Общероссийская общественная организация
«Региональная ассоциация науки»
АНО ДПО «Центр развития образования и науки»

II ВСЕРОССИЙСКАЯ НАУЧНО-ПРАКТИЧЕСКАЯ КОНФЕРЕНЦИЯ
(ШИФР МКРНП)

«РАЗВИТИЕ НАУКИ И ПРАКТИКИ В ВЕК ЦИФРОВОЙ ПЕРЕЗАГРУЗКИ: ТЕКУЩИЕ ТЕНДЕНЦИИ И БУДУЩИЕ ПЕРСПЕКТИВЫ»

29 июля 2024
Москва

Forecasting Population Growth in Hanoi Using Google Earth Engine and Cloud Computing

Dao Dinh Thuan, Doctor of Philosophy
Nguyen Phuong Dong, Doctor of Philosophy
Nguyen Thi Hoa, Doctor of Philosophy
Nguyen Thi Thu Huyen, Master of Science
Hanoi University of Mining and Geology
Vietnam, Hanoi

Abstract

The study evaluates the effectiveness of using linear regression combined with population statistics and remote sensing data for population forecasting in Hanoi. The results show that this method provides highly accurate visualization of population distribution and effectively forecasts population growth over time, aligning with traditional methods. Using remote sensing image data in linear regression helps create long-term population distribution forecast maps, overcoming the limitations of conventional methods. Additionally, this method allows for quick updates and edits to input data, which is beneficial for addressing continuous population fluctuations and providing timely information for policymakers.

Keywords: Google Earth Engine, population, growth, trends

1. Introduction

Hanoi, the capital of Vietnam, has experienced rapid urban growth in recent decades. Hanoi is divided into 12 urban districts, 1 district-leveled town and 17 rural districts. According to the 2009 census, Hanoi's population was 6,451,909 people; by 2020, the population had reached 8,246,500 people [1, 2]. This population growth, driven by urbanization and the migration of people from rural to urban areas, poses significant challenges to urban management, infrastructure planning, public services, and the quality of life of urban residents. Meanwhile, socio-economic development and population changes have a reciprocal relationship that impacts the sustainable development of Hanoi.

Population prediction is crucial for urban planning and resource management. Accurate population forecasts enable policymakers and city managers to develop and implement sustainable development strategies. In Vietnam, specifically in Hanoi, population statistics are collected annually and are vital for various societal functions [3]. However, population statistics are often presented in tabular form, which cannot show the spatial distribution of the population within a specific administrative unit. Remote sensing technology addresses this issue by enabling data visualization on maps, providing managers with a comprehensive view of each area to develop appropriate management plans and solutions.

In the rapidly evolving technology landscape, cloud computing is considered a technological revolution. It is increasingly growing with the formation of new models such as Anything as a Service (XaaS) and serverless computing, allowing any technology to be provided via the internet and enabling users to access, manage, and develop computer resources remotely without physical management [4]. Google Earth Engine (GEE) is a powerful cloud computing platform for big spatial data analysis.

GEE provides flexible tools and computing and storage resources to process big data, perform environmental change analyses, monitor land use changes, and assess factors impacting population growth [5, 6]. In the context of population prediction, cloud computing can process vast amounts of data from various sources, including satellite images and census data. Common population prediction methods use regression models, temporal models, and machine learning techniques to forecast population growth based on historical and current data [5].

The main goal of this study is to predict population growth in Hanoi in the coming years using Google Earth Engine (GEE) technology and cloud computing. Specifically, this research will develop a regression model based on population data from 2000 to 2020 and apply it to project population figures for 2030 and 2050. The resulting map will show population distribution and density in pixels, with each pixel representing the number of people living within the spatial boundaries of the research area.

2. Research methods and data

Data used in this study include:

- **Population data** : Population data from previous population surveys and official reports in Hanoi city statistical yearbook through 2015 and 2020 [2, 7].

- **Satellite Data**: Population dataset generated by the World Pop project at 100m resolution globally. Global high-resolution, contemporary data on human population distributions are a prerequisite for accurately measuring the impacts of population growth, monitoring changes, and planning interventions [8, 9].

The study uses the linear regression method, which is an essential statistical tool often used to determine the correlation between variables, thereby making predictions of future values based on existing data. In this study, the steps to build a regression method are divided into 3 basic steps:

- Limit the research area, in which the research area for population data is determined by the 'geometry' object focusing specifically on the administrative boundaries of Hanoi city.

- Linear regression function is built to evaluate population changes over the years by collecting yearly population data bands. Applying the 'reduceRegion' functions with ' Reducer.sum( )' allows calculating the total population in a specified area, creating a multi-temporal dataset that helps analyze population trends. The 'linearFit( )' function allows for the calculation of two main parameters for each pixel, including the regression coefficient (scale) and the intercept (offset).

```
// Year band
var popYearBand = javaPop.map(function(image){
  var yearBand = ee.Image(ee.Number(image.get('year'))).rename('year');
  return image.addBands(yearBand).select(['year', 'population']).toUint32();
});

// Linear regression
var regression = popYearBand.reduce(ee.Reducer.linearFit());
var scale = regression.select('scale');
var offset = regression.select('offset');
var chartRegression = ui.Chart.image.series({
  imageCollection: popYearBand,
  region: geometry,
  reducer: ee.Reducer.sum(),
  scale: 100,
  xProperty: 'year'
}).setOptions({
  title: 'Linear Regression of Population Over Years',
  hAxis: {title: 'Year'},
  vAxis: {title: 'Population'},
});
print(chartRegression);
```

- Population projections for future years are made using linear regression coefficients (scale and offset) to calculate population values based on future years.

```
// Predict 2030
var pop2030 = regression.select('scale').multiply(2030).add(regression.select('offset'));
Map.addLayer(pop2030, popVis, 'Java Population 2030 Prediction'); // Show 2030 data

// Predict 2050
var pop2050 = regression.select('scale').multiply(2050).add(regression.select('offset'));
Map.addLayer(pop2050, popVis, 'Java Population 2050 Prediction'); // Show 2050 data

```

3. Results

Assessment of population growth in the period 2000-2020

Hanoi is a large city with a high population density and many complex population fluctuations. WorldPop data can provide an overview of population distribution; however, some challenges may be encountered when applying this data to Hanoi. These include differences in spatial resolution and rapid population changes due to urbanization, leading to discrepancies from the original data. Therefore, this study evaluated the correlation and accuracy of WorldPop data compared to actual data sources.

Table 1. Population change in the period from 2000 to 2020 in Hanoi city

Year	Population according to actual statistics (thousands of people)	Population according to satellite image data (thousands of people)
2010	6,617.9	5,659.7
2011	6,779.3	5,798.8
2012	7,128.3	5,943.3
2013	7,265.6	6,090
2014	7,390.9	6,239.7
2015	7,433.6	6,392.1
2016	7,590.8	6,549.4
2017	7,742.2	6,709.8
2018	7,914.5	6,877.4

2019	8,093.9	7,046.9
2020	8,246	7,219.9

Results from WorldPop data indicate that satellite image processing data always tends to be lower than the actual population data of Hanoi city. The linear regression equation, with an R^2 value of 0.98, shows a high correlation between the interpreted values and the actual statistical values. This suggests that values interpreted from remote sensing data can be used to estimate and explain actual values.

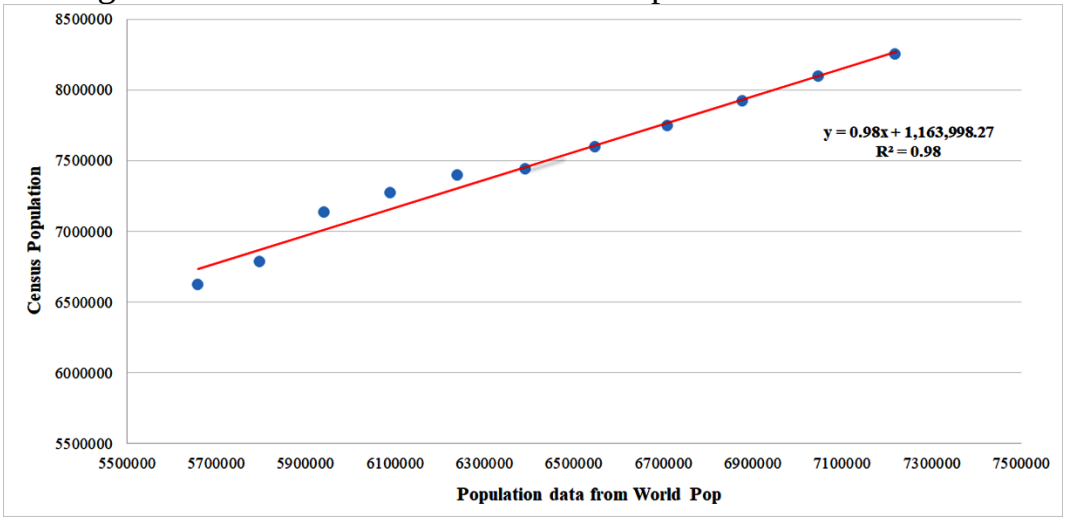

Figure 1. The Correlation between actual Population and WorldPop estimates

Population prediction results

The population prediction results for Hanoi were made using a regression model for 2030 and 2050. The prediction results from the model show that Hanoi's population will increase increasingly according to the current trend, with an estimated population of 9.35 million in 2030 and 12.08 million in 2050. These results are consistent with scenario 1 of Hanoi's Planning project for the period 2021-2030, with a vision to 2050, which forecasts a population of about 9.5 million in 2030 and 11.2 million in 2050.

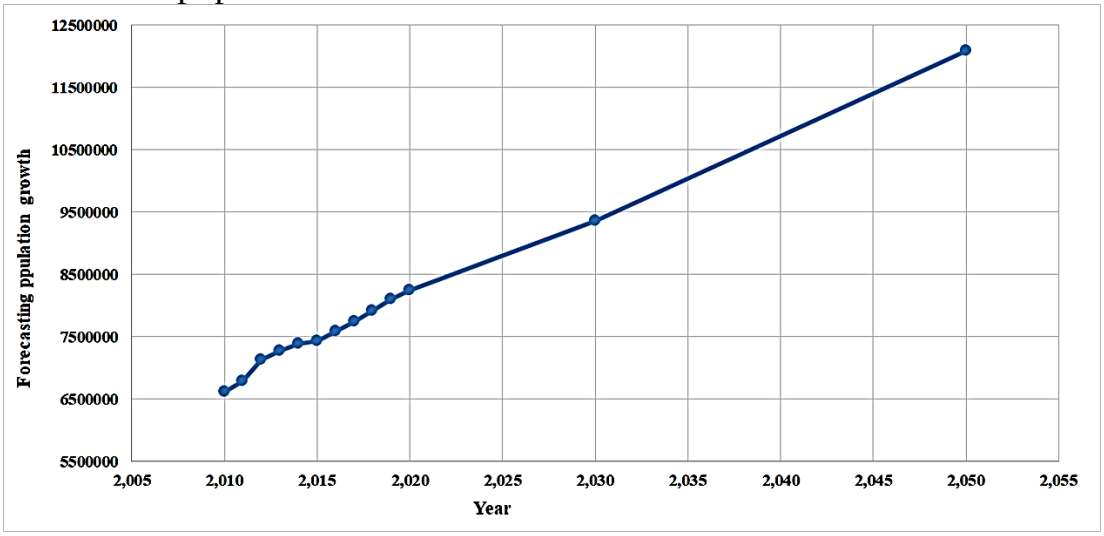

Figure 2. Prediction Population Growth in Hanoi to 2050

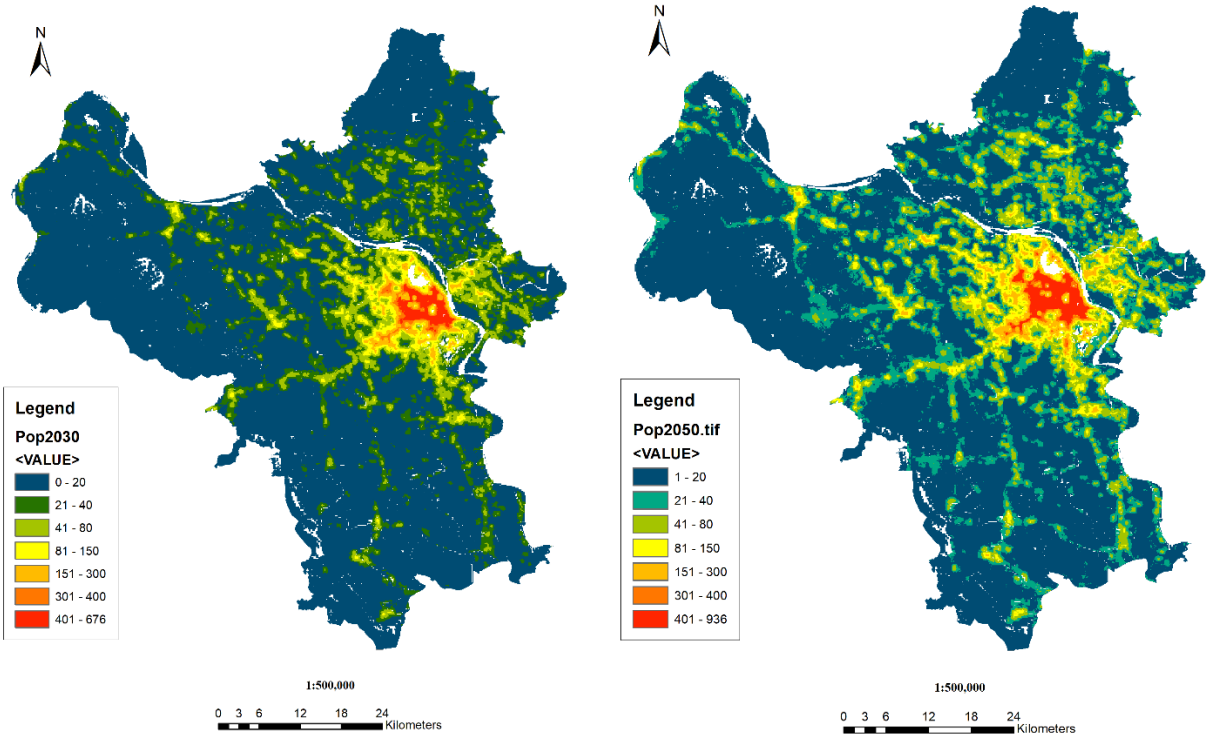

Figure 3. Projected Population Distribution Map of Hanoi City for 2030 and 2050

4. Conclude

The population growth forecasting method, which combines population statistics and WorldPop's remote sensing data, has produced intuitive results that more accurately reflect population distribution and help effectively forecast population growth in future. The forecast population distribution map clearly shows population density and concentration in the area, providing a scientific basis and an essential tool for urban planning and management.

The linear regression method using remote sensing image data helps create long-term population distribution forecast maps, overcoming the limitations of traditional methods. This method also offers the advantage of quickly updating and editing input data, which is helpful in the context of continuous population fluctuations and the need for timely information by policymakers.

The accuracy and intuitiveness of the method depend significantly on the quality of remote sensing data and actual population census data. Therefore, it is necessary to continue researching the scientific and experimental basis of this method to apply accurate forecasts and solve other practical problems related to regional population distribution. Also, continued research and improvement are needed to maximize the effectiveness of this method, particularly in ensuring the quality of input data and applying advanced technologies such as artificial intelligence and machine learning.

References:

1. Hanoi Statistical Office, *2014 Statistical Yearbook*, H.S. Office, Editor. 2015, Statiscal publishing house Hanoi.
2. Statistics Office of Ha Noi City, *Statistical Yearbook of Ha Noi City in 2020*. 2021, Hanoi: Statistical Publishing house 671.

3. Do, T.P.T., V.S. Mai, and V.L. Nguyen, *Intergrate census data and land use for population mapping*. Journal of Geodesy and Cartography, 2018(36): p. 52-58.
4. Miyachi, C., *What is "Cloud"? It is time to update the NIST definition?* IEEE Cloud computing, 2018. **5**(03): p. 6-11.
5. Leong, D., et al., *World Population Prospects 2019. Department of Economic and Social Affairs Population Dynamics. New York (NY): United Nations; 2019 (https://population. un. org/wpp/Download/, accessed 20 September 2020). The decade of healthy ageing. Geneva: World Health Organization.* World, 2018. **73**(7): p. 362k2469.
6. Gorelick, N., et al., *Google Earth Engine: Planetary-scale geospatial analysis for everyone.* Remote sensing of Environment, 2017. **202**: p. 18-27.
7. Hanoi Statistical Office, *2015 Statistical Yearbook*. 2016, Hanoi: Statistical publishing house. 476.
8. Gaughan, A.E., et al., *High resolution population distribution maps for Southeast Asia in 2010 and 2015.* PloS one, 2013. **8**(2): p. e55882.
9. Tatem, A.J., *WorldPop, open data for spatial demography.* Scientific data, 2017. **4**(1): p. 1-4.