



ELSEVIER

Contents lists available at ScienceDirect

## Data in brief

journal homepage: [www.elsevier.com/locate/dib](http://www.elsevier.com/locate/dib)



### Data Article

# A spatial database of CO<sub>2</sub> emissions, urban form fragmentation and city-scale effect related impact factors for the low carbon urban system in Jinjiang city, China



Shaoqing Dai <sup>a, b</sup>, Shudi Zuo <sup>a, b, \*</sup>, Yin Ren <sup>a, c, \*\*</sup>

<sup>a</sup> Key Laboratory of Urban Environment and Health, Institute of Urban Environment, Chinese Academy of Sciences, Xiamen, 361021, China

<sup>b</sup> University of Chinese Academy of Sciences, Xiamen, 361021, China

<sup>c</sup> Ningbo Urban Environment Observation and Research Station-NUEORS, Chinese Academy of Sciences, Ningbo, 315800, China

#### ARTICLE INFO

##### Article history:

Received 31 October 2019

Received in revised form 1 February 2020

Accepted 4 February 2020

Available online 11 February 2020

##### Keywords:

Urban form

CO<sub>2</sub> emissions gridded maps

GIS

Landscape ecology

#### ABSTRACT

This paper presented the spatial database collected in 2013 for mitigating the urban carbon emissions of Jinjiang city, China. The database included the high-resolution CO<sub>2</sub> emissions gridded maps, urban form fragmentation evaluation maps, and city-scale effect related impact factors distribution maps at 30 m and 500 m. We collected the multi-sources data including statistical, vector, and raster data from open-access websites and local governments. We used a general hybrid approach based on global downscaled and bottom-up elements to produce the CO<sub>2</sub> emissions gridded maps. The urban fragmentation was measured by the landscape fragmentation metrics under the feature scale and the accurate identification of the urban functional districts. The percentage of the urban area and the points of interest (POI) density representing the city-scale effect related impact factors were calculated in each grid by the land use and POI data. Our database could be used for the validation of urban CO<sub>2</sub> emissions estimation at the city scale. The landscape metrics and city-scale effect related impact

\* Corresponding author. Key Laboratory of Urban Environment and Health, Institute of Urban Environment, Chinese Academy of Sciences, Xiamen, 361021, China.

\*\* Corresponding author. Key Laboratory of Urban Environment and Health, Institute of Urban Environment, Chinese Academy of Sciences, Xiamen, 361021, China.

E-mail addresses: [sdzuo@iue.ac.cn](mailto:sdzuo@iue.ac.cn) (S. Zuo), [yren@iue.ac.cn](mailto:yren@iue.ac.cn) (Y. Ren).

factors maps can also be used for evaluating the socio-economic status in order to solve the other urban spatial planning problems.

© 2020 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

#### Specifications Table

|                                |   |
|--------------------------------|---|
| Subject                        | Environmental Sciences  |
| Specific subject area          | Landscape ecology, Climate change, Carbon emissions   |
| Type of data                   | Text and Geo-tiff raster  |
| How data were acquired         | The raw data used to produce the CO <sub>2</sub> emissions gridded maps were from the statistical yearbooks (energy consumption by sectors), public remote sensing products (NPP-VIIRS, DEM) and the Jinjiang municipal government department (population, land use and road network). Besides, the other raw data used to identify the urban landscape functional districts were obtained from the Jinjiang municipal government department (city master plan) and the Baidu <a href="http://map.com">map.com</a> (points of the interest data). The analyzed data were produced by authors using data fusion and calculation based on GIS and RS technologies through the raw data. |
| Data format                    | Raw and Analyzed  |
| Parameters for data collection | Raw spatial data included the vector master planning (2010–2030), the vector land use data, point of interest data which were uniformed into 16 urban functional subtypes, road network data with 5 different road levels, digital elevation model, and nightlight imagery. Excepting for them, the population per block, the GDP, energy consumption and related socio-economic factors data were collected and calculated as well. All the data were collected for the year of 2013.  |
| Description of data collection | Raw data were collected from a number of major sources including: NOAA website ( <a href="https://ngdc.noaa.gov/eog/viirs/index.html">https://ngdc.noaa.gov/eog/viirs/index.html</a> ), 2014 Quanzhou statistical yearbook, IPCC report, 2014 China Urban Statistical Yearbook, 2014 China Urban Construction Yearbook, 2013 Jinjiang National Economic and Social Development Statistical Bulletin, Geospatial Data Cloud website ( <a href="http://www.gscloud.cn/">http://www.gscloud.cn/</a> ), Planning Bureau of Jinjiang, Public Security Bureau of Jinjiang, and Baidu.com.   |
| Data source location           | Institution: Institute of Urban Environment, Chinese Academy of Sciences<br>City/Town/Region: Jinjiang<br>Country: China  |
| Experimental factors           | The project coordinate system of the spatial data was the Albers WGS 1984. The data fusion and analysis were run out with Arcmap software.  |
| Data accessibility             | Data identification number: 10.5281/zenodo.3566072.<br>Direct URL to the analyzed dataset: <a href="http://doi.org/10.5281/zenodo.3566072">http://doi.org/10.5281/zenodo.3566072</a> .  |
| Related research article       | Authors' names: Shudi Zuo, Shaoqing Dai, Yin Ren.<br>Title: More fragmented urban form more CO <sub>2</sub> emissions? A comprehensive relationship from the combination analysis across different scales<br>Journal: Journal of Cleaner Production<br><a href="https://doi.org/10.1016/j.jclepro.2019.118659">https://doi.org/10.1016/j.jclepro.2019.118659</a>  |

#### Value of the Data

- The provided dataset of gridded CO<sub>2</sub> emissions could be used for the validation of other CO<sub>2</sub> emissions studies at different resolution scales.
- The urban functional district/zone maps could be used to optimize the urban form and design a carbon emissions mitigation strategy.
- The city-scale effect impact factors could be used to evaluate the socio-economic status of this city.

## 1. Data

A spatial database of low carbon urban system represented the spatial distribution maps of CO<sub>2</sub> emissions, urban form metrics (urban landscape fragmentation), proportion of urban area (PUA) and

points of interest density (POID) at two resolutions (30 m ( $R_{30m}$ ) and 500 m ( $R_{500m}$ )) in 2013 Jinjiang city, China. The data were produced from ArcGIS 10.2, Apack 2.23, Fragstats 4.2 and R 3.5.3. In order to produce the spatial database, 10 types of raw data (Table 1) were preprocessed into the uniformed geographical coordinate system and some of them which were used to calculate the urban form metrics were uniformed into 16 urban functional subtypes by the specific standards (Tables 2 and 3). The high-resolution CO<sub>2</sub> emissions gridded maps contained the emissions from the residential, industrial and transport sectors at two resolutions (Fig. 1). The landscape mixing degree of urban functional districts were classified through the functional district types at two resolutions (Fig. 2). The fragmentation levels at two resolutions were identified by the landscape metrics (Fig. 5) which were calculated at the feature scales (Figs. 3 and 4, Table 4). The spatial distribution of PUA and POID in the spatial database represented the city-scale effect impact factors (Fig. 6). The description of all the data could be seen in the Datadescription.txt.

## 2. Experimental design, materials and methods

### 2.1. Raw data collection and preprocess

Our methodology developed the spatial database of CO<sub>2</sub> emissions, urban form fragmentation, and city-scale related impact factors (PUA and POID) at two resolutions ( $R_{30m}$  and  $R_{500m}$ ). All the raw data were collected to produce the gridded maps of the spatial database. The raw data included the vector master planning spatial data for Jinjiang 2010–2030, the vector land use data for Jinjiang 2013, point of interest (POI) data, road network data, the population per town in 2013, the GDP, energy consumption and related socio-economic factors in 2013, digital elevation model (DEM) at 30 m, and nightlight imagery at 500 m in 2013. The detailed information of raw data sources is shown in Table 1.

We must preprocess the spatial raw data from different sources into the specific and standard input data before developing the spatial database. For example, the master planning spatial data of Jinjiang 2010–2030, land use map of Jinjiang, and Baidu POI data have 49 functional subtypes, 31 land use subtypes, and 10 POI subtypes respectively. We aggregated the master planning spatial data and land use map into 16 urban functional districts according to the “Current land use classification standard” (GB/T 21010-2017). Detailed aggregated information could be seen in Table 2. We reclassified the Baidu POI data into the 16 specific urban functional districts by the name of POI data. Detailed reclassification information could see Table 3. Then, we used simple correction method to adjust the nightlight imagery of NPP-VIIRS, which unified the negative value into 0 and resampled it to the 500-m resolution to correct the data. Then, the Kriging spatial interpolation method was used to downscale the 500 m nightlight imagery to 30 m [1]. Besides, 30 m DEM was resampled to 500 m. So far, all the raster data mentioned in Table 1 had the images at two resolutions. Finally, all the vector and raster were unified into the WGS Albers 1984 projection coordinate system.

### 2.2. High-resolution CO<sub>2</sub> gridded maps

Our study follows the 2006 IPCC guidelines for National Greenhouse Gas Inventories [2]. We produced gridded maps of CO<sub>2</sub> emissions with sizes of 30 x 30 m and 500 x 500 m—resolutions which are typical in urban studies—based on multi-source geospatial CO<sub>2</sub> emissions data.

In this study, gridded maps of CO<sub>2</sub> emissions were constructed using a general hybrid approach based on global downscaled and bottom-up elements (e.g., industrial area). The total CO<sub>2</sub> emissions in each grid were calculated as follows:

$$Grid_i = \sum_{l=1}^3 C_l \cdot Weight_{i,l} = \sum_{l=1}^3 AL_l \cdot EF_l \cdot Weight_{i,l} \quad (1)$$

**Table 1**

The description of raw data sources for spatial database.

| Raw data  | Sources   | Data type   |
|---|---|-------------|
| Nightlight imagery from NASA/NOAA Suomi National Polar-orbiting Partnership (NPP-VIIRS) at 500 m  | <a href="https://ngdc.noaa.gov/eog/viirs/index.html">https://ngdc.noaa.gov/eog/viirs/index.html</a> | Raster      |
| The town-level population of Jinjiang, including the number of households, urban population, non-urban population and total population distributed over the 389 blocks, villages and towns.   | Public Security Bureau of Jinjiang.   | Vector      |
| Value increased of industrial GDP (10 thousand yuan), energy consumption per industrial increased value (standard coal/10 thousand yuan), emission factors of standard coal (2.773).  | 2014 Quanzhou statistical yearbook, IPCC report.  | Statistical |
| Electricity, liquefied gas, coal gas and natural gas consumption of residents in Jinjiang, the heat value of liquefied gas, coal gas and natural gas, the emission factor of power grid (0.8095 t CO <sub>2</sub> · Mwh <sup>-1</sup> ) | 2014 China Urban Statistical Yearbook and 2014 China Urban Construction Yearbook, IPCC Report.      | Statistical |
| Energy consumption of private and public (bus and cab) transportation   | 2013 Jinjiang National Economic and Social Development Statistical Bulletin                         | Statistical |
| DEM of Jinjiang at 30 m   | <a href="http://www.gscloud.cn/">http://www.gscloud.cn/</a>   | Raster      |
| Land use map of Jinjiang in 2013 at the parcel level  | Planning Bureau of Jinjiang   | Vector      |
| Master planning spatial data for Jinjiang 2010–2030 at the parcel level   | Planning Bureau of Jinjiang   | Vector      |
| Baidu POI data  | Baidu Map   | Vector      |
| Baidu road network  | Baidu Map   | Vector      |

**Table 2**

The functional district and their descriptions.

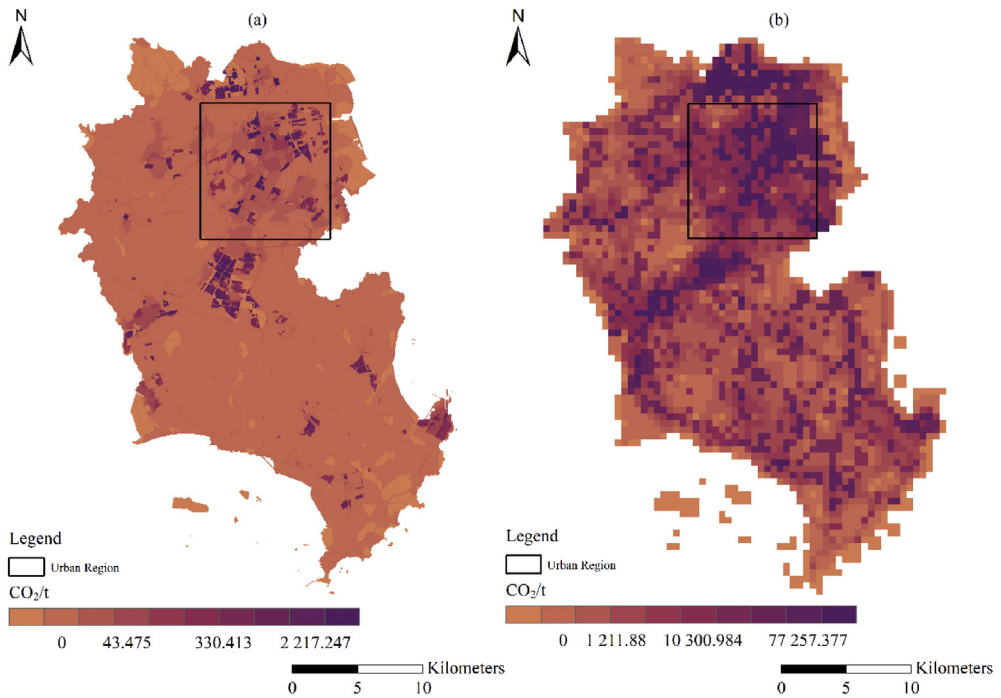
| Functional district                | Detail   |
|------------------------------------|--|
| Roads                              | All the transportation facilities in urban areas, for example, trunk road, expressway, secondary-trunk road, junctions, bus stations and so on |
| Industry                           | The production workshop, warehouse and its auxiliary facilities in industrial and mining enterprises   |
| Water                              | Rivers, lakes, reservoirs, ponds, coastal waters, inland beaches, ditches with water construction, glacier, and permanent snow                 |
| Administration and Public Services | Administrative, cultural, educational, sports, health, and other facilities  |
| Greenland and Plazas               | Public places such as parks, green space, squares and so on  |
| Commercial and Service Facilities  | Business, commercial, entertainment, and other facilities  |
| Municipal Utilities                | Supply, environment, safety and other facilities   |
| Residential                        | Residence and its corresponding facilities   |
| Mixed Function                     | More than three different functional districts   |
| Cropland and Orchard               | Cropland, orchard, forest, grassland, agricultural facilities, rural roads and other kinds of land   |
| Other Non-construction             | Idle land, agricultural land, facilities, raised path, saline soil, swamp, sandy land, bare land   |
| Countryside                        | Construction for rural residential areas   |
| Transportation System              | Railway, highway, airport, port, pipeline and so on  |
| Special Purpose                    | Land of special purposes   |
| Mining                             | Mining, quarrying, sand mining, salt, ground brick kiln production land, and tailings dumps  |
| Logistics Warehousing              | Material reserves, transit, distribution and other kinds of land   |

where  $Grid_i$  is the total CO<sub>2</sub> emissions (unit:  $t$ ) at the  $i^{\text{th}}$  grid ( $i = 1, 2, 3, \dots, n$ ),  $C_l$  (units:  $t$ ) is the total amount of CO<sub>2</sub> emissions from different emission sources,  $AL_l$  (units:  $t$ ) is the total energy consumption from different emission sources, and  $EF_l$  (units:  $t/t$  CO<sub>2</sub>) is the emission factor for different emission sources based on the IPCC method [2] at the  $i^{\text{th}}$  grid ( $l = p, I, T$ , which represent residential, industrial, and transport emissions, respectively), and  $Weight_{i,l}$  is the weight of the specific emission type on-grid  $i$ . In fact,  $Weight_{i,l}$  is the mathematical form of spatial proxies.

**Table 3**

Rules of reclassified POI into different functional districts.

| Functional district                | Types of POI  |
|------------------------------------|---|
| Commercial and Service Facilities  | Hotel, restaurant, supermarket, building, bank, other types |
| Administration and Public Services | School, drugstores, hospital, government, other types       |
| Roads                              | Parking lot   |
| Greenland and Plazas               | Parks   |
| Transportation System              | Toll station, other types                                   |
| Municipal Utilities                | Other types   |
| Residential                        | Other types   |
| Industry                           | Other types   |
| Logistics Warehousing              | Other types   |

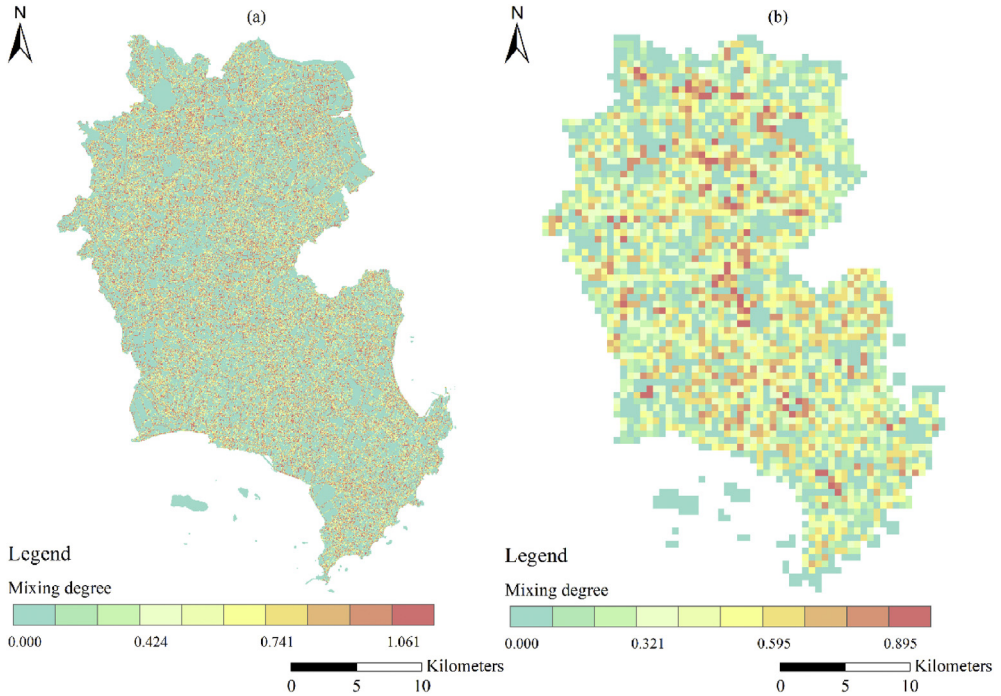
**Fig. 1.** The high-resolution CO<sub>2</sub> gridded maps. (a) and (b) represent the CO<sub>2</sub> emissions gridded maps at R<sub>30m</sub> and R<sub>500m</sub> respectively.

As formula (1) showed, we calculated the total CO<sub>2</sub> emissions based on energy consumption values within the urban geographic boundary of Jinjiang City in 2013 firstly. The total CO<sub>2</sub> emissions could be divided into three sectors: residential, industrial, and transport emissions.

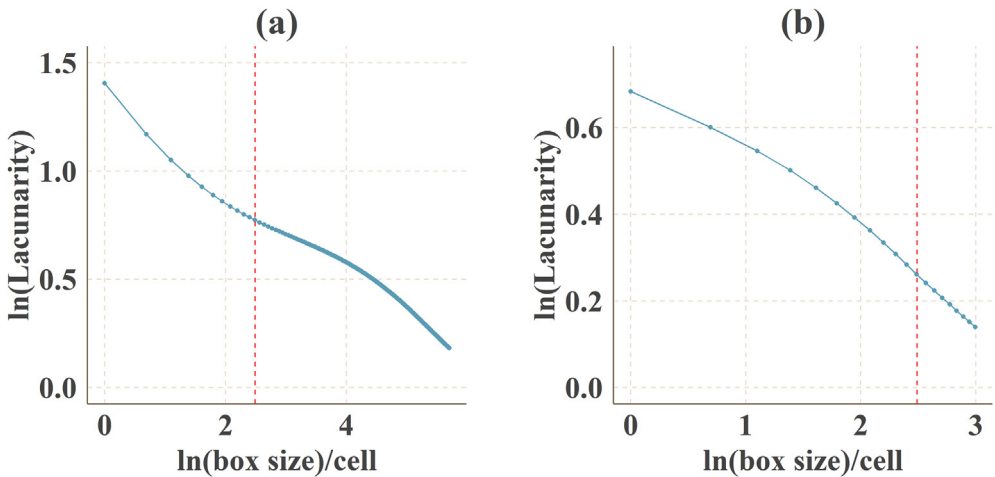
The equation for residential emissions was as follows:

$$C_1 = E_1 \times EF_{1i} / P \quad (2)$$

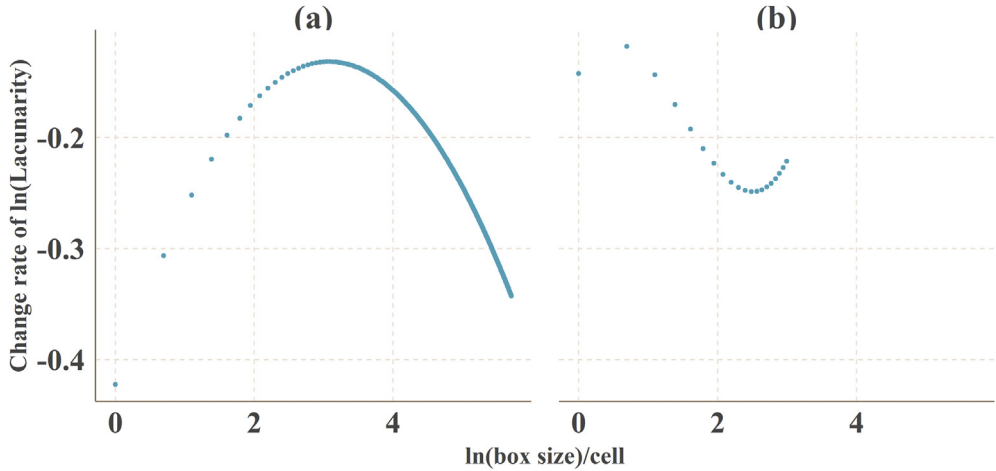
$$C_2 = F_2 \times NVI_1 \times EF_2 / P + F_3 \times NVI_2 \times EF_3 \times M_1 / P + F_4 \times NVI_3 \times EF_4 \times M_2 / P \quad (3)$$



**Fig. 2.** The mixing degree of urban functional district (UFD). (a) and (b) represent the mixing degree of UFD at  $R_{30m}$  and  $R_{500m}$  respectively.



**Fig. 3.** Log-log plots of Lacunarity index versus sliding frame size. (a) and (b) represented  $R_{30m}$  and  $R_{500m}$  respectively. The inflection points were indicated by the dotted red line.



**Fig. 4.** The differential of fitting polynomial for log-log curve of size-Lacunarity index for the feature scale. (a) and (b) represented  $R_{30\text{ m}}$  and  $R_{500\text{ m}}$  respectively.

$$C_p = C_1 + C_2 \quad (4)$$

where  $C_1$  was the electricity  $\text{CO}_2$  emission per capita;  $E_1$  was the household electricity consumption;  $EF_{I1}$  was the carbon emission factor of the power grid, which was  $0.8095 \text{ tCO}_2 \cdot \text{Mwh}^{-1}$  (Fujian Province belongs to the East China regional power grid);  $P$  was the population;  $C_2$  was the gas  $\text{CO}_2$  emissions per capita;  $F_2$ ,  $F_3$  and  $F_4$  were the quantities of household liquefied gas, gas and natural gas consumption;  $NVI_1$  ( $50.179 \text{ MJ} \cdot \text{kg}^{-1}$ ),  $NVI_2$  ( $38.931 \text{ MJ} \cdot \text{m}^{-3}$ ) and  $NVI_3$  ( $38.7 \text{ MJ} \cdot \text{kg}^{-1}$ ) were the heating values of liquefied gas, gas and natural gas, respectively;  $EF_2$  ( $0.06307 \text{ kgCO}_2 \cdot \text{MJ}^{-1}$ ),  $EF_3$  ( $0.0561 \text{ kgCO}_2 \cdot \text{MJ}^{-1}$ ) and  $EF_4$  ( $0.0444 \text{ kgCO}_2 \cdot \text{MJ}^{-1}$ ) were the carbon emission factors of liquefied gas, gas and natural gas, respectively;  $M_1$  ( $0.45 \text{ kg} \cdot \text{m}^{-3}$ ) and  $M_2$  ( $0.717 \text{ kg} \cdot \text{m}^{-3}$ ) were the density of gas and natural gas;  $C_p$  was the mean emissions per capita of  $\text{CO}_2$ .

The equation for industrial emission was as follows:

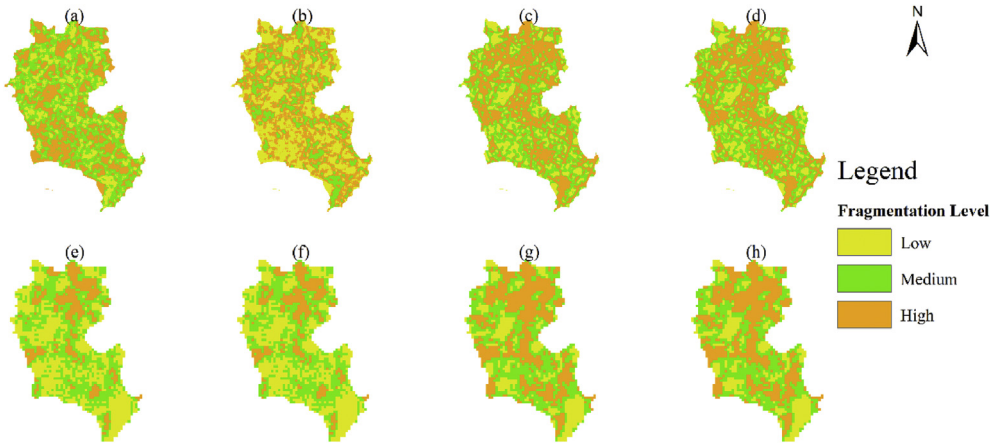
$$C_I = k_1 \times I_{\text{growth}} \times K_I \quad (5)$$

where  $C_I$  was the industrial emission;  $K_I$  was the energy consumption of industrial enterprises above a designated size (t standard coal per million yuan);  $I_{\text{growth}}$  was the increment of the industrial enterprises (million yuan);  $k_1$  was the standard coal  $\text{CO}_2$  emission factor in the city, which was  $2.773 \text{ t CO}_2$  per t standard coal. The standard coal was calculated from the energy consumed during production process that generated greenhouse gases, (e.g. cement and lime production). Besides, energy consumption per industrial increased value was equal to the industrial energy consumption divided by the increased value of industrial GDP, details can be seen at [http://www.stats.gov.cn/tjsj/tjgb/qttjgb/qgqttjgb/201007/t20100715\\_30644.html](http://www.stats.gov.cn/tjsj/tjgb/qttjgb/qgqttjgb/201007/t20100715_30644.html).

The equation for transportemissions was as follows:

$$C_T = Q_1 L_1 \lambda_1 k_2 + Q_2 L_2 \lambda_2 k_2 + Q_3 L_3 \lambda_3 k_2 + T_1 K_{T1} k_1 + T_2 K_{T2} k_1 \quad (6)$$

where  $C_T$  was the transport  $\text{CO}_2$  emissions;  $Q_1$ ,  $Q_2$  and  $Q_3$  were the numbers of buses, taxis and private cars, respectively;  $L_1$ ,  $L_2$ , and  $L_3$  were the annual total mileages of buses, taxis and private cars, respectively;  $\lambda_1$  ( $32 \text{ L} \cdot \text{km}^{-1}$ ),  $\lambda_2$  ( $10 \text{ L} \cdot \text{km}^{-1}$ ) and  $\lambda_3$  ( $10 \text{ L} \cdot \text{km}^{-1}$ ) were the oil consumption factors per hundred kilometers, respectively;  $T_1$  and  $T_2$  were passenger and freight turnover, respectively;  $K_{T1}$  ( $11.6 \text{ kg standard coal per thousand person kilometers}$ ) and  $K_{T2}$  ( $1.9 \text{ kg standard coal per hundred tons}$ )



**Fig. 5.** Fragmentation level based on the landscape metrics. a, b, c, and d are at  $R_{30m}$ . e, f, g, and h are at  $R_{500m}$ . (a) and (e) are NP, (b) and (f) are PD, (c) and (g) are DIVISION, (d) and (h) are MESH.

**Table 4**

Fitting and inflection point of log-log curve of size-Lacunarity.

| Landscape type                      | Fitting curve  | $R^2$ | Inflection point |
|-------------------------------------|--|-------|------------------|
| Functional district patches (30 m)  | $y = -0.422x + 0.094x^2 - 0.010x^3 + 1.406$            | 0.999 | 3.080            |
| Functional district patches (500 m) | $y = -0.143x + 0.061x^2 - 0.049x^3 + 0.008x^4 + 0.684$ | 0.999 | 0.693            |

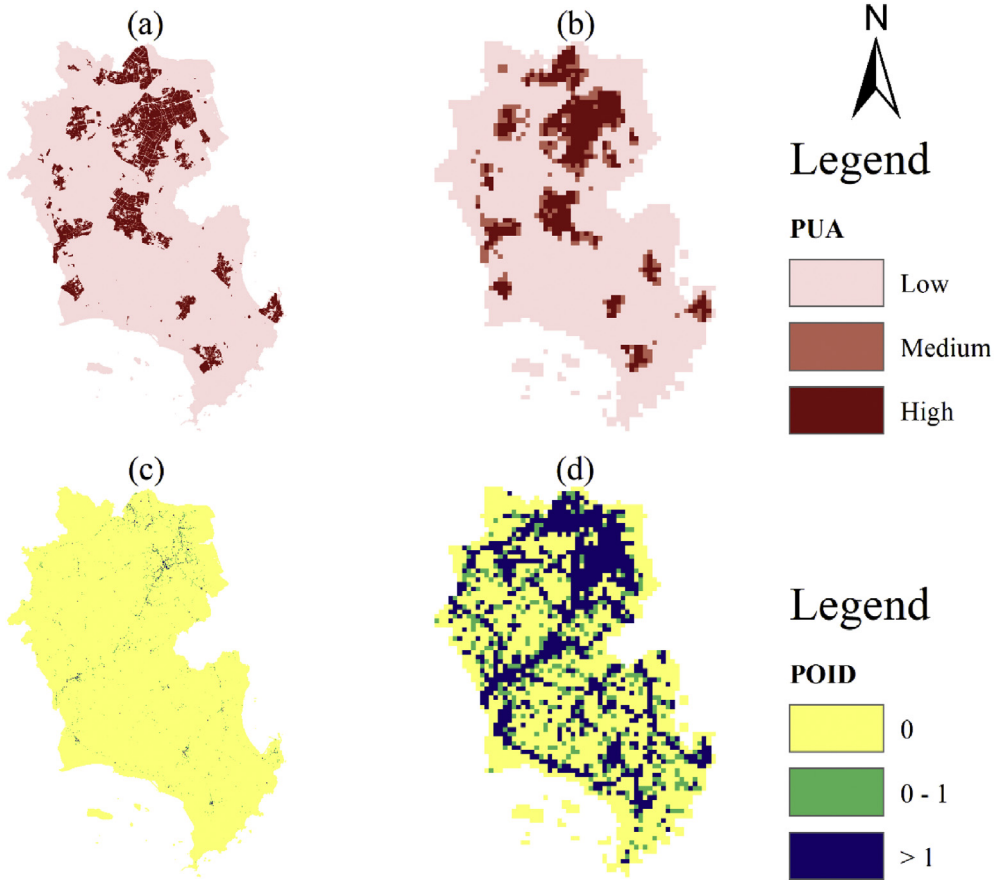
kilometers) were the units of energy consumption of the passenger transport and freight;  $k_1$  (2.773 t CO<sub>2</sub> per t standard coal) was the CO<sub>2</sub> emission coefficients of standard coal and  $k_2$  (2.314 kg·L<sup>-1</sup>) was the gasoline CO<sub>2</sub> emission coefficient. The mileage of buses (70 080 km·year<sup>-1</sup>), taxis (12 000 km·year<sup>-1</sup>) and private cars (20 000 km·year<sup>-1</sup>) was calculated in terms of references [3–6].

To produce gridded maps of CO<sub>2</sub> emissions, we constructed spatial proxies that reflected the distribution of CO<sub>2</sub> emissions and allocated the total emissions of certain regions to each grid according to different weights. The spatial proxies were the high-resolution gridded population map, industrial land map, maps of nighttime light intensity, and the areas of various types of the road which were generated from multi-source data including digital elevation models, nighttime light imagery from the NPP-VIIRS, land use map, POIs and road network map. In order to produce the maps at two spatial resolutions, the spatial proxies were both generated using geospatial data at two resolutions. For instance, the high accuracy population map at 500 m relied on the digital number (DN) value of nighttime light imagery from NPP-VIIRS. However, more proxy variables should be added for the OLS regression models at 30 m such as the elevation and the area of different land-use subtypes. The product of the nighttime light intensity and the industrial land map were used to generate the industrial proxies at two resolutions. Road areas were calculated based on the national road construction standard which states that different classes of road have different road widths at two resolutions [7]. Road areas representing the transport emissions were calculated by multiplying road widths and road lengths in each grid cell using the ArcGIS 10.2 software. Finally, we generate the CO<sub>2</sub> emission gridded maps by overlaying these proxies at two resolutions (Fig. 1).

High-resolution CO<sub>2</sub> emissions gridded maps were obtained by combining global downscaled and bottom-up approaches with spatial analysis models; however, some uncertainties of input variables existed. The uncertainties of input variables were propagated and mapped when obtaining the high-resolution CO<sub>2</sub> emission map. The uncertainties should be explained and analyzed [8].

The first step was to analyze the input variables, the output variables and the whole distribution model. The input variables included total CO<sub>2</sub> estimate and spatial proxies, as shown in formula (1). As to the CO<sub>2</sub> estimate (also known as “magnitude” uncertainty), the activity level and emission factor had





**Fig. 6.** PUA and POID at different resolutions. (a) and (c) are at  $R_{30m}$ , (b) and (d) are at  $R_{500m}$ .

great uncertainties; however, we calculated the number of different emissions using the IPCC guidance, which had high confidence. Besides the magnitude uncertainty, the spatial weight distribution created and propagated the uncertainty. The uncertainty of  $Weight_p$  was the error in the generation of the population map. Uncertainty of  $Weight_t$  depended on identifying the accuracy of industry location and the “blooming” effect of night light images. The area of the road network couldn’t represent the transport emission intensity [9].

Based on the research of Wang et al. [10], greater accuracy could be obtained by comparing the total estimated emissions with the total number of all grids on the map. The maps showed that the absolute errors of total emissions at  $R_{30m}$  and  $R_{500m}$  were 886.70 t and 0.28 t, respectively. The relative errors of total emissions at  $R_{30m}$  and  $R_{500m}$  were 0.005% and less than 0.001%, respectively, especially in residential emissions. RMSEs at  $R_{30m}$  and  $R_{500m}$  were 185.73 t and 4508.19 t, respectively.

### 2.3. Generation of urban form fragmentation

#### 2.3.1. Identification of urban functional districts

With reference to the research of Chi and Long [11], we generated urban functional districts using the land use map, master planning data and POI data that generated the grid frequency density (FD) and grid category ratio (CR).

FD and CR are defined as follows:

$$F_i = \frac{n_i}{N_i} \quad (i = 1, 2, 3 \dots, 6) \quad (7)$$

$$C_i = \frac{F_i}{\sum_{i=1}^9 F_i} \times 100\%, \quad i = 1, 2, 3 \dots, 6 \quad (8)$$

where,  $F_i$  represented the FD of the corresponding classification and  $C_i$  represented the ratio of the FD of the corresponding classification to the FD of all classifications in the grid;  $i$  represented the functional category of POI;  $n_i$  was the number of POI in the  $i$ th category in the grid;  $N_i$  represented the sum of POI in the  $i$ th category. If  $C_i$  exceeded 50%, we defined the grid as the single functional district patch, with the category being the same as the corresponding classification. If  $C_i$  was less than 50%, we defined the grid as the mixing functional district patch.

Since POI was point data, there were some data deficiencies in areas where human activity was rare. Moreover, there were only 10 categories of urban functional areas in the POI data, in which six functional area categories (Water, agriculture and forestry land, other non-construction land, village construction land, special land, mining land) were missing compared to the master plan map data. We spatially superimposed two kinds of data and found that the missing POI data area was labeled as "indefinite" and we replaced it using the master plan map [12]. The indefinite patch was corrected using the land use/cover data. The rules for correcting an indefinite classification of urban functional district patch were as follows:

To identify the parcels whose classification was inconsistent among the POI, land use map and master planning data. No correction was needed for parcels whose classification was all consistent, or at least classification of land use/cover and POI should be the same. The inconsistent parcels were defined as uncertain parcels and were then corrected. The uncertain parcels would be corrected by the land use/cover map. Due to the rare POI deficiency, the single functional district patches would be validated with the land use/cover map as well.

To confirm that urban functional districts were accurately defined, 100 verification points were randomly generated. The Google Earth image in 2013 was selected for visual identification. 84 verification points were consistent. Only 16 points were inconsistent. We revised the inconsistent urban functional district of these corresponding parcels into true urban functional subtypes. The urban mixed and single functional district/zone was defined according to the division of the mixture degree of each grid pixel.

### 2.3.2. Landscape mixed index of urban functional districts

The equation used to calculate the urban functional landscape mixed index was as follows:

$$Landscape_{mix}_i = \frac{-\sum_{k=1}^K p_{k,l} \ln(p_{k,l})}{\ln(K, l)} \quad (9)$$

where  $K$  was the number of urban functional districts types in the  $l$ th pixel and  $p_{k,l}$  represented the area percentage of the  $k$ -type urban functional districts in the  $l$ th pixel. The higher the urban functional districts mixed index was, the higher the fragmentation degree was. The 0 mixed indexes represented the single urban functional district, the area with other values was the mixed urban functional district. The urban functional districts landscape mixed indexes at two resolutions were showed in Fig. 2.

### 2.3.3. Lacunarity index and feature scale determination

We used the Lacunarity index to quantify landscape heterogeneity and select feature scales [13]. The feature scales were the sizes of the moving windows used to calculate the landscape metrics. The landscape metrics refer to a simple quantitative indicator that enriches information on landscape pattern. The metrics reflect the landscape structural composition and provide information regarding some aspects of spatial configuration. Different moving window sizes were set for  $R_{30m}$  and  $R_{500m}$ . As

to  $R_{30m}$ , the moving window size ranged from  $30\text{ m} \times 30\text{ m}$ – $9000\text{ m} \times 9000\text{ m}$ . The moving window of  $R_{500m}$  was from  $500\text{-m} \times 500\text{-m}$  to  $10000\text{ m} \times 10000\text{ m}$ . The Lacunarity index was calculated based on the following equation:

$$\Lambda(r) = S_s^2(r) / \overline{S^2}(r) + 1 \quad (10)$$

where  $\Lambda(r)$  was the Lacunarity index;  $S(r)$  was the mean mixing degree contained in each sample with the moving window;  $S_s^2(r)$  was the variance and  $r$  was the length of the moving window. The inflection point of log-log curve of the size-Lacunarity index was the feature scale, i.e., the maximum value of the differential of the fitted hyperbolic function.

The logarithm of the Lacunarity index decreased as the logarithm of the size increased (Fig. 3). The logarithmic curve dropped greatly in the middle and right sections, indicating that the hierarchical structure of the whole urban mixed landscape and the fractal dimension distribution were in the middle and left sections. Polynomials were used to fit the porosity exponent double logarithmic curve at  $R_{30m}$  and  $R_{500m}$ . The fitted hyperbolic curve showed that  $R^2$  reached 0.999, all the fitted equations were showed in Table 4, of which the fitting effect was excellent. The feature scales of functional area fragmentation analysis were 652.698 m ( $R_{30m}$ ) and 1000 m ( $R_{500m}$ ) (Fig. 4).

### 2.3.4. Landscape metrics

To characterize the fragmentation degree of urban form, Fragstats 4.2 software was used to calculate the number of patches (NP), patch density (PD), division (DIVISION) and effective mesh size (MESH) metrics. Fragstats 4.2 was a spatial analysis program for categorical map and could calculate all the different landscape indexes at different levels. ([https://www.umass.edu/landeco/research/fragstats/downloads/fragstats\\_downloads.html](https://www.umass.edu/landeco/research/fragstats/downloads/fragstats_downloads.html)).

Formulas of the above mentioned indicators were as follows:

$$NP = N_i \quad (11)$$

$$PD = \frac{\sum_{i=1}^M N_i}{A} \quad (PD > 0) \quad (12)$$

$$DIVISION = \left[ 1 - \sum_{j=1}^n \left( \frac{a_{ij}}{A} \right)^2 \right] \quad (0 \leq DIVISION \leq 1) \quad (13)$$

$$MESH = \frac{\sum_{j=1}^n a_{ij}^2}{A} \left( \frac{1}{10000} \right) \quad (14)$$

Where  $N_i$  was the number of patches in the  $i$ th landscape type (e.g. agriculture);  $M$  was the total number of landscape types;  $A$  was the total area of the landscape ( $\text{m}^2$ );  $n$  represented the number of single patches (mixing degree was 0) in the landscape;  $a_{ij}$  was the area of patch  $ij$ .

**Table 5**  
The range of different landscape metric values at  $R_{30m}$  and  $R_{500m}$ .

| Spatial Resolutions | Metric Value | NP | PD     | DIVISION | MESH    |
|---------------------|--------------|----|--------|----------|---------|
| 30 m                | Low          | 4  | 10.078 | 0.117    | 2.094   |
|                     | Medium       | 16 | 40.312 | 0.589    | 9.145   |
|                     | High         | 24 | 60.469 | 0.770    | 16.319  |
| 500 m               | Low          | 2  | 0.889  | 0.198    | 25.000  |
|                     | Medium       | 4  | 2.667  | 0.568    | 58.333  |
|                     | High         | 6  | 3.111  | 0.716    | 108.333 |

High NP, PD and DIVISION values represented high fragmentation, whereas a high MESH value represented low fragmentation. We divided Jinjiang city into low-, mid- and high-fragmentation level areas according to classification maps obtained by the K-Means method (Fig. 5). K-Means method which is a quick clustering method has a low algorithm complexity and a high efficiency in handling the big data [14]. The range of different landscape metric values were showed in Table 5.

#### 2.4. City-scale effect related impact factors of the CO<sub>2</sub> mitigation: PUA and POID

The UFD was the relatively detailed function classification of built-up areas. Non-built-up areas can be treated as a single function classification. The UFD fragmentation of non-built-up areas was not significant enough. In contrast, the phenomenon of urban functional landscape fragmentation in urban built-up areas was more significant. In quantifying the influence of UFD fragmentation on urban CO<sub>2</sub> emissions, we used PUA and POID to control the effects of the urban population, economic scale and aggregation on CO<sub>2</sub> emissions, which are the three most recognized major impact factors of urban CO<sub>2</sub> emissions.

PUA was expressed as the proportion of the urban area to the whole unit per grid. Detailed information on the metrics of urban sprawl was provided by Ref. [15]. POID was expressed as the number of POI to the whole units per grid. PUA was classified into three categories (high, low and middle) using K-Means. POID was classified into three categories (POID=0, POID=1 and POID >1).

#### Acknowledgments

This work was supported by the National Key Research Program of China [grant number 2016YFC0502704]; National Science Foundation of China [grant number 31670645, 31470578, 31200363]; National Social Science Fund [17ZDA058]; Fujian Provincial Science and Technology Project [grant number 2016T3032, 2016T3037 and 2018T3018]; Key Deployment Project of CAS (KFZD-SW-324).

#### Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.dib.2020.105274>.

#### References

- [1] Z.Q. Bai, J.L. Wang, H. Jiang, M.X. Gao, The gridding approach to redistribute population based on multi-source data, *J. Geoinf. Sci.* 17 (2015) 653–660, <https://doi.org/10.3724/SPJ.1047.2015.00653> (in Chinese).
- [2] Intergovernmental Panel on Climate Change (IPCC), *Climate Change 2013: the Physical Science Basis*, Cambridge University Press, Cambridge, 2014, <https://doi.org/10.1017/CBO9781107415324>.
- [3] B.F. Cai, D. Cao, L.C. Liu, Z.S. Zhang, Y. Zhou, Study on CO<sub>2</sub> emission in China road transport, *Energy of China* 33 (2011) 26–30, <https://doi.org/10.3969/j.issn.1003-2355.2011.04.006> (in Chinese).
- [4] M. Zhao, W.G. Zhang, L.Z. Yu, Resident travel modes and CO<sub>2</sub> emissions by traffic in Shanghai city, *Res. of Environ. Sci.* 22 (2009) 747–752, <https://doi.org/10.13198/j.res.2009.06.125.zhaom.019> (in Chinese).
- [5] X.Y. Zhang, X.M. Yang, Y. Yan, Statistical estimation method for energy consumption and carbon emissions by urban transport, *China Soft Sci* 6 (2014) 141–150 (in Chinese).
- [6] Y. Zhang, Y.C. Qin, W.Y. Yan, J.P. Zhang, L.J. Zhang, F.X. Lu, X. Wang, Urban types and impact factors on carbon emissions from direct energy consumption of residents in China, *Geogr. Res.* 13 (2012) 409–417, <https://doi.org/10.11821/yj2012020014> (in Chinese).
- [7] C.K. Gately, L.R. Hutyrta, W. Sue, Cities, traffic, and CO<sub>2</sub>: a multidecadal assessment of trends, drivers, and scaling relationships, *Proc. Natl. Acad. Sci. U.S.A.* 112 (2015) 4999, <https://doi.org/10.1073/pnas.1421723112>.
- [8] M.F. Goodchild, Integrating gis and remote-sensing for vegetation and analysis and modeling - methodological issues, *J. Veg. Sci.* 5 (1994) 615–626, <https://doi.org/10.2307/3235878>.
- [9] W.T. Zhang, B. Huang, D. Luo, Effects of land use and transportation on carbon sources and carbon sinks: a case study in Shenzhen, China, *Landsc. Urban Plann.* 122 (2014) 175–185, <https://doi.org/10.1016/j.landurbplan.2013.09.014>.

- [10] J.N. Wang, B.F. Cai, L.X. Zhang, D. Cao, L.C. Liu, Y. Zhou, Z.S. Zhang, W.B. Xue, High resolution carbon dioxide emission gridded data for China derived from point sources, *Environ. Sci. Technol.* 48 (2014) 7085–7093, <https://doi.org/10.1021/es405369r>.
- [11] J. Chi, L.M. Jiao, T. Dong, Y.Y. Gu, Y.L. Ma, Quantitative identification and visualization of urban functional area based on POI data, *J. Geomatics.* 41 (2016) 68–73, <https://doi.org/10.14188/j.2095-6045.2016.02.017> (in Chinese).
- [12] R. Mercado, A. Páez, Determinants of distance traveled with a focus on the elderly: a multilevel analysis in the Hamilton CMA, Canada, *J. Transport Geogr.* 17 (2009) 65–76, <https://doi.org/10.1016/j.jtrangeo.2008.04.012>.
- [13] R. Plotnick, R.H. Gardner, R.V. Oneill, Lacunarity indexes as measures of landscape texture, *Landsc. Ecol.* 8 (1993) 201–211, <https://doi.org/10.1007/BF00125351>.
- [14] J.A. Hartigan, M.A. Wong, Algorithm as 136: a K-means clustering algorithm, *J. R. Stat. Soc. Ser. C (Appl. Stat.)*. 28 (1979) 100–108, <https://doi.org/10.2307/2346830>.
- [15] J.A.G. Jaeger, R. Bertiller, C. Schwick, D. Cavens, F. Kienast, Urban permeation of landscapes and sprawl per capita: new measures of urban sprawl, *Ecol. Indicat.* 10 (2010) 427–441, <https://doi.org/10.1016/j.ecolind.2009.07.010>.