How many Floating Car Data (FCD) are needed for Traffic Management?

Peter Wagner
Institute of Transportation Systems
German Aerospace Center, Berlin

4th International Symposium Networks for Mobility
Stuttgart, 25./26. September 2008

from: http://nedbatchelder.com

Deutsches Zentrum
für Luft- und Raumfahrt e.V.
in der Helmholtz-Gemeinschaft

What to expect – content

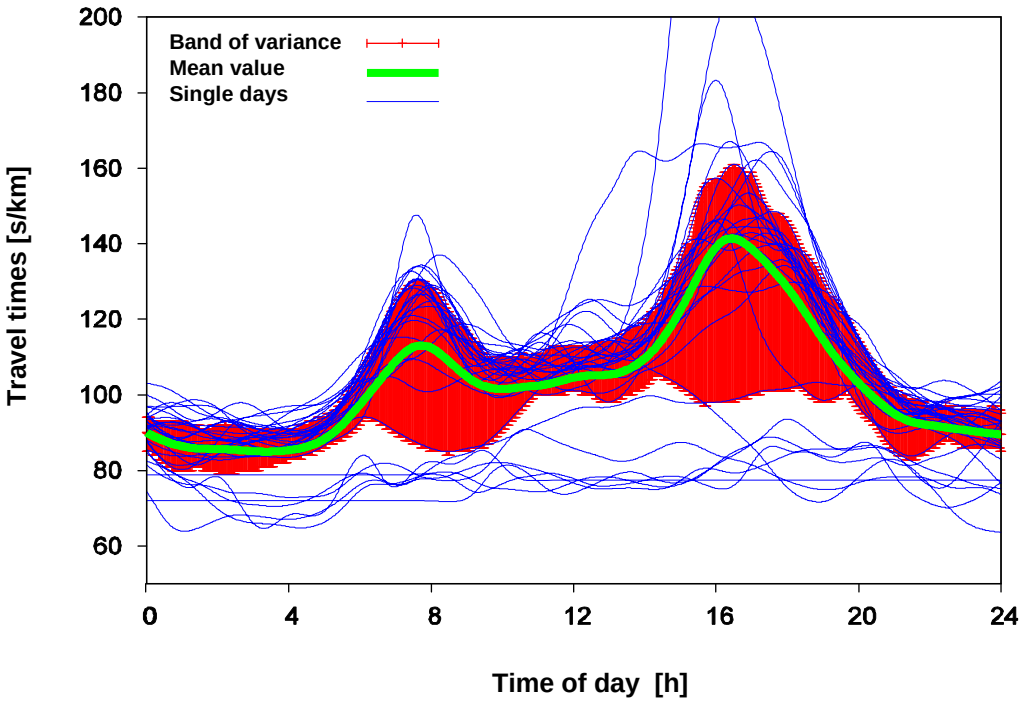
- Introduction: FCD for traffic management
- A short stroll through elementary statistics:
  the story of the standard error
- A few traffic management applications
  – and what FCD can do for them (and what not)
- Summary

Introduction – thesis‘

- Traffic management applications cover a wide range of temporal domains
  - Long-term (days…weeks),
  - Mid-term (hours…days),
  - Short-term (seconds…hours).
- Current FCD fleets can deliver better data for the first two levels:
  - Estimating the queue-length at an intersection,
  - Optimizing a program-based traffic signal control,
  - Less on hard online control.
- The following considerations apply to V2X, too.

Introduction FC Data

- FCD data are being used very successfully for…
  - Traffic surveillance,
  - Navigation & Routing



- Works, because traffic patterns are usually well "reproducible"
  ➔ average travel times can be estimated with ease
- Need to answer: which traffic management applications need which amount of FCD data?

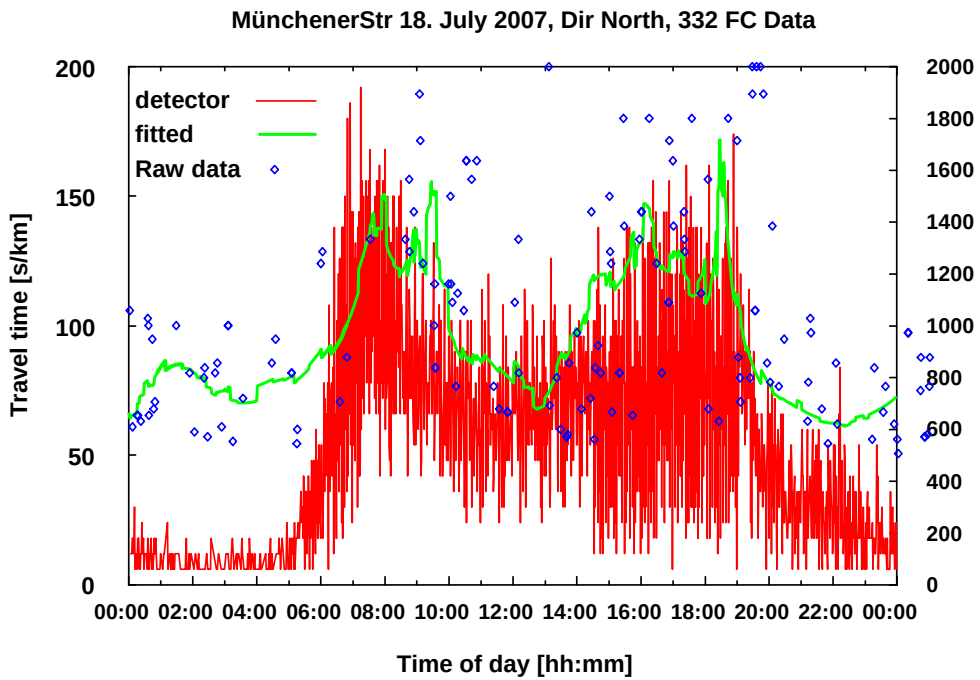
Introduction – is it possible at all?

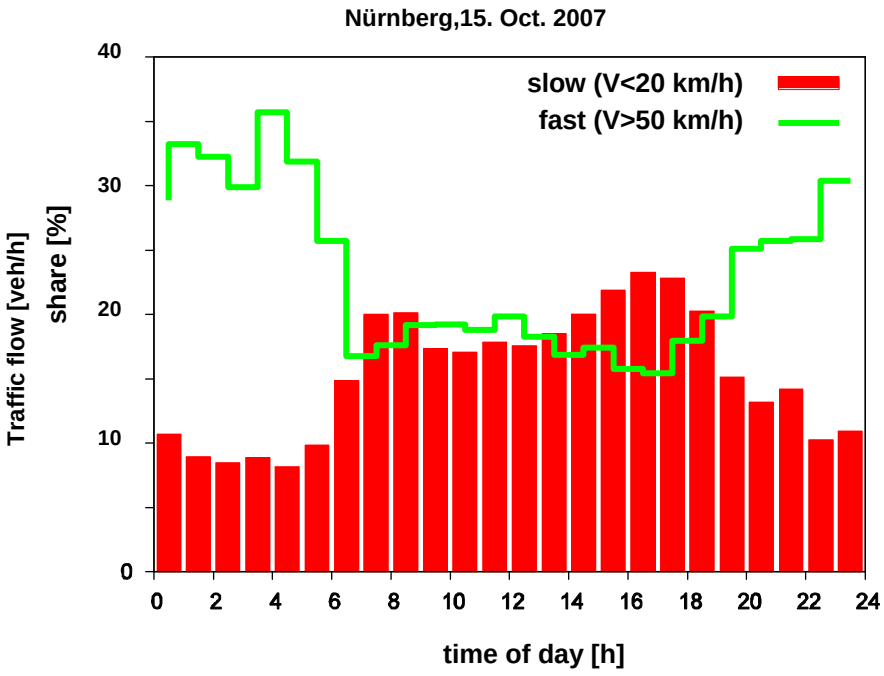
- Pre-condition: traffic management needs a certain information from the data – but what is the appropriate one from FCD?
- Well: FCD data are travel times; do they react on demand at all?
- Two possible answers (of course!): yes and no!
  - Yes: more vehicles need more time e.g. to clear an intersection,
  - No: if the traffic signals are optimally adapted to the demand, no change in travel times could be noticed.
  - (in theory, that is one reason why they are there)
- (Believe it our not: we have seen a clear tendency towards the latter effect in microscopic traffic simulations!)

Introduction – Yes, it works!

The plots below demonstrate:

- Travel times are proportional to demand (may be with a certain delay)
- Larger demand increases the number of standing vehicles
- ➔ FCD travel speeds contain important and useable information for traffic management and traffic control.

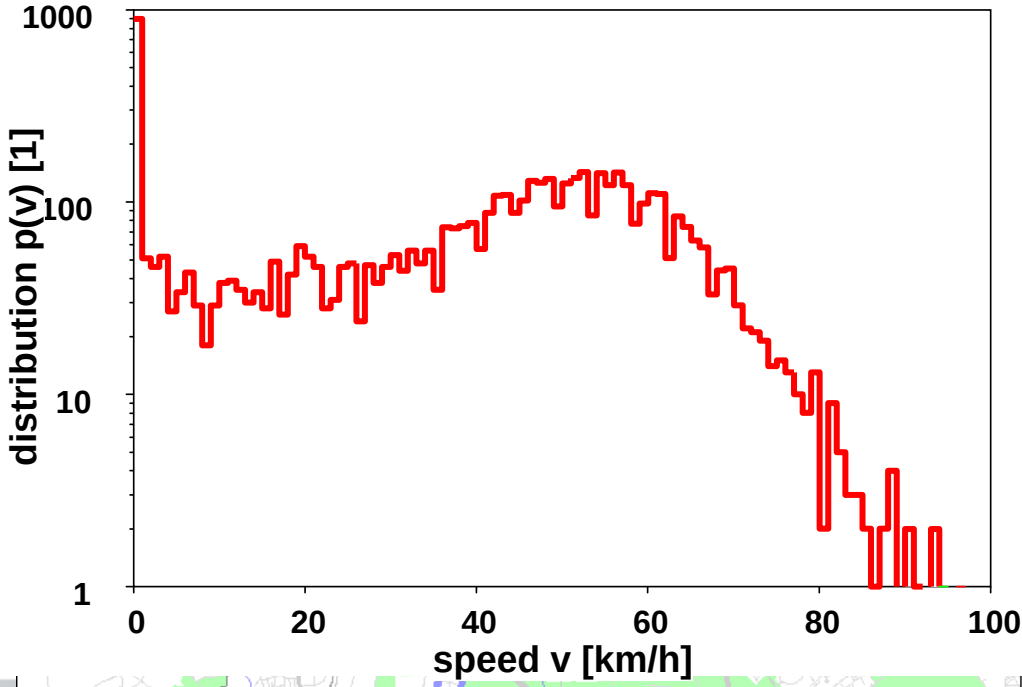




Introduction – Benefits and drawbacks of FC Data

- Advantage: it‘s simple to get an area-wide traffic state (e.g. Nürnberg: 918 links with < 15 min update time with 500 probes)
- Drawback: strong speed scatter anything between 0 and max spe
- Statistics is simple (standard err
- Run Monte Carlo simulations; estimate error as function of
  - Aggregation time,
  - Update time T,
  - Equipment rate η.

- **submitted to TRB 2009**
  Diploma thesis Sascha Krieg (2008)



Statistics of FCD data

Statistics of FCD data

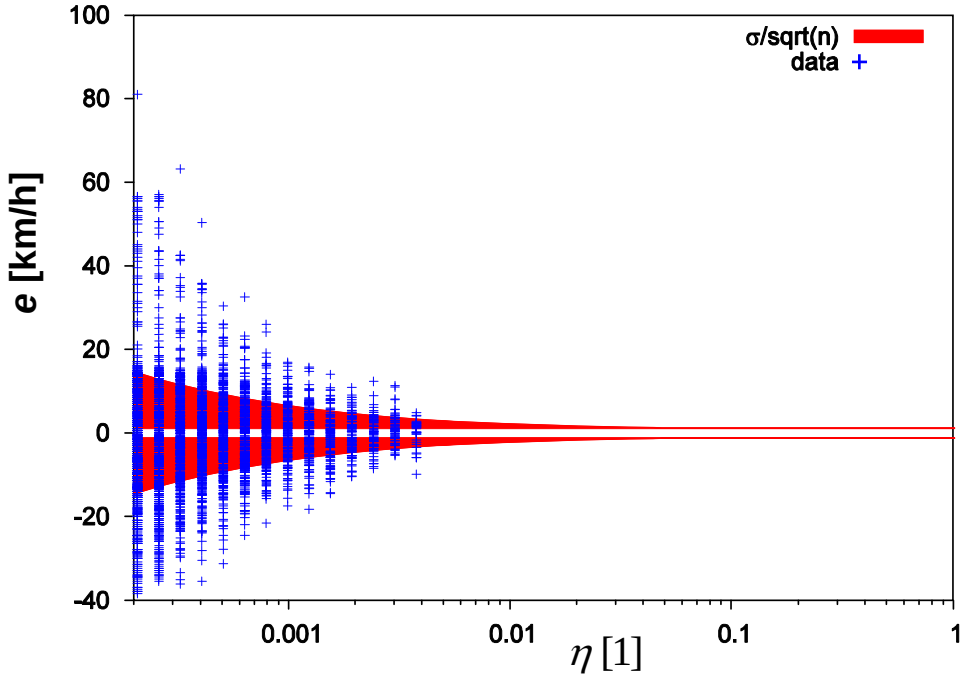
- Assume a stationary situation with a true average travel speed *V* generated by N vehicles
- For a sample $n = \eta N$ of data from this situation, the mean can be computed:

$$\overline{V} = \frac{1}{n}\sum_{i=1}^{n} V_i$$

- repeated drawing yields the blue crosses
- difference between sample mean and true mean is:

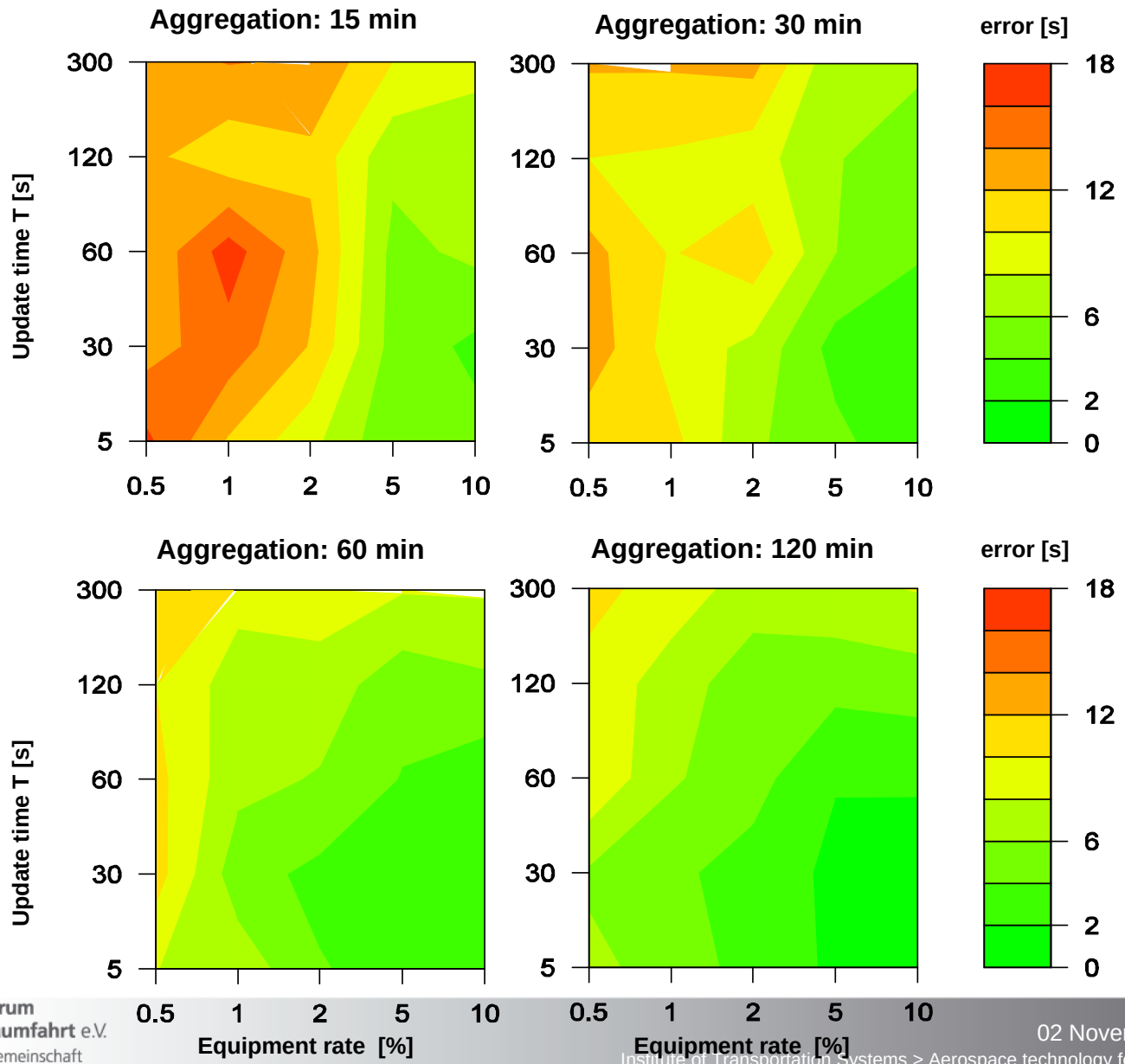
$$e = \overline{V} - V = \frac{\sigma_V}{\sqrt{n}}$$

- …the standard error!



Analyzing a real fleet with microscopic simulation

- Simulation of Nürnberg's VLS area: detailed infrastructure, time-dependent demand, a very well fitted simulation which is close to reality
- A fleet of FCD vehicles, which send any T seconds their current position to a server
- We can vary the equipment rate η to any desired value
- For any pair of (η,T) the simulation samples at least 10 different values to estimate the difference between the true and the processed travel speed / travel time
- results depend on the aggregation time, i.e. how long the data are sampled from a stationary situation
- Yields a picture like the following ones

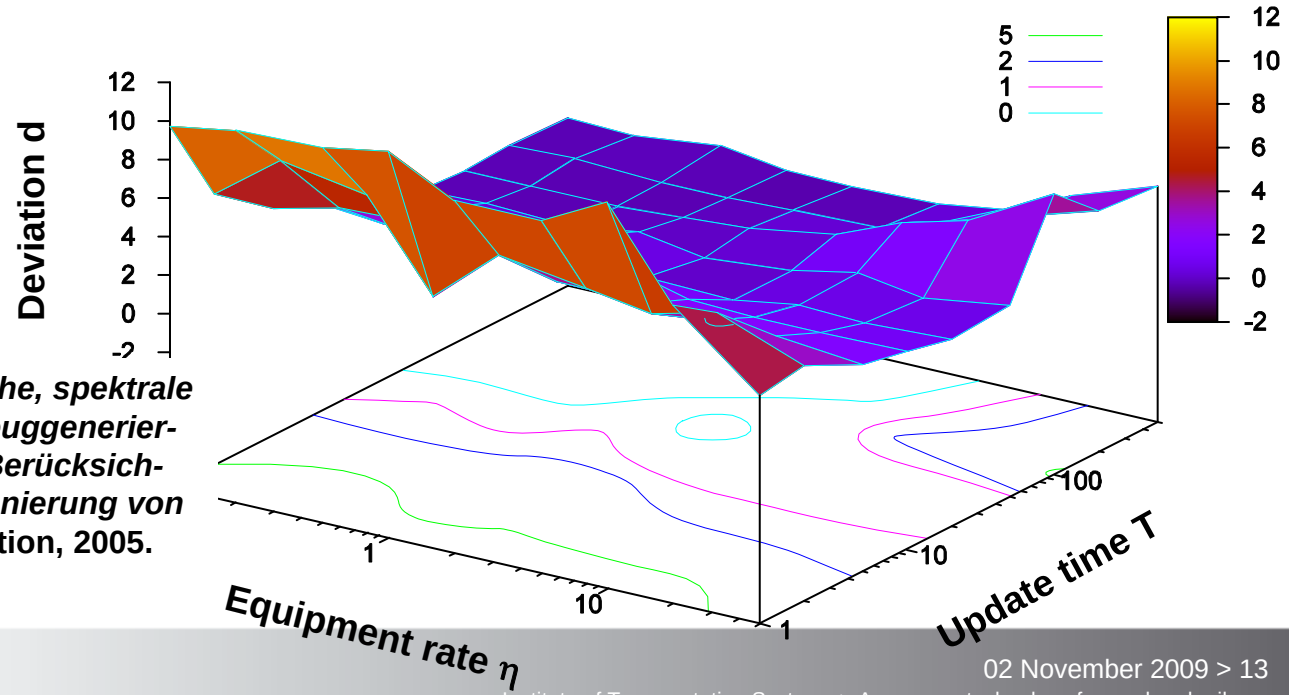

Aggregation: 15 min
Aggregation: 30 min
Aggregation: 60 min
Aggregation: 120 min
error [s]
18
12
6
2
0
Update time T [s]
300
120
60
30
5
0.5
1
2
5
10
Equipment rate [%]

A short “theory”

- What happens here?
  Again, the standard error is the relevant quantity: $e = \overline{V} - V = \frac{\sigma_V}{\sqrt{n}}$
- But *n* is known, it depends on η and *T*: $n = \frac{\eta N}{T}$
- Therefore, the figs above should be described by a simple formula:
- $e = \frac{\sigma_V}{\sqrt{\eta N / T}} = C\sqrt{\frac{T}{\eta}}$
- Let’s see whether this is true or not:

Comparison to simple theory

- Plotting the relative deviation: $d(\eta,T) = \dfrac{e_{\text{sim}}(\eta,T) - e_{\text{stat}}(\eta,T)}{e_{\text{stat}}(\eta,T)}$
- It fits well, except for small T
- but this is understandable: the correlations in the speed effectively decrease *n* ➔ e_{stat} is small, but e_{sim} is not



Gössel, Frank: ***Informationsentropische, spektrale und statistische Untersuchung fahrzeuggenerierter Verkehrsdaten unter besonderer Berücksichtigung der Auswertung und Dimensionierung von FCD-Systemen*****, TU Dresden, Dissertation, 2005.**

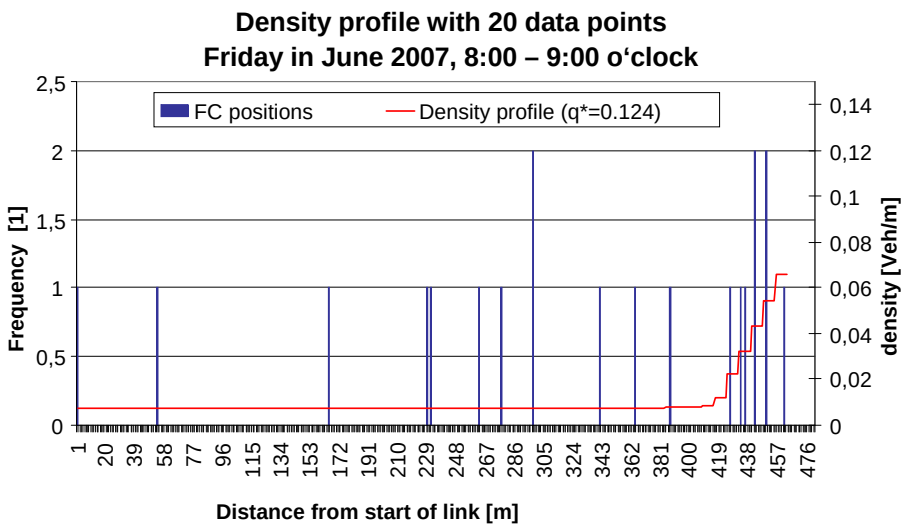
Traffic management applications

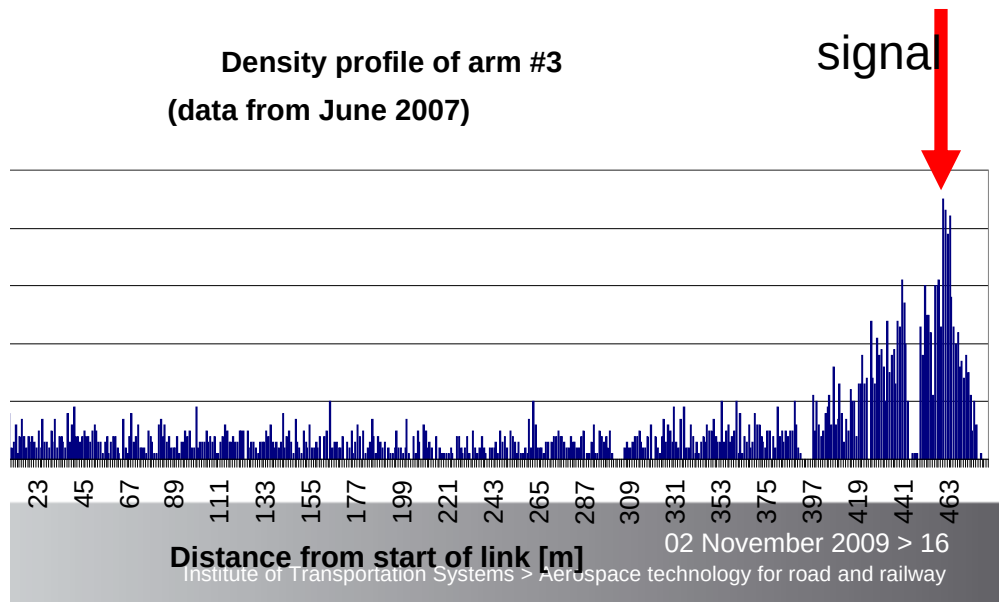
Three levels of traffic management

- Traffic/transport planning (e.g. estimation of average queue-lengthes),
- Program controled: computing better program timings (i.e. which program to use at which traffic situation)
  - offline
  - online
- True online applications in traffic management and control – a few remarks.

Statistical determination of queue lengthes

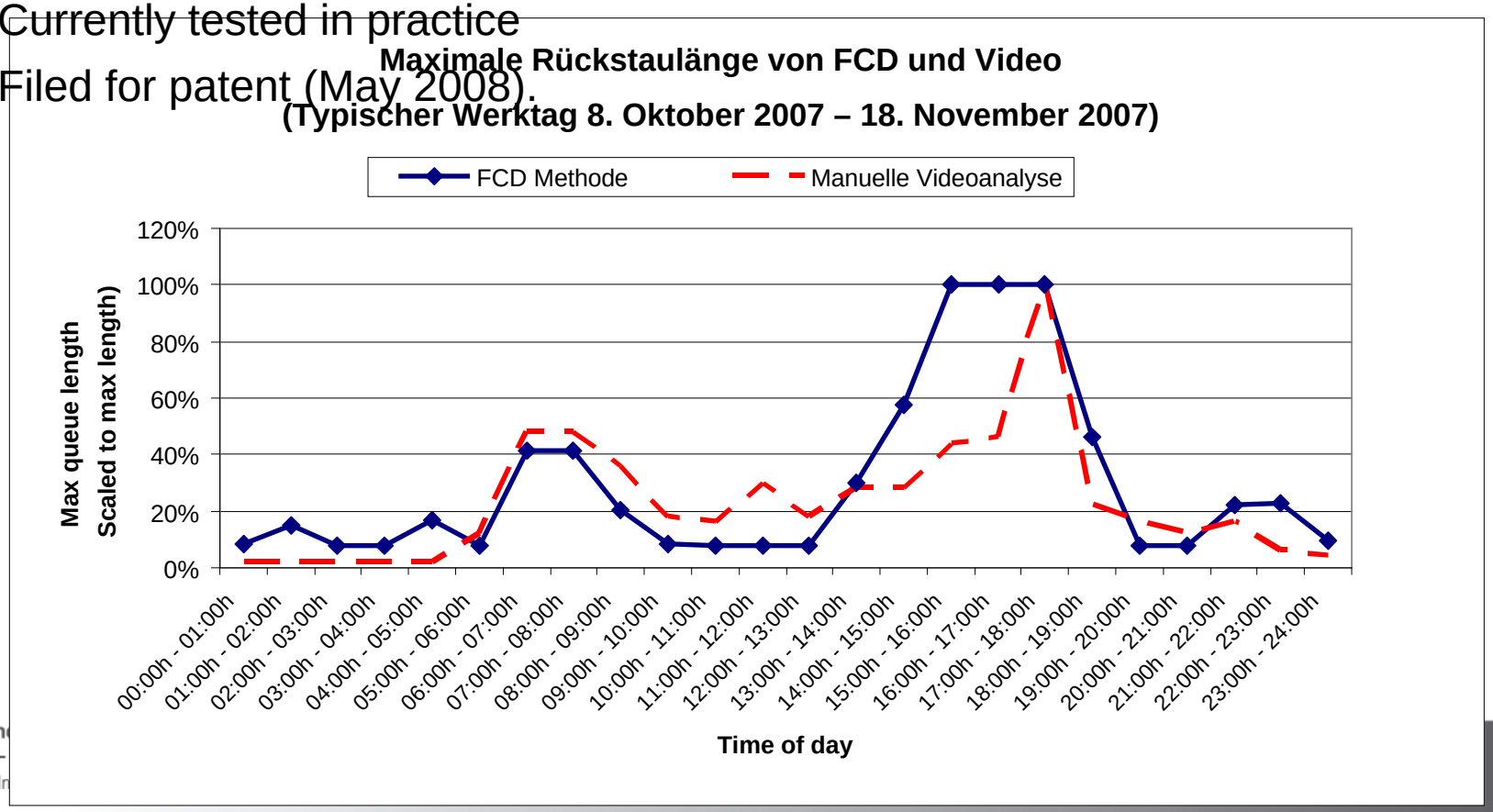
- idea: vehicles have to stop at intersections
- ➔ position data are sampled more frequently near the intersection
- density profile contains important information, even with scarce data…
- E.g. about the average queue length (related to the delays)
- Can be used to improve the surveillance of signal programs





Statistical estimation of queue-lengthes – quality?

- Daily courses look reasonable
- Comparison with real data not simple (here with video)
- Currently tested in practice
- Filed for patent (May 2008).



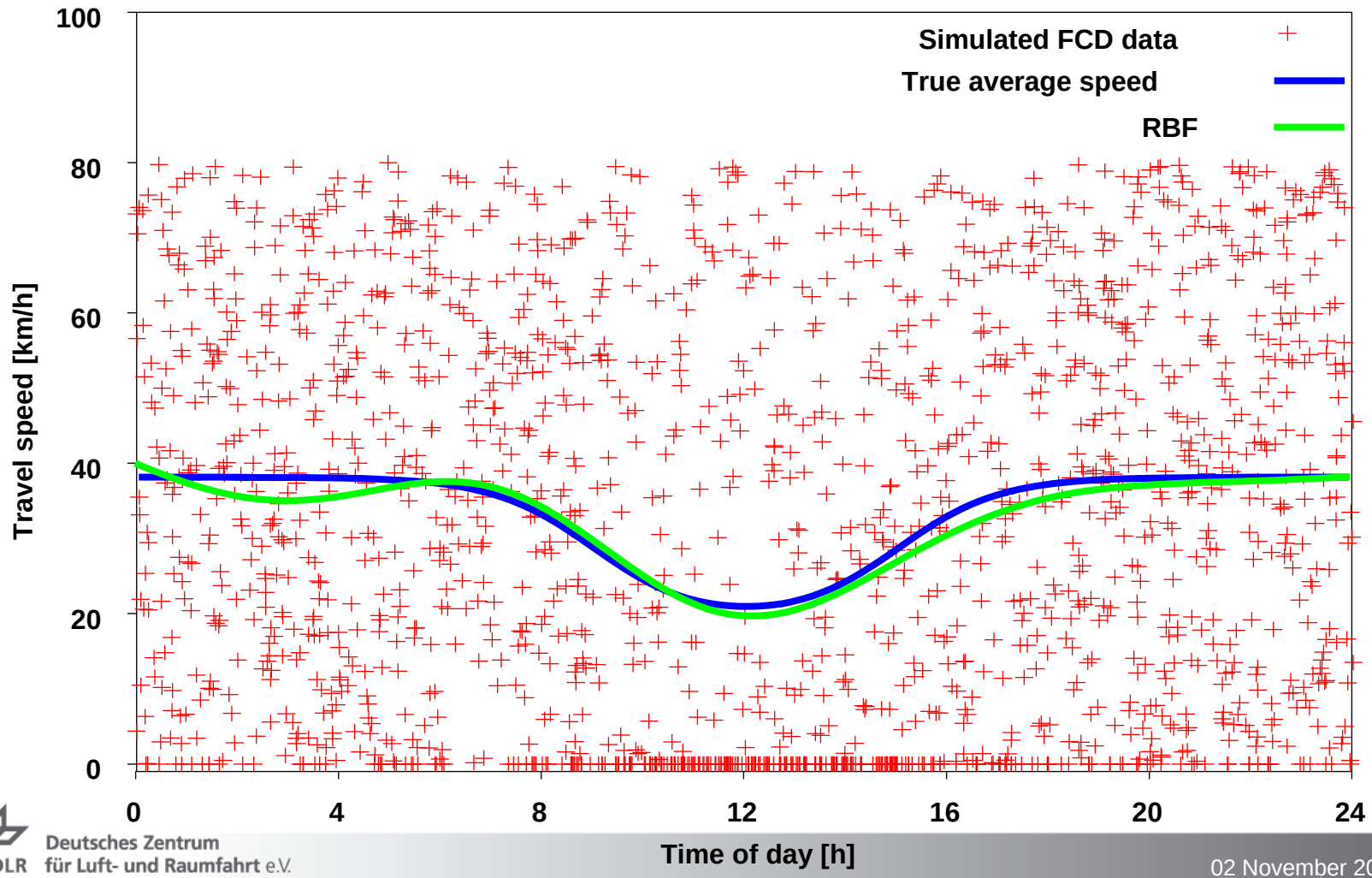
Analysing daily courses

- Simple optimization methods need the daily courses of traffic flow to recognize “patterns“ (morning, afternoon, night pattern to trigger the respective program)
- Big question: is it possible to use the travel time daily courses?
- Yes, but…
- FCD data are a true challenge:
  - They are very noisy,
  - Data are not equi-distant in time (asynchronous)
  - Interpolation is needed to any time stamp

➔ usual methods do not work!

- We have investigated a couple of different methods…

Analysis of daily speed series – how to (1)



Analysis of daily speed series – how to (2)

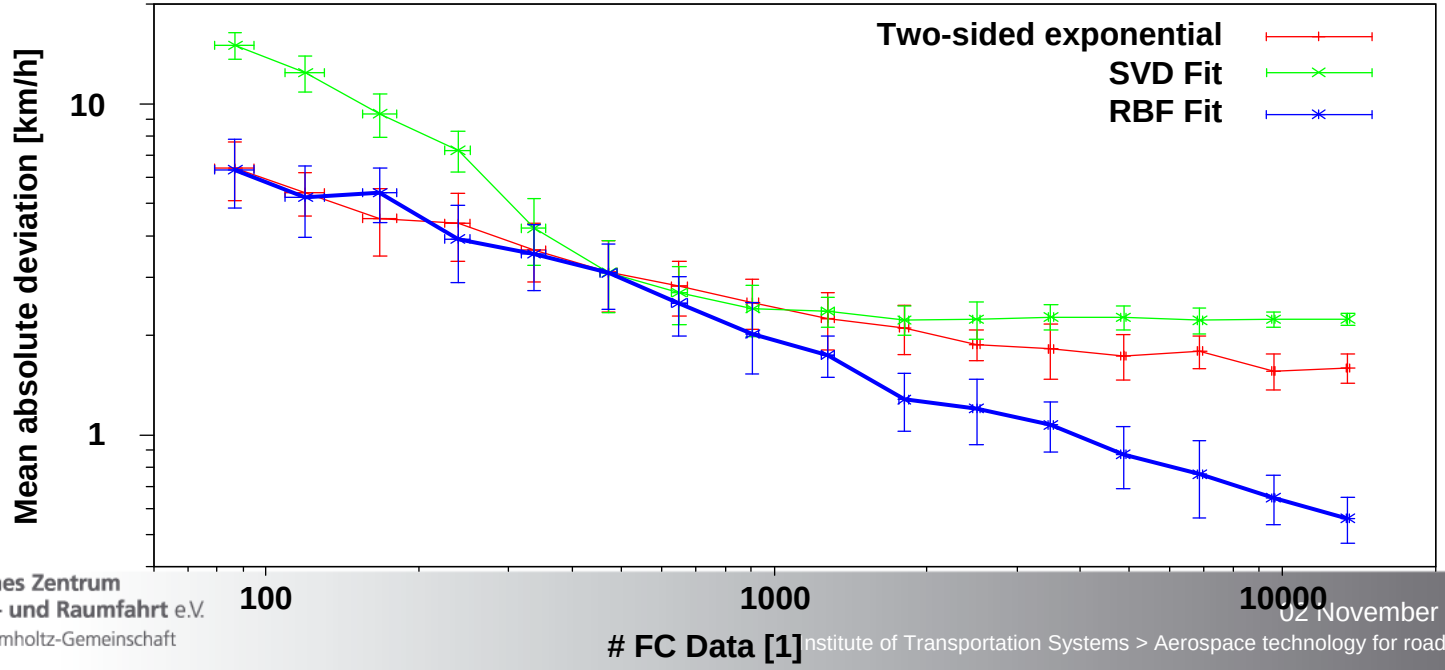
- Local smoothing methods:
  - Symmetric (two-sided) exponential smoothing
  - Local polynomial function fit (SVD fit)
  - Smoothing spline functions
- Global smoothing methods:
  - Lomb periodogramm – which is a Fouriertransformation (not fast) for arbitrary data: yields significance levels
  - Radial basis functions (RBF)

Performance of the methods – offline

- Global methods use the hidden information most effectively
- the best method so far: fit with RBF (radial basis functions)

$$V(t) = V_0 + \sum_{k=1}^{K} V_k \Phi(t - t_k)$$

- Error as function of the number of data-points (mean sampled over 40 independent repetitions) is close to the theoretic optimum (standard error)



Final goal: online application
(this relates to V2X, too!)

- Online methods work with single traffic signals (actuated control, micro-BALANCE, OPAC,…)
- Or with groups of signals (MOTION, SCATS, SCOOTS,…)
- They need uptodate information about headways, flows, speeds, delays
- it is not known how these methods will work with spotty data:
- actuated control needs the time headway which is not available with 10% equipment rate,
- But: if the vehicles simply communicate their experienced delay, even less than 10% equipped vehicles might suffice (this works)
- Put differently: the result depends strongly on the method

➔ there is a strong need for R&D here; which is exactly the topic of V2X!

Summary, outlook, and future prospects

- We have shown, that the statistics of FCD is basically the standard error
- Even a small amount of data (equipment rates well below 1%) can be used for planning and for offline traffic management
- Simple online approaches seem to be within reach with 1% equipment rate – still some tweaking is needed
- True online traffic management and control need equipment rates well above 5% – and other, more appropriate methods
- FCD progress in practice is slow; not only because of this difficult data source, in addition people are careful in accepting it
- However, there are already ideas around to base at least the quality assessment of traffic management on the distribution of “lateness” – this is already used in public transport, why not in individual traffic and in traffic management?

Thank You for listening! Any questions?

From: http://veryspatial.com/wp-content/uploads/2008/05/questions.jpg

Thanks to my colleagues Elmar Brockfeld, Rüdiger Ebendt, Sascha Krieg, Thorsten Neumann, Sten Ruppe, Alexander Sohr, and Louis Touko for making this possible

Summary and outlook

- progress has been made in the appliance of FC Data for traffic management:
- 500 probe vehicles provide sufficient data density and quality for doing statistical queue-length estimation ➔ surveillance of the signals may become easier
- Several methods to estimate stable and reproducible daily courses are in stock and almost ready for practioneers
- Statistics o.k.: these applications need an equipment rate of 1% (or, better, 100 data / day); true online applications need a fleet which is 5 to 10 times bigger
- But: using travel times instead of traffic flows in practice is a slow process
- But: the lateness, which is used for measuring the public transit‘s quality might a good measure of quality for both modes *)

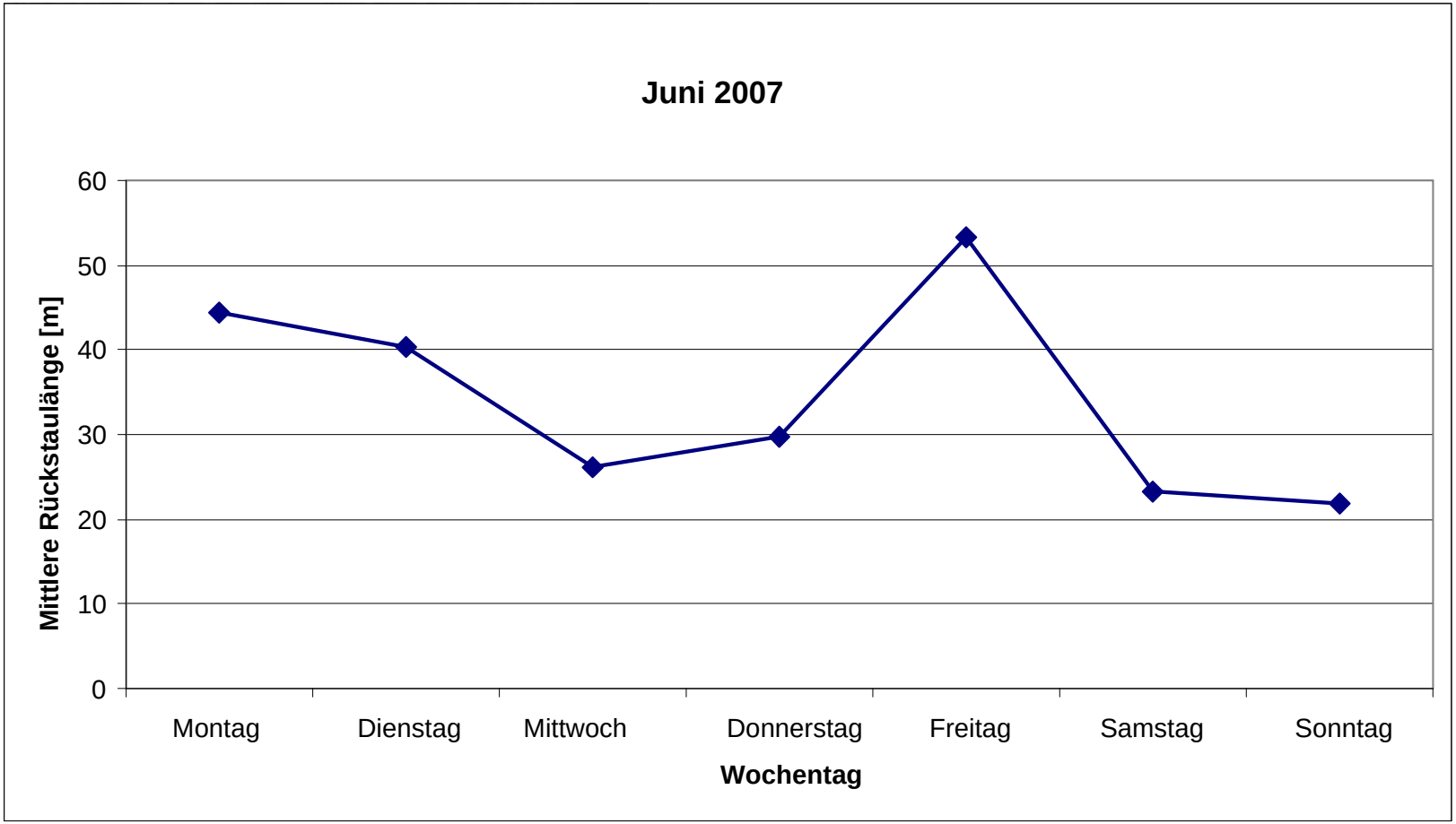
)* proposal by Markus Friedrich

Summary and outlook (2)

- The methods described here still need improvement:
- Queue-length estimation has difficultied with instationary data
- Interpretation of the data for use in practice have to be detailed (what does this result mean?)
- Online methods: have to improve the prediction methods
- Still a lot of work to be done to come to grips with online control: a highly interesting research topic!

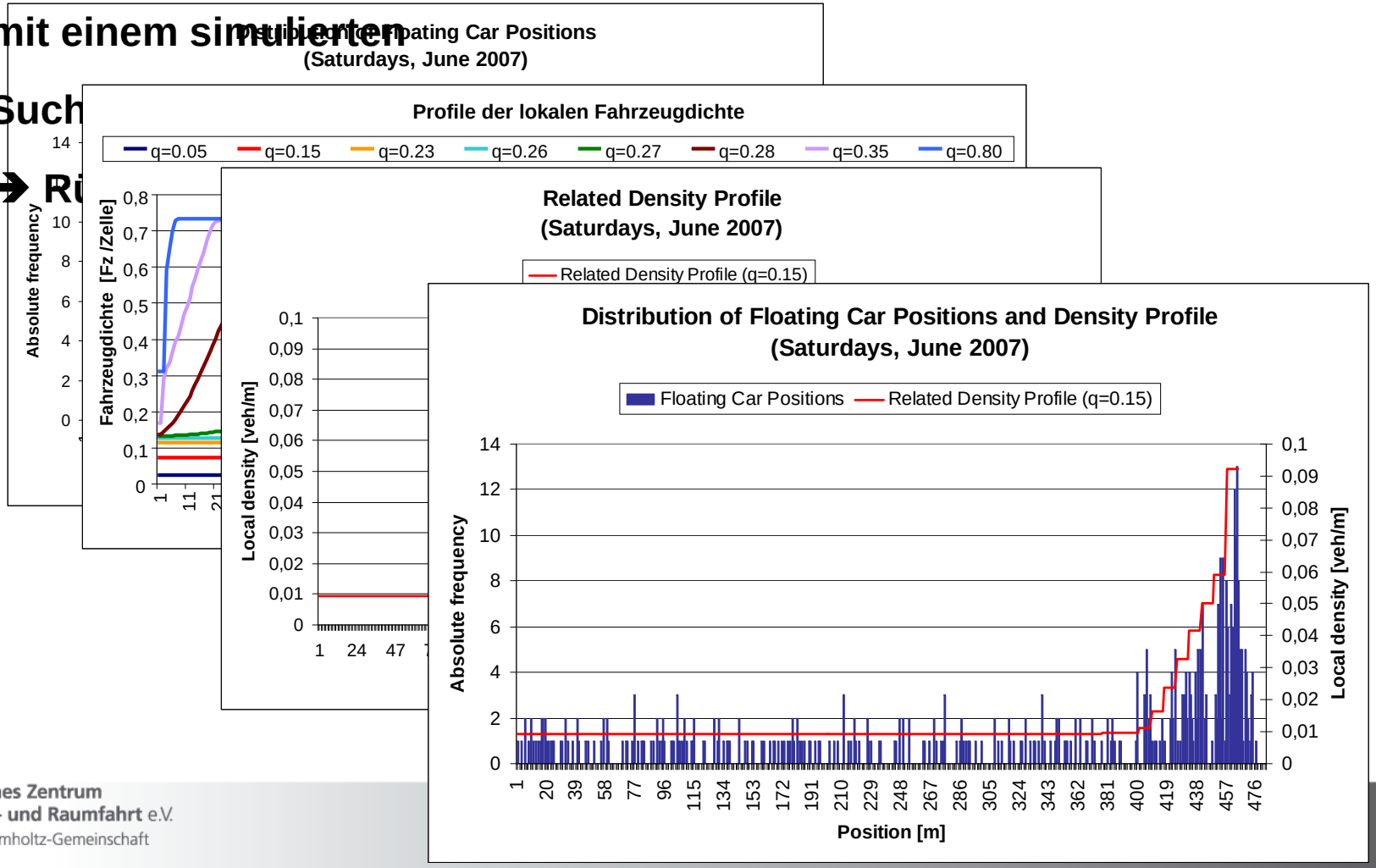
Statistische Rückstaulängenschätzung – Ergebnisse

- **B4 / Nordring**



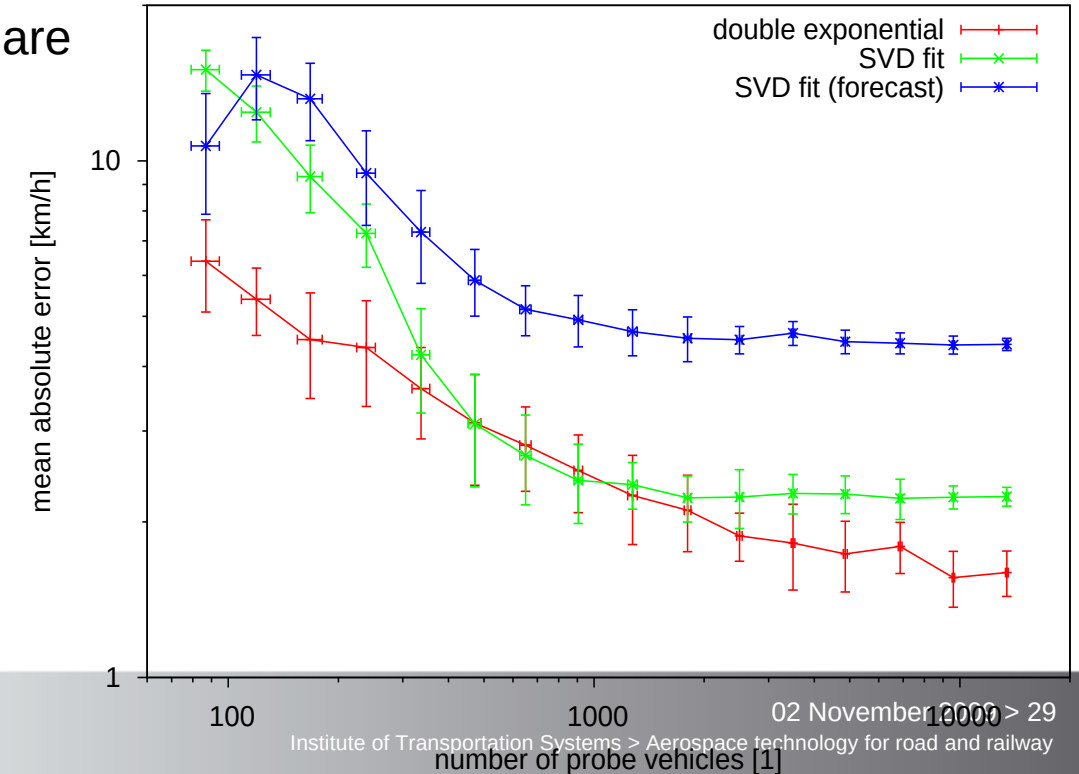
Statistische Rückstaulängenschätzung – wie?

- Vergleiche das empirische Profil
- mit einem simulierten
- Such
- ➔ Rü



Performance of the methods – online

- So far: analysis was offline; need to make it online to recognize the prevailing traffic pattern and react to it. Two difficulties:
- FC Data have an inherent delay of a couple of minutes,
- In addition, a short-term forecast is needed
- Work in progress*, here are preliminary results



- * Alexander Sohr, Dissertation (2009).