






OPEN

DATA DESCRIPTOR

# County-level CO<sub>2</sub> emissions and sequestration in China during 1997–2017

Jiandong Chen<sup>1</sup>, Ming Gao<sup>1</sup>, Shulei Cheng<sup>1</sup>, Wenxuan Hou<sup>2,3</sup>, Malin Song<sup>4</sup>, Xin Liu<sup>5</sup>, Yu Liu<sup>6,7</sup>  & Yuli Shan<sup>8</sup>  

With the implementation of China's top-down CO<sub>2</sub> emissions reduction strategy, the regional differences should be considered. As the most basic governmental unit in China, counties could better capture the regional heterogeneity than provinces and prefecture-level city, and county-level CO<sub>2</sub> emissions could be used for the development of strategic policies tailored to local conditions. However, most of the previous accounts of CO<sub>2</sub> emissions in China have only focused on the national, provincial, or city levels, owing to limited methods and smaller-scale data. In this study, a particle swarm optimization-back propagation (PSO-BP) algorithm was employed to unify the scale of DMSP/OLS and NPP/VIIRS satellite imagery and estimate the CO<sub>2</sub> emissions in 2,735 Chinese counties during 1997–2017. Moreover, as vegetation has a significant ability to sequester and reduce CO<sub>2</sub> emissions, we calculated the county-level carbon sequestration value of terrestrial vegetation. The results presented here can contribute to existing data gaps and enable the development of strategies to reduce CO<sub>2</sub> emissions in China.

## Background & Summary

As one of the largest carbon emitters globally, China has pledged to reach the peak of its carbon emissions by 2030<sup>1–3</sup>, and significant effort has been put into developing a sustainable economy<sup>4–6</sup>. An increasing number of studies have focused on topics such as CO<sub>2</sub> emissions accounts<sup>7–9</sup>, driving forces of CO<sub>2</sub> emissions<sup>10,11</sup>, forecasting future emissions, and more<sup>12,13</sup>. However, most of this research has been conducted at the national, provincial<sup>14–16</sup>, or city level<sup>17,18</sup>. Actually, even within the same province or the same prefecture-level city, there can be obvious differences in CO<sub>2</sub> emissions among counties. Research at the county-level is important for capturing regional heterogeneity and developing policies that can effectively lead to reductions in CO<sub>2</sub> emissions.

Therefore, records of county-level CO<sub>2</sub> emissions are required, which could help to fill the gaps in China's CO<sub>2</sub> emission data and could be used for the development of strategic policies that propose county specific emission reduction actions. However, very few studies have investigated the county-level emissions in China owing to low availability of data sources, and the corresponding studies have limitations in methodology, time span, and geographical coverage. These studies can be classified into two main categories:

- (1) Previous studies calculated the county-level CO<sub>2</sub> emissions on the basis of the published energy use data<sup>19,20</sup>. For example, Cai *et al.*<sup>19</sup> estimated CO<sub>2</sub> emissions of 16 counties in Tianjin, China in 2007 based on the construction of a CO<sub>2</sub> emissions grid, which was derived from the spatial distributions of energy use from the industrial sector, agricultural sector, and residential sector. Additionally, Guan *et al.*<sup>20</sup> adopted the CO<sub>2</sub> emission coefficients provided by the IPCC and 11 types of energy use, such as coal, coke, coal gas,

<sup>1</sup>School of Public Administration, Southwestern University of Finance and Economics, Chengdu, China. <sup>2</sup>School of Finance, Shanghai Lixin University of Accounting and Finance, Shanghai, China. <sup>3</sup>University of Edinburgh Business School, University of Edinburgh, 29 Buccleuch Place, Edinburgh, United Kingdom. <sup>4</sup>School of Statistics and Applied Mathematics, Anhui University of Finance and Economics, Bengbu, China. <sup>5</sup>Curtin University Sustainability Policy Institute, School of Design and the Built Environment, Curtin University, Perth, Australia. <sup>6</sup>Institutes of Science and Development, Chinese Academy of Sciences, Beijing, 100190, China. <sup>7</sup>School of Public Policy and Management, University of Chinese Academy of Sciences, Beijing, 100049, China. <sup>8</sup>Integrated Research on Energy, Environment and Society (IREES), Energy and Sustainability Research Institute Groningen, University of Groningen, Groningen, 9747 AG, Netherlands.  e-mail: [liuyu@casipm.ac.cn](mailto:liuyu@casipm.ac.cn); [y.shan@rug.nl](mailto:y.shan@rug.nl)

and natural gas, to calculate the CO<sub>2</sub> emissions of 18 counties in the Ningxia Hui Autonomous Region during 1991–2011. The county-level CO<sub>2</sub> emissions estimated by these studies relied on the published energy use data. Hence, the estimated emissions are always limited to a small coverage and short period, owing to the insufficient and often incomparable county-level energy use information.

- (2) When dealing with the lack of data in cities and counties, some studies used a top-down approach to calculate their CO<sub>2</sub> emissions<sup>21,22</sup>. Although some scholars have used several socioeconomic variables (such as urbanization ratio and population density) as the cutting index<sup>23</sup>; also the nighttime light data are always selected as proxies to downscale the total energy-related CO<sub>2</sub> emissions<sup>21,22</sup>. For example, Meng *et al.*<sup>21</sup> used the brightness data from Defense Meteorological Satellite Program/Operational Linescan System (DMSP/OLS) satellite imagery to downscale and estimate CO<sub>2</sub> emissions of 287 prefecture-level cities in China during 1995–2010. Similarly, Su *et al.*<sup>22</sup> calculated city-level CO<sub>2</sub> emissions in China during 1992–2012 based on the DMSP/OLS imagery. In the present case, there are primarily two reasons for selecting lighting data: first, the brightness distribution on the surface of the earth at night is closely related to human activities; second, with advancements in remote sensing interpretation technologies, a wider time-span and coverage of global lighting data can be extracted and used. In addition, considering the advantages above, nighttime light data have also been widely accepted and applied in other research fields, such as energy use forecasting, gross domestic productivity evaluation, and population distribution estimation<sup>24–26</sup>.

Thus, based on the available nighttime light data provided by DMSP/OLS images, China's energy-related CO<sub>2</sub> emissions can be calculated at micro-level administration. However, because the DMSP/OLS images are only available up to 2013, the research period was limited. Additionally, even though Suomi National Polar-Orbiting Partnership/Visible Infrared Imaging Radiometer Suite (NPP/VIIRS) images provide another source of nighttime light brightness data after 2012, the evident gaps between the two sets of satellites' data have hindered the construction of long-term nighttime light data sets and calculations of CO<sub>2</sub> emissions. Thus, several studies have made attempts to unify the two sets of satellite data<sup>27–29</sup>. But matching the results proved difficult, and other problems involving discontinuity and saturation were encountered. Hence, there is room for further improvements.

In addition, existing literature on CO<sub>2</sub> emissions reduction has only focused on the energy-related carbon emissions, and the influence of carbon sequestration of vegetation have always been ignored. With regard to the concept of plant carbon sequestration capacity, it is a natural carbon sequestration process, which directly counteracts the processes of emitting CO<sub>2</sub> into the atmosphere. In addition, the natural processes mainly originate from vegetation net primary productivity (NPP) or net ecosystem productivity—vegetation in the ecosystem absorb CO<sub>2</sub> from the air, produce carbohydrates such as glucose through photosynthesis, and release oxygen. Actually, vegetation has a significant effect on CO<sub>2</sub> sequestration, and can account for a major part of the CO<sub>2</sub> emissions associated with energy use. Among them, terrestrial vegetation plays a significant role in CO<sub>2</sub> sequestration, and methods of estimation of its sequestration capacity have advanced and have been widely adopted<sup>29</sup>. Therefore, we also estimate the county-level carbon sequestration values of terrestrial vegetation, which facilitates more comprehensive research on reducing CO<sub>2</sub> emissions in China and evaluating sustainable development<sup>30</sup>.

Thus, our present study makes the following marginal contributions to this field of research: (1) we developed a new model and employed a particle swarm optimization-back propagation (PSO-BP) algorithm to unify the scale of DMSP/OLS and NPP/VIIRS images during 1997–2017, which obtain superior fitting effects than those of previous studies based on original models and normal econometrics; (2) we adopted the PSO-BP algorithm to downscale the provincial energy-carbon emissions based on the nighttime light data, and calculated 2,735 county-level energy-related carbon emissions during 1997–2017; and (3) we estimated the corresponding county-level carbon sequestration values of terrestrial vegetation, which is an issue that has rarely been considered in previous studies and plays a significant role in CO<sub>2</sub> emissions mitigation.

## Methods

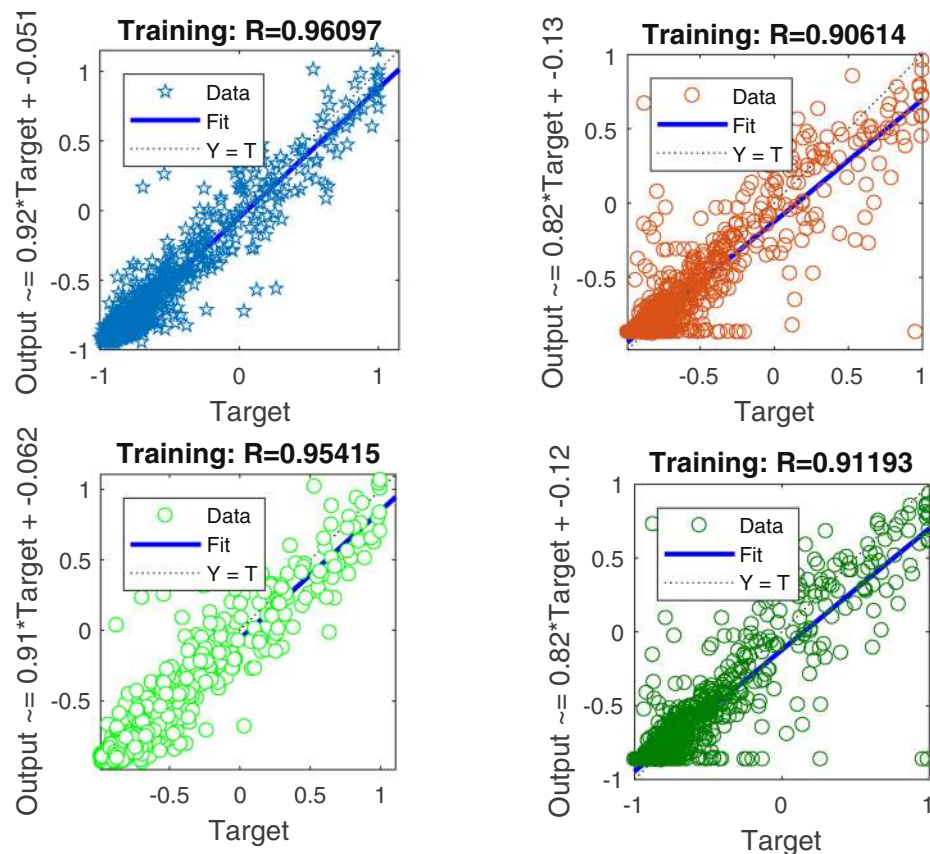
**Study areas and data preprocessing.** Since China is one of the largest CO<sub>2</sub> emitter globally, our aim was to facilitate the determination of carbon reduction status in China and to address current data gaps in China. In addition, our research results could facilitate energy saving and emissions reduction activities and efforts in other countries, especially in other developing countries.

As the most basic governmental unit in China, counties should play important roles in the implementation of emission reduction policies from the central, provincial, and municipal governments. Therefore, we selected county-level CO<sub>2</sub> emissions as a research focus. Our study areas cover 2,735 counties of 30 provinces in China mainland (excluding Tibet, Hong Kong, Macau, and Taiwan) based on the accessibility of data sets, which cover approximately 87% of China's land area, over 90% of the population, and 90% of the GDP.

Our estimated and provided data sets include energy-related CO<sub>2</sub> emissions (1997–2017) and carbon sequestration values for terrestrial vegetation (2000–2017). The estimated boundary for energy-related CO<sub>2</sub> emissions lies in Scope 2 proposed by the Climate Action Registry (CAR) and other institutes, which is consistent with Meng *et al.*<sup>21</sup>, Su *et al.*<sup>22</sup>, and Zhao *et al.*<sup>28</sup>.

Three satellite data sets were used in our study: two types of nighttime light data (provided by DMSP/OLS<sup>31,32</sup> and NPP/VIIRS<sup>33</sup> images) and net primary productivity data of terrestrial vegetation (which were provided by the MODIS NPP products<sup>34</sup>). Considering that there are several problems in the satellite images, such as discontinuities, the white noise, and fill values, these data sets need to be pre-processed before they can be further used.

The DMSP/OLS images were from the period of 1992–2013, and these data had problems stemming from discontinuities, saturation, and incomparability. Hence, we adopted several methods including inter-calibration, radiometric calibration, intra-annual composition, and inter-annual series correction methods proposed by previous studies to obtain continuous and stable DMSP/OLS images<sup>27,28,35</sup>.

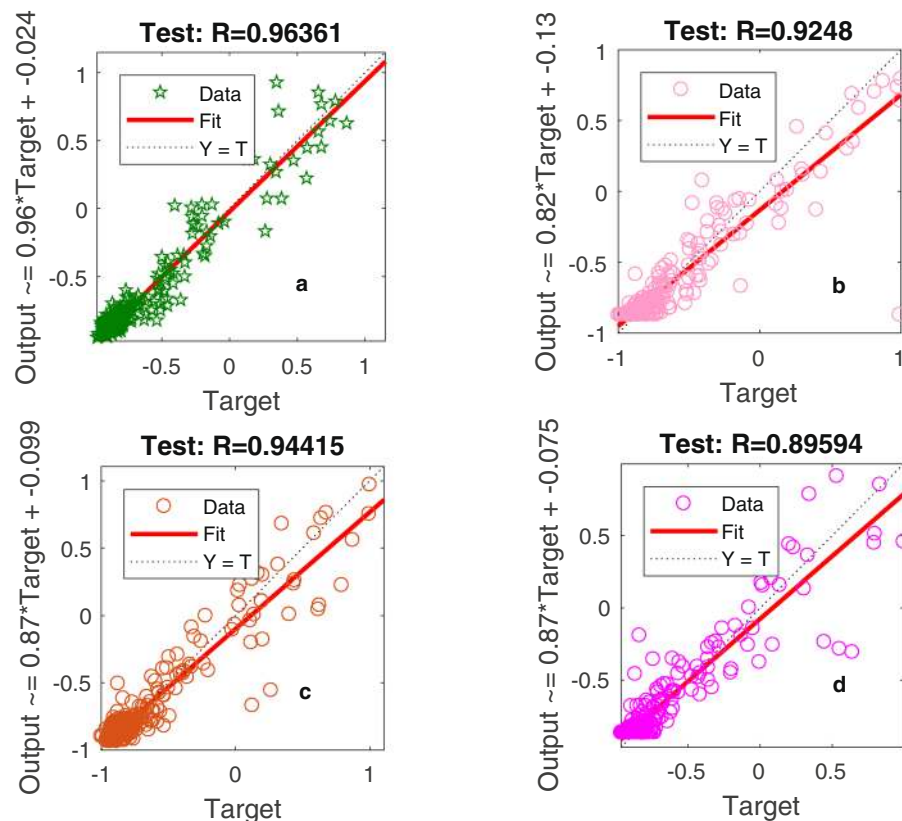


**Fig. 1** Training results for the mean pixel values in 2013. Data represent (a) the results with the supplementary input parameters based on the particle swarm optimization-back propagation (PSO-BP) algorithm, (b) the results for only the mean pixel values of nighttime light values based on the PSO-BP 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.

We used the monthly NPP/VIIRS images, an approach consistent with previous studies<sup>26,27</sup>. Additionally, because of the influence of stray light pollution, lighting data in the mid–high latitudes of China in summer showed large errors; thus, we removed the images from June to August, and we used the remaining monthly data to synthesize the annual data. Then, we applied a Gaussian low-pass filter with a window size of  $5 \times 5$  to mitigate the NPP/VIIRS images' spatial variability and smooth the data to better match the DMSP/OLS images<sup>36</sup>. The  $\sigma$  was set as 1.75 in accordance with studies by Li *et al.*<sup>37</sup> and Zheng *et al.*<sup>38</sup> Moreover, to further reduce the white noise of NPP/VIIRS images, we replaced the negative values with zero and set a threshold of  $0.3 \text{ nW}\cdot\text{m}^{-2}\cdot\text{sr}^{-1}$  in the annual images, which is consistent with earlier work<sup>36</sup>.

The MOD17A3 products provided by the National Aeronautics and Space Administration (NASA) have fill values and need to be multiplied by a 0.0001 conversion factor. Next, following the user guides<sup>39,40</sup>, we obtained the net primary productivity data. Finally, based on the conversion coefficient (i.e.,  $1.62/0.45$ ) used by Chen *et al.*<sup>30</sup>, we obtained the carbon sequestration values of terrestrial vegetation, including Evergreen Needleleaf Forest (ENF), Evergreen Broadleaf Forest (EBF), Deciduous Needleleaf Forest (DNF), Deciduous Broadleaf Forest (DBF), Mixed forests (MF), Closed Shrublands (CShrub), Open Shrublands (OShrub), Woody Savannas (WSavanna), Savannas (Savanna), grassland (Grass), and Croplands (Crop). Based on vector cutting, we obtained the county-level nighttime light values and carbon sequestration values of vegetation. The vector map of county-level cities in China in 2015 was derived from the National Geomatics Center of China, which is consistent with the work of Lv *et al.*<sup>27</sup> and Zhao *et al.*<sup>28</sup>.

**Inter-calibration between DMSP/LOS and NPP/VIIRS Based on PSO-BP.** Because the DMSP/OLS and NPP/VIIRS images were derived from different types of satellites, there are evident gaps in the two sets of data. Specifically, there are discrepancies caused by various factors such as the use of different sensors, different spatial resolutions, different spread functions, and so on<sup>36</sup>. However, the mechanisms for explaining the differences remain like a “Black Box,” and any fixed functional form for the inter-calibration between DMSP/OLS and NPP/VIIRS data may fail to produce a good match between the two sets of data and lead to large errors. Therefore, in the present study, we used an artificial neural network (ANN) to explore the relationship between the DMSP/OLS and NPP/VIIRS data rather than conventional econometric methods, because the conventional methods often fail to model the non-linear relationships<sup>41</sup>.



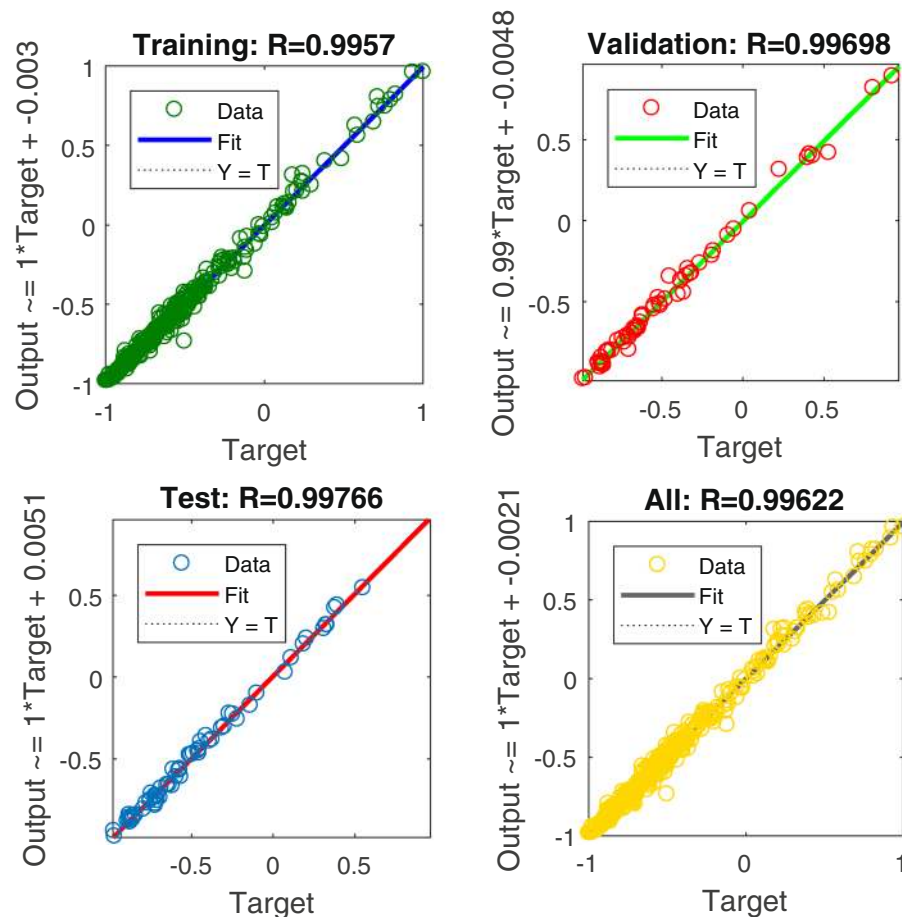
**Fig. 2** Test results for the mean pixel values in 2013. Data represent (a) the results with the supplementary input parameters based on the particle swarm optimization-back propagation (PSO-BP) algorithm, (b) the results with only the mean pixel values of nighttime light values based on the PSO-BP 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.

Variables	LLC		IPS	
	Level	First difference	Level	First difference
sdn	0.7541	-8.0742***	12.3274	-7.2143***
	(0.7746)	(0.0000)	(1.0000)	(0.0000)
c	-2.0716**	-5.2931***	7.0507	-8.2347***
	(0.0191)	(0.0000)	(1.0000)	(0.0000)
Pedroni test	Panel PP statistic	Panel ADF statistic	Group PP statistic	Group ADF statistic
	2.4199***	5.2237***	1.2318	3.7719***
	(0.0078)	(0.0000)	(0.1090)	(0.0001)

**Table 1.** Results for the panel unit root tests and co-integration tests. Notes: (1) Values in parentheses are the p-values; LLC denotes the Levin, Lin, and Chu t test, while IPS denotes the Im, Pesaran, and Shin Wald statistic test. (2) The LLC and IPS tests for all series include an intercept term. (3) The \*\* and \*\*\* symbols denote rejection of the null hypotheses of a unit root at the 5% and 1% significance level, respectively.

Additionally, because the back-propagation (BP) algorithm has performed well in previous studies for constructing regressions and obtaining local optimistic results<sup>42,43</sup>, the BP algorithm was adopted in this research. However, given that the BP algorithm can lead to data at local extremes and training failures, we also combined the BP algorithm with particle swarm optimization (PSO)—PSO has shown great potential in exploring the global optimistic results<sup>44,45</sup>.

As for the input parameters, we followed the approach of Zhao *et al.*<sup>28</sup> and selected the county-level mean pixel values of NPP/VIIRS in 2013 ( $V$ ) as the input. Given the geographical heterogeneity of individual data in mainland China, we made use of the minimum boundary method to obtain each county's central geographic coordinates and used Arcmap 10.5 to obtain the area of each county. Then, we selected the central geographic coordinates ( $X$  and  $Y$ ) and the area of each county ( $A$ ) as the supplementary input parameters, which greatly improved the matching accuracy of the two sets of data and reduced errors. To enhance the accuracy of



**Fig. 3** Training and test results for the relationship between provincial carbon emissions and the sum of digital number (DN) values.

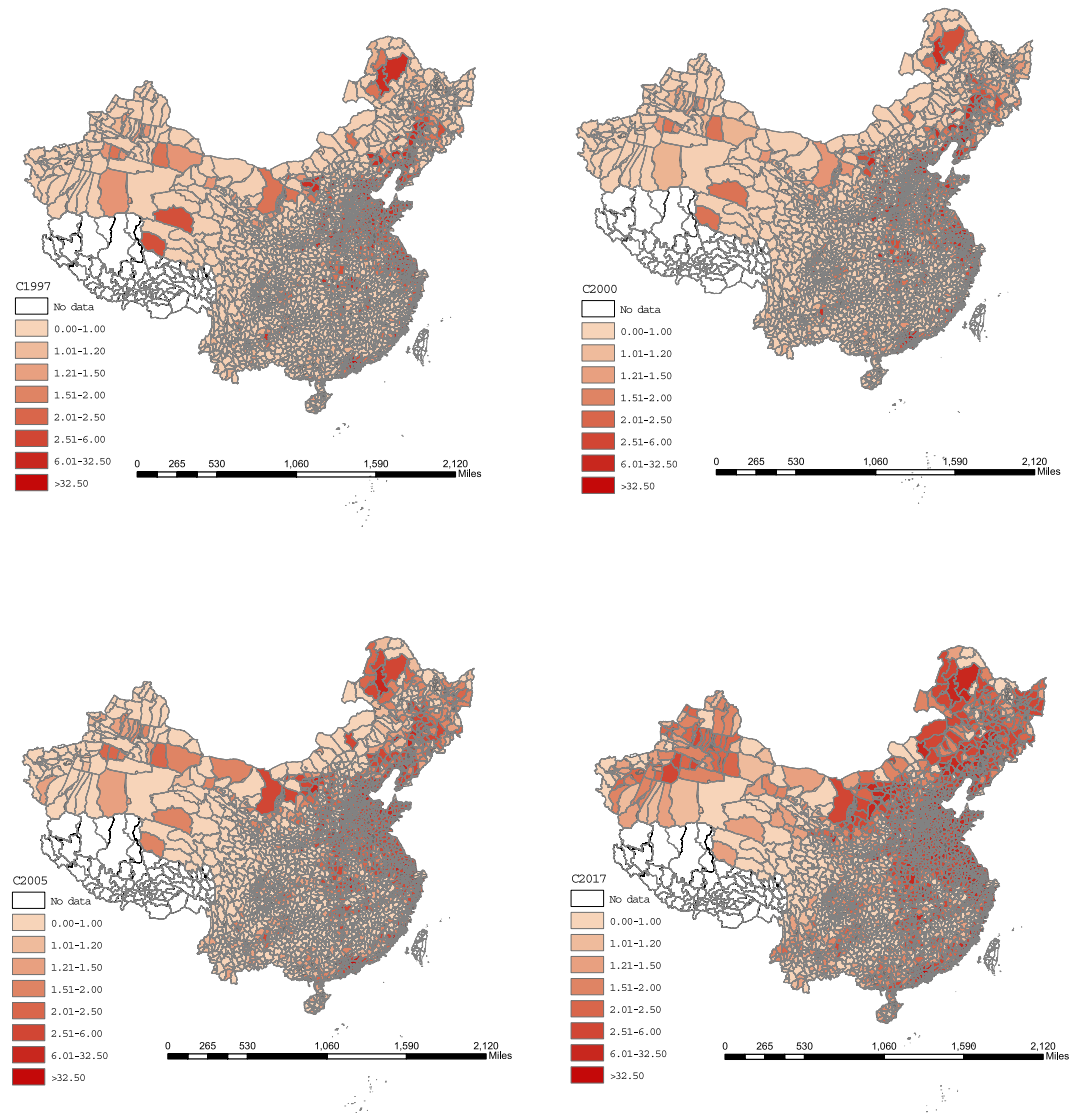
modelling, we use the logarithmic form of the input parameters according to the suggestion of Li *et al.*<sup>37</sup> In addition, with regard to the output parameter, we select the county-level mean pixel values of DMSP/OLS in 2013 (*D*).

Based on initializing the ANN weights with the PSO technique, we set values of  $C_1$  and  $C_2$  both as 2.0, the maximum iteration number as 50, and the population size as 20<sup>45,46</sup>. Additionally, the structure of the model was set as one hidden layer with five nodes in the hidden layer, which is consistent with the work of Mohamad *et al.*<sup>45</sup> The total number of samples was 2,826. Among these, 2,000 samples were randomly selected as training samples, whereas the other 826 samples were used as testing samples. The calculation procedure for the PSO-BP model was consistent with the work of Mohamad *et al.*<sup>45</sup> and Yin *et al.*<sup>47</sup>.

To test the validity of the PSO-BP algorithm and the proposed supplementary input parameters, we also trained the BP algorithm and corresponding algorithm without the supplementary input parameters as control groups. The best training results are presented in Fig. 1. As shown in Fig. 1, all of the correlation coefficient values for the training results of mean pixel values in 2013 were more than 0.9, which indicated that the ANN was advantageous for identifying the potential relationship between DMSP/OLS and NPP/VIIRS data. Among the results, it was evident that the models that considered geographic coordinates and area (i.e., panels a and c) showed comparably better fitting effects than models that only used the county-level mean pixel values of NPP/VIIRS data as input parameters (i.e., panels b and d). Additionally, the correlation coefficient values for the PSO-BP algorithm were higher than those for the BP algorithm, thus indicating that the PSO-BP algorithm was better for determining the potential matching relationship between DMSP/OLS and NPP/VIIRS images.

Figure 2 shows the test performances of the four models, and these results can be used to identify the fitting effects of each model. The highest correlation coefficient value of 0.96361 in model *a* for the testing dataset indicated that the proposed PSO-BP algorithm reliably matched the DMSP/OLS images with NPP/VIIRS images in later years (e.g., 2014 and 2015). The correlation coefficient,  $R^2$ , based on our method, was 0.955, which was significantly higher than values obtained previously, including the 0.8354 of Lv *et al.*<sup>27</sup>, 0.9154 of Zhao *et al.*<sup>28</sup>, and 0.91 of Li *et al.*<sup>36</sup>.

Furthermore, the matching work has not yet been completed. Although the correlation coefficient is close to 1, there were evident and unavoidable faults in some counties in 2013 (i.e., DMSP/OLS data in 2013) and 2014 (i.e., converted NPP/VIIRS data in 2014), which also exist in previous studies. Therefore, we make use of the annual increase amounts of converted NPP/VIIRS data during 2013–2017 to obtain the final simulated DMSP/OLS data during 2014–2017, avoiding the shortcoming of faults and discontinuities in some regions during 2013–2014.



**Fig. 4** China mainland's county-level CO<sub>2</sub> emissions from energy combustion in 1997, 2000, 2005 and 2017 (unit: million tons).

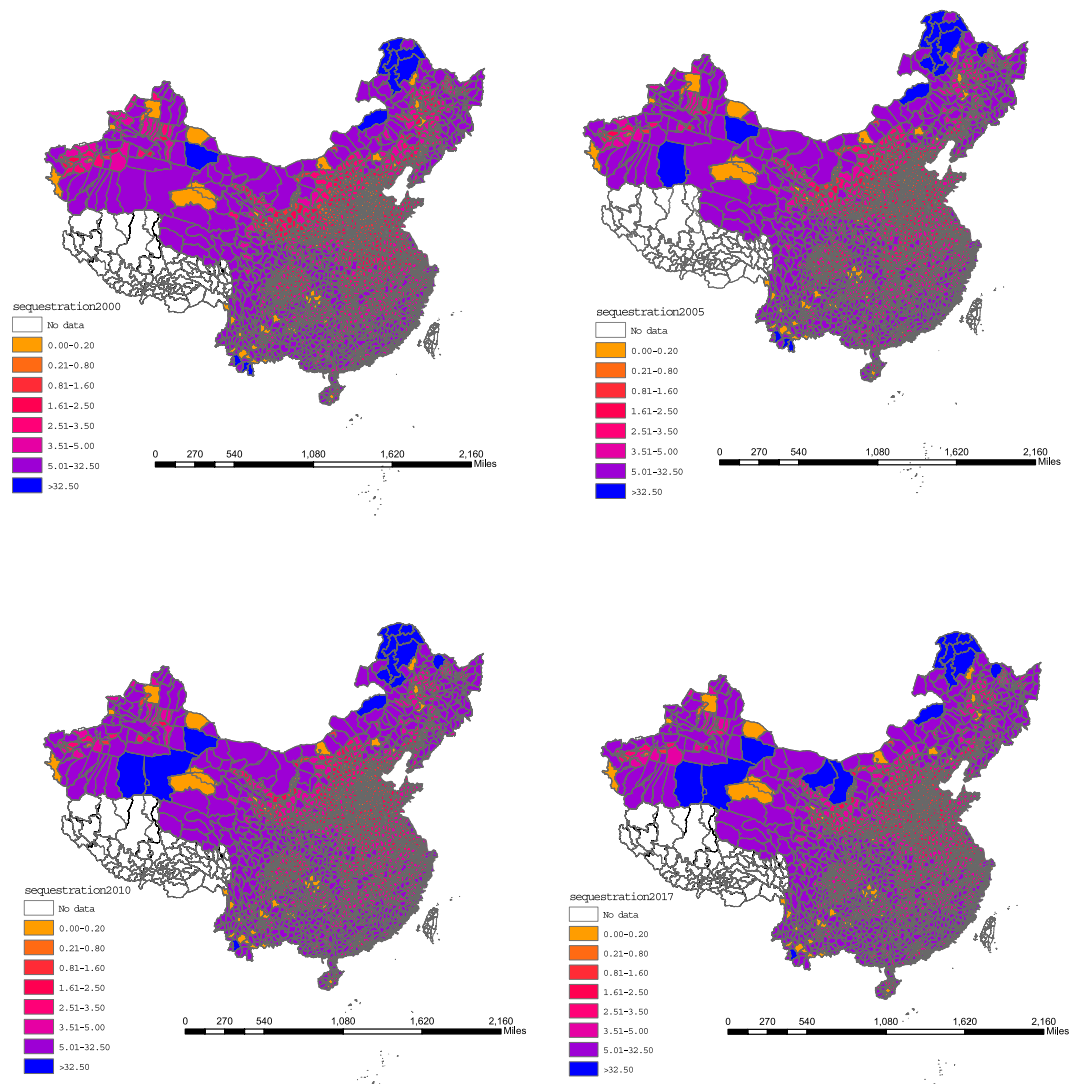
In summary, based on the PSO-BP algorithm, we could confidently convert the scale of NPP/VIIRS data during 2013–2017 to the scales of DMSP/OLS data and obtain stable and continuous county-level nighttime light data during 1997–2017, which lay a foundation for further calculations of county-level CO<sub>2</sub> emissions.

**Calculation of CO<sub>2</sub> emissions based on satellite data.** Because provincial energy balance tables were available and there was a lack of energy use data for various counties, we first established the relationship between provincial CO<sub>2</sub> emissions and nighttime light data (i.e., the sum of the DN values) in this study; then, the sum of the DN values was used as a proxy to estimate the county-level carbon emissions.

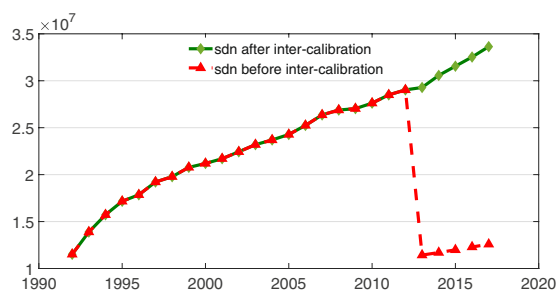
First, the estimations of provincial CO<sub>2</sub> emissions were carried out based on the following method provided by the Intergovernmental Panel on Climate Change (IPCC), which has been widely adopted<sup>4,48,49</sup>:

$$C_E^t = \sum_{i=1}^{30} C_{Direct,i}^t = \sum_{i=1}^{30} \sum_{j=1}^{17} \left[ E_{ij}^t \times LCV_{ij}^t \times CC_{ij}^t \times COF_{ij}^t \times \frac{44}{12} \right] \quad (1)$$

where  $C_{E,i}^t$  represents the provincial CO<sub>2</sub> emissions from energy use (unit: million tons);  $E_{ij}^t$  represents the  $j^{\text{th}}$  type of energy use in province  $i$ ;  $LCV_{ij}^t$  is the low calorific value of the  $j^{\text{th}}$  energy consumption;  $CC_{ij}^t$  is the carbon content of the  $j^{\text{th}}$  energy source; and  $COF_{ij}^t$  is the carbon oxidation factor of the  $j^{\text{th}}$  energy source. In addition, 17 types of fossil fuel used are considered, including raw coal, cleaned coal, other washed coal, briquettes, gangue, coke, coke oven gas, blast furnace gas, converter gas, other gases, other coking products, crude oil, gasoline, kerosene, diesel oil, fuel oil, naphtha, lubricants, paraffin, white spirit, bitumen asphalt, petroleum coke, other petroleum products, liquefied petroleum gas (LPG), refinery gas, and natural gas.



**Fig. 5** China mainland's county level spatial and temporal patterns of carbon sequestration capacity of terrestrial vegetation in 2000, 2005, 2010, and 2017 (unit: million tons).



**Fig. 6** Total trend for the sum of DN values (*sdn*) during 1992–2017 (unit of the sum of DN values:  $10^7$ ).

Subsequently, to avoid spurious regression problems, we adopted the unit root test to verify the relationship between provincial  $\text{CO}_2$  emissions  $c_t$  and sum of DN values  $sdn_t$ . The results are presented in Table 1. It was evident that the sum of DN values and carbon emissions had to be processed at the same time with Eq. (1). Then, the co-integration Pedroni test was adopted<sup>50</sup>, which has been widely accepted in the field of econometrics<sup>51–53</sup>. The majority of tests led to the rejection of the null hypothesis of no co-integration, thus suggesting that there was significant co-integration among the provincial carbon emissions and sum of DN values.

Furthermore, considering that the relationship between provincial  $\text{CO}_2$  emissions and nighttime light data is non-linear, the normal econometric methods may lead to relatively high errors<sup>27</sup>; here, we employed the PSO-BP

Year	Model (1)			Model (2)		
	Slope	AIC values	R <sup>2</sup>	Slope	AIC values	R <sup>2</sup>
1997	0.0036 (0.0000)	17.03	0.87	0.0005 (0.0000)	12.62	0.92
1998	0.0038 (0.0000)	17.13	0.87	0.0005 (0.0000)	12.74	0.91
1999	0.0039 (0.0000)	17.23	0.88	0.0005 (0.0000)	12.86	0.91
2000	0.0043 (0.0000)	17.45	0.88	0.0006 (0.0000)	12.96	0.92
2001	0.0047 (0.0000)	17.65	0.88	0.0006 (0.0000)	13.08	0.92
2002	0.0051 (0.0000)	17.83	0.88	0.0007 (0.0000)	13.30	0.93
2003	0.0058 (0.0000)	18.08	0.89	0.0008 (0.0000)	13.70	0.92
2004	0.0068 (0.0000)	18.41	0.89	0.0009 (0.0000)	13.97	0.92
2005	0.0080 (0.0000)	18.75	0.90	0.0010 (0.0000)	14.20	0.92
2006	0.0091 (0.0000)	19.04	0.90	0.0011 (0.0000)	14.42	0.93
2007	0.0104 (0.0000)	19.37	0.90	0.0012 (0.0000)	14.61	0.93
2008	0.0122 (0.0000)	19.65	0.91	0.0012 (0.0000)	14.70	0.93
2009	0.0132 (0.0000)	19.90	0.90	0.0013 (0.0000)	14.85	0.93
2010	0.0154 (0.0000)	20.23	0.90	0.0015 (0.0000)	15.15	0.93
2011	0.0177 (0.0000)	20.53	0.90	0.0016 (0.0000)	15.34	0.93
2012	0.0191 (0.0000)	20.75	0.90	0.0016 (0.0000)	15.41	0.93
2013	0.0195 (0.0000)	21.11	0.88	0.0016 (0.0000)	15.87	0.91
2014	0.0201 (0.0000)	21.26	0.88	0.0016 (0.0000)	15.90	0.91
2015	0.0205 (0.0000)	21.42	0.88	0.0016 (0.0000)	15.93	0.92
2016	0.0214 (0.0000)	21.66	0.87	0.0016 (0.0000)	16.04	0.91
2017	0.0239 (0.0000)	21.94	0.87	0.0017 (0.0000)	16.16	0.91

**Table 2.** Validity test results for the spatial patterns of nighttime light data based on GDP and power consumption. Notes: The values in parentheses are p-values. Model (1) represents the linear regression of the provincial cross-sectional GDP with the sum of DN values; Model (2) represents the linear regression of the provincial cross-sectional power consumption with the sum of DN values (*sdn*). AIC denotes the Akaike information criterion.

algorithm to fit and train the relationship. We selected the sum of DN values, dummy variables of identity, and year as the input parameters, and the provincial CO<sub>2</sub> emissions were the output parameter. In addition, the other initialized parameters were consistent with those discussed in the earlier section on the inter-calibration. The results are presented in Fig. 3.

The test and training results showed great fitting effects, which were indicative of the high effectiveness of the algorithm. Notably, the coefficient of determination R<sup>2</sup> of 0.9895 was higher than the 0.94 of Meng *et al.*<sup>21</sup> and 0.6922 of Lv *et al.*<sup>27</sup> estimated based on conventional econometric methods. Then, based on the concept of the top-down method and a DN value-based weighted-average strategy<sup>21,22,41</sup>, we obtained the county-level carbon emissions.

## Data Records

A total of 5470 data records (county-level CO<sub>2</sub> emissions caused by energy use and carbon sequestration values of terrestrial vegetation). Among them, there were 2,735 CO<sub>2</sub> emission county data records associated with energy use (1997–2017), and corresponding 2,735 carbon sequestration values associated with terrestrial vegetation (2000–2017).

The present dataset and vector map are made public under Figshare<sup>54</sup>. And the units are million tons. The temporal and spatial changes in the China mainland's county-level energy-related CO<sub>2</sub> emissions and the carbon sequestration value of terrestrial vegetation are presented in Figs. 4 and 5.

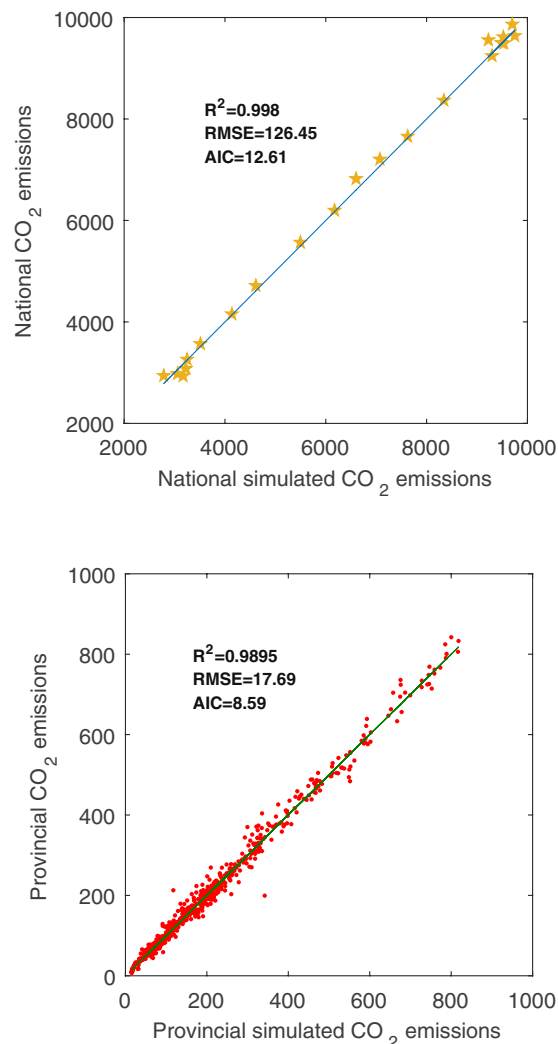
## Technical Validation

**Validity testing for temporal and spatial nighttime light data changes.** On the basis of the inter-calibration method described earlier, we were able to obtain continuous and stable county-level nighttime light DN values during 1997 to 2017, and the sum of DN values is presented in Fig. 6. The red line represents the changes in the sum of DN values before the inter-calibration between DMSP/OLS and NPP/VIIRS images, and the green line represents the changes in the sum of DN values (*sdn*) after the inter-calibration based on the PSO-BP algorithm. Evidently, there was a gap between the scale of DMSP/OLS and NPP/VIIRS images before the matching. Additionally, the trend of our inter-calibrated results continuously increased, which is consistent with previous studies<sup>27,28</sup>.

Subsequently, because nighttime light data tend to be highly consistent with economic output<sup>55,56</sup> and power consumption data<sup>57–59</sup>, we used the provincial cross-sectional gross domestic product (GDP) and power consumption to individually perform linear regressions with the sum of DN values (*sdn*) during 1997–2017. These results are presented in Table 2.

With regard to the Model (1) results shown in Table 2, it was evident that there was a significant positive relationship between the provincial cross-sectional GDP and *sdn* during 1997–2017. All of the R<sup>2</sup> values were





**Fig. 7** Scatter plots of our simulated national and provincial energy-related CO<sub>2</sub> emissions with the CO<sub>2</sub> emissions based on existing literature during 1997–2017 (unit: million tons).

over 0.87, and the AIC values were small, thus implying that inter-calibrated nighttime light data characterized the economic output well. Simultaneously, in regard to the Model (2) results shown in Table 2, the sum of the DN values were evidently consistent with the power consumption data because the slopes were significantly positive, the R<sup>2</sup> values were high, and the AIC values were small.

**Validity testing for CO<sub>2</sub> emissions based on satellite data.** The method of calculation of carbon sequestration value of terrestrial vegetation is consistent with that in Chen *et al.*<sup>30</sup> and, therefore, was deemed a reliable measure. With regard to the validity of energy-related carbon emissions based on the nighttime light data, we made use of the national and provincial energy-related CO<sub>2</sub> emissions provided by existing studies<sup>17,60,61</sup> to conduct a comparison with the summary of our simulated energy-related carbon emissions. These results are presented in Fig. 7. Panels (a) and (b) in Fig. 7 individually show the scatter plots of our simulated national and provincial energy-related CO<sub>2</sub> emissions with the CO<sub>2</sub> emissions based on existing literature during 1997–2017. The results in each graph were highly consistent, thus indicating that the simulated CO<sub>2</sub> emissions based on the nighttime light data are reliable.

**Limitations and future work.** Our datasets have several limitations, which we will address in the future to improve the accuracy of China's county-level emission accounts. First, our estimated county-level CO<sub>2</sub> emissions are based only on the nighttime light data, overlooking other factors such as urbanization rate, population, and earth surface temperature. Secondly, our estimated carbon sequestration values only include carbon sequestration capacity of terrestrial vegetation, without taking into account ocean carbon sequestration capacity<sup>62</sup>.

Therefore, our future work will include two aspects: first, we will combine nighttime light data with other satellite data such as earth surface temperature provided by MOD11A2, and impervious surface data provided by Gong *et al.*<sup>63</sup>, to improve the accuracy of the calculated county-level carbon emissions. We will further analyze the carbon sequestration capacity of mangrove vegetation to enrich the estimated county-level carbon sequestration data.

## Usage Notes

First, the China mainland's 2,735 county-level energy-related carbon emissions data is provided and has the advantages of wide coverage and long time-span. The data set can help fill the existing data gaps and be further used in future research. For example, scholars can use the data to further analyze the driving forces of the CO<sub>2</sub> emissions at county-level rather than the nation<sup>3</sup>, province<sup>4,5</sup> or prefecture-level<sup>21,22,27</sup>. Additionally, it can be used to further evaluate the emission reductions<sup>19,20</sup> or construct budget allocation of CO<sub>2</sub> emission rights in the county.

Second, considering that vegetation plays a significant role in sequestering and reducing CO<sub>2</sub> emissions, the dataset of the 2,735 county-level carbon sequestration values of vegetation can be combined with our provided county-level energy-related CO<sub>2</sub> emissions. Evidently, the dataset could facilitate further comprehensive analyses and research on China's emissions mitigation<sup>30,64,65</sup>, the gap between CO<sub>2</sub> emissions and carbon sequestration, and comprehensive evaluation of sustainable development<sup>65</sup>.

Additionally, the present study was limited by differences in time spans of the energy-related CO<sub>2</sub> emissions and carbon sequestration values for vegetation. Because of the availability of the original data, the energy-related CO<sub>2</sub> emissions data span 1997–2017, while the carbon sequestration values of vegetation data span 2000–2017. Similarly, we have to point out that our study areas only include China mainland. The units of county-level CO<sub>2</sub> emissions and carbon sequestration values provided are million tons.

## Code availability

The programs used to generate all the results were MATLAB (R2017b) and ArcGIS (10.5). The PSO-BP codes for matching the scales of the nighttime light data and modelling the relationships among the provincial energy-related CO<sub>2</sub> emissions are presented in Suppl. File 1.

Received: 28 August 2020; Accepted: 9 October 2020;

Published online: 12 November 2020

## References

- Pearson, P. N. & Palmer, M. R. Atmospheric carbon dioxide concentrations over the past 60 million years. *Nature* **406**, 695–699 (2000).
- Clark, P. U. *et al.* Consequences of twenty-first-century policy for multi-millennial climate and sea-level change. *Nat. Clim. Change* **6**, 360–369 (2016).
- Liu, Y. *et al.* 'Made in China': a reevaluation of embodied CO<sub>2</sub> emissions in Chinese exports using firm heterogeneity information. *Appl. Energy* **184**, 1106–1113 (2016).
- Liu, Y., Tan, X. J., Yu, Y. & Qi, S. Z. Assessment of impacts of Hubei Pilot emission trading schemes in China – A CGE-analysis using TermCO2 model. *Appl. Energy* **189**, 762–769, <https://doi.org/10.1016/j.apenergy.2016.05.085> (2017).
- Wang, S., Zhou, C., Li, G. & Feng, K. CO<sub>2</sub>, economic growth, and energy consumption in China's provinces: investigating the spatiotemporal and econometric characteristics of China's CO<sub>2</sub> emissions. *Ecol. Indic.* **69**, 184–195 (2016).
- Dubey, R. *et al.* Supply chain agility, adaptability and alignment: Empirical evidence from the Indian auto components industry. *Int. J. Oper. Prod. Man.* **38**, 129–148 (2018).
- Ouyang, X. & Lin, B. An analysis of the driving forces of energy-related carbon dioxide emissions in China's industrial sector. *Renew. Sust. Energ. Rev.* **45**, 838–849, <https://doi.org/10.1016/j.rser.2015.02.030> (2015).
- Lin, B., Chen, Y. & Zhang, G. Technological progress and rebound effect in China's nonferrous metals industry: an empirical study. *Energy Pol.* **109**, 520–529, <https://doi.org/10.1016/j.enpol.2017.07.031> (2017).
- Huang, J., Liu, Q., Cai, X., Hao, Y. & Lei, H. The effect of technological factors on China's carbon intensity: new evidence from a panel threshold model. *Energy Pol.* **115**, 32–42, <https://doi.org/10.1016/j.enpol.2017.12.008> (2018).
- Huang, J., Hao, Y. & Lei, H. Indigenous versus foreign innovation and energy intensity in China. *Renew. Sust. Energ. Rev.* **81**, 1721–1729 (2018).
- Tan, X., Liu, Y., Cui, J. & Su, B. Assessment of carbon leakage by channels: An approach combining CGE model and decomposition analysis. *Energy Econ.* **74**, 535–545 (2018).
- Dong, F., Yu, B. & Pan, Y. Examining the synergistic effect of CO<sub>2</sub> emissions on PM<sub>2.5</sub> emissions reduction: Evidence from China. *J. Clean. Prod.* **223**, 759–771 (2019).
- Chen, J., Gao, M., Mangla, S. K., Song, M. & Wen, J. Effects of technological changes on China's carbon emissions. *Technol. Forecast. Soc.* **153**, 119938 (2020).
- Lu, Y., Liu, Y. & Zhou, M. Rebound effect of improved energy efficiency for different energy types: A general equilibrium analysis for China. *Energy Econ.* **62**, 248–256 (2017).
- Wang, S. & Liu, X. China's city-level energy-related CO<sub>2</sub> emissions: Spatiotemporal patterns and driving forces. *Appl. Energy* **200**, 204–214 (2017).
- Wang, Q., Hang, Y., Su, B. & Zhou, P. Contributions to sector-level carbon intensity change: an integrated decomposition analysis. *Energy Econ.* **70**, 12–25 (2018).
- Shan, Y. *et al.* China CO<sub>2</sub> emission accounts 1997–2015. *Sci. Data* **5**, 170201 (2018).
- Shan, Y., Liu, J., Liu, Z., Shao, S. & Guan, D. An emissions-socioeconomic inventory of Chinese cities. *Sci. Data* **6**, 190027 (2019).
- Cai, B. & Zhang, L. Urban CO<sub>2</sub> emissions in China: spatial boundary and performance comparison. *Energy Pol.* **66**, 557–567 (2014).
- Guan, Y., Kang, L., Shao, C., Wang, P. & Ju, M. Measuring county-level heterogeneity of CO<sub>2</sub> emissions attributed to energy consumption: A case study in Ningxia Hui Autonomous Region, China. *J. Clean. Prod.* **142**, 3471–3481 (2017).
- Meng, L., Graus, W., Worrell, E. & Huang, B. Estimating CO<sub>2</sub> (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 (2014).
- Su, Y. *et al.* China's 19-year city-level carbon emissions of energy consumptions, driving forces and regionalized mitigation guidelines. *Renew. Sust. Energ. Rev.* **35**, 231–243 (2014).
- Jing, Q., Bai, H., Luo, W., Cai, B. & Xu, H. A top-bottom method for city-scale energy-related CO<sub>2</sub> emissions estimation: A case study of 41 Chinese cities. *J. Clean. Prod.* **202**, 444–455 (2018).
- Shortland, A., Christopoulou, K. & Makatsoris, C. War and famine, peace and light? The economic dynamics of conflict in Somalia 1993–2009. *J. Peace Res.* **50**, 545–561 (2013).
- Michalopoulos, S. & Papaioannou, E. National institutions and subnational development in Africa. *Q. J. Econ.* **129**, 151–213 (2014).
- Alix-Garcia, J., Walker, S., Bartlett, A., Onder, H. & Sanghi, A. Do refugee camps help or hurt hosts? *J. Dev. Econ.* **130**, 66–83 (2018).

27. Lv, Q., Liu, H., Wang, J., Liu, H. & Shang, Y. Multiscale analysis on spatiotemporal dynamics of energy consumption CO<sub>2</sub> emissions in China: Utilizing the integrated of DMSP-OLS and NPP-VIIRS nighttime light datasets. *Sci. Total Environ.* **703**, 134394 (2020).
28. Zhao, J. *et al.* Spatio-temporal dynamics of urban residential CO<sub>2</sub> emissions and their driving forces in China using the integrated two nighttime light datasets. *Appl. Energy* **235**, 612–624. <https://doi.org/10.1016/j.apenergy.2018.09.180> (2019).
29. Ma, J., Guo, J., Ahmad, S., Li, Z. & Hong, J. Constructing a new inter-calibration method for DMSP-OLS and NPP-VIIRS nighttime light. *Remote Sens.-basel.* **12**, 937 (2020).
30. Chen, J., Fan, W., Li, D., Liu, X. & Song, M. Driving factors of global carbon footprint pressure: Based on vegetation carbon sequestration. *Appl. Energy* **267**, 114914 (2020).
31. National Centers for Environmental Information <https://ngdc.noaa.gov/eog/download.html>.
32. National Centers for Environmental Information [https://ngdc.noaa.gov/eog/dmsp/download\\_radcal.html](https://ngdc.noaa.gov/eog/dmsp/download_radcal.html) (2006).
33. National Centers for Environmental Information <https://ngdc.noaa.gov/eog/download.html>.
34. Running, S. *et al.* MOD17A3 MODIS/Terra Net Primary Production Yearly L4 Global 1km SIN Grid V055. NASA EOSDIS Land Processes DAAC <https://lpdaac.usgs.gov/products/mod17a3v055/> (2011).
35. 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. *Landscape Urban Plan.* **106**, 62–72 (2012).
36. 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 (2014).
37. Li, X., Li, D., Xu, H. & Wu, C. Inter-calibration 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 (2017).
38. 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 (2019).
39. Heinsch, F.A. *et al.* GPP and NPP (MOD17A2/A3) products NASA MODIS land algorithm. *MOD17 User's Guide* (2003).
40. Running, S.W. & Zhao, M. Daily GPP and annual NPP (MOD17A2/A3) products NASA Earth Observing System MODIS land algorithm. *MOD17 User's Guide* (2015).
41. 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 (2020).
42. Zhang, C., Shao, H. & Li, Y. Particle swarm optimisation for evolving artificial neural network. *IEEE International Conference on Systems, Man and Cybernetics* **4**, 2487–2490 (2000).
43. Adhikari, R., Agrawal, R. K. & Kant, L. PSO based neural networks vs. traditional statistical models for seasonal time series forecasting. *IEEE 3rd International Advance Computing Conference (IACC)* 719–725 (2013).
44. 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. of Artif. Intel.* **26**, 2305–2314 (2013).
45. 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 (2018).
46. Shi, Y. & Eberhart, R. C. Parameter Selection in Particle Swarm Optimization. *International Conference on Evolutionary Programming* 591–600 (1998).
47. Yin, X., Cao, F., Wang, J., Li, M. & Wang, X. Investigations on optimal discharge pressure in CO<sub>2</sub> heat pumps using the GMDH and PSO-BP type neural network—Part A: Theoretical modeling. *Int. J. Refrig.* **106**, 549–557 (2019).
48. Wu, J., Wu, Y., Guo, X. & Cheong, T. S. Convergence of carbon dioxide emissions in Chinese cities: A continuous dynamic distribution approach. *Energy Pol.* **91**, 207–219 (2016).
49. Chen, J., Gao, M., Ma, K. & Song, M. Different effects of technological progress on China's carbon emissions based on sustainable development. *Bus. Strateg. Environ.* **29**, 481–492 (2020).
50. Pedroni, P. Critical values for cointegration tests in heterogeneous panels with multiple regressors. *Oxford B. Econ. Stat.* **61**, 653–670 (1999).
51. Pedroni, P. Panel cointegration: asymptotic and finite sample properties of pooled time series tests with an application to the PPP hypothesis. *Econ. Theor.* **20**, 597–625 (2004).
52. Lee, C. C. & Chang, C. P. Tourism development and economic growth: A closer look at panels. *Tourism Manage.* **29**, 180–192 (2008).
53. Kasman, A. & Duman, Y. S. CO<sub>2</sub> emissions, economic growth, energy consumption, trade and urbanization in new EU member and candidate countries: a panel data analysis. *Econ. Model.* **44**, 97–103 (2015).
54. Chen, J. *et al.* County-level CO<sub>2</sub> emissions and sequestration in China. *figshare*. <https://doi.org/10.6084/m9.figshare.c.5136302.v2> (2020).
55. Elvidge, C. D. *et al.* Relation between satellite observed visible-near infrared emissions, population, economic activity and electric power consumption. *Int. J. Remote Sens.* **18**, 1373–1379 (1997).
56. Elvidge, C. D. *et al.* Night-Time Lights of the World: 1994–1995. *ISPRS J. Photogramm.* **56**, 81–99 (2001).
57. Raupach, M. R., Rayner, P. J. & Paget, M. Regional variations in spatial structure of nightlights, population density and fossil-fuel CO<sub>2</sub> emissions. *Energy Pol.* **38**, 4756–4764 (2010).
58. Wu, J. S., He, S. B., Peng, J., Li, W. & Zhong, X. Inter-calibration of DMSP/OLS night-time light data by the invariant region method. *Int. J. Remote Sens.-basel.* **34**, 7356–7368 (2013).
59. Wu, J. S., Ma, L., Li, W. F., Peng, J. & Liu, H. Dynamics of urban density in China: Estimations based on DMSP/OLS nighttime light data. *IEEE J. Stars* **7**, 4266–4275 (2014).
60. Shan, Y., Huang, Q., Guan, D. & Hubacek, H. China CO<sub>2</sub> emission accounts 2016–2017. *Sci. Data* **7**, <https://doi.org/10.1038/s41597-020-0393-y> (2020).
61. Guan, D., Liu, Z., Geng, Y., Lindner, S. & Hubacek, K. The gigatonne gap in China's carbon dioxide inventories. *Nat. Clim. Change* **2**, 672–675 (2012).
62. Guan, D., Shan, Y., Liu, Z. & He, K. Performance assessment and outlook of China's emission-trading scheme. *Engineering-prc.* **2**, 398–401 (2016).
63. Gong, P. *et al.* Annual maps of global artificial impervious area (GAIA) between 1985 and 2018. *Remote Sens. Environ.* **236**, 111510 (2020).
64. Li, D., Gao, M., Hou, W., Song, M. & Chen, J. A modified and improved method to measure economy-wide carbon rebound effects based on the PDA-MMI approach. *Energy Pol.* **147**, 111862 (2020).
65. Chen, J., Gao, M., Li, D. & Song, M. Analysis of the rebound effects of fossil and non-fossil energy in China based on sustainable development. *Sustain. Dev.* **28**, 235–246 (2020).

## Acknowledgements

This work was supported by the Major Program of National Social Science Foundation of China (Grant No. 20ZDA084); The Key Program of National Natural Science Foundation of China (Grant No. 71934001); The National Natural Science Foundation of China (Grant Nos. 71974186, 71934001, 71471001, 41771568, 71533004); the National Key Research and Development Program of China (Grant No. 2016YFA0602500); the Strategic Priority Research Program of Chinese Academy of Sciences (Grant No. XDA23070400); Sichuan Province Social Science High Level Research Team Building.

### Author contributions

J.C. led the project and provided the idea for this research. M.G. collected the raw data, calculated and assembled the data and prepared the initial manuscript. S.C. collected the raw data and participated in the database construction. W.H., M.S. and Y.S. revised the manuscript. Y.L. and Y.S. designed the research. Y.S. and X.L. provided technical support.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41597-020-00736-3>.

**Correspondence** and requests for materials should be addressed to Y.L. or Y.S.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://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 Creative Commons Public Domain Dedication waiver <http://creativecommons.org/publicdomain/zero/1.0/> applies to the metadata files associated with this article.

© The Author(s) 2020