Data-driven traffic flow:
A state-of-the-art report

TENS-PROJECT

Ian Marsh (RISE AB)
Pei-Lun Hsu, Cosar Ghandeharioon (KTH)

RISE report 2021:93
ISBN 978-91-89385-83-2

English abstract

This state-of-the-art report is the *first* of two reports. Both are the result of a Swedish project, the Swedish Energy Authority's TENS project, which spanned from 2018-2021. This report lays out the background, related work, state of the art, related projects as well as mapping out the area for the newcomer to the area. Examples are from Sweden due to the available data, as well as the locally installed infrastructure. We dedicate some space to the data gathering process as well as basic traffic flow principles. It should be seen as introductory material to the topic, getting the laymen up to speed in data processing issues arising from using traffic sensors. A second (sister) report contains technical results from our work in the above named Swedish project.

Keywords: Transport ITS, data-driven, traffic flow.

Svenska sammanfattning

Denna spets teknologi rapport är den första av två rapporter. Båda är resultatet av ett svenskt projekt, Energimyndighetens TENS-projekt, som sträckte sig från 2018-2021. Denna rapport beskriver bakgrunden, relaterat arbete, spets teknologi, relaterade projekt samt kartlägger området för nykomlingen till området. Exempel är från Sverige på grund av tillgänglig data, samt lokalt installerad infrastruktur. Vi ägnar lite utrymme åt datainsamling processen samt grundläggande trafikflöde principer. Det ska ses som ett introduktionsmaterial till ämnet, för att få lekmännen på plats i databehandling frågor som uppstår vid användning av trafik sensorer. En andra (syster)rapport innehåller tekniska resultat från vårt arbete i ovannämnda svenska projekt.

Nyckelord: Transport ITS, datadriven, trafikflöde.

The TENS project
Energimyndigheten
Diarienummer: 2018-006615
Projektnummer: 46963-1

0 Glossary

Term	Description Units	Range Typical
Capacity	Maximum throughput at a particular road section.	-
Demand	Amount of traffic that wants to use a particular road section.	Typically, 1 hour.
Volume	Units of traffic Typically per lane.	Typically, measured in 15 min intervals.
Density	Number of vehicles in a specified area / region. Vehicles per km per lane.	Typical: 20 vehicles per hour per lane *can* lead to overall speed reductions.
Flow	Rate of vehicles per hour that cross a segment or boundary.	Range: 0-8000. Typical: 5600
Velocity	Velocity of individual/group of vehicles in kilometers per hour.	Kilometers per hour. 0-150 km per hour.
Macroscopic flow	Traffic flow of a group (aggregated) vehicles.	-
MCS	Motorway Control System used to measure traffic flow in and around cities.	-
Microscopic flow	Flow of individual (identifiable) vehicles.	-
Peak hour factor (PHF)	Variability of volumes (flow) between 15-min periods.	Typically, in the range of 0.7 to 0.9
Floating (car) data	Raw floating car data provides the geolocation of a vehicle at a certain point in time. Speed and direction of the vehicle can be extrapolated using multiple points of data.	-
Time headway	Microscopic expression of flow.	Time over unit of traffic. Typically seconds per vehicle (s/veh) 3600/h_{avg}

INRIX	Commercial offering of traffic flow deduced from floating car data.	-
Time-mean speed	An observer measures instantaneous speeds at a particular location and obtains the average of those instantaneous speeds at a particular location.	See below.
Space-mean speed	Based on travel times. The harmonic mean of the speeds.	See below.
Travel time	The time to travel from a given origin to a given destination. $tt(h) = (d_{dest} - d_{orig})/ v_{avg}$	The reciprocal of speed.

1 Introduction

Predicting traffic flow congestion is extremely beneficial to society to lower journey times, reduce CO_2, and improve safety. With the increased number of rich traffic datasets, processing platforms plus innovative algorithms, it is now possible to help alleviate traffic congestion. Within Sweden, there is a large state-based program with players from academia and industry to leverage new technologies. Traffic congestion accounts for increased commuting time, pollution, decreased safety, soaring transport costs and increased demands on logistics.

We list some examples of the impact of traffic congestion to name just a few. **i)** An average of nine days/year are spent in jams in the United States. **ii)** Seven billion hours were lost in the UK due to congestion.[1] **iii)** Many cities now ban old diesel vehicles, partly because combustion exhausts are responsible for 5000 premature deaths annually across the UK compared with 1850 in road accidents. iv) Kofi Annan brought road safety to the general assembly in 2004, stating that 1.3 million people are killed annually on the worlds' roads, more than HIV, malaria, and tuberculosis combined. Compared to the seismic cost of traffic problems, relatively little is spent on accident prevention. v) An OECD reports costs of infrastructure, and in many cases, the costs are increasing significantly, see also [*ITF*]. vi) Logistic companies need to know where queues occur, and if they are entering them. A consignment or delivery comprises the contents of several truck loads and estimating the time at a delivery point or port is very valuable to hauliers vii) Frustration and stress due to road congestion is cited as one of the major factors in modern living.

Intelligent traffic systems (ITS) can utilise the best of modern technology to help alleviate congestion, through measurements and traffic control.

2 Society and traffic flow

2.1 UN goals

Based on the Sustainable Development Goals (SDG) proposed by the United Nations (UN) in 2015 [*UN2021*], the TENS project contributes to solving at least four goals:

[1] https://inrix.com/press-releases/traffic-congestion-cost-uk-motorists-more-than-30-billion-in-2016/.

1. *Goal 3. Good Health and well-being*: By applying our flow estimation approach in ITS, we can improve ITS' abilities to monitor and mitigate traffic congestion and emissions. As we know that vehicular emission is one of the leading causes of air pollution, this thesis can indirectly reduce the health risks and deaths resulting from air pollution and congestion.

2. *Goal 9. Industry, Innovation, and Infrastructure:* This work applies information technology, i.e., machine learning, in ITS, which improves the effectiveness and efficiency of the transportation infrastructure innovatively.

3. *Goal 11 Sustainable Cities and Communities:* A city having frequent traffic congestion will suffer from many daily life problems, e.g., noise, long commute time, and bad air quality. Our approach can make cities and communities more livable and sustainable by improving their transportation systems.

4. *Goal 13 Climate Action:* Transportation plays an essential role in greenhouse gas emissions such as carbon dioxide, and greenhouse gas is the main factor causing global climate change. Since an important use-case for our approach is to monitor and control emissions on the roads, our work can help reduce greenhouse gas and fight climate change.

2.2 Stakeholders

Road traffic measurements are performed by many different stakeholders: government agencies, municipalities, researchers, commercial companies, and even crowdsourcing initiatives. Government agencies and municipalities often measure road traffic for the planning of future roads, for operation and maintenance, and for analysis of accident risks and environmental impact. The timescale of interest is often days, months, or years. But sometimes measurements are also collected and used in real-time for traffic control and incident detection. The measurement techniques often include inductive loops, radars, or pneumatic tubes.

The Swedish Transport Administration, "Vägverket", measures road traffic on the government-owned roads in Sweden. They measure for instance traffic volume, "trafikarbete", expressed in vehicle kilometers, to understand the usage of the road network. To monitor traffic trends in the country over time, the Swedish Transport Administration continuously measures traffic on about eighty road sections. These roads are selected to represent the entire state road network in Sweden. Complimentary sampling is also done on many other stretches of road. In practice, measurements are done for varying reasons, on different time scales using a variety of techniques, see [*Allström2017*, *Sharma2017*] for detailed surveys.

2.3 Cost

The cost of road monitoring systems is substantial. Either in procurement, installation, maintenance, personnel, data cost, gathering, processing and presentation. Typically, with floating (GPS or Bluetooth) traffic data is cheaper to gather and process. A subset of vehicles on the road must monitor their position and upload that information at regular intervals. Vehicle fleets, which cover many kilometers in a country, are popular choices for sampling the road traffic. Note one issue with crowdsourcing solutions, European data might leave the EU as the companies gathering from fleets or using apps are owned by US companies (see below).

2.4 Technology

New technology opens for new innovative ways of measuring traffic. This includes video and cameras, Bluetooth, and WiFi measurements [*Forsman2018*]. Future measurement methods may include the use of drones, connected cars using Vehicle-2-vehicle protocols, 5G side channel or communicating with the roadside infrastructure, sometimes called V2I. Techniques that use Bluetooth or WiFi[2], are common for Automatic Vehicle Identification (AVI). The same vehicle can in this way be identified at several locations, which can be used to calculate the travel time between locations. The Bluetooth or WiFi hardware addresses are captured by roadside equipment and then reused/reread to calculate their speed and direction when seen again.

3.0 The data gathering process

3.1 Potential data sources for observing traffic flow

Before we delve into the data available in the TENS project, it is important to point out traffic information is available from several sources, as the next table shows. In the interests of space, we refer the reader to [*French1986, Yatskiv2013*] for a review of the technologies below.

[2] Vehicles use a variant of WiFi known as IEEE 802.11p rather than the, b, g, a, x otherwise used.

Detector Type	Volume / Count	Speed	Occupancy
Inductive Loop Detector	Yes	Yes	Yes
Radar Detector	Yes	Yes	Yes
Active Infrared	Yes	Yes	No
Passive Infrared	Yes	Yes	Yes
Bluetooth	Yes	Yes	Yes
Floating car data	No	Yes	Yes
Video Image Processing	No	Yes	Yes
Satellite imaging	Yes	Yes	Yes

Table 1: A comparison of the data gathering process

Alternative sensor data like vehicle probe data is cheaper to collect, compared to traditional stationary sensors like radar and inductive loops. Also, with the new sensors, a much larger part of the road network can be monitored. But there is an important difference in what type of data the different methods can provide. Vehicle probe data can provide speed and travel time but typically not traffic flow data. If flow data is needed as input to algorithms for traffic control or other calculations, then there is a need to estimate the flow from speed or travel time data. Essentially a problem of estimation and inference, we discuss it in the second report [*Marsh2021TENSII*] and can be found in [*Kim2014*, *Blandin2012*].

3.2 Fixed infrastructure (macroscopic) measurements

The traffic flow in Stockholm is monitored with a Motorway Control System (MCS). Many stationary MCS portals (gantries) have been installed on the E4 plus other roads. The MCS-portals are equipped with radar detectors where they monitor the flow and speed in each lane of the road. The data gives the regional traffic control centre information about the current traffic flow and speeds. The data is also input to a control system that sets variable speed limit signs. The radar measurements are point measurements and give time-mean-speed. This contrasts with probe measurements such as INRIX calculate speed over a distance and provide space-mean speed. [*Nissan2010*] in his Ph.D. thesis provides a technical overview of the Stockholm Motorway Control System, including subsystems for Variable Speed Limits (VSL) and Automatic Incident Detection (AID).

As mentioned, this method uses stationary railing, scaffolding or gantries to hold cameras and radars. Well-known locations are used as measurement points to attain the speed and flow of traffic flowing across the chosen location. Often these points are chosen as strategic points where lanes merge or diverge and known congestion points occur. When large deployments are performed, the providers try to equally space the measurement locations. In Stockholm, there are approximately 2700 locations which are about 300 meters apart. Seen over a road section of 1-5 km this 300m resolution gives an accurate picture of the flows. The data gathering is performed per lane, i.e., multiple cameras are used.

Macroscopic or fixed infrastructure companies have traditionally been transportation-based, their competence is in planning, running, and managing road infrastructure.

3.3 Floating (microscopic) measurements

This method relies on a device in the vehicle that sends periodic measurements of its position and speed. These are then collated and presented at a measurement company. Usually, they are gathered by GPS receivers in selected vehicles. Vehicle fleets, as mentioned above, are common as they are active compared to normal commuters and cover a fair percentage of the road network.

Several commercial companies, including Google, Waze, Here, TomTom, INRIX collect and provide traffic information to drivers. In some cases, they deliver the road state to automotive companies, cities, and road authorities. As an example, Google Traffic relies on crowdsourcing from drivers to collect traffic information. Google collects GPS information from Android phones and calculates the speed of the users on the road. In 2007, Google started to offer live traffic information with Google maps. A color code is used to highlight the speed of traffic on a road as many users are aware of, green – free flowing, orange – congested, and red for stationary traffic.

Waze also relies on crowdsourcing from drivers to provide a real-time traffic service. They collect map and traffic information and allow users to report incidents on the road via a phone app [*WAZE*], Google bought Waze in 2013. Here Technologies is another example of a company that provides map, traffic, and location services to both companies and individuals [*HERE*]. INRIX provides traffic information to road authorities, cities, automotive industries, and individuals. They collect traffic data from many sources including road sensors, connected cars and mobile devices [*INRIX*]. The INRIX data is described below and forms one of the used datasets in our research work.

Also, telecom companies collect data about traffic and mobility patterns, thereby providing services based on the data. In Sweden, an example is the services offered by Telia Crowd Insights [*Telia*]. They do however gather data in batches, aggregated over time. This is to avoid capturing individual vehicles, see privacy integrity in Section 4.7**.** However as is clear, the tech companies dominate this field with their reach and data processing capabilities.

Bluetooth sensors, installed in the roadside of transportation networks, collect travel time of vehicles equipped with such devices. When a Bluetooth-enabled vehicle is detected by sensors, the unique ID in each device and the time of detection is recorded. These media access control (MAC) addresses are used to determine the speed and travel time information for that device. Observations of multiple vehicles provide accurate estimates of traffic conditions, such as several traveltime samples between a pair of Bluetooth sensors.

4 Supporting data systems and considerations

In order to provide an accurate picture of what is happening on the nations' roads, an array of systems is needed. We describe some of them next but point out others may be used in certain circumstances. Air quality monitoring is one example, before and after a major change in the road infrastructure (the Stockholm bypass road for example).

4.1 GIS systems

GIS systems process not only positioning systems described below, but they also capture, store, check, and display data related to all positions of their interest. Transport GIS systems show streets, *roads*, buildings, pipelines, cables as well as many other transport related issues. Whether for placement of traffic sensors, understanding the best positioning, incorporating the landscape, or visualizing the traffic flow, GIS systems are essential. GIS is an enormous field, and we redirect interested readers to a textbook

For solely *mapping*, tools that are known include Google, WAZE, INRIX, HERE maps. For travel *planners*, some use an open file system, GTFS from Google originally, for timetabling. OpenStreetMaps is an open mapping system, and besides the core data, relies heavily on user contributions. Open REST APIs can be built into applications such as the above-mentioned ones. They inevitably have inbuilt misuse detection such as a certain number of requests from a particular source. Again, this is a large area and more reading is needed to obtain a fuller picture. Trafiklab, an open community, regularly runs hackathons to get people up to speed on coding open traffic sources.

4.2 Positioning systems

The Global Positioning System (GPS) is one of the best-known *global* positioning systems. Note others exist, the Russian GLONASS, EU's Galileo, India's NavIC, China's BeDou and Japan's Quasi-Zenith. Local position systems also exist, for example in Sweden the SWEREF99 system indicates the position along a motorway, from the start, 0 indicates the start, the direction, and the distance along the motorway (see below). Trafikverket provides the road position in SWEREF99 and GPS (called POINT), whilst the INRIX dataset uses GPS.

4.3 Accident data

Related to the road conditions, and is often a contributory factor, accidents are a sad artifact of modern road systems. Their impact on congestion is usually significant, usually in both directions of a road, as drivers on the opposite side slow down to observe the scene. An accident system provided by the police in Sweden, is called STRADA [*STRADA*].

4.4 Weather data

Clearly rain, fog, surface dampness/friction affects driving conditions and thereby flow, average speed and density of vehicles. Road agencies have dedicated divisions monitoring and disseminating weather information for their operations as well as road users. In Sweden, Trafikverket's (TrV) weather information is called VviS. TrV provides an Application Programming Interface (API) giving real time information and provides datasets for download/experimentation.

Contributory sources include, TrV's own sources such as the road, as opposed to atmospheric, weather (RWIS), road surfaces (PMSV3) plus many others. Their open data source pages provide more details and how to access them. In Sweden, the Meteorological and Hydrological Institute (SMHI) has provided minutely forecasts to the public since 1874. The open data comes with licensing terms specified in Creative Commons Attribution 4.0. SMHI's weather forecasts are presented in the media, smhi.se and via their app. Importantly, they provide an API to their databases, which can be queried with respect to road conditions.

4.5 Societal events

Often congestion is the result of 'external' factors, which could be rush hours, holidays, concerts, football matches and so on. Therefore, data exploration, analysis should consider these. News, travel apps and navigation systems. Radio bulletins and increasingly notifications from GIS systems warn users of congestion ahead of time or on the road. These cause vehicles to move away from their original route and thereby shift the flows to other channels in the road network.

4.6 Rush hours

One of the simplest phenomena to observe in speed and flow quantities is the presence of two rush hour periods. Again, Trafikverket provides an API to extract road traffic flow, an example is in Section 8.

4.7 Privacy, GDPR and integrity issues

There is a clear advantage to companies offering navigation services, as users divulge their speed, location amongst other information as a fee for using the collated information. Other operators on the other hand, must process groups of users, to preserve individual anonymity. In the EU, companies must also comply with GDPR issues if any personal information is captured and recorded. This includes faces, number plates etc. Any incident that leads to personal data being lost, stolen, destroyed, or changed is considered a data breach. Up to €20 million or 4% of annual global turnover can be fined. Notable GDPR infringements have been made by British Airways, Google and the Marriott hotel group.

5 Floating to macro inference: data-driven flow measurement

According to the classifying method proposed by Seo [*Seo*], approaches can be grouped into three categories:

1. Model-driven
2. Data-driven
3. Streaming-data-driven

They depend on whether the approach needs a predetermined assumption, i.e., a traffic flow model, and the data type they used. The approach adopted by this work belongs to the data-driven approach.

Model-driven approaches use predetermined physical traffic flow models to estimate the traffic state. Those traffic flow models describing the dynamic of traffic were developed based on physical principles and empirical relations. They are well-formulated with a fixed model structure, e.g., functional forms, and a fixed number of parameters. The model parameters, i.e, the functions' parameters, are calibrated by historical data collected on the road segments/networks where the models will be implemented. The calibrated model will then estimate the traffic state in an unobserved region using real-time data as input. The model-driven approach is the most popular type of TSE approach that has been adopted by various studies in TSE, which could estimate traffic states accurately under ordinary traffic conditions.

The data-driven approach uses the extracted dependence model to estimate the current traffic state based on real-time data. The same type of approach is also referred to as the non-parametric approach in the field of AI and traffic prediction [*VanLint2012Short-term*], where machine learning models are extensively used. Non-parametric means both of the model structure, e.g statistics functional form of ML algorithms, and the model parameters, i.e. functions' parameters or algorithms' parameters, are not determined in advanced but determined from the traffic data based on some evaluation metrics, e.g., RMSE and MAPE [*VanLint2012Short-term*]. Typically, a large amount of historical data is needed by data-driven approaches for building the dependence model. Recent ICT advances, e.g. IoT, have generated data increasingly from different sensor sources that are available for developing various data-driven approaches. Data-driven approaches, especially nonlinear ones such as neural networks, are good at modeling complex nonlinear relationships and traffic dynamics in the transportation field [*Seo2017*, *Vlahogianni2004Short*, *VanLint2012Short-term*]. Moreover, it does not need or need less prior knowledge on traffic to estimate the traffic state [3]. The data-driven approach needs less domain-specific knowledge of the traffic process for estimation.

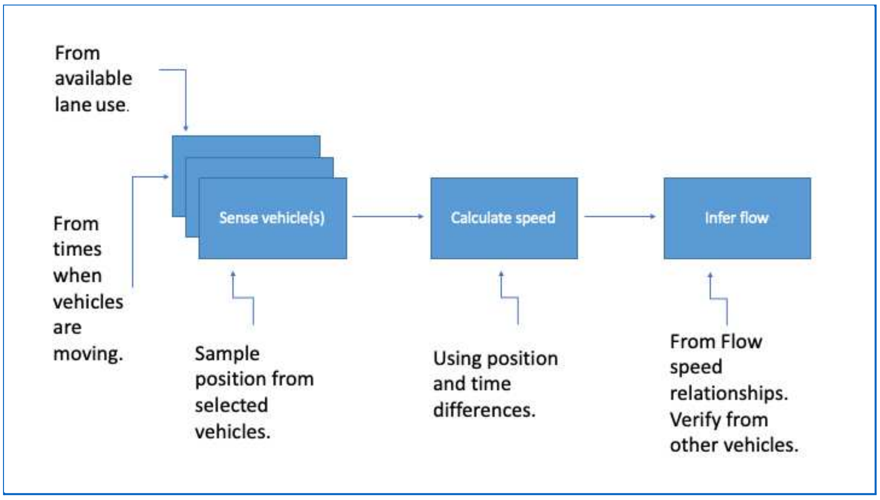

Figure 1: Sampling and inference as part of data-driven traffic flow.

Essentially our work focuses on using mobile devices to infer the traffic flow. This is because one can infer the traffic flow in a simpler, cheaper, deployable, and importantly, flexible manner. Speed and flow data from vehicles can be gathered from anywhere given there are sufficient vehicles to sample the road conditions. Before we embark on this, it is important to state which questions we are attempting to answer.

1. Is it possible (theoretically) to infer the road state from GPS / Bluetooth?
    a. If so, how much?
2. When harvesting microscopic flow and speed data, who should do this?
3. What is the best way to achieve the harvesting?
4. Can one use the macroscopic flow + speed as the ground truth[3]?
5. Does segment microscopic flow + speed correlate well with fixed location macroscopic speed + flow
6. Are the data volumes for each type (microscopic, macroscopic) similar enough?
7. Are the timestamps sufficiently aligned in both data types?
8. Data is gathered on a second resolution, what are suitable averaging / smoothing parameters?
9. Are there integrity issues with the data and processing and inference?

Table 1: Some questions in data-driven inference

Some of the issues are in the second report [*Marsh2021TENSII*], and alternative works have also considered the issues [*Gundlegård2020, Seo2017, Kim2014*].

6 Traffic Theory

6.1 Introduction

The goal of this section is to give lay readers the necessary background to understand the rest of this report. At certain resolutions, road traffic obeys physics and mathematical relationships. In this section, we will consider the most important relationships. Many more details, including derivations can be found in [*Elefteriadou2014*], particularly Chapters 3-5. This section concerns itself with modeling vehicle interactions, and *groups of vehicles* and not individual vehicle models, car-following, and such like. These can be found in the previous reference, Chapter 2.

6.2 Speed, distance, and time

The simplest and best-known relationship:

$$Average\ speed\ (u) = \frac{Distance\ (x)}{Time\ (t)}$$

[3] It is also an estimate of the traffic state.

Is the speed as the proportion to a distance covered in a specific time? The travel time, see below is in effect this equation with the average speed and time exchanged. Although simple in principle, practically the distance for the MCS and INRIX measurement infrastructures provide slightly offset values. See practical considerations, see section 8.

6.3 Travel time

The average speed (u) is defined as the total distance divided by the total time, thus

$$average\ speed\ = \frac{travel\ distance\ (td)}{travel\ time\ (tt)}$$

The average travel time (tt) over a distance l can be found as the average of the time a vehicle travels over a distance l:

$$mean(tt) = mean\left(\frac{l}{v}\right) = mean\left|\frac{l}{v}\ \right|$$

In this equation, tt indicates the travel time. This can be measured for all vehicles passing a road stretch, for instance at a local detector. Note that the mean travel time is not equal to the distance divided by the mean speed:

$$mean(tt)\ != \ \frac{l}{mean(v)}$$

It can be proven that in case the speeds of vehicles are not the same, the average travel time is underestimated if the mean speed is used.

$$mean(tt)\ \ <= \ \frac{l}{mean(v)}$$

The harmonically averaged speed, l divided by the average of (1/v) is a reasonable travel time estimation. If a new term, the pace P_i, is introduced:

$$P_i = \frac{1}{v_i}$$

The harmonically averaged speed is:

$$mean(v)_{harmonic} = \frac{1}{mean(P_i)} = \frac{1}{\frac{1}{v_i}}$$

The same quantity is required to find the space mean speed. We will see the travel time as an important quantity for the INRIX data. This is because the INRIX speed is measured over a segment, i.e., several meters a distance. Using the travel time in some more advanced methods might improve the estimation of flow and speed. One reason is that the travel time as an additional feature may complement speed when the speed value is small. Again, we will explore this further in the second report.

Time grouping in some datasets, the actual recorded, logged, and registered may be subject to time shifts. We will consider single state methods first.

6.4 Traffic model states

Transport engineers consider the observed road to be in a particular state, phase, or state. We will use the term *state*.

1. Free-flow or uncongested state
2. Congested
3. Transition
4. (Stationary)

Models concern themselves with these states and are organised around how many of the states they consider. The transition between the three typically is of particular interest to road traffic operators, data scientists, road, and city planners. We consider the first and simple traffic model first, single state traffic models.

Models use 2-3 variables such as an indicator of the speed up to a certain speed, up to where the density plays a role (see Figure 2 below). For keen readers, a complete review of single and multiple states (with examples) can be found in a recent 2021 publication [*Romanowska2021*].

6.5 Single state traffic models

In the following sections (6.6-6.10) we consider the free flow or congested flow *only*, i.e. a single state of the traffic state. Section 7 considers dual states of the traffic state which lead to the relatively well-known fundamental laws (and diagrams) of traffic flow.

6.6 Greenshields model (1935)

The first traffic stream model was developed by Greenshields who developed a linear speed–density relationship based on data in 1935 [*Greenshields1935*]. He derived a speed–flow relationship based on the assumption of a linear speed–density relationship.

$$v = v_{ff} - \left(\frac{v_{ff}}{D_j}\right)D$$

Where:

v_{ff} = free-flow speed
D_j = jam density
D = density
v = speed

The Greenshields model has been used extensively in transportation analysis as it is relatively easy to use in analytical models. Speed-density traffic models started in the 1930s and have been continually developed until 2017 (at least).

6.7 Greenberg's model (1959)

Using the logarithmic relationship between density and speed, Greenberg proposed [*Greenberg1959*].

$$v = v_{opt}\, ln\left(\frac{D_{max}}{D}\right)$$

Where:

D = density
D_{max} = max density
v = speed
v_{opt} = critical speed

Using Greenberg's model shows where a certain, in some cases critical, density leads to a significant speed decrease. Drop. Often as a knee in a bilinear plot of log(density) (x) versus speed (y). The free flow part is indicated by a horizontal (ish) segment, at the critical density, a tapering off phase to the jam state. One problem

Data from [*Notley2009*] suggests at 20 vehicles / per hour / lane can lead to overall speed reductions. One problem with Greenberg's model is when the density approaches zero, the speed approaches infinity.

6.8 Underwood's model (1961)

Underwood developed the following to avoid the speed tending to infinity as the density tends to 0.

$$v = v_{ff} \bullet e^{\frac{-D}{D_{opt}}}$$

Where:

v = speed

v_{ff} = free-flow speed

D_{opt} = optimal density

D = density

6.9 Van Aerde (1995)

$$D = \frac{1}{c_1 + \frac{c_2}{(v_f - v)} + c3v}$$

Where:

$$c_1 = fixed\ distance\ headway\ constant\ c_1 = mc_2$$

$$c_2 = first\ \ distance\ headway\ constant\ c_2 = \frac{1}{D_{max}(m + 1 / v_t)}$$

$$c_3 = second\ \ variable\ distance\ headway\ constant\ c_3 = \frac{-c_1 + v_{opt} / q_{max} - c_2 / (v_t - v_{opt})}{D_{max}(m + 1 / v_t)}$$

$$m = is\ a\ constant\ \frac{2v_{opt} - v_t}{(v_t - v_{opt})^2}$$

The reason for this complex form is that given some of the variables being set to 0, the formula degenerates to the previous models. For example, when $c_1 = c_3 = 0$ then Van Aerde reduces to Greenshields. For an even more recent model see Kucharski and Drabicki [*Kucharski2017*].

7 Dual state traffic models

Concern themselves with the free-flow *and* congested states. Within dual state schemes people refer to 'fundamental traffic laws' can, given two of (speed, flow, density) one can derive the third, as indicated in the simplified forms below.

```
Flow = Average speed x Density

q = u x k

Vehicles per hour = kilometers per hour x vehicles per minute
```

That is the average speed, the flow and the density of road vehicles follow connected relationships, through formulae in theory and empirical evidence in practice. In reality the terms are functions of space and time, i.e. q(x, t) and q is a function of average speed multiplied by density. When looking at the changes in these quantities we need differential equations and analysing them shows they are non-linear so although the equations above look simple, can quickly lead to solving analytically and numerically non-linear partial differential equations, which is by no means simple.

7.1 Flow and density

One illustration is a plot of flow against density.

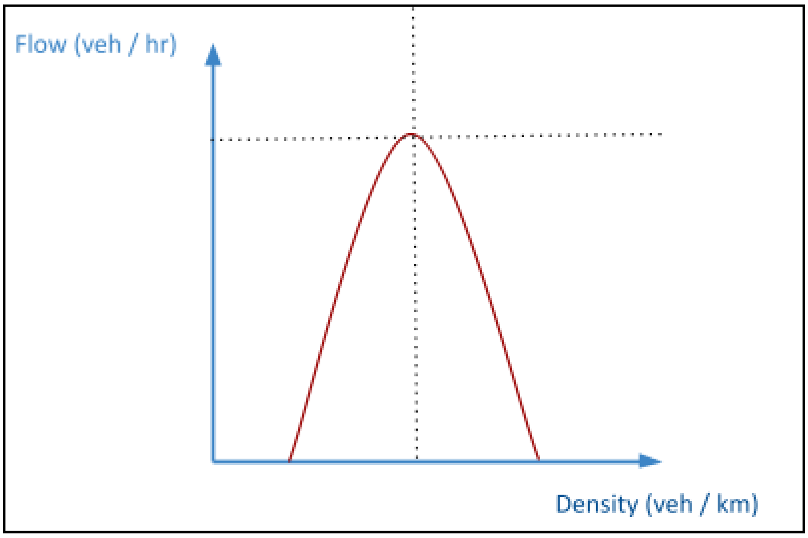

Figure 2: Simplified illustration of flow against density. Green represents free flow whilst red illustrations a congested state.

On the road, the relationship between flow and density can be seen as Figure 3.

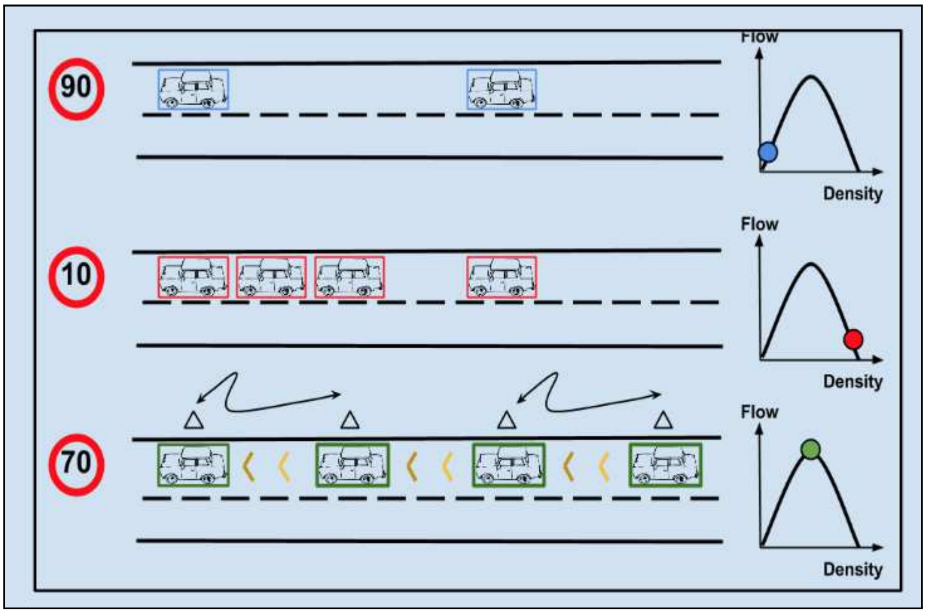

Figure 3: Illustration of the flow in relation to density.

A clearly non-linear relationship for varying speeds. At x=0, there are no vehicles, hence no density, k=0, therefore no flow (recall q=k·u). At maximum density (back-to-back) there is no flow as u=0, again no flow. At fixed speeds this produces the straight dashed lines. In the case of the one with the highest gradient is sometimes called the free-flow case, essentially the density, or congestion does not interfere with the desired speed of the driver. From empirical data up the optimal density this holds. One can see that there is a maximum density k^* at which the density *maximises* the flow. In road planning cases this is where the road 'should' operate, less implies potentially underused capacity and higher densities implies congestion.

7.2 Speed versus flow

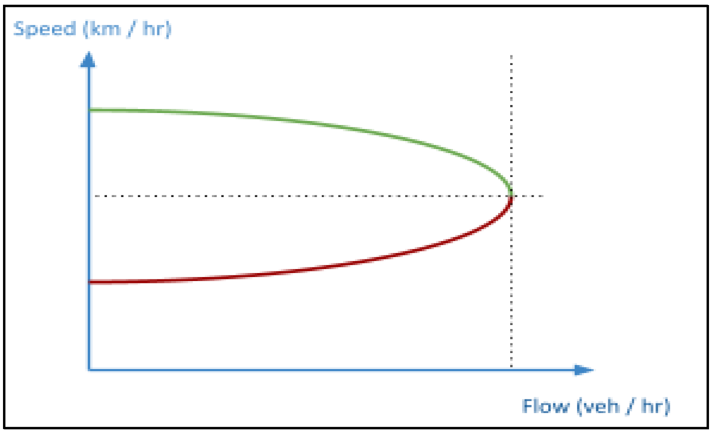

Figure 4: Simplified illustration of speed against flow. Green represents free flow whilst red illustrates a congested state.

Blandin et. al. studied the empirical relation between point speed and point flow for 112 stationary traffic radars in the San Francisco Bay Area, California [*Blandin2012*]. Their empirical measurements show that the speed-flow relation can look very different on different roads; particularly flat and inclined roads have different speed-flow relationships. From a driver perspective, the equation implies that drivers choose their speed according to the traffic density around them. The maximum speed is at 0 density, and conversely the minimum at maximum density. Typically, the equations are helpful for seeing the relationships between flow, density, and speed. Also, the equations help identify traffic flow states as discussed above.

8 Data processing and examples

8.1 Introduction

Once the data from whichever source (Table 1) is gathered it needs to be processed. Here we give examples from MCS and INRIX as examples[4]. In this report, we show what the MCS and INRIX data looks like. The CSV files resemble data in these tables.

8.2 The Motorway Control System (MCS)

ID	Date	Speed	Speed Deviation	Flow	Used Lanes
1164	2018-10-01 00:01:00	16.94	7.18	120	{1}

[4] The second report gives more details on how to infer traffic flow from floating / Bluetooth data.

1164	2018-10-01 00:22:00	13.05	9.53	60	{1}

Table 2: An example of the MCS dataset.

8.3 The INRIX data

Segment ID	TMCID Timestamp (UTC)	Segment Type	Speed	Average	Referenced speed	Travel time (mins)	Score	Cvalue	Speed Bucket	Registered
225525816	2018-10-01 00:00:15	XDS	85	0.501	60	30	56	49	3	2018-10-01 02:00:14
225525816	2018-10-01 00:01:13	XDS	95	0.501	68	30	56	98	3	2018-11-01 00:55:11

Table 3: An example of the INRIX dataset.

Segment ID. An identifier for defining a unique road segment. TMCID, a time segment a 19-digit time format used by INRIX to define the year:month:day:hours:minutes:seconds, 2014-09-30 23:59:33 for September 30, 2014 at the 23rd hour, 59th minute, and 33rd second for each TMC. Speed represents the average speed for a given TMC code, calculated from live data from the most current time slice. Referenced speed is an uncongested “free flow” speed determined for each TMC segment using the INRIX traffic archive. Average speed the historical average mean speed for the reporting segment for that time of day and day of the week in miles per hour. Travel time reported by INRIX based on an aggregation of data provided by GPS probes. Score. A measure of confidence in a given reported speed with three possible values: 10/20/30. Samples with confidence scores larger than 10 are based on real-time data, otherwise are based on historical data. Confidence is an attribute reported by INRIX having three levels: 10, 20, and 30. A confidence of 30 indicates that enough base data were available to estimate traffic conditions in real time, rather than using either historical speed based on time of day and day of week (indicated by confidence of 20) or free-flow speed for the road segment (indicated by a confidence of 10). Cvalue, the confidence value (range 0–100), designed to help agencies determine whether the INRIX value meets their criteria for real-time data. Speed Bucket. Level of congestion according to the speed range.

One of the simplest phenomena to observe in speed and flow quantities is the presence of the rush hour periods during the weekdays. Trafikverket provides an API to extract road traffic flow, an example is in Table 4.

Site Id	Meas. Time	Meas. or Calc. Period	Veh. Type	Veh. Flow Rate	Av. Veh. Speed	County No.	Geom.	Mod. Time	Reg. ID	Spec. Lane	Meas. Side
912	2021-11-29T16:47:59.000+01:00	60	Any Vehicle	1020	83	1	**SWEREF99TM:** POINT (671038.204092097 6580249.92891215) **WGS84:** POINT (18.0061951 59.3263245)	2021-11-29T15:48:25.099Z	4	lane4	unknown
914	2021-11-29T16:47:59.000+01:00	60	Any Vehicle	1380	59	1	**SWEREF99TM:** POINT (671042.379336325 6580244.04022384) **WGS84:** POINT (18.0062637 59.32627)	2021-11-29T15:48:25.099Z	4	lane2	unknown

Table 4: Extract example from Trafikverket's traffic flow API, link.

8.4 Practical issues

In both microscopic and macroscopic cases, the private information of the individual vehicle license plate needs to be hidden, removed, or obfuscated. In Europe GDPR rules apply, therefore no personal information should be used that can be used to identify you. This includes the license plate(s) of course.

That is the distances or road segments are rarely the same regions (as seen on a map). One may be straighter, the other more curved. The recorded times and registered time may also contain inaccuracies when using INRIX to infer the MCS flow.

The time resolution is road traffic analysis tends to be in the order of 10's of seconds to several hours. The spatial resolution tends to be in the 10's of metres. Therefore, at very a few meters or seconds the laws will probably be inaccurate. The spatial resolution can be of the order of a few hundred meters up to several kilometres.

Detector ID 1310 is the E4Z 61,890 which you find on E4 southbound and detector ID 1322 is the E4Z 63,025 here. The used_lanes are indicating which of the lanes are used when the aggregate speed/flow is calculated. For example {1,1,NULL} indicates that there are three lanes, but only two have valid measurements. There are MCS ids 1159 and 1162, see Figure 5 below.

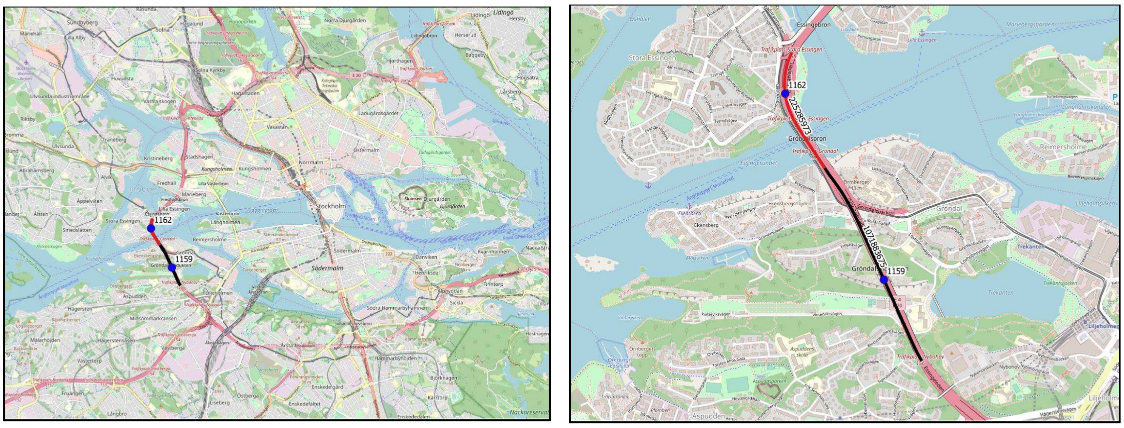

Figure 5: Example locations of floating and fixed traffic flow measurements in Stockholm.

Regarding data preparation, the following techniques of feature engineering for machine learning are adopted for preparing the datasets:

- Filtering: The INRIX datasets contain data points with a low confidence score (10), which always report the same speed value equals the historical reference speed, and thus cannot reflect the real-time traffic condition on the road segment. We filter out these data points in INRIX datasets to ensure the traffic variable values represent the traffic condition at the moment of measuring
- Smoothing: The measurements of traffic variables, i.e., flow, speed, are noisy in INRIX and MCS datasets. We need to remove noise and expose the variable values that signify the real trends to increase the model's estimation accuracy. Moving average with a window width of 30 timesteps is adopted for smoothing the speed, flow, and travel time.
- Shifting: A 6-minute latency of speed is observed in INRIX datasets compared with the MCS speed. This lag-time is within the latency range, from 6 to 8 min, reported by various studies [*Kim2014*]. We eliminate the latency by shifting the INRIX time-series by 360 s to ensure the measurements from both types of datasets reflect the same traffic condition in the same timestamp. Besides, INRIX and MCS timestamps were registered in different timezones, requiring time-series shifting to match the timestamps from the two platforms beforehand.

9 Related projects

Except the citations in this report we will focus on projects here. Again, the sister report will provide additional technical / scientific related works. Most of the German car industry, academic partners plus the government joined forces in the project, 2012-2015 [*FEAE12*]. UR:BAN had three main areas: 'Cognitive Assistance', 'Traffic System', and 'Human factors in traffic' of which there were 16 subprojects. Examples of sub-projects include i) routing tasks through metropolitan areas using information at the network level. Up to now, decisions have been based on static data such as travel time curves and rudimentary knowledge of the current traffic state. ii) A close integration of intelligent infrastructure with intelligent vehicles here is a prerequisite for achieving emissions reductions and enhancing traffic efficiency along a route. Microsoft used multi-year data to infer and forecast traffic flow in the ClearFlow project [*Mic12*]. The work leverages machine learning to build services that make use of both live streams of sensed information and large amounts of heterogeneous historical data. This has led to multiple prototypes and real-world services such as traffic-sensitive directions now used in Bing Maps. Their seminal work stimulated new efforts in related areas, such as privacy and routing [*HASL12*].

Mobile Century and Mobile Millennium [*MobileMillenium*, *Bayen2011*] were influential projects at University of California, Berkeley, in 2008-2011. Mobile Millennium was a research project that included a pilot traffic-monitoring system that used the GPS in cellular phones to gather traffic information, process it, and distribute it back to the phones in real time. The objective of the project was to demonstrate the potential of GPS in cell phones to alter the way traffic data is collected.

The Mobile Millennium Stockholm project [*Allström2011*] studied methods for data fusion. The project adapted and extended the knowledge gained from the Mobile Millennium project at Berkeley to estimate travel times in Stockholm. Examples of how methods developed in the Mobile Millennium Stockholm-project are used to estimate traffic state can for instance be found in [*Grumert2019*]. The report evaluates different control algorithms that determine which speed limits are going to be displayed on variable message signs on two motorway sections in Stockholm. In addition, the report studies how estimated traffic conditions can be used as a complement to fixed detectors installed along the road.

The focus of the current phase of the POST project is to integrate the results of previous MMS projects i) calibration of DTA models, ii) improved tunnel shutdowns and iii) ongoing motorway control strategies to build a framework for scenario evaluation for traffic management. By combining model-based scenario evaluation and optimisation with improved short-term predictions, POST assists traffic management at Trafik Stockholm to develop proactive strategies [*POST*].

10 Potential future directions

Each data method has pros and cons, and macroscopic cross-sectional ones are outlined below. Some are addressed with the probe approach, however ascertaining traffic mixes and accounting for when drivers change lanes lies beyond the scope of the TENS project. This report highlights the basics of data-driven traffic flow. Clearly ITS systems, traffic flow and its effects (mostly negative) are huge multi-billion areas for cities, citizens, companies, lobby groups, politicians, city planners and so on. Some topics for further investigation or are still mostly open are:

10.1 Vehicle lane changes

These are important as the flow typically changes when vehicles switch in and out of lanes. Predicting when drivers change lanes is non-trivial as it includes a human decision. The statistical i.e., macro properties of different lanes, however, can be explored see for example [*Marsh2021EoQ*].

10.2 Lane merges and splitting

The topology of the road network has a significant effect on the flows, lane merges or ramp on / offs are often responsible for choke or congestion points. Work has been done at RISE to analyze the effect of the road structure on traffic flows [*Reginbald2018*, *Sigurdsson2018*].

10.3 Separating the free flow and congested phase(s)

Separating the free-flow state from the congested state is one option, not unlike [*Blandin2012*]. In this approach we use temporal factors and identify the rush-hours and off-peak, as the congested state and free-flow state are predictable at the same time on most (work) days. [*Blandin2012*] states that flow is very difficult (or impossible) to predict in the free-flow state. This is because, in theory, the speed in the free-flow phase is almost constant (a horizontal line in the speed-flow plot, see [*Notley2019*]. But for our dataset, if we plot speed vs flow and look at the free-flow phase, there is steeper slope of the line, which makes it easier to predict the flow from speed, see our sister report.

10.4 Traffic mix

Knowing what types of vehicles, cars, trucks, and buses5 play a significant role on the nations' roads. Trucks occupy certain roads more, and typically are more frequent in the evenings, where goods transport is commonplace. Since they occupy the slowest lane, they also affect the relative speed of traffic in each lane, and of course, induce overtaking (i.e., lane changes). One innovative approach under discussion is to use the camera images from trafiken.nu to classify the type of vehicle at a particular place and time to obtain the mix in real time.

10.5 Using travel times

A topic is to use the travel times from type ii) (probe) as input, and to provide the flow, density or speed as an output given some of the other variables. A comparison between probe and cross sectional can be found in [*Kim2014*].

10.6 Using density as early warnings of congestion

Traditional methods predict speed or flow. Using the density, easily obtainable from the fundamental relationships might be a less explored parameter to infer and predict. High density states are as responsible for congestion and also might be a better predictor of emissions (since a high density of vehicles implies a large number and higher emissions). We explore this aspect a little further in the second report.

10.7 Streaming approaches

The streaming-data-driven approach uses only real-time data, i.e., streaming data, and some weak assumptions to estimate the traffic state in the unobserved region. Weak assumptions are some basic principles, e.g., conservation law in traffic flow theory, supported by physical theory and does not need empirical justification. Streaming-data-driven approaches do not need to extract dependence from historical data like data-driven approaches or calibrate the model parameters using historical data like model-driven approaches. However, it needs a large amount of streaming data to estimate the traffic state and usually has lower estimation accuracy than data-driven and model-driven approaches. On the other hand, because it does not rely on any empirical relations, it is more robust under uncertain phenomena and unpredictable conditions, e.g., traffic accidents. See also [*Cebecauer2017*].

11 Conclusions

Gathering data from road traffic is a complex, intricate process. Often it is costly, involves people working in harsh conditions and needs complex sensors and IT systems. Traffic centers, information need near-time information for many types of road usage. This report has given an overview of some of the basics as well as providing some background (maths, projects and data preparation) on the topic.

For areas where INRIX and MCS measurements are present we believe that INRIX or indeed floating car data (GPS / Bluetooth) can be used to infer the flow. Qualitative data is given in the second report as well as the techniques used to draw this conclusion.

Finally, the authors make a plea for open data on this topic. Even if historical data, time, (some students) and financing such as given in this case make the studies feasible and possible.

12 References

[*Bhouri2019*] Neila Bhouri, Maurice Aron, Habib Haj Salem. *A Data-driven Approach for Estimating the Fundamental Diagram*. Promet - Traffic & Transportation, 2019, 31 (2), pp. 117-128. ff10.7307/ptt. v31i2.2849ff. hal-02110696v2

[*Bertsimas2018*] Dimitris Bertsimas, Arthur Delarue, Patrick Jaillet and Sebastien Martin, *Travel Time Estimation in the Age of Big Data*, Operations Research Center, Massachusetts Institute of Technology May 2018, http://web.mit.edu/jaillet/www/general/travel-time-18.pdf

[*Blandin2012*] Blandin, S., Salam, A., Bayen, A.M., 2012. *Individual speed variance in traffic flow: analysis of Bay Area radar measurements*, in: Transportation Research Board 91st Annual Meeting.

[*Cosar2018*] Cosar Ghandeharioon, *An evaluation of deep neural network approaches for traffic speed prediction*, KTH EECS, 2018, Thesis.

[*Cebecauer2017*] Cebecauer, M., Jenelius, E. and Burghout, W. (2018). *Integrated framework for real-time urban network travel time prediction on sparse probe data*. IET Intell. Transp. Syst 12: 66-74. https://doi.org/10.1049/iet-its.2017.0113.

[*Duan2016Efficient*] Y. Duan, Y. Lv, Y.-L. Liu, and F.-Y. Wang, *An efficient realization of deep learning for traffic data imputation*, Transportation research part C: emerging technologies, vol. 72, pp. 168–181, 2016.

[*Elefteriadou2014*] Eli Elefteriadou, *An Introduction to Traffic Flow Theory*. Book. Springer, 2014.

[*Forsman2015*] Åsa Forsman, Anna Vadeby, David Gundlegård and Rasmus Ringdahl, Utvärdering *av hastighetsmätningar med blåtandssensorer Jämförelse med data från MCS (Motorway Control System),* 2015, VTI rapport 969, Link.

[*French1986*] Traffic Data Collection and Analysis: Methods and Procedures, Transportation Research Board, National Research Council.

[*Gühnemann2004*] Astrid Gühnemann, Ralf-Peter Schäfer, Schäfer Kai-Uwe Thiessenhusen and Peter Wagner. *Monitoring traffic and emissions by floating car data*, 2004.

[*ITF2015*] International Transport Forum, *The ITF transport outlook 2021*[5], link (previous years also available).

[*Kang2019*] *Kang, Traffic Density Estimation Based on Vehicle Speed Profile Data,* Extended Abstract at the 98th Traffic Board Meeting, 2019.

[*Kim2014*] Seoungbum Kim, Benjamin, Coifman, *Comparing INRIX speed data against concurrent loop detector stations over several months*, Transportation Research Part C: Emerging Technologies. Volume 49, 2014, Pages 59-72, ISSN 0968-090X, https://doi.org/10.1016/j.trc.2014.10.002.

[*Gundlegård2016*] David Gundlegård, Clas Rydergren, Nils Breyer and Botond Anti, Tr*avel demand estimation and network assignment based on cellular network data*. Computer Communications, 95:29–42, 2016.

[*Gundlegård2020*] David Gundlegård, *Förstudie kring användning av probedata för skattning av tidsvarierande reseefterfrågan och trafiktillstånd*, TENSE project, 2020-11-30.

[*Jing2016*] Boyu Jing et.al. *Development of a vehicle emission inventory with high temporal–spatial resolution based on NRT traffic data and its impact on air pollution in Beijing part 1: Development and evaluation of vehicle emission inventory*. Atmospheric Chemistry and Physics, 16(5):3161–3170, 2016.

[*Hooper2014Impact*] Hooper, E., Chapman, L. & Quinn, A. *The impact of precipitation on speed–flow relationships along a UK motorway corridor*. *Theor Appl Climatol* 117, 303–316 (2014). https://doi.org/10.1007/s00704-013-0999-5

[*Hsu2021*] Pei-Lun Hsu, *Data-driven macroscopic traffic prediction from microscopic measurements*, KTH, 2021, report.

[*Kim2014*] Seoungbum Kim, Benjamin, Coifman, *Comparing INRIX speed data against concurrent loop detector stations over several months*, Transportation Research Part C: Emerging Technologies. Volume 49, 2014, Pages 59-72, ISSN 0968-090X, https://doi.org/10.1016/j.trc.2014.10.002.

[*Liu2015*] Xiaobo Liu, *Evaluation of floating car technologies for travel time estimation*, DOI: 10.1007/BF03325777, link.

[5] Previous years are also available.

[*Hooper2014*] Hooper, E., Chapman, L. & Quinn, A. The impact of precipitation on speed–flow relationships along a UK motorway corridor. *Theoretical Applications of Climatology* 117**,** 303–316 (2014) https://doi.org/10.1007/s00704-013-0999-5.

[*Kucharski2017*] Rafał Kucharski, Arkadiusz Drabicki, *Estimating Macroscopic Volume Delay Functions with the Traffic Density Derived from Measured Speeds and Flows*, Journal of Advanced Transportation, vol. 2017, Article ID 4629792, 10 pages, 2017. https://doi.org/10.1155/2017/4629792

[*Marsh2012*] I. Marsh, *VANET communication: a traffic flow approach*, PIMRC 2012.

[*Marsh2021*] Ian Marsh, *End of traffic queue detection*, paper in progress.

[*MarshTENS2021I*] Ian Marsh and Pei-Lun Tsu. D*ata-driven traffic flow: A state-of-the-art report*, RISE report 2021:93, ISBN 978-91-89385-83-2 (this report).

[*MarshTENS2021II*] Ian Marsh, Henrik Abrahmsson, Pei-Lun Tsu. *Data-driven traffic flow, predicting macro flows from micro measurements*. TENS Report I, 2021.

[*Mathew2007Introduction*] Tom V. Mathew and K. V. Krishna Rao, *Introduction to Transportation Engineering*, 2007, online.

[*Nissan2010*] Nissan, Albania, *Evaluation of Variable Speed Limits: Empirical Evidence and Simulation Analysis of Stockholm's Motorway Control System*, Ph.D Thesis, KTH, School of Architecture and the Built Environment (ABE), Transport and Economics, Traffic and Logistics. 2010, link, viewed June 10, 2020.

[*Notley2009*] Speed, Flow and Density of Motorway Traffic, Transport Research Laboratory, ISR INS003, 2009.

[*Polson2015*] Polson, Nicholas; Sokolov, Vadim. *Bayesian analysis of traffic flow on interstate I-55: The LWR model*. Ann. Appl. Stat. 9 (2015), no. 4, 1864–1888. doi:10.1214/15-AOAS853. https://projecteuclid.org/euclid.aoas/1453993096.

[*POST*] *Prediction and Scenario-based Traffic management*, LiU KTH, & Trafikverket. The Centre for Traffic Research, (CTR), link.

[*Raadsen2016*] Mark P.H. Raadsen, Michiel C.J., Bliemer Michael and G.H. Bell. *An efficient and exact event-based algorithm for solving simplified first order dynamic network loading problems in continuous time*. Transportation Research Part B, 92:191–210, 2016.

[*Reim2012Analysis*] Erich Reim, *Analysis and visualization of historical traffic data collected on the Stockholm highway system*, 2013-10-30.

[*Reginbald2018*] Reginbald Ivarsson, Jón, Scalable *System-Wide Traffic Flow Prediction Using Graph Partitioning and Recurrent Neural Networks*, KTH EECS, 2018, Thesis.

[*Romanowska2021*] Romanowska, A. Jamroz, K. *Comparison of Traffic Flow Models with Real Traffic Data Based on a Quantitative Assessment*. Appl. Sci. 2021,11,9914. https://doi.org/10.3390/app11219914.

[*Seo2017*] Toru Seo et.al, *Traffic state estimation on highway: A comprehensive survey*, Annual Reviews in Control. Volume 43, 2017, Pages 128-151, ISSN 1367-5788, https://doi.org/10.1016/j.arcontrol.2017.03.005.

[*Sigurdsson2018*] Thorsteinn Thorri Sigurdsson, *Road traffic congestion detection and tracking with Spark Streaming Analytics*, KTH EECS, 2018, Presentation, Thesis.

[*STRADA*] https://www.transportstyrelsen.se/sv/vagtrafik/statistik/olycksstatistik/om-strada (in Swedish)

[*Tsanakas2019*] Tsanakas N, Ekström J, Gundlegård D, Olstam J, Rydergren C, *Data-driven network assignment*, Swedish National Transport Conference, Linköping, Oct 2019.

[*UN2021*] The United Nations, *The 17 goals, sustainable development*, accessed:2021-05-16. Available: https://sdgs.un.org/goals.

[*VanLint2012Short-term*] J. Van Lint and C. Van Hinsbergen, *Short-term traffic and traveltime prediction models*, Artificial Intelligence Applications to Critical Transportation Issues, vol. 22, no. 1, pp. 22–41, 2012.

[*Vlahogianni2004Short*] I. Vlahogianni, J. C. Golias, and M. G. Karlaftis, *Short-term traffic forecasting: Overview of objectives and methods*, Transport reviews,vol. 24, no. 5, pp. 533–557, 2004.

[*Yang2019*] Shu Yang Yao and Jan Wu, *Data-Driven Approaches for Estimating Travel Time Reliability*, 10.1016/B978-0-12-817026-7.00004-7, 2019.

[*Yatskiv2013*] I. Yatskiv, A. Grakovski, El Yurshevich, *An Overview of Different Methods Available to Observe Traffic Flows Using New Technologies*, NTTS 2013.

[*Yperman2007*] I. Yperman. *The link transmission model for dynamic network loading*. Ph.D. Thesis, Katholieke Universiteit Leuven, 2007.

[*Zhan2016*] Xianyuan Zhan, Yu Zheng, Xiuwen Yi and Satish V. Ukkusuri. *Citywide traffic volume estimation using trajectory data*. IEEE Transactions on Knowledge and Data Engineering, 29(2):272–285, 2016.