

The Prediction of Finely-Grained Spatiotemporal Relative Human Population Density Distributions in China

ZHI ZHENG^{1,2,3,4} AND GUANGYUAN ZHANG⁵

¹Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences (CAS), Beijing 100101, China

²Key Laboratory of Regional Sustainable Development Modeling, Chinese Academy of Sciences (CAS), Beijing 100101, China

³College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China

⁴Department of Geography, National University of Singapore, Singapore 119077

⁵School of Electronic Engineering and Computer Science, Queen Mary University of London, London E1 4NS, U.K.

Corresponding author: Guangyuan Zhang (guangyuan.zhang@qmul.ac.uk)

This work was supported in part by the Queen Mary University of London, and in part by the China Scholarship Council (CSC).

ABSTRACT China's transportation industry has been experiencing huge changes and the travelling frequency of citizens becomes higher and higher and more and more diversified in a short time period. The analysis and deep research on the short-term change of population density in the city-level spatial resolution are worthy of further exploration. In this study, we first used two linear regression models to build relationships between the 2010 census density, predicted 2020 census density and the Tencent density respectively to test the usability of Tencent positioning data. The Pearson's correlation coefficients $r = 0.58$ ($p < 0.01$) and 0.54 ($p < 0.01$) demonstrates good positive correlations between the ground truth (census data) and the geographic spatiotemporal big data (Tencent positioning data), which could be used to represent the relative human population density distribution in China. Then we use the human population distribution based on the Tencent positioning dataset of China in every hour of the first 21 days, to predict the hourly distribution in the next week by seasonal autoregressive-integrated-moving-average (SARIMA) and a convolutional long short-term memory (ConvLSTM) models respectively, with the 50×50 km of spatial resolution. The total average of the ConvLSTM model's Root Mean Square Error (RMSE) is 139.0, while the SARIMA's one is three times greater than the value. And the coefficient of determination (R^2) values of ConvLSTM model is higher than 0.9, while the other ones are about 0.78. Comparing the two results in both time and space concludes that: the evaluation results reflected by RMSE and R^2 showed that the two models are both suitable for the prediction of Tencent density distribution in finely-grained time and space. Nevertheless, the predicted density correlated much better with the tested data at temporal and spatial scales when using ConvLSTM compared to SARIMA, and the capability of prediction in space by ConvLSTM model is more stable.

INDEX TERMS Prediction, human population density distribution, SARIMA, ConvLSTM, Tencent positioning data, deep learning, geographic spatiotemporal big data.

I. INTRODUCTION

Before the end of the human population growth in the foreseeable future is likely to come [1], the future development of population size and structure is of importance since planning in many areas of politics and business is conducted based on expectations about the future makeup of the population [2]. Furthermore, there is another similar concept

The associate editor coordinating the review of this manuscript and approving it for publication was Guangcun Shan.

with the population size, population density, also play a vital role in the related fields. With the internal structure of cities getting more complex, the study of urban population density distribution, as a result, also becomes more [3]. For example, it is of considerable significance for urban planning, such as the division of urban functional areas, the allocation of public service facilities and the urban traffic planning [4], and the protection of urban public safety and crisis management [5]. Thus, in terms of the administrator of the nation, predicting the changings of human population density distribution

would be necessary to understand the situation and development of countries, and manage the policy of society. For instance, foreknowing the hot spot reflected by people density can provide urban and transportation planning suggestions for the government [6]. Besides the situation of the urban area of the nationals, the human who stays at the rural areas is more worthy to be understood and predicted because they are always away from the administrators and with less part of the nationals in the modern world. Thus, the global situation of nations in the future put a significant impact on the management of the government, and bring the public a safe and useful information, especially in a large-area country with more areas of rural regions such as China, which is the first most populous country in the world with the high rate of increase in population.

China's transportation industry has been experiencing huge changes in a leap-forward speed with brilliant achievements since 1978 [7]. Although the rapid expansion of High-Speed Rail (HSR) has caused the falling air passengers' number in competition routes, the amount of the passengers for the travelling over the cities, provinces and countries have increased significantly [8], which means the travelling frequency of citizens in China becomes higher and higher and more and more diversified in a short time period such as a half-day. This travelling phenomenon worth deeper analysis, measurement, and even prediction.

In detail, we analyze the travelling passenger volumes by the two main transportations, train and flight. Based on a report of the 2019 development of the high-speed railway in China¹ released by World Bank Group,² The passenger mileage competition between civil aviation and HSR is mainly concentrated in the travelling distance of 800-1200 km, therefore, we simply divide the travelling crowd into those who are less than 1000 km by train (including both HSR and traditional train) and those who exceed by plane. According to the Civil Aviation Industry Development Statistical Bulletin of China,³ in 2019, the total passenger traffic volume of the whole industry was 659.9342 million, which means on average, at least 75000 people travel between two cities more than 1000 kilometres apart per hour. While in terms of the passenger who prefer take traditional trains or HSR, the total volume of the passenger is about 3.66 billion, which is equivalent to that at least 417 thousand people travel between two cities less than 1000 kilometres apart per hour, according to the Railway Statistical Bulletin in 2019 of China.⁴ Thus, the analysis and deep research on the short-term change of population density in the city-level spatial resolution (defined as the smallest cell represents an area similar to that of a city) is worthy of further exploration.

¹<https://documents.worldbank.org/curated/en/69739156777005065/pdf/Chinas-High-Speed-Rail-Development.pdf>

²<https://www.worldbank.org/>

³<http://www.mot.gov.cn/tongjishuju/minhang/202006/P020200702602086869690.pdf>

⁴<http://www.nra.gov.cn/xxgkml/xxgk/xxgkml/202004/P020200430612271004598.pdf>

At present, many studies focus on the short-term prediction of population density. However, predicting the population size or density distribution in both finely-grained spatial and temporal scales in a large area, such as China, with a land area of about 9.6 million square kilometres, had not received much attention because the related data in a large area over a dynamic time period was hard to obtain in the past several years until the geographic spatiotemporal big data [9] appeared recently. Note that the finely-grained space and time are a relative quantity, which depends on the ratio of resolution to the total study area. For example, the resolution is a block or a building while the study area is city-wide. In the same way, the resolution is a city while the study area is country-wide could also be regarded as a finely-grained work, which has been defined and used in our study.

Geographic spatiotemporal big data is defined as due to continuous spatiotemporal data are extracted and stored through discrete samples in geography so that it can be seen as a time series correlated in space [10], namely spatiotemporal series. With the constant improvement of space data acquisition and data storage capacity [11], more and more spatiotemporal data are added to analyze interactions among them in geospatial, such as time series of remote sensing images and different kinds of socio-economic statistical data. That multi-source of spatiotemporal series constitutes geographic spatiotemporal big data. Several types of geographic spatiotemporal big data are suitable to extract the people density in finely-grained spatial and temporal scales, for example, the Call Detail Record (CDR) data [12] which is obtained by Subscriber Identity Module (SIM) card in the smartphone, and the GNSS data-oriented from applications in a smartphone such as Baidu heat map data [13]. However, the basis of the usage of these types of big data to compute the global human population density is the much high penetration of the smartphones [14], while there are insufficient researches to compare the datasets with the ground truth to test their robustness because of there is no ground truth of dynamic human population density distribution in finely-grained spatial, temporal scales. Because the penetration of the smartphone cannot reach absolute 100% in the real world, we assume that the population using mobile phones is spatially randomly distributed among all people, and named the human population as the relative human population.

Furthermore, in the related research, there are also many defects in the use of data and the selection of prediction models, especially the usability verification of spatiotemporal big data and the accuracy comparison of prediction models. In addition, the current research based on the big data and deep learning to predict high-precision (relative) spatiotemporal human population density distribution has not appeared, which means our research will be devoted to filling the gap in this field.

In this study, based on a geographic spatiotemporal big data, Tencent positioning big data, we compute the relative human population distribution based on the dataset in every hour of the first three weeks (21 days), to predict the

hourly distribution in the next week (7 days) by SARIMA and ConvLSTM models respectively, and finally to compare the accuracy. The spatial resolution in this study is set as 50 km, which is considered as the city-level resolution. However, as Tencent positioning big data belongs to Volunteered Geographic Information (VGI) [15], its usability needs to be confirmed because of its diversity of authorship [16]. However, up to now, there is no open-source data source with high spatial-temporal accuracy in the databases covering the whole of China, so we can only consider other similar data for comparative analysis. Census data is a kind of open-source and national dataset, which can be compared with Tencent big data through certain processing to prove its usability. The novel contributions of this article are as follows.

- 1) We first detect the correlation between Tencent positioning data and census data, to test the usability and robustness to use Tencent density to represent relative human population density in China (including territorial land area only, excluding territorial sea area. And in the last part of this article, we only use China to represent this).
- 2) We first predict the human population distribution in a finely-grained spatial and temporal resolutions for the whole of China by a traditional model SARIMA and a novel deep learning model ConvLSTM, and then to compare the results to conclude the most accurate one.

The remainder of this article is organized as follows: Section II presents the related works while Section III introduced the dataset and the method we used. In Section IV, we evaluate their accuracy and compare these three methods' abilities to predict the relative human population distribution both in time and space. In Section V, we present our conclusions and our thoughts for future research.

II. RELATED WORKS

The initial research on population prediction dates back to 1789, Malthus presents that the population will exhibit exponential growth characteristics without any restrictions [17]. Then plenty of model of prediction of people population raised. Verhulst proposed the Logistic Function for the relationship between population size, population growth rate and environmental carrying capacity, which considers the constraints of the environment and resources on the population based on the theory of Malthus [18]. Later, Markov Chain model was applied in the population prediction to solve the potential errors brought by the changings population growth rate [19]. Nathan pioneered the idea of using matrix multiplication for population prediction and established a matrix model [20] which considers the gender, age, fertility and survival rate of people. The method was then improved to Leslie matrix [21]. However, these studies do not consider the spatial factor and the spatial resolution only remains at an administrative unit such as country or state level. And the temporal scale and the resolution of them are coarse-grained such as predicting people in the future decades with a year interval. At the start, more finely-grained prediction for the

population distribution both in space and time focus on the small area, such as at the minor civil division (MCD) governmental level in Wisconsin, the USA since 1960 [22] because population change in small areas is more volatile than in more extensive areas because migration often accounts for a more substantial fraction of population change at the local level than it does at the national level [23]. There are some works that focus on the human population density prediction in a ground-truth level based on the video and image datasets from surveillance networks. For example, Hiroaki et al presented a new visual forecasting task named crowd density forecasting by modelling patch-based dynamics, but the prediction region is only camera-scale [24].

Finely-grained prediction works in space and time have taken place on the crowd flow, which provides similar scientific merits with population density changing by focusing on a speed that objects in a crowd move in space, i.e., Samuel et al proposed an approach to instantly predict the long-term flow of crowds in arbitrarily large, realistic environments [25]. Ziqian et al propose DeepSTN+, a deep learning-based convolutional model, to predict crowd flows in the metropolis. The spatial resolution in the study is about 20 km, while the time intervals are 30 minute and one hour, but the whole study areas are two cities, Beijing and New York [26]. The study [27] proposed DeepFlowFlex, a graph-based model to jointly predict inflows and outflows for each region of arbitrary shape and size in a city. The spatial and temporal resolutions are real finely-grained in the prediction, which are 2 km and half an hour, but the study regions are one city as well.

In terms of the model of predicting population size or the population density, after the combination of the autoregressive (AR) and moving average (MA), autoregressive-moving-average (ARMA) model [28] has achieved excellent results in short-term population forecasting and has been widely used [12]. But in terms of a stationary time series dataset, whose statistical properties such as mean, variance, autocorrelation, etc. are all constant over time, Autoregressive-integrated-moving-average (ARIMA) model can eliminate some errors by stationarizing a time series through differencing [29]. While in terms of the time series dataset that has seasonality that is the presence of variations that occur at specific regular intervals less than a year, such as weekly, monthly, or quarterly, this repeating cycle may obscure the signal that we wish to model when forecasting, and in turn may provide a strong message to our predictive models. But seasonal ARIMA (SARIMA) is an improved model to solve the problem [29]. These traditional methods above are based on the principle of statistics, which are mature, simple and easy to promote. However, considering limited influencing factors of the growth of population, accuracy requirements for training data make them generate more errors about the forecasted results. But the emergence of the artificial neural network model [30] gave the researchers a new thought of the prediction of population. The characteristics of high learning precision, fast calculation speed and

strong robustness make it fashionable to apply [12]. With the development of machine learning and deep learning, there are many prediction models to predict time series data, e.g., long short-term memory (LSTM) model [31], while a Convolution Long Short-Term Memory (ConvLSTM) is the first deep learning model which consider both spatial-temporal autocorrelations. Thus, the model is expected to play an excellent performance on the prediction work based upon spatial-temporal data [32].

VGI are a rich and, often inexpensive, source of information but still exists some trusting issue such as the integration with authoritative and other datasets [33]. VGI could be used as the main dataset to analyze and provide key information, but also be regarded as an auxiliary to adjust none-VGI, i.e., census data. As an auxiliary case, a convolutional neural network (CNN) architecture is proposed to validate landslide photos collected by citizens or nonexperts and integrated into mobile- and web-based Geographic Information System (GIS) environment designed specifically for a landslide CitSci project [34]. The role of VGI in this study is providing a larger amount of training data to improve identification accuracy. To be as the main datasets, the assessing or usability analysis is necessary and a large body of work proposes using trust as a proxy for VGI quality [35], [36]. For example, the research [16] presented a formulaic model that addresses VGI quality issues, by quantifying trust in VGI. Further, a review [37] has classified the related literature into 30 methods that can be used to assess one or more of the 17 quality measures and indicators coming across in the literature for a map, image, and text-based VGI, respectively. Among the methods, comparison with ground truth data to test the VGI's quality with corresponding comparative analysis is a popular way. However, in terms of some specific dataset, i.e., Tencent Position data, there is no actual ground truth, but the similar datasets, such as census data, could be utilized after adequate comparative analysis.

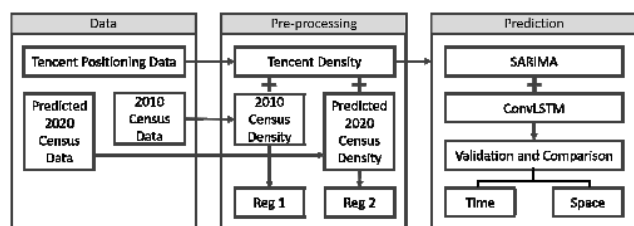


FIGURE 1. Data and method overview to perform a spatiotemporal prediction of the relative human population density distributions in China.

III. DATA AND METHOD

The structure of this section consists of three parts. Firstly, we introduce the dataset we use, and then to explain the pre-processing on the dataset. Finally, the methods of the prediction are reported with the validation and comparison strategies. The overview of this section is shown in figure 1.

In this study, we mainly use two datasets including the Tencent positioning data, which is the primary input data

into the prediction models, and the census data that is used to explain the authenticity of the Tencent positioning data. In the second part, we map two types of density distributions that are computed by the two datasets and then use regression model illustrates the practical significance of the Tencent density distribution. And in the final part, we apply two prediction models, including a traditional SARIMA model and a deep learning ConvLSTM model on the Tencent density distributions. Furthermore, the validation and comparison strategies of the two results would also be introduced in this part.

A. DATA

1) TENCENT POSITIONING DATA

WeChat from Tencent Company is currently the most popular social software in China, and according to the WeChat Annual Data Report 2018,⁵ there have been more than 1.01 billion users log in to WeChat every day. In larger cities in China such as Beijing, Shanghai and Guangzhou, the penetration rate of WeChat has reached at least 96% of citizens according to Kantar China Social Media Impact Report 2018.⁶ Besides WeChat software, there are several other mobile applications produced by Tencent Company such as social application QQ, Tencent Video and QQ Browser, which are also popular with high penetration rates. All these applications need their users to turn on positioning permissions so that Tencent Company can get the location of users using their smartphone GNSS once we install one of them. Based on the data of users' location, Tencent presents a real-time platform to show the heat map of their users⁷ and the density distribution could be easily mapped by GIS software such as ArcGIS.⁸

The spatial structure of the Tencent positioning data consists of vector points that are stored as SHP format file,⁹ and they regularly scatter at the map of China. Under the projected coordinate system of WGS 1984 World Mercator,¹⁰ they are arranged evenly shown in figure 2 (a) and (b), each point is 5566 meters away from the point on its east or west (figure 2 (b), WE Line), while 7235 meters away from the point on its north or south (figure 2 (b), NS Line). For each SHP file, there are millions of points that cover the whole China land at a specific time, for example, 6:00 A.M, 01 September 2019. For each point, there is a feature named Recorded Number of People (RNP) that represents the related people number on this point at this time. However, because in this study, we focus on China land region, we convert the projected coordinate system of all SHP files from WGS

⁵<http://media.people.com.cn/n1/2019/01/10/c40606-30513560.html>

⁶<https://cn-en.kantar.com/media/social/2018/kantar-china-social-media-impact-report-2018/>

⁷<https://heat.qq.com/index.php>

⁸<https://www.esri.com/en-us/arcgis/about-arcgis/overview>

⁹<https://desktop.arcgis.com/en/arcmap/10.3/manage-data/shapefiles/what-is-a-shapefile.htm>

¹⁰<https://desktop.arcgis.com/en/arcmap/latest/map/projections/mercator.htm>

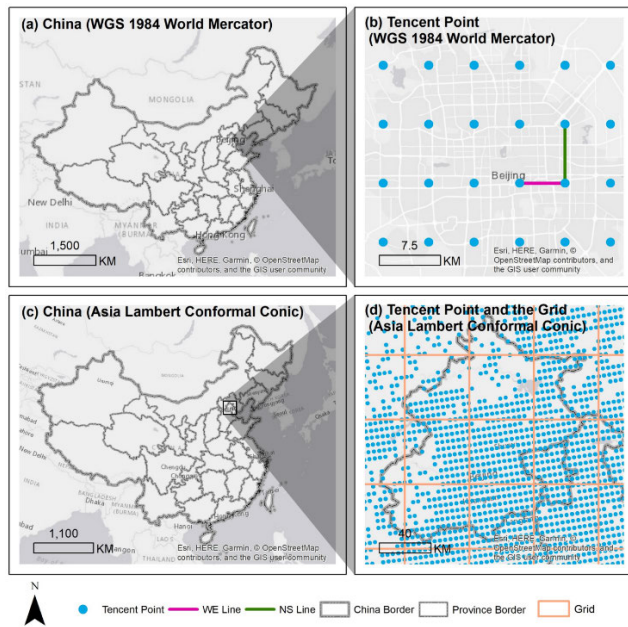


FIGURE 2. The map of China under different projected coordinate systems with details of the distributions of Tencent Point and the grid: (a) is the China map under WGS 1984 World Mercator projection, (b) is an example place (one region in Beijing) to illustrate the distribution of the Tencent Point under the WGS 1984 World Mercator projection. (c) shows the China map under Asia Lambert Conformal Conic projection, while (d) regards whole Beijing as an example to show the detail of the Tencent Point and the grid.

1984 World Mercator to Asia Lambert Conformal Conic,¹¹ which is a general projection system for China (shown in figure 2(c)). It is worth mentioning that not all region has evenly points distribution, such as the mountain and forest area, because there is less activity of people.

In terms of the temporal structure of the Tencent positioning data, the resource of the data updates secondly, which means every second, there is an SHP file generated from the website. But in this study, we use the SHP files that are produced at the first second of every hour to represent the Tencent people distribution on this hour. We get the data from 1 September 2019 to 7 October 2019 (37 days), which means SHP files are representing 888 hours' distributions of Tencent recorded people.

2) CENSUS DATA

In this study, we used census data from China's sixth census in 2010¹² which is the newest census in China, to compare with the Tencent positioning data. The spatial unit of this census data is the prefecture-level city, which has 333 units in 2010,¹³ we call them administrative units in the last part of the paper. For each administrative unit, there is a corresponding number of human populations. The spatial distribution of the units is shown in figure 4(a), and the further process of

¹¹<https://desktop.arcgis.com/zh-cn/arcmap/latest/map/projections/lambert-conformal-conic.htm>

¹²<http://www.stats.gov.cn/tjsj/pcsj/>

¹³<http://xzqh.mca.gov.cn/statistics/2010.html>

the comparison with the Tencent recorded people would be introduced in section III.A.

However, there are two key problems when we use census data from 2010. First, we notice that the Tencent data we use is in 2019, which is 9 years after currently the newest census data in China. Second, the census data reflects the relative static of the population distribution while the Tencent one is relative dynamic. Thus, before the comparison, the comparability between the two datasets need more analysis. In terms of the two issues, we present two corresponding strategies to address them.

First, although there is nine years' gap between the two datasets, according to the National Population Development Plan (2016-2030)¹⁴ that was released by The State Council of the People's Republic of China, the expected population of China in 2020 is 1.42 billion while in 2010 there was 1.37 billion, which means there would be only 3.5% of changing of the population in 2020 from 2010. However, China has experienced nearly 30 years of rapid socio-economic development, and with the change of natural population growth and spatial migration, the spatial pattern of population in different provinces or cities is also evolving synchronously [38], which means we need to consider the inner changes not only the global changing of China to measure the comparability. According to a previous study about the spatial pattern with the evolution of Chinese provincial population [39], the provinces or cities of rapid population change, which are classed by the index of population change intensity (is computed based on the variance of population density over the years), include Beijing, Tianjin, Jiangsu, Shanghai, Zhejiang, Guangdong, Henan, Anhui, Chongqing, Hubei, Liaoning and Shandong. In this study, the authors predicted the final population density in each of 12 provinces or cities in 2020. Thus, we just need to update the densities of the 12 areas of the census data in 2010, then to compare Tencent density and the predicted population density.

Second, the census data reflect a relative static density distribution for each administrative unit such as a province or city because the Census Bureau of a government only counts the people that settle down somewhere, not other places such as a temporary residence when travelling. However, the Tencent data can represent the real-time distributions of people, which means they are relatively dynamic. Thus, to eliminate the difference between the dynamic and static, we use the average relative population density distribution that is computed by the Tencent positioning data in the study period to compared with census data or predicted census data.

While we use the dataset of Tencent in September and October 2019 that is much closer to 2020, thus, we consider these two datasets to be comparable.

B. PRE-PROCESSING

In terms of the census data, the area of the administrative unit ranges from 1440 km² (Zhoushan, Zhejiang) to 263953 km²

¹⁴http://www.gov.cn/zhengce/content/2017-01/25/content_5163309.htm

(Hulunbuir, Inner Mongolia). In order to reflect the human population distribution more intuitively, we created a fishnet-like grid to divide the whole of China (excluding the territorial sea area) into square cells. We define a frame region which covers the China land area as a 5000×5000 km square area. To make the data structure more suitable for the prediction models, the grid in this area consists of 100×100 cells, and each cell is a 50×50 km square. Because the area of each cell is 2500 km^2 where there are 12 administrative unit is smaller than this, which means the cells in the grid are smaller than 96% of units. Thus, the spatial resolution of the grid is much higher compared to the census data.

In the next step, for each cell, we cumulative the RNP of all Tencent points in the corresponding cell, to create a new feature for the cell, named RNP of the cell (RNPC). Then we calculated the human population density for each cell and named the final distribution as the Average Tencent Density, by the equation:

$$\text{Tencent Density}_i = \text{RNPC}_i / 2500 \text{ km}^2 \quad (1)$$

where i is the index of the 10,000 cells. Finally, we map a new type of human population density distribution. The same process is operated on all 888 files, thus there are 888 distributions of Tencent Density which would be input into the prediction models.

However, because we need to compare the 2010 census data, predicted 2020 census data with the Tencent data respectively, first, the predicted 2020 census data would be computed based on the previous research [39]. In detail, for the 12 provinces or cities that would have more rapidly changed, we use the final predicted results of the research, and for the other provinces or cities, we just use the densities of the 2010 census data. Then, we need to unify the space and the time of these three datasets. In the temporal scale, we calculate the average Tencent Density of 888 files for every cell, map a final Average Tencent Density (ATD) distribution to compared with the census datasets. In the spatial scale, first, we calculate the density of each administrative unit by the equation:

$$\text{Census Density}_j = \text{RNP}_j / A_j \quad (2)$$

where j is the index of the 333 administrative units and the A_j is the corresponding area of the unit. In the next step, we convert the ATD into the same spatial format as Census Density, naming it as the Converted Census Density (CCD), by the equation:

$$\text{CCD}_j = \sum_{k=1}^n (\text{Tencent Density}_{i,k} \times A_k) / A_j \quad (3)$$

where n is the number of components that administrative unit j is divided by Tencent grid, i is the index of the 10,000 cells and the k is the index of these components. A_k is the area of the component k , and finally, the CCDs of all units have been computed as a new distribution.

Because the CCD still represents the relative density of the human population, we cannot compare the CCD with the Census Density by their absolute values. We classify and visualize the density of the 333 administrative units for both two distributions into 32 levels by the Natural breaks (Jenks) method [40], where 32 is the maximum quantity of this method in the ArcGIS symbology system.¹⁵ And now we have a new feature of the density level value for each administrative unit for both density distributions and the results would be shown in section III.A.. Finally, we extract 333 tuples that each tuple consists of CCD density level, 2010 Census Density and predicted 2020 Census Density level for the unit. We build two simple linear regression [41], [42] models to these datasets to see if there are linear relationships between the Tencent density distributions and other two census density distributions respectively.

C. PREDICTION

In our study, we apply two prediction models on the dataset, SARIMA and ConvLSTM. The input of data is the 888 Tencent Density distributions. However, the approaches of inputting data to these two models are different. In terms of the SARIMA model, we regard each cell as a single input, which means every cell has a time series consists 888 density values, and the model would be built for 10,000 times because there are 10,000 cells. On the other hand, because the convolution would be used in the training of ConvLSTM model, the input of this process would regard a grid as an entirety. This section would introduce the two models in detail.

1) SARIMA MODEL

The SARIMA model applied in this research is the most general form of a univariate class of models originally presented by Box and Jenkins [29]. It has been extensively studied and used in different fields, such as economy, industry and, more recently, in public health. In this study, this model is utilized in the prediction of relative human density time series.

Because human population density distribution changed with time in a specific spatial area, such as one city or province, we need to determine that if they have a trend component in the dataset. The initial process is that a first-order difference of the series z_t is given by w_t , the difference between points in the series on unit apart, calculated as $w_t = z_t - z_{t-1}$. We may also write w_t in terms of the backward shift operator B as $w_t = (1 - B)z_t$, and so the d th order differencing is obtained as $(1 - B)^d z_t$.

Besides the judgement of the trendiness by Augmented Dickey-Fuller (ADF) test [43], seasonality needs to be exhibited as well. In our case, the population density distribution which is extracted from Tencent positioning data has obvious daily seasonality because the people's usage of their smartphone is in line with their basic schedule, for example, when people go to sleep at night, they would not use their

¹⁵<https://pro.arcgis.com/en/pro-app/help/mapping/layer-properties/symbolize-feature-layers.htm>

phone which causing the less information has been recorded by Tencent software, and when they get up in the morning, the smartphone will run again, and people's position would be updated and recorded by the software with the activities getting more. Thus, Box and Jenkins have extended the above idea of ordinary differencing by forming seasonal differences $w_t = z_t - z_{t-1} = (1 - B^s)z_t$, where s is the seasonal period of the data. Therefore, the most general Box-Jenkins model, the seasonal autoregressive integrated moving average (SARIMA), has the following form:

$$\phi(B)\Phi(B^s)(1-B^s)^D(1-B)^d z_t = \theta(B)\Theta(B^s)a_t \quad (4)$$

with

$$\begin{aligned} \phi(B) &= 1 - \phi_1 B - \dots - \phi_p B^p \\ \theta(B) &= 1 - \theta_1 B - \dots - \theta_q B^q \\ \Phi(B^s) &= 1 - \Phi_1 B^s - \dots - \Phi_p B^{sP} \\ \Theta(B^s) &= 1 - \Theta_1 B^s - \dots - \Theta_Q B^{sQ} \end{aligned} \quad (5)$$

where p is the autoregressive order, q the moving average order, d is the number of differencing operations, and P, D and Q are the corresponding seasonal orders.

After removal of trend and seasonal components, the process of model fitting involves identification, parameter estimation and diagnostic validation. At the identification stage, a tentative autoregressive moving average (ARMA) process is developed based on the estimated autocorrelation function (ACF) and the estimated partial autocorrelation function (PACF). The shape of the ACF and PACF of the population density time series is compared with the shape of the theoretical model. This process of comparison allows the definition of p and q , the orders of the ARMA model.

It worth mentioning that in our study, for each validation (total we have 10), because we need to model 10,000 times for 10,000 cells' density time series by SARIMA, there would be a huge computation work that consumes plenty of time and power of computers. In order to simplify the process, for every distribution that consists of 10,000 cells, we compute the average Tencent density. 888 average density values are regarded as the final time series to calculate the parameters of the SARIMA, and then we still model 10,000 times of the 10,000 times series for every cell but use the same parameters that are computed by the final average time series.

2) CONVLSTM MODEL

Long short-term memory (LSTM) is a type of recurrent neural network (RNN) node structure known to have good performance when handling time-series data with temporal autocorrelations [44]. It was used to successfully learn and generalize the properties of time sequences such as traffic flow [45] and financial stock options return [46]. The core concept of LSTM is the cell state affected by various inter-linked gates. The cell state acts as a transport highway that transfers relative information down the sequence chain as the "memory" of the network. The cell state can carry relevant

information throughout the processing of the sequence. Thus, even information from earlier time steps can make its way to later time steps, reducing the effects of short-term memory. As the cell state evolves, information gets added or removed via gates, acting as a type of neural network that decides which information is allowed to exist for the cell state by learning what information is relevant (during training) [47]. In an LSTM network, at each time step t , the hidden state h_t is updated by the current data, i.e., at the same time step X_t , the hidden states at the previous time step h_{t-1} , the input gate i_t , the forget gate f_t , the output gate o_t , and a memory cell C_t are updated as well [48]. The fundamental principle of the model is similar to that of ConvLSTM; thus, its equations will not be repeated here, as they are given in the introduction of the ConvLSTM model below.

The ConvLSTM model is a variation of LSTM to handle spatiotemporal prediction problems, which were first introduced by Xingjian *et al.* [32] for precipitation nowcasting, where nowcasting is a technique for very short-range forecasting of the current state using an estimate of speed and direction of movement. In this article, we follow the formulation of ConvLSTM as in Reference [32], which includes inputs X_1, \dots, X_t , cell outputs C_1, \dots, C_t , hidden states h_1, \dots, h_t , and gates i_t, f_t, o_t , and uses a three-dimensional (3D) tensor structure. The first two dimensions of the three-dimensional spatiotemporal tensor of each input feature of a ConvLSTM network are the spatial dimensions and the third dimension is time. The input-to-state and state-to-state transitions of the ConvLSTM cell involve convolutional operations that three-dimensional output tensors, as with the original LSTM model. This model can be further formulated using the following equations, where '*' denotes the convolution operation and 'o' denotes the Hadamard product.

$$\begin{aligned} i_t &= \sigma(W_{xi} * X_t + W_{hi} * h_{t-1} + W_{ci} \circ h_{t-1} + b_i) \\ f_t &= \sigma(W_{xf} * X_t + W_{hf} * h_{t-1} + W_{cf} \circ h_{t-1} + b_f) \\ o_t &= \sigma(W_{xo} * X_t + W_{ho} * h_{t-1} + W_{co} \circ h_{t-1} + b_o) \\ C_t &= f_t \circ C_{t-1} + i_t \circ \tanh(W_{xc} * X_t + W_{hc} * h_{t-1} + b_c) \\ h_t &= o_t \circ \tanh(C_t) \end{aligned} \quad (6)$$

In the above equations, i_t, f_t and o_t are the outputs of the input gate, forget gate, and output gate for time step t . C_t is the cell output at time step t . h_t is the hidden state of a cell at time step t . Sigmoid (σ) is used as the gating function for the three gates since it outputs a value between 0 and 1. It can either let no flow or a complete flow of information through the gates. On the other hand, to overcome the vanishing gradient problem, which is a difficulty found in training artificial neural networks with gradient-based learning methods and backpropagation, a function is needed (\tanh) whose second derivative can be sustained for a longer range before going to zero. W and b are weight matrices and bias vector parameters which need to be learned during training. Then we input the Tencent density distributions grids into the ConvLSTM model to predict for 10 times as the validation strategies that would be introduced in the next section.

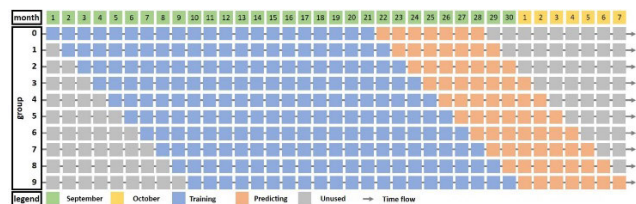


FIGURE 3. The validation strategy for time series of Tencent density distribution.

3) VALIDATION AND COMPARISON

Due to the dataset in this study is time series, we present a cross-validation strategy, as shown in figure 3. We validate the prediction models for 10 times, which utilizes 10 different data groups that are extracted from the original dataset. In every time there are three weeks' data (21 days) that is used to train the models, and the next one week's (7 days) density distributions are prepared to be tested, which means we use $21 \times 24 = 504$ hours' distributions to train models and the next $7 \times 24 = 168$ hours' density distributions to be tested.

After training and predicting the two models for 10 times, we assess the results by calculated the Root Mean Square Error (RMSE) and the R^2 (coefficient of determination) regression score function [49] of the tested data and the predicted data. The RMSE indicates the absolute fit of the model to the data how close the observed data points are to the model's predicted values, while the R^2 reflects the relative fit result. We compute the results of these two indicators as the average results for each cell at the one-time interval (one hour). Nevertheless, for the purpose of assessment in different scales to compare the two models, we graph and assess the two results in time and space respectively by only using RMSE reflecting the absolute value which makes easy to distinguish.

In the temporal scale, for every fishnet (grid) that consists to 10,000 cells, there are three features on every cell including the original density value (F1), predicted density value by SARIMA (F2) and predicted density value by ConvLSTM (F3). We calculate the RMSE between the 10,000 F1 and 10,000 F2, the result is defined as the R1. And then calculate the RMSE between the 10,000 F1 and 10,000 F3, the result is defined as the R2. Because for every validation group there are one week's hours to be tested, there are 168 R1 and 168 R2 that can be arranged by time flow. While we model 10 times for validation, there are 10 groups' results. And in every group, there are 168 R1 and 168 R2. R1 and R2 that reflected the accuracy of the two predicted results can be assessed and compared between the two prediction models in temporal scale.

In the spatial scale, according to the discussion above, there are three features for each cell in every different fishnet, in terms of the same cell in different the fishnet, we extract all 168 F1, 168 F2 and 168 F3. Then we calculate the RMSE between the 168 F1 and 168 F2, named the result as R3, and

calculate the RMSE between the 168 F1 and 168 F3, called the result as R4. Evenly there are 10,000 R3 and 10,000 R4. Because we validate the models for 10 times, there would be 10 groups' results. R3 and R4 that reflected the accuracy of the two predicted results can be assessed and compared between the two prediction models in spatial scale.

IV. RESULTS AND DISCUSSION

In this section, we introduce the simple linear regression results of the two density levels and then present the prediction results. Finally, the comparison of the used two models is proposed.

A. SIMPLE LINEAR REGRESSION

In the pre-processing, we compute the ATD for every cell of the 100×100 grid. In figure 4(c) we visualize the distribution of ATD by 32 classifications natural break method that has been mentioned in section II.B., while the 2010 census density is classified and visualized by the same method, shown as figure 4(a), and the same output of predicted 2020 census density is shown in figure 4(b). The bigger the level number, the higher the density in the two figures. With the comparison of these two visualized maps, we can understand the high similarity of them intuitively. For example, in North China that covers Beijing, Tianjin, Hebei province, Shanxi province and Inner Mongolia, and the Shandong province, Henan province, the people density is relatively high compared to other regions, wherein both two distributions map, much more red-coloured region that means higher than 25 levels cover them at the same time.

Based on the ATD, we calculated the CCD by equation 3, and then extract the relative densities of each administrative unit in two census data and Tencent data. Then we model 2 simple linear regressions for the two groups of density data.

The regression results are shown in figure 5, where (a) shows the relationship between 2010 census density and ATD (regression 1) while (b) illustrates the relationship between predicted 2020 census density and ATD (regression 2). Pearson's correlation coefficients r of the linear regressions are 0.58 and 0.54, which means relatively positive correlations between the census densities and Tencent density for both two models. And the p-values are 9.8×10^{-29} and 4.6×10^{-25} , which are much less than 0.001 means the results of the relationship are reliable.

The reason why the r value of regression 2 is lower than regression 1 may from following aspects: first, the predicted results have errors; Second, in the pre-processing, because the predicted results are only with province-level, which means the predicted density for the each of 12 provinces only has one value, where one province may have more than 2 cities. For example, Guangdong province consists of 21 cities, in 2010 census data, 21 cities have different exact values of density, but in predicted 2020 census data, they have only one value. In terms of other provinces that not included in these 12 ones, i.e., Hebei province, has 13 different density values in predicted 2020 census data for each inside city. And this

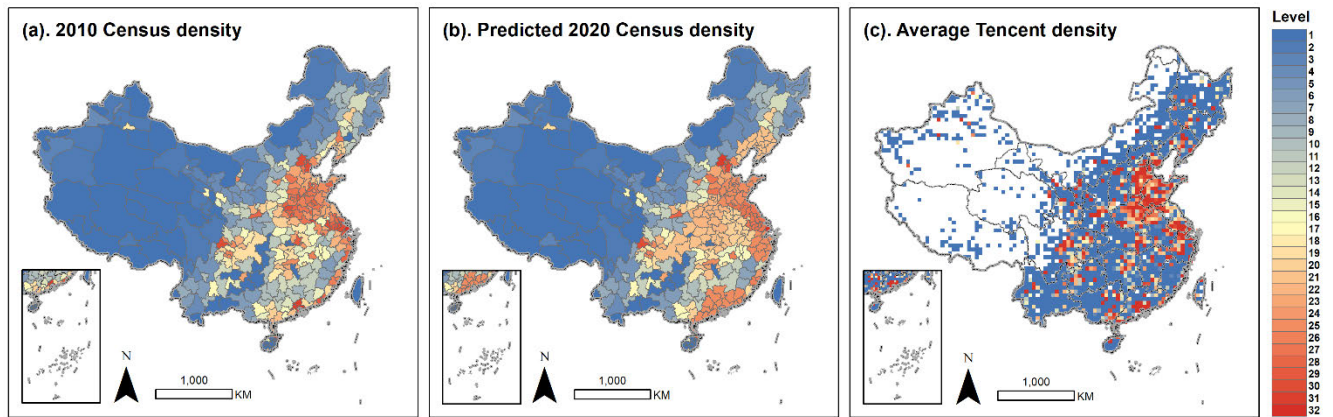


FIGURE 4. The spatial density-level distributions of Census and Tencent. (a) shows the census density level in the administrative unit in 2010, and (b) shows the predicted census density level in 2020, while (c) illustrates the average Tencent density level in the grid unit. The bigger the level number, the higher the density.

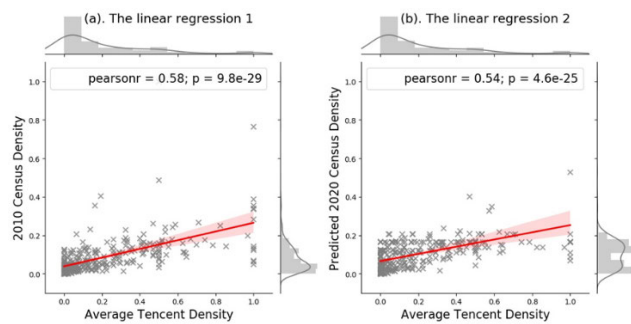


FIGURE 5. The result of the two simple linear regressions. (a) represents the result between the 2010 Census and Tencent densities, named regression 1; (b) illustrates the result between predicted 2020 Census and Tencent densities, defined as regression 2. The “pearsonr” means Pearson’s correlation coefficient r .

is why in figure 4 (b), all areas in Guangdong province have the same colour, while in Heilongjiang province’ area, the colours differ. These two reasons may lead to the lower r value of regression 2 compared to regression 1. But the 0.04 difference for the two values is not so large.

Thus, the similarity of the two r values which are both higher than 0.5 has tested the usability of the Tencent positioning big data, which can reflect the relative human population density as a high-quality VGI. Although the absolute value of the density in Tencent dataset is different from the actual situation, in terms of the reflecting the changing and hot spot or cold spot of the human density distribution, Tencent datasets can be utilized with trustworthy.

B. PREDICTION

1) SARIMA

As the method described in section II.C.1), we model 10 SARIMA for the 10 validation in this section.

First, we need to focus on the trend and seasonality of time series, which could be determined by time series decomposition [50]. The results of the decompositions for 10 groups of time series are shown in figure 6.

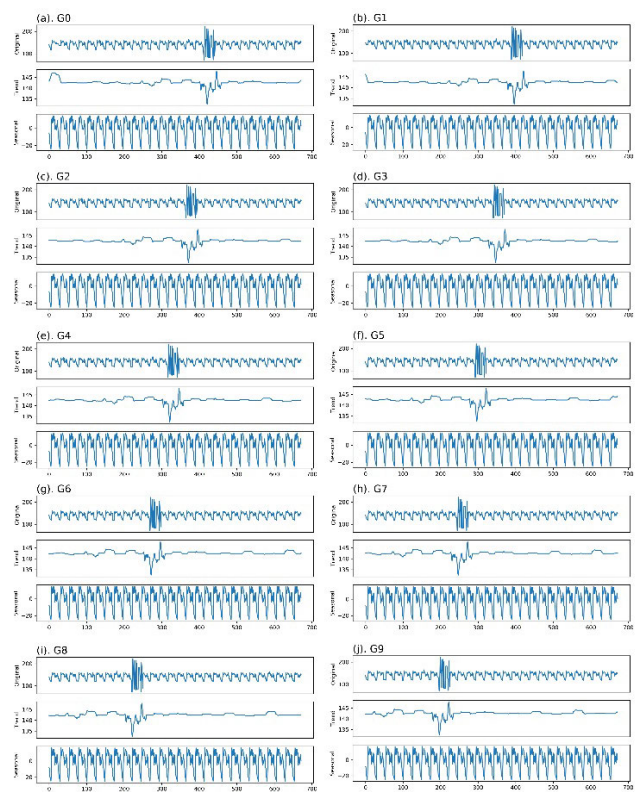


FIGURE 6. The result of the ADF for time series in each group. (a) to (j) shows the 10 groups’ results respectively. Every subplot has 3 components, where the top one is the original time series, the middle one is the trend component of the time series, while the bottom one is the seasonal component. The values on the x-axis are the date index was changing the specific dates into the number sequence.

In the figure, the seasonality is very obvious for all 10 groups, the frequency is 24 (each day has 24 hours while our data is with hourly temporal resolution), which means we need to take seasonal factor into consideration when modelling. However, the trendiness in the figure is not obvious. Thus, we need to use ADF to test if the time series have

TABLE 1. The ADF Test Results in 10 validation groups.

Model	G0	G1	G2	G3	G4	G5	G6	G7	G8	G9
Test Statistic	-14.57	-16.20	-16.49	-16.45	-16.46	-16.10	-16.20	-16.21	-16.21	-16.20
p-value	4.6*10 ⁻²⁷	4.0*10 ⁻²⁹	2.2*10 ⁻²⁹	2.4*10 ⁻²⁹	2.3*10 ⁻²⁹	5.1*10 ⁻²⁹	4.0*10 ⁻²⁹	4.0*10 ⁻²⁹	4.0*10 ⁻²⁹	4.0*10 ⁻²⁷
Critical Value (1%)	-3.44	-3.44	-3.44	-3.44	-3.44	-3.44	-3.44	-3.44	-3.44	-3.44
Critical Value (5%)	-2.87	-2.87	-2.87	-2.87	-2.87	-2.87	-2.87	-2.87	-2.87	-2.87
Critical Value (10%)	-2.57	-2.57	-2.57	-2.57	-2.57	-2.57	-2.57	-2.57	-2.57	-2.57

trendiness for the spatial average Tencent distribution. The results are shown in Table 1.

From the table, all test statistics are lower than 1% critical value in the 10 validation processes, while the p-values are much lower than 0.01, which means in all groups when the H0 hypothesis of ADF test is the existence of unit root, the results reject the original hypothesis. Thus, the time series in the 10 groups have no trendiness. The reason why there are no trends in them may be that our study focuses on a fine-grained population density changing in a relatively short period (37 days). At the same time, the normal trendiness would exist in the year or decade data period, which means we do not need to model our data with a difference.

Thus, because there are no trends for all 10 groups, the parameter *d* in all validations is set to be 0. And in the modelling processing, we need to eliminate seasonal effects. Furthermore, the parameters determined by ACF process are shown in Table 2. Then we build the SARIMA for every 10,000 cells' time series and repeat it for 10 times.

TABLE 2. The parameters of SARIMA models for 10 validation groups.

Para	G0	G1	G2	G3	G4	G5	G6	G7	G8	G9
p	24	12	18	18	18	6	18	6	18	18
d	0	0	0	0	0	0	0	0	0	0
q	5	3	5	5	5	4	5	4	5	5
P	1	1	1	1	1	1	1	1	1	1
D	1	0	0	0	0	0	0	0	0	0
Q	1	1	1	1	1	1	1	1	1	1
s	24	24	24	24	24	24	24	24	24	24

“Para” means parameters.

2) CONVLSTM

For each round of the prediction, the parameters of the model included the kernel size, which we proposed to be 3 × 3, with 40 convolutional filters that can extract important features from the convolution layers, with five units for each. In order to improve the generalization ability and to prevent overfitting (which is the production of an analysis that corresponds too closely or exactly to a particular set of data, and may, therefore, fail to fit additional data or predict future observations reliably in machine learning or deep learning models), the recurrent weight dropout was set to 0.2 in the model; the number of training times (epochs) was set as 500, whilst the Adam optimizer [51] was used with a learning rate of 0.001 and a decay rate of 0.9.

The overall accuracy of the prediction results can be reflected by the RMSE and R² of the tested and predicted values. The total average RMSEs of the two prediction results for the 10,000 cells in the 10 validations have been reported in Table 3. It is obvious that the accuracy of the prediction results of ConvLSTM model is much higher than the SARIMA one's, where the total average of the ConvLSTM model's RMSE is 139.0, while the SARIMA predicted RMSE is 1587.5. Here we notice that the RMSE of the results by SARIMA model in the first validation group G0 (12210.0) is much higher than the other 9 results, it is because after computing the parameters of the model in this round, p was set to 24 while in other 9 groups the most value of p is 18, which lead to the total difference of the predicting accurate between the others based on every 10,000 cells' density time series used the parameters but not all series in some cells are suitable for the parameters. However, even we regard the value as an outlier and exclude it, the average RMSE of the other 9 values is 407.3, which is still about 3 times higher than the ConvLSTM one.

The R² results are reported in Table 4, where the structure is similar to RMSE's. All R² values of the ConvLSTM are beyond 0.95, indicating a very high fit of the predicted to the test values. Except for the results in G0 that it regards as an outlier of the 10 validations, the overall R²s of the SARIMA results are between 0.7 and 0.8, which is also high but comparing with the ConvLSTM model, the accuracy is lower.

The S means the SARIMA model, and the C means the ConvLSTM model.

However, because the structure of the overall process of this study is multi-layered both in spatial and temporal scales, in order to compare the two prediction results clearly, we compute the RMSE results by the method in section II.C.3). Then we assess and discuss the results separately in time and space as follows.

C. COMPARISON

1) COMPARISON IN TIME

We calculate the RMSE between the 10,000 F1 and 10,000 F2, then get the R1. And then calculate the RMSE between the 10,000 F1 and 10,000 F3 to get the R2. The 168 R1 and 168 R2 over every 7 days' hours in 10 validations are shown in figure 7.

In figure 7, we can easily learn that the RMSE of the predicted results by ConvLSTM is overall higher than the ones

TABLE 3. The average RMSE of the prediction results for all cells in 10 validation groups.

Model	G0	G1	G2	G3	G4	G5	G6	G7	G8	G9	Mean
S	12210.0	415.6	409.7	403.6	400.9	412.7	406.8	409.0	402.4	404.4	1587.5
C	142.4	143.7	125.0	131.4	130.7	139.5	144.7	140.8	142.1	148.9	139.0

The S means the SARIMA model, and the C means the ConvLSTM model.

TABLE 4. The R² score of the prediction results for all cells in 10 validation groups.

Model	G0	G1	G2	G3	G4	G5	G6	G7	G8	G9
S	< 0.5	0.78	0.78	0.79	0.79	0.78	0.78	0.78	0.79	0.79
C	0.96	0.95	0.97	0.96	0.96	0.95	0.95	0.96	0.96	0.95

The S means the SARIMA model, and the C means the ConvLSTM model.

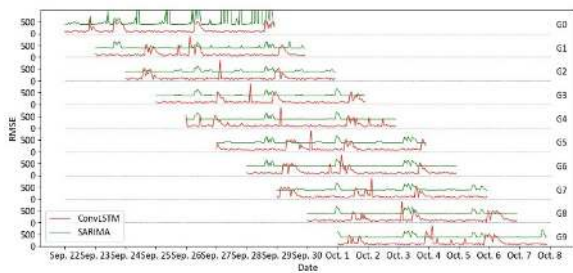


FIGURE 7. The RMSE results of SARIMA and ConvLSTM models in time. All groups' data share the same x-axis, while in the y-axis, there are 10 bins that every bin span the RMSE from 0 to 1000, where 1000 could also be the start of the next bin.

by SARIMA over the 108 hours (7 days) in all 10 validations, which means the accuracy of the prediction of ConvLSTM is overall higher than the SARIMA in temporal scale. In terms of the details, more sudden fluctuation happens with the time going on in G0 resulting from the cumulative errors, which also can explain why the RMSE error is much higher than others' in G0 shown in Table 3. Moreover, both the results of SARIMA and ConvLSTM in other 9 groups have similar trendiness. In most of the hours, the RMSEs of ConvLSTM model are lower than SARIMA one and keeping around 140, but in some periods the opposite is exact. For example, during the 22:00 October 3 to 1:00 October 4, the RMSEs of ConvLSTM are around 700 while the SARIMA's are stable at about 500. However, the RMSE of ConvLSTM is lower than SARIMA in more than 90% of hours, which indicates the smaller error and more accurate of the capability of the prediction of ConvLSTM model.

2) COMPARISON IN SPACE

In the spatial scale, we compute the 168 F1, 168 F2 and 168 F3. Then we calculate the RMSE between the 168 F1 and 168 F2 to get R3, and calculate the RMSE between the 168 F1 and 168 F3 to reach R4. Evenly there are 10,000 R3 and 10,000 R4 in 10 validations. We output them as the map of China where the different colours in

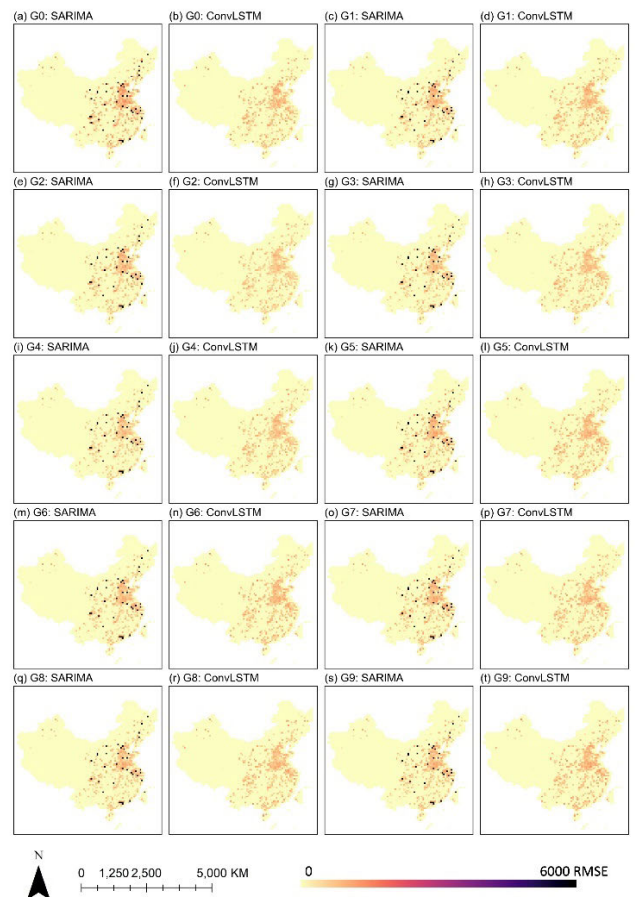


FIGURE 8. The RMSE results of SARIMA and ConvLSTM models in space. (a) to (t) alternately show the RMSE maps of SARIMA and ConvLSTM in China.

the cell represent the RMSE of the prediction. The results are illustrated in figure 8. The whiter regions in the graphs correspond to a higher error (RMSE) of prediction for user density. In contrast, the blacker zones mean the error is lower. Thus, the distribution of the error scan shows the effect of the prediction models in space.

Intuitively, in all prediction RMSE maps of SARIMA, there are plenty of the whitest cells that represent the highest error of the prediction, which are surrounded by the cells with much blacker colour. The reason why they are so different from the others is that most of them represent the big cities such as Beijing (North China) and Guangzhou (South China), which have higher actual human densities. After using SARIMA model to predict them, although the accuracy of the results is similar with the prediction of other regions with lower density, the RMSE that is the accumulative errors of all hours' predicted values would be much higher while they have the sizeable cardinal number. Thus, the specific cells in the error distribution look so independent within the whole map. However, while we use ConvLSTM to predict the Tencent density distribution, because the convolution is used when we train the models, which can consider the surrounded cells' values, there is not the situation like the SARIMA. It is precisely because the ConvLSTM model considers the spatial autocorrelation that it is more suitable for prediction of Tencent density distribution, which proves that ConvLSTM model is much more accurate than the SARIMA model in this study.

V. CONCLUSION

In this study, we first used two linear regression methods to model the relationships between the 2010 census density, predicted 2020 census density and the Tencent density respectively to test the usability of Tencent positioning data. The Pearson's correlation coefficients r 0.58 and 0.54, while the much less than 0.001 p -values demonstrate good positive correlations between the ground truth data and a GNSS-oriented open resource, which means the Tencent positioning data can be used in the study that is related to the human density distribution. Then we used a ConvLSTM model to predict the relative human density distributions with a traditional time-series predicted model, SARIMA for comparison. The total average of the ConvLSTM model's RMSE is 139.0, while the SARIMA predicted RMSE is three times greater than 139.0. And the R^2 values of ConvLSTM model are higher than 0.9, while the SARIMA ones are about 0.78. On the one hand, in the temporal scale, the SARIMA is as stable as ConvLSTM model but with overall lower accurate than ConvLSTM. On the other hand, ConvLSTM model's prediction results are more durable than the SARIMA one in spatial scale while the latter one has much more outliers with much higher RMSEs in big cities' areas.

To summary, the two models are both suitable for the prediction of Tencent density distribution in finely-grained time and space, nevertheless, the predicted density correlated much better with the tested data at both the temporal and spatial scales used when using ConvLSTM as compared to the SARIMA, while the capability of prediction by ConvLSTM model is more stable than the latter one in space.

In future works, we would focus on the following three studies. First, considering other factors that affect the human density into the prediction models; second, testing the

capability of the models with different resolutions of the time and space; finally, combining the first and second ones at the same time.

REFERENCES

- [1] W. Lutz, W. Sanderson, and S. Scherbov, "The end of world population growth," *Nature*, vol. 412, no. 6846, p. 543, 2001.
- [2] P. Vanella and P. Deschermeier, "A probabilistic cohort-component model for population forecasting—The case of Germany," *J. Population Ageing*, vol. 13, pp. 1–33, Jan. 2020.
- [3] P. Jia, Y. Qiu, and A. E. Gaughan, "A fine-scale spatial population distribution on the high-resolution gridded population surface and application in Alachua County, Florida," *Appl. Geography*, vol. 50, pp. 99–107, Jun. 2014.
- [4] C. Deng, C. Wu, and L. Wang, "Improving the housing-unit method for small-area population estimation using remote-sensing and GIS information," *Int. J. Remote Sens.*, vol. 31, no. 21, pp. 5673–5688, Nov. 2010.
- [5] W. Zhong, D. Wang, D. C. Xie, and L. Yan, "Dynamic characteristics of Shanghai's population distribution using cell phone signaling data," *Geographical Res.*, vol. 36, no. 5, pp. 972–984, 2017.
- [6] G. Zhang, X. Rui, S. Poslad, X. Song, Y. Fan, and B. Wu, "A method for the estimation of finely-grained temporal spatial human population density distributions based on cell phone call detail records," *Remote Sens.*, vol. 12, no. 16, p. 2572, Aug. 2020.
- [7] X. Qiu, "Development in Chinese transportation construction," in *China 40 Years Infrastructure Construction*. Cham, Switzerland: Springer, 2020, pp. 79–98.
- [8] H. Li, J. Strauss, and L. Lu, "The impact of high-speed rail on civil aviation in China," *Transp. Policy*, vol. 74, pp. 187–200, Feb. 2019.
- [9] W. Song, L. Wang, Y. Xiang, and A. Y. Zomaya, "Geographic spatiotemporal big data correlation analysis via the Hilbert–Huang transformation," *J. Comput. Syst. Sci.*, vol. 89, pp. 130–141, Nov. 2017.
- [10] L. Curry, "Univariate spatial forecasting," *Econ. Geography*, vol. 46, pp. 241–258, Jun. 1970.
- [11] X. Li, W. Chen, X. Cheng, and L. Wang, "A comparison of machine learning algorithms for mapping of complex surface-mined and agricultural landscapes using ZiYuan-3 stereo satellite imagery," *Remote Sens.*, vol. 8, no. 6, p. 514, Jun. 2016.
- [12] G. Zhang, X. Rui, S. Poslad, X. Song, Y. Fan, and Z. Ma, "Large-scale, fine-grained, spatial, and temporal analysis, and prediction of mobile phone Users' distributions based upon a convolution long short-term model," *Sensors*, vol. 19, no. 9, p. 2156, May 2019.
- [13] J. Li, J. Li, Y. Yuan, and G. Li, "Spatiotemporal distribution characteristics and mechanism analysis of urban population density: A case of Xi'an, Shaanxi, China," *Cities*, vol. 86, pp. 62–70, Mar. 2019.
- [14] P. Deville, C. Linard, S. Martin, M. Gilbert, F. R. Stevens, A. E. Gaughan, V. D. Blondel, and A. J. Tatem, "Dynamic population mapping using mobile phone data," *Proc. Nat. Acad. Sci. USA*, vol. 111, no. 45, pp. 15888–15893, Nov. 2014.
- [15] M. F. Goodchild, "Citizens as sensors: The world of volunteered geography," *GeoJournal*, vol. 69, no. 4, pp. 211–221, Nov. 2007.
- [16] J. Severinsen, M. de Roiste, F. Reitsma, and E. Hartato, "VGTrust: Measuring trust for volunteered geographic information," *Int. J. Geographical Inf. Sci.*, vol. 33, no. 8, pp. 1683–1701, Aug. 2019.
- [17] T. R. Malthus, *An Essay on the Principle of Population*. London, U.K.: Dent, 1872.
- [18] N. Bacaër, "Verhulst and the logistic equation (1838)," in *A Short History of Mathematical Population Dynamics*. Cham, Switzerland: Springer, 2011, pp. 35–39.
- [19] E. Seneta, "Markov and the birth of chain dependence theory," *Int. Stat. Rev./Revue Internationale de Statistique*, vol. 64, no. 3, pp. 255–263, Dec. 1996.
- [20] N. Keyfitz and H. Caswell, *Applied Mathematical Demography*. Cham, Switzerland: Springer, 2005.
- [21] P. H. Leslie, "Some further notes on the use of matrices in population mathematics," *Biometrika*, vol. 35, nos. 3–4, pp. 213–245, 1948.
- [22] G. Chi and P. R. Voss, "Small-area population forecasting: Borrowing strength across space and time," *Population, Space Place*, vol. 17, no. 5, pp. 505–520, Sep. 2011.
- [23] N. Keyfitz, "On future population," *J. Amer. Stat. Assoc.*, vol. 67, no. 338, pp. 347–363, 1972.

- [24] H. Minoura, R. Yonetani, M. Nishimura, and Y. Ushiku, "Crowd density forecasting by modeling patch-based dynamics," 2019, *arXiv:1911.09814*. [Online]. Available: <http://arxiv.org/abs/1911.09814>
- [25] S. S. Sohn, S. Moon, H. Zhou, S. Yoon, V. Pavlovic, and M. Kapadia, "Deep crowd-flow prediction in built environments," 2019, *arXiv:1910.05810*. [Online]. Available: <http://arxiv.org/abs/1910.05810>
- [26] Z. Lin, J. Feng, Z. Lu, Y. Li, and D. Jin, "Deepstn+: Context-aware spatial-temporal neural network for crowd flow prediction in metropolis," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, 2019, pp. 1020–1027.
- [27] X. Wang, Z. Zhou, Y. Zhao, X. Zhang, K. Xing, F. Xiao, Z. Yang, and Y. Liu, "Improving urban crowd flow prediction on flexible region partition," *IEEE Trans. Mobile Comput.*, early access, Aug. 20, 2019, doi: [10.1109/TMC.2019.2934461](https://doi.org/10.1109/TMC.2019.2934461).
- [28] P. Whittle, *Hypothesis Testing in Time Series Analysis*. Stockholm, Sweden: Almqvist & Wiksells, 1951.
- [29] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control*. Hoboken, NJ, USA: Wiley, 2015.
- [30] X. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *Bull. Math. Biophys.*, vol. 5, no. 4, pp. 115–133, Dec. 1943.
- [31] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [32] S. Xingjian, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 802–810.
- [33] S. Hecker, R. Bonney, M. Haklay, F. Hölker, H. Hofer, C. Goebel, M. Gold, Z. Makuch, M. Ponti, A. Richter, L. Robinson, J. R. Iglesias, R. Owen, T. Peltola, A. Sforzi, J. Shirk, J. Vogel, K. Vohland, T. Witt, and A. Bonn, "Innovation in citizen science—perspectives on science-policy advances," *Citizen Sci., Theory Pract.*, vol. 3, no. 1, pp. 1–14, 2018.
- [34] R. Can, S. Kocaman, and C. Gokceoglu, "A convolutional neural network architecture for auto-detection of landslide photographs to assess citizen science and volunteered geographic information data quality," *ISPRS Int. J. Geo-Inf.*, vol. 8, no. 7, p. 300, Jul. 2019.
- [35] H. Vahidi, B. Klinkenberg, and W. Yan, "Trust as a proxy indicator for intrinsic quality of volunteered geographic information in biodiversity monitoring programs," *GIScience Remote Sens.*, vol. 55, no. 4, pp. 502–538, Jul. 2018.
- [36] J. Golbeck, "Trust on the world wide Web: A survey," *Found. Trends Web Sci.*, vol. 1, no. 2, pp. 131–197, 2006.
- [37] H. Senaratne, A. Mobasher, A. L. Ali, C. Capineri, and M. Haklay, "A review of volunteered geographic information quality assessment methods," *Int. J. Geographical Inf. Sci.*, vol. 31, no. 1, pp. 139–167, Jan. 2017.
- [38] K. H. Zhang and S. Song, "Rural–urban migration and urbanization in China: Evidence from time-series and cross-section analyses," *China Econ. Rev.*, vol. 14, no. 4, pp. 386–400, Jan. 2003.
- [39] L. S. Deng Yu, C. Jianming, and L. Xi, "Spatial pattern and its evolution of Chinese provincial population and empirical study," (in Chinese), *Acta Geographica Sinica*, vol. 69, no. 10, pp. 1473–1486, 2014.
- [40] M. J. De Smith, M. F. Goodchild, and P. Longley, *Geospatial Analysis: A Comprehensive Guide to Principles, Techniques and Software Tools*. Kibworth, U.K.: Troubador Publishing Ltd, 2007.
- [41] K. H. Zou, K. Tuncali, and S. G. Silverman, "Correlation and simple linear regression," *RSNA Radiol.*, vol. 227, no. 3, pp. 617–628, 2003.
- [42] H. J. Seltman. (2012). *Experimental Design and Analysis*. [Online]. Available: <http://www.stat.cmu.edu/hjseltman/309/Book/Book.pdf>
- [43] D. A. Dickey and W. A. Fuller, "Distribution of the estimators for autoregressive time series with a unit root," *J. Amer. Stat. Assoc.*, vol. 74, no. 366a, pp. 427–431, Jun. 1979.
- [44] Z. Yuan, X. Zhou, and T. Yang, "Hetero-ConvLSTM: A deep learning approach to traffic accident prediction on heterogeneous spatio-temporal data," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jul. 2018, pp. 984–992.
- [45] R. Fu, Z. Zhang, and L. Li, "Using LSTM and GRU neural network methods for traffic flow prediction," in *Proc. 31st Youth Academic Annu. Conf. Chin. Assoc. Autom. (YAC)*, Nov. 2016, pp. 324–328.
- [46] K. Chen, Y. Zhou, and F. Dai, "A LSTM-based method for stock returns prediction: A case study of China stock market," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Oct. 2015, pp. 2823–2824.
- [47] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with LSTM," *Neural Comput.*, vol. 12, no. 10, pp. 2451–2471, 2000.
- [48] R. Zhao, R. Yan, J. Wang, and K. Mao, "Learning to monitor machine health with convolutional bi-directional LSTM networks," *Sensors*, vol. 17, no. 2, p. 273, Jan. 2017.
- [49] J. L. Devore, *Probability and Statistics for Engineering and the Sciences*. Boston, MA, USA: Cengage Learning, 2011.
- [50] X. Wang, K. Smith, and R. Hyndman, "Characteristic-based clustering for time series data," *Data Mining Knowl. Discovery*, vol. 13, no. 3, pp. 335–364, Sep. 2006.
- [51] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: <http://arxiv.org/abs/1412.6980>



ZHI ZHENG received the B.Sc. degree in urban planning from Peking University, China, in 2016. He is currently pursuing the Ph.D. degree with the Chinese Academy of Sciences. He is also a Visiting Student with the National University of Singapore. His research interests include regional development, global production networks, GIS, and remote sensing.



GUANGYUAN ZHANG received the B.Sc. degree in GIS from the Hefei University of Technology, China, in 2016, and the M.Sc. degree in GIS from the University of Chinese Academy of Sciences (UCAS), China, in 2019. He is currently pursuing the Ph.D. degree in computer science with the Queen Mary University of London, London, U.K. His research interests include spatial-temporal big data mining, GIS and remote sensing, urban computing, and the Internet of Things (IoT).

...