



OPEN

DATA DESCRIPTOR

City- and county-level spatio-temporal energy consumption and efficiency datasets for China from 1997 to 2017

Jiandong Chen¹, Jialu Liu¹, Jie Qi¹, Ming Gao¹ , Shulei Cheng¹ ✉, Ke Li² & Chong Xu¹

Understanding the evolution of energy consumption and efficiency in China would contribute to assessing the effectiveness of the government's energy policies and the feasibility of meeting its international commitments. However, sub-national energy consumption and efficiency data have not been published for China, hindering the identification of drivers of differences in energy consumption and efficiency, and implementation of differentiated energy policies between cities and counties. This study estimated the energy consumption of 336 cities and 2,735 counties in China by combining Defense Meteorological Satellite Program/Operational Line-scan System (DMSP/OLS) and Suomi National Polar-Orbiting Partnership/Visible Infrared Imaging Radiometer Suite (NPP/VIIRS) satellite nighttime light data using particle swarm optimization-back propagation (PSO-BP). The energy efficiency of these cities and counties was measured using energy consumption per unit GDP and data envelopment analysis (DEA). These data can facilitate further research on energy consumption and efficiency issues at the city and county levels in China. The developed estimation methods can also be used in other developing countries and regions where official energy statistics are limited.

Background & Summary

China has been the world's highest energy consumer since 2007, followed by the United States^{1–3}. Optimizing the energy mix and improving energy efficiency is crucial for China itself and the world response to climate change. Accurate measurement of China's energy consumption and efficiency, particularly at the sub-national level, is essential to accelerate the achievement of the energy transition and energy efficiency improvement goals.

China's national energy consumption data is published by the Chinese National Bureau of Statistics (CNBS)^{4–6}, International Energy Agency (IEA)^{7–9}, British Petroleum (BP)^{10–12}, Energy Information Administration (EIA)^{13–15}, World Bank^{16,17}, and other international organizations, while China's provincial-level energy consumption data is mainly published by the CNBS^{18–20}. These widely accepted energy consumption data have substantially contributed to research on international energy consumption comparisons^{13,15,21}, overall and inter-provincial energy consumption drivers^{22–24}, and socioeconomic effects of overall and inter-provincial energy consumption in China^{5,25,26}.

The availability of China's national and provincial-level energy consumption data also indirectly provides key support for energy efficiency studies. Energy efficiency not only reflects the coupling relationship between energy consumption and economic development, but also implies the energy policy appeal to maximize economic output based on energy input. To deal with climate change under the sustainable development goals, it is necessary to drive sustained economic development with as little energy consumption as possible. The measurement of energy efficiency can be classified into three categories:

First, the indicator measurement analysis (IMA) is used to measure energy efficiency by energy consumption per unit GDP⁴. The advantages of this measurement are mainly reflected in its low requirements for other socioeconomic data, simple calculation and easy to be regarded as a policy target¹⁷. Most countries, including China,

¹School of Public Administration, Southwestern University of Finance and Economics, Chengdu, Sichuan, 611130, China.

²School of Statistics, Southwestern University of Finance and Economics, Chengdu, Sichuan, 611130, China. ✉e-mail: chengsl@swufe.edu.cn

Satellite	Year	d	e	R^2
F10	1992	1.158	1.17	0.7955
F10	1993	2.105	0.896	0.8103
F10	1994	1.264	1.127	0.9093
F12	1994	3.074	0.765	0.8909
F12	1995	1.300	1.073	0.9151
F12	1996	1.491	1.029	0.9018
F12	1997	0.947	1.192	0.8952
F12	1998	0.884	1.167	0.8220
F12	1999	1.473	1.008	0.8657
F14	1997	2.264	0.97	0.8619
F14	1998	1.485	1.077	0.8629
F14	1999	1.454	1.137	0.8395
F14	2000	1.282	1.108	0.8628
F14	2001	1.179	1.129	0.8978
F14	2002	1.598	0.987	0.9242
F14	2003	1.394	1.038	0.8729
F15	2000	1.196	1.075	0.8514
F15	2001	1.341	1.033	0.8423
F15	2002	0.889	1.151	0.8411
F15	2003	1.824	0.974	0.8427
F15	2004	1.236	1.07	0.8838
F15	2005	1.764	0.937	0.8564
F15	2006	1.225	1.061	0.9066
F15	2007	1.324	1.067	0.9103
F16	2004	0.807	1.143	0.8790
F16	2005	1.058	1.119	0.9191
F16	2006	1.005	1.117	0.9352
F16	2007	0.965	1.109	0.8911
F16	2008	0.855	1.108	0.9480
F16	2009	0.543	1.175	0.9176
F18	2010	0.273	1.304	0.8914
F18	2011	0.694	1.09	0.8351
F18	2012	0.474	1.179	0.8743
F18	2013	0.418	1.153	0.8501

Table 1. Coefficients of the power function for nighttime light data (RCNTL in 2006 as a reference image). Note: d , e , and R^2 represent the constant term, index, and coefficient of determination, respectively.

have set reducing energy consumption per unit GDP as a government policy target to tackle climate change¹⁷. However, it cannot reveal the improvement path of energy efficiency.

Second, the non-parametric method, in which energy, capital and labor are regarded as inputs, GDP and pollutants are regarded as expected output and undesirable output respectively, is applied to obtain energy efficiency. Energy efficiency means that reducing energy consumption as much as possible in exchange for increased economic output and fewer pollutants²⁴. The closer the actual output increase and pollutant reduction in exchange for the reduction of energy consumption are to the theoretical potential output increase and pollutant reduction, that is, the closer the actual energy performance is to the potential energy performance, the higher the energy efficiency will be. The data envelopment analysis (DEA) is the most widely used non-parametric method²⁴. The advantages of this measurement are mainly reflected in the convenience of locking the drivers of energy efficiency improvement, the ease of comparison of the relative energy efficiency between regions, and the need for no prior setting of the estimation function. However, it not only requires more socioeconomic data except energy consumption, but also is relatively complicated to calculate.

Third, the parameter method, which takes energy, capital and labor as independent variables and GDP as dependent variable, is chosen to estimate energy efficiency²⁴. The stochastic frontier analysis (SFA) is the most widely used parameter method²⁷. However, because of relying on the prior setting of the regression function and requiring more socioeconomic data except energy consumption, this measurement is relatively few adopted.

Although research on energy efficiency has practical values, unlike energy consumption data, neither Chinese officials nor other international organizations have published energy efficiency data. This is mainly because energy efficiency data is the use of energy consumption data. If Chinese officials and other international organizations publish energy consumption data, scholars and institutions can use the above methods to obtain energy efficiency data. However, the widely accepted energy efficiency data only focus on the national

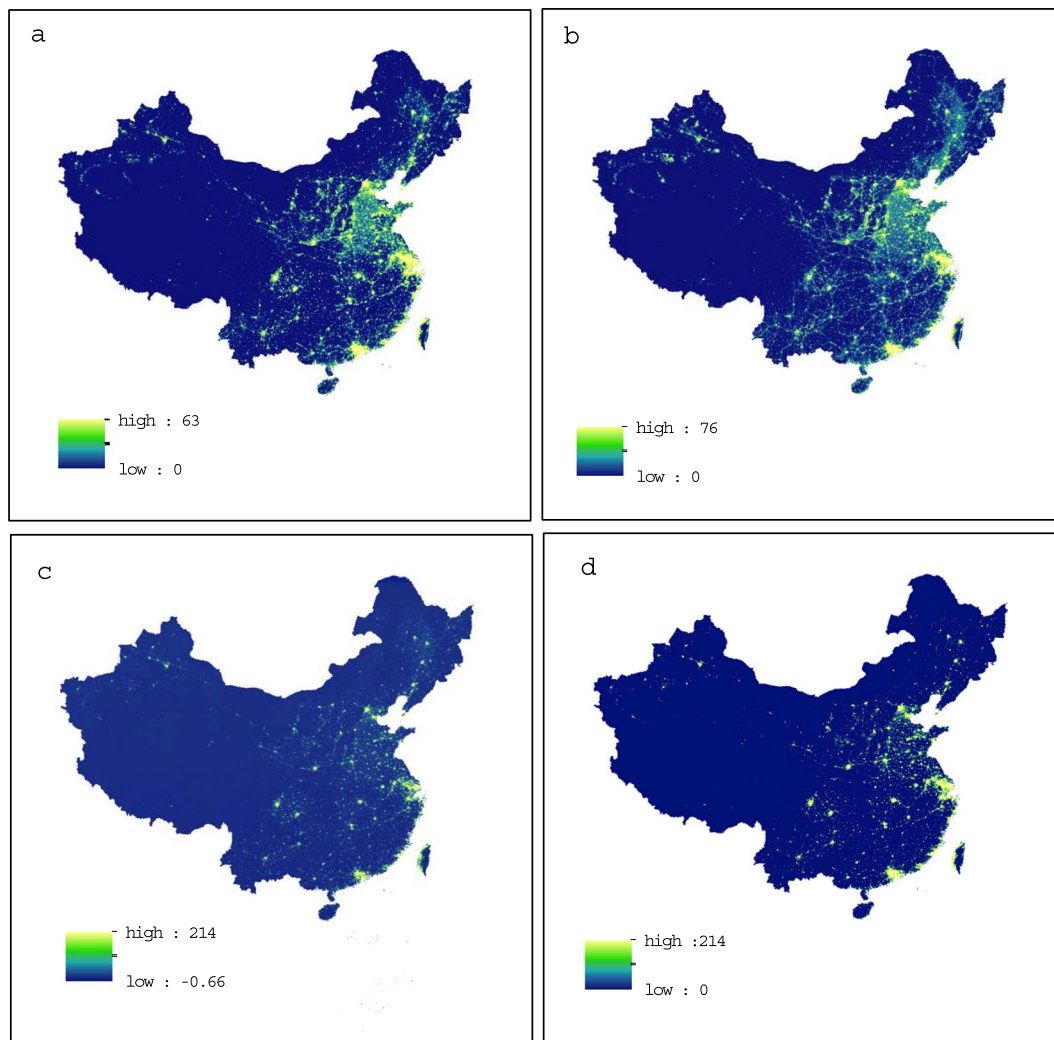


Fig. 1 Differences before and after pre-processing of DMSP/OLS and NPP/VIIRS data. **(a)** The original DMSP/OLS images in 2013 before pre-processing, **(b)** the DMSP/OLS images in 2013 after pre-processing, **(c)** the original NPP/VIIRS images in 2017 before pre-processing, **(d)** the NPP/VIIRS images in 2017 after pre-processing.

and provincial levels, because the energy consumption data published by the CNBS and other international organizations is at national and provincial levels^{24,27}.

Socioeconomic development and energy consumption vary considerably between cities and counties in sub-provincial China^{27–29}. For example, in 2010, energy consumption was approximately 30 times higher in Tangshan than in Suqian³⁰. Meanwhile, even cities with similar scales of energy consumption have different levels of economic development. For example, in 2010, the energy consumption of Wenzhou is similar to that of Jiayuguan, the GDP of the former is nearly 15 times that of the latter³⁰. However, official data are unavailable on city-and-county-level spatio-temporal energy consumption and efficiency in China. This hinders microscopic-level research on the drivers of energy consumption and assessment of energy efficiency in China. Moreover, this limits central and provincial governments in setting differentiated energy efficiency targets and city and county governments in adopting targeted energy efficiency improvement initiatives. Therefore, numerous studies have recently attempted to overcome these research constraints. These studies can be broadly classified into the following three categories.

First, city-level energy consumption data are collated from provincial and city statistical yearbooks^{31–33}. Although the data from these studies are highly reliable as they are published by official agencies, they lack annual energy consumption data for all cities, and the energy types counted by some city statistics departments vary. This is mainly because the lower the government level, the less efficient their data disclosure³⁴ and the weaker the accounting capacity of energy statistics departments, particularly in developing cities.

Second, city-and-county-level energy consumption surveys are conducted to obtain energy consumption data at these levels, and at the enterprise and household levels^{35–38}. The data obtained from these surveys are more microscopic, accurate, and complete; however, annual energy consumption data cannot be obtained for all cities and counties, and the energy consumption status and change characteristics for both cities and counties cannot be directly generalized. This is primarily due to the high labor and financial resources required to

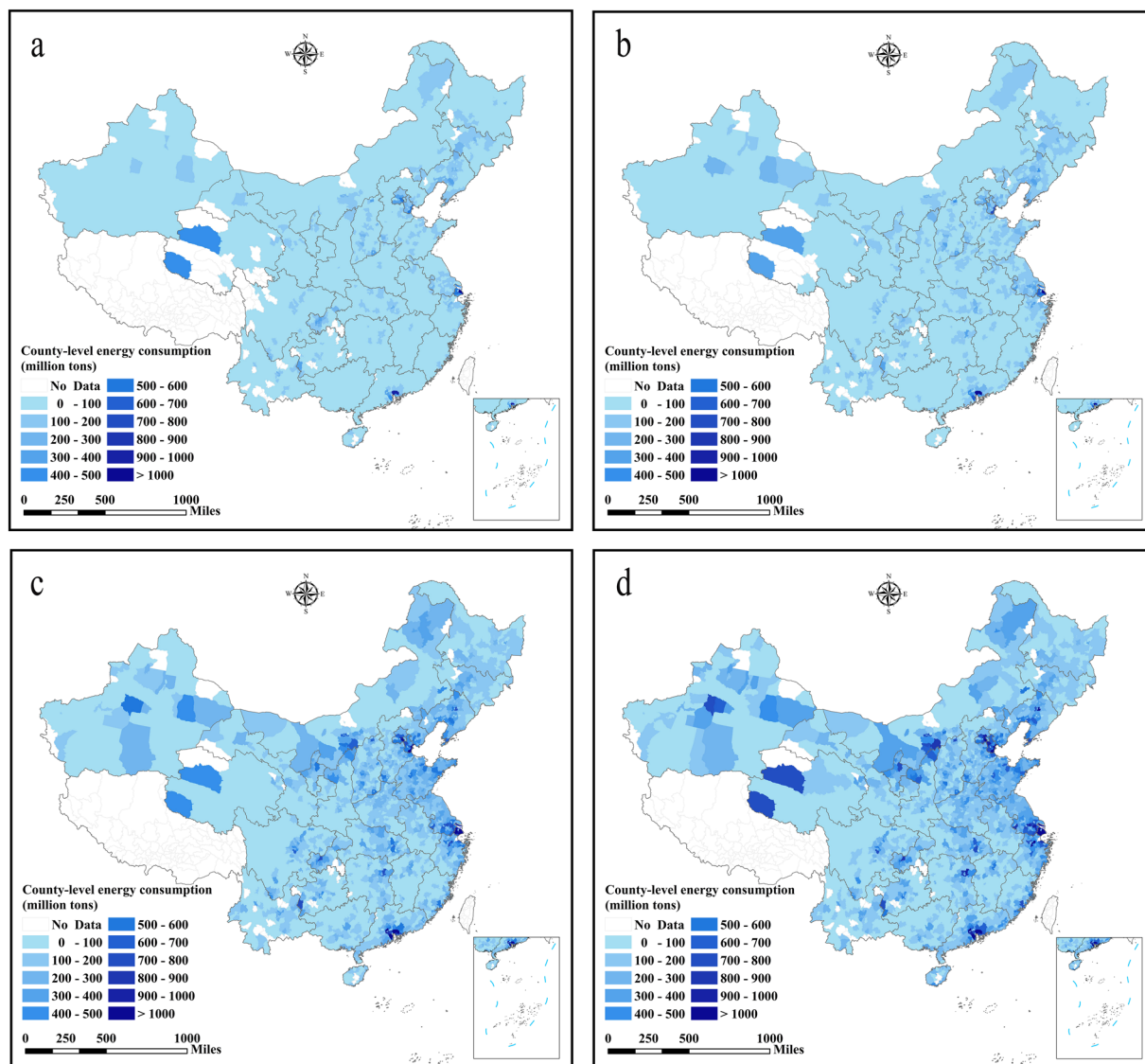


Fig. 2 County-level energy consumption in China (unit: million tons of coal equivalents). (a) The energy consumption in 1997, (b) the energy consumption in 2003, (c) the energy consumption in 2010, (d) the energy consumption in 2017.

conduct these surveys³⁵, the difficulty of including all cities and counties, and the difficulty of conducting annual surveys.

Third, satellite data and machine learning methods are used to determine city-level CO₂ emissions related to energy consumption. For example, Chen *et al.*^{39,40} employed particle swarm optimization-back propagation (PSO-BP) to construct quantitative relationships between provincial nighttime light data and statistical provincial CO₂ emission data in China, and applied top-down data inversion to derive CO₂ emission data for all cities and counties in China, by using the total lighting brightness in these cities and counties as weights. Yue *et al.*⁴¹ used econometric regression and two nighttime light data to obtain a 1-km resolution energy consumption data on the regional scale for China. Although these studies do not provide energy consumption data at the city and county levels, they provide a useful way to obtain annual energy consumption data for cities and counties in China.

Our study filled this gap through inversion of satellite data and machine learning methods to obtain spatio-temporal energy consumption data for 336 cities and 2,735 counties in China from 1997 to 2017 by combining Defense Meteorological Satellite Program/Operational Line-scan System (DMSP/OLS) and Suomi National Polar-Orbiting Partnership/Visible Infrared Imaging Radiometer Suite (NPP/VIIRS) satellite nighttime light datasets using PSO-BP. Based on these energy consumption and related socioeconomic data, the study also provides spatio-temporal energy efficiency data at the city and county levels by using the energy consumption per unit GDP and the ratio of actual to potential energy performance in the non-radial directional distance function (NDDF) derived from DEA. The proposed methodology and datasets can be widely used in energy consumption and efficiency studies at the city and county levels in China, and can provide a reference

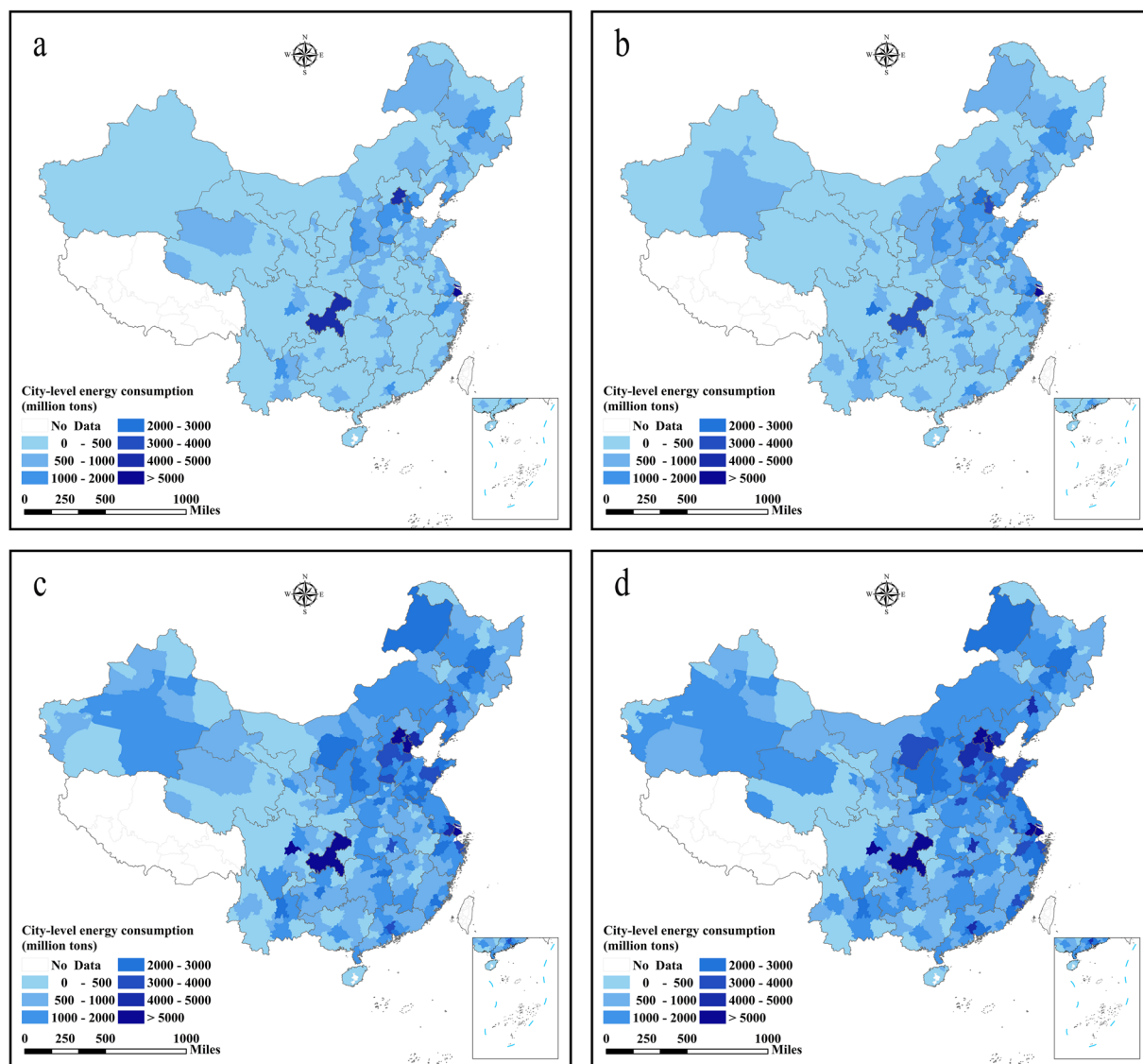


Fig. 3 City-level energy consumption in China (unit: million tons of coal equivalents). (a) The energy consumption in 1997, (b) the energy consumption in 2003, (c) the energy consumption in 2010, (d) the energy consumption in 2017.

for other developing countries and regions with limited energy statistics to analyze their sub-national energy consumption and efficiency.

Methods

Data scopes. In this study, the measured energy consumption and efficiency datasets were based on city and county levels in China. Cities and counties are the grassroots administrative regions in China and the basic units in the implementation of energy policies by the Chinese government^{23,42,43}. Energy consumption at city-and-county-level excludes the production side of the jurisdiction and the energy consumption transferred across administrative regions.

Our study provides data on energy consumption for 336 cities (including 332 prefecture-level cities and 4 municipalities) and 2,735 counties in mainland China from 1997 to 2017 (excluding Hong Kong, Macau, Taiwan, and Tibet) based on the accessibility of the identifiable satellite nighttime light data with a certain resolution, which covers approximately 87% of the land area, over 90% of the population, and 90% of the GDP. The definitions of cities and counties are published in 2010 by the Ministry of Civil Affairs of China. Considering the names of some cities and counties have changed after 2010, our study used their names of 2010 to keep consistency. More detailed explanation was shown in the data files⁴⁴.

Our study also provides two energy efficiency datasets: First, energy efficiency data was calculated by energy consumption per unit GDP, which covered 336 cities and 2,489 counties (1997–2017). It includes fewer cities and counties in some years, due to the lack of GDP data. Second, energy efficiency data was calculated by the ratio of actual to potential energy performance (also known as energy efficiency performance index, EEPI) using

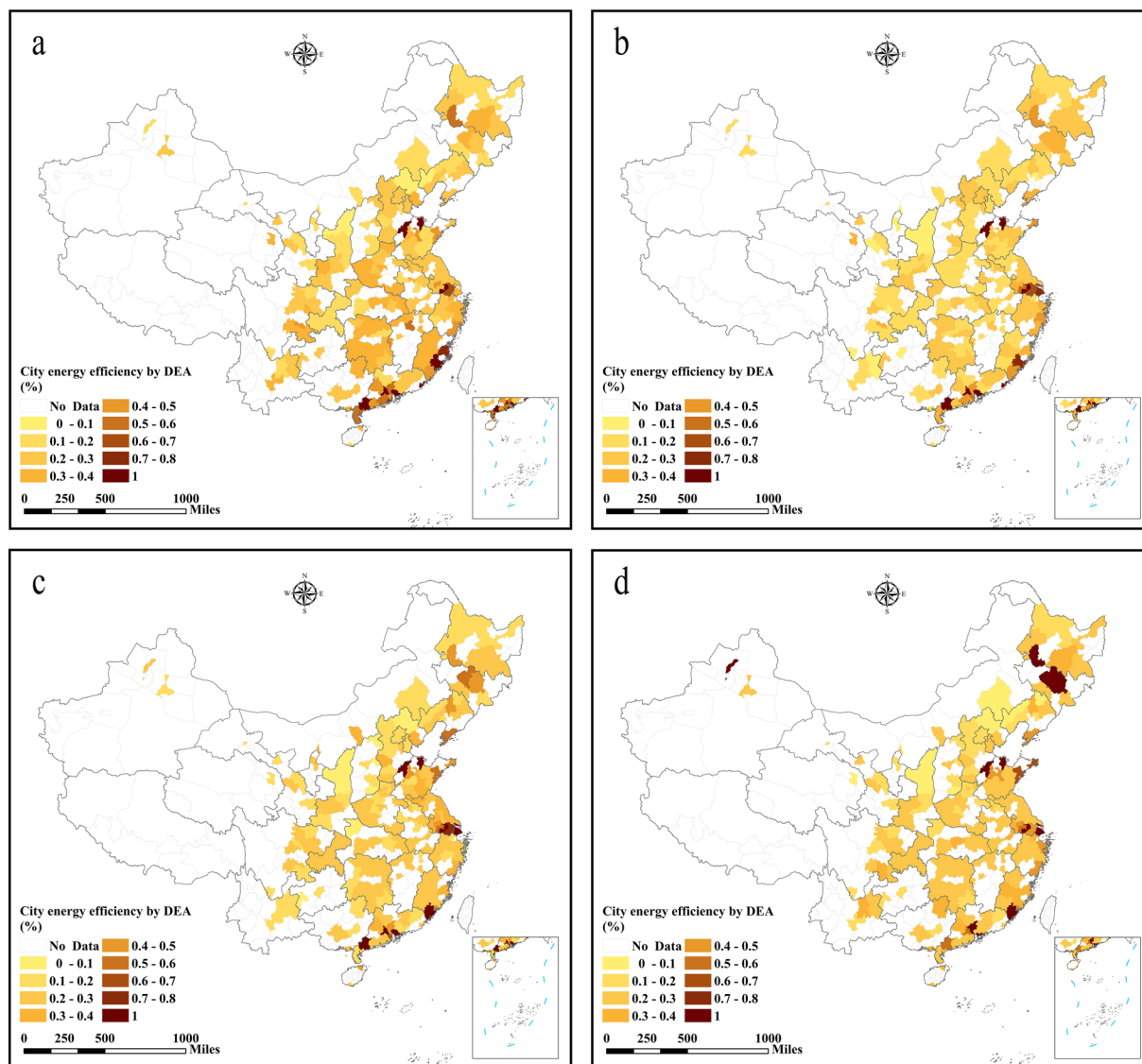


Fig. 4 City-level energy efficiency in China obtained using DEA. (a) The energy efficiency in 2003, (b) the energy efficiency in 2007, (c) the energy efficiency in 2012, (d) the energy efficiency in 2016.

the NDDF of DEA, which covered 189 cities (2003–2016). It includes fewer years and cities, and none counties, mainly due to the lack of basic socioeconomic data considered as input variables of the NDDF, for example, fixed asset investment and fixed asset investment price index.

Matching nighttime light data from two types of satellites. The city-and-county-level energy consumption data in this dataset were inferred from two satellite nighttime light datasets: the DMSP/OLS and the NPP/VIIRS⁴⁵. The DMSP/OLS data are derived from the OLS scanners from 1992 to 2013^{45,46}, and NPP/VIIRS data from the VIIRS scanners from 2012⁴⁷. Although the above two accessible nighttime light data have the advantages of long time span and wide space coverage, and are widely accepted and applied, they cannot be directly mixed together. It is mainly because the two nighttime light images belong to different sensors, different time and space of nighttime light image collection and differences in pixel levels. For example, the inconsistencies in the acquisition time and cloud cover of nighttime light images in 2013 resulted in a large gap between DMSP/OLS and NPP/VIIRS images at the pixel level. Moreover, the two nighttime light data also have the different discontinuities and saturation levels, and the differences in spillover and white noise. Thus, to obtain a long and continuous nighttime light data (including those before and after 2013), we spliced the two satellite nighttime light datasets following the procedures adopted by Liu *et al.*⁴⁸, Lv *et al.*⁴⁹, and Chen *et al.*³⁹.

First, inter-calibration⁵⁰, radiometric calibration⁵¹, intra-annual composition^{52,53} and inter-annual series correction methods⁵³ were adopted for correcting DMSP/OLS data to eliminate pixel under-saturation and spillover as well as discontinuities, incomparability, and instability. Inter-calibration method is an on-board calibration and additional processes on the annual composites of nighttime light data. Considering both satellite sensors

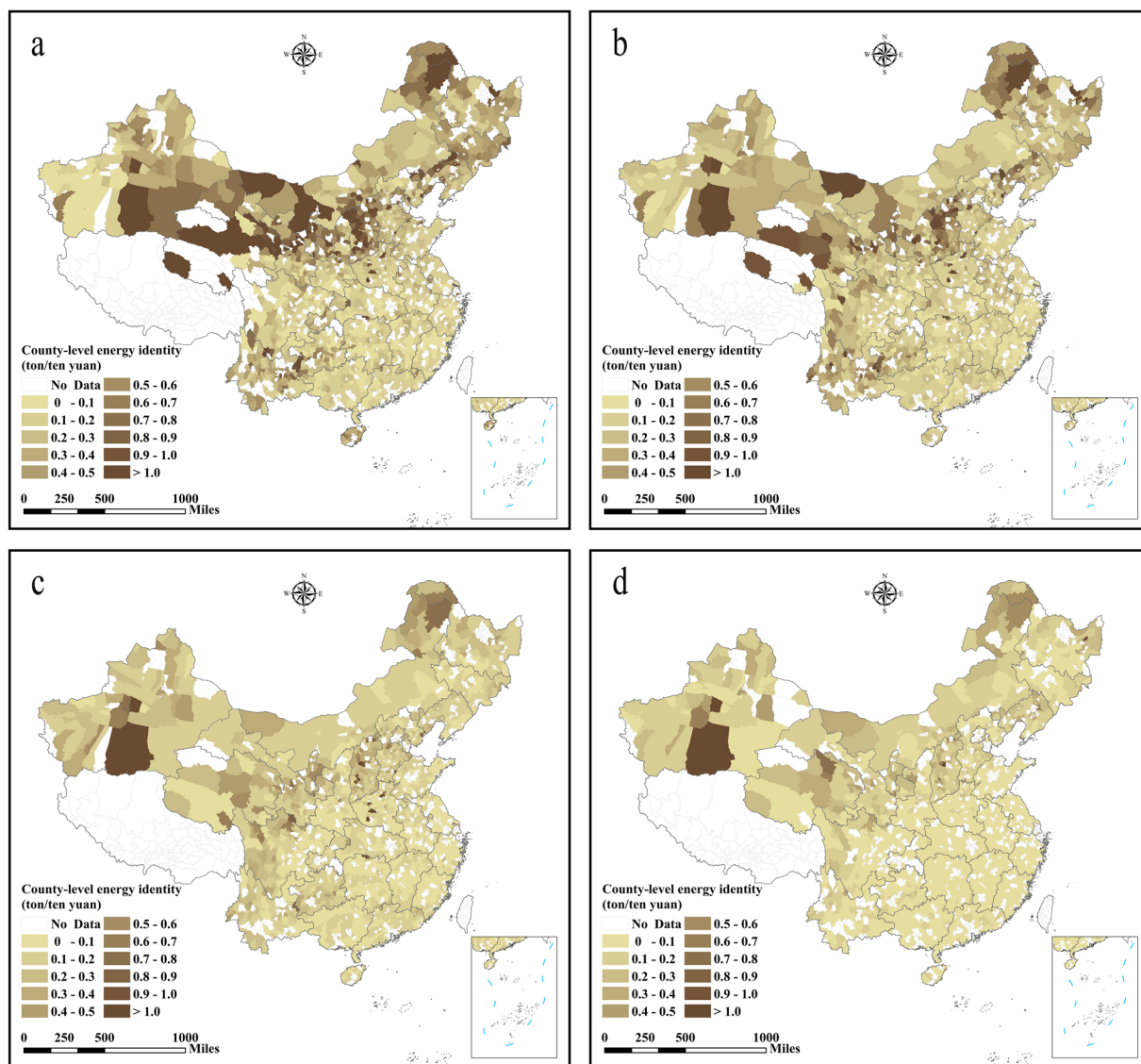


Fig. 5 County-level energy efficiency in China obtained from energy consumption per unit GDP. **(a)** The energy efficiency in 1997, **(b)** the energy efficiency in 2003, **(c)** the energy efficiency in 2010, **(d)** the energy efficiency in 2017.

transferring nighttime light data in the same year, such as F14 and F15 in 2001, we applied the intra-annual composition and to improve the stability of lit pixels, as follows:

$$DN_{(n,i)} = \begin{cases} 0, & DN_{(n,i)}^a = 0 \text{ and } DN_{(n,i)}^b = 0 \\ \frac{DN_{(n,i)}^a + DN_{(n,i)}^b}{2}, & \text{else} \end{cases} \quad (1)$$

where $DN_{(n,i)}$ represents the digital number (DN) values of lit pixel i from two types of satellite sensors in year n .

Radiometric calibration method is used for eliminating the differences between nighttime light data in the same year from two satellites when conducting comparative time series analysis. We applied the invariant region method to inter-calibrate DMSP/OLS images, which is consistent with Wu *et al.*⁵¹. It has better inter-calibration accuracy. Power function form was chosen and the global radiance calibrated nighttime light (RCNTL) in 2006 was treated as a reference image. The estimated parameters are shown in Table 1.

Inter-annual series correction method is used for stabilizing the inter-annual variability of the same satellite. According to Hu & Huang⁵³, it was performed by assuming that the DN value of each lit pixel would not decrease over time with the economic development, as follows:

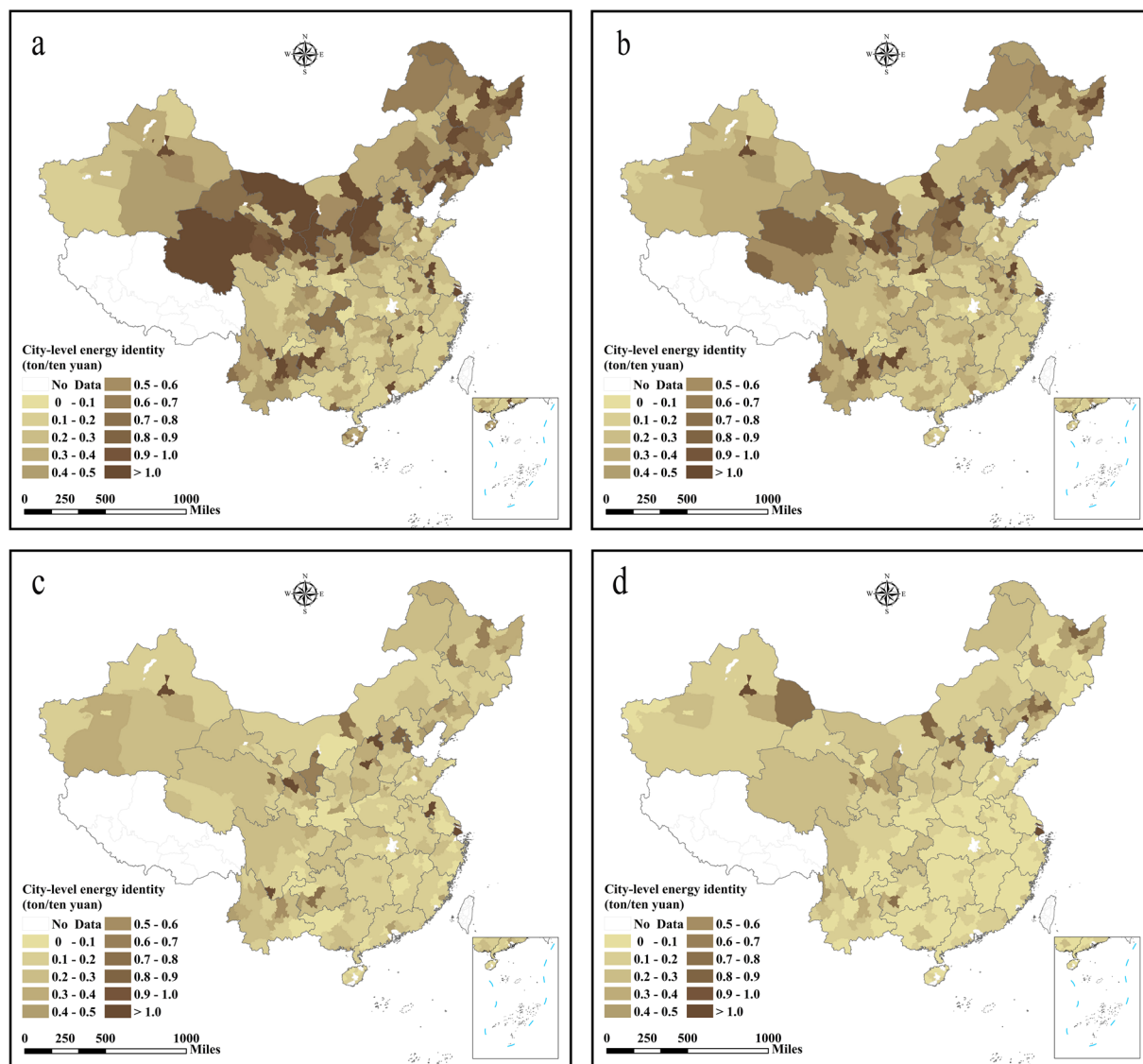


Fig. 6 City-level energy efficiency in China obtained from energy consumption of per unit GDP. (a) The energy efficiency in 1997, (b) the energy efficiency in 2003, (c) the energy efficiency in 2010, (d) the energy efficiency in 2017.

$$DN_{(n,i)} = \begin{cases} DN_{(n-1,i)}, & DN_{(n-1,i)} \geq DN_{(n,i)} \\ DN_{(n,i)}, & \text{else} \end{cases} \quad (2)$$

Moreover, as the higher intensity summer light in high latitudes can cause increased interference with the accuracy of the NPP/VIIRS data⁴⁹, the monthly NPP/VIIRS data for June, July, and August were excluded and the annual data was synthesized as an arithmetic mean based on the nighttime light data for the remaining nine months. To avoid the influence of deleting abnormal nighttime light data of the above three months on the accuracy of matching, noise pollution was removed in the process of converting NPP/VIIRS data into NPP/VIIRS data, which is consistent with Lv *et al.*⁴⁹. Accordingly, we used a Gaussian low-pass filter with a 5×5 pixel window to smooth the NPP/VIIRS annual data and reduce spatial variability⁵⁴ for better matching with the DMSP/OLS annual data. The Gaussian low-pass filter δ was set to 1.75^{54,55} and the negative value was replaced with zero. To further reduce the effect of white noise, a threshold $0.3nWcm^{-2}sr^{-1}$ was set in the annual image, and data smaller than this threshold were excluded⁵². Fig. 1 reported the differences by comparing the DMSP/OLS images in 2013 and the NPP/VIIRS images in 2017 before and after pre-processing, respectively. By comparing Fig. 1a and 1b, the processed images have solved the saturation issue of the original images that all DN values of these saturated pixels were 63. Meanwhile, compared to Fig. 1c, the processed images in Fig. 1d have eliminated negative values and maintained better matching with the DMSP/OLS images in terms of the distribution of nighttime light values.

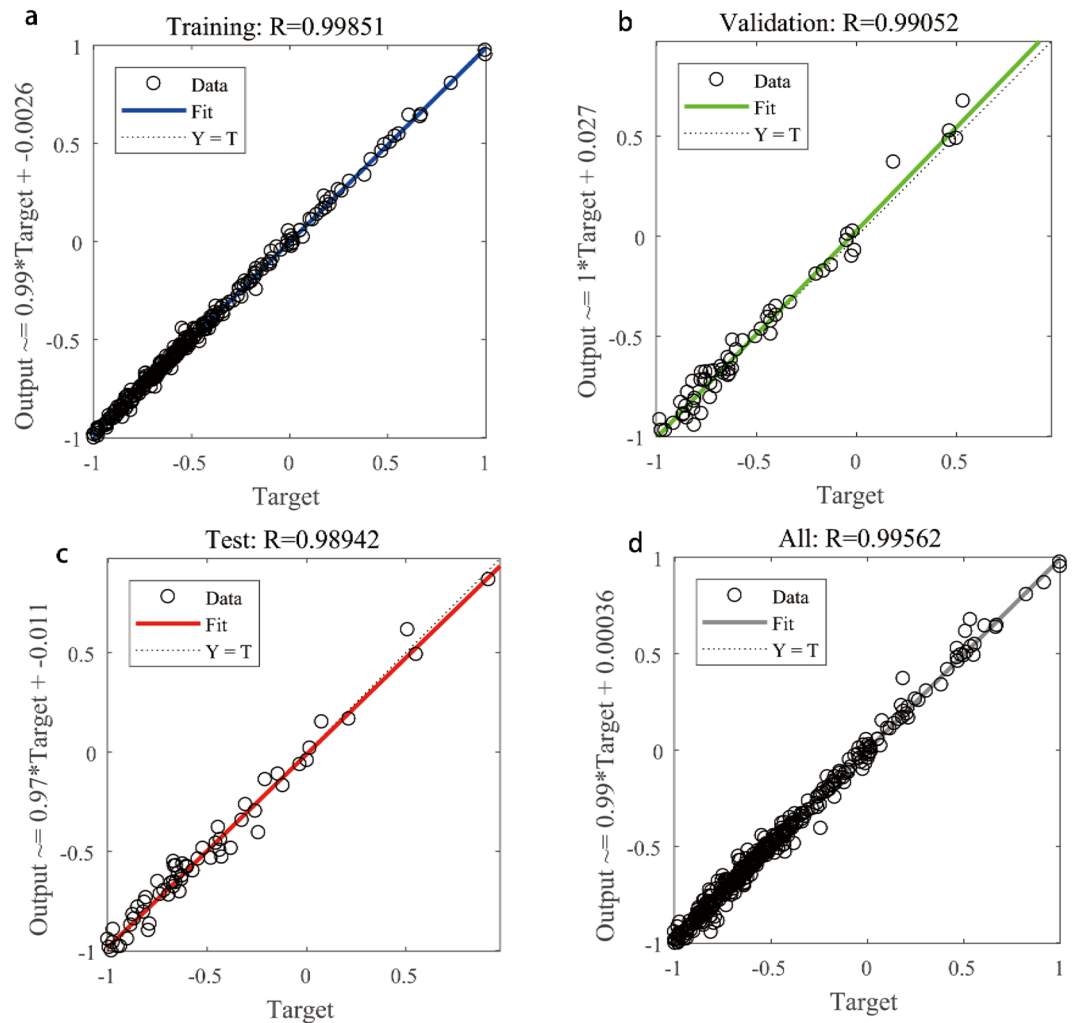


Fig. 7 Test results for the mean pixel values in 2013. Data represent (a) the results with the supplementary input parameters based on the PSO-BP algorithm, (b) the results for only the mean pixel values of nighttime light values based on the PSO-BP algorithm, (c) the results with the supplementary input parameters based on the BP algorithm, and (d) the results with only the mean pixel values of nighttime light values based on the BP algorithm.

Second, PSO-BP was used to match the DMSP/OLS, which has differences in sensors and resolution, with the NPP/VIIRS satellite nighttime lighting dataset. The PSO, proposed by Kennedy & Eberhart⁵⁶, is an intelligent global optimization algorithm based on the laws of collaborative mechanisms in bird populations. Each particle in the swarm represents a possible solution to the problem, and each particle corresponds to a fitness value whose position and velocity will be determined by the optimal position⁵⁷. After each position is updated, its fitness value is recalculated, and its position and velocity are updated again, so that all particles reach the optimal solution after repeated iterations⁵⁷. Rumelhart *et al.*⁵⁸ developed a multilayer feedforward neural network algorithm with a three-layer network topology, also known as BP network algorithm, which included an input layer, an output layer, and an implicit layer. The number of neurons in the implicit layer and the transfer function between the nodes are related to the specific object under study, and there is no unified principle for selecting them^{59,60}; however, they can generate local data at local extremes and cause training failures⁶¹. The BP algorithm is a local algorithm with good performance in constructing regressions and obtaining local optimum optimistic results, however, data at local extremes and training failures may occur. The PSO algorithm is a global algorithm which has the potential to explore the global optimistic results⁶². The combination of these two algorithms helps to improve the computational effectiveness⁶³. Ismail *et al.*⁶⁴, Mohamad *et al.*⁶³, Lee & Cheng⁵⁷, and Wang *et al.*²⁸ have successively combined these two algorithms and proposed the PSO-BP neural network algorithm. This improved algorithm can overcome the shortcomings of the traditional BP neural network, such as poor learning stability, low reliability, and easily attained local minima. The specific steps are as follows^{63,65}.

To build the neural network model, we set the county-level NPP/VIIRS pixel averages for 2013 as input parameters, and derived the central geographical coordinates of each county in China using the minimum boundary method, following Chen *et al.*⁴⁰. By applying the Lambert projection and the ArcGIS zone statistics,

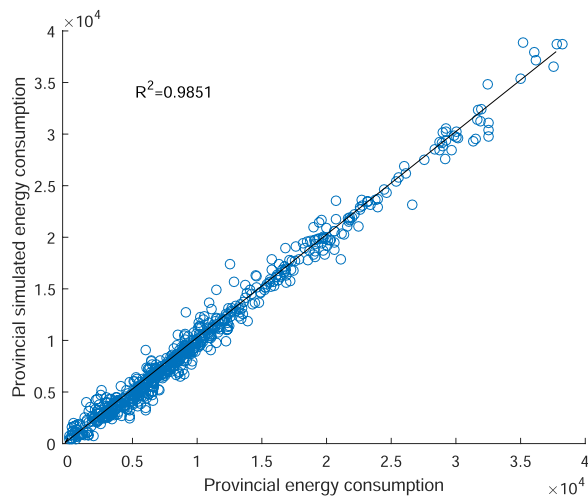


Fig. 8 Training and test results for the relationship between provincial energy consumption and the sum of DN values of provincial night-light data (unit: million tons of coal equivalents).

we obtained the area of each county using Arcmap 10.5, and used the abovementioned coordinate data as well as the area data of each county as supplementary data parameters. This approach reduces matching errors and improves the integration accuracy of the two sets of satellite nighttime lighting data. We logarithmized the input parameters according to Li *et al.*⁵⁴. When setting the output parameters, the 2013 county DMSP/OLS pixel averages are used as the output parameters.

The computational program developed by Ismail *et al.*⁶⁴ and Mohamad *et al.*⁶³ was used as a reference for the inter-calibration of the night-light data from DMSP/OLS and NPP/VIIRS. The number of layers of BP neural network is set to three, in which the hidden layer is set to five nodes, the maximum number of iterations is set to 50, and the population size is 20⁵⁴. There are 2,735 counties in the Chinese mainland; therefore, we randomly selected 2,000 counties as our training set and 735 counties as our test set. However, even though the algorithm results obtained a high coefficient of determination, a large amount of data might also cause problems such as anomalous value, fluctuations, and jumps and faults in the conversion process. Thus, continuous satellite nighttime light data might not be obtained, leading to large errors in the estimated energy consumption data. To reduce this error, we further converted NPP/VIIRS data from 2013 to DMSP/OLS data, and calculated the annual growth of NPP/VIIRS data from 2013 to 2017 after conversion. For example, the annual growth of NPP/VIIRS data for 2014 and the existing DMSP/OLS data for 2013 provided the simulated 2014 DMSP/OLS data. Then, we can calculate the smooth DMSP/OLS data obtained from the 2014–2017 simulation⁴⁰, as shown in the following equations.

$$\Delta DV_{i,t} = DV_{i,t} - DV_{i,t-1} \quad (3)$$

$$D_{i,t} = D_{i,t-1} + \Delta DV_{i,t} \quad (4)$$

where $D_{i,t}$ is the nighttime light data of DMSP/OLS, $DV_{i,t}$ is the DMSP/OLS nighttime light data transformed by NPP/VIIRS nighttime light data, $\Delta DV_{i,t}$ is the annual increment of NPP/VIIRS nighttime light data after transforming, i represent counties ($i=1, 2, 3, \dots, 2735$), and t represent years ($t = 2014, 2015, 2016, 2017$).

Estimation of satellite-based energy consumption at the city and county levels. For the inversion of county-level energy consumption data, the provincial energy balance sheet data published in the *China Energy Statistics Yearbook* and the previously mentioned PSO-BP neural network algorithm were used to establish the quantitative relationship between provincial energy consumption data and provincial nighttime light data. The values of sum of the DN (SDN) value, and dummy variables of identity and year were selected as input parameters, and the provincial energy consumption were selected as output parameters. The reason why dummy variables of identity and year were selected in addition to the SDN values is to control the influence of province differences that do not change with time and time differences that do not change with time on provincial energy consumption. For example, although the DN values of two provinces are the same in a certain year, the different levels of energy consumption may still occur because of other parameters that have not been observed and vary with provinces. Similarly, although a province has the same DN value in two years, it may still have different energy consumption due to other parameters that have not been observed over time. The total energy consumption sample was 630 (21 years \times 30 provinces), of which 400 samples were randomly selected as the training set, and the remaining 230 samples were selected as the test set. County energy consumption data^{66–68} were obtained using the top-down method and a DN value-based weighted-average strategy. Finally, 2,735 county-level energy consumption data were aggregated to obtain 336 city-level energy consumption data. The equation is as follows:

$$CE_{i,t} = \frac{SDN_{i,t}}{SDN_{j,t}} PE_{i,t} \quad (5)$$

where $CE_{i,t}$ represents county energy consumption, $SDN_{i,t}$ represents county lighting brightness, $SDN_{j,t}$ represents province lighting brightness, $PE_{i,t}$ represents province energy consumption, i represents county ($i = 1, 2, 3, \dots, 2735$), j represents province ($j = 1, 2, 3, \dots, 30$), and t represents year ($t = 1997, 1998, 1999, \dots, 2017$).

Calculation of energy efficiency based on two methods. This dataset uses the DEA and the IMA to measure the energy efficiency data at the city and county levels in China. DEA is a non-parametric method widely used to measure efficiency^{69–71}. Researchers worldwide have adopted and constantly improved this approach in their studies on energy efficiency^{72–75}. The energy efficiency can be calculated using this method without setting the production function between input and output, which could be regarded as its advantage. According to the studies by Chen *et al.*⁷⁶, Jebali *et al.*⁷⁷, Wang *et al.*⁷⁸, and Wu *et al.*⁷⁹, we selected labor, capital, and energy as input variables; GDP as expected output variable; and particulate matter (PM_{2.5}) as undesirable output variable.

We chose the NDDF with an undesired output as the non-parametric functions. This is because, unlike the traditional Sheppard distance function, the NDDF with undesirable output considers both the expected output and the undesirable output (usually the pollutant), and it does not require price-specific data. Moreover, compared with the traditional directional distance function, NDDF with undesirable output can avoid the problem of overestimating efficiency due to the existence of slacks. Furthermore, NDDF can calculate the inefficiency value of each input and output factor^{80–82}. Therefore, the use of NDDF is widely accepted in measuring energy efficiency^{83–85}. By referring to the studies by Zhou *et al.*⁸⁶, Zhang *et al.*⁸⁷, and Lin & Chen⁸⁸, we set the collection of weights for energy efficiency $EEPI$ as $(0, 0, 0, 1/3, 1/3, 1/3)$ and the set of directions as $(0, 0, 0, 0, 0, 0)$, details as follows:

$$\rightarrow_{D_E}(K, L, E, Y, B; g) = \max w_E \beta_E + w_Y \beta_Y + w_B \beta_B \quad (6)$$

$$\begin{aligned} s.t. \quad & \sum_{n=1}^N Z_n K_n \leq K \\ & \sum_{n=1}^N Z_n L_n \leq L \\ & \sum_{n=1}^N Z_n E_n \leq E - \beta_E g_E \\ & \sum_{n=1}^N Z_n G_n \geq Y + \beta_Y g_Y \\ & \sum_{n=1}^N Z_n B_n = B - \beta_B g_B \end{aligned} \quad (7)$$

$$Z_n \geq 0, n = 1, 2, \dots, N \quad (8)$$

$$\beta_E, \beta_Y, \beta_B \geq 0 \quad (9)$$

$$EEPI = \frac{1/2[(1 - \beta_E^*) + (1 - \beta_B^*)]}{1 + \beta_Y^*} = \frac{1 - 1/2(\beta_E^* + \beta_B^*)}{1 + \beta_Y^*} \quad (10)$$

where $\rightarrow(\cdot)$ represents the directional distance function; and K represents the capital stock, which is calculated by adopting perpetual inventory method based on the fixed asset investment data in the *China City Statistical Yearbook*: $K_t = I_t + (1 - \alpha_t)K_{t-1}$, where I_t is the current fixed asset investment, and α_t is the depreciation rate of fixed assets. Referring to the set of Meng *et al.*⁸⁹, α_t is assumed to be 9.6%, and K_{t-1} is the capital stock of the previous period. After using the fixed asset investment price index to deal with the fixed asset investment, the capital stock at the beginning of the period is set as $K_{2003} = I_{2003}/(\alpha_t + p_n)$ according to Li⁹⁰, where K_{2003} is the capital stock at the beginning of 2003, I_{2003} represents the value of fixed asset investment in 2003 after adjusting for prices, and p_n is the geometric mean of the growth rate of real fixed asset investment. L represents the labor, as measured by the number of employees in urban units in the *China City Statistical Yearbook*; E represents energy consumption; Y represents the gross product of a city, derived from the *China City Statistical Yearbook* and the provincial statistical yearbooks; B represents PM_{2.5}, derived from the real time air quality release system of the Ministry of Environmental Protection of China, that is the National Control Site Hour Overall Data. We first found the average daily concentration of PM_{2.5} per hour, then multiplied the hourly concentration of PM_{2.5} by 24 to get the daily concentration of PM_{2.5}, and finally divided the total daily concentration of PM_{2.5} by 365 to

get the annual concentration of $PM_{2.5}$, g is direction vector; w_E , w_Y , and w_B represent the city's energy consumption, GDP, and the standardized weight vector for $PM_{2.5}$, respectively. β_E , β_Y , and β_B represent energy consumption, GDP, and level factor vector of $PM_{2.5}$ in cities, respectively. In the constraint conditions represented by s , Z_n is the weight coefficient; n is the city; and β_E^* , β_Y^* , and β_B^* are the energy consumption, GDP, and the inefficiency value of $PM_{2.5}$, respectively.

The energy efficiency derived from the DEA method is generally considered as more scientific and reliable; however, its calculations must be supported by other socioeconomic data, such as the capital stock and labor. These data are not published by the statistical agencies of some cities and counties. Therefore, we could only provide energy efficiency data for 189 cities using the DEA.

Energy consumption per unit GDP has also been used to approximately represent energy efficiency by some studies, such as Bor⁹¹, Duro & Padilla⁹², Cheng *et al.*⁹³, and Cheng *et al.*⁹⁴. Lower energy consumption, which drives economic growth, is often considered to be more energy efficient. Accordingly, we used the energy consumption per unit GDP to estimate the energy efficiency of 336 cities and 2,489 counties in China.

Data Records

This dataset provides a total of 6,085 data records (energy consumption and energy efficiency)⁴⁴. These include: 336 city energy consumption data (1997–2017), 2,735 county energy consumption data (1997–2017), 189 city energy efficiency data (2003–2016) based on DEA, 336 city energy efficiency data (1997–2017) based on energy consumption per unit GDP, and 2,489 county energy efficiency data (1997–2017). The datasets and figures published herein are publicly available in the data files. Energy consumption data was measured in million tons of coal equivalents. The spatial and temporal distribution of energy consumption and energy efficiency at the city and county levels in China are shown in Figs. 2–6.

Technical Validation

Validity testing for energy consumption based on nighttime light data from satellites. The fitting results of PSO-BP neural network algorithm are shown in Fig. 7. The coefficient of determination of the training set was 0.999, while that of the test and verification sets were 0.989 and 0.991, respectively. This shows that the model has good fitting effect and validity in the training, testing, and verification stages, and the overall coefficient of determination of the model was high at 0.9956, which is higher than that of Yue *et al.*⁴¹. In their work, the coefficient of determination using the econometrics estimation method was 0.734.

Furthermore, we fit the county-level energy consumption data according to the province-level division by using the sum of the provincial simulated energy consumption, and the results were compared with official statistical provincial energy consumption data in the *China Energy Statistics Yearbook* from 1997 to 2017. The latter was published by the Chinese official agency, CNBS, and was used as the validation data for comparisons, as shown in Fig. 8. The provincial energy consumption data fitted in this study was highly correlated with the official statistical provincial energy consumption data, and the decisive coefficient was 0.9851. This shows that the energy consumption data obtained from fitting of satellite nighttime light data is reliable.

Limitations and future work. There are three deficiencies in our datasets, which will provide future avenues for more accurate estimates of energy consumption and efficiency at the city and county levels in China. First, the city-and-county-level energy consumption data retrieved from satellite nighttime light data only comprises the total energy consumption data, and cannot identify the change of energy mix. Second, the provincial energy consumption data spliced using satellite nighttime light data do not include energy consumption data from land-use change and use of non-commercial biomass fuels, which results in underestimation of the energy consumption data of cities and counties. Third, considering the unavailability of socioeconomic data for some cities and all counties, such as labor and capital stock, even though data on energy consumption was provided for 336 cities and 2,735 counties from 1997 to 2017, the energy efficiency data based on the DEA included only 189 cities between 2003 and 2016.

Therefore, our future studies will focus on three aspects: First, in the retrieval of city-and-county-level energy consumption data, the disaggregated energy consumption data and other socioeconomic variables should be explored and introduced. Second, a provincial total energy consumption data covering more energy types should be obtained before comprehensively inverting the city and county energy consumption data through investigation. Third, the socioeconomic data of cities and counties should be obtained to support the use of DEA in evaluating the energy efficiency of all cities and counties in China.

Usage Notes

The data of city-and-county-level energy consumption and efficiency provided in this study are valuable with practical applications in the field of energy economics, management, and policy, including the following: First, the provided data have the characteristics of wide spatial coverage and long time span. This unique panel-structured dataset can be used to observe the trajectories and spatial differences of energy consumption and efficiency on a micro-level than at national and provincial levels. Therefore, it can also be used to analyze the factors driving the changes and spatial differences of energy consumption and efficiency in cities and counties. Second, the panel-structured dataset of energy consumption and efficiency can be used to match other socioeconomic data at the city and county levels, and studies such as the economic effects of energy consumption, the environmental and social effects of energy consumption, and the coupling relationship between energy efficiency and economic development may be conducted. Third, the development of energy consumption and efficiency data at the city and county levels can not only contribute to the energy management at China's

grassroot-level governments, such as in the formulation and implementation of road maps for energy transition and energy efficiency improvement, but also provide a basis for the central and provincial governments to allocate the energy rights of cities and counties under the constraints of “carbon peak” and “carbon-neutral” targets. Fourth, the development of energy consumption and efficiency data at city and county levels could provide a more accurate assessment of the impact of the central government’s energy saving, emissions reduction, and low-carbon green policies, as well as other socioeconomic policies, for example, assessment of the impact of the “central heating,” “coal to electricity,” low-carbon pilot city, and carbon emissions trading right pilot policies on energy consumption and energy efficiency. Fifth, the method of retrieving micro-level energy consumption data by using satellite nighttime light data can also provide a reference for other developing countries and regions with limited energy statistics to evaluate their energy consumption and efficiency at the sub-national level.

Code availability

MatLab (R2017b), STATA (16), and ArcGIS (10.5) are the major applications used to obtain the energy consumption and efficiency data. The code for PSO-BP matching the relationship between the two sets of satellite nighttime light data and the inversion of city and county energy consumption is provided in the Appendix. Codes and datasets for the DEA method, which measures energy efficiency in 189 cities, are also provided in the Appendix.

Received: 14 September 2021; Accepted: 4 March 2022;

Published online: 24 March 2022

References

- BP. <https://www.bp.com/content/dam/bp/business-sites/en/global/corporate/pdfs/energy-economics/energy-outlook/bp-energy-outlook-2020-country-insight-china.pdf> (2020).
- Magazzino, C., Mele, M. & Schneider, N. A machine learning approach on the relationship among solar and wind energy production, coal consumption, GDP, and CO₂ emissions. *Renew. Energy* **167**, 99–115, <https://doi.org/10.1016/j.renene.2020.11.050> (2021).
- Li, J. D. *et al.* How to make better use of intermittent and variable energy: a review of wind and photovoltaic power consumption in China. *Renew. Sustain. Energy Rev.* **137**, 110626, <https://doi.org/10.1016/j.rser.2020.110626> (2021).
- Zhao, J. C. & Liu, Q. Q. Examining the driving factors of urban residential carbon intensity using the LMDI method: evidence from China’s county-level cities. *Int. J. Environ. Res. Public Health* **18**, 3929, <https://doi.org/10.3390/ijerph18083929> (2021).
- Zheng, W. & Walsh, P. P. Economic growth, urbanization and energy consumption: a provincial level analysis of China. *Energy Econ.* **80**, 153–162, <https://doi.org/10.1016/j.eneco.2019.01.004> (2019).
- Si, S. Y., Lyu, M. J., Lawell, C. Y. C. L. & Chen, S. The effects of energy-related policies on energy consumption in China. *Energy Econ.* **76**, 202–227, <https://doi.org/10.1016/j.eneco.2018.10.013> (2018).
- Kaivo-oja, J., Vehmas, J. & Luukkanen, J. Trend analysis of energy and climate policy environment: comparative electricity production and consumption benchmark analyses of China, Euro area, European Union, and United States. *Renew. Sustain. Energy Rev.* **60**, 464–474, <https://doi.org/10.1016/j.rser.2016.01.086> (2016).
- Le Quere, C. *et al.* Drivers of declining CO₂ emissions in 18 developed economies. *Nat. Clim. Change* **9**, 213–217, <https://doi.org/10.1038/s41558-019-0419-7> (2019).
- Zhou, W., Chen, Q. J., Luo, D. X., Jiang, R. & Chen, J. Global energy consumption analysis based on the three-dimensional network model. *IEEE Access* **8**, 76313–76332, <https://doi.org/10.1109/ACCESS.2020.2989186> (2020).
- Khan, A. M. & Osinska, M. How to predict energy consumption in BRICS countries? *Energies* **14**, 2749, <https://doi.org/10.3390/en14102749> (2021).
- Hoesly, R. M. & Smith, S. J. Informing energy consumption uncertainty: an analysis of energy data revisions. *Environ. Res. Lett.* **13**, 124023, <https://doi.org/10.1088/1748-9326/aebc3> (2018).
- Wu, W. Q., Ma, X., Wang, Y., Cai, W. & Zeng, B. Predicting China’s energy consumption using a novel grey Riccati model. *Appl. Soft Comput.* **95**, 106555, <https://doi.org/10.1016/j.asoc.2020.106555> (2020).
- Khan, I., Hou, F. J., Zakari, A. & Tawiah, V. K. The dynamic links among energy transitions, energy consumption, and sustainable economic growth: a novel framework for IEA countries. *Energy* **222**, 119935, <https://doi.org/10.1016/j.energy.2021.119935> (2021).
- Shahbaz, M., Raghutla, C., Chittedi, K. R., Jiao, Z. L. & Vo, X. V. The effect of renewable energy consumption on economic growth: evidence from the renewable energy country attractive index. *Energy* **207**, 118162, <https://doi.org/10.1016/j.energy.2020.118162> (2020).
- Bhattacharya, M., Paramati, S. R., Ozturk, I. & Bhattacharya, S. The effect of renewable energy consumption on economic growth: evidence from top 38 countries. *Appl. Energy* **162**, 733–741, <https://doi.org/10.1016/j.apenergy.2015.10.104> (2016).
- Rauf, A., Zhang, J., Li, J. K. & Amin, W. Structural changes, energy consumption and carbon emissions in China: empirical evidence from ARDL bound testing model. *Struct. Change Econ. Dyn.* **47**, 194–206, <https://doi.org/10.1016/j.strueco.2018.08.010> (2018).
- Chen, J., Zhou, C. S., Wang, S. J. & Li, S. J. Impacts of energy consumption structure, energy intensity, economic growth, urbanization on PM_{2.5} concentrations in countries globally. *Appl. Energy* **230**, 94–105, <https://doi.org/10.1016/j.apenergy.2018.08.089> (2018).
- Wei, W. D., Cai, W. Q., Guo, Y., Bai, C. Q. & Yang, L. Z. Decoupling relationship between energy consumption and economic growth in China’s provinces from the perspective of resource security. *Resour. Policy* **68**, 101693, <https://doi.org/10.1016/j.resourpol.2020.101693> (2020).
- Fan, W. Y. & Hao, Y. An empirical research on the relationship amongst renewable energy consumption, economic growth and foreign direct investment in China. *Renew. Energy* **146**, 598–609, <https://doi.org/10.1016/j.renene.2019.06.170> (2020).
- Shan, Y. L. *et al.* New provincial CO₂ emission inventories in China based on apparent energy consumption data and updated emission factors. *Appl. Energy* **184**, 742–750, <https://doi.org/10.1016/j.apenergy.2016.03.073> (2016).
- Agbede, E. A., Bani, Y., Azman-Saini, W. N. W. & Naseem, N. A. M. The impact of energy consumption on environmental quality: empirical evidence from the MINT countries. *Environ. Sci. Pollut. Res.* **1–20**, <https://doi.org/10.1007/s11356-021-14407-2> (2021).
- Zi, C., Qian, M. & Gao, B. Z. The consumption patterns and determining factors of rural household energy: a case study of Henan Province in China. *Renew. Sustain. Energy Rev.* **146**, 111142, <https://doi.org/10.1016/j.rser.2021.111142> (2021).
- Shan, Y. L. *et al.* Chinese cities exhibit varying degrees of decoupling of economic growth and CO₂ emissions between 2005 and 2015. *One Earth* **1**, 124–134, <https://doi.org/10.1016/j.oneear.2020.12.004> (2021).
- Guo, R. & Yuan, Y. Different types of environmental regulations and heterogeneous influence on energy efficiency in the industrial sector: evidence from Chinese provincial data. *Energy Policy* **145**, 111747, <https://doi.org/10.1016/j.enpol.2020.111747> (2020).
- Xing, J. *et al.* The quest for improved air quality may push China to continue its CO₂ reduction beyond the Paris Commitment. *PNAS* **117**, 29535–29542, <https://doi.org/10.1073/pnas.2013297117> (2020).

26. Dogan, E., Seker, F. & Bulbul, S. Investigating the impacts of energy consumption, real GDP, tourism and trade on CO₂ emissions by accounting for cross-sectional dependence: A panel study of OECD countries. *Curr. Issues Tour.* **20**, 1701–1719, <https://doi.org/10.1080/13683500.2015.1119103> (2017).
27. Li, Y., Chiu, Y. H., Lu, L. C. & Chiu, C. R. Evaluation of energy efficiency and air pollutant emissions in Chinese provinces. *Energy Effic.* **12**, 963–977, <https://doi.org/10.1007/s12053-018-9733-6> (2018).
28. Wang, S. W. *et al.* Natural gas shortages during the “coal-to-gas” transition in China have caused a large redistribution of air pollution in winter 2017. *PNAS.* **117**, 31018–31025, <https://doi.org/10.1073/pnas.2007513117> (2020).
29. Chen, L., Xu, L. Y., Cai, Y. P. & Yang, Z. F. Spatiotemporal patterns of industrial carbon emissions at the city level. *Resour. Conserv. Recycl.* **169**, 105499, <https://doi.org/10.1016/j.resconrec.2021.105499> (2021).
30. Shan, Y. L. *et al.* City-level climate change mitigation in China. *Sci. Adv.* **4**, eaaq0390, <https://doi.org/10.1126/sciadv.aaq0390> (2018).
31. Zhu, J. P. & Lin, B. Q. Convergence analysis of city-level energy intensity in China. *Energy Policy* **139**, 111357, <https://doi.org/10.1016/j.enpol.2020.111357> (2020).
32. Pei, J. *et al.* Spatial-temporal dynamics of carbon emissions and carbon sinks in economically developed areas of China: a case study of Guangdong Province. *Sci. Rep.* **8**, 13383, <https://doi.org/10.1038/s41598-018-31733-7> (2018).
33. Wang, H. *et al.* China’s CO₂ peak before 2030 implied from characteristics and growth of cities. *Nat. Sustain.* **2**, 748–754, <https://doi.org/10.1038/s41893-019-0339-6> (2019).
34. Zhang, H., Fan, L. W. & Zhou, P. Handling heterogeneity in frontier modeling of city-level energy efficiency: the case of China. *Appl. Energy* **279**, 115846, <https://doi.org/10.1016/j.apenergy.2020.115846> (2020a).
35. Wu, S., Zheng, X. & Wei, C. Measurement of inequality using household, energy consumption data in rural China. *Nat. Energy* **2**, 795–803, <https://doi.org/10.1038/s41560-017-0003-1> (2017).
36. Meng, W. *et al.* Energy and air pollution benefits of household fuel policies in northern China. *PNAS.* **116**, 16773–16780, <https://doi.org/10.1073/pnas.1904182116> (2019).
37. Kushal, T. & Chandra, V. Climate co-benefits of air quality and clean energy policy in India. *Nat. Sustain.* **4**, 305–313, <https://doi.org/10.1038/s41893-020-00666-3> (2020).
38. Tong, K. *et al.* Measuring social equity in urban energy use and interventions using fine-level data. *PNAS.* **118**, 2023554118, <https://doi.org/10.1073/pnas.2023554118> (2021).
39. Chen, H., Zhang, X., Wu, R. & Cai, T. Revisiting the environmental Kuznets curve for city-level CO₂ emissions based on corrected NPP-VIIRS nighttime light data in China. *J. Clean. Prod.* **268**, 121575, <https://doi.org/10.1016/j.jclepro.2020.121575> (2020).
40. Chen, J. D. *et al.* County-level CO₂ emissions and sequestration in China during 1997–2017. *Sci. Data* **7**, 391, <https://doi.org/10.1038/s41597-020-00736-3> (2020).
41. Yue, Y., Tian, L., Yue, Q. & Wang, Z. Spatiotemporal variations in energy consumption and their influencing factors in China based on the integration of the DMSP-OLS and NPP-VIIRS nighttime light datasets. *Remote Sens.* **12**, 1151, <https://doi.org/10.3390/rs12071151> (2020).
42. Ramaswami, A. *et al.* Urban cross-sector actions for carbon mitigation with local health co-benefits in China. *Nat. Clim. Change* **7**, 736–742, <https://doi.org/10.1038/nclimate3373> (2017).
43. He, G., Liu, T. & Zhou, M. Straw burning, PM_{2.5}, and death: evidence from China. *J. Dev. Econ.* **145**, 102468, <https://doi.org/10.1016/j.jdeveco.2020.102468> (2020).
44. Chen, J. D. *et al.* City- and county-level spatio-temporal energy consumption and efficiency datasets for China from 1997 to 2017. *figshare. Dataset.* <https://doi.org/10.6084/m9.figshare.19196780.v1> (2022).
45. NOAA. DMSP & VIIRS Data Download. <https://ngdc.noaa.gov/eog/download.html> (2020).
46. Yang, W. M. *et al.* A new global anthropogenic heat estimation based on high-resolution nighttime light data. *Sci. Data* **4**, 170116, <https://doi.org/10.1038/sdata.2017.116> (2017).
47. Sun, Y. B., Zheng, S., Wu, Y., Schlink, U. & Singh, R. P. Spatiotemporal variations of city-level carbon emissions in China during 2000–2017 using nighttime light data. *Remote Sens.* **12**, 2916, <https://doi.org/10.3390/rs12182916> (2020).
48. Liu, Z., He, C., Zhang, Q., Huang, Q. & Yang, Y. Extracting the dynamics of urban expansion in China using DMSP-OLS nighttime light data from 1992 to 2008. *Landsc. Urban Plan.* **106**, 62–72, <https://doi.org/10.1016/j.landurbplan.2012.02.013> (2012).
49. Lv, Q., Liu, H., Wang, J., Liu, H. & Shang, Y. Multilevel analysis on spatiotemporal dynamics of energy consumption CO₂ emissions in China: utilizing the integrated of DMSP-OLS and NPP-VIIRS nighttime light datasets. *Sci. Tot. Environ.* **703**, 134394 (2020). [10.1016/j.scitotenv.2019.134394](https://doi.org/10.1016/j.scitotenv.2019.134394).
50. Kumar, P. *et al.* Analyzing trend in artificial light pollution pattern in India using NTL sensor’s data. *Urban Clim.* **27**, 272–283, <https://doi.org/10.1016/j.uclim.2018.12.005> (2018).
51. Wu, J. S., He, S. B., Peng, J., Li, W. F. & Zhong, X. H. Intercalibration of DMSP-OLS night-time light data by the invariant region method. *Int. J. Remote Sens.* **34**, 7356–7368, <https://doi.org/10.1080/01431161.2013.820365> (2013).
52. Ma, T., Zhou, C., Pei, T., Haynie, S. & Fan, J. Responses of Suomi-NPP VIIRS-derived nighttime lights to socioeconomic activity in China’s cities. *Remote Sens. Lett.* **5**, 165–174, <https://doi.org/10.1080/2150704X.2014.890758> (2014).
53. Hu, T. & Huang, X. A novel locally adaptive method for modeling the spatiotemporal dynamics of global electric power consumption based on DMSP-OLS nighttime stable light data. *Appl. Energy* **240**, 778–792, <https://doi.org/10.1016/j.apenergy.2019.02.062> (2019).
54. Li, X., Li, D., Xu, H. & Wu, C. Intercalibration between DMSP/OLS and VIIRS night-time light images to evaluate city light dynamics of Syria’s major human settlement during Syrian Civil War. *Int. J. Remote Sens.* **38**, 5934–5951, <https://doi.org/10.1080/01431161.2017.1331476> (2017).
55. Zheng, Q., Weng, Q. & Wang, K. Developing a new cross-sensor calibration model for DMSP-OLS and Suomi-NPP VIIRS night-light imageries. *ISPRS J. Photogramm.* **153**, 36–47, <https://doi.org/10.1016/j.isprsjprs.2019.04.019> (2019).
56. Kennedy, J. & Eberhart, R. Particle swarm optimization. *Proceedings of ICNN’95-international conference on neural networks. IEEE.* **4**, 1942–1948 (1995). [1109/ICNN.1995.488968](https://doi.org/10.1109/ICNN.1995.488968).
57. Lee, C. Y. & Cheng, Y. H. Motor fault detection using wavelet transform and improved PSO-BP neural network. *Process.* **8**, 1322. [10.3390/pr8101322](https://doi.org/10.3390/pr8101322) (2020).
58. Rumelhart, D., Hinton, G. & Williams, R. Learning representations by back-propagating errors. *Nature.* **323**, 533–536, <https://doi.org/10.1038/323533a0> (1986).
59. Zhang, C., Shao, H. & Li, Y. Particle swarm optimisation for evolving artificial neural network. *IEEE International Conference on Systems, Man and Cybernetics. ‘Cybernetics evolving to systems, humans, organizations, and their complex interactions’.* **4**, 2487–2490, <https://doi.org/10.1109/ICSMC.2000.884366> (2000).
60. Adhikari, P., Agrawal, R. K. & Kant, L. PSO based neural networks vs. traditional statistical models for seasonal time series forecasting. *IEEE Int. Adv. Comput. Conf.* **3**, 719–725, <https://doi.org/10.1109/IADCC.2013.6514315> (2013).
61. Li, R. Y., Liu, X. N. & Li, X. Q. Estimation of the PM_{2.5} pollution levels in Beijing based on nighttime light data from the defense meteorological satellite program-operational linescan system. *Atmos.* **6**, 607–622, <https://doi.org/10.3390/atmos6050607> (2015).
62. Jakubik, J., Binding, A. & Feuerriegel, S. Directed particle swarm optimization with Gaussian-process-based function forecasting. *Eur. J. Oper. Res.* **295**, 157–169, <https://doi.org/10.1016/j.ejor.2021.02.053> (2021).
63. Mohamad, E. T., Armaghani, D. J., Momeni, E., Yazdavar, A. H. & Ebrahimi, M. Rock strength estimation a PSO-based BP approach. *Neural Comput. Appl.* **30**, 1635–1646, <https://doi.org/10.1007/s00521-016-2728-3> (2016).

64. Ismail, A., Jeng, D. S. & Zhang, L. L. An optimised product-unit neural network with a novel PSO-BP hybrid training algorithm: applications to load-deformation analysis of axially loaded piles. *Eng. Appl. Artif. Intel.* **26**, 2305–2314, <https://doi.org/10.1016/j.engappai.2013.04.007> (2013).
65. Yin, X., Cao, F., Wang, J., Li, M. & Wang, X. Investigations on optimal discharge pressure in CO₂ heat pumps using the GMDH and PSO-BP type neural network-part a: theoretical modeling. *Int. J. Refrig.* **106**, 549–557, <https://doi.org/10.1016/j.jrefrig.2019.04.027> (2019).
66. Meng, L., Graus, W., Worrell, E. & Huang, B. Estimating CO₂ (carbon dioxide) emissions at urban scales by DMSP/OLS (Defense Meteorological Satellite Program's Operational Linescan System) nighttime light imagery: Methodological challenges and a case study for China. *Energy* **71**, 468–478, <https://doi.org/10.1016/j.energy.2014.04.103> (2014).
67. Su, Y. *et al.* China's 19-year city-level carbon emissions of energy consumptions, driving forces and regionalized mitigation guidelines. *Renew. Sustain. Energy Rev.* **35**, 231–243, <https://doi.org/10.1016/j.rser.2014.04.015> (2014).
68. Yang, D., Luan, W., Qiao, L. & Pratama, M. Modeling and spatio-temporal analysis of city-level carbon emissions based on nighttime light satellite imagery. *Appl. Energy* **268**, 114696, <https://doi.org/10.1016/j.apenergy.2020.114696> (2020).
69. Ouenniche, J. & Carrales, S. Assessing efficiency profiles of UK commercial banks: a DEA analysis with regression-based feedback. *Ann. Oper. Res.* **266**, 551–587, <https://doi.org/10.1007/s10479-018-2797-z> (2018).
70. Cantor, V. J. M. & Poh, K. L. Integrated analysis of healthcare efficiency: a systematic review. *J. Med. Syst.* **42**, 1–23, <https://doi.org/10.1007/s10916-017-0848-7> (2018).
71. Song, M. L. & Li, H. Estimating the efficiency of a sustainable Chinese tourism industry using bootstrap technology rectification. *Technol. Forecast. Soc. Change* **143**, 45–54, <https://doi.org/10.1016/j.techfore.2019.03.008> (2019).
72. Cui, Q., Kuang, H., Wu, C. & Li, Y. The changing trend and influencing factors of energy efficiency: the case of nine countries. *Energy* **64**, 1026–1034, <https://doi.org/10.1016/j.energy.2013.11.060> (2014).
73. Moon, H. & Min, D. Assessing energy efficiency and the related policy implications for energy-intensive firms in Korea: DEA approach. *Energy* **133**, 23–34, <https://doi.org/10.1016/j.energy.2017.05.122> (2017).
74. Yang, Z. & Wei, X. The measurement and influences of China's urban total factor energy efficiency under environmental pollution: based on the game cross-efficiency DEA. *J. Clean. Prod.* **209**, 439–450, <https://doi.org/10.1016/j.jclepro.2018.10.271> (2019).
75. Zhao, H., Guo, S. & Zhao, H. Provincial energy efficiency of China quantified by three-stage data envelopment analysis. *Energy* **166**, 96–107, <https://doi.org/10.1016/j.energy.2018.10.063> (2019).
76. Chen, Y., Cook, W. D., Du, J., Hu, H. & Zhu, J. Bounded and discrete data and Likert levels in data envelopment analysis: application to regional energy efficiency in China. *Ann. Oper. Res.* **255**, 347–366, <https://doi.org/10.1007/s10479-015-1827-3> (2017).
77. Jebali, E., Essid, H. & Khraief, N. The analysis of energy efficiency of the Mediterranean countries: a two-stage double bootstrap DEA approach. *Energy* **134**, 991–1000, <https://doi.org/10.1016/j.energy.2017.06.063> (2017).
78. Wang, K. L., Ding, L. L., Wang, J. M. & Yang, M. Analysis of provincial total-factor air pollution efficiency in China by using context-dependent slacks-based measure considering undesirable outputs. *Nat. Hazards* **104**, 1899–1921, <https://doi.org/10.1007/s11069-020-04253-9> (2020). b.
79. Wu, X. *et al.* A study of allocative efficiency of PM_{2.5} emission rights based on a zero sum gains data envelopment model. *J. Clean. Prod.* **113**, 1024–1031, <https://doi.org/10.1016/j.jclepro.2015.11.025> (2016).
80. Lin, B. & Benjamin, N. I. Green development determinants in China: a non-radial quantile outlook. *J. Clean. Prod.* **162**, 764–775, <https://doi.org/10.1016/j.jclepro.2017.06.062> (2017).
81. Zhang, N., Wu, Y. & Choi, Y. Is it feasible for China to enhance its air quality in terms of the efficiency and the regulatory cost of air pollution? *Sci. Tot. Environ.* **709**, 136149, <https://doi.org/10.1016/j.scitotenv.2019.136149> (2020b).
82. Zhang, Y., Li, X., Jiang, F., Song, Y. & Xu, M. Industrial policy, energy and environment efficiency: evidence from Chinese firm-level data. *J. Environ. Manag.* **260**, 110123, <https://doi.org/10.1016/j.jenvman.2020.110123> (2020c).
83. Cheng, Z., Li, L., Liu, J. & Zhang, H. Total-factor carbon emission efficiency of China's provincial industrial sector and its dynamic evolution. *Renew. Sustain. Energy Rev.* **94**, 330–339, <https://doi.org/10.1016/j.rser.2018.06.015> (2018).
84. Xie, H. *et al.* Spatial-temporal disparities and influencing factors of total-factor green use efficiency of industrial land in China. *J. Clean. Prod.* **207**, 1047–1058, <https://doi.org/10.1016/j.jclepro.2018.10.087> (2019).
85. Nie, Y., Cheng, D. & Liu, K. The effectiveness of environmental authoritarianism: evidence from China's administrative inquiry for environmental protection. *Energy Econ.* **88**, 104777, <https://doi.org/10.1016/j.eneco.2020.104777> (2020).
86. Zhou, P., Ang, B. W. & Wang, H. Energy and CO₂ emission performance in electricity generation: a non-radial directional distance function approach. *Eur. J. Oper. Res.* **221**, 625–635, <https://doi.org/10.1016/j.ejor.2012.04.022> (2012).
87. Zhang, N., Kong, F., Choi, Y. & Zhou, P. The effect of size-control policy on unified energy and carbon efficiency for Chinese fossil fuel power plants. *Energy Policy* **70**, 193–200, <https://doi.org/10.1016/j.enpol.2014.03.031> (2014).
88. Lin, B. & Chen, X. Environmental regulation and energy-environmental performance: empirical evidence from China's non-ferrous metals industry. *J. Environ. Manag.* **269**, 110722, <https://doi.org/10.1016/j.jenvman.2020.110722> (2020).
89. Meng, F., Su, B., Thomson, E., Zhou, D. & Zhou, P. Measuring China's regional energy and carbon emission efficiency with DEA models: a survey. *Appl. Energy* **183**, 1–21, <https://doi.org/10.1016/j.apenergy.2016.08.158> (2016).
90. Li, M. Decomposing the change of CO₂ emissions in China: a distance function approach. *Ecol. Econ.* **70**, 77–85, <https://doi.org/10.1016/j.ecolecon.2010.07.034> (2010).
91. Bor, Y. J. Consistent multi-level energy efficiency indicators and their policy implications. *Energy Econ.* **30**, 2401–2419, <https://doi.org/10.1016/j.eneco.2007.11.005> (2008).
92. Duro, J. A. & Padilla, E. Inequality across countries in energy intensities: an analysis of the role of energy transformation and final energy consumption. *Energy Econ.* **33**, 474–479, <https://doi.org/10.1016/j.eneco.2010.12.008> (2011).
93. Cheng, S. L. *et al.* Potential role of fiscal decentralization on interprovincial differences in CO₂ emissions in China. *Environ. Sci. Technol.* **55**, 813–822, <https://doi.org/10.1021/acs.est.0c04026> (2021).
94. Cheng, S. L., Wang, P., Chen, B. Y. & Fan, W. Decoupling and decomposition analysis of CO₂ emissions from government spending in China. *Energy* **243**, 122741, <https://doi.org/10.1016/j.energy.2021.122741> (2022).

Acknowledgements

This work was supported by the National Key Social Science Foundation of China [grant number 21ATJ008].

Author contributions

J.C. led the project and provided the idea for this research. J.L., J.Q., M.G., K.L. and C.X. collected the raw data, calculated and assembled the data and prepared the initial manuscript. S.C. revised the manuscript. J.C. and S.C. designed the research. M.G., J.L. and C.X. provided technical support.

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/s41597-022-01240-6>.

Correspondence and requests for materials should be addressed to S.C.

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 4.0 International License, which permits use, sharing, adaptation, 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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022