Land mapping with least median of squares regression using landsat imagery: a case study Jakarta and surrounding area

Bagus Mulyawan*, Nicholas Jovianto, Janson Hendryli, and Dyah Erny Herwindiati
Faculty of Information Technology, Universitas Tarumanagara, Jakarta 11440, Indonesia
*bagus@untar.ac.id; *Corresponding author’s email

Abstract. Impervious surface is an area that has undergone natural or semi-natural substitution of artificial land cover which is usually waterproof and relatively permanent, including settlements, industrial buildings, road networks, railroad networks, high voltage electricity networks, airports air and sea ports. Changes cause the narrowing of rivers and open land which become water catchment areas. One of the problems that triggered an increase in impervious land is urbanization. This study aims to map land changes in the Jakarta and surrounding areas using the least median of squares regression method. Based on data from land classification results using 4 classes, Jakarta and Tangerang are areas that have more impervious land than non-impervious land. Meanwhile, Bogor, Bekasi, and Depok are areas that have more non-impervious land.

1. Introduction
Impervious surface (built up land) is an area that has undergone natural or semi-natural land cover substitution which is artificial cover which is usually waterproof and relatively permanent, including settlements, industrial buildings, road networks, railroad networks, high voltage electricity networks, airports air and sea ports [2]. According to Stankowski (1972) and Arnold & Gibbons (1996), population growth and urbanization are closely related to an increase in the percentage of developed land that cannot absorb water. Expansion of residential areas can result in a decrease in open land to build land that cannot absorb water. Flood disaster in urban areas is one of the real examples of the decline in open land as a water catchment area. In addition to flooding, various other environmental problems can arise as a result of the change in open land, including erosion disasters, landslides and water crises.

2. Method
The existence of outliers in linear regression which is a problem because outliers can cause the formation of a regression parameter model to be less accurate. In overcoming this problem, statisticians try to find alternative estimates of other parameters that are better at overcoming outliers. One of the methods suggested in muscular regression is Least Median of Squares (LMS). Least Median Squares (LMS) is a parameter estimation method with a high breakdown point that was introduced by Peter J. Rousseeuw in 1984. This method predicts the regression parameters by minimizing the median square of the remainder of the observations. The LMS estimator is defined as follows:

$$\hat{b}_{\text{LMS}} = \arg \min_{b} Q_{\text{LMS}} (b).$$

(1)

where $Q_{\text{LMS}}(b)$ is the median squared remaining of $h$ observations, $\text{med}(e_h^2)$. 
In general, the algorithm for applying the LSM Regression method can be summarized in the steps below. Suppose that a data group of size N is given, and you want to expect a vector $\theta$ of dimension p that contains the parameters of the data group. The steps that need to be done are:

1. Determine the size of the subset $n$, the number of subsets $M$ according to the number of classes.
2. Randomly, take $M$ subset size $n$ from sample size N. Find the estimated parameter $\hat{\theta}_j$ for each subset.
3. Find the median squared error $e_{ij}^2$ from each subset. $i$ is the index for the example, and index $j$ is the index for the subset.
4. Define:
   \[
   m = \arg \min \left[ \text{med} (e_{ij}^2) \right] \]  
   So the subset $\hat{\theta}_m$ is a subset with the median least square error and $e_{im}$ is the error vector generated by that subset.
5. Count:
   \[
   S_0 = 1.4826 \left( 1 + \frac{5}{n-p} \right) \sqrt{\text{med}_{i}(e_{im}^2)} \]  
6. Calculate the weight of $w_i$, for example with $w_i = 1, \frac{|e_i|}{\text{median}}$ and $w_i = \frac{S_0}{|e_i|}$.
7. Give $w_i$ the weight of each example.
8. Make a fit using the weighted least squares method using $w_i$ as the weight to get the final $\hat{\theta}$.

The method used in building this application is the least median of square regression method. This method is a regression method that is very suitable to be applied to a calculation of the regression equation model when the data held have outliers. This method will form a regression parameter based on the median value of the data so that outliers will not be very influential on calculations. making this application program will require input data in the form of spatial data in the form of Landsat 8 satellite imagery for the Greater Jakarta area. Landsat 8 satellite imagery data that will be used on the system can be downloaded on the USGS Global Visualization Viewer website page (https://glovis.usgs.gov/).

The Landsat 8 imagery downloaded will include two intersection files (map codes 122064 and 122065) for each time to cover the whole Jabodetabek area. The map that is used as data in making this land use change detection application program will cover band 1, band 2, band 3, band 4, and band 5. The choice of using these bands is based on the spectrum represented by each of these bands. Band 1 represents the coastal / aerosol spectrum, making it suitable for detecting territorial waters. Band 2, band 3, and band 4 represent RGB images that are easily captured by the human eye. Band 5 representing the Near-Infrared (NIR) spectrum is commonly used to detect vegetation density when combined with band 4, for example by implementing it in NDVI or SAVI calculations.

### 3. Result and Discussion

Testing with 4 (impervious land, green land, water, and vacant land) class will use testing data for the Greater Jakarta area in the 2014, 2015 and 2017 periods. Comparison between the actual area and the number of pixels per Jabodetabek scene in Landsat 8, can be seen in Table 1.

| City        | Area (km$^2$) | Number of pixels per scene |
|-------------|---------------|----------------------------|
| Jakarta     | 659.26        | 732,514                    |
| Bogor       | 3,102.35      | 3,447,053                  |
| Depok       | 180.01        | 200,015                    |
| Tangerang   | 1,353.45      | 1,503,838                  |
| Bekasi      | 1,483.90      | 1,648,775                  |
In image detection using Landsat 8 data, it is known that the resolution of each pixel in the Landsat 8 data represents the original area with a size of 30x30 meters. So that every 1 pixel of Landsat 8 data will represent an actual area of 0.0009 km². The results of the calculation of land area based on each region in 2014, 2015, and 2017 with observations of 4 classes can be seen in Table 2.

| City     | Impervious (km²) | Green (km²) | Water (km²) | Clear (km²) | Area (km²) |
|----------|------------------|------------|-------------|-------------|------------|
| Jakarta  | 570.2382         | 76.6107    | 11,7153     | 0.6984      | 659.26     |
| Bogor    | 302,1255         | 2,460.9663 | 17,7795     | 321.4764    | 3,102.35   |
| Depok    | 82.1691          | 95.0427    | 0.9243      | 1.8774      | 180.01     |
| Tangerang| 706.6944         | 581.7204   | 50,1795     | 14.8599     | 1,353.45   |
| Bekasi   | 697.8627         | 596.8539   | 53,2593     | 135.9216    | 1,483.90   |
| Jabodetabek | 2,359.0899   | 3,811.1940 | 133,8579    | 474.8337    | 6,778.98   |

If categorized as impervious and non-impervious land (green land, water, and vacant land), the comparison of the percentage of impervious and non-impervious land detection results for each region based on 4 class observations can be seen in Table 3.

| Years | City     | Impervious | Non-impervious |
|-------|----------|------------|----------------|
| 2014  | Jakarta  | 84.78 %    | 15.22 %        |
| 2015  | Jakarta  | 77.54 %    | 22.46 %        |
| 2017  | Jakarta  | 86.49 %    | 13.51 %        |
| 2014  | Bogor    | 8.06 %     | 91.94 %        |
| 2015  | Bogor    | 10.26 %    | 89.74 %        |
| 2017  | Bogor    | 9.75 %     | 90.25 %        |
| 2014  | Depok    | 43.84 %    | 56.16 %        |
| 2015  | Depok    | 39.87 %    | 60.13 %        |
| 2017  | Depok    | 45.64 %    | 54.36 %        |
| 2014  | Tangerang| 52.41 %    | 47.59 %        |
| 2015  | Tangerang| 58.11 %    | 41.89 %        |
| 2017  | Tangerang| 52.21 %    | 47.79 %        |
| 2014  | Bekasi   | 36.23 %    | 63.77 %        |
| 2015  | Bekasi   | 49.45 %    | 50.55 %        |
| 2017  | Bekasi   | 47.02 %    | 52.98 %        |

Based on data from land classification results using 4 classes, Jakarta and Tangerang are areas that have more impervious land than non-impervious land. Meanwhile, Bogor, Bekasi, and Depok are areas that have more non-impervious land. Percentage of the amount of impervious and non-impervious land every year experiences fluctuating changes. In the Bogor, Tangerang and Bekasi regions, the number of impervious land has increased in 2015 and decreased in 2017. While in the Jakarta and Depok area, impervious land has decreased in 2015 and experienced an increase in 2017.

4. Conclusion
In general, the most frequent changes in land use in most Jabodetabek areas are land use changes from the type of green land to impervious land. The following are some regional data samples that experienced changes in land use in the period 2014 to 2017 which can be seen from Figure 1 and Figure 2.
5. References

[1] Herwindiati, Dyah E., Hendryli, Janson, & Hiryanto, Lely. (2017). Impervious Surface Mapping Using Robust Depth Minimum Vector Variance Regression. *European Journal of Sustainable Development* 6(3) 29-39

[2] Rousseeuw, P. J. (1984). Least Median of Squares Regression. *American Statistical Association*, 79(388), 871-880

[3] U. S. Geological Survey. (2016). Landsat 8 Data User Handbook. Retrieved from [https://landsat.usgs.gov/sites/default/files/documents/Landsat8DataUsersHandbook.pdf](https://landsat.usgs.gov/sites/default/files/documents/Landsat8DataUsersHandbook.pdf)

[4] Draper, N. R. & Smith, Harry. (2014). *Applied Regression Analysis, Third Edition*. New York: John Wiley

[5] Rousseeuw, P. J. & Leroy, A.M. (1987). *Robust Regression and Outlier Detection*. New York: John Wiley