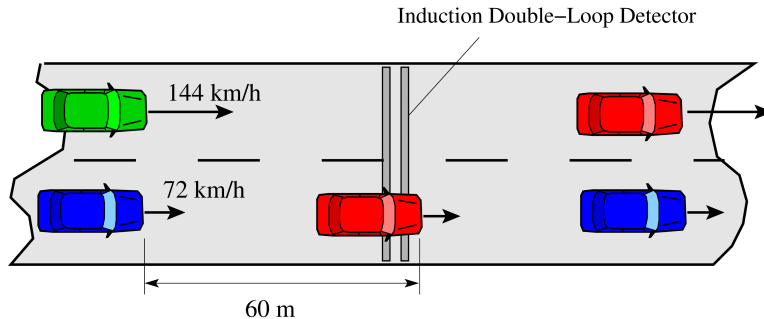


Lecture 03: Cross-Sectional Data Analysis

- ▶ 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:



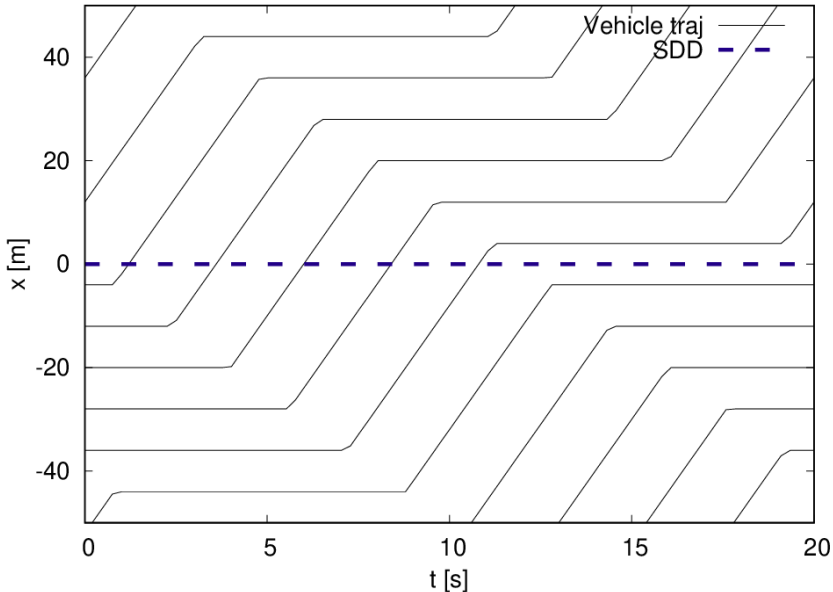
? determine Q^{tot} , V and $\rho^{\text{tot}} = Q^{\text{tot}}/V$ for the total cross-section of two lanes and compare with the “true” density and spatial mean speed

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

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

true value: $3 \text{ veh}/120 \text{ m} = 1 \text{ veh}/(40 \text{ m})$

Another example of a bias



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

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

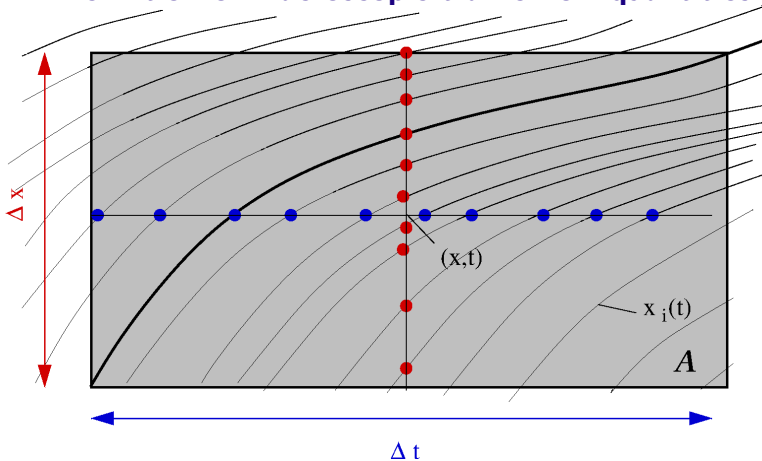
$$V = 10 \text{ m/s}$$

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



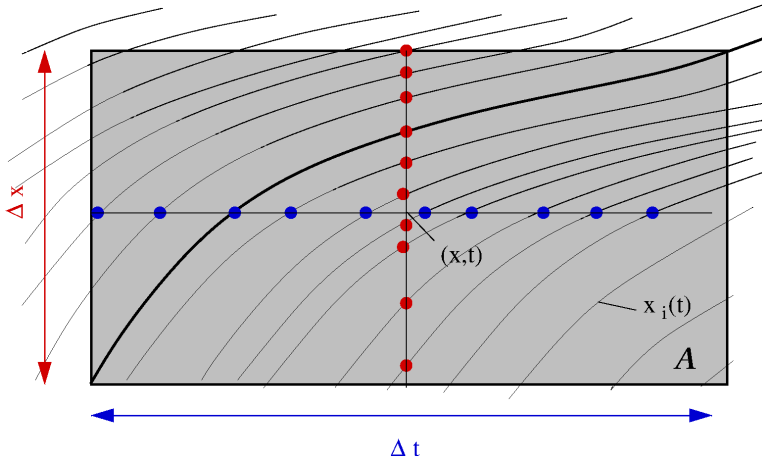
- ▶ Local average or time mean of quantity y (stationary detectors)

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

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

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

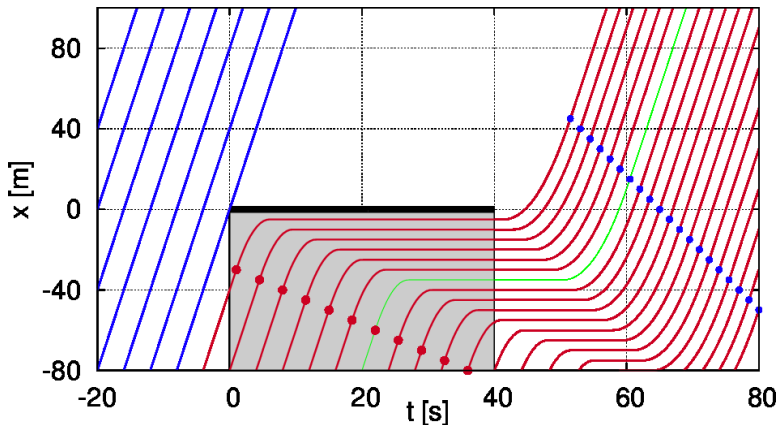
Definition of macroscopic traffic flow quantities II: Edie's definitions



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

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

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

Leutzbach relation between space and time mean speed

Assume a steady state at a spatial or **instantaneous speed density function** $f(v)$. Then,

- ▶ the *partial density* of a speed layer is given by $d\rho = \rho f(v) dv$. $f(v)$ and $w(v)$ blackboard
- ▶ Since the number of stationary detector recordings (time mean!) is proportional to the flow, the temporal or **local speed density function** $w(v)$ relevant for detector measurements is proportional to the *partial flow* $dQ = v d\rho$:

$$w(v) dv = \frac{dQ}{Q} = \frac{\rho v f(v) dv}{\int \rho v f(v) dv} = \frac{v f(v) dv}{\int v f(v) dv} = \frac{v f(v) dv}{V_s} \Rightarrow w(v) = \frac{v f(v)}{V_s}$$

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

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

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

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

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 \rightarrow 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 $w(v) dv = vf(v)/V_s dv$ as

$$E(y_i) = \int yw(v) dv = \frac{1}{V_s} \int yvf(v) dv$$

- ▶ With $y_i = 1/v_i$, we obtain

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

$$V_s = V_H$$

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.

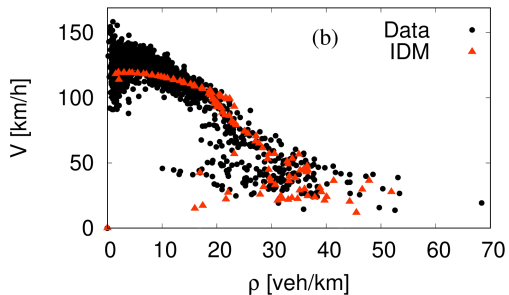
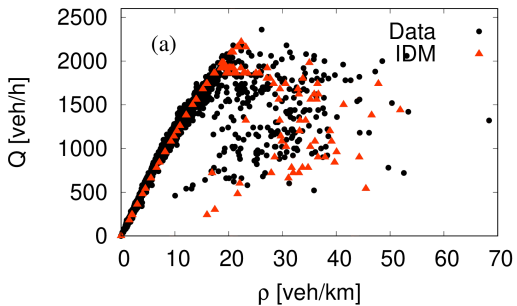
$$\begin{aligned}\frac{1}{\hat{\rho}} &= E(d_i) = E(v_{i-1}\Delta t_i) \\ &\approx E(v_i\Delta t_i) \\ &= E(v_i)E(\Delta t_i) + \text{Cov}(v_i, \Delta t_i) \\ &= \frac{V}{Q} + \text{Cov}(v_i, \Delta t_i)\end{aligned}$$

⇒ unbiased estimator $\hat{\rho}$ as a function of the “usual” estimator $\rho = Q/V$:

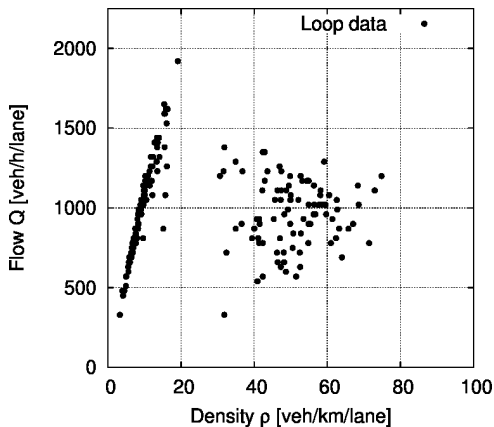
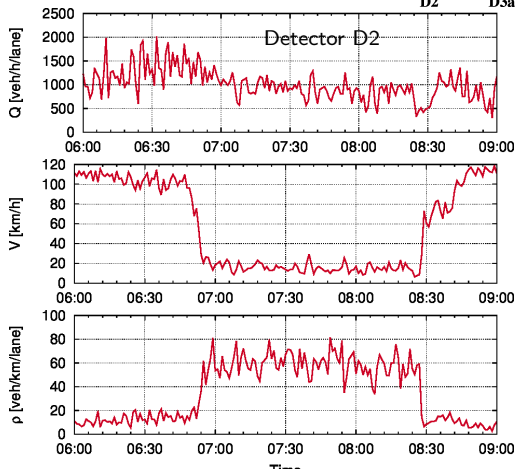
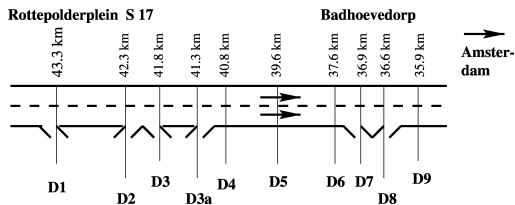
$$\hat{\rho} = \rho \left(\frac{1}{1 + \rho \text{Cov}(v_i, \Delta t_i)} \right)$$

? 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?

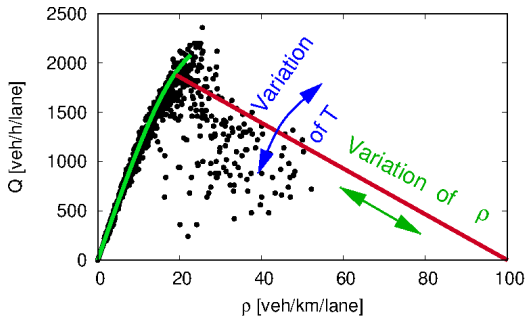
Comparison with a model reveals systematic bias



3.2. Analysis I: Local Flow Characteristics

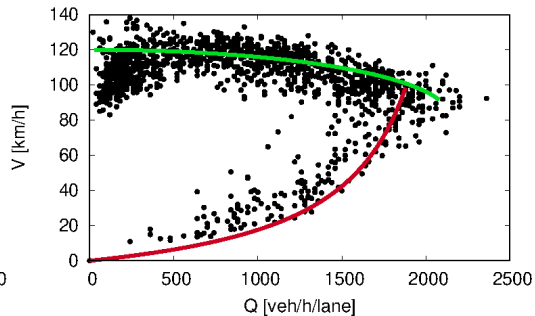
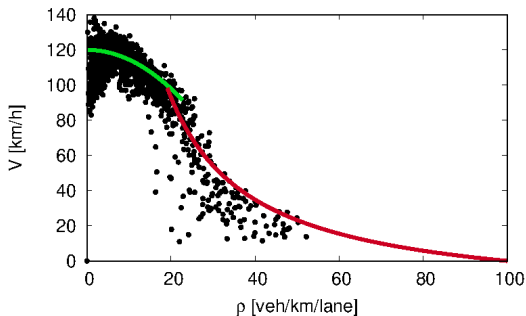


Flow-density-speed data and fundamental diagram

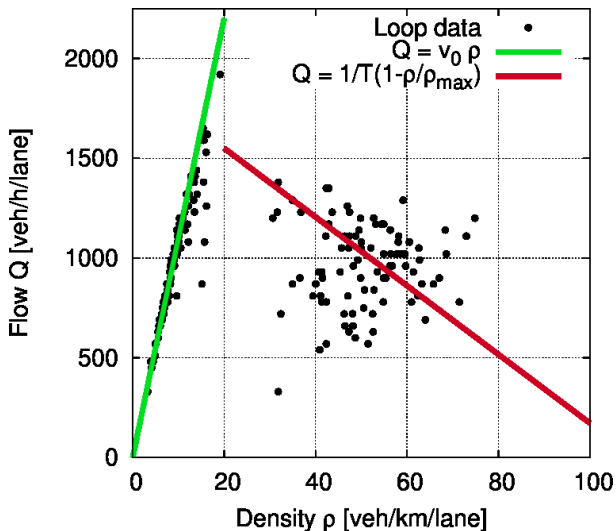


Free and congested Regimes:

Why is the red congested line of the FD not a regression line of the congested data points?



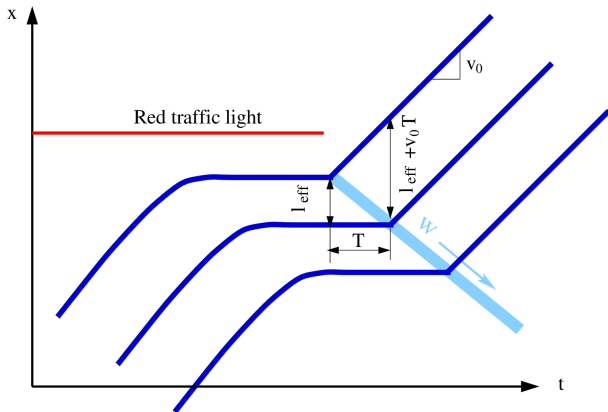
Why is the fundamental diagram so “fundamental”



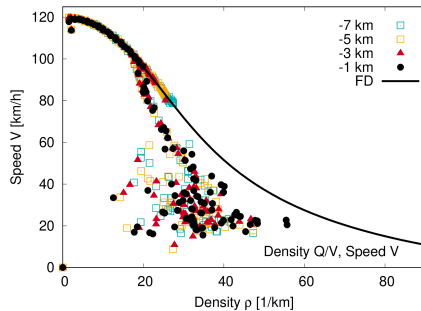
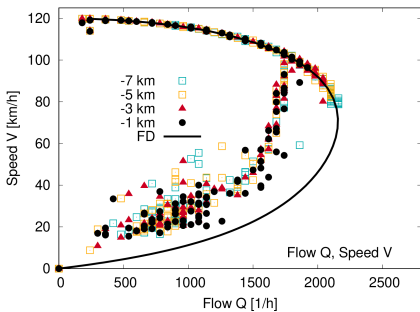
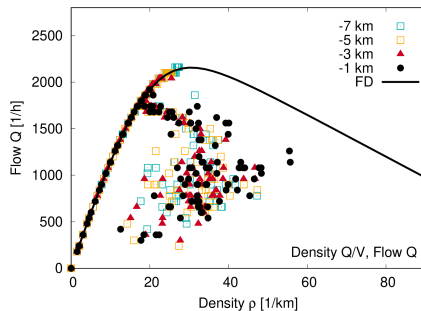
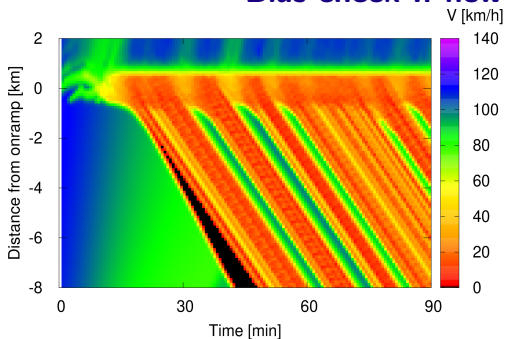
- ▶ **free traffic:**
 $Q(\rho) = V_0 \rho$
- ▶ Intersection with the abszissa:
 $\Rightarrow l_{\text{eff}} = 1/\rho_{\max}$
- ▶ Slope $w = Q'_c(\rho) = -l_{\text{eff}}/T \Rightarrow$
 wave speed w , time gap T
- ▶ Intersection with Q_f :
 \Rightarrow estimate for capacity
 $Q_{\max} = V_0/(V_0 T +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

Derivation of the wave speed

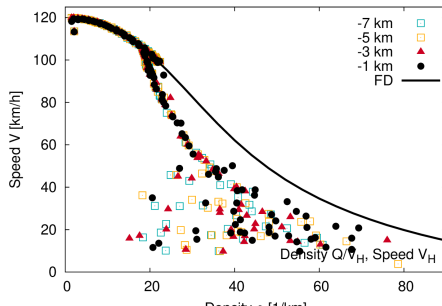
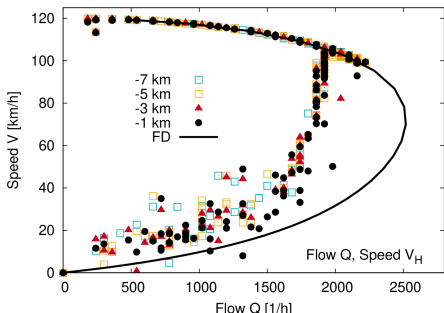
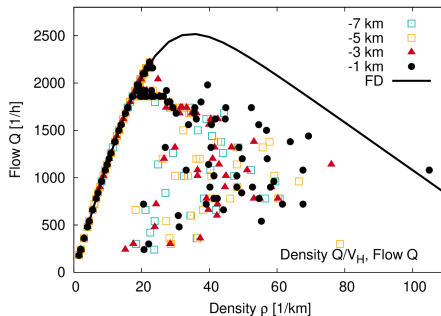
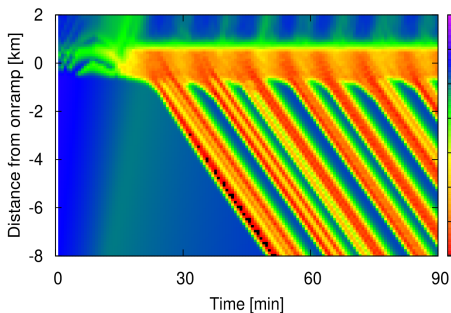
- ▶ The sequential starting of vehicles once a traffic light turns green has nothing to do with reaction time but that a moving vehicle needs more space headway ($l_{\text{eff}} + vT$) than a standing one (l_{eff})
- ▶ Extrem case: Zero reaction time, infinite acceleration to the desired speed v_0 one the space headway $\Delta x = l_{\text{eff}} + v_0 T$
- ▶ Reasoning also valid for the general congested case (\Rightarrow Newell's model)
- ▶ Wave speed equal to gradient of the congested part of the FD \Rightarrow later



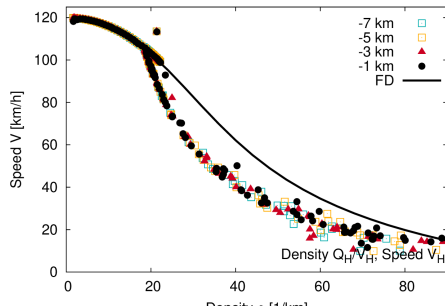
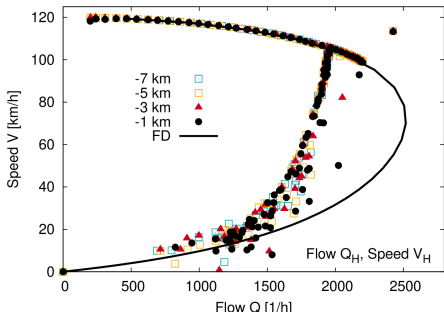
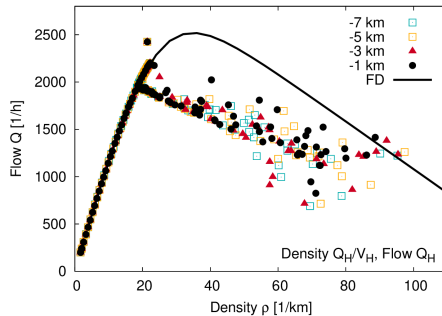
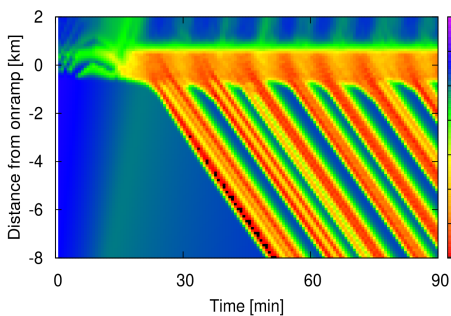
Bias check I: flow Q , speed V , density $\rho = Q/V$



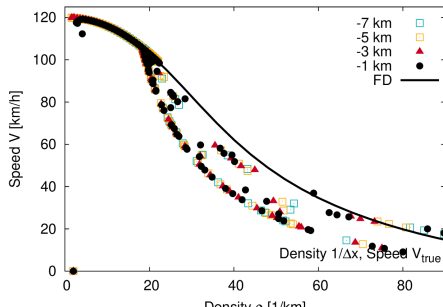
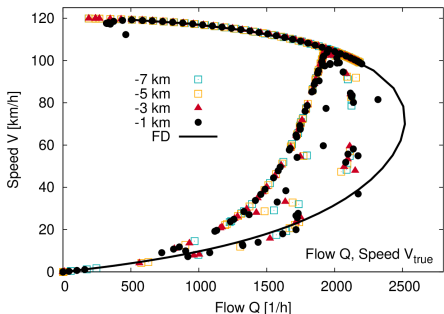
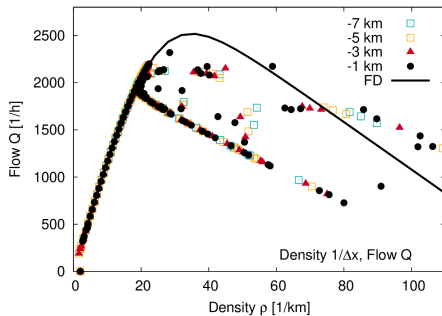
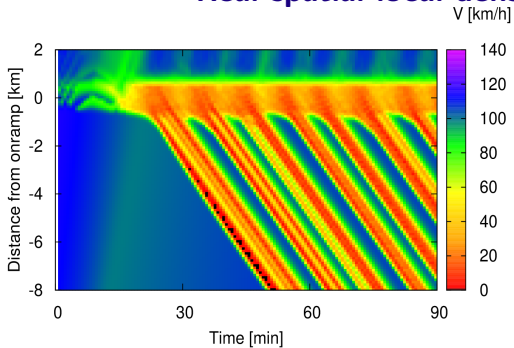
Bias II: flow Q , speed V_H , density Q/V_H



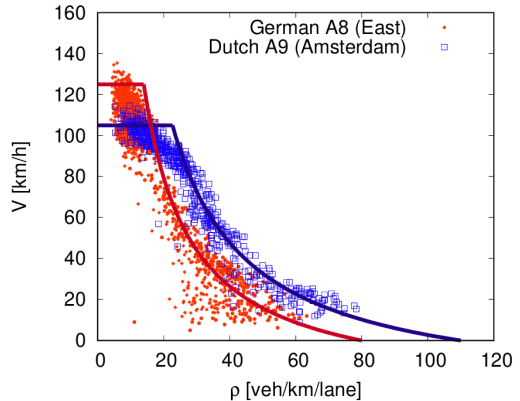
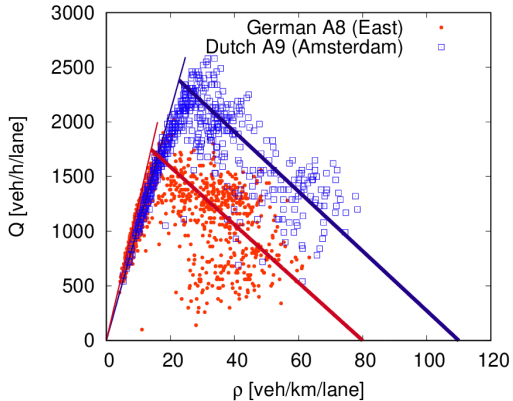
Bias III: flow Q , speed V_H , density $\rho = Q_H/V_H = E(1/\Delta t_i)/V_H$



Real spatial local density, spatio-temporal local speed

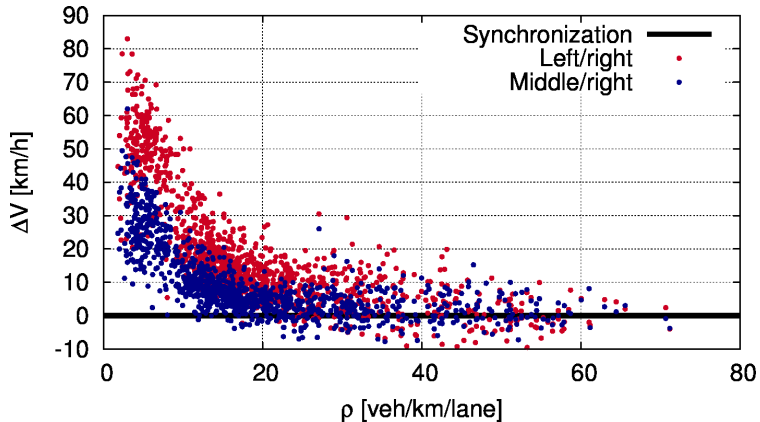


Regional and infrastructural differences



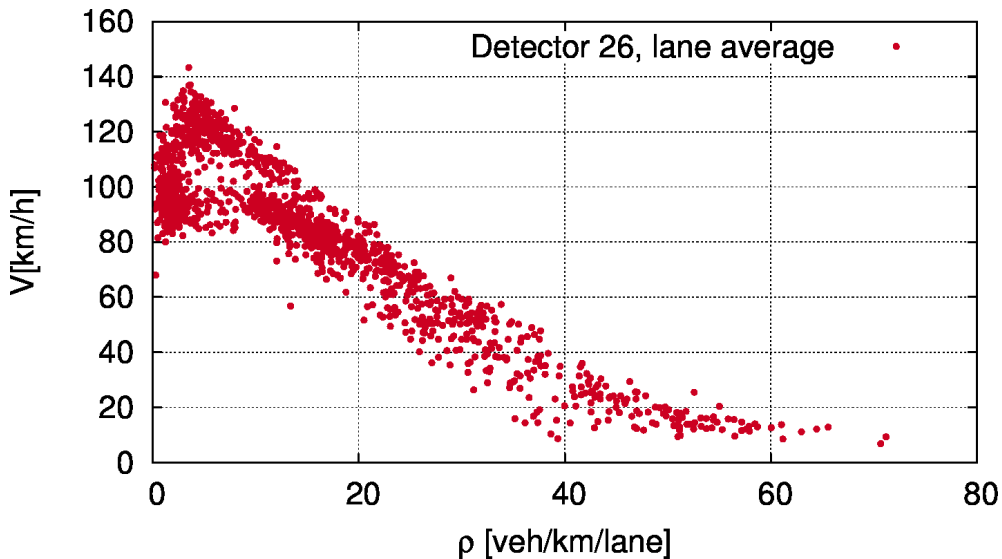
- ▶ German A8-East near Munich: Higher maximum speed and lower capacity compared to the Dutch A9 near Amsterdam

Speed synchronisation across lanes



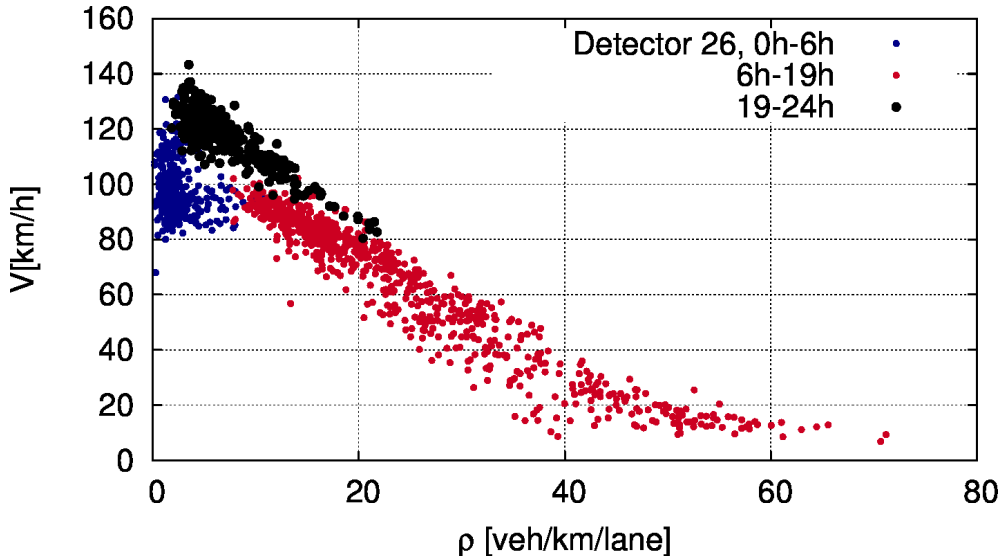
- ▶ Low densities \rightarrow little interactions \rightarrow nearly everybody can drive at his/her desired speed choosing the suitable lane ("fast", "middle", or "slow") $\rightarrow \Delta v$ large;
- ▶ densities near capacity: still no congestion but much interaction \rightarrow small Δv values;
- ▶ 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

- ? Explain Simpson's effect for exam ratings: Languages: females 80 %, avg grade 2.5, males 20 %, avg grade 2.0; STEM: females 20 %, avg grade 3.5, males 20 %, avg grade 3.0.
- ! In each department, males have a better average. Still, mixing the departments together, the women are better (average 2.7) than the men (2.8).
- ? 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: $V = 0.5(80 + 120) \text{ km/h} = 100 \text{ km/h}$
 - (ii) before the rush hour: $V = 0.1 * 80 + 0.9 * 110 \text{ km/h} = 107 \text{ km/h}$

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

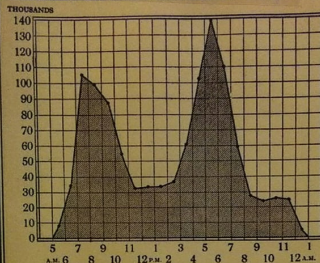
! Time series:

- ▷ 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



THE UNDERGROUND

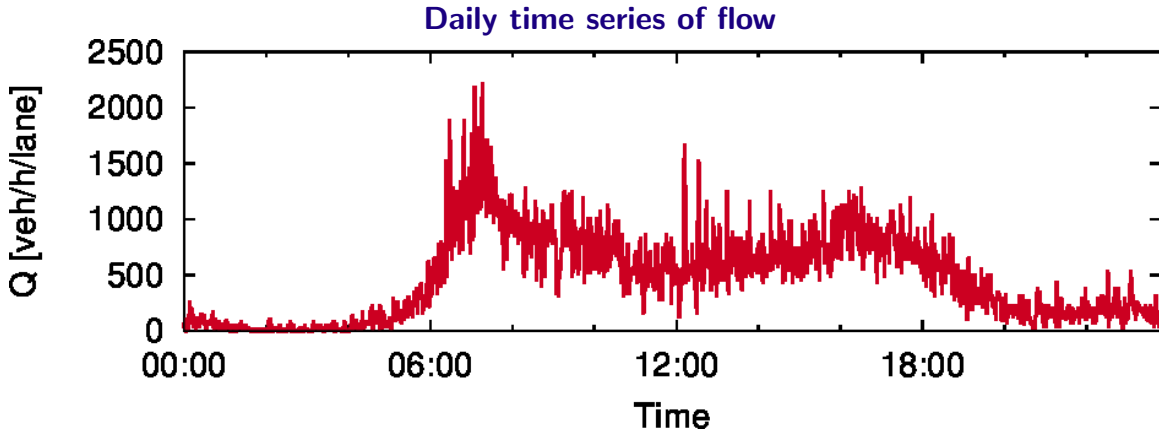
Doors open 5 a.m.

A.M.		
5-6	The Early Birds . . .	8,100
6-7	More Early Birds . . .	34,250
7-8	The Workers . . .	105,960
8-9	The Business Man . . .	99,030
9-10	More Business Men . . .	87,810
10-11	The Late Comers . . .	54,610
11-12		32,380
NOON	The Shoppers and Pleasure-	
12-1	Seekers are now abroad, and	33,480
P.M.	it's the best time too, as the	
1-2	Business Folk are at work	33,480
2-3	and there is more room in	37,220
3-4	the Trains	61,720
4-5		102,040
5-6	The Business Man and	139,620
6-7	the Workers return	110,290
7-8	Theatres, Cinemas and	
	Restaurants IN . . .	59,000
8-9	Not all Patrons are punctual	27,570
9-10	A quiet hour. London is recreating	24,430
10-11	Cinemas OUT . . .	26,800
11-12	Theatres and Restaurants OUT	25,210
12- 1	The Night Birds . . .	5,200

Total Passengers. 1,108,200

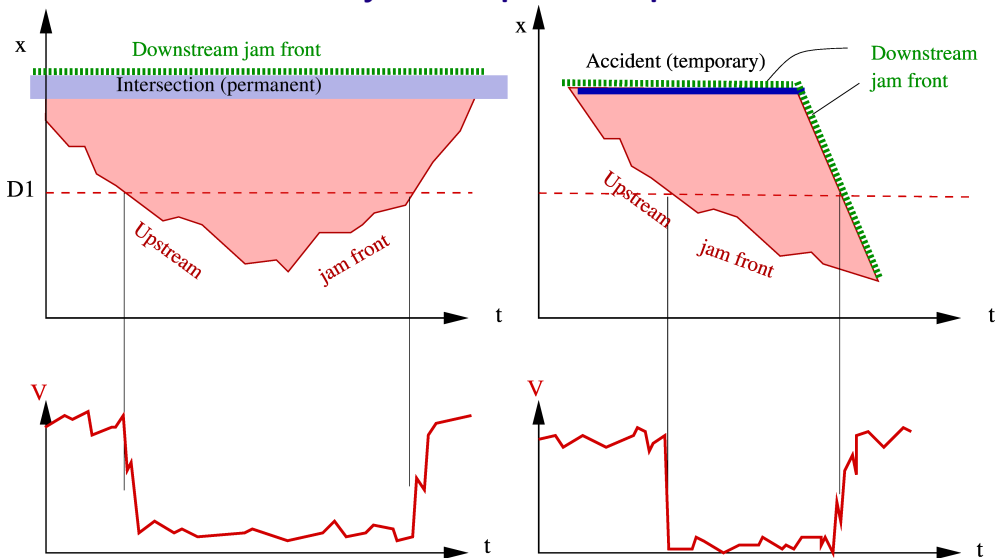
Doors close 1.30 a.m.

WALK IN AND SEE THE SHOW
NEVER A DULL MOMENT IF YOU
TRAVEL



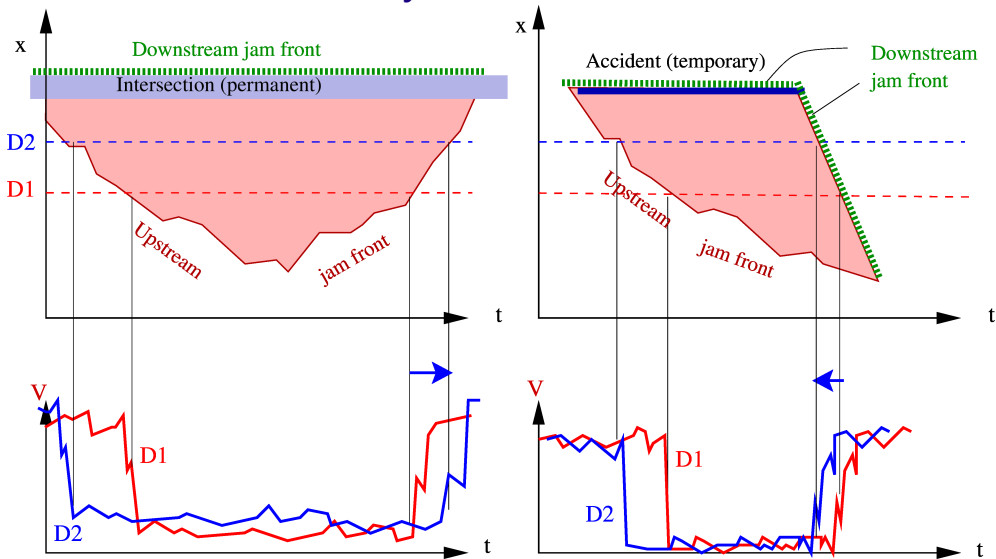
- ▶ Only single-loop detectors needed
- ▶ Unless there is congestion, this data reflects the traffic demand (why this restriction?)
- ▶ Application mainly in transportation/traffic planning and traffic politics \Rightarrow DTV
- ▶ Traffic flow application: **historic data base** to improve traffic state 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

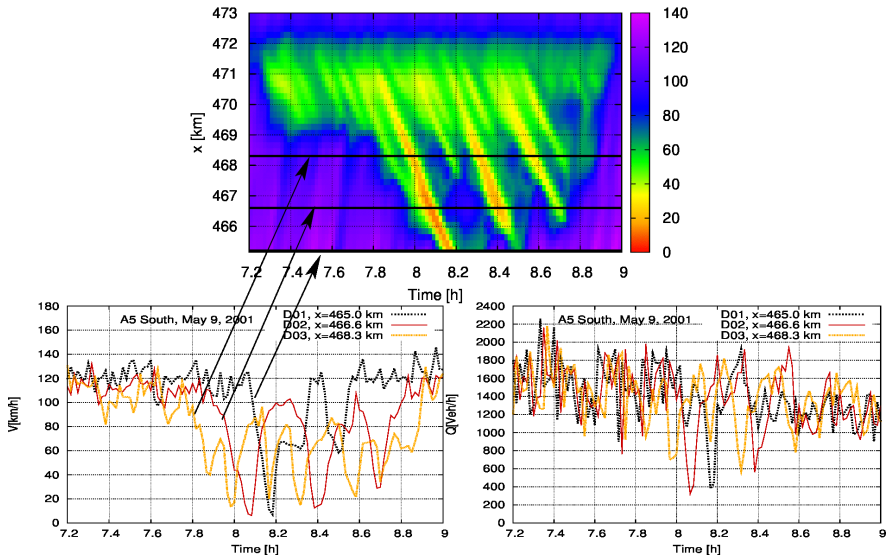
Resolution by two or more cross-sections



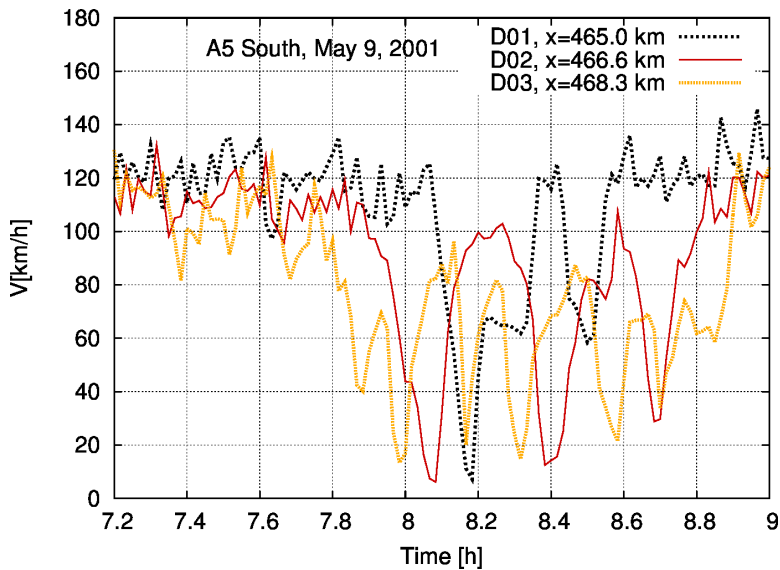
Ambiguity resolved!

Determining the wave velocity

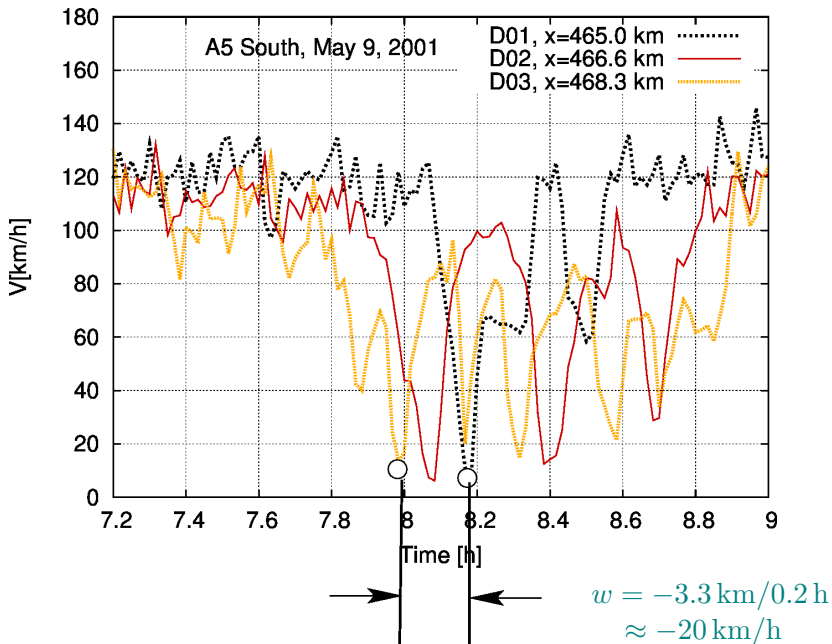
A5 South, May 9, 2001

 V [km/h]


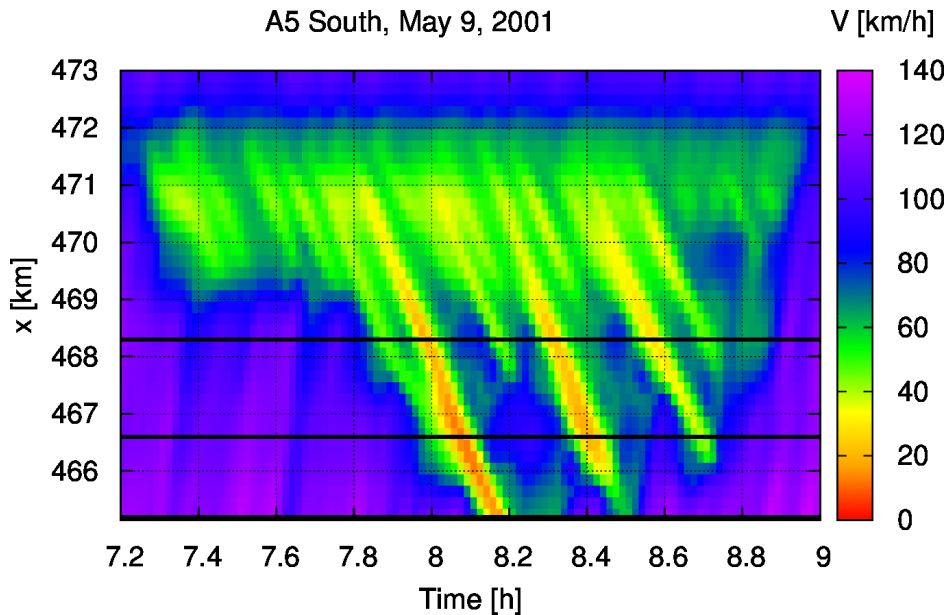
Problem: Determine the wave velocity by speed time series



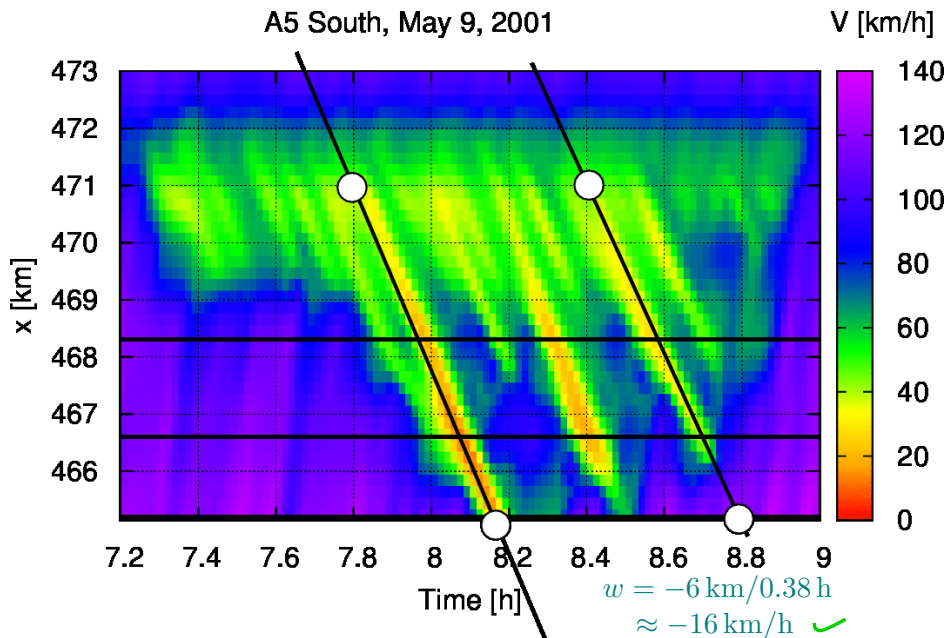
Solution



Check results by approximate ground truth \Rightarrow 2.9

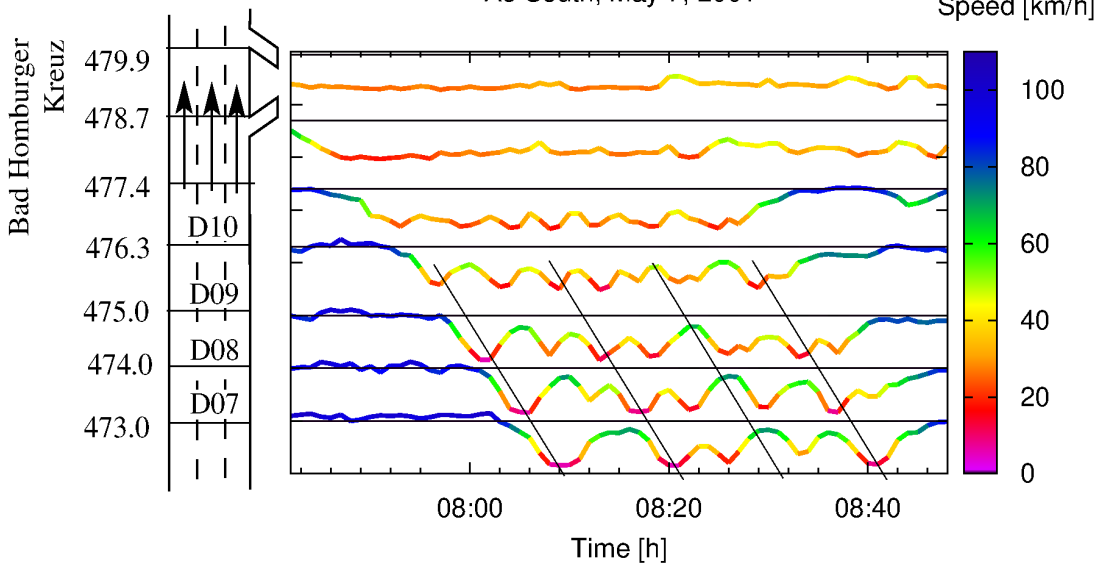


Solution



Determining the wave speed w statistically

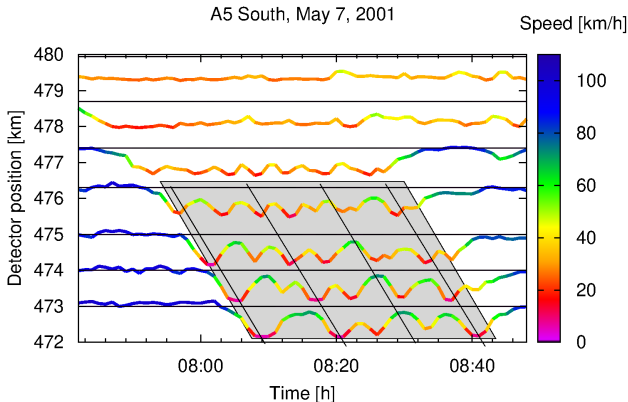
A5 South, May 7, 2001



Statistical procedure for determining w

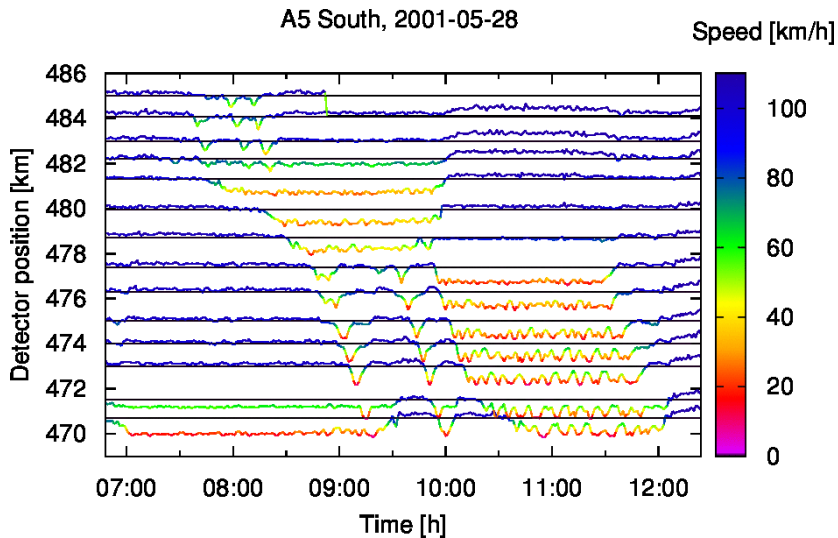
- ▶ Determine the oscillation area
- ▶ Estimate inside the oscillation area the **speed cross correlation functions (CCF)** between the detectors k and l

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



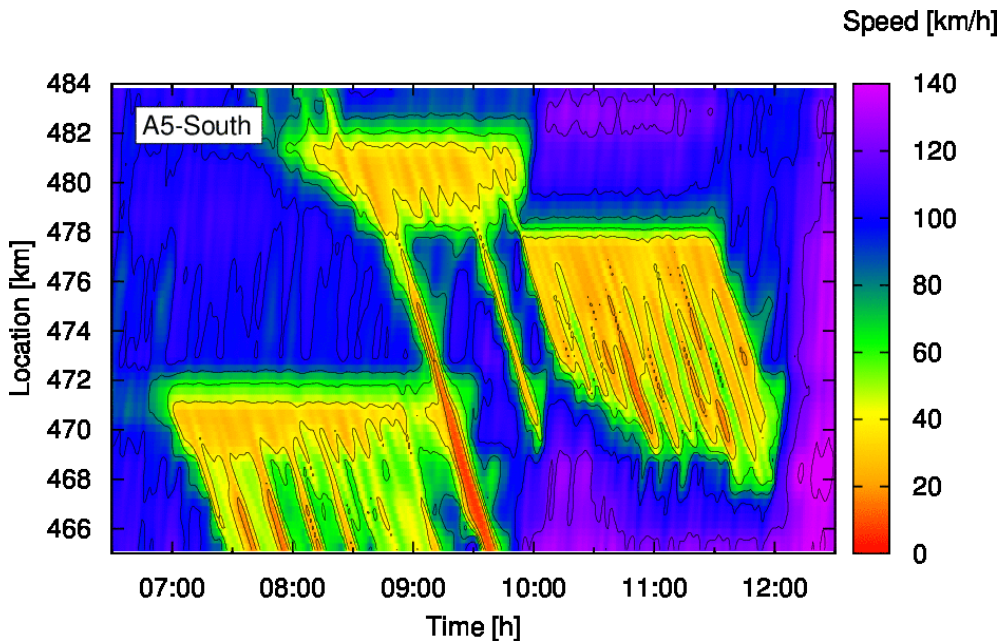
- ▶ Determine the time shifts τ_{kl} for the maxima of the CCF and the individual wave speed estimates $w_{kl} = \Delta x_{kl} / \tau_{kl}$ with $\Delta x_{kl} = x_l - x_k$ the distance between the respective detector stations
- ▶ The estimate W is the (weighted) mean of the w_{kl}

Another example

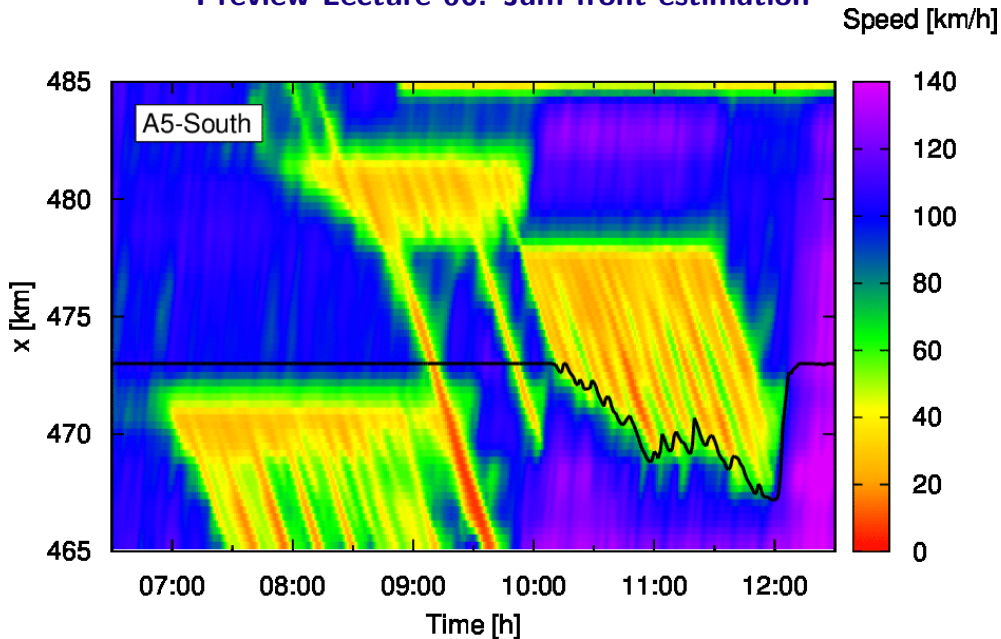


? Discuss what you see on this graphics!

Preview Lecture 04: state reconstruction



Preview Lecture 06: Jam-front estimation



You will learn how to do that three lessons ahead!