Data-driven modeling of transportation systems and traffic data analysis during a major power outage in the Netherlands
Melnikov, V.R.; Krzhizhanovskaya, V.; Boukhanovsky, A.V.; Sloot, P.M.A.

Published in:
Procedia Computer Science

DOI:
10.1016/j.procs.2015.11.039

Citation for published version (APA):
Melnikov, V. R., Krzhizhanovskaya, V. V., Boukhanovsky, A. V., & Sloot, P. M. A. (2015). Data-driven modeling of transportation systems and traffic data analysis during a major power outage in the Netherlands. Procedia Computer Science, 66, 336-345. DOI: 10.1016/j.procs.2015.11.039
Data-driven modeling of transportation systems and traffic data analysis during a major power outage in the Netherlands

Valentin R. Melnikov¹, Valeria V. Krzhizhanovskaya¹,²,³, Alexander V. Boukanovsky⁴, Peter M.A. Sloot¹,²,⁵

¹ ITMO University, Saint-Petersburg, Russia
² University of Amsterdam, The Netherlands
³ St. Petersburg State Polytechnic University, Russia
⁴ Netherlands Institute for Advanced Study, The Netherlands
⁵ Nanyang Technological University, Singapore

mail@valmelnikov.ru, V.Krzhizhanovskaya@uva.nl, avb_mail@mail.ru, P.M.A.Sloot@uva.nl

Abstract
Efficient methods and tools for road network planning and traffic management are critically important in the ever more urbanized world. The goal of our research is the development of a data-driven multiscale modeling approach for accurate simulation of road traffic in real-life transportation networks, with applications in real-time decision support systems and urban planning. This paper reviews the multiscale traffic models, describes the traffic sensor data collected from 25000 sensors along the arterial roads in the Netherlands, and discusses the applicability of sensor data to model calibration and validation on each modeling scale. We also present a road network graph model and the reconstructed Dutch road network. Finally, we show the results of traffic data analysis during the major power outage in North Holland on 27 March 2015, paying special attention to one of the most affected locations around the A9/E19 interchange near Amsterdam airport Schiphol.

Keywords: transportation systems, data-driven modeling, complex networks, traffic flow, multiscale modeling, traffic sensor data, power outage

1 Introduction
Transportation systems are the blood vessels of modern civilization. To cure the chronic thrombosis of these vessels, significant efforts are put into development of efficient methods and tools for road network planning and traffic management. The goal of our research is the development of a data-driven multiscale modeling approach for accurate simulation of road traffic in real-life transportation networks, with applications in real-time decision support systems and urban planning.
Transportation systems consist of many intertwined components: road infrastructure (roads, lanes, intersections, traffic lights) with their technical subsystems; vehicles (cars, buses, trains) with their complex motors and control systems; and - most importantly - humans (drivers, passengers, traffic controllers) with their inexplicably complex brains and sometimes unpredictable behavior. On top of that, many other factors influence the traffic: weather, social events and even fashion. Interacting, these components form a large-scale, integrated, open, complex system. Many components and their interactions cannot be adequately described by a mathematical equation or simple agent, for instance vehicles interaction, drivers' behavior and other human-imposed events [1]. Advanced models and computational techniques are required to understand the nature of traffic flow and to solve the challenging transportation problems of prediction, monitoring, managing and planning.

All traffic flow models are usually divided into three levels representing the underlying processes: macro, meso and micro scales [2]. Some researchers also consider the fourth scale: the submicroscopic (picoscopic) level [3]. Accurate prediction of the traffic flow model parameters requires data-driven modeling. The main principle of data-driven modeling is to calibrate (train) the models in order to minimize the model error [4], as schematically shown in Figure 1.

Obtaining complete and reliable empirical data for model validation is an everlasting problem for computer modelers [5]. The Dutch National Data Warehouse (NDW) for traffic information [6] recently presented a unique opportunity by providing traffic data from thousands of sensors installed along the arterial roads throughout the Netherlands: vehicles speed, traffic intensity and travel time.

The first steps of our work reported in this paper included the study of data completeness, accuracy, road network coverage, accessibility, retrieval, format, and API. We then investigated the existing models of traffic flow and road networks, and found ways to apply the NDW data for model calibration and validation. Finally, we analyzed a set of sensor data and identified the data describing "normal" road conditions and some anomalies. The most interesting recent example of a critical event was the “Blackout Friday” on 27 March 2015, when a massive power outage in North Holland disrupted electricity-driven public transportation (train, metro and tram) for hours, thus forcing people to travel by car. In this paper we report our findings of the blackout impact on road traffic performance.

The rest of the paper is organized as follows: In section 2 we describe the NDW sensor data. Section 3 reviews the multiscale traffic models and discusses the applicability of the NDW data on each scale. In Section 4 we present the road network graph model and the results of constructing the Dutch road network. Section 5 shows the results of NDW data analysis during the “Blackout Friday” on 27 March 2015. With Section 6 we conclude the paper and propose future research directions.

2 NDW sensor data description

The Netherlands is a country with a well-developed and dense road network that faces a strong traffic growth. In 2008, the Dutch National Data Warehouse (NDW) was established for collecting and providing road traffic data [6]. In 2015, some data become open for public research. Recent results from complexity science indicate that the Dutch road highway network has an optimal topology [7] that provides efficient emergency exit opportunities in case of severe flooding [8].

The NDW collects traffic performance characteristics (speed, intensity and travel time) with a 1-minute interval from more than 25 thousand sensors located along the arterial roads throughout the Netherlands. The NDW sensors also distinguish vehicles by their length, as shown in Figure 2.
The NDW sensor data is provided in two ways:
1. real-time data can be downloaded from a public FTP server in XML files refreshed every 1 minute;
2. historical data is provided by a web service (available after registration), with an option of data filtering and aggregation (by time and location).

Information about each sensor characteristics (location, data measured, accuracy) is stored in a separate file. In addition to the absolute geographical coordinates, the sensors are also attributed to the nodes of the road network database VILD [9] (see Section 4.2).

3 Multiscale traffic flow modeling

There is a big variety of traffic flow models that differ by the underlying principles used to reproduce the traffic flow behavior. Depending on the level of detail, these models are categorized by three scales: macro-, meso- and microscale [2]. Different-level models need different traffic data to be calibrated. Below we briefly describe the modeling scales and discuss the suitability of the NDW sensor data for model validation and calibration.

3.1 Macroscale models

Being the least detailed, macroscale models reproduce only the integral traffic flow performance characteristics [10]:
- traffic density $\rho$ — the number of vehicles on the observed part of the road at instant moment of time (instantaneous characteristic);
- traffic flow (or intensity) $q$ — the number of vehicles that passed the control point during the period of observation;
- average speed (velocity) $v$ — vehicles speed averaged over time or space.

These three flow characteristics are related via the fundamental equation of traffic flow modeling, the continuity equation:

$$q = \rho \cdot v \quad (1)$$

The kinematic wave model of traffic flow theory described by the equation of conservation of vehicles (2) is expressing the principle of balance: the number of vehicles within an elementary road segment changes according to the difference of inflow and outflow on its boundaries.

$$\frac{\partial \rho}{\partial t} + \frac{\partial q}{\partial x} = 0 \quad (2)$$

The system of two equations for the three unknowns needs the third equation to describe the traffic flow dynamics. Existing macroscale models propose different equations to deal with this problem, for
instance [11] proposed a velocity dynamics description. Further developments of this model led to a
class of Payne-type models, for example [12]. One of the extensions of Payne model is the Helbing's
model [13], where an additional partial differential equation for velocity variance is introduced.
Operating with the average characteristics, macromodels benefit best from the NDW traffic data
applied to model parameter calibration. In our study, we plan to combine the benefits of the
macroscale with the level of detail of the microscale.

3.2 Mesoscale models

At the intermediate level of description, the mesoscale models use the probability distribution
function to describe each vehicle velocity at each instant moment. These models can be divided into
three main groups: headway distribution models, cluster models, and gas-kinetic models [10]. The
most advanced models of the first group include individual distributions for different classes of
vehicles and road lanes. Cluster models group vehicles into clusters with a dynamically changing
size and velocity. These clusters usually appear in some road conditions (speed limits, overtaking
prohibition, poor weather conditions), when overtaking is limited. Gas-kinetic models introduce
dynamic velocity distribution functions of vehicles. The first model of this class was proposed by
Prigogine and Herman [14] and later extended, resulting in a number of complex models taking into
account different aspects of vehicle interactions and traffic flow dynamics [15], [13]. That provided
more realistic results in reproducing main flow characteristics. Theoretical developments of mesoscale
models became a starting point for many macromodels.

For validation and calibration of the mesoscale models, a detailed car-tracking data needs to be
collected. Currently, it is not feasible in a large-scale transportation system such as the complete road
network of the Netherlands. In our work, we skip this modeling level in favor of the more general
macroscale level and a more detailed microscale.

3.3 Microscale models

On the microscale level, the behavior of each vehicle is simulated. The most developed class of
micromodels is the car-following approach [16]. It solves an equation of distance dynamics between
the preceding and following vehicle depending on several factors: velocity of the preceding vehicle,
vehicle length, desired velocity, driver personal traits, characteristics of vehicles, etc. With growing
computational power, microscale modeling of millions of vehicles becomes possible, which makes the
algorithmically simple micromodels very attractive.

Detailed agent-based models require "personalized" data about the vehicle dynamics, driving
routes, driver behavior (careful, aggressive, or hasty) and even driver's physical condition (tired or
drunk). Only small-scale experiments of data collection are conducted by car-industry research groups,
looking at the vehicle dynamics and driving routes. Unfortunately, such data is not available to us,
partly due to the privacy issues. Monitoring the psycho-medical driver's condition is at its infancy yet.
Therefore today we have to guess the microscale model parameters (using previous theoretical
investigations), but we can calibrate the model by comparing the integral simulation results (average
speed, intensity, travel time) with the macroscale empirical data. The NDW macrodata are very
valuable from this point of view. In our work we aim to use a popular MatSim simulation package for
microscale agent-based simulation.

3.4 Multiscale modeling and the corresponding NDW sensor data

Table 1 summarizes the underlying principles and data necessary for each scale of the models. The
last column indicates whether the sensor data from NDW is applicable to run models of the particular
level. The NDW data can be used as input data for the macroscale models and as validation data to be
compared with micro- and mesoscale modeling results.
Table 1: Scales of traffic flow models

<table>
<thead>
<tr>
<th>Scale</th>
<th>Underlying principle</th>
<th>Model parameters</th>
<th>NDW data used for</th>
</tr>
</thead>
<tbody>
<tr>
<td>Macro</td>
<td>System of partial differential equations: fluid dynamics, gas kinetics</td>
<td>Average flow characteristics (speed, intensity, density, travel time)</td>
<td>Model input parameters</td>
</tr>
<tr>
<td>Meso</td>
<td>Gas kinetics</td>
<td>Distributions of velocities and headway of vehicles</td>
<td>Model validation by comparing the averaged model output</td>
</tr>
<tr>
<td>Micro</td>
<td>Car following, agent based</td>
<td>Individual vehicle behavior characteristics (acceleration, gaps, lane changing, etc.)</td>
<td></td>
</tr>
</tbody>
</table>

All traffic flow models, no matter what the scale is, need the road network to run on. Next section discusses two road network models and shows the results of graph construction of the Dutch road network.

4 Road network modeling

Building a road complex network model from scratch is a laborious process. We built the road network of the Netherlands from OpenStreetMap (OSM) data. Each road in OSM is a sequence of points, which is hard to use for traffic modeling. Transformation of OSM data to the acceptable format is a non-trivial process [17], [18]. The big advantage of NDW is that their sensors are linked to the nodes of the Dutch traffic information location database VILD [9].

In this work we consider a classical graph model of the road network. A more advanced model based on Markov chains [19] employs turning probabilities in each intersection, thus reflecting the stochastic nature of traffic flow at the micro and mesoscale modeling levels. Deriving the turning probabilities from NDW data would allow building the Markov chain model of the road network.

4.1 Graph model of a road network

A classical approach to describe road networks is a graph [20]. Road graph is a bidirectional weighted graph, where nodes are the road intersections/junctions and edges are the parts of the roads between intersections. Edge weight is the segment length or travel time. Figure 3 shows a graphical representation of a small road network and the adjacency matrix describing the graph connectivity. From this matrix, the main road network characteristics can be calculated: degree distribution, clustering, path length, connection patterns, diameter and network robustness [21], [22].

Figure 3: (a) Example of a road network. (b) Adjacency matrix describing this road graph. The matrix may be non-symmetric for one-way roads or weights based on the travel time or number of lanes.
4.2 Dutch roads graph construction

To construct the road graph, the following data is needed: nodes location, network topology (nodes interconnections) and length of the edges (road parts). This information was extracted from the VILD database [9]. It consists of two shape files (Esri SHP): one with points reproducing the nodes and another one with lines reflecting the geometry of roads. Points from the first shape file are interconnected, forming a network, and linked to the road lines from the second shape file.

The edge length values extracted from the VILD database appeared partially incorrect (length of some road segments was shorter than the straight-line distance between the two points). We are currently looking into the possible causes of this misfit. To calculate the edge length, Euclidian distance between the node points would give too coarse an estimate of the real road length. Instead, we used a more accurate method of calculating the curve length between the node points, using the Lines layer in the SHP file. Since node points do not coincide with the starting or ending points of the curves, and often lay outside these curves, some approximation and interpolation is required, with projection of the node point to the road curve.

The result of reconstruction of the road graph in Amsterdam metropolitan area covered by the NDW traffic sensors is shown in Figure 4. It is important to note that only arterial (A and N type) roads are covered by the NDW sensors. For a detailed traffic simulation inside the cities, we need to develop a modeling approach that will take the known sensor data from the equipped roads and "propagate" this information along the smaller roads not covered by sensors. The combination of micro- and macroscale modeling seems the best solution in the conditions of incomplete information.

![Figure 4: Visualization of the reconstructed graph of arterial roads covered by the NDW traffic sensors. Amsterdam metropolitan area.](image)

5 Traffic data analysis during the major power outage in the Netherlands

We analyzed a large volume of the NDW traffic data and found anomalies standing out from the normal traffic patterns. The most striking anomaly was recently observed on a “Blackout Friday”, 27 March 2015, when a massive power outage in North Holland disrupted electricity-driven public transportation for several hours and caused severe delays in Schiphol airport departures [23]. We put forward the hypothesis that this critical event will have the following effect on traffic performance: (1) road traffic flow (intensity) shall increase, because trains and trams were stuck for hours, thus forcing
people to take a car; (2) traffic speed shall decrease due to congestions in oversaturated roads and due to the failure of regulatory infrastructure (traffic lights) in the affected area. Let's test this hypothesis.

Analyzing the number of operating sensors in Amsterdam metropolitan area, we determined the time bounds of the blackout. Figure 5 shows that the cascading power failure started around 8:30, peaked at 9:30 and recovered by 12:00. These findings are confirmed by an official report of the Dutch government [24]. A sharp drop at 12:00 is common for all days and can be explained by a scheduled system backup or reboot.

![Figure 5](image_url)

**Figure 5:** Number of operating sensors during the Blackout Friday (major power outage on 27 March 2015) in Amsterdam metropolitan area, compared to a "normal" Friday a week earlier. Data aggregated over 5 minutes.

Next, we analyzed the traffic flow and speed data during the Blackout Friday and several "normal" Fridays before and after the event. To avoid possible interpolation errors, we relied only on data from the sensors that stayed operational during the blackout. That means that we did not "see" the consequences of the power failure in the spots where sensors were powered down for an extended period of time (longer than the emergency batteries hold). Instead, we observed the traffic anomalies down the road. With this partial information, one can accurately reconstruct a complete picture by running an agent-based microscale model or by solving equations of the meso- or macroscale models.

Figure 6a shows that traffic flow during the blackout was higher than normal almost everywhere in the arterial road network. At the same time, traffic speed decreased only in the south of Amsterdam, where the power failure actually took place (see Figure 6b). The most affected junctions of the roads are marked by red boxes in Figure 6b. Note that no data is available along some roads adjacent to these critical junctions (shown by grey points), especially in and around box 3, in the vicinity of the town of Duivendrecht. Animation of the flow and speed anomaly dynamics demonstrated the propagation of traffic jams and illustrated the power failure spread and recovery over time (see supplementary materials online).

Figure 7 shows the dynamics of traffic speed and flow intensity during the Blackout Friday and three normal Fridays for comparison, in one of the most affected parts of the road network, the A9/E19 interchange near Amsterdam airport Schiphol. A 40% drop in traffic speed lasted over 2 hours (see Figure 7a), whereas flow intensity got higher than normal only after an initial decrease before 10:30 (see Figure 7b). This initial reduction is obviously linked to the sharp velocity drop, but that alone cannot explain this effect, since the roads were not saturated to their maximum capacity at that time. Most likely another factor played a role: the failure of traffic lights and electronic road signs in the towns along the highways definitely caused a lower car influx in the beginning of the blackout. An intermediate recovery of the traffic flow for 30 minutes (9:00-9:30) can be explained by the temporary power recovery: In many places, electricity was going on and off several times before the final recovery of the entire power grid around 12:00. As we see in Figure 7a, traffic speed near Schiphol airport recovered to normal around 11:00, but traffic intensity (Figure 7b) stayed higher than normal till the end of the day, which is logical: in the morning many people had to commute to work by car instead of a train or metro, thus in the evening they all had to return by car as well.
Figure 7: Traffic speed and flow intensity dynamics during the Blackout Friday compared to three "normal" Fridays at the intersection near Amsterdam airport Schiphol (A9/E19 interchange).

Each data point is the average from 21 sensors located in box 2 in Figure 6.
The results of the blackout impact analysis generally proved our hypothesis, except of the initial traffic flow decrease explained by the abrupt speed reduction and by the disrupted inflow from the hindered traffic in the towns.

6 Conclusions and future work

The goal of our project is the development of a data-driven multiscale modeling approach for simulation of road traffic in real-life transportation networks, with applications in real-time decision support systems and urban planning. In this paper we reviewed the multiscale traffic models, described the traffic sensor data collected from over 25 thousand NDW sensors along the arterial roads in the Netherlands, and discussed the applicability of sensor data to model calibration and validation on each modeling scale. NDW traffic data is highly valuable for defining the macroscale model parameters. It can also be used for reverse-engineering and calibrating the micromodel parameters by comparing their integral simulation results with the sensor data.

Further, we presented the road network graph model and the Dutch road network reconstructed from the VILD database. An alternative approach that could be tried in the future is the Markov chain model, which requires a smart method of deriving the turning probabilities from available data. The constructed road network graph will be used in microscale agent-based simulation and macroscale transportation modeling. First, we plan to test the MatSim agent-based transport modeling package.

The sensors we analyzed were located only along the arterial (A and N type) roads. Later we found that another type of sensors is available in smaller streets within some cities. Investigating new data and developing the methods to use the different data types in one model is the subject of future work. We will need to develop a modeling approach that will take the known sensor data from the equipped roads and "propagate" this information into the other roads not covered by this type of sensors. The combination of micro- and macroscale modeling seems to be the best solution in the conditions of incomplete information.

Finally, we showed the results of traffic data analysis during the major power outage in North Holland on 27 March 2015 that lasted around 3.5 hours (8:30-12:00), paying special attention to one of the most affected locations around the A9/E19 interchange near Amsterdam airport Schiphol. This critical event had a profound impact on the transportation system: the electricity-driven public transportation (train, metro, tram) was disrupted for hours, thus forcing people to travel by car. To make things worse, the regulatory infrastructure (traffic lights and electronic road signs) were switched off in many areas, causing some chaos and congestion.

We observe that: (1) The road traffic flow (intensity) first dropped down, apparently because of the disabled traffic lights and local congestions. After the initial decrease, the flow recovered (in some places going through a series of ups and downs) and remained higher than normal till the end of the day in the whole road network. (2) Traffic speed decreased by 40% in the south of Amsterdam metropolitan area, where the power failure took place. It can be explained by the failure of regulatory infrastructure and later by congestions in oversaturated roads, when thousands of citizens had to commute by car.

These results are the first step in a long-term project. The next step in our research is the data-driven scenario-based modeling of critical events [25] on micro- and macrolevels, to understand and control the causes and consequences. After validation of the models on available historical sensor data (both normal and critical conditions), we can use the real-time data for early detection of the tell-tales of critical events [26], for situation prediction and for individualized route planners that would save time of the particular driver, while also alleviating the symptoms of the global road congestion. Real-time traffic flow analysis will be further developed to monitor and control the situations influencing the global transportation systems.

Acknowledgements. This work is supported by the Russian Science Foundation project #14-21-00137 and the Dutch Science Foundation NWO project #629.002.101.
References