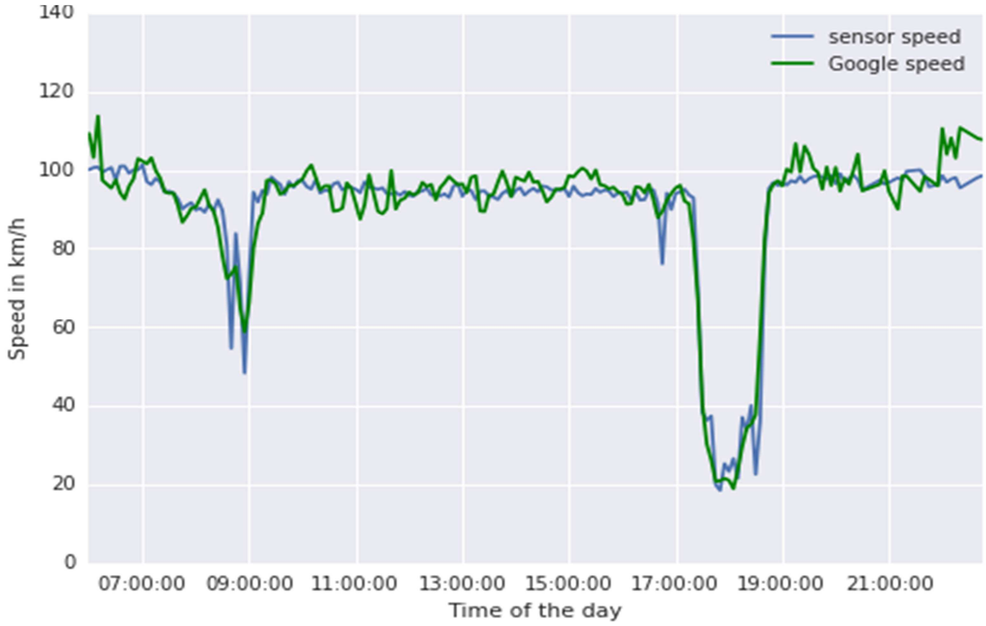
Validation of Google floating car data for applications in traffic management

Executive Summary

Authors: W.P. van den Haak MSc, M.F. Emde MSc

The whitepaper describes a study by TNO in cooperation with Google to validate the use of Google traffic statistics for the purpose of traffic management. The Google data, also called *floating car data* (FCD), originates from Google-enabled devices carried by persons travelling in vehicles on the road. For this study, we have compared Google floating car data with ground truth data coming from the dense infrastructural sensor network that is installed and used for traffic management in the major and minor roads in The Netherlands. These sensors measure traffic flow and average traffic speed. The whitepaper considers speed as the main indicator for traffic states on roads, but in further studies we also look at travel time on major arterial roads. Using the study results, we will also look at the potential use of Google FCD data as a way to augment infrastructural sensor data so that investments for road sensors can be lowered because less road sensors are required per length of road.

We have defined, implemented and executed a method to validate the quality of Google's anonymized traffic indicators by comparing them to a ground truth dataset (2253 sensors, 26 million data points). The main question here is can Google statistics describe real traffic conditions on the road accurately and reliably? The figure below shows the sensor speed measurements compared with the estimated speed based on Google statistics on the A10 highway for one day.



The main outcomes of the study are:

- Google floating car data can be adequately used for analyzing traffic management scenarios and for informing and signaling users traveling on roads. The quality of traffic state indicators derived from Google data is however not sufficient to be used for operational traffic management., e.g. for controlling traffic lights.
- Road coverage of Google floating car data is appropriate as long as traffic intensity is high enough (higher than one vehicle per 2 minutes), In situations roads are busy and where traffic management is most important, coverage of Google floating car data is good.
- Replacing part of all road sensors with Google data has limited and acceptable impact on quality. For example on the A12 we found that replacing 20% of the sensors with Google data diminishes data quality by a mere 3% on accuracy.
- Replacing part of all road sensors with Google data has the potential to lead to substantial cost reductions. For instance, replacing 20% of all sensors with Google data over a 30km stretch of highway in The Netherlands could reduce costs by 364 thousand Euros over the (typical) lifespan of the road. In a greenfield situation where 20% of sensors would not be installed in the first place, 700 thousand Euros could be saved of the lifespan of the road.

In more detail, in our study, we categorize the quality according to three quality levels, A, B and C, where A is the highest level. The levels correspond to the quality required for specific traffic management tasks. C is a quality level that is sufficient for informing (e.g. route advice, expected travel times) and signaling (e.g. dynamic maximum speed) end users. B is a higher quality level sufficient for selecting and adapting traffic management scenarios for network management. In our study, the highest level we consider is A. It corresponds to a level that is sufficient for all tactical traffic management applications as well as some basic operational traffic management applications (although operational traffic management applications also have additional requirements like timeliness which is not taken into account in this study).

In the study, we investigated the following important criteria and found the following results:

- <u>Similarity</u>:
  What is the precision of estimates for traffic state indicators based on Google floating car data (in this case speed) as compared to ground truth data?
  <u>Result</u>:
  Similarity for speed is within level C for 92% of all locations, within level B for 83%. However, only 21% of all locations reach level A. This means, that for most locations, Google FCD data can be used for analyzing traffic management scenarios and for informing and signaling users on the road. Google FCD data reaches level A for only a smaller minority of all locations.
- <u>Reliability:</u>
  What proportion of traffic state estimations based on Google floating car data (in this case speed) per location falls within the category A, B and C?
  <u>Result:</u>
  The proportion of estimates per location that scores level A is around 60% which means that for this level the Google data is unreliable. The Google data is considered reliable for level B and C, since the proportion of the estimates per location that reach the quality level is around 90% for level B and reaches 100% for level C.
- <u>Coverage</u>:
  Coverage in time: at all locations, during which periods of time can Google data be used for determining traffic state indicators to what extent. I.e. for each location, can Google data be used all of the time, or only part of the time (e.g. all day or only during peak hours)
  Coverage in space: over all times, for what part of the road network are traffic state indicators based on Google data adequate for traffic management?

Result:
Coverage in space and time is good at locations and at times where flow is higher than one vehicle per two minutes. At locations where flow is lower, estimates from Google data have too low quality and cannot be used for traffic management. However, typically traffic management is most important when the roads are busy and flow is high. At these times, coverage is good.

- Sensitivity:
  To what extent can we maintain the quality of estimated traffic state indicators when we lower the number of infrastructural sensors per length of road (i.e. stretching the distance between sensors) by augmenting road sensor data with Google floating car data? As ground truth comparison is not possible for non-existing sensors (that have virtually been removed), in order to investigate sensitivity, we studied the extent to which augmented road sensor data can be used to detect traffic jams.
  Result:
  Replacing all sensor data with Google data leads to errors . Replacing parts of the sensor data with Google data yields lower errors than replacing all sensor data with Google data. For example: on the A12 we found that replacing 20% of the sensors with Google data has limited and acceptable effects on quality.
- Costs:
  What costs (operational costs and capital investments) can be prevented when data from a (less-than-usual densely populated) road sensor network is combined with Google data?
  Result:
  Replacing 20% of all sensors with Google data on a 3 lane highway in the Netherlands in both driving directions over a 30 km stretch reduces operational costs and capital investments of the sensor network. The operational cost reduction is 41 thousand Euros per year. Considering a time span of 12 years (average lifetime of a highway) this reduction would be 364 thousand Euros.
  Sensors are being replaced after pavement milling, when the lifetime of a highway is reached. If less sensors would be (re)installed the reduction on capital investments and operational costs would be 700 thousand Euros over a period of 12 years. Scaling up these findings to the entire Dutch highway network (6500 km) the reduction could be 161 million Euros over the lifetime of highways.

The traffic indicator speed based on Google FCD is reliably similar as the infrastructure sensor data for level B and C for a vast majority of locations and can therefore be used for traffic management applications that require quality level B and C respectively. Parts of the sensor network can be replaced by Google FCD with an acceptable decrease in quality. The results from this study show the applicability of Google FCD for the purpose of traffic management. If parties embrace this new data source for traffic management, effort is being needed in integration with traffic control centers as well as the provisioning of an (almost) real time data stream.