Traffic loop data for transport statistics

Peter Struijs

International Conference on Big Data for Official Statistics, Beijing, China, 28 – 30 October 2014
Topics covered

➢ Characteristics of the data source
➢ Issues when using traffic loop data
➢ Solutions to issues
➢ Results
The main roads
Road sensors

Road sensor (traffic loop) data

- Each minute (24/7) the number of passing vehicles is counted in around 20,000 highway ‘loops’ for different length classes

- No identification of vehicles

- Big Data: around 230 million records a day
Sensors in a road segment

- Road sensor

Road segment:
- Main route
- Exit ramp
- Entrance ramp
- Other
A special dike
Road sensors in the dike
Minute data of one sensor for 196 days
Researching the data

Cross correlation between sensor pairs
- Used to validate metadata

Trajectory speed vs. point speed
- Average speed is 98 Km/h
Small, medium-sized & large vehicles

![Graph showing the total number of vehicles detected over time for different vehicle sizes. The x-axis represents time in hours, ranging from 0 to 20. The y-axis represents the total number of vehicles detected, normalized. There are three lines: black for small vehicles, red for medium vehicles, and green for large vehicles. The graph shows peaks and troughs for each type of vehicle at different times, indicating variations in vehicle detection.]
Issues and non-issues

Non-issues:
- Privacy
- Data acquisition

Issues:
- Methodology
  - Selectivity
  - Quality
- Infrastructural needs
- Other issues
  - Skills needed
Data options

Historical database
- Request data via web interface
- Minute data for all highways
  - 48 variables, around 2.5 TB (Jan 2010-April 2014)
  - Data at a higher aggregation level is edited

Data stream
- Every minute, all data for **all** active sensors
- Has to be continuously collected
Process of road sensor based statistics

- Select sensors on Dutch highways
- Preprocessing
  - Remove non-informative variables
  - Remove bad records
  - Calculate number of vehicles (per minute)
  - Quality indicators for daily data per sensor
- Dimension reduction of daily data
  - Exclude bad sensors
  - Reduce dimensions on same road and region
  - Obtain number of vehicles for each road and region
- Calculate traffic index
  - Calculate indices per region
Conclusions

- Invest in methodological research and play with the data to get a grip on quality
- Traffic loop data are an ideal Big Data source
- Develop good relations with the data provider