Data: Historical Urban Development in Jakarta

Allan Hsiao
Princeton University

June 9, 2023
(most recent version here)

I digitize Dutch colonial maps of Jakarta to measure the evolution of built-up land from 1887 to 1945. This file describes the primary data and digitization procedure. I link the original paper here.

Citation

Hsiao, Allan. Sea Level Rise and Urban Adaptation in Jakarta. 2023.

Output data

The processed data record built-up area at 50m and 100m resolutions for the years 1887, 1897, 1904, 1910, 1920, 1930, 1937, and 1945. I provide data in .tif format with georeferencing by coordinate reference system WGS84 (EPSG:4326). The 50m data are as digitized, and the 100m data aggregate over the 50m data to reduce noise. For 50m raster files, the band value is one (developed areas) or zero (undeveloped areas). For 100m raster files, the band value records the number of 50m pixels that are developed and thus takes values from zero to four.

Primary data

Historical maps of Jakarta – formerly Batavia – come from the digital collections of Leiden University Libraries. Table 1 lists years and sources. I select eight maps based on ease of digitization and a desire for consistent coverage throughout the study period, but the table lists all available maps. I georeference and digitize the maps, then overlay them to form a panel. These data capture the extensive margin of built-up land development, but not the intensive margins of density or height.

Email: ajhsiao@princeton.edu. Wenqing Yu provided exceptional research assistance.

Table 1: Dutch colonial maps

Year	Source
1887	Visser & Co. (link)
1897	Topographisch Bureau (link)
1904	Seyffardt's Boekhandel (link)
1910	Official Tourist Bureau (link)
1920	Topografische Dienst (link)
1930	Official Tourist Bureau (link)
1937	G. Kolff & Co. (link)
1945	AFNEI Headquarters Survey Department (link)

Source: Leiden University Library Digital Collections. Maps are also available for 1890 (link), 1905 (link), 1914 (link), 1938 (link), and 1942 (link).

Georeferencing

I georeference each map by overlaying it onto an OpenStreetMap base layer. I do so by selecting and matching five ground control points, as shown in figure 1. I select these points to prioritize accuracy in the vicinity of the National Monument and the West Flood Canal, with a modified set of points before the canal is constructed. I implement the overlay with first-order polynomial (affine) transformation and nearest-neighbor resampling. This affine transformation preserves the collinearity of points by applying only rotation, scaling, and translation, avoiding image distortions but ruling out the exact matching of more than two control points.

Digitizing

I digitize maps with unsupervised machine learning. In each map, red shading denotes built-up areas, while green and write denote undeveloped lands. I divide maps into 50m cells, then I take the modal R, G, and B values across pixels in each cell to obtain one RGB code per cell. I apply a k-means clustering algorithm on these RGB codes to group cells with similar colors. I choose k to obtain no more than one grouping of red cells, and I code these cells as built-up. This approach reduces noise in the image files, which contain red in many different shades. The 1910 map marks built-up areas with red dots instead of shading, and so I apply shading manually then digitize it as above. Figure 2 overlays the image inputs and the digitization outputs, which together illustrate the accuracy of this procedure.

Figure 1: Ground control points for georeferencing

(a) Pre-1918 maps

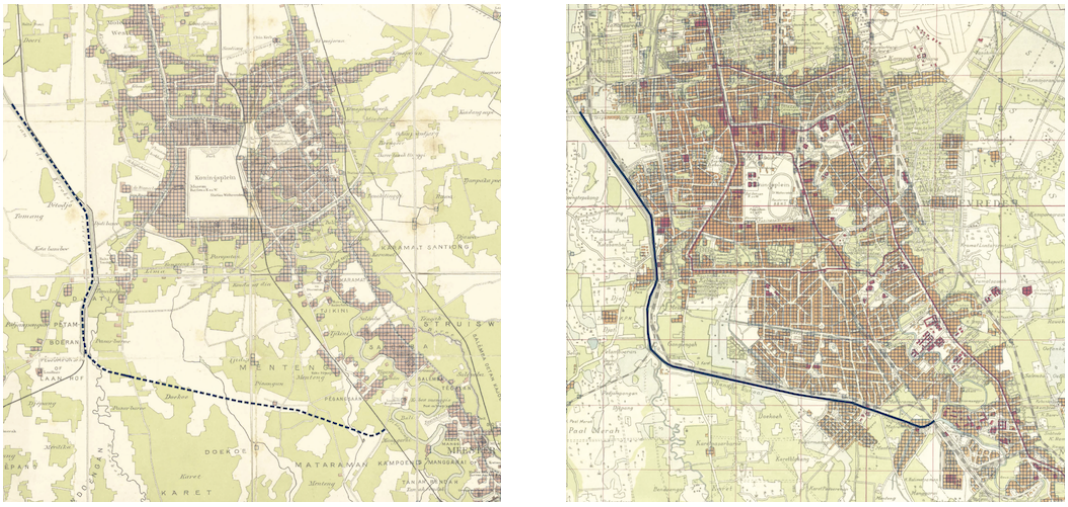
(b) Post-1918 maps

Red stars mark the five ground control points used for georeferencing.

Figure 2: Land development and the West Flood Canal

(a) Before (1904)

(b) After (1945)



Red shading denotes developed lands, and square boxes mark 50m cells that are coded as developed. I mark the West Flood Canal with a black curve – dotted in 1904 before its construction in 1918, and solid in 1945 after its construction.