Traffic Data

Trajectory and Car floating Data

Introduction

First, we will look at the relationship between a real system and a simulated system, as shown in Figure 1. As indicated earlier, validation intends to determine how well a simulation model replicates a real system. In calibration, the outputs of the simulation and the real system are also compared, but the parameters of the simulated system are optimized until the difference between both outputs is minimal or at least meets specific minimum requirements. Ideally, the inputs of the real and simulated systems should be identical. Therefore, both the input variables and outputs of the real system should be observed.

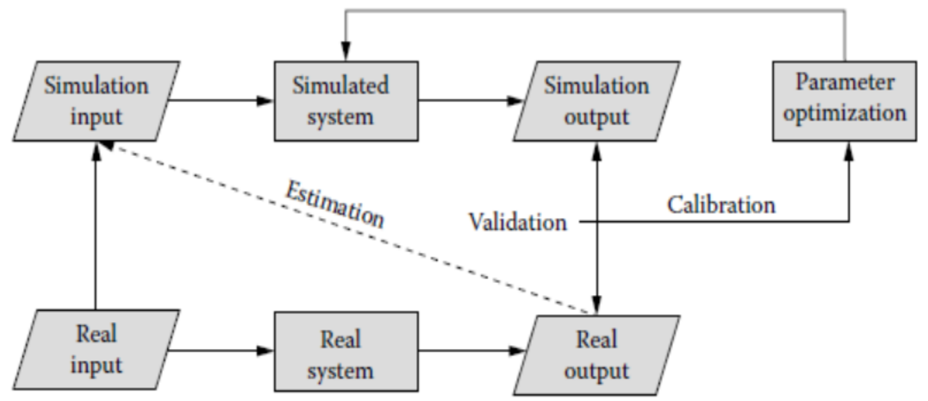


Fig. 1 Relationship of simulated and real systems and locations of calibration and validation processes. (Source: Toledo, T. and Koutsopoulos, H. 2004. *Transportation Research Record, 1876, 142–150.)*

A framework for calibration and validation of traffic simulation models is shown in Figure 2. Calibration and validation of traffic simulation models involve two steps (Toledo et al., 2003).

<u>First Step</u>
The individual models of the simulator (e.g., driving behavior and route choices) are estimated using disaggregated data. Disaggregate data include detailed driver behavior issues such as vehicle trajectories. These individual models may be tested independently, for example, using a holdout sample. The disaggregate analysis is performed by statistical software and does not involve the use of a simulation model.

Second Step

The simulation model as a whole is calibrated and then validated using aggregate data (e.g., flows, speeds, occupancies, time headways, travel times, and queue lengths). Aggregate calibration and validation are important both in developing the model and applying it. The role of aggregate calibration is to ensure that the interactions of the individual models within the simulator are captured correctly and to refine previously estimated parameter values.

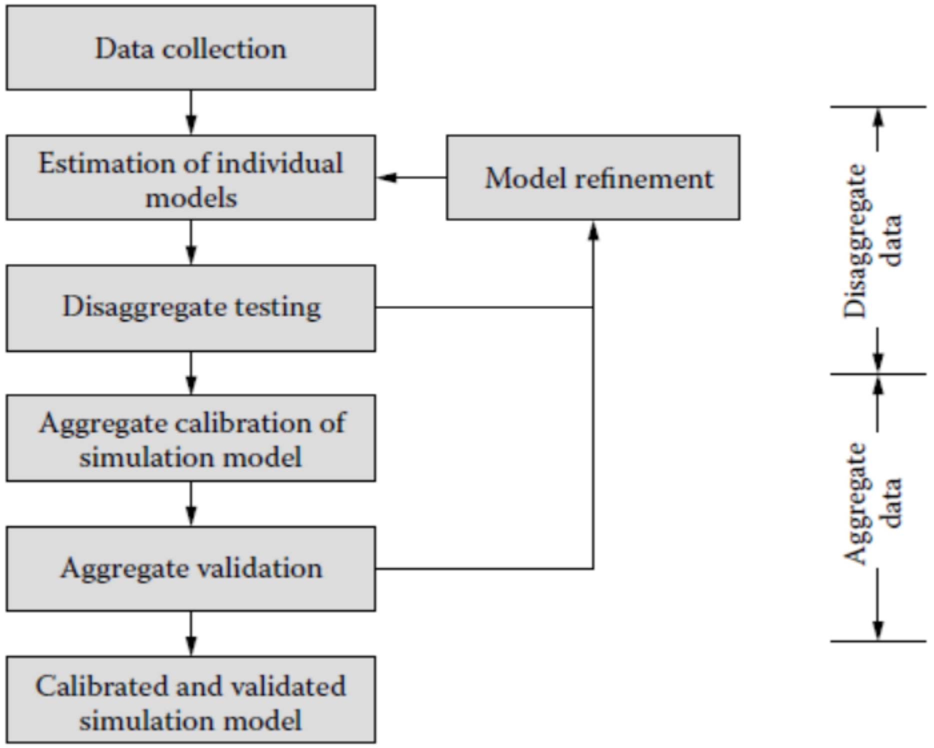


Fig. 2 Calibration and validation framework of traffic simulation models. (Source: Toledo, T., Koutsopoulos, H.N., Davol, A. et al. 2003. Transportation Research Record, 1831, 65–75.)

Dynamic Traffic Flow

There are several aspects of why a mathematical description of traffic flow dynamics nevertheless makes sense. Firstly, a huge amount of traffic flow data is available ranging from the acceleration characteristics of single drivers and vehicles to macroscopic data obtained by stationary detectors, supplemented by a rapidly growing amount of data obtained by GPS, wireless LAN, and mobile phones inside the vehicles. The associated measurements—corresponding to experiments in the fields of the natural sciences—serve as the basis of any mathematical modeling see Figure 3.

Traffic flow models describe the dynamics of vehicles and drivers in terms of mathematical equations. Predictions are obtained by running the model simulation. The values of the model's parameters are chosen so that the simulation produces the best fit with the data (model calibration). Once calibrated, the model can be used for traffic flow prediction and other applications

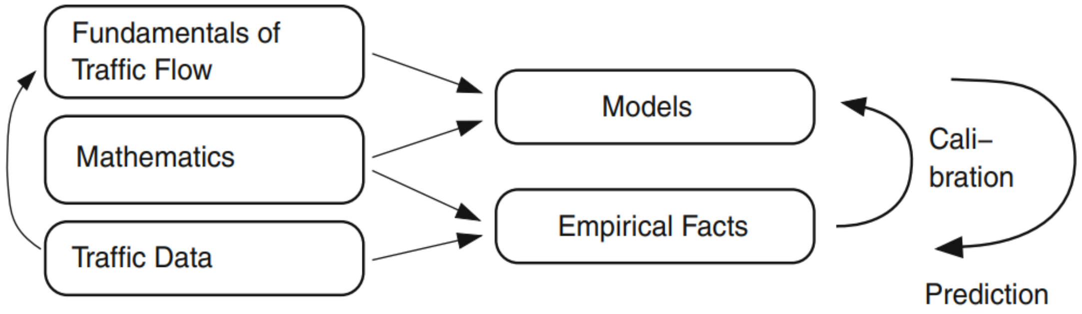


Fig. 3 Traffic Flow Modeling.

Delimitation of traffic flow dynamics

One can distinguish traffic flow dynamics from other fields of traffic science by the time scales given in Table 1. Traffic flow dynamics include time scales ranging from about one second to a few hours. Human reaction times and the time gap between two vehicles following each other are of the order of 1s while braking and acceleration maneuvers typically take several seconds. In city traffic, the period of one red-green cycle of traffic lights is of the order of 1 min while, on freeways, the period of traffic oscillations and stop-and-go waves is between 5 and 20 min. Finally, the traffic demand serving as an exogenous variable (model input) for traffic flow models varies on time scales of one hour, as illustrated by the term “rush hour”.

Table 1. Delimitation of traffic flow dynamics from vehicular dynamics and transportation planning.

Time scale	Field	Models	Aspect of traffic (examples)
<0.1 s	Vehicle dynamics	Sub-microscopic	Control of engine and brakes
1 s			Reaction time, time gap
10 s	Traffic flow dynamics	Car-following models	Acceleration and deceleration
1 min		Macroscopic models	Cycle period of traffic lights
10 min			Stop-and-go waves
1 h			Peak hour
1 day		Route assignment traffic demand	Daily demand pattern
1 year	Transportation planning		Building/changing infrastructure
5 years		Statistics age pyramid	Socioeconomic structure
50 years			Demographic change

Longer time scales ranging from hours to years are the domain of transportation planning. This includes the very long time scales of variations in traffic demand caused by demographic change. Transportation planning and traffic flow dynamics complement each other: The endogenous variables (model output) of the classical four-step scheme of transportation planning and its modern dynamical variants are the traffic demand (vehicles per hour) on each link of the considered network. For traffic flow simulations, in turn, these variables are exogenous (externally given), typically in form of boundary conditions.

Data Collection

Traffic data collection can and should be used for the calibration and validation of traffic simulation models. There are big differences in the availability of data from different sources. Some types of data such as loop detector data are widely available and used. Some can be measured with additional effort, for example, travel time data from GPS probe vehicles. Some types such as trajectory data are available only in rare situations such as research projects.

This means that a simulation study carried out as part of a traffic engineering project, having a restricted budget, typically must rely on existing loop data or can at most utilize some GPS probe drives. The objective of calibration and validation in a traffic engineering project is mainly to check whether a model of a specific area replicates

- at a desired level of detail
- the macroscopic traffic conditions (flow, speed, travel time) for a certain traffic demand.

Consequently, data for calibration and validation in traffic engineering projects typically need not be microscopic. Conversely, data generated with much more effort (e.g., trajectory data) are typically used by researchers to investigate driver behavior in general Analysis of driving behavior such as car following and lane changing requires highly detailed data to generate adequate insight into the traffic features to be modeled. These data are typically very expensive and/or laborious to acquire. An interesting point is the expected quality of the data. However, there is some ambiguity in existing studies because the "performance of a data collection system" is a result of several factors (hardware and software used, sensor configuration, and environmental and traffic conditions).

Data categories

Trajectory data: full (x_i, Y_i, t) coverage of all vehicles i in a certain spatiotemporal region, typically taken from external camera images I.

Floating-car (FC) data: GPS records from inside a vehicle i Extended FC (xFC) data: Other inside-vehicle sensors recorded as well.

Cross-sectional data: Typically recorded from stationary detector stations.

Event-oriented data: Accident and traffic jam information i Infrastructure data: Besides the road network the traffic-light phases etc.

Trajectory and Floating-Car Data

Different aspects of traffic dynamics are captured by different measurement methods. In this lecture, we discuss trajectory data and floating-car data, both providing space-time profiles of vehicles. While trajectory data captures all vehicles within a selected measurement area, floating-car data only provide information on single, specially equipped vehicles. Furthermore, trajectory data is measured externally while, as the name implies, floating-car data is captured inside the vehicle.

Data Collection Methods

Traffic can be directly observed by cameras on top of a tall building or mounted on an airplane. Tracking software extracts trajectories $x_\alpha(t)$, i.e. the positions of each vehicle α over time, from the video footage (or a series of photographs). If all vehicles within a given road section (and period) are captured in this way, the resulting dataset is called trajectory data. Thus, trajectory data is the most comprehensive traffic data available. It is also the only type that allows direct and unbiased measurement of traffic density and lane changes.

However, camera-based methods involve complex and error-prone procedures which require automated and robust algorithms for vehicle tracking and thus are often the most expensive option for data collection. Furthermore, a simple camera can cover a road section of at most a few hundred meters since smaller vehicles are occluded behind larger ones if the viewing angle is too low.

A different method uses probe vehicles that “float” in the traffic flow. Such cars collect geo-referenced coordinates via GPS receivers which are then “map-matched” to a road on a map-the speed is a derived quantity determined from the spacing (on a map) between two GPS points. This type of data is called floating-car data (FCD).

Some more recent navigation systems also record (anonymized) trajectories and send them to the manufacturer. The probe vehicles can be equipped with other sensors (e.g. radar) to record the distance to the leading vehicle and its speed (however, such equipment is expensive). FCD augmented in this way is also referred to as extended floating-car data (xFCD). One problem of FCD is that many equipped vehicles are taxis or trucks/vans of commercial transport companies which, due to their lower speeds, are not representative of the traffic as a whole.
Fortunately, this bias vanishes just when the FCD information becomes relevant: In congested situations, free-flow speed differences do not matter.

Both trajectory and floating-car data record the vehicle location $x_\alpha(t)$ as a function of time, yet they differ substantially:

- Trajectory data records the spatiotemporal location of all vehicles within a given road segment and time interval while FCD only collects data on a few probe vehicles.
- Contrary to trajectory data, FCD does not record which lane a vehicle is using since present GPS accuracy is not sufficient for lane-fine map-matching.

- FCD may contain additional information such as the distance to the leading vehicle, the position of the gas/brake pedals, activation of turning signals, or the rotation angle of the steering wheel (xFCD). In principle, every quantity available via the CAN bus (The CAN bus is a micro-controller communication interface present in all modern vehicles) can be recorded as a time series. This kind of data is naturally missing in trajectory data due to the optical recording method.

Time-Space Diagrams

Figures 4 and 5 are examples of trajectory data of a single lane visualized in a space-time diagram. By convention, we will always plot time on the x-axis vs. space on the y-axis. The following information can be easily read from the diagrams:

- The local speed at (front-bumper) position x and time t are given by the gradient of the trajectory. A horizontal trajectory corresponds to a standing vehicle.
- The time headway, or simply headway, Δt_αbetween the front bumpers of two vehicles following each other is the horizontal distance between two trajectories (The time headway is composed of the (rear-bumper-to-front-bumper) time gap plus the occupancy time interval of the leading vehicle).
- Traffic flow, defined as the number of vehicles passing a given location per time unit, is the number of trajectories crossing a horizontal line denoting this time interval. It is equal to the inverse of the time mean of the headways.
- The distance headway between two vehicles is the vertical distance of their trajectories. It is composed of the distance gap between the front and the rear bumpers plus the length of the leading vehicle.
- The traffic density, defined as the number of vehicles on a road segment at a given time, is the number of trajectories crossing a vertical line in the diagram and thus the inverse of the space means of the distance headways.
- Lane changes to and from the observed lane are marked by beginning and ending trajectories, respectively.
- The gradient of the boundary of a high-density area indicates the propagation velocity of a traffic jam. The congestions in Figs.4 and 5 are stop-and-go waves that are moving upstream and thus have a negative propagation speed.

If not only the longitudinal positions $x_\alpha(t)$ (along the road) but also the lateral positions $y_\alpha(t)$ (across the lanes) are recorded, one can generate a two-dimensional trajectory diagram from which one can deduce lateral accelerations and the duration of lane changes.

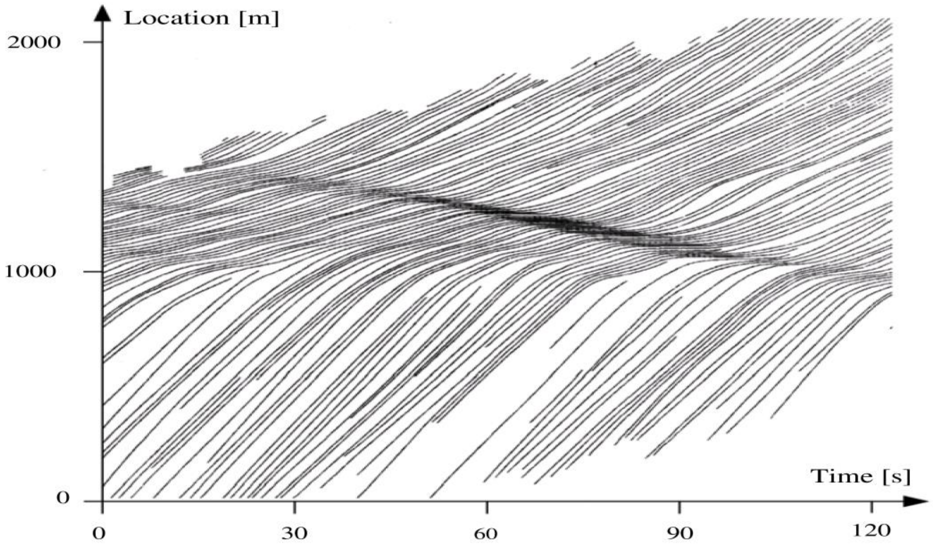


Fig. 3 The original: trajectories showing a stop-and-go wave on a British motorway segment. [Adapted from: J. Treiterer et al. (1970)].

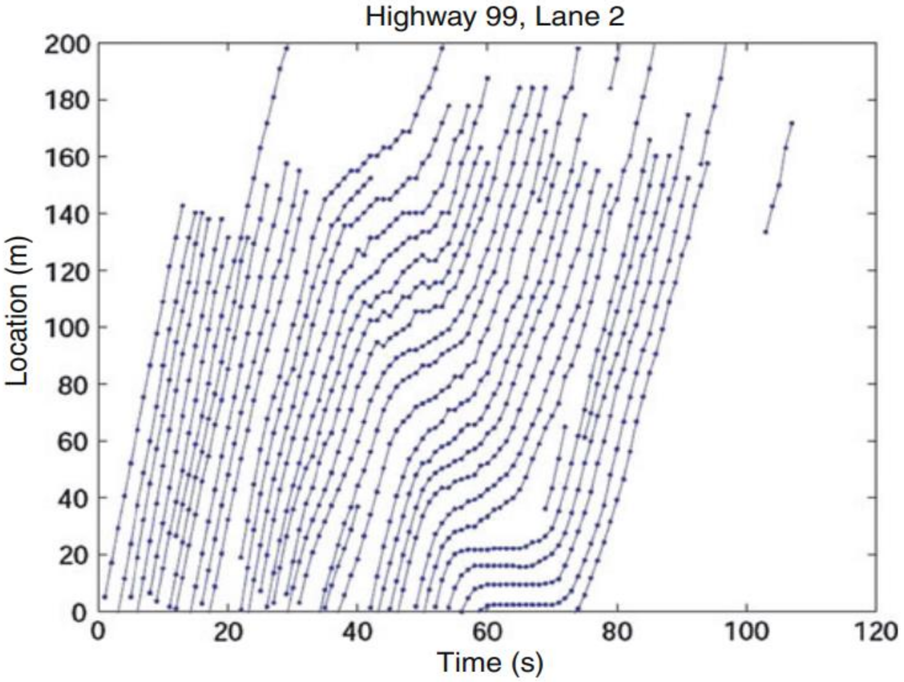
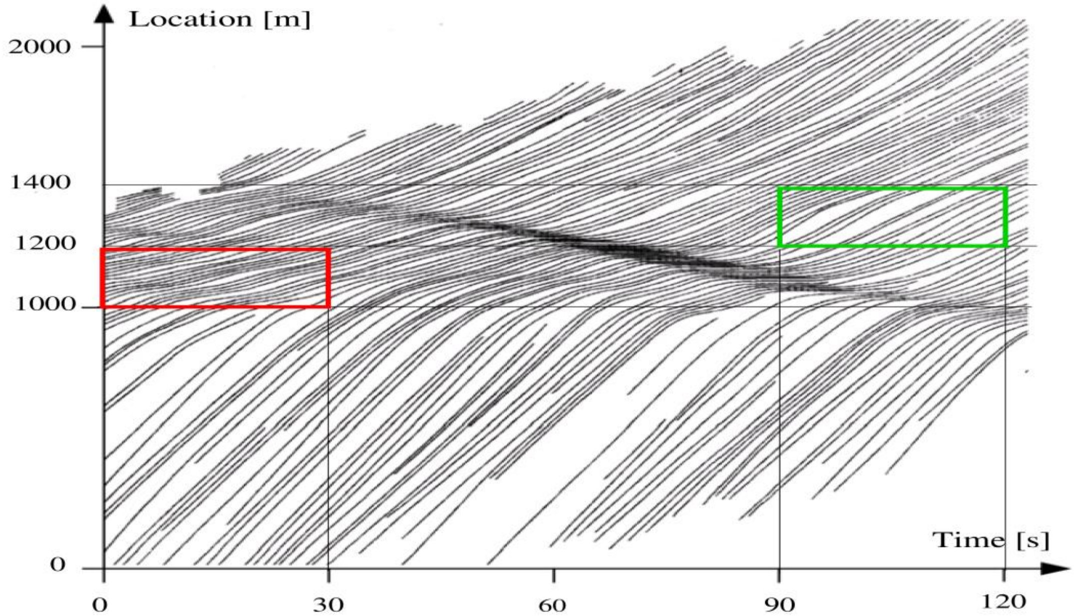


Fig. 3 Trajectories with moving stop-and-go waves on the California State Route 99.

Determining the State from Trajectory Data

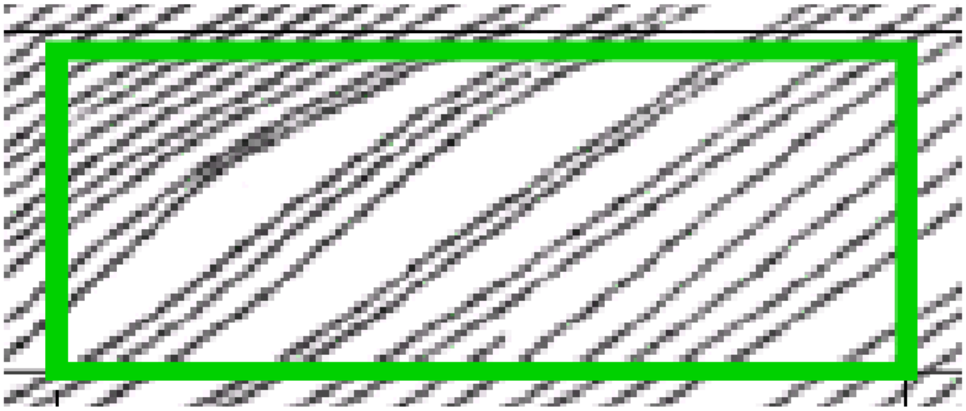


Estimate the wave velocity and the lane-changing intensity (#changes/h/km)

Region 1: $[0\,\mathrm{s}, 30\,\mathrm{s}] \times [1\,000\,\mathrm{m}, 1\,200\,\mathrm{m}]$

1. Estimate the local density $\rho(\bar{x}, \bar{t})$ by trajectory counting
   (15+13)/2 per 200 m=70 veh/km
2. estimate the local flow $Q(\bar{x}, \bar{t})$by counting
   (11+14)/2 per 30 s=25 veh/min=1 500 veh/h
3. estimate the local speed $V(\bar{x}, \bar{t})$ by the gradient method and compare it with the hydrodynamic relation V = Q/ρ
   Gradient method; take Diagonal: 180 m/28 s ≈ 6.5 m/s = 23 km/h;
   Hydrodynamic relation: V = Q/ρ = 1500 veh/h/70 veh/km = 21 km/h

Determining the local state of the boxed regions II



Region 2: $[90\,\mathrm{s}, 120\,\mathrm{s}] \times [1\,200\,\mathrm{m}, 1\,400\,\mathrm{m}]$

1. Estimate the local density $\rho(\bar{x},\bar{t})$ by trajectory counting
   (10+8)/200 m=45 veh/km
2. estimate the local flow $Q(\bar{x},\bar{t})$ by counting
   (14+16)/2 veh/30 s=1 800 veh/h
3. estimate the local speed $V(\bar{x},\bar{t})$ by the gradient method and by the hydrodynamic relation
   Gradient method: 200 m/18 s=11 m/s=40 km/h;
   Hydrodynamic relation: $V = Q/\rho$ = 1 800 veh/h/45 veh/km = 40 km/h

Differences between Trajectory and FC Data

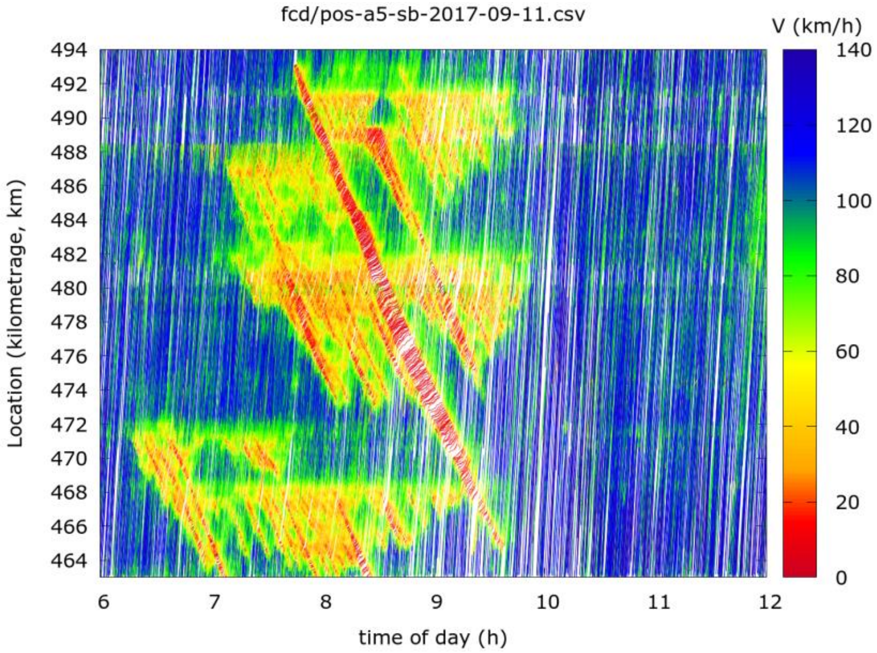
- Position of the data source: stationary vs. mobile; out-of-system vs. in the system
- Sampled spatiotemporal regions: very small vs. huge.
- Scalability: Non-existent (at present) vs. huge (essentially all roads with significant traffic, worldwide).
- Completeness of sampling: 100 % vs. unknown and small (the penetration rate is a few percent, at most)
- Computation requirements: huge (extract trajectories from images) vs. small (just map-match the transmitted coordinates).
- Precision: a few centimeters (depending on pixel size) vs. about 10 m (depending on how many satellite systems are being used).
- Sampling frequency: high (every 100 ms) vs. low (several seconds).

Description of xFCD Data

The sampling interval is typically between 10 ms and 100 ms. At present, only used in probe vehicles for special, mainly scientific, applications:

- Directly measured speed and speedometer reading.
- Sometimes precise DGPS measurements.

- Throttle and/or brake pedal pressure.
- Steering angle.
- Lights on/off left winker on/off, etc.
- Distance gap to the leader or other surrounding objects I relative speed to leader
- Desired ACC deceleration.



Floating-Car (FC) Data: Data sources: Everybody uses a smartphone-based or built-in navigation system (Google Maps); additionally probe vehicles from logistics companies.