Environmental Pollution 238 (2018) 874–883

Contents lists available at ScienceDirect

Environmental Pollution

journal homepage: www.elsevier.com/locate/envpol

How do people in different places experience different levels of air pollution? Using worldwide Chinese as a lens[☆]

Bin Chen [a, b], Yimeng Song [c], Mei-Po Kwan [d, e], Bo Huang [c], Bing Xu [a, f, g, *]

[a] Ministry of Education Key Laboratory for Earth System Modelling, Department of Earth System Science, Tsinghua University, Beijing 100084, China
[b] Department of Land, Air, and Water Resources, University of California, Davis, CA 95616, USA
[c] Department of Geography and Resource Management, The Chinese University of Hong Kong, Shatin, Hong Kong
[d] Department of Geography and Geographic Information Science, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
[e] Department of Human Geography and Spatial Planning, Utrecht University, 3508 TC Utrecht, The Netherlands
[f] State Key Laboratory of Remote Sensing Science, College of Global Change and Earth System Science, Beijing Normal University, Beijing 100875, China
[g] Department of Geography, University of Utah, 260 S. Central Campus Dr., Salt Lake City, UT 841112, USA

ARTICLE INFO

Article history:
Received 25 January 2018
Received in revised form
17 March 2018
Accepted 26 March 2018

Keywords:
$PM_{2.5}$
Exposure risk
Spatiotemporal difference
Worldwide chinese
Mobile phone location data

ABSTRACT

Air pollution, being especially severe in the fast-growing developing world, continues to post a threat to public health. Yet, few studies are capable of quantifying well how different groups of people in different places experience different levels of air pollution at the global scale. In this paper, we use worldwide Chinese as a lens to quantify the spatiotemporal variations and geographic differences in $PM_{2.5}$ exposures using unprecedented mobile phone big data and air pollution records. The results show that Chinese in South and East Asia suffer relatively serious $PM_{2.5}$ exposures, where the Chinese in China have the highest $PM_{2.5}$ exposures (52.8 $\mu g/m^3$/year), which is fourfold higher than the exposures in the United States (10.7 $\mu g/m^3$/year). Overall, the Chinese in Asian cities (35.5 $\mu g/m^3$/year) experienced the most serious $PM_{2.5}$ exposures when compared with the Chinese in the cities of other continents. These results, partly presented as a spatiotemporally explicit map of $PM_{2.5}$ exposures for worldwide Chinese, help researchers and governments to consider how to address the effects of air pollution on public health with respect to different population groups and geographic locations.

1. Introduction

Air pollutants, especially fine particulate matter such as $PM_{2.5}$ (particles with an aerodynamic diameter of less than 2.5 μm), have been the focus of increasing public concern because of their potential adverse impacts on human health (Apte et al., 2015; Franklin et al., 2007; Kioumourtzoglou et al., 2016; Kloog et al., 2013; Pope III et al., 2009). Numerous epidemiologic studies have established robust associations between long-term exposure to $PM_{2.5}$ and premature mortality associated with various health conditions—such as heart disease, cardiovascular and respiratory diseases, and lung cancer—that substantially reduce life expectancy (Apte et al., 2015; Franklin et al., 2007; Kioumourtzoglou et al., 2016; Kloog et al., 2013; Pope III et al., 2009). Previous air pollution exposure studies have worked to obtain refined exposure estimates with fine spatiotemporal resolutions (Apte et al., 2015; Han et al., 2016; Ma Z 2016; Park and Kwan, 2017; Van Donkelaar et al., 2010) in order to better address public health issues associated with $PM_{2.5}$ exposure (Di et al., 2017; Kioumourtzoglou et al., 2016; Kloog et al., 2013; Pope III et al., 2009). However, assessing how people in different places experience different levels of air pollution is still a major challenge, especially for specific groups of population at the regional or global scale.

Currently, demographic data based on administrative boundaries is the most widely used data for estimating people's exposures to air pollution (Fleischer et al., 2014; Gray et al., 2014). It provides accurate population census information over a certain period based on the smallest administrative unit (e.g., census block). However, such kind of data has limitations for comparing the exposures of the people in different countries since the data collection procedures used to collect demographic information may not be consistent among different nations. In addition, census data

☆ This paper has been recommended for acceptance by Haidong Kan.

* Corresponding author. Ministry of Education Key Laboratory for Earth System Modelling, Department of Earth System Science, Tsinghua University, Beijing 100084, China.
E-mail address: bingxu@tsinghua.edu.cn (B. Xu).

https://doi.org/10.1016/j.envpol.2018.03.093
0269-7491/

was collected based on census units. It does not consider the spatial variability in population distribution and has very low update frequency (5–10 years), which inevitably introduce considerable uncertainty to the accurate exposure assessment. In contrast, global gridded population data such as LandScan Global Population (Dobson et al., 2000), and Gridded Population of the World (GPW) series (Center for International Earth Science Information Network, 2016) are able to depict the geographic distribution of humans at large scales (e.g., covering many countries) with finer spatial details. However, the main data source of these gridded population data is census population data. Therefore, these datasets will still be affected by the potential concerns mentioned above (Center for International Earth Science Information Network, 2016; Dobson et al., 2000). More importantly, these datasets just regard population as a homogeneous entity and thus cannot be used to identify and differentiate population groups with respect to ethnicity or other socio-economic attributes, which prevents previous studies from investigating the spatial differences and the long-term cumulative differences in air pollution exposures for specific groups of the population. Fortunately, the popularization of mobile computing platforms (e.g., smartphones) and rapid growth of mobile apps (applications designed to run on smartphones and other mobile devices for social networking, navigation, shopping, and dining) produce large amounts of geotagged information and provide researchers with unprecedented opportunities to discover the spatiotemporal characteristics of human activities (Cheng et al., 2011; Lee and Sumiya, 2010; Preoţiuc-Pietro and Cohn, 2013). Moreover, the high geographic correlation between the distribution of geo-spatial big data (e.g., mobile phone records, social media check-in records, taxi trajectories, and smart card records) and human distribution has been widely revealed in previous studies (Fang et al., 2013; Lelieveld et al., 2015; Van Donkelaar et al., 2015; Zhou et al., 2010). The advent of geospatial big data has led to the promising direction of incorporating the socio-economic attributes and mobility of human beings in environmental exposure assessments to discover more specific facets of population exposure to air pollution. From GPS trajectory data of cars to social media check-in records and mobile phone data, a growing number of data sources have been used in relevant studies (Dewulf et al., 2016; Gariazzo et al., 2016; Nyhan et al., 2016). By identifying individuals' behavioral patterns in continuous space-time, geospatial big data may also be used to address the uncertain geographic context problem (UGCoP), which is a common problem in environmental health research because data with coarse spatial and temporal resolution cannot accurately assess individuals' actual environmental exposures (Kwan, 2012; Park and Kwan, 2017).

Meanwhile, some existing caveats concerning the use of geospatial big data should be pointed out here. First, geotagged information derived from multi-source platforms tends to contain considerable noises and discrepancies caused by the different number and composition of active users in terms of ethnicity, culture, education, occupation, income, and age groups. It is thus difficult to obtain geospatial datasets adequately characterize the socio-economic attributes of different groups of population at the regional or global scales. Second, most open-source geospatial datasets have limited spatial and temporal coverage, which hinders the characterization of long-term human activity patterns at the global scale. Third, despite the above issues, limited studies have attempted to combine geospatial big data to investigate how different groups of people who live in different places (at a global scale) experience different levels of air pollution.

Globalization of the 21st century has ushered in an era of fluidity and openness, in which changes in transportation, technology and culture are encouraging people to move across national borders with multiple purposes (e.g., work, settlement, study, professional advancement, marriage, retirement, or lifestyle change) (Castles, 2010), thus leading to the movement of different groups of people among different geographic areas at various spatial scales. For example, with roughly a 12.6% increase (~5 million) in overseas Chinese population during the period 2001–2011, this trend is significant for China (Poston and Wong, 2016). However, knowledge of the fine-resolution distribution of Chinese worldwide still remains limited despite national or international consensus on the distribution and size of oversea Chinese population. Similarly, knowledge of the geographic distribution of other population groups at the global scale is also highly limited to date.

This study seeks to address the challenges in assessing air pollution risks from the perspective of a specific population group and the difficulties in characterizing the geographic distribution of different population groups at the global scale. It uses the global Chinese population as a lens to quantify how people living in different places in the world experience different levels of air pollution. Specifically, this paper presents a spatiotemporal analysis of $PM_{2.5}$ exposures for the global Chinese population using unprecedented mobile phone big data and air pollution records.

2. Data and methods

2.1. Mobile phone location-based big data

In this study, we use the location information in a big mobile phone dataset from Tencent (China) to portray the geographic distribution of the global Chinese population. All of the location data is produced by Tencent through retrieving real-time locations of active mobile phone users when they are using Tencent applications and Tencent's location-based service (LBS) invoked by other mobile apps. As one of the world's largest internet service providers for ethnic Chinese, and given the widespread use of Tencent's service and apps (e.g., Wechat, QQ, etc.), the daily location records have reached 38 billion from more than 450 million users globally in 2016 (Tencent, 2016). It can be argued that the geography of Tencent location data presents a unique geographic distribution of worldwide Chinese. The dataset in this study was collected from March 14, 2016, to August 13, 2016. It has a spatial resolution of 36 arc-second (~1.2 km) and a temporal resolution of 5-min, and is retrieved and updated using the application program interface (API) from the Tencent location big data website (http://heat.qq.com). All the information regarding users' identities and privacies were removed from the public released dataset.

2.2. Global $PM_{2.5}$ concentration dataset

The time-series $PM_{2.5}$ observation records used in this study came from the global $PM_{2.5}$ concentration dataset (Van Donkelaar et al., 2010). Using a simulation of GEOS-Chem chemical transport model, this $PM_{2.5}$ concentration dataset was estimated from an integration of Moderate Resolution Imaging Spectroradiometer (MODIS) and Multi-angle Imaging SpectroRadiometer (MISR) aerosol optical depth (AOD) data with aerosol vertical profiles and scattering properties (Van Donkelaar et al. 2010, 2013). It has a spatial resolution of 10 km and a temporal coverage from 1999 to 2011 as a 3-year moving average, which was applied to reduce the retrieval biases of the annual $PM_{2.5}$ concentrations. Additionally, this dataset has been validated with ground observations at the global scale (Van Donkelaar et al. 2010, 2013, 2015). Experimental tests show that there is a significant agreement between satellite-based estimates and ground-based measurements across different continents (Van Donkelaar et al. 2010, 2013, 2015). The dataset thus provides us a spatially explicit and temporally consistent $PM_{2.5}$ concentration dataset for this study.

2.3. Representing the geographic distribution of Chinese worldwide

Locations in the original dataset were the real-time locations of active mobile users recorded as pairs of geodetic coordinates (longitude, latitude). To utilize this location information, we first constructed a grid with a spatial resolution of 36 arc-second (~1.2 km) (which is consistent with the spatial resolution of the location data). Using this grid structure, we assigned people to particular grid cells based on the location in their mobile phone records. In this way, the geographic distribution of the global Chinese population was represented and visualized using the grid structure. Generally, people are mobile and the geographic distribution of the population is not spatially stationary and temporally constant. As the location data was recorded with a 5-min update frequency, thereby providing dense time-series dynamics of population movements from daily, to weekly or monthly temporal scales. Here we incorporated all the location data in the phone records to achieve the general geographic distribution of global Chinese. Specifically, we aggregated the 5-min location records per day to produce the daily sum of all location records, which represents the daily geographic pattern of global Chinese. We then averaged all the daily records from March 14, 2016 to August 13, 2016 to get the final geographic pattern, which incorporates all the daily-, weekly-, and monthly-fluctuations of population movement and average them to obtain the general pattern.

2.4. Global administrative boundaries and world urban areas

Since $PM_{2.5}$ concentrations and the population distribution pattern obtained using the method describe above vary across space, the spatial heterogeneity of both $PM_{2.5}$ concentrations and human distribution should be considered to better estimate the $PM_{2.5}$ exposures at different spatial scales. Meanwhile, the modifiable areal unit problem (MAUP) illustrates the need for considering scale in real-world analysis. The scale at which we choose to analyze $PM_{2.5}$ exposures, be it for the entire country, by state, by county, or by city, may produce different results. Therefore, three different spatial scales were analyzed in this study. Specifically, level-0 administrative division boundaries (country; see Fig. 1a) and level-2 administrative division boundaries (county; see Fig. 1b) were adopted to extract the smallest units of $PM_{2.5}$ concentrations and population density. Taking the Unite States as an example (Fig. 1c), for the level-0 administrative division, the entire national boundary will be the smallest unit for extracting the $PM_{2.5}$ concentrations and calculating population density. For the level-2 administrative division, the county (e.g., specific counties in Texas) will be the smallest unit for extracting $PM_{2.5}$ concentrations and calculating population density.

In addition, it should be noted that different countries may have different administrative division systems and thus administrative divisions may not be consistent among different countries. Therefore, for convenience, we named the level-0 administrative division as "high-level administrative unit," while the level-2 administrative division is called "low-level administrative unit" in this study. All the administrative boundaries were downloaded at http://www.gadm.org/version2.

Urban air quality has been recognized to be closely correlated with citizens' health (Lelieveld et al., 2015; Pascal et al., 2013). Thus, we further focused on investigating the spatiotemporal differences in $PM_{2.5}$ exposures for Chinese in different cities in the world. Here we adopted the global urban areas dataset (https://www.arcgis.com/home/item.html?id=2853306e11b2467ba0458bf667e1c584) to extract $PM_{2.5}$ concentrations and calculate population density at the city level. This dataset maps the world's major urban areas with populations than 10,000 as discrete polygons and classifies all the urban areas (i.e., cities) into four categories, ranking them from 1st to 4th according to the level of importance (LOI). The rank is calculated based on the source field LOI which numerically ranks feature importance using a DeLorme numbering scheme (DeLorme Publishing Company, 2010). The lower rank, the more importance the city is.

2.5. Estimation of population-weighted $PM_{2.5}$ concentrations

Since the levels of $PM_{2.5}$ concentration and population density are spatially varied (see the test of spatial stratified heterogeneity (Wang et al., 2016; Wang et al., 2010) in Supplementary Materials), the population-weighted metric is likely to be more representative of exposure of population at different spatial scales to ambient air pollution. Here we adopted the population-weighted method (Equation (1)) to estimate the actual $PM_{2.5}$ concentrations.

$$PM^{j}_{Exp} = \sum_{i=1}^{N}(pop_i \cdot pm_i) \Big/ \sum_{i=1}^{N} pop_i \tag{1}$$

where pop_i and pm_i denote the population (derived from mobile phone location records in this study) and $PM_{2.5}$ concentration volume in the ith pixel, N is the total number of pixels within the corresponding administrative unit. PM^{j}_{Exp} is the population-weighted $PM_{2.5}$ concentration volume in the jth unit.

Given the differences of physical environment and socio-economic development in different areas, the population-weighted $PM_{2.5}$ concentrations estimated at the national- or continental-scale will undoubtedly result in the underestimation of Chinese density in rural areas and overestimation of that in urban areas, thereby raising uncertainties and biases in the estimates of actual $PM_{2.5}$ concentrations. Here we first calculated the population-weighted $PM_{2.5}$ concentrations based on the low-level administrative units (i.e., level-2 administrative units). Specifically, for the estimates based on low-level administrative units, both the geographic distribution map of Chinese and the time-series annual $PM_{2.5}$ concentrations dataset were divided into 46579 regions, and then the subsequent divisions will be used to compute the population-weighted $PM_{2.5}$ concentrations according to Equation (1) unit by unit.

Similarly, we further calculated the population-weighted $PM_{2.5}$ concentrations based on the high-level administrative units (i.e., level-0 administrative units). Specifically, both the geographic distribution map of Chinese and the time-series annual $PM_{2.5}$ concentrations dataset were divided into 256 nations/districts, and then the subsequent divisions will be used to compute the population-weighted $PM_{2.5}$ concentrations unit by unit.

Regarding the spatiotemporal difference in $PM_{2.5}$ exposures for Chinese across different cities at the global scale, we extracted 664 major cities in the world from the world urban areas dataset to calculate their population-weighted $PM_{2.5}$ concentrations. All the selected cities were categorized with labels from 1 to 3 that represents their importance in the function of urbanities. In this way, we could not only compare the differences in $PM_{2.5}$ exposure risk for Chinese across global different cities, but also investigate the differences in $PM_{2.5}$ exposure along with different levels of city development.

2.6. Temporal trend of $PM_{2.5}$ concentrations

To detect the temporal trend of $PM_{2.5}$ concentrations over the entire study period (1999–2011), a least-square linear regression model was applied as follows:

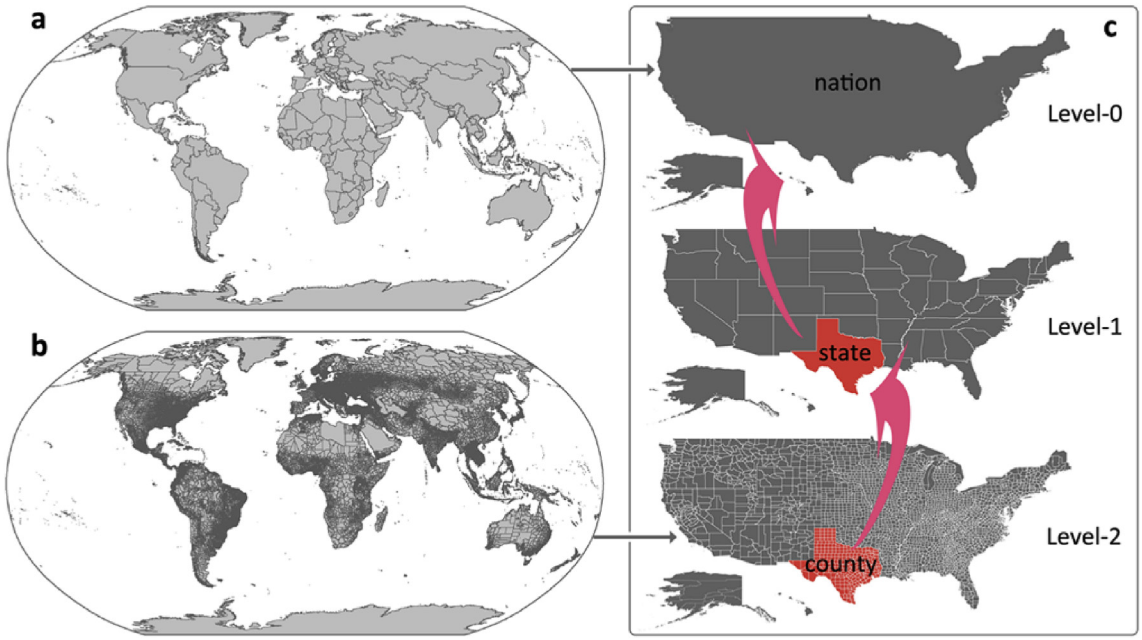

Fig. 1. Global administrative division boundaries. (a) Level-0 administrative boundaries (e.g., countries). including 256 regions (high-level administrative unit) (b) Level-2 administrative boundaries (e.g., counties); including 46579 regions (low-level administrative unit); and (c) taking the United States as an example.

$$y = a + bt + \varepsilon \tag{2}$$

where y represents annual $PM_{2.5}$ concentration volume, t is year, a and b are the least-square fitted coefficients (a is the intercept and b is the trend slope), and ε is the residual bias.

3. Results

3.1. Geographic distribution of worldwide Chinese

Here we derived the geographic distribution of worldwide Chinese (Fig. 2) at a spatial resolution of 36 arc-second (~1.2 km) using mobile phone location data. The distribution and density map of worldwide Chinese population provides us a spatially explicit visualization of the spread of Chinese across the world (Fig. 2). While the vast majority of Chinese population are found within mainland China, the remaining proportion spreads all over the world with several highlighted hot spots, e.g., the Southeast Asian region (including Singapore, India, Thailand, Malaysia, and Indonesia), the East Asian region (including South Korea and Japan), the Middle East area, Europe (including Southeast England, Germany, France, Italy, and Spain), the eastern and western seaboards of the United States and Mexico, and so forth. Globally, cities have higher density of Chinese than areas outside cities, such as rural areas (Fig. 2), and city centers generally have the highest density (subplot in Fig. 2).

To validate the performance of the location data in revealing the actual geographic distribution of Chinese at the global scale, we selected four experimental sites including China, the United States, the United Kingdom, and Japan to quantify the correlation between the geographic pattern of Chinese derived from the mobile phone location data and census data (detailed description of the methods and results are provided in the Supplementary Material). The county-level census datasets of Chinese population in these four experimental sites were collected to compare with the county-level aggregation of the mobile phone location data. As shown in Fig. 3, the spatial patterns between the county-level aggregation of the mobile phone location data (Fig. 3b, e, h, and k) and the county-level census data (Fig. 3c, f, i, and l) are very similar in all testing sites. The correlation coefficient is 0.78 for China, 0.96 for the United States, 0.73 for England, and 0.97 for Japan, thus verifying the reliability of Tencent-based location data for identifying the pattern of geographic distribution of global Chinese.

3.2. Estimates of $PM_{2.5}$ exposure risk based on low-level administrative units

By combining the geographic location of Chinese worldwide and the long-term annual $PM_{2.5}$ concentration records during 1999–2011, we calculated county-level spatiotemporal variation of $PM_{2.5}$ exposures for Chinese worldwide. Globally, Southeast China, North India, the Middle East, and North Africa are the hotspots with the highest $PM_{2.5}$ exposures while South/North America, Europe, South Africa, Southeast Asia, and Australia/Oceania are regions with relatively low $PM_{2.5}$ exposures (Fig. 4a). Regarding the annual temporal changes of $PM_{2.5}$ exposures (Fig. 4b), many areas experienced an elevated rate, especially for the Southeast China (>2 μg/m^3/year), the North India (>2 μg/m^3/year), and the Middle East area (1–1.5 μg/m^3/year).

3.3. Estimates of $PM_{2.5}$ exposure risk based on high-level administrative units

To investigate the spatiotemporal differences of $PM_{2.5}$ exposures for Chinese living in different countries/districts, we calculated the annual exposures to $PM_{2.5}$ for global Chinese based on the low-level administrative units. Results showed that the majority of Chinese in Asian countries experienced relatively high $PM_{2.5}$ exposures (Fig. 5), and the temporal variation of $PM_{2.5}$ concentrations (from upper to bottom chart in Fig. 5) also indicated higher $PM_{2.5}$ exposures for most countries and districts. Based on the 13-year average population-weighted $PM_{2.5}$ concentrations for the 40 countries having the largest location records, we identified the top five countries for Chinese in order of $PM_{2.5}$ exposures as China (52.8 μg/

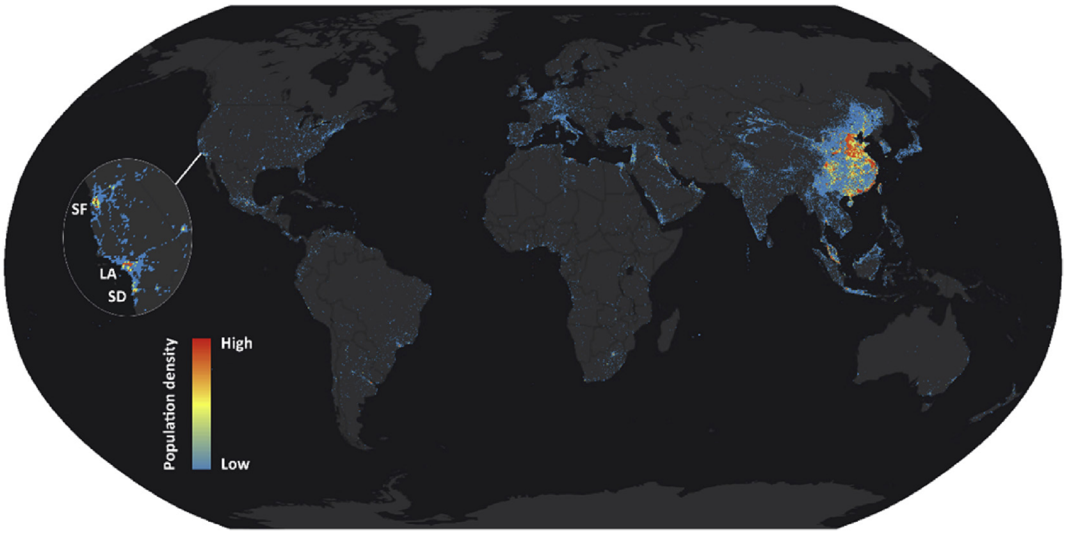

Fig. 2. The geographical distribution and population density of worldwide Chinese derived from the mobile-phone location big data.

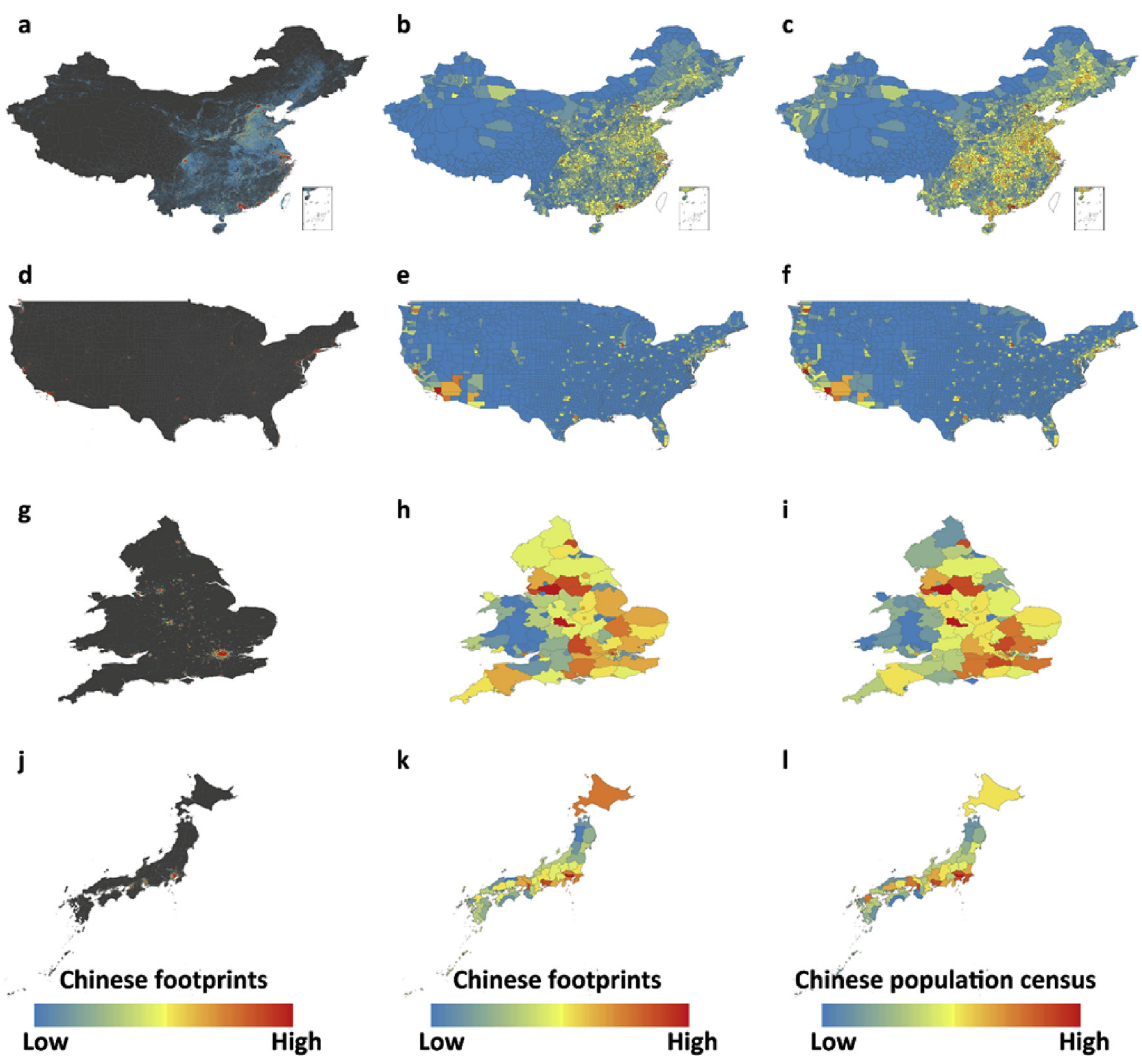

Fig. 3. Comparison between Tencent-based Chinese distribution, Tencent-based clustered Chinese distribution, and Chinese population census data in China (a–c), United States (d–f), United Kingdom (g–i), and Japan (j–l). The left panel (a, d, g, j) represents the pixel-based Chinese density derived from Tencent location data, the middle panel (b, e, h, k) represents the clustered sum of Chinese in terms of the administrative division corresponding to the census data of Chinese population in the right panel (c, f, i, l).

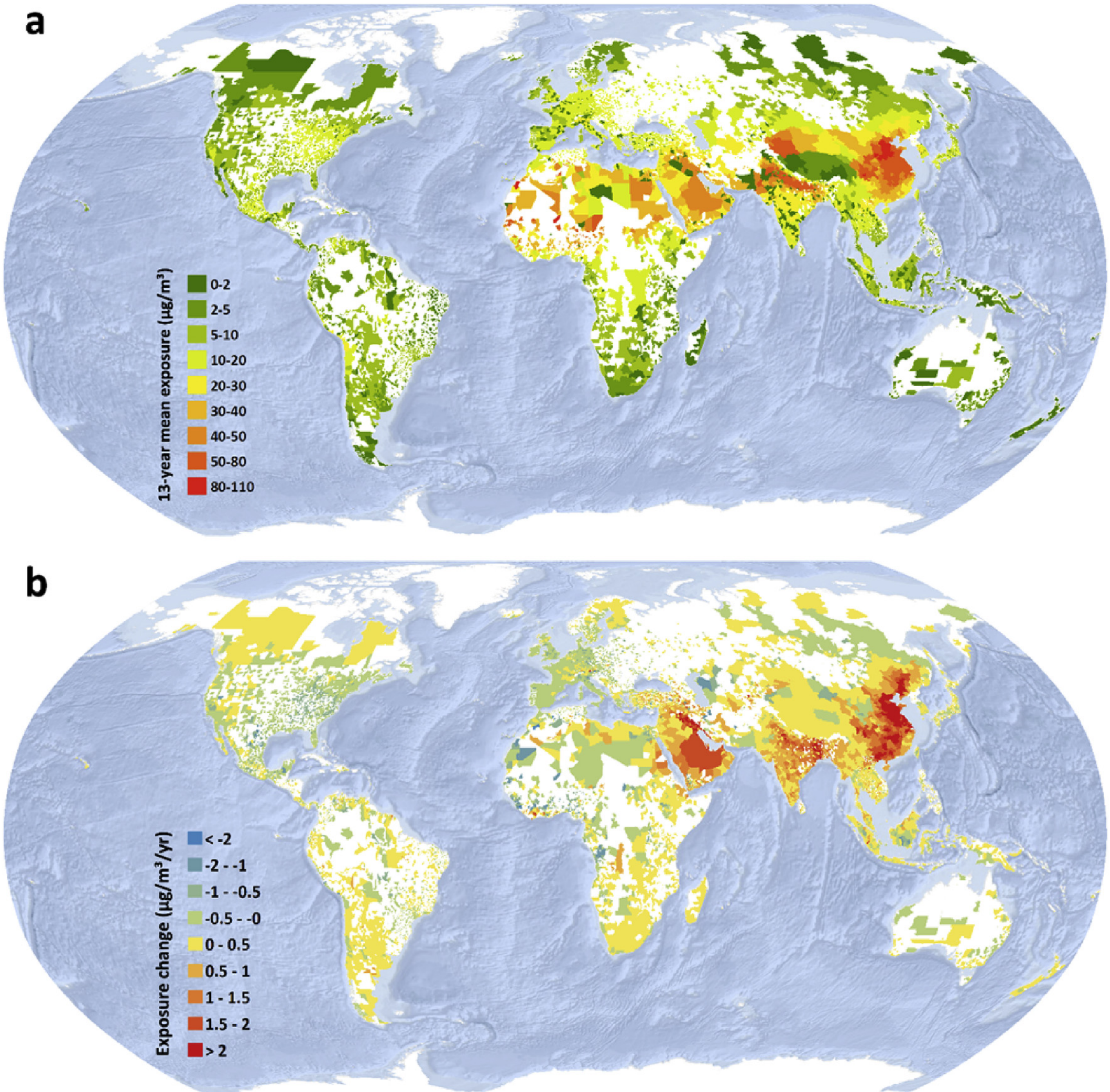

Fig. 4. Estimates of $PM_{2.5}$ exposure risk and its temporal trend based on low-level administrative units. (a) The 13-year mean $PM_{2.5}$ exposure levels estimated by combining mobile phone locating-request big data and air pollution records in the administrative division of worldwide 46579 counties. (b) The linear regression trend of (a) over the period 1999–2011, and the significance test is shown in Figure S1.

m^3/year), Iraq (41.9 μg/m^3/year), United Arab Emirates (37.6 μg/m^3/year), Saudi Arabia (37.5 μμg/m^3/year), India (36.1 μg/m^3/year); while the five countries with the lowest $PM_{2.5}$ exposures are Australia (3.8 μg/m^3/year), Singapore (4.8 μg/m^3/year), South Africa (7.5 μg/m^3/year), Argentina (7.5 μg/m^3/year), and Brazil (8.2 μg/m^3/year). Compared with the Chinese living in the U.S. (10.7 μg/m^3/year) and Canada (9.2 μg/m^3/year), Chinese who live in China suffered approximately five-to six-fold of their exposures. The ratio of $PM_{2.5}$ exposures between China and European countries with considerable Chinese habitants could also reach four and five, e.g., Italy (19.0 μg/m^3/year), Spain (11.5 μg/m^3/year), France (13.5 μg/m^3/year), U.K. (11.6 μg/m^3/year), Germany (16.1 μg/m^3/year).

3.4. $PM_{2.5}$ exposure risk in major cities in the world

As shown in Fig. 2, the majority of Chinese are concentrated in major cities in the world. However, the high $PM_{2.5}$ concentration levels in urban environments pose greater health risks for most Chinese. We further investigated the annual exposures to $PM_{2.5}$ for the Chinese distributed in 664 major cities in the world (Fig. 6). The $PM_{2.5}$ exposures in these cities, aggregated with respect to continent, are ranked in order of average $PM_{2.5}$ concentration levels: Asia (35.5 μg/m^3/year), Africa (17.2 μg/m^3/year), Europe (14.6 μg/m^3/year), North America (9.0 μg/m^3/year), South America (6.6 μg/m^3/year), and Australia and Oceania (3.1 μg/m^3/year). Within each continent, $PM_{2.5}$ exposures varied across locations. For example, the Eastern seaboard cities in the U.S. exhibited higher $PM_{2.5}$ exposures than Western seaboard cities; Southern and Eastern European cities exhibited higher $PM_{2.5}$ exposures than cities in Northern and Western Europe; North African cities suffered distinctly higher $PM_{2.5}$ exposures than cities in South Africa; and three obvious city agglomerations with high $PM_{2.5}$ exposures are located in the Middle East area, South Asia, and East Asia. To differentiate the impacts on $PM_{2.5}$ exposures, we analyzed the spatiotemporal difference in $PM_{2.5}$ exposures across cities with different levels of importance (ranks from 1 to 3). The results

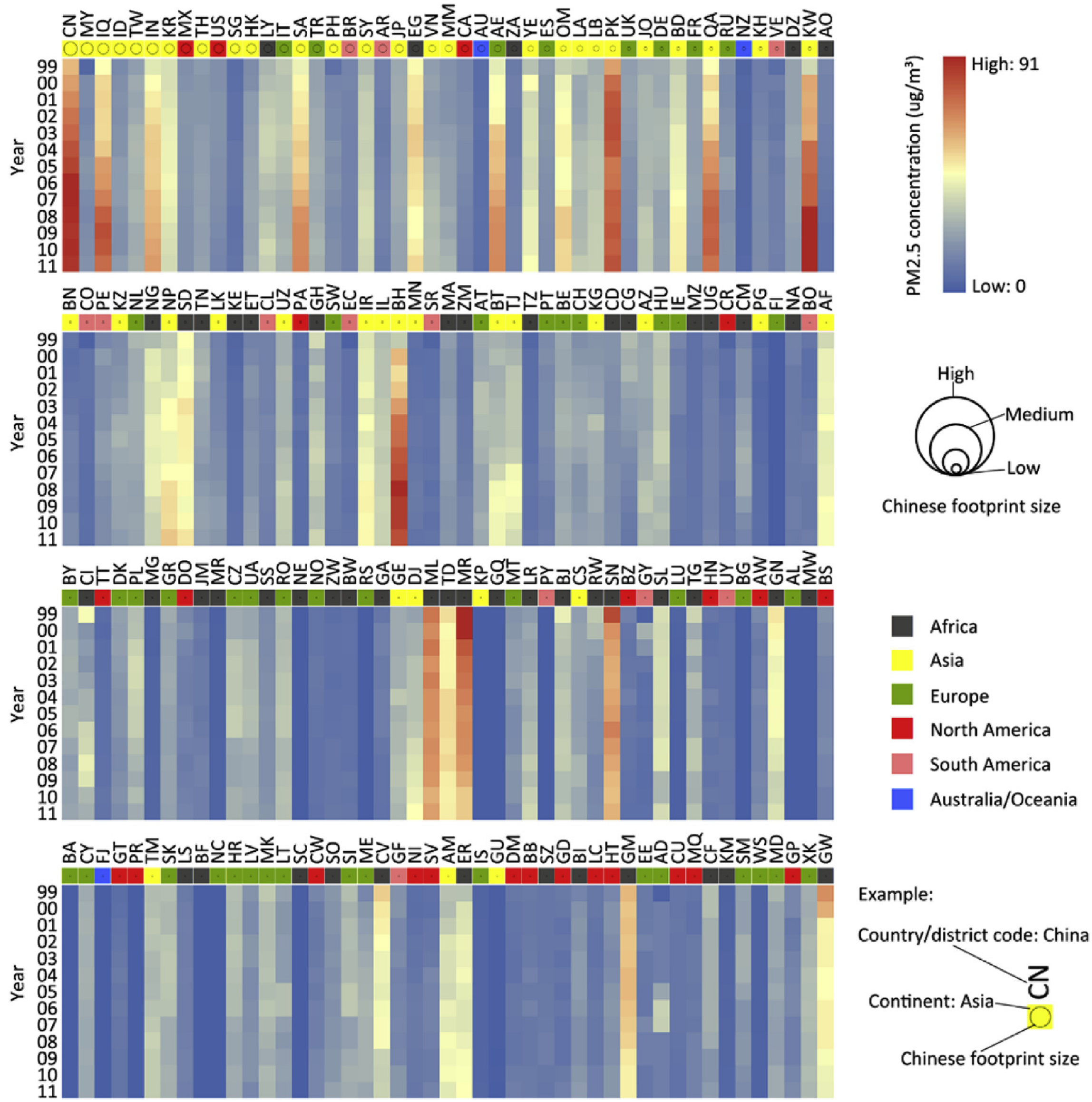

Fig. 5. Estimates of $PM_{2.5}$ exposure risk and its temporal trend based on high-level administrative units.

showed that only Asia exhibited obvious differences in $PM_{2.5}$ exposures between cities of different ranks according to the 13-year mean $PM_{2.5}$ concentrations: third-rank cities (46.1 μg/m^3/year), first-rank cities (33.9 μg/m^3/year), and second-rank cities (19.3 μg/m^3/year), whereas there are no significant difference in $PM_{2.5}$ exposures between cities of different ranks in other continents: specifically, first-rank (17.6 μg/m^3/year) and second-rank (17.4 μg/m^3/year) cities were slightly higher than third-rank cities (14.4 μg/m^3/year) in $PM_{2.5}$ concentrations in Africa, third-rank (16.2 μg/m^3/year) and first-rank (14.9 μg/m^3/year) cities were slightly higher than second-rank (13.4 μg/m^3/year) cities in Europe, first-rank (9.3 μg/m^3/year) and third-rank (9.1 μg/m^3/year) cities were slightly higher than second-rank (8.3 μg/m^3/year) cities in North America, third-rank (7.9 μg/m^3/year) cities were slightly higher than first-rank (6.1 μg/m^3/year), second-rank (6.5 μg/m^3/year) in South America, and first-rank (3.3 μg/m^3/year) cities were slightly higher than third-rank (2.4 μg/m^3/year) cities in Australia and Oceania.

4. Discussion

4.1. The spatiotemporal differences in $PM_{2.5}$ exposures for global Chinese

Estimates based on the three different spatial scales showed that ambient $PM_{2.5}$ exposures for Chinese vary spatially across counties, countries, and cities. That is, for Chinese, living in different places or for different periods will lead to differences in $PM_{2.5}$ exposures and accumulated $PM_{2.5}$ exposures, suggesting that Chinese habitants in different areas may have different levels of health concerns. At the continental level, Chinese who live in Asia, especially Central, South, and East Asia, suffered higher $PM_{2.5}$ exposures than those living in other continents. At the national level, Chinese who live in China suffered the highest $PM_{2.5}$ exposures, and the 13-year average population-weighted $PM_{2.5}$ concentrations from 1999 to 2011 in China were four to five times higher than that in the U.S. and Canada, and three to four times higher than that in

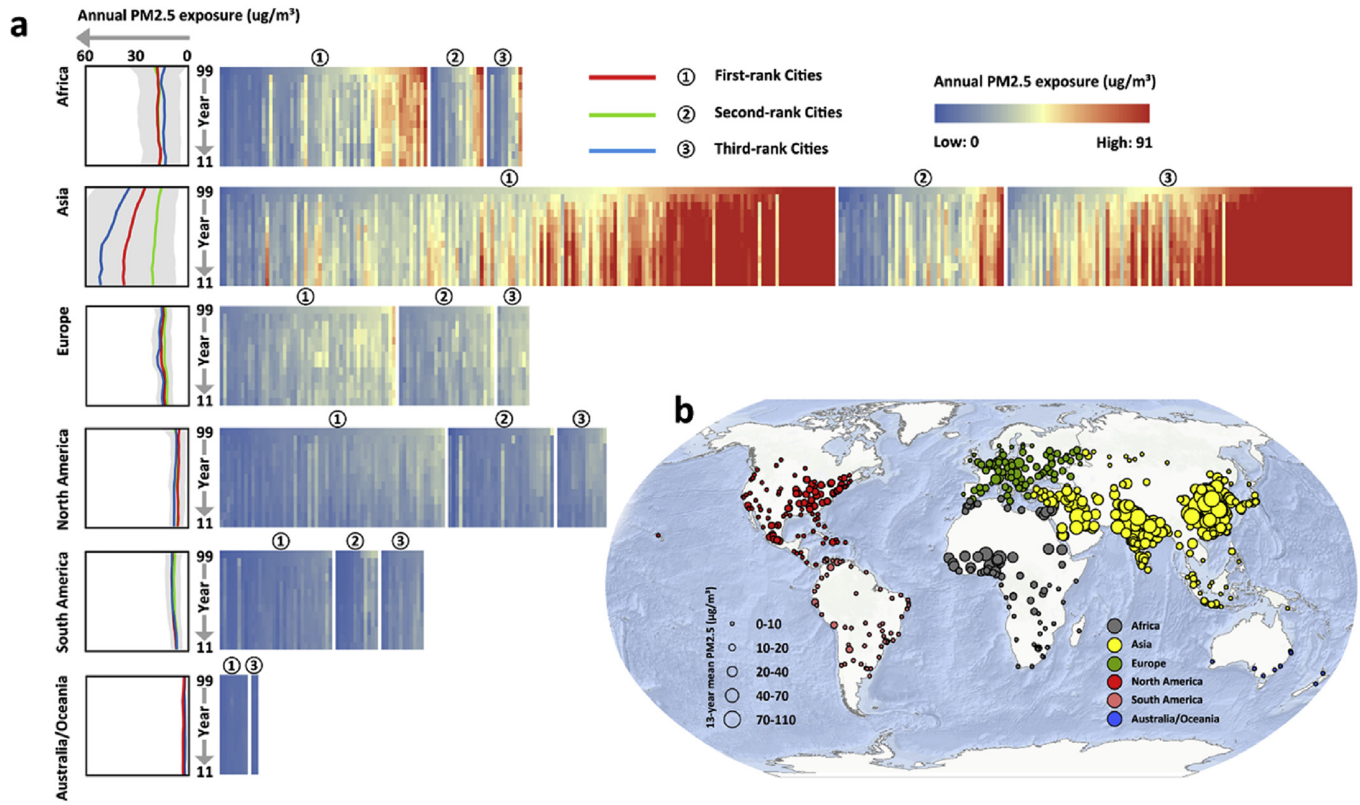

Fig. 6. $PM_{2.5}$ exposure risk and its change trend in global major cities. (a) Annual $PM_{2.5}$ exposure change in different continents by the division of first-, second-, and third-rank cities. (b) The geographic difference of the 13-year mean $PM_{2.5}$ exposure in global major cities in the division of continents.

European countries (Fig. 5). The same situation was also identified at the city level. Furthermore, the estimated $PM_{2.5}$ exposures witnessed an increasing trend from 1999 to 2011. In the front of worsening air pollution, there remains to be possible to lead a new form of global immigration from the polluted world to the clean world. Based on the official statistics from the United Nations Population Division (United Nations, 2015), it can be found out that the Chinese immigrant population are continuously increasing from 5.7 to 8.6 million during 1990–2015 (United Nations, 2015), and North America, Asia, Europe, and Oceania are the most popular immigration destinations (Fig. 7a). Focusing on the period from 2000 to 2010, we could find out that globally, the United States, Canada, Australia, New Zealand, Japan, European countries, and southeast Asian countries are the top destinations for Chinese immigrants (Fig. 7b), accounting for the vast majority of oversea Chinese. With the comparison of the destination of Chinese immigrants (Fig. 7b) and the estimates of $PM_{2.5}$ exposures for global Chinese (Fig. 4a), we can find that these top destinations for Chinese immigrants are all enjoying relative better air quality with low $PM_{2.5}$ exposures than the mainland China. As shown in Fig. 7b, it is interesting to figure out that the Chinese immigrant population in the North America, Europe, and Oceania continued to be increasing during 2000–2010, whereas the Chinese immigrant population witnessed an obvious decrease in Southeast Asia. Besides other socio-economic factors, the decreased air quality in Southeast Asia will also have the potential to account for the cooling-off attractiveness to Chinese immigrants.

4.2. Implementing regulations and improving air quality

In addition to particulate matters $PM_{2.5}$ and PM_{10}, air pollutants such as carbon monoxide (CO), sulphur dioxide (SO_2), nitrogen

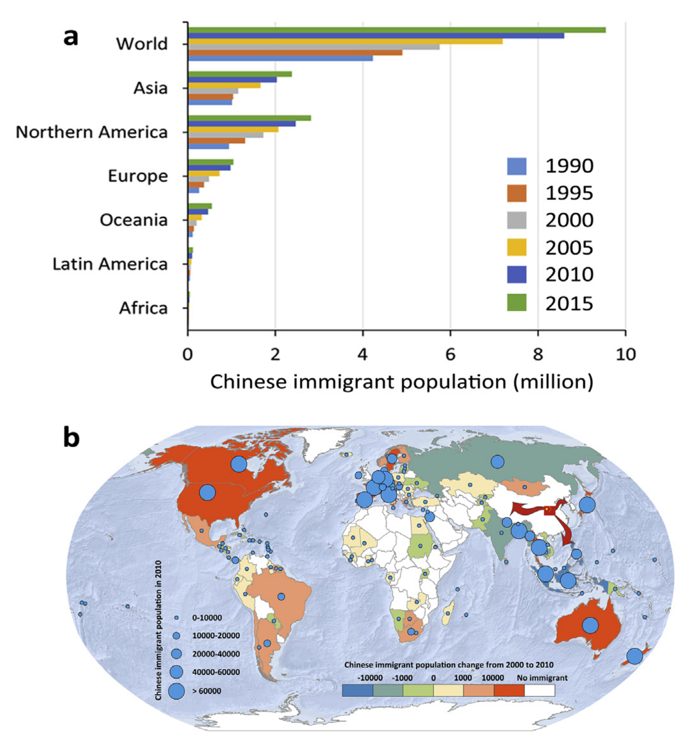

Fig. 7. Chinese immigrant population in the world. (a) Chinese immigrant population change in the division of continents, 1990–2015. (b) The geographic distribution of Chinese immigrant in 2010, and Chinese immigrant population change from 2000 to 2010.

dioxide (NO_x), and ozone (O_3) are emitted by combustion processes that also contribute to emissions of greenhouse gas (Requia et al., 2017; Shah et al., 2013; Watts et al., 2017). Therefore, public actions and policy-driven regulations should be put into effects to reduce ambient air pollutant emissions. For example, as shown in Fig. 6, $PM_{2.5}$ concentrations in most cities are well above the level of $PM_{2.5}$ concentrations at the national level, with particularly high concentrations in cities in Central, South, and East Asia. Therefore, urban air quality management—including traffic control, phasing out vehicles that do not meet emission standards, promoting the use of clean-energy vehicles, and increasing urban greenspace—is highly required. At the same time, additional steps are needed to ensure that countries cooperate and effectively take actions. First, to increase transparency, there needs to be timely and accurate public databases of detailed emission sources that will allow governments and international communities to differentiate and assess the socio-environmental impacts of different emission sources to advance related polices. Second, voluntary actions and policy-driven regulations should be supported and surveilled by spatially-explicit and temporally-consistent observational records. Such surveillance will allow for more accurate and timely assessments of air quality dynamics, which is a prerequisite for establishing, implementing, assessing, and adjusting polices to mitigate emissions from energy, industry, transport, and other sectors. Third, developed countries should provide adequate, predictable, and sustainable financial resources and technologies to support the implementation of emission reduction in developing countries to balance the trade-offs among economic development, local ecosystem services, and human livelihoods. Reducing air pollutant emissions will be a long-term investment that contributes to green development and ultimately yields substantial benefits. Implementing these regulations and strategies would help underpin polices that aim at improving global ambient air pollution and in particular offer a better opportunity of reversing decade-long trajectories of severe air pollution in Asian countries.

4.3. Incorporating geospatial big data into public health issues

Geospatial big data have been widely recognized as being able to provide great support and opportunities for public health research (Dewulf et al., 2016; Gariazzo et al., 2016; Nyhan et al., 2016). In this study, with the unique geotagged information from the active Tencent apps and location-based service users, we inferred the geographic distribution of Chinese in the world with unprecedented spatial details at various spatial scales, thereby providing a new way to assess how specific groups of people live in different places experience different levels of air pollution at varied spatiotemporal scales. To our best knowledge, it is the first time long-term estimates of exposure to $PM_{2.5}$ for Chinese at the global scale is provided, which can be used as a reference for assessing air pollution risks of global Chinese, especially in alarming the emerging need in the mitigation of air pollutants that contribute to higher $PM_{2.5}$ concentrations in Asian countries. Moreover, the proposed method and dataset can be easily extended to estimate other ambient pollution exposures, and they may be used to address the health effects of air pollution with respect to different population groups and different geographic locations.

Meanwhile, some potential concerns about our research methods and results should be addressed. First, the distribution pattern of Chinese at the global scale is assumed to be stable. In this study, the $PM_{2.5}$ concentrations data set was retrieved from 1999 to 2011, whereas the daily record for Chinese population (to identify the distribution and density) were from 2016 (March–August) because of the difficulty in obtaining such data before 2016. Even the validation shows the relative stability of the distribution pattern of Chinese at the global scale to some extent (see detailed validations in Supplementary Material), it is still an assumption which may deviate from the facts, especially for some local areas that experience surges of incoming Chinese. Second, the estimates of $PM_{2.5}$ exposures are computed with a yearly scale. However, air pollution and population distribution need to be investigated at much finer spatiotemporal resolutions in order to obtain accurate exposure estimates (Park and Kwan, 2017). In additions, the volunteer-produced geospatial big data such as the mobile-phone location data used in this study tend to leave out some population groups because children, elderly, and the poor are less frequent active users of mobile phones. Nevertheless, such data can still provide us opportunities for capturing the dynamics of population movements and distribution, which can be closely related to the temporal pace of air pollution monitoring. Third, it will be more interesting to investigate the differences in air pollution exposure between different groups of population at the global scales. However, in this study, we mainly used worldwide Chinese as a lens as a preliminary attempt to address this challenge. Identifying the socioeconomic attributes of different population groups by making use of social media and mobile phone big data (such as Twitter, Facebook, WeChat, Google Maps, Baidu Maps, and Weibo), and combining this geotagged information in the study of public health issues will be important topics for future research.

5. Conclusions

Having better knowledge of how people in different places experience different levels of air pollution is of great importance to both researchers and the public. Taking worldwide Chinese as a lens, this study investigated the spatiotemporal variation of their $PM_{2.5}$ exposures using unprecedented mobile phone location big data and air pollution records. The results showed that $PM_{2.5}$ exposures for global Chinese exhibited considerable differences between geographic regions. That is, Chinese living in different locations or for different periods will experience different levels of $PM_{2.5}$ exposures and accumulative $PM_{2.5}$ effects. This will likely lead to different public health concerns for the Chinese habitants living in different locations. Globally, Chinese who live in mainland China suffered the most severe $PM_{2.5}$ exposures over the decade 1999–2011, which is fourfold higher than the exposures in the United States and Canada, and threefold higher than the exposures in European countries. Our work presents spatiotemporally explicit estimates of $PM_{2.5}$ exposures for worldwide Chinese, which is a new attempt to address the health effects of air pollution with respect to different population groups and geographic locations.

Acknowledgement

The authors thank Tencent Inc. for making the mobile phone location data public available. We also thank Aaron van Donkelaar, R. V. Martin, M. Brauer, and B. L. Boys for making the global annual $PM_{2.5}$ grids product in the socioeconomic data and applications center (SEDAC) available online. This work was supported by the Ministry of Science and Technology of China under the National Key Research and Development Program (2016YFA0600104), and was also supported by a project funded by the China Postdoctoral Science Foundation (2017M620739). In addition, Mei-Po Kwan was supported by a grant from the National Natural Science Foundation of China (41529101). The authors also thank two anonymous reviewers and the editor for providing valuable suggestions and comments, which have greatly improved this manuscript.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.envpol.2018.03.093.

References

Apte, J.S., et al., 2015. Addressing global mortality from ambient PM2. 5. Environ. Sci. Technol. 49, 8057–8066.
Castles, S., 2010. Understanding global migration: a social transformation perspective. J. Ethn. Migr. Stud. 36, 1565–1586.
Center for International Earth Science Information Network, C.C.U, 2016. Gridded population of the world, version 4 (GPWv4): population count adjusted to match 2015 revision of UN WPP country totals. In: Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC).
Cheng, Z., et al., 2011. Exploring millions of footprints in location sharing services. ICWSM 81–88, 2011.
DeLorme Publishing Company, Esri, I., 2010. World Urban Areas.
Dewulf, B., et al., 2016. Dynamic assessment of exposure to air pollution using mobile phone data. Int. J. Health Geogr. 15, 14.
Di, Q., et al., 2017. Air pollution and mortality in the medicare population. N. Engl. J. Med. 376, 2513–2522.
Dobson, J.E., et al., 2000. LandScan: a global population database for estimating populations at risk. Photogramm. Eng. Remote. Sens. 66, 849–857.
Fang, Y., et al., 2013. Impacts of 21st century climate change on global air pollution-related premature mortality. Clim. Change 121, 239–253.
Franklin, M., et al., 2007. Association between PM2. 5 and all-cause and specific-cause mortality in 27 US communities. J. Expo. Sci. Environ. Epidemiol. 17, 279–287.
Gariazzo, C., et al., 2016. A dynamic urban air pollution population exposure assessment study using model and population density data derived by mobile phone traffic. Atmos. Environ. 131, 289–300.
Han, L., et al., 2016. Fine particulate (PM2.5) dynamics during rapid urbanization in Beijing, 1973–2013. Sci. Rep. 6, 23604.
Kioumourtzoglou, M.-A., et al., 2016. Long-term PM2. 5 exposure and neurological hospital admissions in the northeastern United States. Environ. health Perspect. 124, 23.
Kloog, I., et al., 2013. Long-and short-term exposure to PM2. 5 and mortality: using novel exposure models. Epidemiol. Camb. Mass 24, 555.
Kwan, M.-P., 2012. The uncertain geographic context problem. Ann. Assoc. Am. Geogr. 102, 958–968.
Lee, R., Sumiya, K., 2010. Measuring geographical regularities of crowd behaviors for Twitter-based geo-social event detection. In: Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks, pp. 1–10. ACM.
Lelieveld, J., et al., 2015. The contribution of outdoor air pollution sources to premature mortality on a global scale. Nature 525, 367–371.
Ma, Z., H.X., Sayer, A.M., Levy, R., Zhang, Q., Xue, Y., Tong, S., Bi, J., Huang, L., Liu, Y., 2016. Satellite-based spatiotemporal trends in PM2.5 concentrations: China, 2004–2013. Environ. Health Perspect. 124, 184–192.
Nyhan, M., et al., 2016. "Exposure track"—the impact of mobile-device-based mobility patterns on quantifying population exposure to air pollution. Environ. Sci. Technol. 50, 9671–9681.
Park, Y.M., Kwan, M.-P., 2017. Individual exposure estimates may be erroneous when spatiotemporal variability of air pollution and human mobility are ignored. Health & place 43, 85–94.
Pascal, M., et al., 2013. Assessing the public health impacts of urban air pollution in 25 European cities: results of the Aphekom project. Sci. Total Environ. 449, 390–400.
Pope III, C.A., et al., 2009. Fine-particulate air pollution and life expectancy in the United States. N. Engl. J. Med. 2009, 376–386.
Poston, D.L., Wong, J.H., 2016. The Chinese diaspora: the current distribution of the overseas Chinese population. Chin. J. Sociol. 2, 348–373.
Preoţiuc-Pietro, D., Cohn, T., 2013. Mining user behaviours: a study of check-in patterns in location based social networks. In: Proceedings of the 5th Annual ACM Web Science Conference, pp. 306–315. ACM.
Requia, W.J., et al., 2017. Global association of air pollution and cardiorespiratory diseases: a systematic review, meta-analysis, and investigation of modifier variables. Am. J. Public Health e1–e8.
Shah, A.S., et al., 2013. Global association of air pollution and heart failure: a systematic review and meta-analysis. Lancet 382, 1039–1048.
Tencent, 2016. Annual Report.
United Nations, Department of Economic and Social Affairs, Population Division, 2015. Trends in International Migrant Stock: Migrants by Destination and Orifin (United Nations Database, POP/DM/MIG/STOCK/Rev.2015).
Van Donkelaar, A., et al., 2015. Use of satellite observations for long-term exposure assessment of global concentrations of fine particulate matter. Environ. Health Perspect. 123, 135.
Van Donkelaar, A., et al., 2010. Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: development and application. Environ. health Perspect. 118, 847.
Van Donkelaar, A., et al., 2013. Optimal estimation for global ground-level fine particulate matter concentrations. J. Geophys. Res. Atmos. 118, 5621–5636.
Wang, J., et al., 2016. A measure of spatial stratified heterogeneity. Ecol. Indic. 67, 250–256.
Wang, J., et al., 2010. Geographical detectors-based health risk assessment and its application in the neural tube defects study of the Heshun Region, China. Int. J. Geogr. Inf. Sci. 24, 107–127.
Watts, N., et al., 2017. The Lancet Countdown on health and climate change: from 25 years of inaction to a global transformation for public health. Lancet.
Zhou, Y., et al., 2010. Risk-based prioritization among air pollution control strategies in the Yangtze River Delta, China. Environ. Health Perspect. 118, 1204.

<u>Supplementary Material</u>

How do people in different places experience different levels of air pollution? Using worldwide Chinese as a lens

Bin Chen, Yimeng Song, Mei-Po Kwan, Bo Huang, Bing Xu

Table of Contents

Figure S1. The significance test (p-value) of least-square regression for Fig. 4b.

Measure of spatial stratified heterogeneity

Comparability validation between geographic pattern of Chinese derived from Tencent-based location data and census data

Spatiotemporal stability in the geographic distribution of Chinese worldwide

Figure S2. The weibo check-in records collected in 2014 (a) and 2016 (b).

Figure S3. Scatter plots of grid-based Chinese footprint proportion in 2014 (horizontal axis) versus 2016 (vertical axis) in selected nations. i.e., (a) Germany, (b) Italy, (c) New Zealand, (d) South Korea, (e) Thailand, and (f) Japan.

Figure S4. Scatter plots of grid-based Chinese footprint proportion in 2014 (horizontal axis) versus 2016 (vertical axis) in selected cities. i.e., (a) Bangkok, (b) New York, (c) Paris, (d) Rome, (e) Seoul, (f) Shenzhen, (g) Taipei, (h) Tokyo, and (i) Japan.

Figure S5. Geographic distribution of correlation coefficients between grid-based Chinese footprint proportion in 2014 and 2016 in city-scale.

Figure S6. Geographic distribution of correlation coefficients between grid-based Chinese footprint proportion in 2014 and 2016 in nation-scale. Noted the background is the Geographic distribution of global Chinese footprints.

Figure S7. Histogram of correlation coefficients between grid-based Chinese footprint proportion in 2014 and 2016 at city- (a) and nation- (b) scales.

References

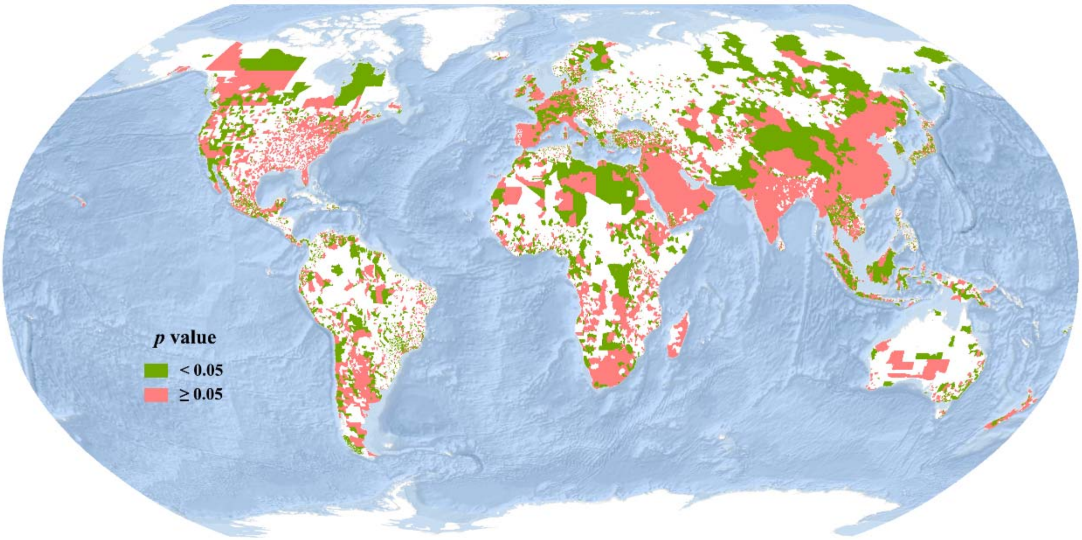

Figure S1. The significance test (*p*-value) of least-square regression for Fig. 4b.

Measure of spatial stratified heterogeneity

A global measurement for the spatial stratified heterogeneity regarding the population and $PM_{2.5}$ concentrations was quantified by the q-statistic, an indicator designed to measure spatial stratified heterogeneity (*1*, *2*).

Given the targeted study area is composited of N units (pixels in this study) and is grouped into $h = 1, 2, \ldots, L$ stratum; stratum h is composed of N_h units; Y_i and Y_{hi} denote the value of unit i and the sum of values in stratum h, respectively; the stratum mean is $\bar{Y}_h = \frac{\sum_{i=1}^{N_h} Y_{hi}}{N_h}$; the stratum variance is $\sigma_h^2 = \frac{\sum_{i=1}^{N_h} (Y_{hi} - \bar{Y}_h)^2}{N_h}$; the total mean of the target study is $\bar{Y} = \frac{\sum_{i=1}^{N} Y_i}{N}$ and its corresponding variance is $\sigma^2 = \frac{\sum_{i=1}^{N} (Y_i - \bar{Y})^2}{N}$.

Thus, the q-statistic can be calculated as follows,

$$q = 1 - \frac{\sum_{h=1}^{L} N_h \sigma_h^2}{N\sigma^2}$$

The value of the q-statistic is between [0,1], and it increases along with the significance of the stratified heterogeneity increases. By adjusting Eq. (1) into three different spatial scales (i.e., level-0 administrative division, level-2 administrative division, and global major cities' division), we are able to measure the spatial stratified heterogeneity for each scale as follows.

For the level-0 administrative division,

$$q_1 = 1 - \frac{\sum_{h_1=1}^{L_1} N_{h1} \sigma_{h1}^2}{N_1 {\sigma_1}^2}$$

where N_1 represents the 46579 stratums (i.e., counties in this scale).

For the level-2 administrative division,

$$q_2 = 1 - \frac{\sum_{h_2=1}^{L_2} N_{h2} \sigma_{h2}^2}{N_2 {\sigma_2}^2}$$

where N_2 represents the 256 stratums (i.e., nations/districts in this scale).

For the global major cities' division,

$$q_3 = 1 - \frac{\sum_{h_3=1}^{L_3} N_{h3}\sigma_{h3}^2}{N_3{\sigma_3}^2}$$

where N_3 represents the 664 stratums (i.e., cities in this scale).

Table S1. The q-statistic for the worldwide Chinese and $PM_{2.5}$ concentrations across different scale.

	Level-0	Level-2	City level
Population	0.55	0.65	0.69
$PM_{2.5}$	0.86	0.60	0.75

Comparability validation between geographic pattern of Chinese derived from Tencent-based location data and census data

To validate the performance of Tencent location data in revealing the actual geographic distribution of Chinese at the global scale, we selected four experimental sites including China, the United States, the United Kingdom, and Japan to quantify the correlation between geographic pattern of Chinese derived with the Tencent-based location data and census data. Specifically, (*i*) in China, the latest county-level population census of China in 2014 obtained from the national scientific data sharing platform for population and health (http://www.ncmi.cn) was used. This dataset was established and maintained by infectious disease network reporting system, and it was derived based on population census released by the State Statistics Bureau. It collected all population census including permanent resident and registered resident at the county level since 2004. Thus, we first extracted the Mainland-China area from the global Chinese distribution (Figure 3a), and then aggregated the Tencent-based Chinese into the county-based division (Figure 3b) corresponding to the county-based census data (Figure 3c).

Similarly, (*ii*) in the United States, the county-based Chinese population census data in 2010 was collected from Pew Research Center based on USA 2010 census (http://www.pewsocialtrends.org/asianamericans/) (Figure 3f); (*iii*) in England, the county-level Chinese population data based on UK 2010 census (Figure 3i) was collected from the Nomis, a web-based database of labor market statistics of University of Durham (https://www.nomisweb.co.uk/default.asp); and (*iv*) in Japan, the county-based Chinese population census data in 2010 (Figure 3l) was collected from the National Statistics Center of Japan (http://www.nstac.go.jp/index.html). Then, the number of ethnic Chinese recorded by census data (Figure 3c,f,i,l) and the volume of Tencent-based MPL records (Figure 3b,e,h,k) was quantitatively compared county by county for each selected country.

Despite the existing time-lag between the census data (2010) and Tencent-based location data (2016), their overall correlations in four selected counties are relatively high and stable (i.e., correlation coefficient: 0.78 for mainland China; 0.97 for the United States, 0.73 for England, and 0.97 for Japan), thereby revealing that Tencent-based location data could help to represent the geographic distribution of Chinese worldwide.

Spatiotemporal stability in the geographic distribution of Chinese worldwide

Although it has been witnessed to show a great significance with roughly 12.6 % gain (~ 5 million) in oversea Chinese population over the period 2001-2011 (*3*), the general spatiotemporal distribution of Chinese in oversea countries and cities seems to be stable, characterized by absolute gains in oversea Chinese population and relative stabilities in oversea Chinese distribution and density. In order to validate this assumption in this study, we used another social media data source "Weibo check-in records" (Figure S2) from 2014-2016 to quantify the dynamics of global Chinese footprints distribution. As the Tencent-based mobile phone location data was launched since late 2015 without inter-annual records over the past years, here we used the check-in records from the Weibo (microblog), which is also one of the most popular social media platforms in domestic and oversea Chinese, as Twitter in America. Its monthly active users reached 222 million in September 2015 and mobile Weibo users cover 85% of the total users (*4*). Although the check-in records of Weibo may not present a complete distribution characteristic of Chinese footprints as Tencent-based location data does, the Weibo data can also perfectly represent a part of Chinese who are active users of Weibo. That is to say that Weibo check-in records can be seen as a subset of Chinese or Tencent-based location data. Therefore, it is reasonable for us to use Weibo check-in records for a substitution to test the spatiotemporal stability of Chinese distribution.

We compared the grid-based Chinese proportion between 2014 and 2016 at the nation- and city- scales. For each grid, we calculate its Chinese proportion p_i according to Equation (S1),

$$p_i = \frac{R_i}{\sum_{i=1}^{N} R_i} \qquad \text{(S1)}$$

where R_i denotes the number of check-in records in the *i*th grid, *N* is the total number of the grids within the administrative boundary (i.e., city or nation in this study).

The spatiotemporal stabilities of Chinese distribution in selected sample nations and cities were shown as scatter plots in Figures S3-4. Results showed that the grid-based Chinese proportion keeps a relatively high consistency along with the period from 2014 to 2016 (e.g., $r^2 > 0.7$ for the selected nations, and $r^2 > 0.8$ for the selected cities). We further applied this method in worldwide nations and cities, and results in Figures S5-6 also confirmed the assumption that global Chinese kept spatiotemporal stabilities in distribution and density. Specifically, as the histogram shown in Figure S7, the majority of cities (Figure S7a) and nations (Figure S7b) achieved high correlation coefficients ($r > 0.60$) of grid-based Chinese proportion between 2014-2016. In the lack of mobile-phone location data over the study period 1999-2011, we used the Tencent location data in 2016 to reveal the global distribution of Chinese footprints and assumed it to be stable during this period.

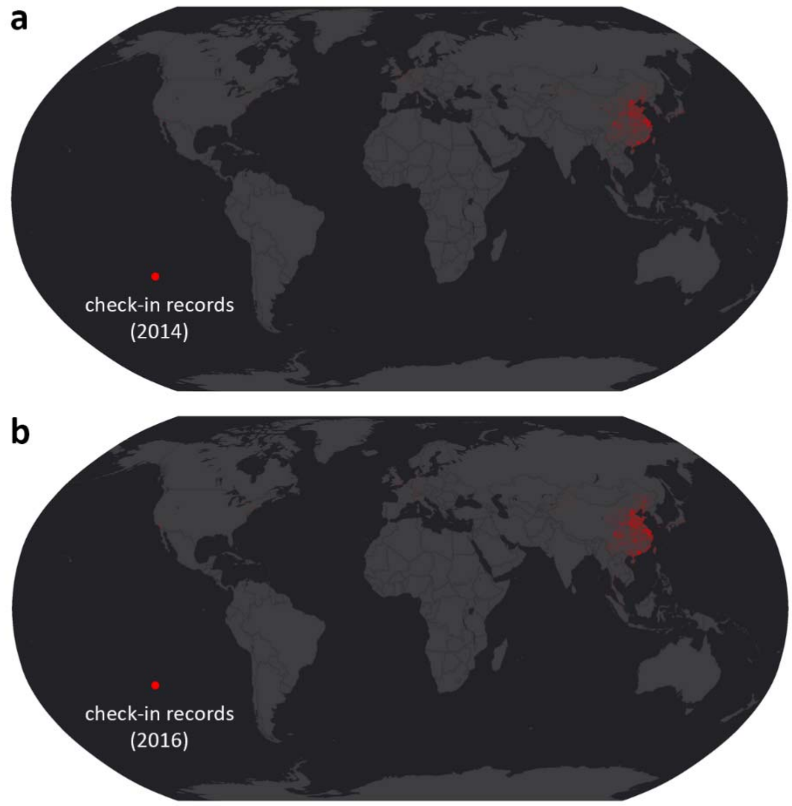

Figure S2. The weibo check-in records collected in 2014 (a) and 2016 (b).

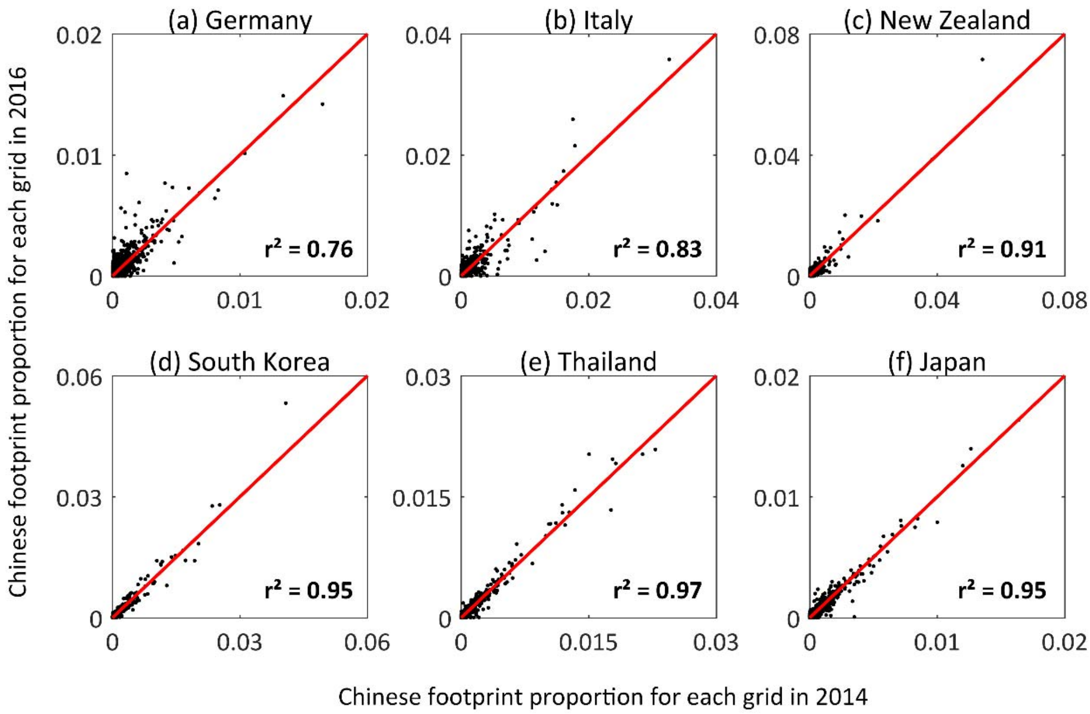

Figure S3. Scatter plots of grid-based Chinese footprint proportion in 2014 (horizontal axis) versus 2016 (vertical axis) in selected nations. i.e., (a) Germany, (b) Italy, (c) New Zealand, (d) South Korea, (e) Thailand, and (f) Japan.

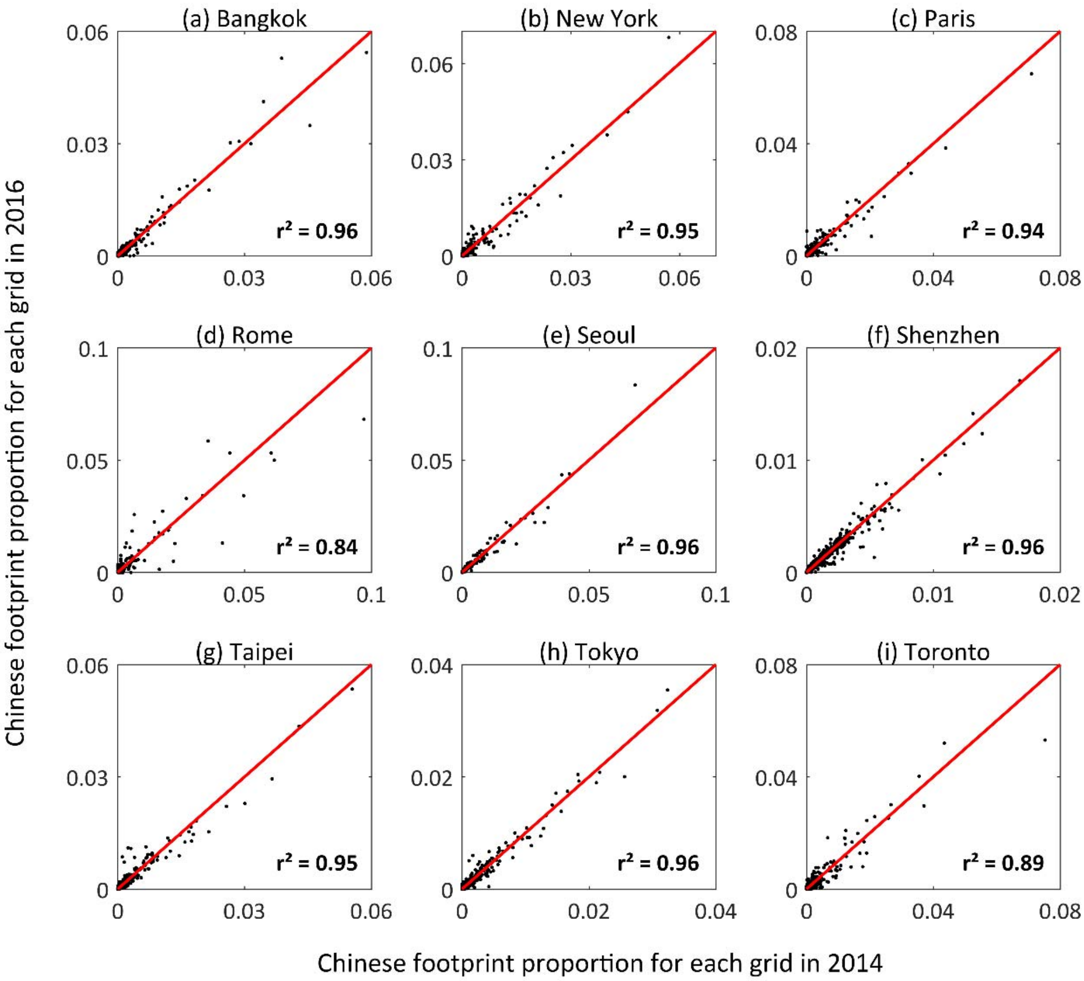

Figure S4. Scatter plots of grid-based Chinese footprint proportion in 2014 (horizontal axis) versus 2016 (vertical axis) in selected cities. i.e., (a) Bangkok, (b) New York, (c) Paris, (d) Rome, (e) Seoul, (f) Shenzhen, (g) Taipei, (h) Tokyo, and (i) Japan.

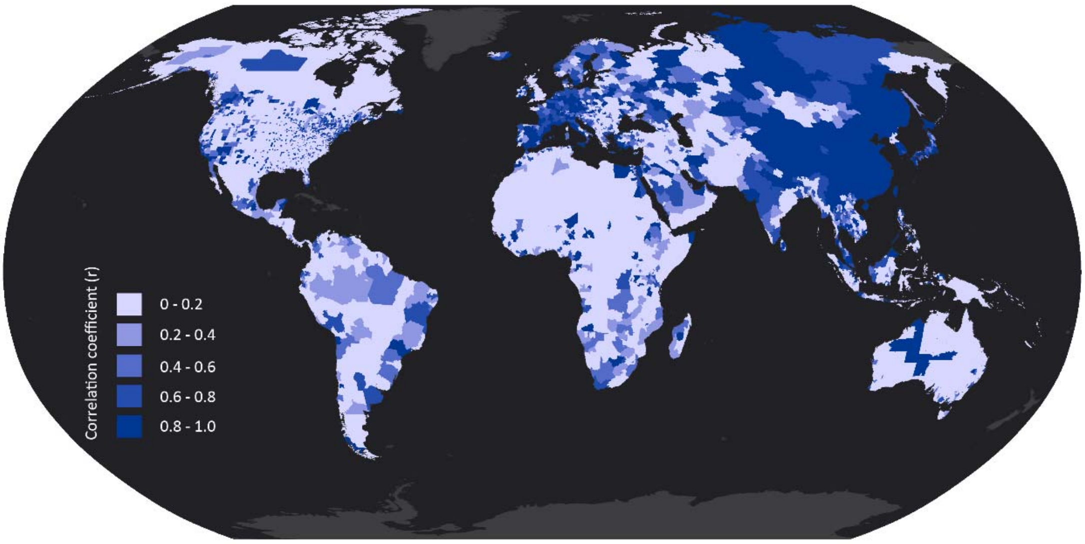

Figure S5. Geographic distribution of correlation coefficients between grid-based Chinese footprint proportion in 2014 and 2016 in city-scale. Note that q-statistic is 0.67 for 2014 and 0.63 for 2016.

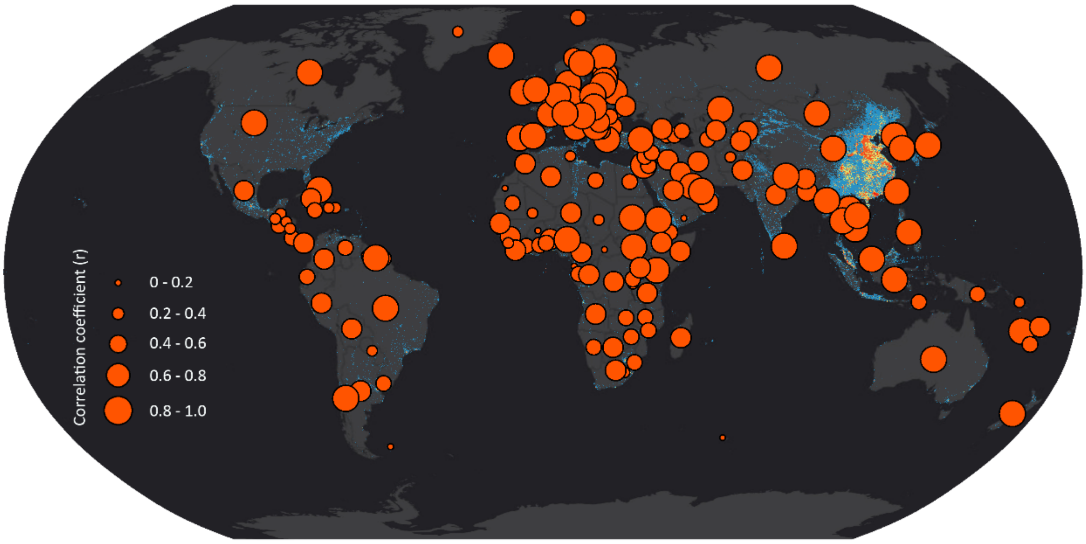

Figure S6. Geographic distribution of correlation coefficients between grid-based Chinese footprint proportion in 2014 and 2016 in nation-scale. Noted the background is the Geographic distribution of global Chinese footprints.

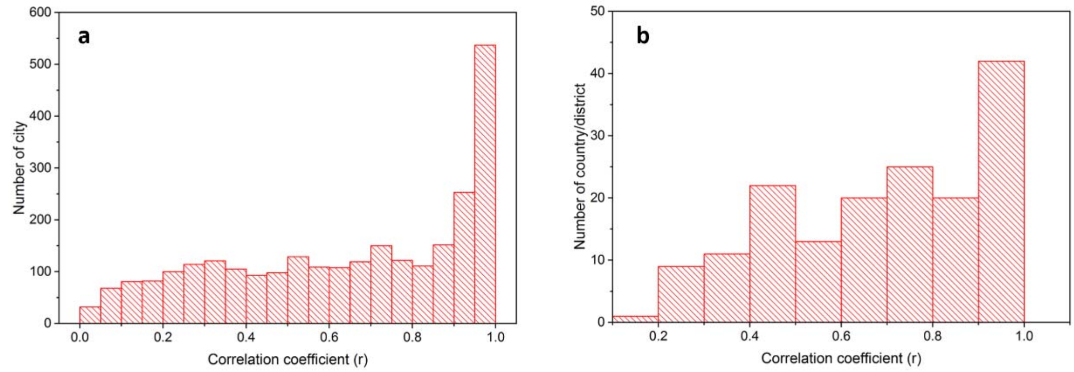

Figure S7. Histogram of correlation coefficients between grid-based Chinese footprint proportion in 2014 and 2016 at city- (a) and nation- (b) scales.

References

1. J. F. Wang *et al.*, Geographical detectors‐based health risk assessment and its application in the neural tube defects study of the Heshun Region, China. *International Journal of Geographical Information Science* **24**, 107-127 (2010).
2. J. F. Wang, T.-L. Zhang, B.-J. Fu, A measure of spatial stratified heterogeneity. *Ecological Indicators* **67**, 250-256 (2016).
3. D. L. Poston, J. H. Wong, The Chinese diaspora: The current distribution of the overseas Chinese population. *Chinese Journal of Sociology* **2**, 348-373 (2016).
4. Weibo. Corporation, "Weibo Reports Third Quarter 2015 Results," (2015).