Projection of road sensors to the Dutch road network

Keywords: Visual inspection, spatial data, big data, traffic statistics

1. Introduction

Road sensors measure the number of passing vehicles every minute. In the Dutch network of highways, there are approximately 20 thousand of those road sensors, resulting in a huge data source. A key step in the production of traffic statistics from this data source is to project the geographic locations of the road sensors on the Dutch road network. To achieve this, road segments have to be defined based on the locations of the road sensors and subsequently the lengths of these segments.

The method that is described in this paper consists of two main parts. First, the main routes per highway per direction are deduced from the detailed road network. Second, these main routes split into road segments based on the locations of the road sensors, and the entrance and exit ramps. We illustrate that visual inspection throughout the whole process is crucial.

All geographic data inspection and editing is done in R, especially with the recently developed package tmap [1]. This package contains a flexible plotting method that is similar to ggplot2 [2], but tailored to spatial data. Also, some processing functions that were needed for this project were added to the package.

2. Method

The first part of the process is the necessary pre-processing of geographic locations of the road sensors and the road network which is described in section 2.1. The extraction of the main routes is discussed in section 2.2. Finally, in section 2.3, we describe the calculation of the road segment lengths.

2.1. Pre-processing geographic location data

For statistical interference of geographic data, it is important to use a proper map projection. Unprojected map coordinates, known as latitude-longitude coordinates, often lead to inaccurate measurements for distances, area sizes, and directions. A good map projection preserves one or more of these properties. For the task at hand we use the Dutch National Grid (Rijksdriehoekstelsel), which is a Cartesian coordinate system that is optimized for the Netherlands. It preserves distances, which means that the difference between any two coordinates in the Netherlands corresponds approximately to the real distance in meters.

Besides the geographic coordinates, the metadata of the road sensors consist of the road names, the direction, and the type of carriageway. Only the road sensors from the main carriageways of the main Dutch highways were selected for the further process.

The information of the Dutch road network is contained in a ESRI shape file. It consists of almost ten thousand polylines that represent different road segments. In Figure 1, the road network including road sensors around an interchange is illustrated. Obviously, the main route segments (coloured green) are selected in order to determine the main roads (see section 2.2). The exit and entrance ramp segments are needed to determine which road sensors correspond to which main road segments (see section 2.3).

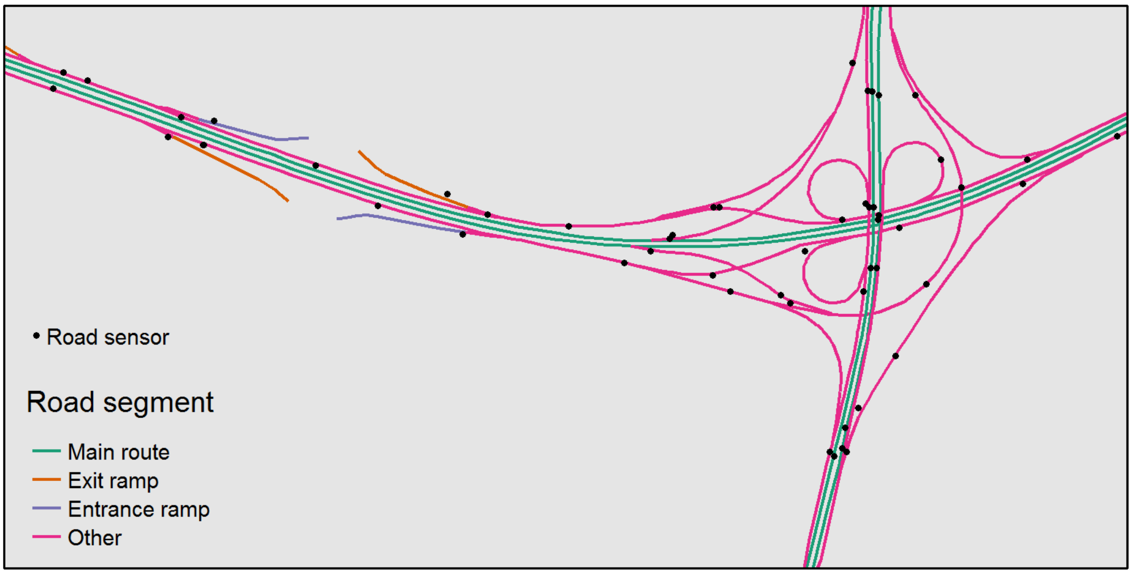

Figure 1. Road sensors on the interchange between the A12 (horizontal) and the A27 (vertical)

2.2. Main routes

Although the main routes in the road network shape file (depicted in green in Figure 1) seem fluent polylines, they cannot be used immediately to calculate main road lengths by two reasons. First of all, each main road consists of multiple polylines which should be glued together. Also, the polylines may represent parallel carriageways in the same direction, which is undesirable since our aim is to calculate the length of the main route.

We introduce the following algorithm to create the main route polylines per road per direction:

i. All polylines get additional points at every *d* meters. By default, let *d*=100.

ii. The points from all polylines are collected, and undoubled based on their location.

iii. A minimum spanning tree is created from all points.

iv. Edges that are longer than *d* are removed. This is only the case if the road consists of disjoint parts. The remaining components represent the disjoint parts of the road.

v. The longest path in each component is the polyline that represents the main route. It is reversed if necessary.

Next, the geographic locations of the road sensors are confronted with the corresponding road polylines. As for the majority of the road sensors, their locations are not too far from the polyline. In those cases, the locations on the polylines at the shortest distance are determined. However, in some cases, the locations of the road sensors are many kilometres away from the highway polyline. This is mainly caused by metadata inconsistence due to errors or difference in reference period. This inconsistence leaves two options. Either the far off road sensors are removed from the data, or the main route polylines are updated with the locations of the road sensors. In the latter case, the locations are added to the points that are collected in step ii. Since road sensors may be farther away from each other than *d* meters, step v is skipped.

2.3. Road segments

Our aim is to define the lengths of the road segments that correspond to the road sensors. Each main route is considered as a straight, one-dimensional, line. An example is given in Figure 2a. This highway consists of one exit ramp, one entrance ramp, and five road sensors.

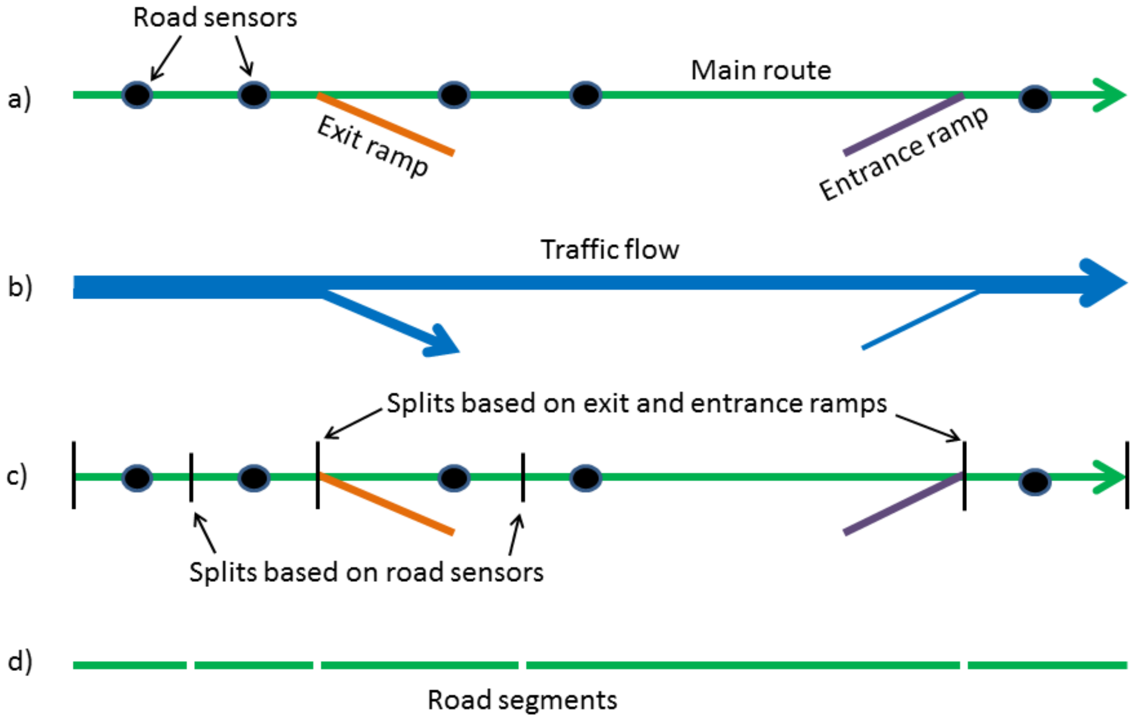

Figure 2. Deriving road segments: a) simplified example of a highway, b) addition of splits, c) derived road segments.

Observe that the traffic flow along this road only depends on the exit and entrance ramps. This is illustrated in Figure 2b, where the traffic flow has three levels. Accordingly, the road is split into three segments. Each segment that contains more than one road sensor is subdivided in such a way that the middles are taken as split points. In our example, the first and the second road segment are split since each contains two road sensors (see Figure 2c). It also may occur that a part does not contain any road sensors. In that case, the statistical outcomes for that part are imputed. Hence, the five resulting road segments are depicted in Figure 2d. The lengths of these road segments correspond approximately to the real length in metres.

In order to project the road sensors to the original main routes, which is needed prior to the analysis above, a series of main route points are created, one at every 10 meters. Each road sensor is projected at the closest 10-meter-point. The points at which the exit and entrance ramps respectively diverge and converge (points of bifurcation) are determined in the same way.

3. RESULTS

First results show that the road network is nicely segmented based on the location of the road sensors. However, on some roads, no loops were found between two ramps, leading to a not completely covered network. Therefore, it is necessary to impute the statistical outcomes regarding the uncovered road segments.

Furthermore, also loops were found on roads that were not completely included in the road network shape file. These roads segments of these roads have been constructed with the locations of the road sensors as described as in section 2.3. However, the quality of these road segments will have to be assessed.

Finally, the total lengths of the road segments per highway have been compared to official figures [3]. The algorithm to determine the main routes, as described as in section 2.2, appears to be accurate at first sight for most highways. However, for some highways, the found differences in road length were considerably large. This problem is probably due to inconsistences in highway definitions or reference periods.

4. CONCLUDING REMARKS

The method that is described in this paper can directly be used to obtain the lengths of the road segments that the road sensors represent. Therefore, it is possible to translate the processed vehicle count data that the road sensors generate, to traffic statistics, such as vehicle-kilometres, traffic intensities and traffic densities. Furthermore, it is possible to produce regional numbers, for instance based on the NUTS (Nomenclature of territorial units for statistics) [4] classification.

Obviously, the processing of the road network only needs to be executed when there are changes in the road network, or periodically, say every quarter of a year. As for the second part of the method, the definition of the road segments, it is important to know whether the road sensors function properly. If a road sensor does not work, the road segments need to be redefined.

It is important to realise that highways are not independent of each other. There are many interchanges between them, especially in the Netherlands. By the same reasoning as in section 2.3, the traffic flow on the main routes within the interchange is lower than on the main routes outside the interchange, since part of the traffic flows to the interchanging highway at the start of the interchange, and new traffic enters the main route from the interchanging highway at the end of the interchange. These middle parts can also be regarded as road segments. Analogue to the methodology described in section 2.3, the statistical outcomes of these segments can be imputed if road sensors are missing.

REFERENCES

[1] M. Tennekes, tmap: Themeatic Maps. R package version 0.6. http://CRAN.R-project.org/package=tmap (2014)

[2] H. Wickham, ggplot2: elegant graphics for data analysis. Springer New York (2009)

[3] Rijkswaterstaat, Actuele Wegenlijst RWS per 1 juli 2013 (2013) https://nis.rijkswaterstaat.nl/portalcontent/logon/p2_32.html

[4] European Commission, NUTS - Nomenclature of territorial units for statistics (2012) http://epp.eurostat.ec.europa.eu/portal/page/portal/nuts_nomenclature/introduction