Geographic information system-based spatial analysis of population distribution in Banten province - Indonesia

Mohd Robi Amri1*, Sesa Wiguna2, and Ridwan Yunus3

1 Geography Science (Magister), University of Indonesia, Kampus Baru UI, Depok, 16424, Indonesia
2 Geography Science (Magister), University of Auckland, Auckland, 1010, New Zealand
3 Member National Assistance, BNPB, Graha BNPB Building, Jl. Pramuka Kav. 38, Jakarta Timur, 10150, Indonesia

Abstract. Population distribution is one of the of disaster vulnerability parameters needed in a disaster risk assessment. The analysis approaches to determine the spatial population distribution can use many methodological alternatives. The general approach used in Indonesia is based on the results of a survey or census, where the number of population density is distributed evenly within the administrative borders. Another approach using Random Regression Tree model-based Forest Mapping is used by Worldpop. Both methodologies have their respective advantages and disadvantages. This study was conducted by combining these two methods where some data and parameters are added as driving factors on the scale spatial resolution analysis (grid size) 0.000833333 decimal degrees (approximately 100 m in the equatorial region) for a case study in Banten Province. Data processing is performed by raster analysis approach and GIS. The results are more affordable, meet cost requirements, and can be utilized to calculate the level of disaster risk in the area.

1 Introduction

In order to conduct a disaster assessment, spatial distribution data that illustrates the distributions of the human population are needed. In disaster management context, population distribution is significant to estimate the element of the risk to reduce the incident of fatalities [1]. The accurate population information can be achieved by considering the estimated Indonesia population for next 10 years since 2010, which more than 50% of the growth will be absorbed into urban areas. The easily-updatable population data for research and others are requested. This request can be satisfied by the use of remote sensing and other geospatial data sets to refine estimation of population data. But, a new approach is needed to achieve accurate spatial population distributions.

In general, the population is provided in the form of statistical tables and presented spatially in a reference map. The spatial presentation also commonly referred to as mapping choropleth. The disadvantage of this method is that this method can’t describe in detail the occupied area. In fact, the population distribution through choropleth mapping does not represent actual population distribution [1]. For the disaster risk analysis purposes, this method is less feasible to be used, especially to describe the populations exposure.

In order to have population data in more detail, some scholars such as [2] and [3] have employed a dasymetric mapping method. The Dasymetric mapping is one area-based thematic mapping method that produces more detailed spatial information [4]. Dasymetric mapping has an advantage in generating more realistic spatial population distribution compared to choropleth mapping.

Dasymetric mapping has been well developed along with the development of spatial technology. One of the spatial data were freely available to be used is data from WorldPop, especially for the region of Indonesia. The WorldPop data are generated using random regression tree model-based forest mapping along with various assumed variables which effect the population density pattern. The predicted value is then used as a surface weight values to redistribute dasymetric population density per grid up to villages level. The WorldPop spatial resolution data is 0.000833333 decimal degrees (approximately 100 m in the equatorial region).

In particular cases in Indonesia, the WorldPop data has some disadvantages. First example, the IndoPop population (Indonesia Population data produced by WorldPop) data in 2015 is overestimate compared to the Indonesia Statistical Agency (BPS) at the Provincial level. Second, the distribution of population density values is not only in residential areas but also outside the areas. Then the IndoPop data quality is necessary to be improved.

This research introduces a new method to improve population distribution data by combining the two earlier mentioned methods for more reliable population data as well as its accuracy.
2 Material and method

2.1. Data

The data used in the form of spatial data consists of:

Table 1. Data types

| No. | Data Type                                  | Forms Data       |
|-----|--------------------------------------------|------------------|
| 1.  | Village Administration Boundaries          | Vector (Polygon) |
| 2.  | Indonesian Population Distribution (IndoPop) | Raster          |
| 3.  | Residential Distribution                    | Vector (Polygon) |
| 4.  | Amount of Population (Village level)        | Tabular          |

The official source data are available on the internet. The village-level boundary data and the population data are obtained from BPS. The IndoPop data is available online and the residential distribution is extracted from 1:25,000 topography map which is also freely available (tanahair.indonesia.go.id) which can be opened through ArcGIS - ArcMap program.

2.2 Methodology

The methods used in this study are data pre-processing, improving IndoPop data, and spatial distribution of the population.

2.2.1 Data pre-processing

This step aim to prepare the initial data for a further analysis. Since the data is run in a raster format, the data must be common in the format, cell size, and coordinate system. All data must use a Mercator projection such as WGS 1984 World Mercator or Universal Transverse Mercator (UTM). The transformation of the coordinate system can be done using ArcToolbox - Projections and Transformations - Project (for vector data) or ArcToolbox - Projections and Transformations - Raster - Raster Project.

2.2.2 Improving IndoPop data

Since the data have inaccurate estimated population and distribution, it is necessary to improve the IndoPop data. The basic concept of this step is to distribute the population only in the settlement area. This step can be run through the following procedures:

- Ensure that the data polygon settlement is distributed for all administrative boundaries of the village.
- Change the grid size (pixel/cell) of Indo Pop to 100 with Resample tools. Use resample technique "BILINEAR".
- Overlay the polygon boundary settlement and study area data with Union tools.
- Add a column (field) with the attribute of Short Integer.
- Please rate 1 for the residential attribute and give 0 for the non-residential attribute.
- Convert to raster polygon overlay by using Polygon to Raster tools and set the size of the cell to 100 (uniformity with IndoPop data).
- Get Indo Pop pieces of data based on study area boundaries using tools Extract by Mask or Clip Raster.
- Use the Raster Calculator tool to perform the data Indo Pop settlement based on data from existing polygons by using the syntax:

  \[
  \text{Con ("raster_A" == 1, "raster_B", "raster_A")}
  \]

  Where "raster_A" is settlement raster, and "raster_B" is IndoPop raster.

- The result of these steps is an improved IndoPop data.

2.2.3 Spatial distribution of population

This step deals with the population size. The population data from the IndoPop will be adjusted according to the BPS data. In this case, the spatial data distribution will be updated based on BPS in 2015. The procedure can be done by following scheme:

- Use the tools Zonal Statistics to count the number of inhabitant village of the Indo Pop data. Select the type of statistics SUM to obtain the number of residents in each village based on administrative boundary data.
- Convert polygon administrative boundaries of the village into raster data based on the number of people by using the tools Polygon to Raster. Set the size of the cell to 100 (uniformity with IndoPop data).
- Use the Raster Calculator tool to redistribute the value of the total population based on data Indo Pop improvement and the existing population in 2015 by using the formula:

  \[
  P = \frac{\sum_{j=1}^{N} p_j}{\sum_{j=1}^{N} p_1} \cdot x \cdot p_j
  \]

  where \(P\) is the number of people per grid; \(p_j\) is the number of people form WorldPop data \(i\); \(p_1\) is the number of people form BPS by village data \(j\), and \(j\) is the village.
3 Results and discussion

3.1 Administrative-based of population distribution

The map shows the population distribution of Banten Province in village-level basis. Each color in the map represents the total population of the villages. In general, the north parts of the area are denser than the south one.

![Fig 1. Population distribution in administrative bases](image)

3.2 Random regression tree model-based forest mapping

The map (Fig 2) illustrates the population distribution in the study area generated through the Random Regression Tree model-based Forest Mapping method. Generally, this map displays the same population distribution as the previous map in which the population concentrates in the northern part of the study area.

![Fig 2. Population distribution is generated from the Random Regression Tree model-base Forest Mapping](image)

3.3 Final results

The map shows the population distribution of Banten region modeled through the combination between the administrative-based mapping method and the Random Regression Tree model-based Forest Mapping method. It shows the population distribution pattern agree with the administrative and WorldPop map.

![Fig 3. Population distribution produced by the combination method.](image)

Even though the three maps show the similar population distribution pattern, each map actually has the different accuracy. For instance, in the first map (Fig.1) the amount of population is distributed evenly in administrative bases. The map is also able to distinguish the number of people live in the area. However, this choropleth map is insufficient to show the population distribution precisely since the people are not distributed evenly but concentrated in settlement areas.

Figure 2 displays the population proportionally based on the land use classification. It addresses the issue of population unevenness. However, when it comes to the population calculation, it may overestimates or underestimates the amount of population (see Table 1). The combination of administrative-based and the Random Regression Tree model-base Forest Mapping does not only deal with the distribution issues but also provides more reliable amount of population compared to the statistics data.

Table 1 shows an underestimate of the total population using WorldPop data. One of the possible reasons is the Random Regression Tree model-base Forest Mapping utilizes a medium-resolution satellite imagery to identify the human presence. As a result, some settlement areas may not be captured, especially for one that smaller than image spatial resolution. In contrast, the combined method uses a settlement dataset derived from the topographic map (1: 25,000) is more detail. Therefore, it can distribute the population data precisely within the settlement area.

In order to illustrate the population distribution data effects the disaster risk assessment, the analysis results of three different methods are compared each other. Figure shows the study area overlayed by flood hazard map. From the figure the number of flood exposure can be estimated as displayed in Table 2.
There may be some distortions if one intends to use this method relies on WorldPop data, all data need to be adjusted accordingly to the WorldPop data structure. This method has some limitations. Furthermore, in case of an updating and distribution mechanism for the census data, the population density using land cover combination method, Fig 4.

Fig 4. Flood hazard map (inrisk.bnpb.go.id)

Compared to the BantenPop, which uses the combination method, the choropleth mapping method is overestimated but the WorldPop data (Random Regression Tree model-base Forest Mapping method) is underestimated. Disaster risk analysis requires an accurate data and information as it will affect the disaster management plan. The overestimate calculation will result in more resource needs such as logistics, housing, and sanitation. On the other hand, an underestimate calculation may result in a lack of resource in case of disaster.

The method used in this study shows the possibility to adjust the WorldPop before it is used. It also provides an updating and distribution mechanism for the census data which usually exists in administrative format. Although the developed method is able to provide more reliable distribution and population information, this method has some limitations. Firstly, since this method relies on WorldPop data, all data need to be adjusted accordingly to the WorldPop data structure especially the cell size (approximately 100 x 100 m). There may be some distortions if one intends to use this method in a finer pixel resolution. Therefore, it requires further study to examine the significance of this issue as well as some possible solutions. Secondly, this method also depends on the quality of input data.

Table 3. Amount of population in the comparison between WorldPop, BantenPop (combined method), and statistics data.

| No  | Distric Name      | WorldPop     | BantenPop (combined method) | Statistics Data |
|-----|-------------------|--------------|-----------------------------|-----------------|
| 1   | Pandeglang        | 1.157.849    | 1.194.547                   | 1.194.547       |
| 2   | Lebak             | 1.270.450    | 1.263.211                   | 1.263.211       |
| 3   | Tangerang         | 3.321.367    | 3.453.444                   | 3.453.444       |
| 4   | Serang            | 1.579.502    | 1.472.883                   | 1.472.883       |
| 5   | Tangerang City    | 2.064.925    | 2.046.998                   | 2.046.998       |
| 6   | Cilegon           | 425.577      | 412.058                     | 412.058         |
| 7   | Serang City       | 499.372      | 643.135                     | 643.135         |
| 8   | South Tangerang   | 1.491.431    | 1.543.143                   | 1.543.143       |
|     | Grand TOTAL       | 11.810.473   | 12.029.419                  | 12.029.419      |

4 Conclusions

Disaster risk analysis requires accurate population distribution data to calculate the vulnerability. The combined methods of administrative-based and the Random Regression Tree model-base Forest Mapping can be an excellent alternative to overcome the population data.

GIS is an effective tool to deal with spatial data. Therefore, it can play a greater role in disaster management practices.

The study is a part of National Disaster Risk Analysis project conducted by the National Disaster Management Authority of Republic Indonesia (BNPB) in 2015.

References

1. Lung, T., Lübker, T., Ngochoch, J. K., & Schaab, G. Human population distribution modeling at regional level using very high-resolution satellite imagery. Applied Geography, 41, 36-45. (2013)
2. Tian, Y., Yue, T., Zhu, L., & Clinton, N. Modeling population density using land cover...
3. Yulianto, F., Tjahjono, B., & Anwar, S. The applications of Monte Carlo algorithm and energy cone model to produce the probability of block-and-ash flows of the 2010 eruption of Merapi volcano in Central Java, *Indonesia Arabian Journal of Geosciences*, 8(7), 4717-4739. (2015)

4. Khomarudin, M. R., Strunz, G., Ludwig, R., Zoßeder, K., Post, J., Kongko, W., & Pranowo, W. S. Hazard analysis and estimation of people exposure as contribution to tsunami risk assessment in the west coast of Sumatra, the south coast of Java and Bali. *Zeitschrift für Geomorphologie, Supplementary Issues*, 54(3), 337-356. (2010)