

# A METHOD FOR ESTIMATING THE NUMBER OF HOUSEHOLDS IN A REGION FROM THE NUMBER OF BUILDINGS ESTIMATED BY DEEP LEARNING WITH THE ADJUSTMENT OF ITS NUMBER USING ANCILLARY DATASETS: CASE STUDY IN DJAKARTA

R. Ito<sup>1</sup>, K. Hara<sup>1</sup>, Y. Shimazaki<sup>1</sup>, N., Mori<sup>1</sup>, Y. Kani<sup>1</sup>, A. Okabe<sup>2</sup>

<sup>1</sup> PASCO CORPOLATION, Tokyo, Japan - (roithi7072, kaaty6781, yiaksa9955, niorro6598, yionsh3531)@pasco.co.jp

<sup>2</sup> Aoyama Gakuin University, Tokyo, Japan - atsu\_okabe@sccs.aoyama.ac.jp

**KEY WORDS:** Deep learning, Convolutional neural network, Household estimation, Urban classification

## ABSTRACT:

The high resolution statistical data such as the number of households in small areas are indispensable for urban planning, disaster prevention and many kinds of business activities. However, it is difficult to obtain the number of households in small areas because census data are usually aggregated in municipal districts. Techniques for automatically analyzing statistical data, e.g., land cover, population density, and the number of households obtained from satellite/aerial images have been continuously studied. In recent years, many methods using deep learning have been proposed in the related literature. In estimating the number of households, the use of buildings, the number of floors and that of rooms are also important information, but it is difficult to obtain such information from only image analysis using deep learning. This study proposes a method for estimating the number of households in 100 meter grid cells from satellite images using deep learning, and adjusting it using ancillary data obtained from a few statistical datasets. The application of this method to Djakarta shows that the difference between the estimated values and the corresponding values of census is less than 10%.

## 1. INTRODUCTSION

Estimation of population distribution and the number of households in small areas over a region is frequently required for the analysis of spatial activities or regional planning, for example, urban land use zoning, disaster prevention in hazardous areas, environmental preservation along rivers, economic development in deprived neighborhoods, local area marketing, and so forth. To analyze these activities, statistical data in small areas are increasingly demanded in recent years (Smith, 2002). In practice, however, small-area data are not always easily available. Therefore, most studies use the data aggregated into municipal districts, typically, census data. The aggregated data are, however, often too coarse to analyze activities in small areas. The time resolution is also important but that of census data is too coarse, especially in developing countries, because census survey is carried out every 5-10 years in these countries. The coarse resolution makes timely analysis difficult. To overcome this limitation, a number of methods for producing small-area data is under development.

A method for estimating population or the number of households in small areas is to utilize land use classification maps published by the government. However, these maps are usually updated at every other year, and hence it is difficult to obtain the latest land uses. To overcome this difficulty, alternative methods are proposed in the related literature. A method is: first, to generate a land use classification map from remote sensing data, and second, to estimate population or the number of buildings from the resulting map (Lu, 2006; Silvan-Cardenas, 2010). The land use classification maps made from satellite images are also applied to many urban analyses, for example, the estimation of urban growth rate (Martinuzzi, 2006) and disaster recovery (Sheykhmousa, 2019). The procedure in these studies is: first, to collect images and statistical data; second, to generate of a land use classification map; third, to

measure objects such as artefacts, and forth, to estimate population, city growth rates, and so forth.

Besides land use classification maps, various kinds of statistical data are used for estimation of population or the number of households. For instance, Bast et al. (2015) estimated population by a regression model using OpenGeoDB, census and Open Street Map (OSM) data, which include polygons of buildings, points of interest (PoI), road networks, and land use maps. Robinosn et al. (2017) proposed a method which first classified patches of satellite images into 14 classes according to building densities, and second, estimated population by weighting a factor on each resulting class. Although this method required neither a land use classification map nor statistical data, it has a limitation in that the population distribution between residential and non-residential areas were not explicitly taken into account. As a result, estimated population was likely to be different from census data.

In recent years, the deep learning method attracts attention in satellite/aerial image analysis. One of the advantages of this method is that it can automatically extract and identify features using supervised data (Krizhevsky, 2012). In remote sensing, the deep learning method is applied to various analyses, such as population estimation (Robinson, 2017; Silvan-Cardenas, 2010), land cover classification (Martin, 2015; Marco, 2015), and change detection in land use (Mou, 2018; Lebedev, 2018).

Although the deep learning method applied to satellite images analysis is powerful, it should be noted that satellite images are not sufficient enough to estimate the number of households, because they do not indicate the number of floors of high-rise buildings, building uses, and area characteristics that are indispensable for estimating the number of households. In this study, we propose a method for estimating the number of households in small areas (smaller than census tracts) through the following two phases:

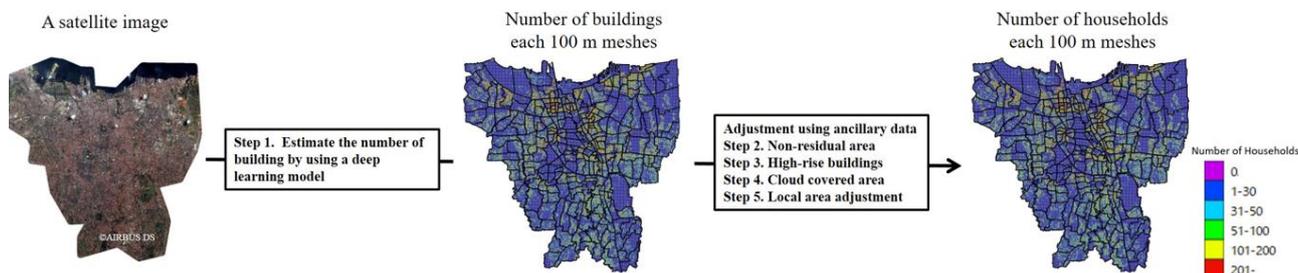


Figure 1. The outline of a proposed method.

- (i) to estimate the number of buildings from satellite images using the deep learning method,
- (ii) to convert the number of buildings to the number of households with adjustment processing using ancillary data obtained from POI, census, OSM, and other statistical data. This method was tested in Djakarta city.

The remainder of this paper is organized as follows. Section 2 introduces a method for estimating the number of households using the deep learning method with adjustment processing. Section 3 describes the results and discusses the accuracy of the proposed method. The paper ends in Section 4 with concluding remarks for future studies.

## 2. METHOD

### 2.1 Dataset

The target area was Djakarta city. The satellite images used for our analysis were optical satellite images (SPOT) and synthetic-aperture radar (SAR) images (TerraSAR-X). The resolutions of SPOT and TerraSAR-X images are 1.5 meters and 0.5 meters, respectively. SAR images are particularly useful for analyzing Southeast Asian countries, because these countries have many cloudy days. However, the visual resolution of SAR is lower than that of optical images. Therefore, it is difficult to count the number of buildings, especially in densely built areas. In this study, optical images were used together with a residential map (land cover map) created by SAR images.

As for statistical data, we used the OSM (open street maps) in Indonesia (updated in 2017), the local zone map released by the government, the POI of high-rise buildings, and the census of Indonesia (updated in 2010).

The OSM was used as supervised data for estimating the number of buildings in each cell, and POI for adjusting the number of households in a high-rise apartment building. To examine estimation accuracy, the number of households in each cell obtained from census data was regarded as the true number, which was compared with our estimated number. The result showed that there were a few areas where discrepancy was large. This discrepancy was corrected manually if we could find its causes.

### 2.2 Estimation of the number of buildings by the deep learning method

Our proposed method consisted of five steps shown in Figure 1. Section 2.2 introduces an application of the deep learning method to the estimation of the number of buildings in 100 meter grid cells using satellite images.

Techniques for detecting building shapes from high-resolution aerial photographs or satellite images have been proposed by

Bischke (2019), Hamguchi (2018) and others. As is noticed from Figure 2, in the case of SPOT optical images, the resolution was 1.5 meters, which was not sufficient enough to detect building boundaries accurately. Therefore, we did not employ a method for detecting the shape of buildings. Instead, we directly estimated the numbers of buildings in 100 meter grid cells using supervised data obtained from OSM data. Stated explicitly, the supervised data were pairwise 100 m grid cells data consisting of patches of SPOT images, and the number of buildings obtained from OSM data. Data in several local areas with low OSM accuracy were corrected by census data or visual confirmation.



Figure 2. SPOT image and building polygons from OSM.

Figure 3 shows the architecture of our deep learning model for estimating the number of buildings. The input data were the patches of 100 m grid cells cut out from optical satellite images, and the output data were the estimated number of buildings in 100 m grid cells. The model consisted of six convolutional layers, three pooling layers and one fully connected layer. The convolutional layer and the pooling layer were designed to learn the attribute values that were possibly effective for estimating the numbers of buildings. The fully connected layer was designed to estimate the number of buildings within each input patch. The activation function of the final layer was a liner function for regression type estimation (not classification type). In the learning process, the minimum squared error function was used to calculate the error between the numbers of buildings (ground truths) and the corresponding estimated values.

At this intermediate phase, the number of households was equal to the number of buildings. This estimated number was adjusted by ancillary data in the following steps.

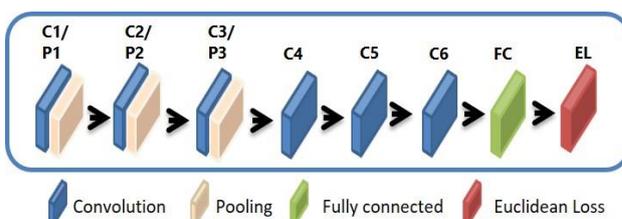


Figure 3. A deep learning network architecture of estimating the number of buildings.

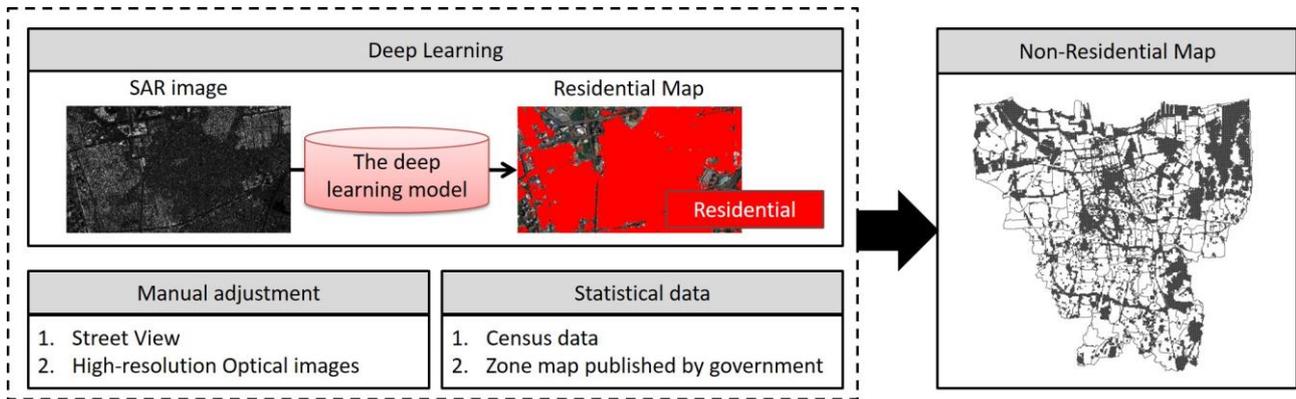


Figure 4. Flow chart of generating non-residential area map

### 2.3 Adjustment for non-residential area

The numbers of households living in buildings varied according to whether buildings were residential use or non-residential use (specifically, commercial, industrial, or governmental uses). Therefore, non-residential use areas were adjusted by the map made from ancillary data shown in Figure 4. Stated explicitly, the map was constructed from OSM, the Google Street View, the local land use map provided by the government of Indonesia and the residential map generated from TerraSAR-X images (resolution is 0.5 m).

The residential map referred to in the above was generated by the deep learning method that was applied to land cover classification (Hu, 2015; abdikan, 2016; Xu, 2017). We employed this method with the specification that supervised data consisted of two classes, namely urban or others. In a few areas in Djakarta, the training data were constructed from manually-made residential maps.

The deep learning model for a non-residential map was almost similar to that for the residential map in Figure 3 except that the activation layer was replaced with the softmax function and the loss function was replaced with the cross-entropy loss function. Figure 5 shows the residential map obtained from the deep learning model.

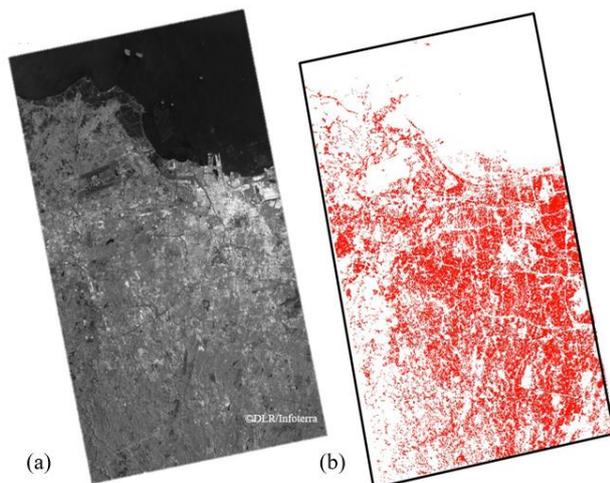


Figure 5. SAR image map and the generated residential map. (a) Input SAR image and (b) residential map generated by the deep learning model (the red points are residential)

Note that the above model was the classification type model (not the regression type model), and that the estimated numbers in grid cells in non-residential areas were set zero.

### 2.4 Adjustment for high-rise apartment

The number of households in a high-rise apartment building was different from that in a detached house. Therefore, the number of households in a grid cell including high-rise apartments (which were known from POI) were adjusted by the number of rooms per floor and that of floors obtained from local real estate websites or booklets; that is, the value was adjusted by the number of households per floor multiplied by the number of floors.

### 2.5 Adjustment for cloud covered area

In processing satellite images, the numbers of buildings in the areas covered with clouds were adjusted by identifying the number of buildings known from OSM data.

### 2.6 Adjustment for large deviation rates

The number,  $n_i$ , of households in each municipal district estimated by our method was compared with that,  $m_i$ , of census, and the deviation rate,  $(m_i - n_i) / m_i$  was calculated across all municipal districts. The deviation rate is positive if the number of households reported in the census data is larger than our estimated value, and negative if it is less than our estimated value. The comparison showed that deviation rates were small in 254 municipalities, while, as shown in Figure 6(a), they were large in ten municipalities. We re-examined these municipalities using high-resolution satellite images and related statistical data. As a result, we found that in seven municipalities, the density of buildings was extremely higher than that in other municipalities. Moreover, there were commercial and industrial facilities or high-rise apartment houses which we should have adjusted in the adjustment processing referred to in the above. Because we had found why the large deviations appeared, we could adjust them manually.

Figure 6(b) shows the adjusted deviation rates. The deviation rates were reduced in seven municipalities.

In three municipalities, the household numbers of the census were extremely larger than observed data. However, in the municipalities, there were no features that would cause a large number of households such as high-rise apartment.

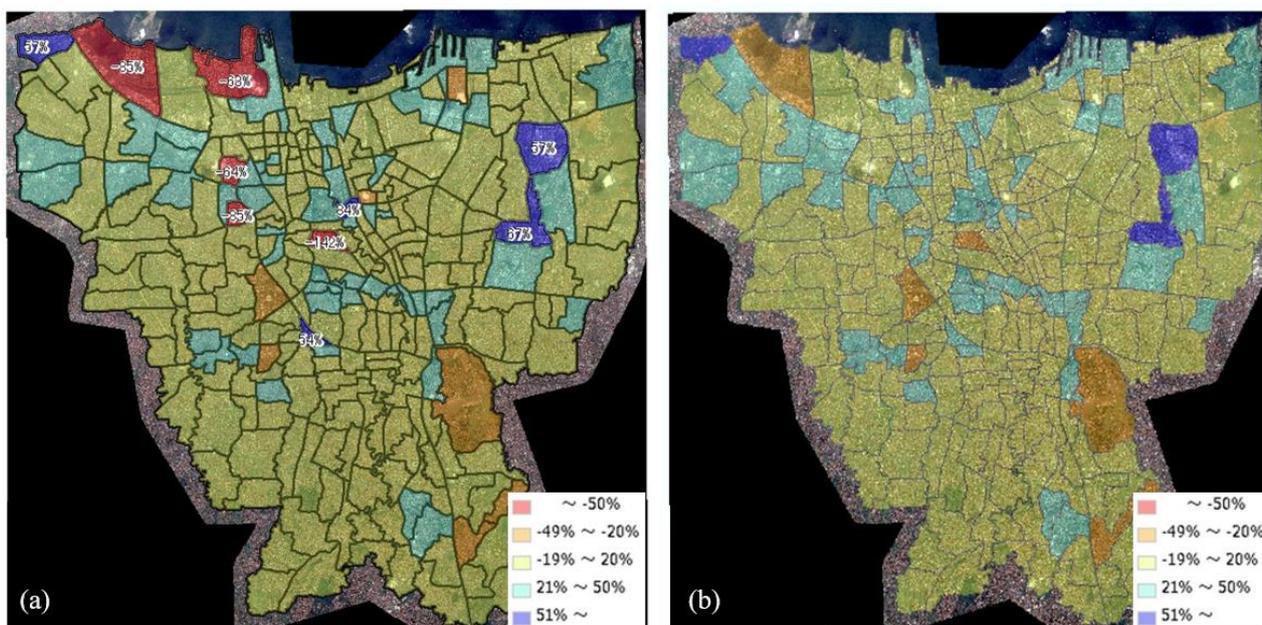


Figure 6. The divergence maps each municipalities. (a) Before adjustment and (b) after adjustment.

### 3. RESULTS AND DISCUSSION

Figures 7 and 8 show the number of households in 100m grid cells estimated by our deep learning method with several adjustments mentioned in Section 2. Note that in Figure 8, the grid cells where buildings existed but the number of households was zero imply the cells where buildings were public, industrial, and commercial facilities. Our results also showed that the number of households was large in high building density areas (which were known from SPOT images), and small in low density areas. This result suggests that the estimation of the number of households by our method is effective to estimate not only the number of households but also that of buildings.

In numerical terms, the number of households estimated by our method was 2,164,945, and that of the census was 2,404,745. The total deviation rate was around 10%, which looks fairly good. This deviation might have resulted from the fact that the distribution of buildings and their surrounding environments in Indonesia have a special kind of provinciality, which was difficult to adjust in our method. The deviation might also resulted from the fact that the updated year of the OSM, that of satellite images used as supervised data and that of the ancillary data were different from that of the census by several years. Therefore, the actual deviation rate may be slightly different.

### 4. CONCLUSIONS

The application of the proposed method to Djakarta shows that the method is practically useful for estimating the number of households in 100 m grid cells. This method would become more practical if the following limitations are overcome. First, this method required manual adjustments. To minimize this work, the deep learning model for identifying buildings should be improved. The accuracy of the identification hinges on the resolution of optical satellite images. As is noticed from Figure 2, the boundaries of buildings in SPOT images (the resolution of 1.5m) were blurred in high density areas. To estimate more accurately, finer resolution of satellite images

should be used to detect the boundaries of buildings distinctively.

Second, the ancillary data used in our method are not always easily available in developing countries. Therefore, an adjustment method with easily available satellite images and open data should be developed.

In future, we wish to develop a deep learning method for a completely automated estimation of the number of buildings in small areas with easily available data.

### REFERENCES

- Abdikan, S., Sanli, F. B., Ustuner, M., Cald, F., 2016: Land cover mapping using sentinel-1 SAR data. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 41, 757.
- Bast, H., Storandt, S., Weidner, S., 2015: Fine-grained population estimation. *In Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp. 1-10.
- Bischke, B., Helber, P., Folz, J., Borth, D., Dengel, A. 2019: Multi-task learning for segmentation of building footprints with deep neural networks. *2019 IEEE International Conference on Image Processing*. pp. 1480-1484.
- Hamaguchi R., Hikosaka S., 2018: Building detection from satellite imagery using ensemble of size-specific detectors. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*.
- Hu, F., Xia, G. S., Hu, J., & Zhang, L., 2015: Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. *Remote Sensing*, 7(11), 14680-14707.
- Krizhevsky, A., Sutskever, I., Hinton, G. E., 2012. Imagenet classification with deep convolutional neural networks. *In Advances in neural information processing systems*, pp. 1097-1105.

Lebedev, M. A., Vizilter, Y. V., Vygolov, O. V., Knyaz, V. A., & Rubis, A. Y., 2018: Change detection in remote sensing images using conditional adversarial networks. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 42(2).

Lu, D., Weng, Q., Li, G., 2006: Residential population estimation using a remote sedetectensing derived impervious surface approach. *International Journal of Remote Sensing*, 27(16), 3553-3570.

Martinuzzi, S., Gould, W. A., Gonzalez, O. M. R., 2007: Land development, land use, and urban sprawl in Puerto Rico integrating remote sensing and population census data. *Landscape and Urban Planning*, 79(3-4), 288-297.

Mou, L., BruzOzone, L., Zhu, X. X., 2018: Learning spectral-spatial-temporal features via a recurrent convolutional neural network for change detection in multispectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 57(2), 924-935.

Robinson, C., Hohman, F., Dilkina, B. 2017: A deep learning approach for population estimation from satellite imagery. In *Proceedings of the 1st ACM SIGSPATIAL Workshop on Geospatial Humanities*, pp. 47-54.

Sheykhmousa, M., Kerle, N., Kuffer, M., Ghaffarian, S., 2019: Post-disaster recovery assessment with machine learning-

derived land cover and land use information. *Remote sensing*, 11(10), 1174.

Silvan-Cardenas, J. L., Wang, L., Rogerson, P., Wu, C., Feng, T., Kamphaus, B. D., 2010: Assessing fine-spatial-resolution remote sensing for small-area population estimation. *International Journal of Remote Sensing*, 31(21), 5605-5634.

Smith, S. K., Nogle, J., Cody, S., 2002: A Regression approach to estimating the average number of persons per household. *Demography*, 39, pp. 697-712.

Xu, Z., Wang, R., Zhang, H., Li, N., Zhang, L., 2017: Building extraction from high-resolution SAR imagery based on deep neural networks. *Remote Sensing Letters*, 8(9), 888-896.

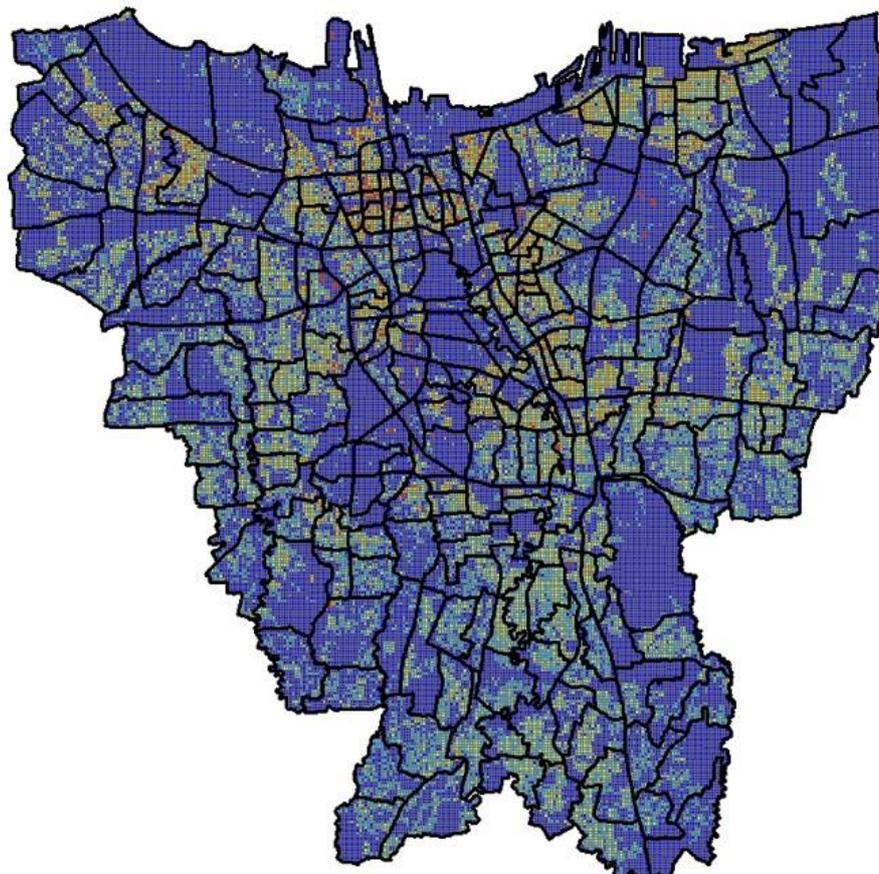


Figure 7. Results of the estimated number of households.

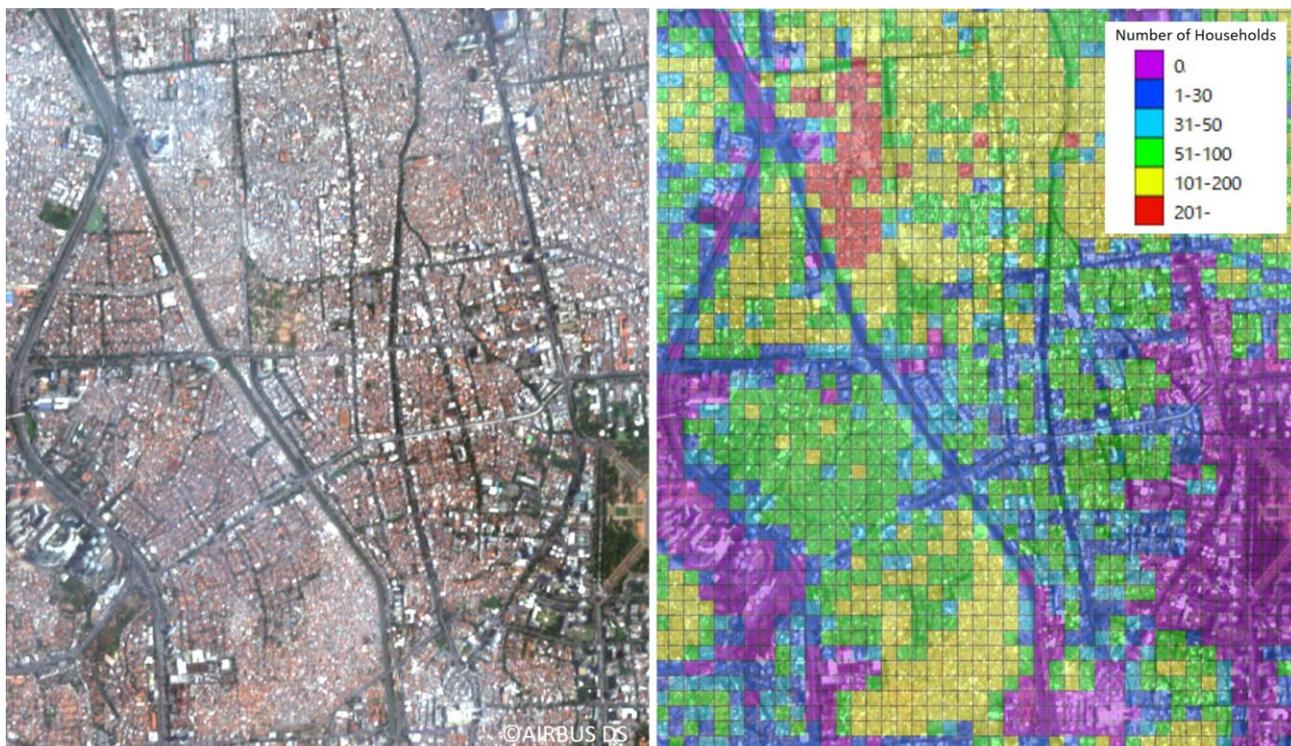


Figure 8. Results of the estimated number of households in a local area in Djakarta..