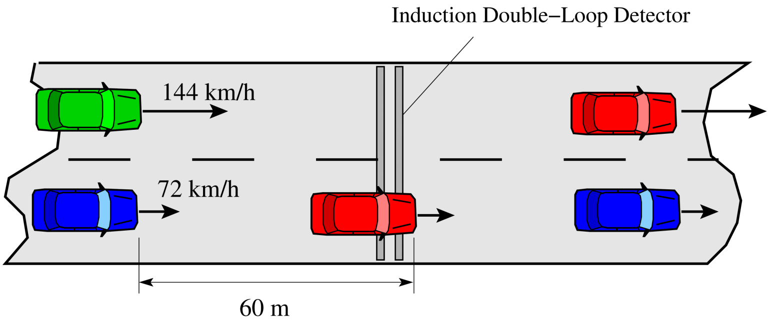
Lecture 03: Cross-Sectional Data Analysis

- 3.1. Estimating Spatial Quantities
- 3.2. Analysis I: Local Flow Characteristics
- 3.3. Analysis II: Time Series
- 3.4. Analysis III: Spatio-Temporal State

3.1. Estimating Spatial Quantities from SDD

Following example shows how biased the arithmetic mean speed and "density=flow/speed" can be when naively estimating spatial quantities:



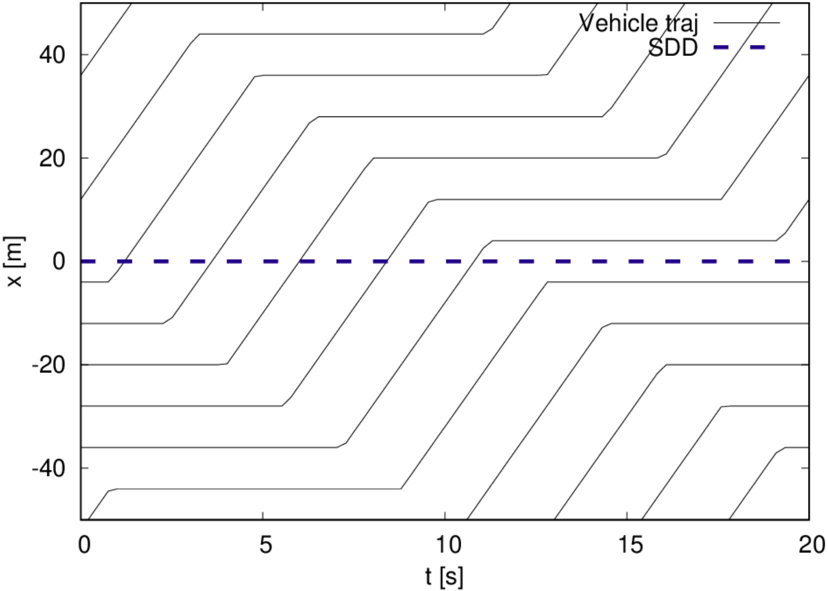
? determine Q^{tot}, V and $\rho^{\text{tot}} = Q^{\text{tot}}/V$ for the total cross section of two lanes and compare with the "true" density and spatial mean speed

Solution: left: $V_l = 40\,\mathrm{m/s}$, $Q_l = 1/3\,\mathrm{veh/s}$; right: $V_r = 20\,\mathrm{m/s}$, $Q_r = 1/3\,\mathrm{veh/s}$;

total: $Q^{\text{tot}} = Q_l + Q_r = 2/3\,\mathrm{veh/s}$, $V = 1/2(V_l + V_r) = 30\,\mathrm{m/s}$, $\rho^{\text{tot}} = Q^{\text{tot}}/V = 1\,\mathrm{veh/(45\,m)}$,

true value: 3 veh/120 m=1 veh/(40 m)

Another example of a bias



Determine flow and speed over an aggregation interval $\Delta t_{\text{aggr}} = 20\,\text{s}$

$Q = 0.25\,\text{veh/s}$,

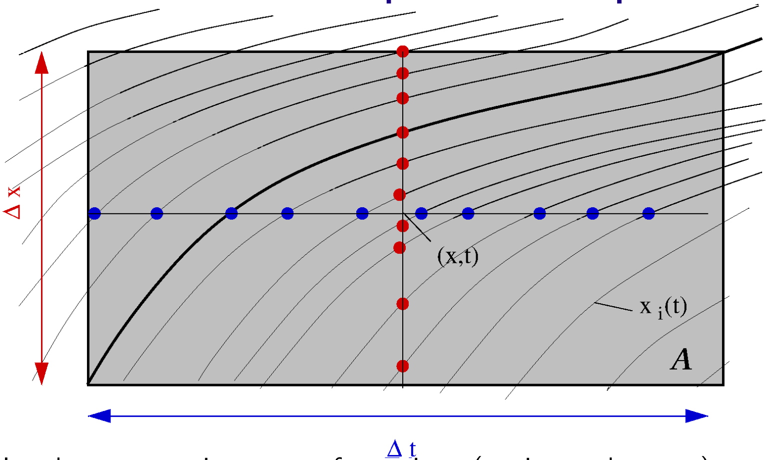
$V = 10\,\text{m/s}$

Compare flow divided by speed with the true spatial density aggregated over 100 m

$Q/V = 25\,\text{veh/km}$,

$\rho^{\text{true}} = 70\,\text{veh/km}$

Definition of macroscopic traffic flow quantities



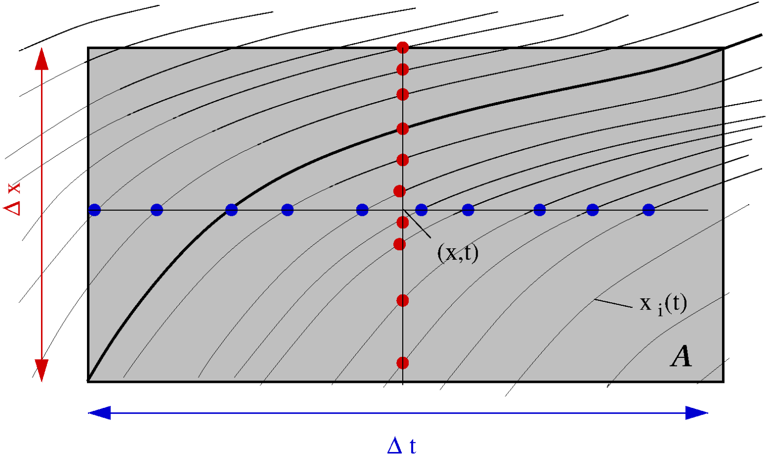
- Local average or time mean of quantity y (stationary detectors)

$$Y = E(y_i|x \text{ fixed}) := E(y_i) = 1/N_{\Delta t} \sum_i y_i(x, t_i)$$

- Instantaneous average or space mean of quantity y (snapshot)

$$Y_s = E(y_i|t \text{ fixed}) := E_s(y_i) = 1/N_{\Delta x} \sum_i y_i(x_i, t)$$

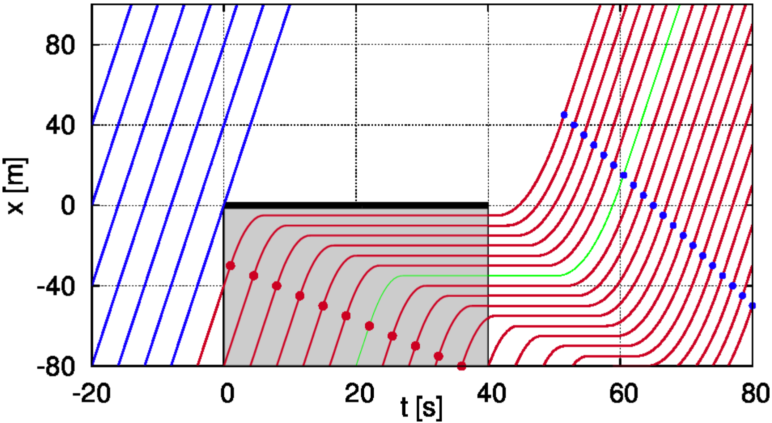
Definition of macroscopic traffic flow quantities II: Edie's definitions



Spatiotemporal mean (Edie's definition) of density, flow, speed:

$$\rho_{\text{Edie}} = t^{\text{tot}}/A, \quad Q_{\text{Edie}} = x^{\text{tot}}/A, \quad V_{\text{Edie}} = Q_{\text{Edie}}/\rho_{\text{Edie}}$$

Problems



? Show that Edie's definition of the speed is just the total distance travelled in A divided by the total time spent in A or, equivalently, the time mean speed over all the trajectories

? Consider the spatiotemporal region $A = [-80\,\mathrm{m}, 0\,\mathrm{m}] \times [0\,\mathrm{s}, 40\,\mathrm{s}]$ and estimate ρ_{Edie}, Q_{Edie}, and V_{Edie} $t^{\text{tot}} = (2+11)/2\ 40\,\mathrm{s} = 260\,\mathrm{s}$, $x^{\text{tot}} \approx 10 * 40\,\mathrm{m} = 400\,\mathrm{m}$,
$\rho_{\text{Edie}} = 260\,\mathrm{s}/3\,200\,\mathrm{sm} = 80\,\mathrm{veh/km}$, $Q_{\text{Edie}} = 0.125\,\mathrm{veh/s}$,
$V_{\text{Edie}} \approx 1.5\,\mathrm{m/s}$

? Estimate Q, the time mean V and $\hat{\rho} = Q/V$ in A at $x = -40\,\mathrm{m}$
$Q = 7/40\,\mathrm{s}$, $V \approx V_0 = 10\,\mathrm{m/s}$,
$\rho = 17.5\,\mathrm{veh/km}$

? Estimate ρ and the space mean V_s in A at $t = 20\,\mathrm{s}$
$\rho = 7/80\,\mathrm{m} \approx 87\,\mathrm{veh/km}$,
$V_s = 2/7\ 10\,\mathrm{m/s} \approx 2.9\,\mathrm{m/s}$

Leutzbach relation between space and time mean speed

Assume that the *instantaneous* (spatial) speed distribution has a density $f(v)$. Then,

- the *partial density* of a speed layer is given by $\mathrm{d}\rho = \rho f(v)\,\mathrm{d}v$.
- Since the number of stationary detector recordings (time mean!) is proportional to the flow, the speed weighting of detector measurements is proportional to the *partial flow* $\mathrm{d}Q = v\,\mathrm{d}\rho$:

$$w(v)\,\mathrm{d}v = \frac{\mathrm{d}Q}{Q} = \frac{\rho v f(v)\,\mathrm{d}v}{\int \rho v f(v)\,\mathrm{d}v} = \frac{v f(v)\,\mathrm{d}v}{V_s}$$

- For the time-mean speed V as a function of the space-mean speed V_s, we obtain

$$V = \int v w(v)\,\mathrm{d}v = \frac{1}{V_s}\int v^2 f(v)\,\mathrm{d}v = \frac{E_s(v_i^2)}{V_s}$$

- With the general relation $E(X^2) = \mathrm{Var}(X) + (E(X))^2$ also valid for spatial averages $E_s(.)$, we have $E_s(v_i^2) = \mathrm{Var}_s(v_i) + V_s^2$, so

$$V = \frac{\mathrm{Var}_s(v_i) + V_s^2}{V_s} = V_s + \frac{\mathrm{Var}_s(v_i)}{V_s} \qquad \text{Leutzbach relation}$$

Estimating space mean speed by harmonic averages

- *Both time and space means can be applied to any function y_i of recorded single-vehicle data* such as $y_i = v_i$ or $y_i = 1/v_i$:
  - temporal arithmetic average: $V = E(v_i)$
  - temporal harmonic average: $V_H = 1/E(1/v_i)$
  - spatial arithmetic average: $V_s = E_s(v_i)$
- Derivation of the Leutzbach relation → any expected time average $E(y_i)$ of data y_i can be written in terms of the spatial (!) speed distribution function $f(v)$ via the weighting $w(v)\,\mathrm{d}v = vf(v)/V_{\mathrm{S}}\,\mathrm{d}v$ as

$$E(y_i) = \frac{1}{V_s} \int yvf(v)\,\mathrm{d}v$$

- With $y_i = 1/v_i$, we obtain

$$\frac{1}{V_H} = E(1/v_i) = \frac{1}{V_s} \int f(v)\,\mathrm{d}v = \frac{1}{V_s}$$

$$\boxed{V_s = V_{\mathrm{H}}}$$

The harmonic time mean speed is an unbiased estimator of the space mean speed provided stationarity (in the statistical sense, i.e., $f(v)$ is unchanged over averaging space and time).

Estimating the density by stationary detector data (SDD)

Problem: The density is a spatial quantity but SDs provide temporal quantities.

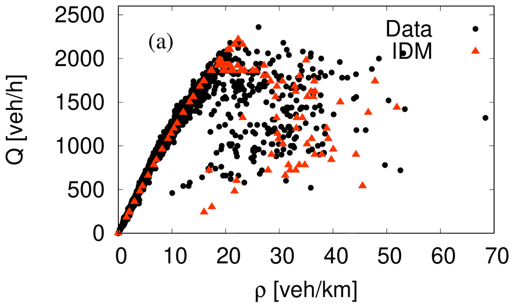
$$
\begin{aligned}
\frac{1}{\hat{\rho}} &= E(d_i) = E(v_{i-1}\Delta t_i) \\
&\approx E(v_i \Delta t_i) \\
&= E(v_i)E(\Delta t_i) + \mathrm{Cov}(v_i, \Delta t_i) \\
&= \frac{V}{Q} + \mathrm{Cov}(v_i, \Delta t_i)
\end{aligned}
$$

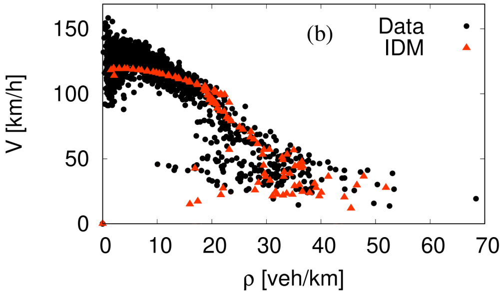
$\Rightarrow$ unbiased estimator $\hat{\rho}$ as a function of the "usual" estimator $\rho = Q/V$:

$$
\boxed{\hat{\rho} = \rho \left(\frac{1}{1 + \rho\, \mathrm{Cov}(v_i, \Delta t_i)} \right)}
$$

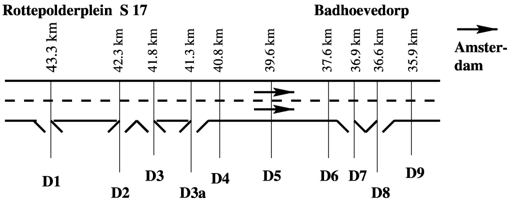
? Show that the expected true density $\hat{\rho}$ is generally underestimated by $\rho = Q/V$. In which situations this bias becomes pronounced? Hint: what is the expected sign of this covariance?

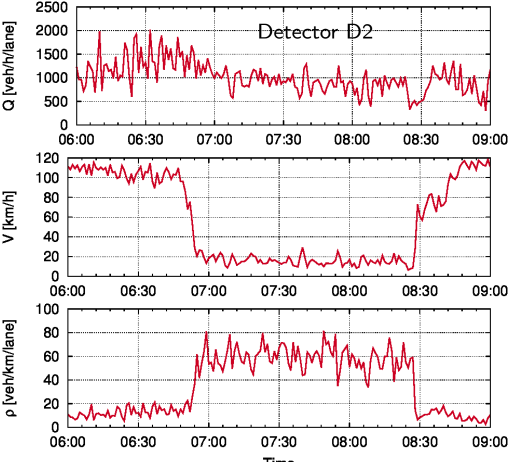
Comparison with a model reveals systematic bias

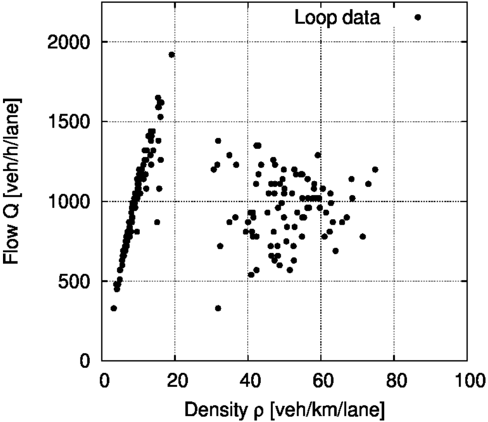




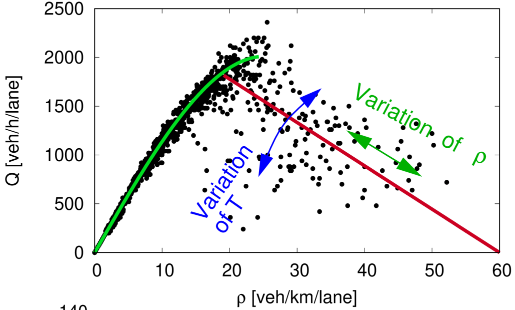
3.2. Analysis I: Local Flow Characteristics





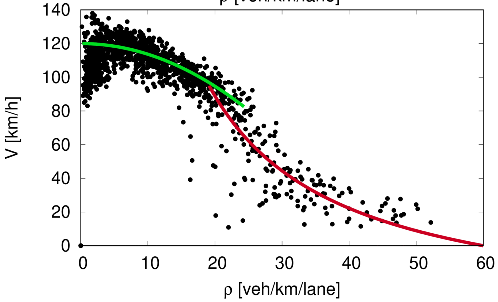


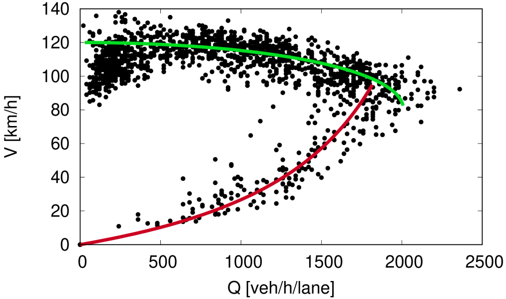
Flow-density-speed data and fundamental diagram



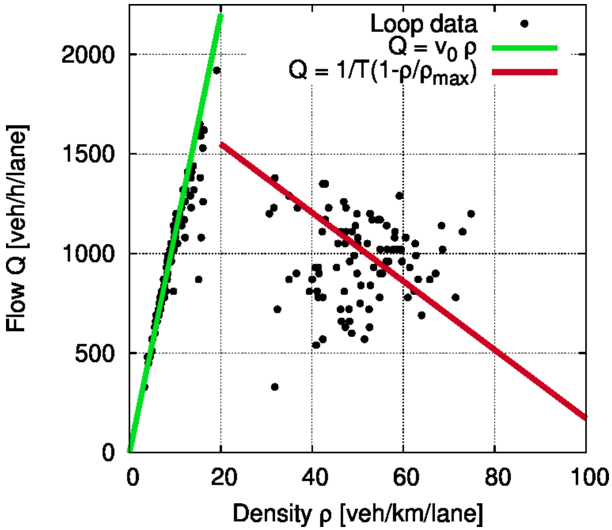
Free and **congested** Regimes:

Which minimum time span is plausibly covered?



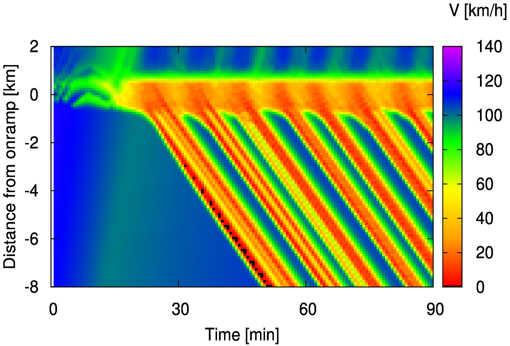


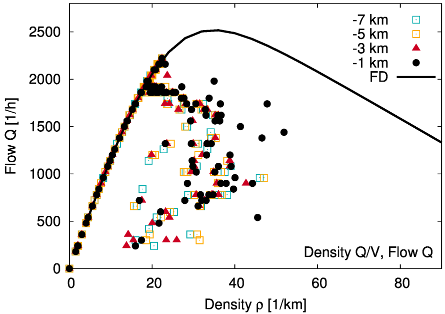
Why is the fundamental diagram so "fundamental"

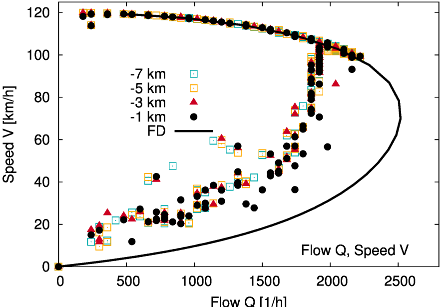


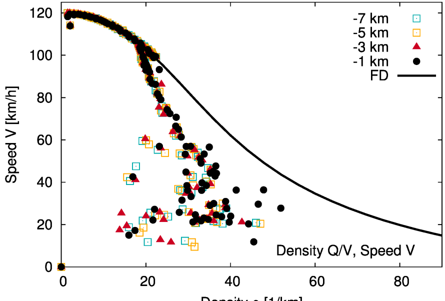
- **free traffic**:
  $Q(\rho) = V_0 \rho$
- Intersection with the abszissa:
  $\Rightarrow l_{\text{eff}} = 1/\rho_{\text{max}}$
- Slope $w = Q'_{\text{c}}(\rho) = -l_{\text{eff}}/T \Rightarrow$
  wave speed w, time gap T
- Intersection with Q_{f}:
  $\Rightarrow$ estimate for capacity
  $Q_{\text{max}} = V_0/(V_0 T + l_{\text{eff}})$

Bias check I: flow Q, speed V, density $\rho = Q/V$

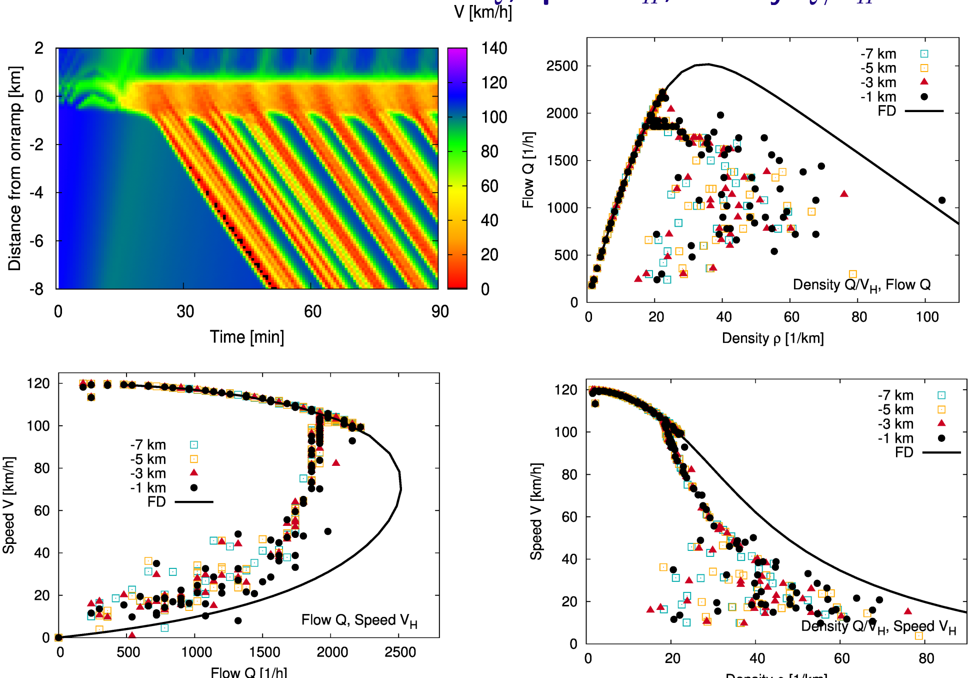




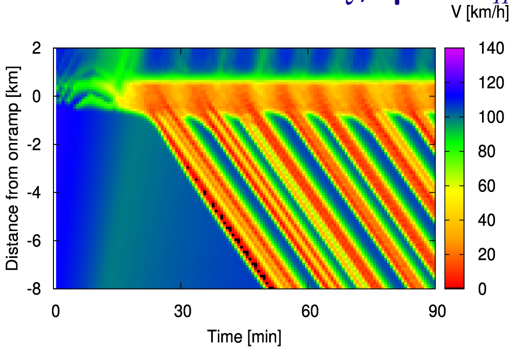


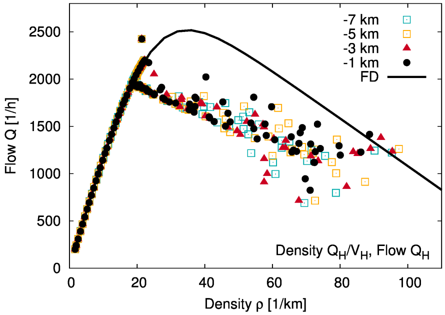


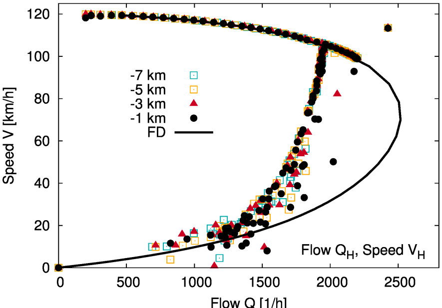
Bias II: flow Q, speed V_H, density Q/V_H

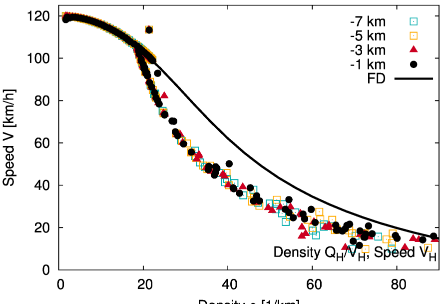


Bias III: flow Q, speed V_H, density $\rho = Q_H/V_H = E(1/\Delta t_i)/V_H$

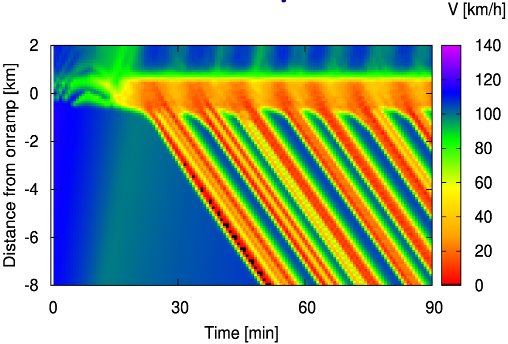


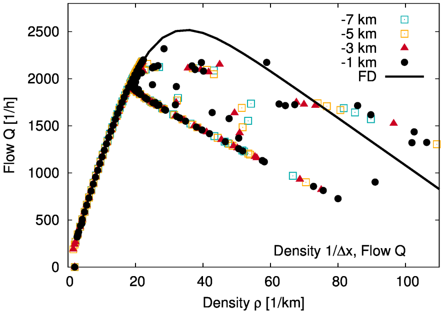


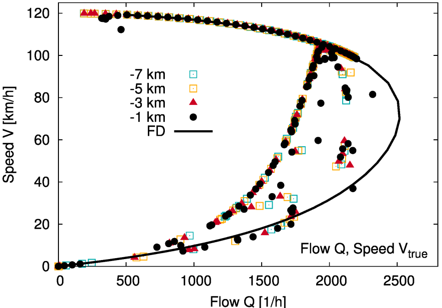


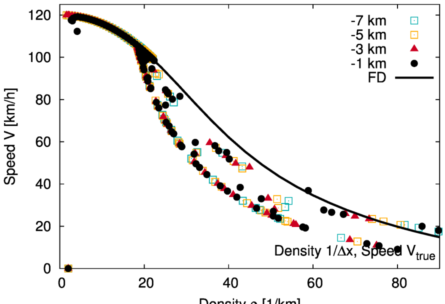


Real spatial local density, spatio-temporal local speed

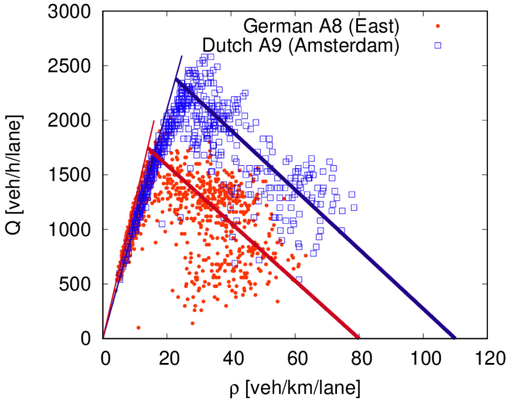


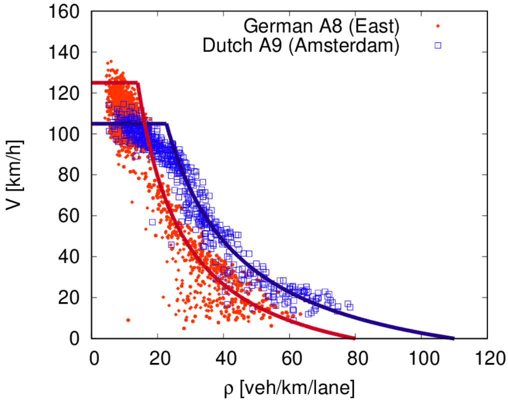






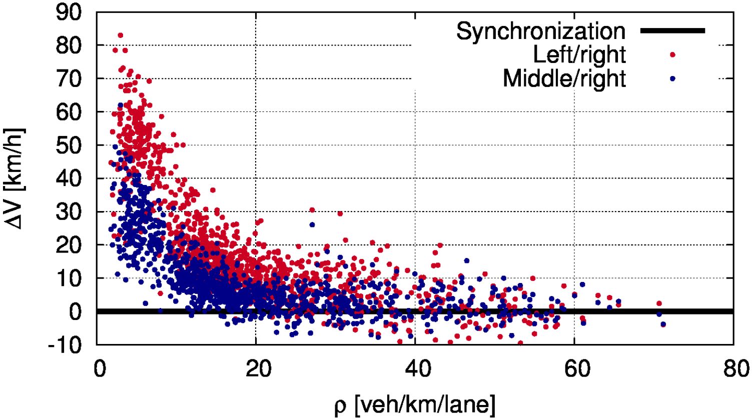
Regional and infrastructural differences





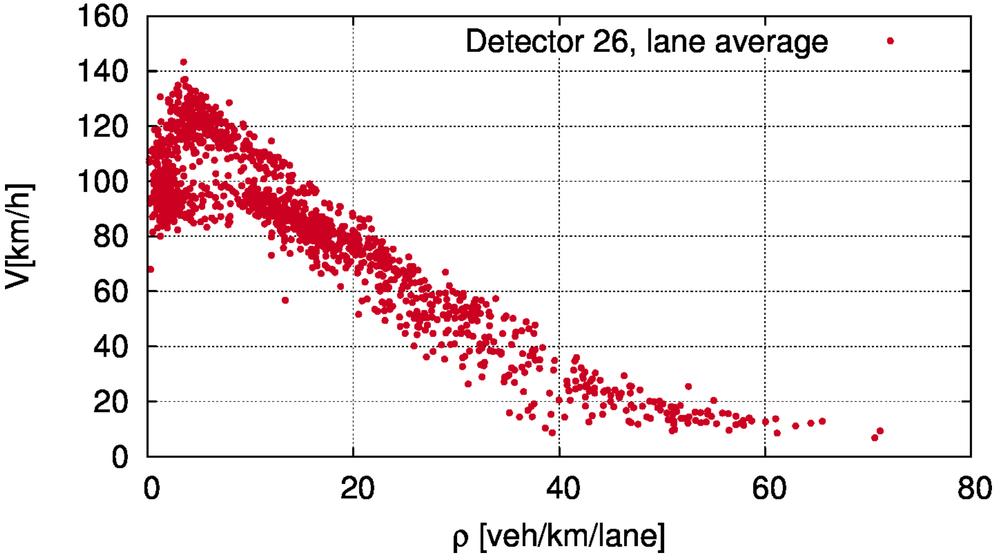
- German A8-East near Munich: Higher maximum speed and lower capacity compared to the Dutch A9 near Amsterdam

Speed synchronisation across lanes



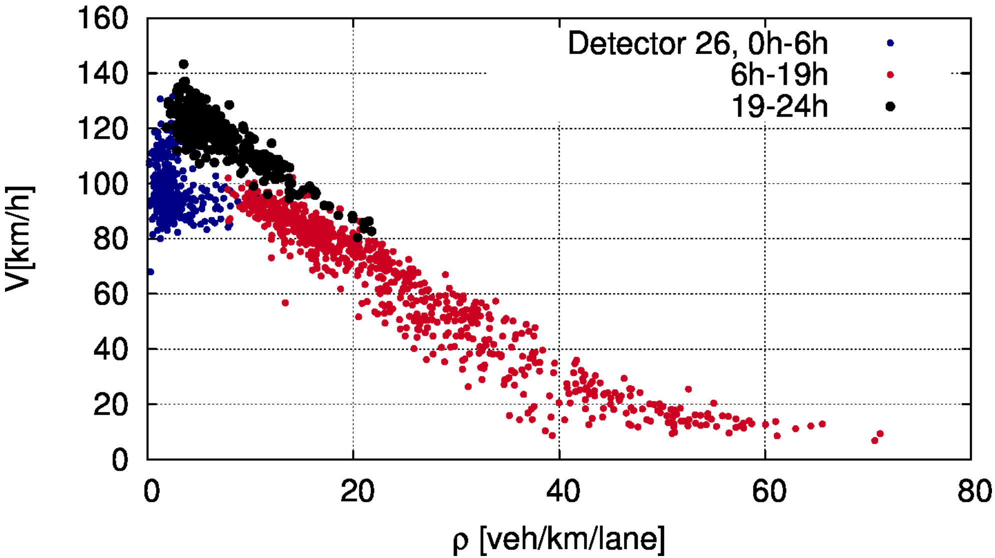
- Low densities $\rightarrow$ little interactions $\rightarrow$ nearly everybody can drive at his/her desired speed chosing the suitable lane ("fast", "middle", or "slow") $\rightarrow$ Δv large;
- densities near capacity: still no congestion but much interaction $\rightarrow$ small Δv values;
- jammed region: $\Delta v \approx 0$ ("in a jam, everybody is equal")

Horror Vacui



? Is there a "fear of the empty Autobahn" (*Horror Vacui*)?

Horror Vacui explained: Simpson's Effect/Paradox



! Different weather and/or traffic composition and/or speed limits in the three time intervals → **Simpson's Effect**

Problem: Simpson's effect in local flow characteristics

? In traffic flow data, Simpson's effect is relevant if the time variable is eliminated such as in speed-density scatter plots. Why?

! because (i) the vehicle and driver composition, visibility/road conditions and possibly traffic regulations change during the daytime, (ii) these changes are correlated with flow, density, and speed $\Rightarrow$ sampling of heterogeneous data with heterogeneities correlated to the variables of interest $\Rightarrow$ Simpson's effect.

? Assume (i) at night a density of 1 veh/h/lane and traffic consisting to 50 % of trucks (temporal average!), (ii) before the rush hour (still little interactions) a density of 10 veh/h/lane and 10 % of trucks. Plot the corresponding two speed-density points demonstrating the apparent *horror vacui*. Assume as (average) desired speed 120 km/h for cars and 80 km/h for trucks and a reduction of 10 km/h for cars in case (ii).

! (i) At night: $V = 0.5(80 + 120)\,\mathrm{km/h} = 100\,\mathrm{km/h}$
(ii) before the rush hour: $V = 0.1 * 80 + 0.9 * 110\,\mathrm{km/h} = 107\,\mathrm{km/h}$

At night, the average speed is lower than before the rush hour although neither vehicle type drives more slowly and the cars even faster: Simpson;s paradox!

Problem: List of biases in stationary detector data

? Summarize all discussed biases affecting

(i) flow, time-mean speed V, space-mean speed V_S, density ρ estimated by vehicle count and arithmetic speed mean as obtained from SDD,

(ii) speed-density and speed-flow scatter plots obtained from SD data

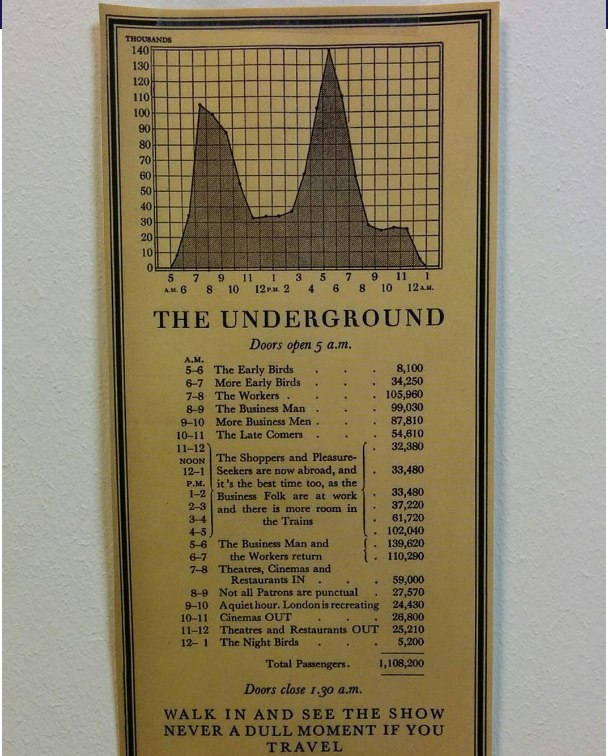
! Time series:

- Flow $Q = n_\text{veh}/\Delta t_\text{aggr}$ and time-mean speed V: none since SDD imply time means
- Space-mean speed V_s: Overestimated by V (Leutzbach relation); would be unbiased if estimated by V_H and there is stationarity in the statistical sense; however, V_H is not available
- True density ρ_real according to Edie's definition: Underestimated by $\rho = Q/V$; unbiased if estimated by Q/V_H and stationarity applies; partial correction for nonstationarity by Q_H/V_H or if $\text{Cov}(v_i, \Delta t_i)$ can be estimated

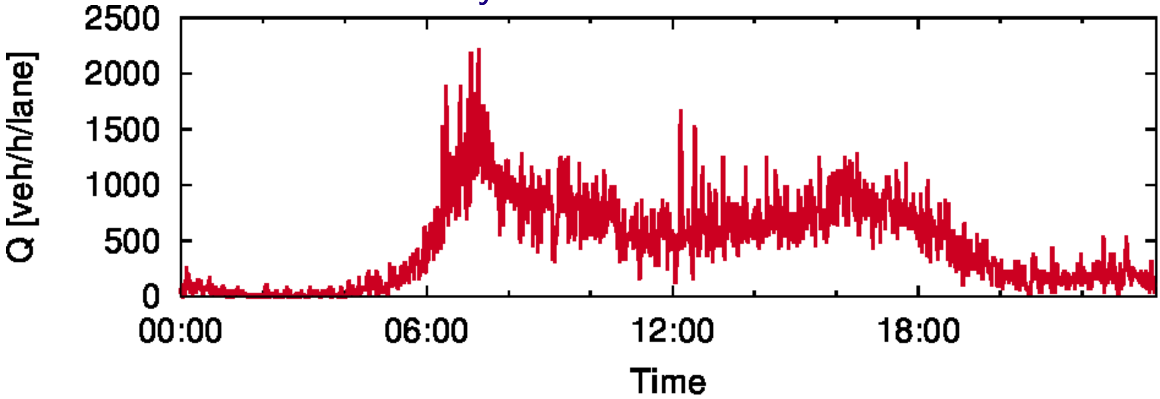
! Scatter plots: additionally, Simpson's effect applies

Notice: random errors are added to all estimates

3.3. SDD
Time Series

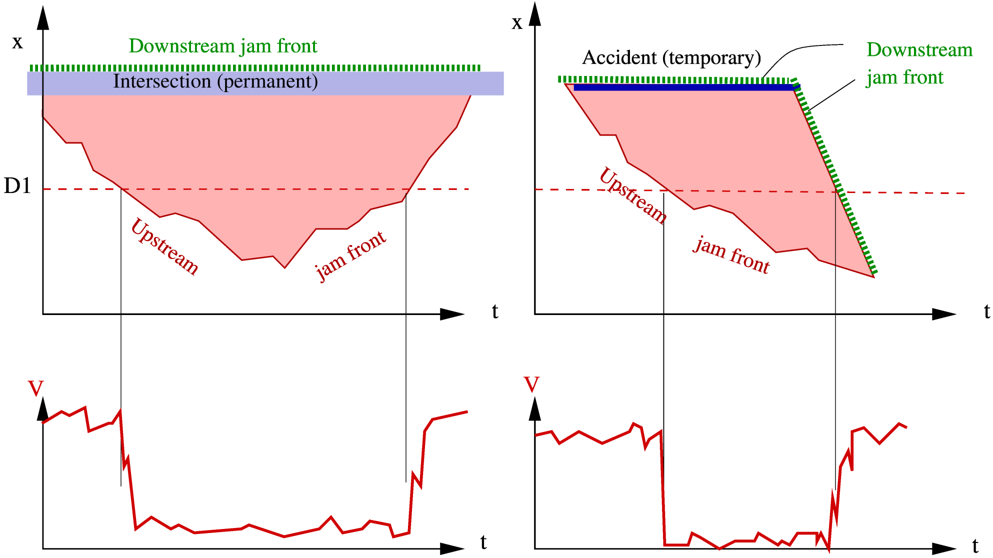


Daily time series of flow



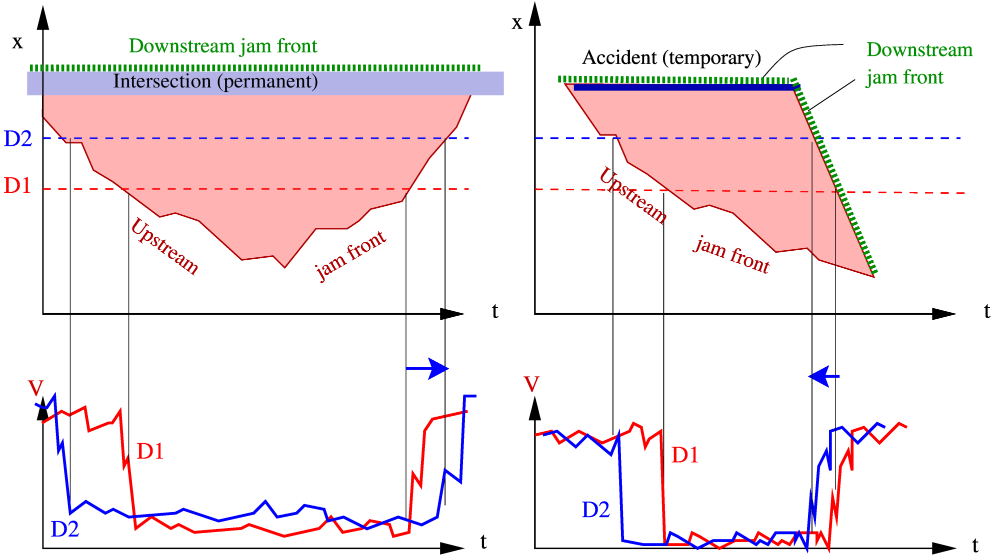
- Only single-loop detectors needed
- Unless there is congestion, this data reflects the traffic demand (why this restriction?)
- Application mainly in transportation/traffic planning and traffic politics ⇒ DTV
- Traffic flow application: **historic data base** to improve traffic state estimation/short-term prediction for dynamic navigation

3.4. Analysis III: Spatio-Temporal State



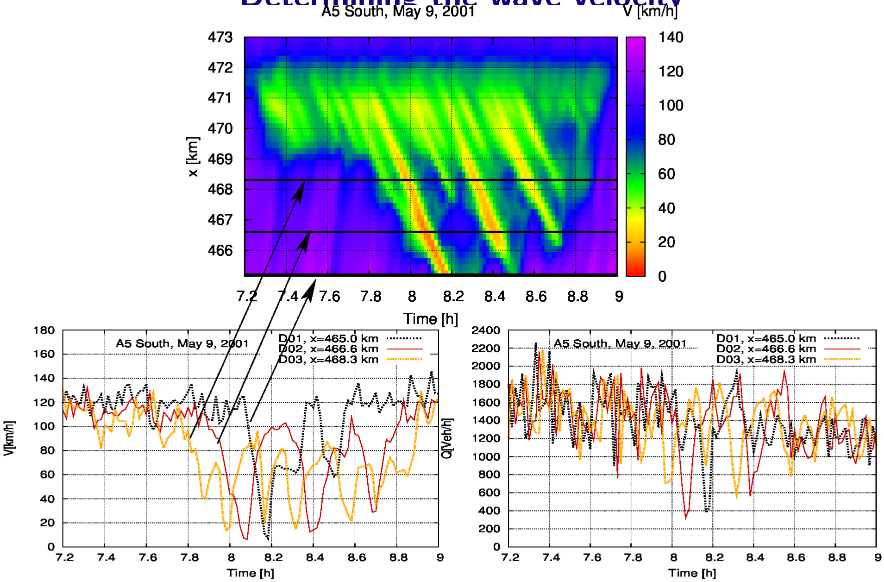
Analysis of a single detector station cannot resolve the upstream-downstream ambiguity when traffic gets free again at the cross section

Resolution by two or more cross sections

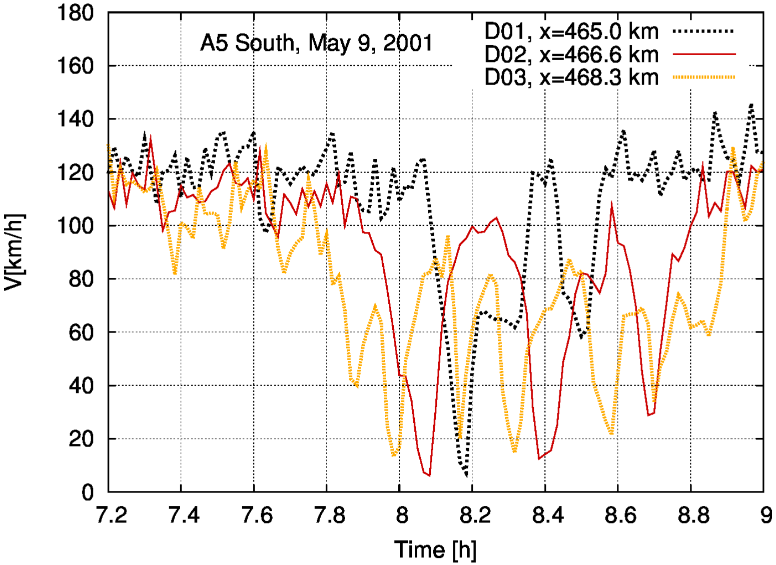


Ambiguity resolved!

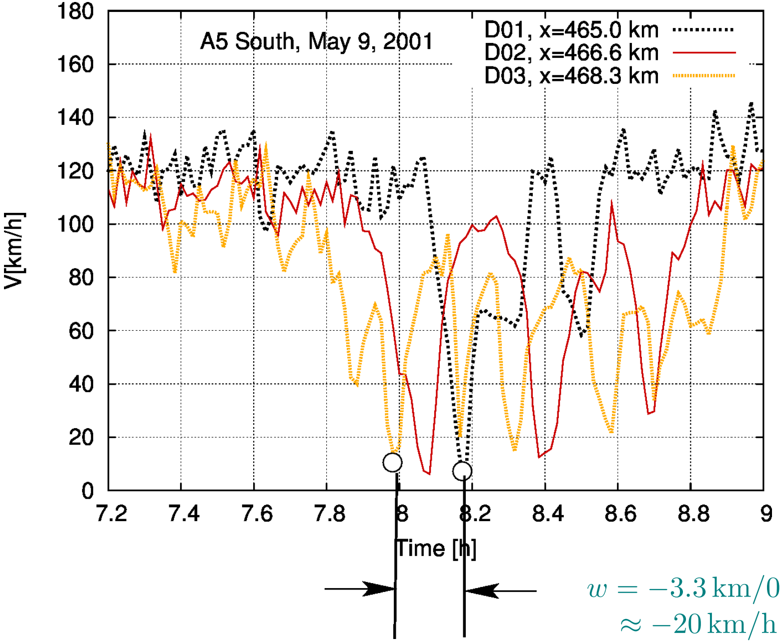
Determining the wave velocity



Problem: Determine the wave velocity by speed time series

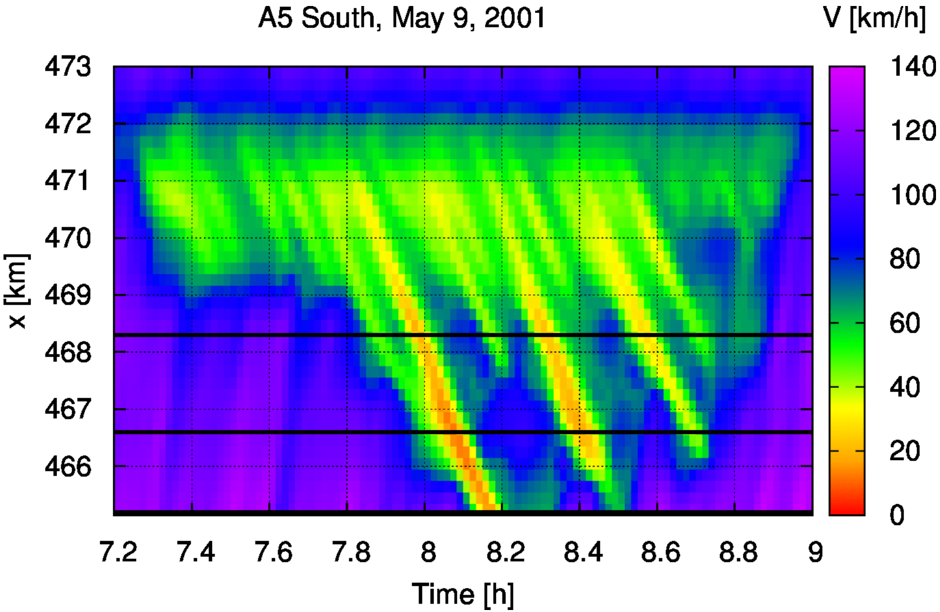


Solution

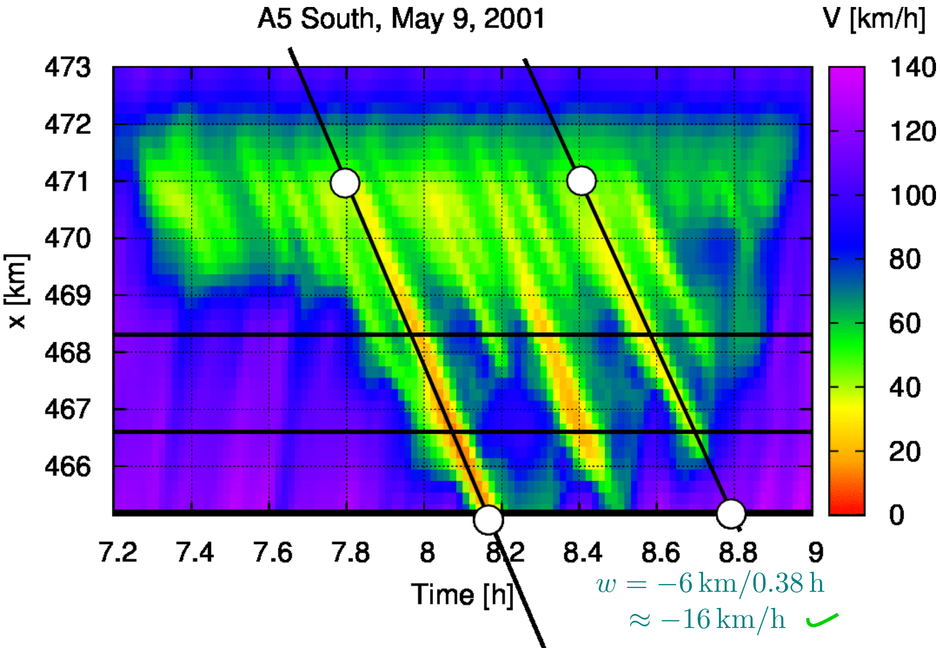


$$w = -3.3\,\mathrm{km}/0.2\,\mathrm{h}$$
$$\approx -20\,\mathrm{km/h}$$

Check results by approximate ground truth ⇒ 2.9

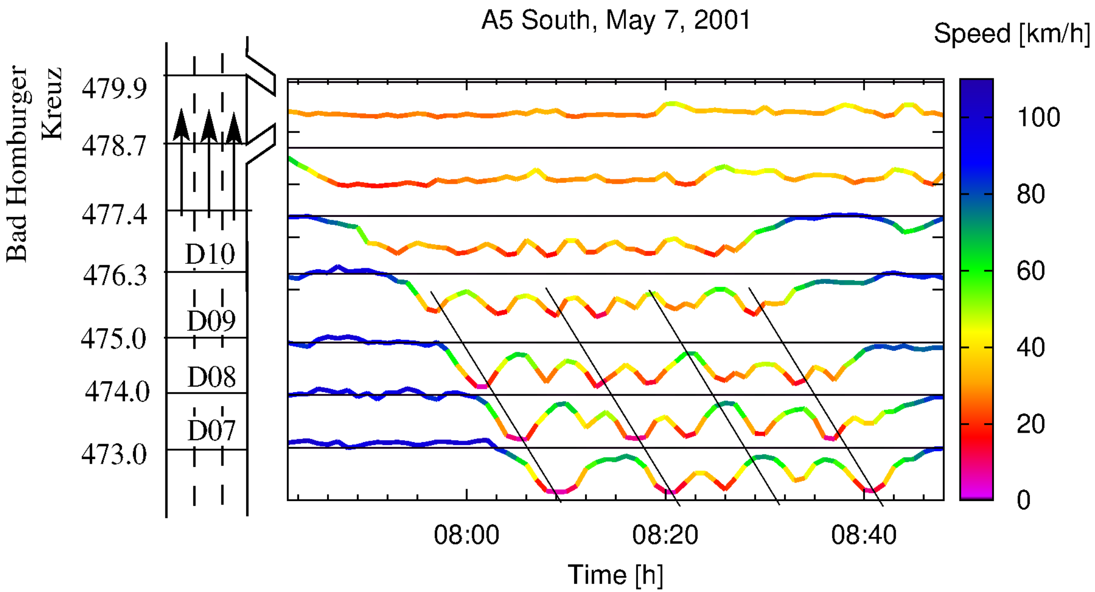


Solution

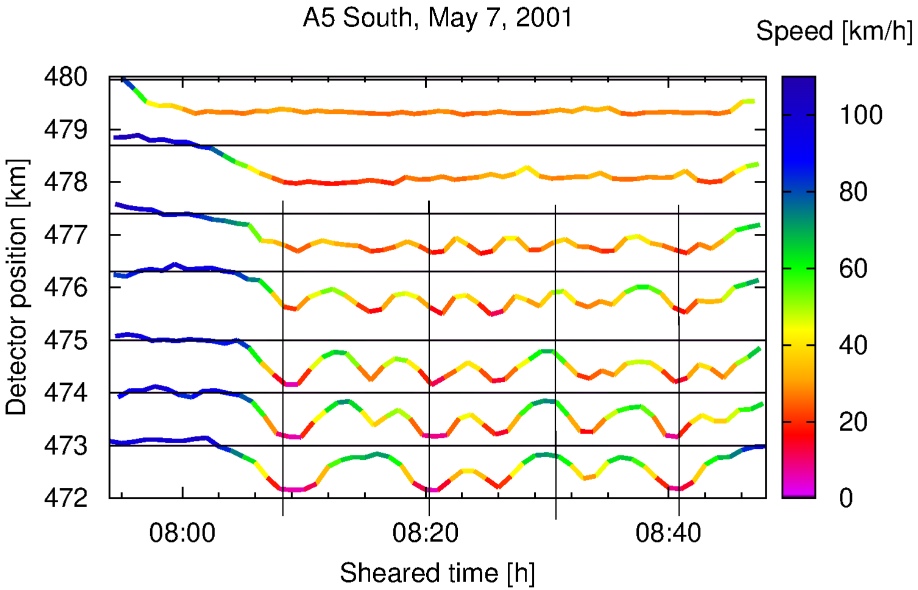


$$w = -6\,\mathrm{km}/0.38\,\mathrm{h} \approx -16\,\mathrm{km/h}$$

Determining the wave speed w statistically



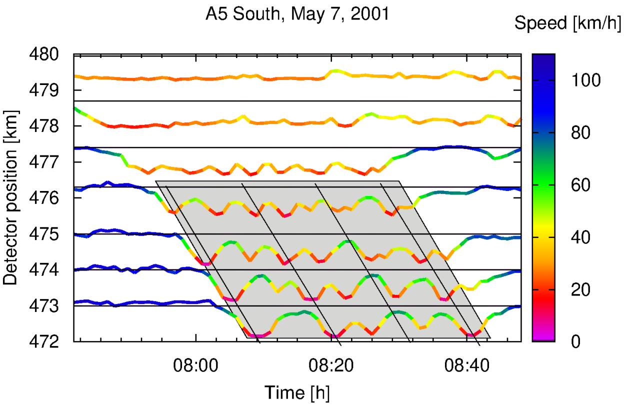
Rectification by skewed time



Statistical procedure for determining w

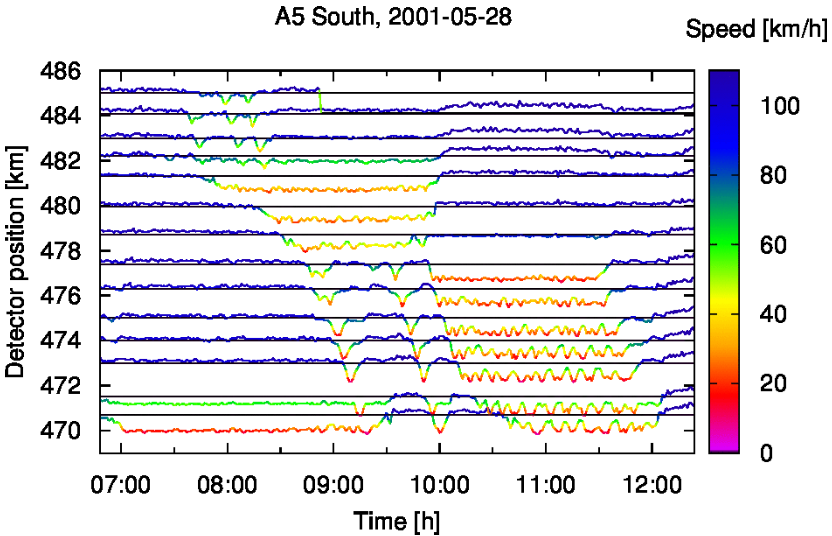
- Determine the oscillation area
- Estimate inside the oscillation area the **speed cross correlation functions (CCF)** between the detectors k and l

$$r_{kl}(\tau) = \frac{E\big((V_k(t) - E(V_k))(V_l(t+\tau) - E(V_l))\big)}{\sqrt{\text{Var}(V_k)\text{Var}(V_l)}}$$



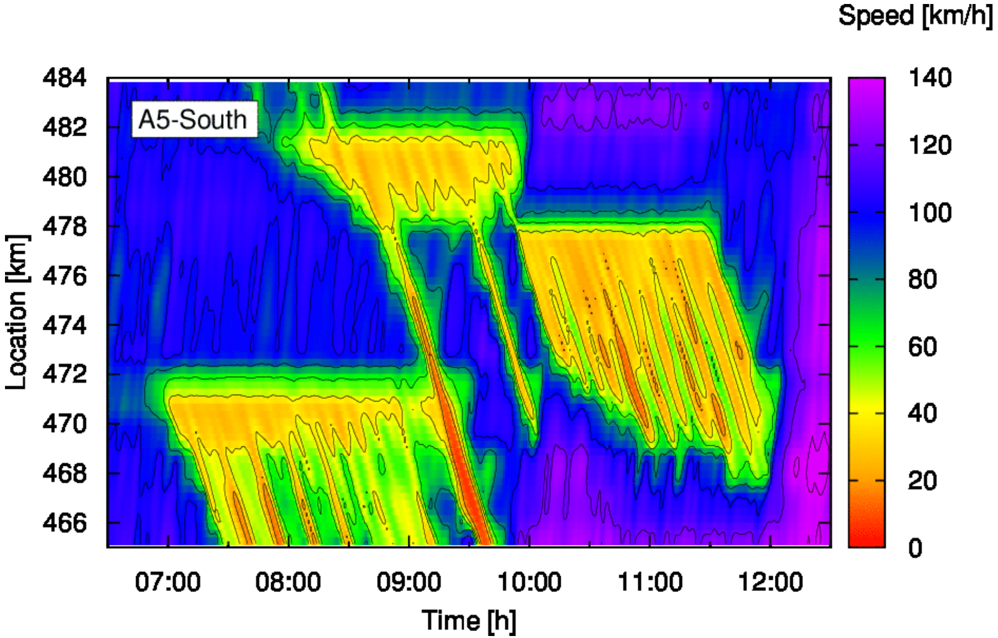
- Determine the time shifts τ_{kl} for the maxima of the CCF and the individual wave speed estimates $w_{kl} = \Delta x_{kl}/\tau_{kl}$ with $\Delta x_{kl} = x_l - x_k$ the distance between the respective detector stations
- The estimate W is the (weighted) mean of the w_{kl}

Another example

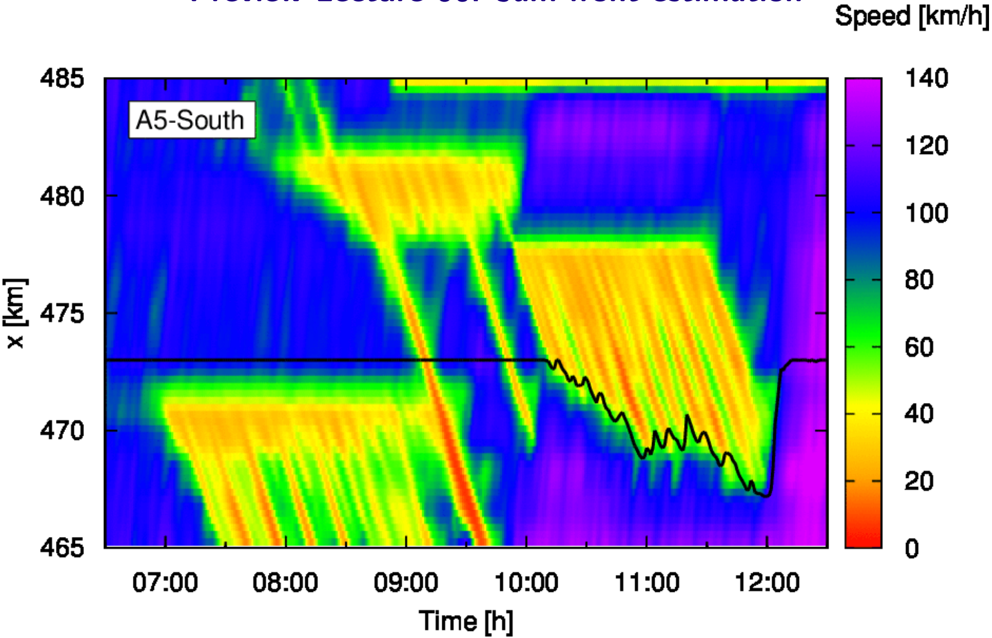


? Discuss what you see on this graphics!

Preview Lecture 04: state reconstruction



Preview Lecture 06: Jam-front estimation



You will learn how to do that three lessons ahead!