Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Data-driven framework for delineating urban population dynamic patterns: Case study on Xiamen Island, China

Lei Fanga,b, Jinliang Huanga,*, Zhenyu Zha, Vilas Nitivattananonc

a Fujian Key Laboratory of Coastal Pollution Control, Xiamen University, 361102, Xiamen, China
b College of Computer and Information Engineering, Xiamen University of Technology, 361024, Xiamen, China
c Urban Environment Management Program, Asian Institute of Technology, Pathumthani 12120, Thailand

ARTICLE INFO

Keywords:
Urban population
Dynamic patterns
Baidu heat map
Deep mining approach

ABSTRACT

The effective data mining of social media has become increasingly recognized for its value in informing decision makers of public welfare. However, existing studies do not fully exploit the underlying merit of big data. In this study, we develop a data-driven framework that integrates machine learning with spatial statistics, and then use it on Xiamen Island, China to delineate urban population dynamic patterns based on hourly Baidu heat map data collected from August 25 to September 3, 2017. The results showed that hot grids are primarily clustered along the main street through the downtown area during working days, whereas cold grids are often observed at the edge of the city during the weekend. The mixed use (of commercial and life services, restaurants and snack bars, offices, leisure areas and sports complexes) is the most significant contributing factor. A new cold grid emerged near conference venues before the Brazil, Russia, India, China, and South Africa Summit, revealing the strong effects of regulations on population dynamics and its evolving patterns. This study demonstrates that the proposed data-driven framework might offer new insights into urban population dynamics and its driving mechanism in support of sustainable urban development.

1. Introduction

Urban areas are the busiest places for human activities, such as transportation, education, catering, shopping, sports, and leisure. Hence, the spatiotemporal distribution of urban populations differs significantly from that of rural areas (Yang, Liu, Li, & Li, 2018). The dynamics of a population within a particular spatial zone can promote the re-aggregation and diffusion of social and economic activities (Jacobs, 1961; Jacobs-Crisioni, Rietveld, Koomen, & Tranos, 2014). China has been experiencing rapid urbanization over the past few decades, and the number of urban population reached 58.52 % by the end of 2017 (He, Zhou, Tang, Fan, & Guo, 2019). The significant change in urban population in China is a key driver of urbanization, informatization, industrialization, and globalization in the future, which can reshape the dynamic patterns of the population (de Haas, 2010; Pan & Lai, 2019). Therefore, accurate mapping and understanding regarding the dynamic patterns of urban population as well as epidemiological modeling and disaster prevention are essential for a sustainable urban planning and development (Catlett, Cesario, Talia, & Vinci, 2019; Liu, Wang, Xiao, & Gao, 2012; Santos-Reyes & Olims-Peña, 2017; Zhang, Huang, Duarte, & Zhang, 2016).

Conventional methods to acquire urban population data were accomplished by questionnaires and census sampling, which covered approximately only 1% of the population and focused on specific time segments or macroscale dynamic patterns over a long time period (Dong, Pu, & Wang, 2013; Fan, 2005). Interpolation, choropleth mapping, and dasymetric mapping, for example, are widely used methods to estimate the spatial distribution of populations. However, these maps create uniform distributions of populations in space according to geographic units (census tracts or block groups), causing the values of the mapped population to change abruptly in the boundaries of spatial units, although the dasymetric method can partially amend this problem using statistical surfaces (Nikola, Branislav, Milan, & Dragutin, 2011). Therefore, the current approaches cannot acquire and map spatiotemporal population data completely and efficiently, and precise and quantitative studies regarding population dynamic patterns are rare.

In recent years, with the rapid development in information and communication technology as well as the extensive use of smartphones with various location-based applications (typically known as “Apps”) or social media platforms (e.g., Twitter, Foursquare, Weibo, Baidu heat map, etc.), the ability to capture accurate location information and
point of interest (POI) has become easier and cheaper than before. Numerous studies regarding population or urban dynamics are currently based on location-related information acquired from application programming interfaces or social media platforms (García-Palomares, Salas-Olmedo, Moya-Gómez, Condego-Melhorado, & Gutiérrez, 2018; Jiang, Alves, Rodrigues, Ferreira, & Pereira, 2015; Li, Li, Yuan, & Li, 2019; Shen & Karimi, 2016; Zhen, Cao, Qin, & Wang, 2017). Twitter data, for example, are a promising alternative to big data source. Researchers have used geo-tagged (active check-in) Twitter data, which contained precise spatial and temporal information to perform a city dynamics analysis (García-Palomares et al., 2018). However, researchers have indicated that geo-tagged Tweets constituted only 1% of all posted messages, and that the users were primarily “young people”; furthermore, “check-in” activities typically occurred in scenic, educational, or commercial spots at specific times, rendering the results biased and incomplete (Blanford, Huang, Savelevy, & MacEachren, 2015). Therefore, a wider collection of big data sets is required to compensate for the inherent data biases from the Twitter dataset.

Unlike Twitter in terms of information interaction, the Baidu heat map was developed passive location information collection (Li et al., 2019; Lyu & Zhang, 2019; Zhou, Pei, & Wu, 2018). It will not only constantly record the location information in the background, but also integrate the location information streamed from various widely used LBS (location based services) apps in the smartphone market, most of which were developed or purchased by the Baidu Company. The information includes almost every aspect of daily life, such as navigation, local services, and takeaway services (e.g., Baidu Map, Baidu Nuomi, Baidu Search. Visit the official website for more information: http://home.baidu.com/home/index/company). In contrast to “check-in” or “check-out” data, the Baidu heat map is updated every 15 min to reflect the real-time population distribution and uses an eight-bit map value (i.e., 256 values) to represent the relative population density (from 0 to 255). Although this value could not reflect the real population, it is a good big data source for population distribution and dynamics assessment.

The Baidu heat maps obtained were handled by the batch processing of georeferencing, time aggregation, gridding, and population density index (PDI) calculation. However, the geographical coordinate information of the Baidu heat maps obtained were insufficient. To match the heat map with other typically used remote sensing images (such as Landsat), the WGS 84 coordinate system was selected as the coordinate system. All the heat maps obtained have the same scope and size. Therefore, they have the same georeference coordinates in terms of latitudes and longitudes. First, a Baidu heat map registration process was performed with the ArcGIS georeference tool. Subsequently, the information of this geo-referenced coordinate was used in other Baidu heat maps through a batch processing program.

Because the number of active users varied considerably, considerable fluctuations were observed in the Baidu heat map (Leng, Ying, Huang, & Zheng, 2015). Therefore, the PDI was applied to normalize the data, as follows (1):

$$PDI = \frac{Q_{th}}{\sum Q_{th}}$$

where $PDI$ is the normalized population density index; $Q_{th}$ is the summary of the heat map value in zone $h$ at time $t$; $\sum Q_{th}$ represents the summary of the heat map value of all zones at time $t$.

To realize the merit of these datasets and to facilitate analysis, the data were aggregated into the following four time slots: morning (07:00 to 10:59), noon (11:00 to 14:59), afternoon (15:00 to 18:59), and night (19:00 to 22:59), and the average PDI of each time slot were calculated daily.

Compared with previous investigations conducted at the city or regional level (Chen et al., 2017; He et al., 2018; Huang & Wong, 2016), 400 m $\times$ 400 m spatially connected grids were created in this study to reflect the heterogeneity of the population dynamics. Consequently, 1096 grids covering the entire island were constructed to support this analysis (Fig. 1).

However, it was difficult to manually calculate the PDI value of each grid for all the heat maps, hence, an iterative calculation program was
developed using Model Builder, ESRI ArcGIS 10.3 software (Arcgis Pro, 2019). Model Builder is a visual programing language for building geoprocessing workflows, where a model is represented as a diagram that chains together sequences of processes and geoprocessing tools using the output of one process as the input to another process. The PDI value of all the heat maps will be calculated automatically at each iteration. The workflow diagram is shown in Fig. 3.

A web mining technique was developed to automatically collect 149,074 POI data of Xiamen Island from the Baidu map. The POI refers to a geographic entity that can be abstracted as a point, containing precise spatial information (latitude and longitude). A name or description and a category for POI is typically included. POI categories are similar to land-use categories and the preferences and social functions of people can be well represented by POIs. The type and density of POIs at a specific location can directly or indirectly reflect land-use and functional zoning (Wu, Ye, Ren & Du, 2018). These points were classified into 12 types according to the classification of urban land use and planning standards of development land (GB50137-2011). Some POI points that were not related to land use, such as public toilets and newsstands were removed, whereas some categories were aggregated into a major one. Therefore, six major categories were classified: commercial and life service points of interest (CPOIs), office points of interest (OPOIs), education and health points of interest (EPOIs), scenic and green space points of interest (SPOIs), leisure and sports points of interest (LPOIs), and restaurant and snack bar points of interest (RPOIs). The number of POIs in each category is shown in Fig. 4, and their spatial distribution in each grid is shown in Fig. 5.

2.2. Analysis of spatiotemporal distribution

Spatiotemporal distribution analysis was applied to Baidu heat maps at different time slots, including local spatial autocorrelation analysis, which was used to assess the local clustering characteristics of each grid; consequently, statistically significant hot/cold grids can be recognized. Meanwhile, a 3D space–time cube approach was used to visualize the spatiotemporal distribution of the statistically significant hot/cold grids.

Getis-Ord Gi* was used to evaluate the local clustering characteristics at each time slot; subsequently, the spatiotemporal distribution of hot and cold grids was statistically identified (Getis & Ord, 2010; Ord & Getis, 1995). The null hypothesis for this statistical test was complete spatial randomness (CSR), which postulates that the observed spatial phenomena represent one of many possible spatial arrangements (Fig. 6). For example, if we select all the grids with different PDI values and throw them down randomly after infinite times, grids with highest PDI values would occasionally be accidentally thrown into the same area, and the probability (p-value) would be small (less than 0.01); consequently, the null hypothesis would most likely be rejected; therefore, a hot grid can be distinguished. The z-score (also known as Gi* value) is the standard deviation. The resultantz-scores and p-values indicate that the grids with either high or low values cluster or disperse spatially. The associated equations are as follows:

\[
Gi* = \frac{\sum_{j=1}^{n} \omega_{ij} \bar{X}_j - \bar{X} \sum_{j=1}^{n} \omega_{ij}}{\sqrt{\sum_{j=1}^{n} \omega_{ij}^2} \left( \sum_{j=1}^{n} \omega_{ij} \bar{X}_j \right)^2} 
\]

(2)
where $x_j$ is the PDI for grid $j$; $\omega_{ij}$ is the spatial weight between grids $i$ and $j$; $n$ is equal to the total number of grids, which is 1096.

To ensure the credibility and reliability of this procedure, some preconditions must be clarified, such as the method to define the spatial relationship between the grids and their values. Compared with traditional (nonspatial) statistics, the foundation of this spatial statistical test is the CSR hypothesis. We wish to determine if the observed spatial pattern (hot or cold spot) represents one of many ($n!$) possible spatial arrangements such that we can reject the CSR hypothesis. Some previous studies have indicated that multiple testing techniques are suitable for obtaining the optimal parameters (Goovaerts, 2010; He et al., 2019). Therefore, we conducted 10,000 iterations of random sampling to test the possibility of the observed spatial patterns. Finally, the spatial relationship was set to a fixed distance and its neighborhood search threshold was set to 400 m. For a more stringent and credible statistical test, the false discovery rate (FDR) procedure was applied, which can potentially reduce the critical p-value thresholds shown in Fig. 6.

A 3D space–time cube technique was used to statistically visualize the spatiotemporal distribution of hot/cold grids. Consider the morning slot from August 25 to September 3 as an example (Fig. 7). Each cube represents the clustering status at one location during a specific time slot per day. The column with different colors represents the daily variation from August 25 to September 3. The slice represents the spatial distribution of the hot/cold grids on the same day.

**2.3. Analysis of dynamic patterns**

Dynamic pattern analysis was performed on each time series of the hot/cold grids at different time slots. The trends of those hot/cold grids in each column were evaluated using the Mann–Kendall trend test (Hamed, 2009; Kossack & Kendall, 1950; Mann, 1945). The Mann–Kendall trend test is a non–parametric test; therefore, it is applicable to all distributions. This method is typically applied to detect increasing
Fig. 5. Spatial distribution of POIs in each category.
or decreasing trends in a time series data-set. The null hypothesis for this test is that no monotonic trend exists in the series, and the alternate hypothesis is that a trend exists. The equations involved are as follows:

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} \text{sgn}(X_j - X_k)$$ (5)

$$\text{sgn}(x) = \begin{cases} 
1 & \text{if } x > 0 \\
0 & \text{if } x = 0 \\
-1 & \text{if } x < 0 
\end{cases}$$ (6)

The mean of $S$ is $E[S] = 0$ and $\text{var}(S) = n(n-1)(2n+5)/18$. If a trend is present, the sign values will tend to increase or decrease constantly. Every value is compared to every value preceding it in the time series.

The test can be used to obtain trends for as few as four samples. However, with only a few data points, the test has a high probability of not obtaining a trend when one is present if more points are provided. The more the data points, the more likely is a true trend is obtained in the test (as opposed to one obtained by chance). The minimum number of recommended measurements is between 8 and 10. Therefore, 10 days of continuous data (August 25 to September 3, 2017) were used to guarantee the credibility of the results.

The dynamic patterns for each column were then identified according to the trend characteristics. Seventeen types of dynamic patterns were identified (Arcgis Pro, 2019). A new hot spot is a grid that is a statistically significant hot spot for the final time step and has never been a statistically significant hot spot before. An intensifying hot spot means a grid that has been a statistically significant hot spot for 90% of the time-step intervals including the final time step, in which the intensity of clustering of high counts in each time step is increasing and that increase is statistically significant. Detailed information regarding other types of dynamic patterns are presented in Table B1 in Appendix B.

2.4. Analysis of driving mechanism

A driving mechanism analysis was performed in terms of spatial functional differentiation, timing of urban activities, and regulation and local rules to determine the factors driving population dynamics and their evolving patterns on Xiamen Island. Local clustering analysis (Gi*) based on three significance levels (i.e., $p = 0.01, 0.05,$ and 0.1) was performed to detect the hot spots for each POI category, by which the effects of spatial functional differentiation on population dynamic patterns can be determined. We used the PDI for four time slots (i.e., morning, noon, afternoon, and night), which were defined in subsection 2.2, to detect the effects of the timing of activities on the population dynamic patterns. The population dynamic patterns before and during the BRICS Summit were used to measure the subsequent effect from regulations and local rules.

3. Results

3.1. Spatiotemporal distribution of population

The spatiotemporal distribution of the hot and cold grids at different time slots was represented in the form of space–time cube. Each space–time cube represents one day in a 400 m × 400 m area. Each column contained 10 cubes that represented the timeline from August 25 to September 3 (Fig. 8).

As shown in Fig. 8, most of the cold grids were distributed in the north and south of Xiamen Island, and the hot grids appeared in an “X” shape along the major streets through the downtown area. The remaining insignificant grids were observed at the junction between the hot and cold areas. In some areas, the grid clustering type varied over time, which means that the grid clustering type can change from hot to cold or vice versa. A series of dynamic changes was observed at the east coast of the island, especially in terms of clustering type at the summit venue. Some grids changed from hot to cold during the test slot.

The clustering types varied substantially among the different areas (Fig. 9). The total number of grids with different clustering types in the...
subdistricts indicated a single-type distribution but the majority was shown by hot or cold grids, e.g., the Heshan, Jiangtou, Yuandang, Binhai, Lujiang, Jialian, and Zhonghua subdistricts. However, mixed-type distributions appeared in the Dianqian, Jiahsan, Huli, Lianqian, Xiagang, Wucun, and Kaiyuan subdistricts. The Binhai subdistrict had 1153 cold grids at 99% confidence interval, followed by the Dianqian subdistrict with 676 cold grids and the Lianqian subdistrict with 640 cold grids. The total number of cold grids reached 1261, 965, and 801 for these three subdistricts, considering 90% and 95% confidence intervals. The Heshan subdistrict had 549 hot grids, followed by the Jiangtou subdistrict (526 hot grids). Insignificant grids were observed at the Lianqian, Jiahsan, Yuandang, Dianqian, Heshan, and Huli subdistricts. This suggests that the population in these subdistricts was randomly and even distributed.

Population dynamic clusters across the entire island under various time slots are presented in Fig. 10. Hot grids were primarily observed on August 30–31 (Wednesday and Thursday), accounting for 31.8%–41.9% of the total grids. However, they were rarely observed on August 27 (Sunday), which constituted 16.0%–24.5% of the total grids. These findings suggest that people were resting rather than gathering in public places. Furthermore, on the first half of September 3, which coincided with the BRICS opening morning, hot grids reduced significantly to 26.1% and 25.7%. Local residents were advised not to be out on the morning of the BRICS opening day for security reasons. However, the hot grids for the afternoon and night slots did not decrease significantly compared with the same time slot of the previous days. Furthermore, a slightly increasing trend was observed during the night slot, suggesting that local residents preferred to gather after the opening. The dynamics of the cold grids was almost opposite those of the hot grids. The cold grids were observed the most on August 27 (Sunday) in the morning and noon slots, which constituted 42.9%–52.0% of the total grids, respectively. This finding indicates that the local people were relaxing at home on Sunday rather than working or commuting. Another interesting phenomenon observed was that the number of insignificant grids (or hybrid dynamics) remained at almost the same level for the entire day, and the insignificant grids were normally distributed at the boundary of the hot and cold areas (Fig. 8).
Fig. 9. Total number of hot and cold grids at different subdistricts.

(a) Morning  
(b) Noon  
(c) Afternoon  
(d) Night

Fig. 10. Population dynamics clusters at various time slots.
3.2. Population dynamic patterns at different time slots

A total of 17 population dynamic patterns were identified at various time slots from August 25 to September 3 (Fig. 11). The detailed spatial results are shown in Table 1. The population dynamics was classified into the following patterns: consecutive, intensifying, persistent, and sporadic hot spots; consecutive, intensifying, persistent, and sporadic cold spots; and no pattern. However, no oscillating or historical hot spots were observed. Only a few grids of new and diminishing hot spots; and new, diminishing, oscillating, and historical cold spots were observed throughout the entire test timeline.

As shown in Fig. 11, in general, hot dynamic patterns, such as the consecutive, intensifying, and persistent hot spots appeared in the center of the island along the sides of the main street. Cold dynamic patterns appeared frequently at the edge of the island. The no pattern spots appeared primarily at the boundary of the hot/cold grids throughout the island, indicating that no clear hot or cold dynamic patterns appeared in these grids.

Fig. 11. Population dynamic patterns at various time slots.
Table 1
Spatial distribution of population dynamic patterns in subdistricts at each time slot.

| Time slot    | Subdistrict | New Hot Spot | Consecutive Hot Spot | Intensifying Hot Spot | Persistent Hot Spot | Diminishing Hot Spot | Sporadic Hot Spot | Oscillating Hot Spot | Historical Hot Spot |
|--------------|-------------|--------------|-----------------------|-----------------------|----------------------|----------------------|-------------------|---------------------|----------------------|
| **Morning**  | Dianqian (HL) | —            | —                     | 1                     | 20                   | —                    | 8                 | —                   | —                   |
|              | Heshan (HL)   | —            | —                     | 21                    | —                    | 17                   | —                 | —                   | —                   |
|              | Huli (HL)     | —            | —                     | 8                     | —                    | 6                    | —                 | —                   | —                   |
|              | Jiangtou (HL) | —            | —                     | 33                    | —                    | 11                   | —                 | —                   | —                   |
|              | Jinshan (HL)  | —            | —                     | 4                     | —                    | 11                   | —                 | —                   | —                   |
|              | Binhai (SM)   | —            | —                     | —                     | —                    | —                    | —                 | —                   | —                   |
|              | Jialian (SM)  | 2            | —                     | 24                    | —                    | 7                    | —                 | —                   | —                   |
|              | Kailiyuan (SM) | —           | —                     | 12                    | —                    | 7                    | —                 | —                   | —                   |
|              | Lianqian (SM) | —            | —                     | 5                     | —                    | 10                   | —                 | —                   | —                   |
|              | Lujiang (SM)  | —            | —                     | 11                    | —                    | 2                    | —                 | —                   | —                   |
|              | Wucun (SM)    | —            | —                     | 19                    | —                    | 10                   | —                 | —                   | —                   |
|              | Xiagang (SM)  | —            | —                     | 1                     | —                    | 2                    | —                 | —                   | —                   |
|              | Yuandang (SM) | —            | —                     | 3                     | —                    | 20                   | —                 | —                   | —                   |
|              | Zhonghua (SM) | —            | —                     | 11                    | —                    | —                    | —                 | —                   | —                   |
| **Noon**     | Dianqian (HL) | 9            | 9                     | 2                     | —                    | 2                    | —                 | —                   | —                   |
|              | Heshan (HL)   | —            | —                     | 18                    | —                    | 7                    | —                 | —                   | —                   |
|              | Huli (HL)     | —            | —                     | 7                     | —                    | 2                    | —                 | —                   | —                   |
|              | Jiangtou (HL) | —            | —                     | 45                    | —                    | 6                    | —                 | —                   | —                   |
|              | Jinshan (HL)  | —            | —                     | 6                     | 2                    | 4                    | —                 | —                   | —                   |
|              | Binhai (SM)   | —            | —                     | —                     | —                    | —                    | —                 | —                   | —                   |
|              | Jialian (SM)  | —            | —                     | 30                    | —                    | 2                    | —                 | —                   | —                   |
|              | Kailiyuan (SM) | 1           | 1                     | 19                    | —                    | 1                    | —                 | —                   | —                   |
|              | Lianqian (SM) | —            | —                     | 8                     | 2                    | 5                    | —                 | —                   | —                   |
|              | Lujiang (SM)  | —            | —                     | 13                    | —                    | —                    | —                 | —                   | —                   |
|              | Wucun (SM)    | —            | —                     | 27                    | —                    | —                    | —                 | —                   | —                   |
|              | Xiagang (SM)  | —            | —                     | 3                     | —                    | 1                    | —                 | —                   | —                   |
|              | Yuandang (SM) | —            | —                     | 18                    | —                    | 11                   | —                 | —                   | —                   |
|              | Zhonghua (SM) | —            | —                     | 11                    | —                    | —                    | —                 | —                   | —                   |
| **Afternoon**| Dianqian (HL) | 1            | 2                     | 19                    | —                    | 6                    | —                 | —                   | —                   |
|              | Heshan (HL)   | —            | 1                     | 29                    | —                    | 6                    | —                 | —                   | —                   |
|              | Huli (HL)     | —            | 2                     | 2                     | 9                    | —                    | 3                 | —                   | —                   |
|              | Jiangtou (HL) | —            | 1                     | 2                     | 40                   | 3                    | 6                 | —                   | —                   |
|              | Jinshan (HL)  | —            | 1                     | 6                     | 4                    | —                    | —                 | —                   | —                   |
|              | Binhai (SM)   | —            | —                     | —                     | —                    | —                    | —                 | —                   | —                   |
|              | Jialian (SM)  | —            | —                     | 29                    | —                    | 2                    | —                 | —                   | —                   |
|              | Kailiyuan (SM) | —           | —                     | 10                    | —                    | 2                    | —                 | —                   | —                   |
|              | Lianqian (SM) | —            | 8                     | 9                     | 10                   | 2                    | —                 | —                   | —                   |
|              | Lujiang (SM)  | —            | 3                     | 11                    | —                    | 1                    | —                 | —                   | —                   |
|              | Wucun (SM)    | —            | —                     | 27                    | 2                    | 1                    | —                 | —                   | —                   |
|              | Xiagang (SM)  | —            | —                     | 2                     | —                    | 1                    | —                 | —                   | —                   |
|              | Yuandang (SM) | —            | 3                     | 15                    | 1                    | 1                    | —                 | —                   | —                   |
|              | Zhonghua (SM) | —            | —                     | 11                    | —                    | —                    | —                 | —                   | —                   |
| **Night**    | Dianqian (HL) | 8            | 20                    | 7                     | —                    | 8                    | —                 | —                   | —                   |
|              | Heshan (HL)   | 14           | 29                    | 7                     | —                    | 2                    | —                 | —                   | —                   |

(continued on next page)
| Time slot | Subdistrict | New Hot Spot | Consecutive Hot Spot | Intensifying Hot Spot | Persistent Hot Spot | Diminishing Hot Spot | Sporadic Hot Spot | Oscillating Hot Spot | Historical Hot Spot |
|-----------|-------------|--------------|----------------------|----------------------|---------------------|---------------------|-------------------|---------------------|---------------------|
| Morning   | Hull (HL)   | —            | 7                    | 7                    | —                   | 2                   | —                 | —                   |
|           | Jiangtou (HL) | —            | 1                    | 35                   | 13                  | —                   | 4                 | —                   |
|           | Jinshan (HL) | —            | 10                   | 9                    | —                   | 1                   | —                 | —                   |
|           | Binhai (SM) | —            | —                    | —                    | —                   | —                   | —                 | —                   |
|           | Jialian (SM) | —            | 3                    | 24                   | 4                   | —                   | 2                 | —                   |
|           | Kaiyuan (SM) | —            | 8                    | 9                    | —                   | —                   | 2                 | —                   |
|           | Lianqian (SM) | —            | 14                   | 3                    | 11                  | 5                   | 5                 | —                   |
|           | Lujing (SM) | —            | 3                    | 10                   | —                   | —                   | 1                 | —                   |
|           | Wucun (SM) | —            | 10                   | 20                   | 2                   | —                   | —                 | —                   |
|           | Xiang (SM) | —            | 3                    | 24                   | 4                   | —                   | 2                 | —                   |
|           | Yuandang (SM) | —            | —                    | 1                    | —                   | —                   | 1                 | —                   |
|           | Zhonghua (SM) | —            | 16                   | 7                    | 2                   | —                   | 2                 | —                   |
| Noon      |             |              |                      |                      |                     |                     |                   |                     |
|           |              |              |                      |                      |                     |                     |                   |                     |
|           |              |              |                      |                      |                     |                     |                   |                     |
|           |              |              |                      |                      |                     |                     |                   |                     |
| Afternoon |              |              |                      |                      |                     |                     |                   |                     |
|           |              |              |                      |                      |                     |                     |                   |                     |
|           |              |              |                      |                      |                     |                     |                   |                     |
|           |              |              |                      |                      |                     |                     |                   |                     |

(continued on next page)
Table 1 (continued)

| Time slot | New Cold Spot | Consecutive Cold Spot | Intensifying Cold Spot | Persistent Cold Spot | Diminishing Cold Spot | Sporadic Cold Spot | Oscillating Cold Spot | Historical Cold Spot | No Pattern |
|-----------|---------------|------------------------|------------------------|----------------------|-----------------------|--------------------|-----------------------|----------------------|-----------|
| 6         | 1             | —                      | 1                      | —                    | 3                     | —                  | —                     | —                    | 52        |
| 3         | 7             | —                      | 11                     | —                    | 12                    | —                  | —                     | —                    | 30        |
| —         | —             | —                      | —                      | —                    | —                     | —                  | —                     | —                    | —         |
| 1         | 12            | 7                      | 35                     | —                    | 10                    | —                  | —                     | —                    | 48        |
| —         | 7             | 11                     | 106                    | —                    | 1                     | —                  | —                     | —                    | 7         |
| —         | —             | —                      | —                      | —                    | —                     | —                  | —                     | —                    | —         |
| 6         | 3             | 17                     | 47                     | —                    | 17                    | 1                  | —                     | —                    | 72        |
| —         | —             | —                      | —                      | —                    | —                     | —                  | —                     | —                    | 3         |
| —         | 4             | 3                      | —                      | —                    | 1                     | —                  | —                     | —                    | 8         |
| —         | —             | —                      | —                      | —                    | —                     | —                  | —                     | —                    | 12        |
| 14        | —             | —                      | —                      | —                    | 7                     | —                  | —                     | —                    | 36        |
| —         | —             | —                      | —                      | —                    | —                     | —                  | —                     | —                    | —         |
| Night     | —             | —                      | 28                     | 58                   | 1                     | 14                 | —                     | —                    | 32        |
| —         | —             | 3                      | 2                      | —                    | —                     | —                  | —                     | —                    | 42        |
| —         | —             | 1                      | 12                     | —                    | 9                     | —                  | —                     | —                    | 41        |
| —         | —             | —                      | —                      | —                    | —                     | —                  | —                     | —                    | 4         |
| —         | —             | 8                      | 27                     | —                    | 6                     | —                  | —                     | —                    | 63        |
| —         | —             | 21                     | 99                     | 3                    | 5                     | —                  | —                     | —                    | 4         |
| —         | —             | 2                      | 10                     | 5                    | —                     | —                  | —                     | —                    | 8         |
| —         | —             | 15                     | 41                     | 8                    | 18                    | —                  | 1                     | 71                   | —         |
| —         | —             | —                      | —                      | —                    | —                     | —                  | —                     | —                    | 5         |
| —         | —             | —                      | 6                      | 1                    | —                     | —                  | —                     | —                    | 7         |
| —         | —             | —                      | —                      | —                    | —                     | —                  | —                     | —                    | 13        |
| —         | —             | —                      | 3                      | —                    | —                     | —                  | —                     | —                    | 50        |
| —         | —             | —                      | —                      | —                    | —                     | —                  | —                     | —                    | —         |
Fig. 12. Hot spots of POIs in each category.
The persistent hot grids were the dominant patterns in the center area, including the Lujiang, Zhonghua, Wucun, Jialian, Heshan, and Dianqian subdistricts during the morning, noon and afternoon slots (15.7%–20.0% of the total 1096 grids). Meanwhile, intensifying hot grids were typical during the night slot (16.7% of the total 1096 grids). This trend indicates that people living in and around those grids were quite active at night.

Most of the consecutive, persistent, and intensifying cold grids appeared at the edge of the island (i.e., Dianqian, Jinshian, Lianqian, Binhai, and Huili subdistricts) at four time slots (i.e., morning, noon, afternoon, night), accounting for 30.0%, 33.5%, 34.5% and 30.4% of the total grids, respectively. These cold grids were commonly associated with industrial area, port, newly-built-up, and mountainous areas.

A new hot block was observed in the the Dianqian district in the afternoon and night slots, indicating that people were increasingly attracted to the city under the urbanization. One historical cold block and 18 diminishing cold grids were identified in the Lianqian district in the night slot, suggesting that the population intensities in these cold grids were increasing, and one of those grids did not show any cold characteristics on the last day. In total, 16, 28, and 32 new cold grids were observed for the morning, noon, and afternoon slots, respectively. This means that these grids had never been cold until the last day. Among them, a new cold grid was located exactly near the Xiamen International Convention and Exhibition Center in the noon slot, demonstrating the powerful effect of the regulations on population movement.

3.3. Functional detection for each POI category

The hot spots for each POI category were detected (Fig. 12). The greatest spatial autocorrelation was observed in the CPOIs, followed by the RPOIs, OPOIs, LPOIs, SPOIs, and EPOIs. Most of the hot spots for the commercial and life services, restaurants and snack bars were located in the downtown area from the southwest to the center of the island. Meanwhile, most of the hot spots for scenic and green spaces were distributed in the southwest of the island, which is the traditional downtown area of Xiamen city. The hot spots for offices, leisure, and sports were more scattered around the island, and some hot spots for offices were observed in the north of the island compared with other categories. Hot spots for education and health, however, were primarily distributed along the west coast or east coast of the island. Fig. 12 indicates that most of the insignificant hot spots (cold spots) for each category were in the north and south of the island; these areas were limited to mountains, airports, industrial areas, or wetland parks.

4. Discussions

4.1. Effectiveness of the proposed framework

A new deep mining framework using a nearly unbiased and real-time big data set is proposed herein to characterize the population dynamics and its evolving patterns. Historically, questionnaires, sampling, physical modelling, and theory-driven based extrapolation beyond the observed data have been widely used in demography. Choropleth mapping and dasymetric mapping, for example, are based on Tobler's First Law of Geography, i.e., everything is related to everything else, but near things are more related to each other (Tobler, 1970). They can extrapolate any quantitative variable based on geographical units such as distance to the observed phenomenon. However, these methods always create an impression of an overly uniform area on the edge of two different land-use types (such as urban area and water body), or change abruptly at two continuous subdistricts on the calculated population distribution. Other researchers have recognized these problems (Mennis, 2003; Nikola et al., 2011).

In the big data era, a deluge of earth system data has become available. For instance, social networking and user-generated web content, which has been termed volunteered geographic information (VGI; Goodchild, 2007), has gained more attention in recent years. In particular, the mobile phone has become an important platform by which interactions between individuals and their geographic space can be observed. Large volumes of data are already well beyond dozens of petabytes, with rapidly increasing transmission rates exceeding hundreds of terabytes per day (Agapiou, 2017). Therefore, multiple VGI data have provided a well-informed demographic behavior data source owing to their nearly real-time and full coverage. However, the ability of information processing methods has not increased in pace with data availability. Several major challenges must be addressed, such as the method to extract knowledge from data, as well as the method to derive models and predictions that learn much more from data than traditional physical modeling approaches. This point of view has been emphasized by Reichstein, Camps-Valls, and Stevens (2019).

Although population distributions have been conducted using multiple VGI data previously (Baidu heat map, Twitter, Weibo, Facebook, POI), traditional procedures such as analysis per data frame (static analysis) and descriptive analysis are still the mainstream methods; therefore, the true value of data are still not yet elucidated. For instance, Li et al. (2019) investigated the distribution of urban populations based on the Baidu heat map and urban POI data in Xi’an, China. Two days’ worth of data was used in their study as basic data, i.e., February 18, 2017 (Saturday) as an off; day and February 20, 2017 (Monday) as a working day. Relatively static results were used to represent the population distribution in these two days. However, more data are required to understand the population dynamic pattern and its underlying driving mechanism using big data and deep mining techniques.

Deep/machine learning and data-driven approaches are used increasingly to characterize patterns and gain insights with the support of the ever-increasing stream of big data, which has been garnering attention. For instance, Beer, Reichstein, and Tomelleri (2010) spatially exploited global data-driven estimates of photosynthesis based on machine learning, although the estimates of photosynthesis indicated an overestimation of photosynthesis in the tropical rainforest by climate models. Related data-driven carbon cycle estimates have enabled the calibration of vegetation models and facilitated the explanation of the conundrum regarding the increasing seasonal amplitude of CO2 concentration at high latitudes (Forkel et al., 2016). However, to our knowledge, reports on the spatiotemporal distribution of urban populations and their dynamic patterns using data-driven and deep mining approaches are not available.

Compared with other related studies, our proposed framework offers several advantages. First, we developed a data-driven approach to characterize the population dynamic patterns at each grid from the Baidu heat maps with a total volume of 1.83 GB from August 25 to September 3, 2017 (Fig. A1 in Appendix A) and to identify the underlying driving mechanism based on 149,074 POIs (Figs. 4 and 5), which were acquired from the Baidu map. Second, we developed a machine learning method integrating a multiple testing operation with an FDR procedure to optimize the method to identify the spatiotemporal distribution of the population. Third, a space-time cube method was designed to statistically represent the hot/cold spots at one grid per day at a specific time slot (Fig. 8).

The framework proposed in this study can be easily applied for extracting useful information from other spatiotemporal big data sets. Using the precipitation radar data as an example, we should focus more on the consecutive, intensifying, and persistent hot areas to mitigate against flooding. However, the drought probabilities in consecutive, intensifying, and persistent cold areas are higher than those in other places. This approach may offer a fresh data-driven pertaining to water-related hazards modeling and forecasting compared with traditional physic-based models.
4.2. Pattern to process in population dynamics

Based on the results of this study, we discovered that urban spatial functional differentiation is the most essential factor that can affect the spatial distribution of a population (Figs. 11 and 12). In summary, the distribution patterns of a population are significantly spatially correlated with the distribution patterns of POIs. However, the hot spots of a population are not singly related to each type of POI, and mixed use is the most significant driving force on population distribution. Since the 1980s, mixed use has regained attention by theories such as sustainable development and new urbanism (Grant, 2002). This point of view was advocated by Yue et al. (2017). China has witnessed rapid development since the reform and the opening up policy decades before. However, infrastructures, such as life services, hospitals, universities, finance, and business offices are primarily located downtown in urban areas, resulting in uneven social or economic developments compared with rural areas. He et al. (2019) reported the same results. Meanwhile, the cold spots of the population were highly consistent with the cold spots of POIs for each category, which were concentrated in mountains, industrial parks, and newly developed residential areas.

The timing of various urban activities is an internal factor that determines the population dynamics. Hot grids were primarily observed on work days. During the weekend, however, people typically relax at home or visit relatives and friends. These types of lifestyles resulted in a more scattered clustering type (Figs. 9 and 10); similar findings have been reported in previous studies (Leng et al., 2015; Li et al., 2019). This investigation also revealed that a fixed number of people distributed and moved randomly and evenly throughout the area and time, resulting in many hybrid grids. These hybrid grids were typically observed at the boundary of hot and cold areas, which was not reported in previous studies.

Regulations and local rules are external factors that significantly affected the spatiotemporal distribution of the population. Quantitative evidence from this investigation revealed that the number of people decreased before and during the BRICS meeting days, whereas cold grids were often observed at the edge of the city during the weekend. The distribution patterns of the population were significantly spatially correlated to the distribution patterns of POIs: the hot spots of the population were closely associated with the commercial and life services, restaurants, snack bars, offices, and their overlays; the cold spots of the population were highly consistent with the cold spots of the POIs, which were concentrated in the mountains, industrial parks, and newly developed residential areas. The BRICS Summit 2017 held in Xiamen significantly affected the spatiotemporal distribution of the population at that time. The framework proposed in this study demonstrated its effectiveness in delineating urban population dynamics and promising potential for informing urban preparedness for emergencies and sustainable urban development of interactions between population distribution and land use types.

5. Conclusion

The data-driven framework proposed in this study offers new insights into the urban population dynamics and its driving mechanism in Xiamen Island. Hot grids were primarily clustered along the main street through the downtown area during working days, whereas cold grids were often observed at the edge of the city during the weekend. The distribution patterns of the population were significantly spatially correlated to the distribution patterns of POIs: the hot spots of the population were closely associated with the commercial and life services, restaurants, snack bars, offices, and their overlays; the cold spots of the population were highly consistent with the cold spots of the POIs, which were concentrated in the mountains, industrial parks, and newly developed residential areas. The BRICS Summit 2017 held in Xiamen significantly affected the spatiotemporal distribution of the population at that time. The framework proposed in this study demonstrated its effectiveness in delineating urban population dynamics and promising potential for informing urban preparedness for emergencies and sustainable urban development of interactions between population distribution and land use types.

Declaration of Competing Interest

The authors declared that they have no conflicts of interest to this work. We declare that we do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted.

Acknowledgements

This research was supported by the National Natural Science Foundation of China (Grant No. 4161101257; Grant No. 41971231; Grant No. 41471154), and the Fujian Provincial Key Laboratory of Data-Intensive Computing. We thank Dr. Wang Shih-Chi from the U.S. EPA for his assistance with English writing and anonymous reviewers for constructive feedbacks and comments that helped improve this paper.
Appendix A

Fig. A1. Part of the captured Baidu heat maps at different time.
## Table B1

### Explanation on the dynamic patterns.

| Pattern name               | Detailed explanation                                                                                                                                 |
|----------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------|
| New Hot Spot               | A grid that is a statistically significant hot spot for the final time step and has never been a statistically significant hot spot before.              |
| Consecutive Hot Spot       | A grid with a single uninterrupted series of statistically significant hot spot in the final time-step intervals. The grid has never been a statistically significant hot spot prior to the final hot spot series and less than ninety percent of all time series are statistically significant hot spots. |
| Intensifying Hot Spot      | A grid that has been a statistically significant hot spot for ninety percent of the time-step intervals, including the final time step, and the intensity of clustering in each time step is increasing overall and that increase is statistically significant. |
| Persistent Hot Spot        | A grid that has been a statistically significant hot spot for ninety percent of the time-step intervals with no discernible trend indicating an increase or decrease in the intensity of clustering over time. |
| Diminishing Hot Spot       | A grid that has been a statistically significant hot spot for ninety percent of the time-step intervals, including the final time step, and the intensity of clustering in each time step is decreasing overall and that decrease is statistically significant. |
| Sporadic Hot Spot          | A grid that is an on-again then off-again hot spot. Less than ninety percent of the time-step intervals have been statistically significant hot spots and none of the time-step intervals have been statistically significant cold spots. |
| Oscillating Hot Spot       | A statistically significant hot spot for the final time-step interval that has a history of also being a statistically significant cold spot during a prior time step. Less than ninety percent of the time-step intervals have been statistically significant hot spots. |
| Historical Hot Spot        | The most recent time period is hot, and at least ninety percent of the time-step intervals have been statistically significant hot spots. |
| New Cold Spot              | A grid that is a statistically significant cold spot for the final time step and has never been a statistically significant cold spot before.              |
| Consecutive Cold Spot      | A grid with a single uninterrupted series of statistically significant cold spot in the final time-step intervals. The grid has never been a statistically significant cold spot prior to the final cold spot series and less than ninety percent of all time series are statistically significant cold spots. |
| Intensifying Cold Spot     | A grid that has been a statistically significant cold spot for ninety percent of the time-step intervals, including the final time step, and the intensity of clustering in each time step is increasing overall and that increase is statistically significant. |
| Persistent Cold Spot       | A grid that has been a statistically significant cold spot for ninety percent of the time-step intervals with no discernible trend, indicating an increase or decrease in the intensity of clustering over time. |
| Diminishing Cold Spot      | A grid that has been a statistically significant cold spot for ninety percent of the time-step intervals, including the final time step, and the intensity of clustering in each time step is decreasing overall and that decrease is statistically significant. |
| Sporadic Cold Spot         | A grid that is an on-again then off-again cold spot. Less than ninety percent of the time-step intervals have been statistically significant cold spots and none of the time-step intervals have been statistically significant hot spots. |
| Oscillating Cold Spot      | A statistically significant cold spot for the final time-step interval that has a history of also being a statistically significant hot spot during a prior time step. Less than ninety percent of the time-step intervals have been statistically significant cold spots. |
| Historical Cold Spot       | The most recent time period is not cold, but at least ninety percent of the time-step intervals have been statistically significant cold spots. |
| No Pattern                 | Does not fall into any of the hot or cold spot patterns defined above.                                                                                   |
Mennis, J. (2003). Generating surface models of population using dasymetric mapping. Professional Geographer, 55(1), 31–42.

Nikola, K., Branislav, B., Milan, K., & Dragutin, T. I. (2011). Modelling the spatial distribution of Vojvodina’s population by using dasymetric method. Spatial. 2011(24), 45–50.

Ord, J. K., & Getis, A. (1995). Local spatial autocorrelation statistics: Distributional issues and an application. Geographical Analysis, 27(4), 286–306.

Pan, J., & Lai, J. (2019). Spatial pattern of population mobility among cities in China: Case study of the national day plus mid-autumn festival based on tencent migration data. Cities, 94, 55–69.

Reichstein, M., Camps-Valls, G., Stevens, B., et al. (2019). Deep learning and process understanding for data-driven earth system science. Nature, 566, 195–204. https://doi.org/10.1038/s41586-019-0912-1.

Santos-Reyes, J., & Olmos-Peña, S. (2017). Analysis of the “News divine” stampede disaster. Safety Science, 91, 11–23.

Shen, Y., & Karimi, K. (2016). Urban function connectivity: Characterisation of functional urban streets with social media check-in data. Cities, 55, 9–21.

Standardization Administration of the People’s Republic of China (2011). Code for classification of urban land use and planning standards of development land (GB50137–2011).

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region.

Yang, Y., Liu, Y., Li, Y., & Li, J. (2018). Measure of urban-rural transformation in Beijing-Tianjin-Hebei region in the new millennium: Population-land-industry perspective. Land Use Policy, 79, 595–608.

Yue, Y., Zhuang, Y., Yeh, A. G. O., Xie, J., Ma, C., & Li, Q. (2017). Measurements of POI-based mixed use and their relationships with neighbourhood vibrancy. International Journal of Geographical Information Science, 31(4), 658–675.

Zhang, N., Huang, H., Duarte, M., & Zhang, J. J. (2016). Dynamic population flow based risk analysis of infectious disease propagation in a metropolis. Environment International, 94, 369–379.

Zhen, F., Cao, Y., Qin, X., & Wang, B. (2017). Delineation of an urban agglomeration boundary based on Sina Weibo microblog “check-in” data: A case study of the Yangtze river delta. Cities, 66, 180–191.

Zhiqiang, W., & Zhongnan, Y. (2016). Research on urban spatial structure based on Baidu heat map: A case study on the central city of Shanghai. City Planning Review, 40(4), 30–40.

Zhou, J., Pei, H., & Wu, H. (2018). Early warning of human crowds based on query data from Baidu maps: Analysis based on Shanghai stampede. In Z. Shen, & M. Li (Eds.). Big data support of urban planning and management: The experience in China (pp. 19–41). Cham: Springer International Publishing.