of electricity, accounting for 31% of electricity consumption in 2022, with operational inefficiencies and structural irrationalities[2]. In this context, high spatio-temporal resolution data can help to reveal such problems to formulate regional energy transition policies[6]. Official statistic data are only available at the county level and above, which makes it challenging to estimate consumption distribution and analyze spatio-temporal dynamics at fine scales[7]. Therefore, it is critical to use advanced methods to accurately estimate high spatio-temporal resolution electricity data.

High spatio-temporal electricity consumption estimation includes "bottom-up" and "top-down" approaches[8,9,10]. The "bottom-up" method accumulates data from individual units to higher aggregates, ensuring high accuracy at granular levels[11]. For example, total regional demand is calculated by aggregating electricity usage data from buildings[12]. However, the "bottom-up" approach is labor-intensive and time-consuming, which is not practical for processing massive long-term datasets. This becomes particularly challenging when swift data generation is essential for decision-making or emergency responses[13].

The "top-down" approach offers an effective alternative[14,15]. This approach utilizes open-source big data, such as satellite imagery and socio-economic data (e.g., gross domestic product (GDP), population density), serving as proxies for data downscaling[16–20]. For instance, Zhou *et al.*[21] developed a spatial disaggregation index using multi-source variables to estimate high-resolution building energy consumption. Furthermore, machine learning techniques effectively extract non-linear relationships between variables, enabling the "top-down" method to estimate electricity consumption accurately with minimal loss, thus improving the model's precision. Chen *et al.*[22] used the particle swarm optimization-back propagation (PSO-BP) algorithm to estimate electricity consumption at the 1 km × 1 km grid using nighttime lights as a proxy. This study demonstrated the synergy between big data and machine learning in downscaling electricity consumption. However, most existing datasets

[1]Institute of Remote Sensing and Geographic Information System, Peking University, Beijing, China. [2]Beijing Key Lab of Spatial Information Integration & Its Applications, Peking University, Beijing, China. ✉e-mail: huangzhou@pku.edu.cn

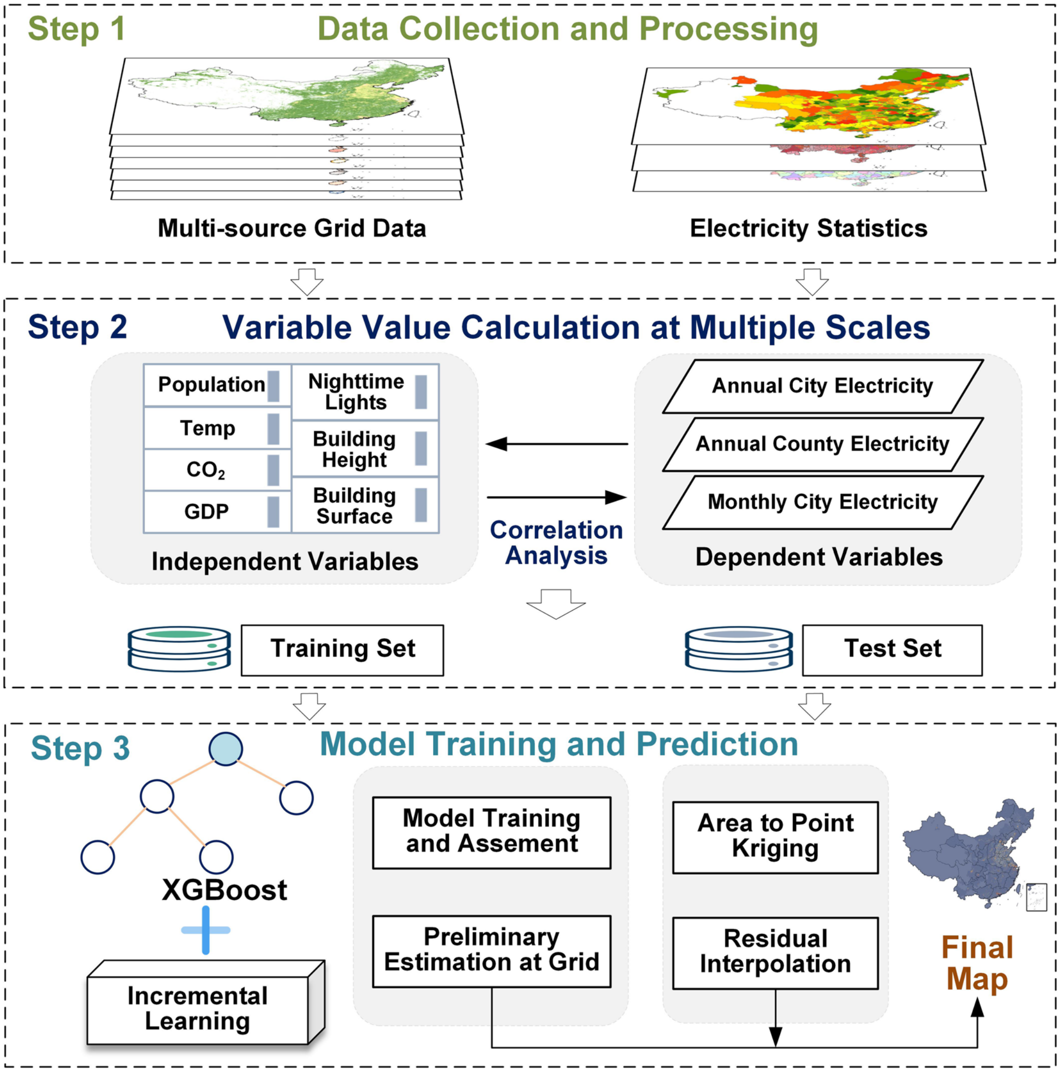

Fig. 1 The framework of the study.

provide only annual data, limiting applications that require monthly data (e.g., energy demand forecasting[23] or short-term effect analysis[24]). According to our survey, monthly electricity consumption data with high spatial resolution are not currently available in China.

Machine learning can effectively extract non-linear variable relationships[25,26] but struggles to capture spatial correlations[27], which are vital for accurately rendering detailed geographic information in high-resolution analyses[28]. Spatial correlation implies that geographically proximate objects are more likely to share similar attributes, and overlooking this factor would lead to increased predictive errors[29]. Kriging interpolation, an advanced technique grounded in spatial statistics, effectively identifies spatial correlations by analyzing distances and relationships among objects[30,31]. This method has proven invaluable in various downscaling applications, including estimations of population[32,33],emission[34], precipitation[35] and temperature[36,37]. Therefore, it is crucial to further integrate kriging interpolation to address the shortcomings of machine learning and improve prediction accuracy.

To address the aforementioned shortcomings, this study introduces a hybrid downscaling model that integrates machine learning with kriging interpolation[38]. This model estimates the electricity consumption of 280 major cities from April 2012 to December 2019 at 1 km × 1 km spatial resolution using multi-source data. This research highlights three contributions: (1) we created the first fine-grained electricity consumption dataset with monthly 1 km × 1 km in China; (2) the proposed method can extract complex variable correlations, which improves the estimation accuracy at different spatio-temporal scales; (3) kriging interpolation can characterize the spatial correlations, and also address the challenge of mismatch between predicted values and statistical data effectively.

Methods

The workflow of high spatio-temporal resolution of electricity data estimation in China is shown in Fig. 1. Firstly, we obtained electricity statistics and high spatio-temporal multi-source data as variables from the open platform. Secondly, by integrating machine learning with kriging interpolation techniques, we developed a "top-down" hybrid model for generating the final results. Finally, we verify the effectiveness of the data through spatio-temporal analysis.

Dataset	Temporal resolution	Spatial resolution	Description
Annual city electricity	Annual	City	Electricity consumption at multiple spatio-temporal scales, including urban and rural areas (data from local statistical offices).
Annual county electricity	Annual	County	
Monthly city electricity	Monthly	City	
Nighttime lights	Monthly	500 m	Reflects human activity on the surface, most notably human lighting at night (https://eogdata.mines.edu/products/vnl/).
Temperature	Monthly	1 km	Average ambient temperature (Peng *et al.*[53]).
CO_2	Monthly	1 km	Open-source data inventory for anthropogenic CO_2 (https://odiac.org/).
Population	Annual	100 m	Population distribution (http://www.worldpop.org/).
GDP	Annual	1 km	Economic development of the region (Chen *et al.*[54]).
Building height	2010, 2015, 2020	500 m	The distribution of the building heights (https://ghsl.jrc.ec.europa.eu/).
Building surface	2010, 2015, 2020	500 m	The distribution of built-up surfaces (https://ghsl.jrc.ec.europa.eu/).

Table 1. Information on datasets for high spatio-temporal electricity estimation.

Data collection and processing. This section summarizes the data products and corresponding preprocessing used to estimate high spatio-temporal electricity consumption data. Table 1 provides information on the resolution, source, and role of all datasets.

Statistical electricity data. Total electricity consumption data comes from Statistical Bureaus across China, covering 280 major prefecture-level cities (including urban and rural areas). It comprises three types of spatio-temporal resolutions. Firstly, annual electricity data at the prefecture-level city, achieving a coverage rate of nearly 95%, with missing values filled in by linear interpolation. The second and third datasets are monthly electricity consumption for prefecture-level cities and annual data for counties, respectively, providing more detailed spatial or temporal resolution. Only approximately 1,500 and 2,000 records were obtained due to data availability limitations.

Multi-source high spatio-temporal data. Previous studies usually focus on the effect of a single data (e.g., nighttime lights[19]) on electricity consumption, ignoring the integration of various factors. To accurately estimate electricity consumption on a fine scale, we have incorporated seven high spatio-temporal resolution variables: nighttime lights, average temperature, CO_2, population distribution density, GDP, building height, and building surface. These variables characterize the urban economic development, material stock and technological progress in multiple dimensions. They have demonstrated strong correlations with electricity data and are drivers for accurately capturing electricity consumption patterns[39–42]. In particular, the collected building height and surface data span over five years. To address temporal inconsistencies, we use these data to represent two adjacent years. For example, 2020 building datasets were used for both 2018 and 2019. This approach accounts for building change cycles, minimizing temporal variation impacts and ensuring accurate feature representation[43]. Due to the inconsistencies in coordinate systems and resolutions among different data sources, we converted all the data to the Albers equal-area coordinate system and resampled them to 1 km resolution to facilitate attribute extraction and model training.

Property calculations. Considering that electricity consumption is generated almost from built-up land, this study filters out built-up areas based on land use data[44] to reduce the estimation error. Additionally, given the significant quantitative differences in electricity consumption at different spatio-temporal scales, this study takes Cheng *et al.*[38] to use electricity intensity (monthly 1 km electricity consumption) as the dependent variable. For example, in the annual data electricity dataset, the electricity consumption is divided by the built-up area and twelve months. The same approach is applied to the dependent variable. In this way, all variables are in a uniform feature space, ensuring that model training is efficient and robust.

Constructing fine spatial and temporal scale methods for electricity estimation. This section introduces a hybrid downscaling model for electricity estimation, combining an improved XGBoost (eXtreme Gradient Boosting) algorithm with kriging interpolation (see Supplementary Figure S1 for details). We utilize processed electricity data as the dependent variable and multi-source spatio-temporal data as independent variables. The XGBoost model, coupled with incremental learning, is employed to extract features across various spatio-temporal scales for training. Subsequently, the output grid results are refined using kriging interpolation to capture geographic autocorrelation features and perform corrections, resulting in high-resolution electricity data products.

Estimation model of electricity consumption. The low resolution of annual electricity data may lead to substantial estimation errors when relying solely on these data for model training. Therefore, we employ finer-grained electricity data combined with incremental learning approach to progressively process the data streams in order to merge, refine and enhance the accumulated information[45]. This approach enables the model to comprehensively reveal electricity consumption patterns at different spatio-temporal scales.

Specifically, the first step is to build a base model using XGBoost and train it using annual city electricity data. The XGBoost is a gradient boosting tree model that integrates multiple weak classifiers, and improves model generalization performance by preventing overfitting through regularization. The second step is spatial feature incremental learning (XGBoost-SIL). We maintain the trained model's tree node weights and add several trees to fine-tune the model by incorporating annual county data. The same approach was applied in the third step and monthly city data is further added for temporal incremental learning (XGBoost-STIL). To analyze the effectiveness of spatio-temporal incremental learning of the method, 20% of the annual county data (spatial-test data) and monthly city data (temporal-test data) are selected as test datasets for the three models. The study uses parametric grid search and five-fold cross-validation to improve the model generalization performance and stability in the training process.

As the kriging interpolation method is only capable of spatial interpolation, is not able to disaggregate the annual data into monthly data. Therefore, this study used the above model to generate monthly county electricity density data as an intermediate product. This dataset is further adjusted by dasymetric mapping[46] to be used as baseline dataset for subsequent area-to-point kriging interpolation.

Electricity correction and mapping. Machine learning methods excel at capturing nonlinear relationships between variables but fall short in capturing the geographical spatial correlations. We used kriging interpolation to fill this gap. Specifically, area-to-point kriging is a modification of ordinary kriging that allows for low to high resolution redistribution[47]. This ensures the consistency of residuals across all grids within the same area. The residual for each grid is determined through a weighted linear combination of nearby coarse areas, following the unbiased weighting constraint.

$$\begin{cases} \hat{e}_p = \sum_{k=1}^{K} \lambda(v_k) e(v_k) \\ \sum_{k=1}^{K} \lambda(v_k) = 1 \end{cases} \tag{1}$$

where $\hat{e}_p$ signifies the estimated residual for the grid, K denotes the number of considered counties, $e(v_k)$ is the known area residual for county v_k, and $\lambda(v_k)$ represents the weight allocated to each county, calculated as follows:

$$\begin{cases} \sum_{l=1}^{K} \lambda(v_l) \bar{C}(v_k, v_l) - X_p = \bar{C}(p, v_k), \quad k = 1, \ldots, K \\ \sum_{l=1}^{K} \lambda(v_l) = 1 \end{cases} \tag{2}$$

where $\bar{C}(v_k, v_l)$ indicates the covariance between areas v_k and v_l, while $\bar{C}(p, v_k)$ describes the covariance between the target grid and the county. The covariance is derived from coarse to fine resolution by deconvolution. Deconvolution is generating variograms from discrete points of input area data by minimizing the variogram difference between the theoretical regularization and the input area data. Post-correction with kriging interpolation reduces errors by ensuring that aggregate of estimated results for grids within county matches the total electricity consumption.

Accuracy assessment. The coefficient of determination (R^2), the root mean square error (RMSE) and the spatio-temporal coefficient (F_I) to evaluate the model performance by calculating the error between the predicted values ($\hat{y}_i$) and true values (y_i). The coefficient (R) is applied to data correlation analysis. F_I drawing inspiration from the F1-score, make R^2 and RMSE comprehensive to assess the model's validity across temporal-test data (TD) and spatial-test data (SD):

$$F_I = 2 \times \frac{I_{SD} \times I_{TD}}{I_{SD} + I_{TD}} \tag{3}$$

where I is the R^2 or RMSE value. The final electricity consumption data are compared with related datasets to validate accuracy. Given the unavailability of data on the same scale, the study use national monthly data on total, residential, and industrial electricity consumption from statistical yearbooks for quantitative verification. Additionally, representative cities from diverse geographic locations–Beijing (North), Shanghai (East), Shenzhen (South), Chengdu (West), and Wuhan (Center)–are chosen for spatial comparison analysis with annual grid electricity data (AGED) created by Chen *et al.*[22].

Data Records

The study estimated high-resolution total electricity consumption data for 280 major Chinese cities based on multi-source data availability, which account for 90.6% of China's electricity consumption (https://www.stats.gov.cn/). The dataset is stored in Geotiff (.tif) format in the folder "China_1km_Ele_201204_201912.zip" and spatially projected using the Albers equal area method. The folder contains 93 .tif files, each labeled with the year and month, describing the monthly electricity consumption in China. Cities details are also provided in the folder in .csv format. The dataset[48] is publicly available for free on Figshare (https://doi.org/10.6084/m9.figshare.25398559.v1).

Technical Validation

The technical validation of this study encompasses three main parts: (1) analysis of the correlation between independent and dependent variables ; (2) assessment of the model's performance; and (3) comparative analysis between our dataset and existing related datasets; (4) analysis of the spatio-temporal patterns of high-resolution electricity consumption in China.

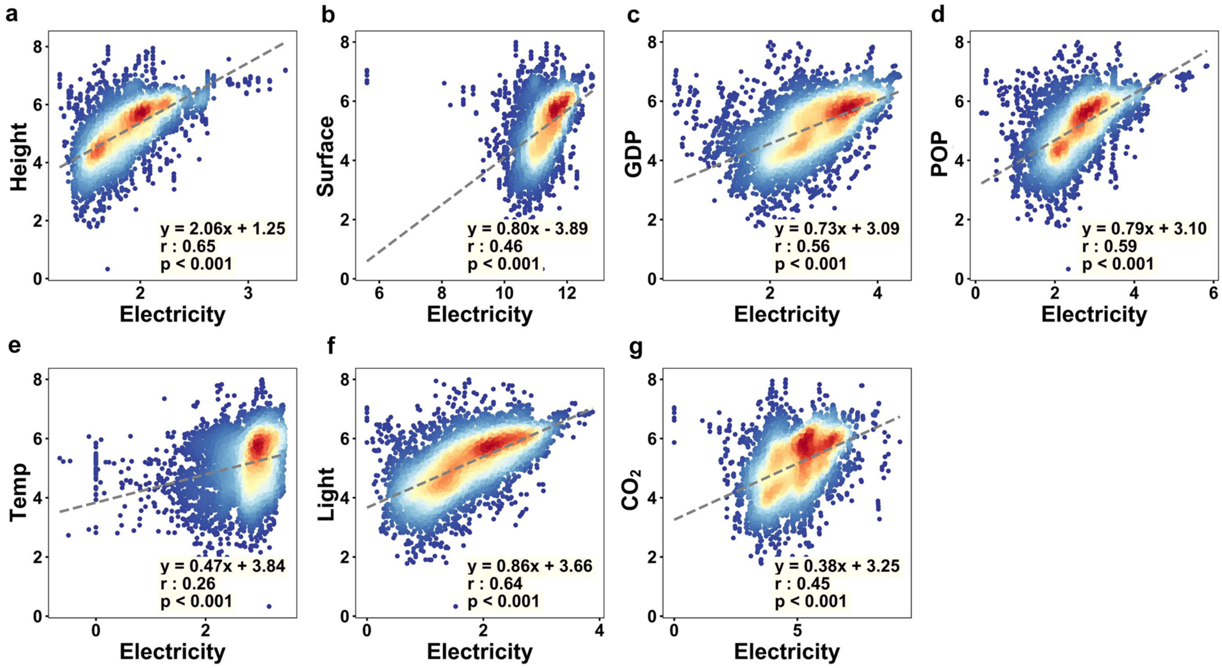

Fig. 2 Correlation analysis of electricity consumption with (**a**) Building height, (**b**) Building surface, (**c**) GDP, (**d**) POP, (**e**) Temperature, (**f**) Nighttime lights, and (**g**) CO_2.

Method	R^2_{SD}	R^2_{TD}	F_{R^2}	$RMSE_{SD}$	$RMSE_{TD}$	F_{RMSE}
XGBoost	0.678	0.706	0.690	239.561	137.072	174.371
XGBoost-SIL	0.895	0.667	0.763	54.777	134.864	77.909
XGBoost-STIL	0.905	0.919	0.911	57.863	62.484	60.084

Table 2. Performance of temporal-test dataset (TD) and spatial-test dataset (SD) in different methods (average results of five-fold cross-validation).

Variable correlation analysis. Firstly, the correlation between the independent variables and electricity consumption is analyzed to provide a solid foundation for accurate estimation of electricity consumption. As shown in Fig. 2, the results indicated that all independent variables have statistically significant correlations with the dependent variable, with p-values less than 0.001 and an average correlation coefficient of 0.52. Building height (0.65) and nighttime lights (0.64) demonstrated the strongest correlations with electricity consumption, underscoring the critical role of urbanization and economic activities in electricity demand. The correlation coefficients for GDP, POP, building surface, and CO_2 fall within the range of 0.45 to 0.6, signifying the considerable influence of economic development, demographics, urban configuration, and environmental factors on the patterns of electricity consumption. Although the correlation between temperature and electricity consumption was lower (0.26) than others, the control variable experiments (Supplementary Table S1) have verified that temperature can further improve accuracy, which may be attributed to the effect of temperature in specific events such as summer cooling. Furthermore, controlled variable experiments were conducted to verify the validity of each variable, as detailed in the Supplementary Information.

Model performance analysis. Table 2 shows the performance of the models based on machine learning and incremental learning in this study. The baseline model XGBoost achieved R^2 of 0.678 and RMSE of 239.561 on the spatial dataset, and R^2 of 0.706 and RMSE of 137.072 on the temporal dataset. Additionally, the F_{R^2} and F_{RMSE} were 0.690 and 174.371, respectively. After integrating spatial incremental learning (XGBoost-SIL), the performance of the spatial dataset is significantly improved with R^2 and F_{R^2} increasing to 0.895 and 0.763, while F_{RMSE} decreases to 77.909. Based on this, the XGBoost-STIL model performance is optimized by further integrating temporal incremental learning, both datasets improved the R^2 to above 0.9, while the RMSE was reduced to around 60. The comprehensive enhancement is further demonstrated by F_{R^2} of 0.911 and F_{RMSE} of 60.084, highlighting the model's improved ability to accurately capture complex electricity consumption patterns across diverse datasets. These improvements underscore the significant impact of integrating spatial and temporal incremental learning, offering a robust framework that outperforms traditional methodologies.

Dataset validation. Further validation and comparisons were conducted using official statistics and existing datasets from quantitative and qualitative perspectives, respectively. In the absence of electricity data at same

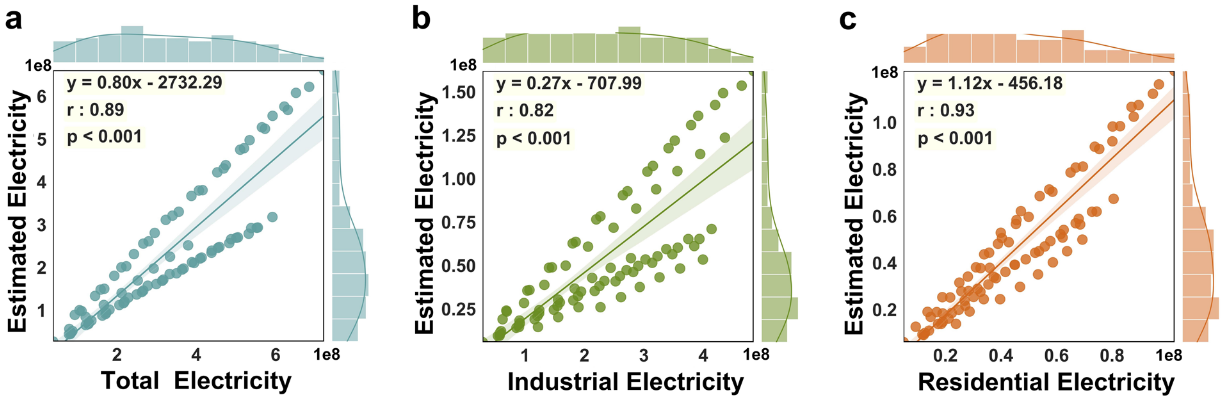

Fig. 3 Correlation analysis of downscaling results with official statistics on (**a**) total electricity consumption, (**b**) industrial electricity consumption, and (**c**) residential electricity consumption.

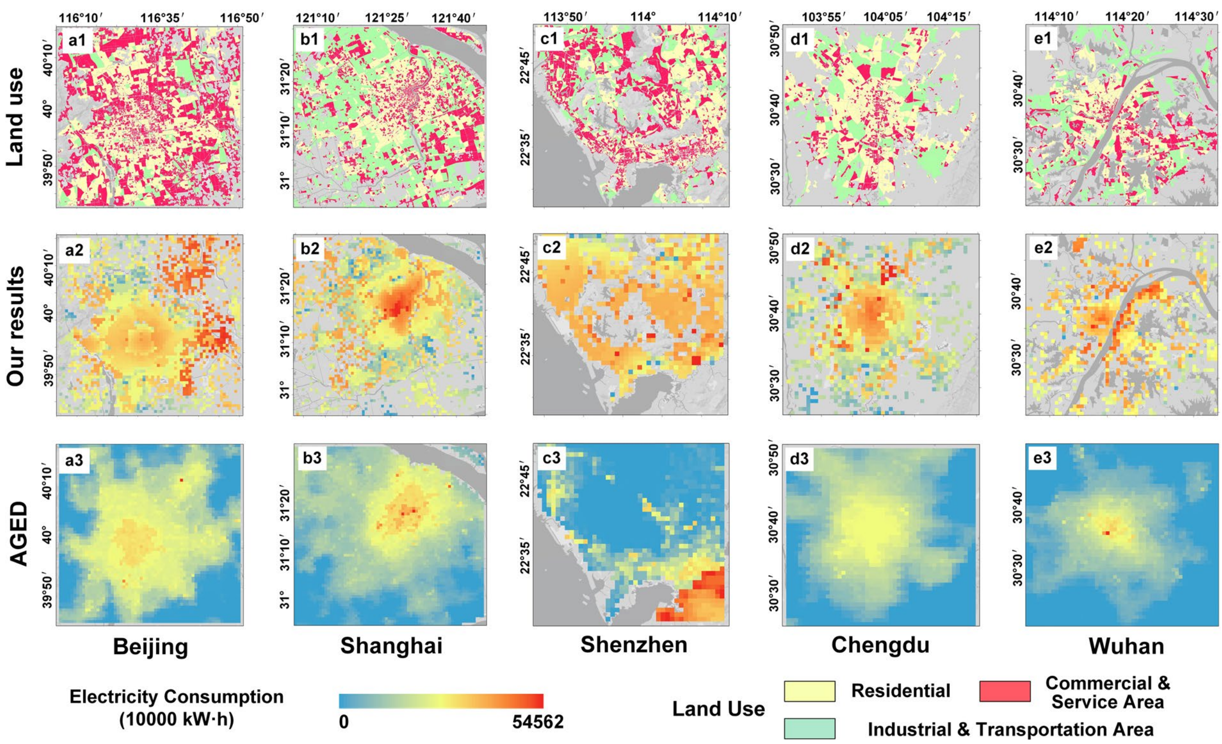

Fig. 4 Distribution patterns of electricity consumption from this study and AGED in different urban land use: Beijing (China's Capital, Northern), Shanghai (International economic center, Eastern), Shenzhen (China's first special economic zone, Southern), Wuhan (Central Transportation Hub) and Chengdu (Western China's leading city).

resolutions, we used the total, residential and industrial electricity consumption of the country at monthly periods for a quantitative correlation analysis. Subsequently, we conducted a comparative validation with the AGED to evaluate our dataset's reliability. Fig. 3 shows the correlation of our results with official statistics for validation. The correlation is 0.89 for total electricity consumption, 0.82 for industrial electricity consumption, and 0.93 for residential electricity consumption, with all p-values less than 0.001. These results confirm the model's effectiveness in accurately reflecting actual electricity consumption patterns across different sectors. Such statistically significant correlations affirm the robustness of our dataset when compared with established benchmarks, providing a solid foundation for its application in energy research and policy development.

We further compared with the AGED in five large cities in different regions by incorporating land use data[49]. As shown in Fig. 4, we observed that AGED displayed a flat distribution, failing to distinguish adequately between consumption patterns across different regions. A critical shortfall of their dataset is the inability to differentiate between built-up and non-built-up areas, mistakenly attributing electricity demand to non-built-up areas like vegetation and water bodies. In addition, their methodology lacks correction in conjunction with actual official statistics, which would lead to errors. In contrast, our data can effectively capture

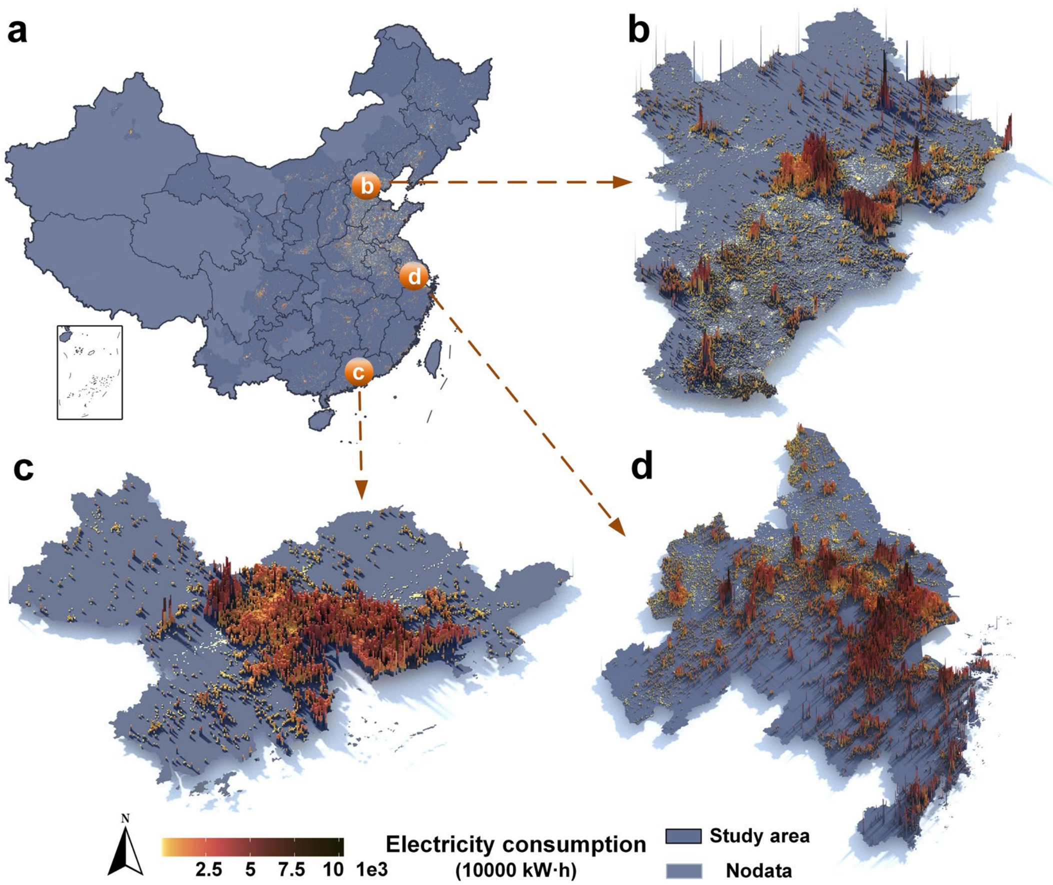

Fig. 5 December 2019 distribution patterns of electricity consumption in China's 1 km (**a**) and typical urban agglomerations: (**b**) Beijing-Tianjin-Hebei, (**c**) Pearl River Delta, and (**d**) Yangtze River Delta.

the electricity consumption patterns of different land use types, and avoid incorrectly estimating electricity use on non-built-up zones. By incorporating kriging interpolation, our method corrects estimations, and capture spatial heterogeneity across high-resolution grids and ensuring our electricity results are precise.

This study also reveals the diversity of electricity consumption patterns in various functional zones (e.g., residential, industrial, and commercial zones) within the city. Take shanghai as an example, which has the highest China's GDP in 2019. The high electricity demand areas are mainly located in downtown Shanghai, which includes the city's central business district (CBD) and various commercial centers. The prosperity of these areas directly influences their substantial electricity demands. Similarly, Shenzhen, known for its high-tech industries, experiences uniformly high levels of electricity consumption across the city. This is particularly pronounced in industrial zones and coastal logistics hubs, reflecting the city's vibrant industrial production and international trade activities. The areas in Wuhan with high electricity demand are mainly found along the Yangtze River, which is the central hub of the city with clusters of Grade A office buildings. There is also high electricity consumption in the northwest, primarily driven by the airport and industrial areas. These high-resolution analyses of electricity consumption patterns provide an insight into urban energy consumption disparities, which can help optimize the allocation of energy resources.

Spatio-temporal patterns of electricity consumption in China. In this study, we estimated the electricity consumption of 1 km × 1 km grid from April 2012 to December 2019. December 2019 was chosen to visualize high-resolution electricity distribution patterns in China (Fig. 5). The highly concentrated pattern of electricity consumption in the North China Plain reflects a dense population with a thriving service and manufacturing industries. Northeastern China, despite economic restructuring, shows a medium density of hotspots as a traditional industrial area. The central and southern regions have a dispersed pattern of electricity consumption due to terrain.

Additionally, the results also show that high electricity consumption patterns are concentrated in the Beijing-Tianjin-Hebei (BTH), Yangtze River Delta (YRD) and Pearl River Delta (PRD) urban agglomerations. The electricity demand in these areas not only reflects their advanced levels of industrialization and urbanization but also their pivotal role in the national economy. The BTH as a hub of political and cultural significance in China, with key industries such as government services, finance, and information technology creating high-energy consumption patterns. The YRD and the PRD, as the centers of China's manufacturing and export sectors, have high electricity consumption pattern, highlighting the concentration of industrial activity and substantial energy needs.

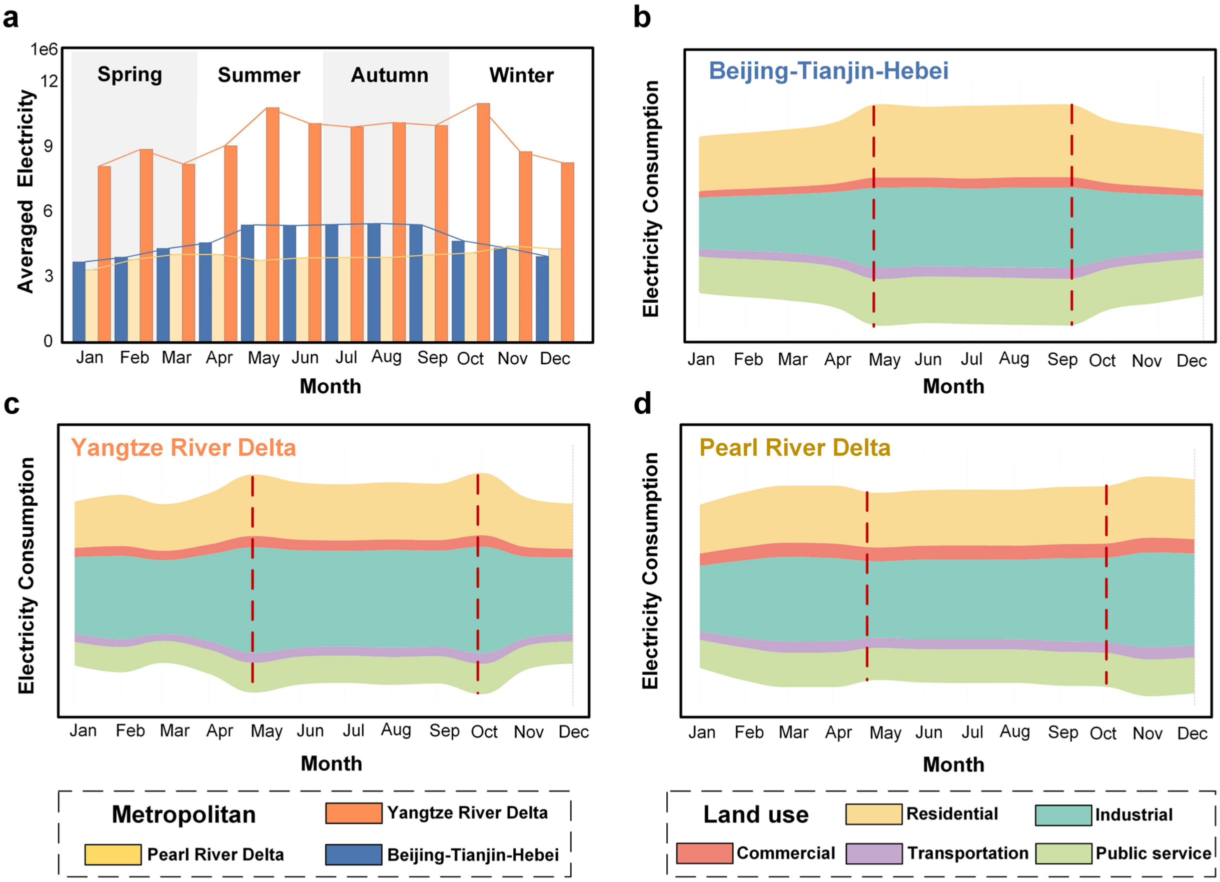

Fig. 6 Monthly electricity consumption in the three urban agglomerations in 2019, as well as details in various land use types. (**a**) represents the monthly electricity consumption of the three urban agglomerations. The YRD's electricity consumption outstrips that of the PRD and BTH, a trend driven by its superior GDP of 29.03 trillion yuan, driven by Shanghai and three prosperous provinces, compared to the PRD's 7.8 trillion yuan and BTH's 6.9 trillion yuan. (**b**,**d**) Represent the monthly electricity consumption of the three urban agglomerations under different land use types in streamgraph.

In terms of temporal dynamics, our meticulous monthly data analysis has captured the seasonal fluctuations and trend variations in electricity consumption across the three urban agglomerations, as shown in Fig. 6. BTH and YRD, with an increase from May to October and a subsequent decrease, may reflect the impact of climatic variations on electricity demand. In contrast, the PRD demonstrates a stable monthly electricity consumption trend, a discrepancy that may be attributed to the distinct industrial structures of each region. Temporal patterns of electricity consumption were further analyzed with land use data. Industrial areas recorded the highest proportion of electricity consumption, accounting for 43.2%, indicating that the industrial production has a high demand for electricity throughout the year. In particular, residential electricity consumption shows seasonal variations, especially during the summer peak season. Commercial and transport facility areas have relatively low electricity throughout the year and have no significant seasonal fluctuations.

This study creates a high spatio-temporal resolution electricity data for China, effectively filling an important data gap. The dataset reveals the intricate dynamics of electricity consumption, providing a reliable data support for sustainable development research. Future studies can use this data to explore diverse energy scenarios, optimize prediction models, and formulate strategies to shift the world toward a more sustainable and efficient energy future.

Uncertainties and limitations. There are several aspects of uncertainties in this study. Firstly, we mainly use socio-economic and environmental variables to estimate electricity consumption, without fully considering geographic factors. This may limit the model's ability to comprehensively capture the electricity consumption patterns in diverse regions, such as the differences of electricity consumption in southern and northern China due to heating and cooling demands. Northern cities have higher heating demands in winter, while southern cities have higher cooling demands in summer[50]. Although our model considers temperature data, it cannot directly reflect these seasonal differences. Future study should consider more geography-related variables, such as Heating Degree Days (HDD) and Cooling Degree Days (CDD)[51]. In addition, regional modeling can be performed based on climate zones to reduce geographic uncertainty and improve model accuracy.

The input dataset uncertainty also challenged this study. Although we considered variables with full-coverage and availability as much as possible, there are also some relevant data not included. For example, we combined land use data in analysis but without integrating it into the downscaling model, which could improve the results[52]. Energy prices and types should also be considered. Moreover, the spatio-temporal differences in the original variables (e.g., the GHSL data spans 5 years) may affect the results. However, the unavailability of spatio-temporal datasets limits the integration of these data in this study. Currently, our dataset covers 2012 to 2019 at the 1 km $\times$ 1 km scale. In the future, we will continue to focus on the availability of relevant data, optimize our approach by incorporating more valuable data and dynamically update the spatio-temporal scales of the dataset.

Code availability

The software used to create the dataset were ArcGIS (10.2), Python 3.8, and R 4.3.2. The code is available on GitHub (https://github.com/kkxiaoqin/electricity_downscaling).

Received: 20 March 2024; Accepted: 25 July 2024;
Published online: 13 August 2024

References

1. Bouckaert, S. *et al*. Net zero by 2050: A roadmap for the global energy sector. International Energy Agency (2021).
2. International Energy Agency (IEA). Electricity market report 2023. https://www.iea.org/reports/electricity-market-report-2023 (2023).
3. Chapman, A. J. & Itaoka, K. Energy transition to a future low-carbon energy society in Japan's liberalizing electricity market: Precedents, policies and factors of successful transition. *Renewable and Sustainable Energy Reviews* **81**, 2019–2027 (2018).
4. Bogdanov, D. *et al*. Low-cost renewable electricity as the key driver of the global energy transition towards sustainability. *Energy* **227**, 120467 (2021).
5. European Commission. National energy and climate plans. https://ec.europa.eu/energy/topics/energy-strategy/national-energy-climate-plans_en (2020).
6. Guo, J., Ma, J., Li, Z. & Hong, J. Building a top-down method based on machine learning for evaluating energy intensity at a fine scale. *Energy* **255**, 124505 (2022).
7. Chen, M. *et al*. Fine-scale population spatialization data of China in 2018 based on real location-based big data. *Scientific Data* **9**, 624 (2022).
8. da Silva, F. L., Oliveira, F. L. C. & Souza, R. C. A bottom-up bayesian extension for long term electricity consumption forecasting. *Energy* **167**, 198–210 (2019).
9. Yan, Y. *et al*. A factor-based bottom-up approach for the long-term electricity consumption estimation in the Japanese residential sector. *Journal of Environmental Management* **270**, 110750 (2020).
10. Bao, Y. *et al*. High-resolution quantification of building stock using multi-source remote sensing imagery and deep learning. *J. Ind. Ecol.* **27**, 350-361 (2023).
11. Panão, M. J. O. & Brito, M. C. Modelling aggregate hourly electricity consumption based on bottom-up building stock. *Energy and Buildings* **170**, 170–182 (2018).
12. Wiesmann, D., Azevedo, I. L., Ferrão, P. & Fernández, J. E. Residential electricity consumption in portugal: Findings from top-down and bottom-up models. *Energy policy* **39**, 2772–2779 (2011).
13. Gurney, K. R. *et al*. Comparison of global downscaled versus bottom-up fossil fuel CO_2 emissions at the urban scale in four us urban areas. *Journal of Geophysical Research: Atmospheres* **124**, 2823–2840 (2019).
14. Bao, Y. *et al*. High-resolution mapping of material stocks in the built environment across 50 Chinese cities. *Resour. Conserv. Recycl.* **199**, 107232 (2023).
15. Jiang, Y. *et al*. Local-global dual attention network (LGANet) for population estimation using remote sensing imagery. *Resour. Environ. Sustain*. **14**, 100136 (2023).
16. Deville, P. *et al*. Dynamic population mapping using mobile phone data. *Proceedings of the National Academy of Sciences* **111**, 15888–15893 (2014).
17. Shi, K. *et al*. Detecting spatiotemporal dynamics of global electric power consumption using DMSP-OLS nighttime stable light data. *Applied Energy* **184**, 450–463 (2016).
18. Shi, K. *et al*. Evaluating spatiotemporal patterns of urban electricity consumption within different spatial boundaries: A case study of Chongqing, China. *Energy* **167**, 641–653 (2019).
19. Wang, J. & Lu, F. Modeling the electricity consumption by combining land use types and landscape patterns with nighttime light imagery. *Energy* **234**, 121305 (2021).
20. Sun, Y., Wang, S., Zhang, X., Chan, T. O. & Wu, W. Estimating local-scale domestic electricity energy consumption using demographic, nighttime light imagery and twitter data. *Energy* **226**, 120351 (2021).
21. Zhou, X. *et al*. High-resolution estimation of building energy consumption at the city level. *Energy* **275**, 127476 (2023).
22. Chen, J. *et al*. Global 1 km $\times$ 1 km gridded revised real gross domestic product and electricity consumption during 1992–2019 based on calibrated nighttime light data. *Scientific Data* **9**, 202 (2022).
23. Deb, C., Zhang, F., Yang, J., Lee, S. E. & Shah, K. W. A review on time series forecasting techniques for building energy consumption. *Renewable and Sustainable Energy Reviews* **74**, 902–924 (2017).
24. Su, B. & Ang, B. Structural decomposition analysis applied to energy and emissions: Frameworks for monthly data. *Energy Economics* **126**, 106977 (2023).
25. Hengl, T., Nussbaum, M., Wright, M. N., Heuvelink, G. B. & Gräler, B. Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ* **6**, e5518 (2018).
26. Ye, R., Huang, Z., Li, L. & Shan, X. GeoUNet: A novel AI model for high-resolution mapping of ecological footprint. *International Journal of Applied Earth Observation and Geoinformation* **112**, 102803 (2022).
27. Janowicz, K., Gao, S., McKenzie, G., Hu, Y. & Bhaduri, B. GeoAI: spatially explicit artificial intelligence techniques for geographic knowledge discovery and beyond (2020).
28. Liu, P. & Biljecki, F. A review of spatially-explicit GeoAI applications in urban geography. *International Journal of Applied Earth Observation and Geoinformation* **112**, 102936 (2022).
29. Zhu, A.-X., Lu, G., Liu, J., Qin, C.-Z. & Zhou, C. Spatial prediction based on third law of geography. *Annals of GIS* **24**, 225–240 (2018).
30. Meng, Q., Liu, Z. & Borders, B. E. Assessment of regression kriging for spatial interpolation–comparisons of seven GIS interpolation methods. *Cartography and Geographic Information Science* **40**, 28–39 (2013).
31. Shukla, K., Kumar, P., Mann, G. S. & Khare, M. Mapping spatial distribution of particulate matter using kriging and inverse distance weighting at supersites of megacity delhi. *Sustainable Cities and Society* **54**, 101997 (2020).
32. Liu, X., Kyriakidis, P. C. & Goodchild, M. F. Population-density estimation using regression and area-to-point residual kriging. *International Journal of Geographical Information Science* **22**, 431–447 (2008).

33. Chen, Y., Zhang, R., Ge, Y., Jin, Y. & Xia, Z. Downscaling census data for gridded population mapping with geographically weighted area-to-point regression kriging. *IEEE Access* **7**, 149132–149141 (2019).
34. Ma, X. *et al.* A regional spatiotemporal downscaling method for CO_2 columns. *IEEE Transactions on Geoscience and Remote Sensing* **59**, 8084–8093 (2021).
35. Chen, C., Hu, B. & Li, Y. Easy-to-use spatial random-forest-based downscaling-calibration method for producing precipitation data with high resolution and high accuracy. *Hydrology and Earth System Sciences* **25**, 5667–5682 (2021).
36. Wu, T. & Li, Y. Spatial interpolation of temperature in the united states using residual kriging. *Applied Geography* **44**, 112–120 (2013).
37. Shtiliyanova, A. *et al.* Kriging-based approach to predict missing air temperature data. *Computers and Electronics in Agriculture* **142**, 440–449 (2017).
38. Cheng, Z., Wang, J. & Ge, Y. Mapping monthly population distribution and variation at 1-km resolution across China. *International Journal of Geographical Information Science* **36**, 1166–1184 (2022).
39. Bessec, M. & Fouquau, J. The non-linear link between electricity consumption and temperature in Europe: A threshold panel approach. *Energy Economics* **30**, 2705–2721 (2008).
40. Huebner, G., Shipworth, D., Hamilton, I., Chalabi, Z. & Oreszczyn, T. Understanding electricity consumption: A comparative contribution of building factors, socio-demographics, appliances, behaviours and attitudes. *Applied Energy* **177**, 692–702 (2016).
41. Lin, B., Omoju, O. E. & Okonkwo, J. U. Factors influencing renewable electricity consumption in China. *Renewable and Sustainable Energy Reviews* **55**, 687–696 (2016).
42. Li, X., Zhou, Y., Zhao, M. & Zhao, X. A harmonized global nighttime light dataset 1992–2018. *Scientific Data* **7**, 168 (2020).
43. Ji, S., Lee, B. & Yi, M. Y. Building life-span prediction for life cycle assessment and life cycle cost using machine learning: A big data approach. *Building and Environment* **205**, 108267 (2021).
44. Xu, X. *et al.* China's multi-period land use land cover remote sensing monitoring data set (CNLUCC). *Resource and Environment Data Cloud Platform: Beijing, China* (2018).
45. Wu, Z. *et al.* CEDUP: Using incremental learning modeling to explore spatio-temporal carbon emission distribution and unearthed patterns at the municipal level. *Resources, Conservation and Recycling* **193**, 106980 (2023).
46. Baynes, J., Neale, A. & Hultgren, T. Improving intelligent dasymetric mapping population density estimates at 30m resolution for the conterminous united states by excluding uninhabited areas. *Earth System Science Data* **14**, 2833–2849 (2022).
47. Wang, Q., Shi, W., Atkinson, P. M. & Zhao, Y. Downscaling modis images with area-to-point regression kriging. *Remote Sensing of Environment* **166**, 191–204 (2015).
48. Yan, X. & Huang, Z. Monthly electricity consumption data at 1 km × 1 km spatial resolution for 280 cities in China from 2012 to 2019. *figshare* https://doi.org/10.6084/m9.figshare.25398559.v1 (2024).
49. Gong, P. *et al.* Mapping essential urban land use categories in China (EULUC-China): Preliminary results for 2018. *Science Bulletin* **65**, 182–187 (2020).
50. Fan, J.-L., Zeng, B., Hu, J.-W., Zhang, X. & Wang, H. The impact of climate change on residential energy consumption in urban and rural divided southern and northern China. *Environmental Geochemistry and Health* **42**, 969–985 (2020).
51. De Rosa, M., Bianco, V., Scarpa, F. & Tagliafico, L. A. Heating and cooling building energy demand evaluation; a simplified model and a modified degree days approach. *Applied Energy* **128**, 217–229 (2014).
52. Yao, Y. *et al.* Classifying land-use patterns by integrating time-series electricity data and high-spatial resolution remote sensing imagery. *International Journal of Applied Earth Observation and Geoinformation* **106**, 102664 (2022).
53. Peng, S. 1-km monthly mean temperature dataset for China (1901–2017). *National Tibetan Plateau Data Center: Beijing, China* (2019).
54. Chen, J. D. & Gao, M. Global 1 km × 1 km gridded revised real gross domestic product and electricity consumption during 1992–2019 based on calibrated nighttime light data. *Figshare* https://doi.org/10.6084/m9.figshare.17004523.v1 (2021).

Acknowledgements
This work was supported by the National Key Research and Development Program of China (Grant No. 2023YFB3906802).

Author contributions
X.Y.: Data processing, conducting experiments, drawing figures and tables, writing original manuscript. Z.H.: Supervision, scientific discussion, manuscript revision, funding support. S.R.: Results analysis, graph optimization, manuscript revision. G.Y.: Results analysis, scientific discussion. J.Q.: Data collection, graph optimization.

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/s41597-024-03684-4.

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

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.