

# Statistical modeling for bus network performance assessment

Sylvain Lassarre

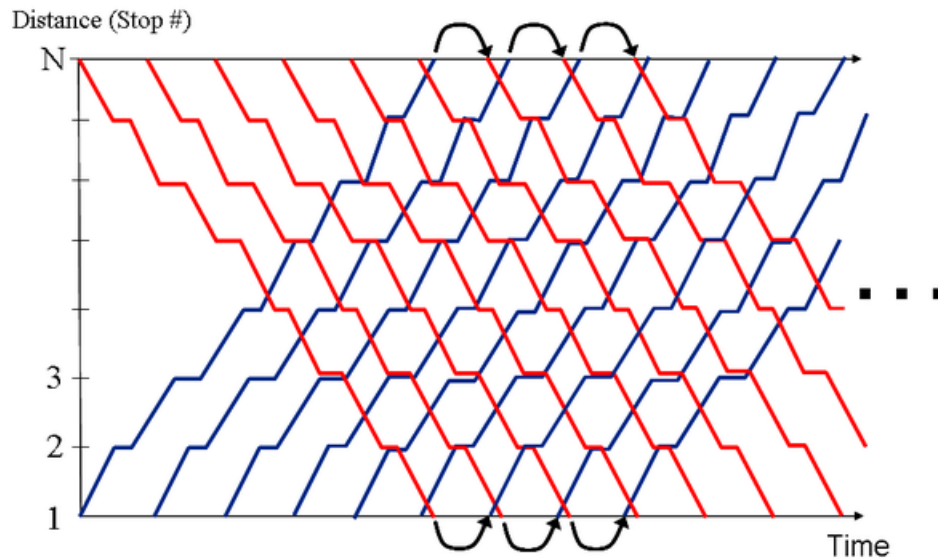
GRETTIA/COSYS/IFSTTAR

- Spatio-temporal processes for buses
  - Deterministic
  - Stochastic
    - Travel time and dwell time
    - Daganzo's model
- Trajectories and missing data
- Statistical models for processes
- Spatio-temporal processes for passengers
- Conclusion

- Most of the reliability indicators are derived from statistics on bus movements data
- For exemple
  - Coefficient of variation
  - Percentile
- without knowing the nature of the distribution of the stochastic variables such as travel time, headway, speed, ... and the stochastic process which governs the dynamics

# Spatio-temporal processes for buses

- The travelling of the bus along the route from origin to destination (#start and end terminal) as a moving particle  $(x,t)$  which makes a pause at bus stops along the route
- Deterministic process = scheduled travelling = time-table



$$t_{n,s+1} = t_{0,0} + nH + \sum_0^s p_i$$

- $t_{n,s}$  time arrivals of bus n at bus stops s (or control points)
- H service headway
- $t_{0,0}$  departure at origin
- $p_i$  target travel time between i and i+1
- Deterministic and stationary process

$$t_{n,s+1} = t_{n,s} + c_s$$

- $p_s = c_s$  average (global) travel time from s to s+1,  
sum of
  - $\pi_s$  Inter-stop travel time between bus stops s and s+1
  - $d_s$  Dwell time or delay due to bus stop s

# Stochastic processes and models (1)

- Of the arrival times  $a_{n,s}$  because of random disturbances (traffic, passenger demand, driver behavior, ...) affecting  $t_{n,s}$
- By means of (Hans et al., 2014)
  - $u_{n,s} = a_{n,s+1} - a_{n,s}$  global travel time for segment  $s, s+1$ 
    - dwell time models (boarding and (alighting) passengers)
    - travel time models

$$u_{n,s+1} = a_{n,s+1} - a_{n,s} = d_{n,s} + \pi_{n,s}$$

$$a_{n,s+1} = a_{n,s} + d_{n,s} + \pi_{n,s}$$

# Dwell time models

- $L_{n,s}$  load of the bus
- $A_{n,s}$  number of alighting passengers off the bus
- $B_{n,s}$  number of boarding passengers in the bus

$$L_{n,s+1} = L_{n,s} - A_{n,s} + B_{n,s} \quad \text{Infinite capacity}$$

$$d_{n,s} = \max(aA_{n,s}, bB_{n,s}) + c$$

$$d_{n,s} = aA_{n,s} + bB_{n,s} \quad \text{Times to alight /board/ opening door}$$

$$A_{n,s} = d_s L_{n,s-1} \quad \text{Alighting proportion}$$

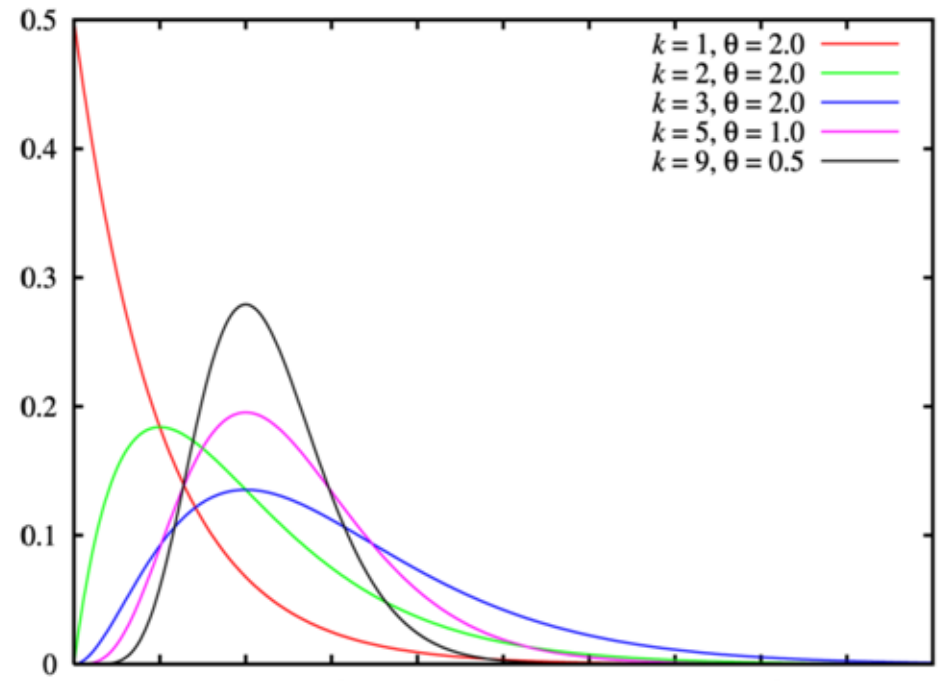
$$B_{n,s} = \left( \lambda_s - \frac{1}{\tau_s} \right) h_{n,s} \quad \text{Arrival rate of passengers}$$

# Travel time and speed models

- Gamma distribution
- skewed to the right, is equal to the exponential distribution when the scale is equal to 1 and tends towards the normal distribution when the scale gets infinite. The variance is proportional to the square of the mean  $E(Y)=v\lambda$ .

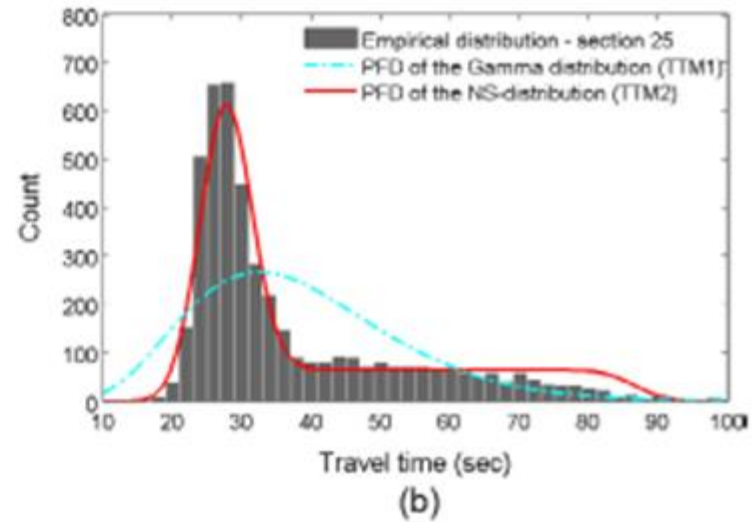
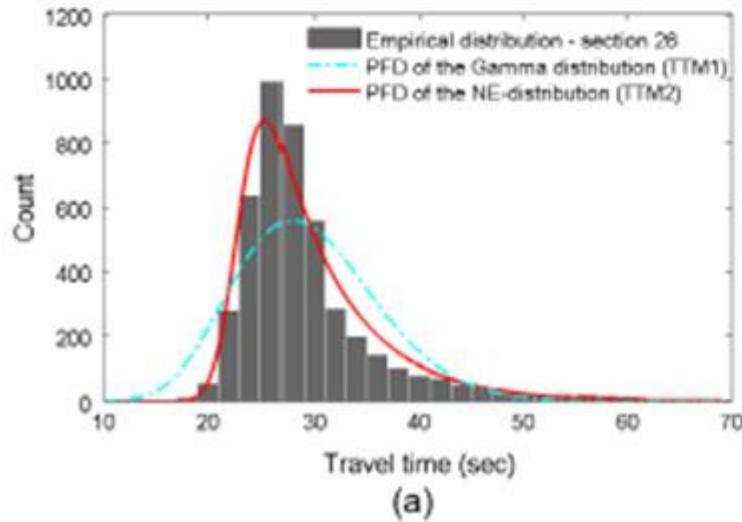
- an inverse-gamma distribution

$$\Gamma(v_{s+1}, 1/L_{s+1}\lambda_{s+1})$$





- Normal-exponential distribution
  - Compound of a normal distribution by an exponential distribution



- Generalized extreme distribution

Portland data At each stop door opening and closing times are recorded, and the APC records the number of boarding and alighting passengers

**Table 1**

Model components with their characteristics.

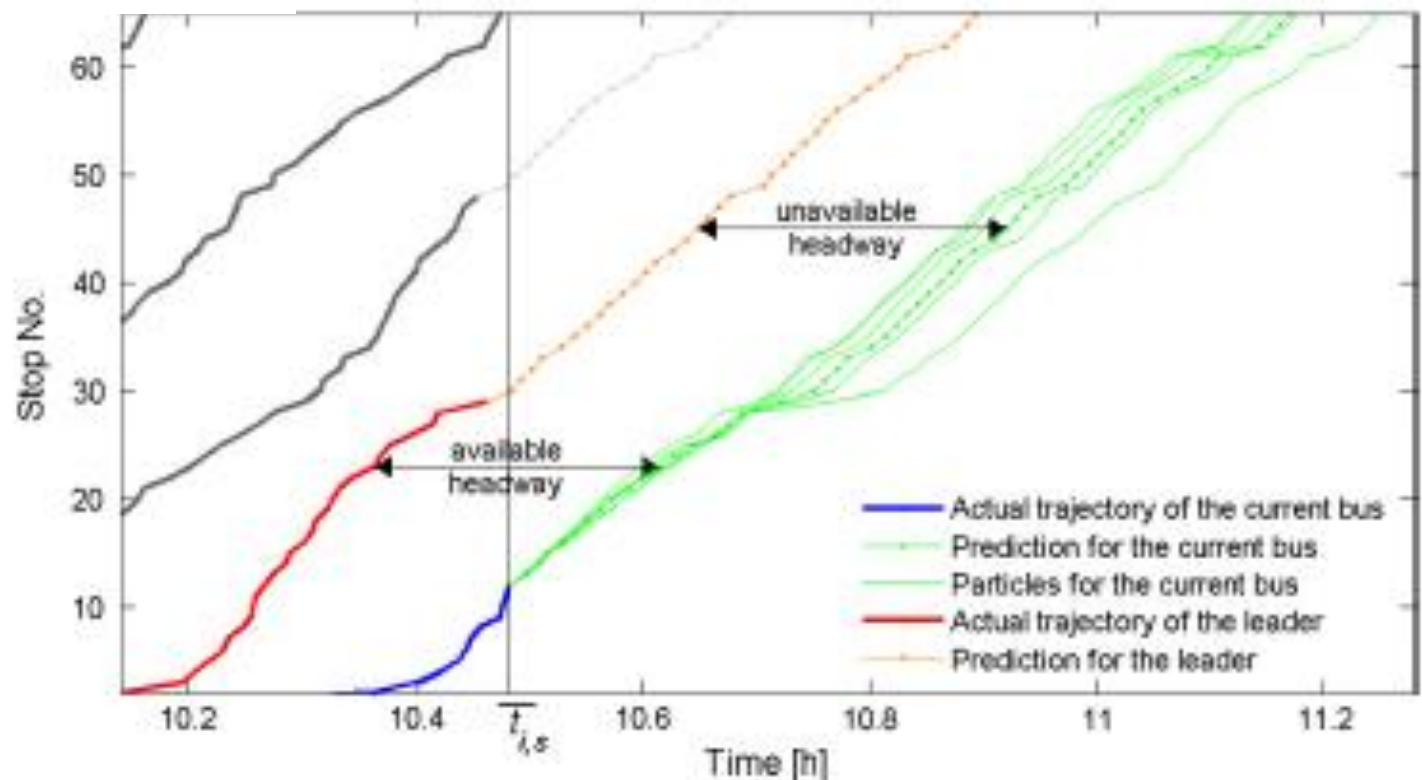
Model	Description	Data requirement	Stochastic form	Deterministic form	
Dwell time	1 Use distributions for dwell times	Archived bus dwell times from AVL	$d_{i,s} \rightarrow \mathcal{E}(c + \beta_s h_{i,s})$	$d_{i,s} = c + \beta_s h_{i,s}$	
	2 Use distributions for numbers of passengers	Door opening detection Archived bus dwell times from AVL Automatic passenger counters (APC)	$B_{i,s} \rightarrow P(\lambda_s h_{i,s})$ $A_{i,s} \rightarrow \text{Bi}(L_{i,s}, \eta_s)$	$B_{i,s} = \lambda_s h_{i,s}$ $A_{i,s} = L_{i,s} \eta_s$	
Travel time	1 Use travel time distributions	Archived bus travel times from AVL	(a) $\pi_{i,s} \rightarrow \mathcal{N}(\mu_s, \sigma_s)$ (b) $\pi_{i,s} \rightarrow \mathcal{L}(\mu_s, \sigma_s)$ (c) $\pi_{i,s} \rightarrow \mathcal{G}(\mu_s, \sigma_s)$		
	2	Archived bus travel times from AVL Signal locations	(a) Without a signal: $\pi_{i,s} \rightarrow \text{NE}(\mu_s, \sigma_s, \lambda_s)$ (b) Otherwise: $\pi_{i,s} \rightarrow \text{NS}(\mu_s, \sigma_s, r_s, g)$	(a) $\pi_{i,s} = \mu_s + \lambda_s$ (b) $\pi_{i,s} = \mu_s + \frac{d_s^2}{2(r_s + c_s)}$	
	3	Travel time based on synthetic and realistic trajectories	Signal location and settings	$\pi_{i,s,n} \rightarrow \mathcal{N}\left(\frac{\Delta x_{i,n}}{v}, \sigma_s\right)$	$\pi_{i,s,n} = \frac{\Delta x_{i,n}}{v}$
	4		Velocity characteristic of buses Signal location and settings Flow from loop detectors Velocity characteristic of buses	$\pi_{i,s,n} \rightarrow \mathcal{N}\left(\frac{\Delta x_{i,n}}{v}, \sigma_s\right)$ $N_{i,s,n} \rightarrow P(q_s \Delta t)$	$\pi_{i,s,n} = \frac{\Delta x_{i,n}}{v}$ $N_{i,s,n} = q_s \Delta t$

# Particle filter as prediction method

- Generate K trajectories (particles) based on models

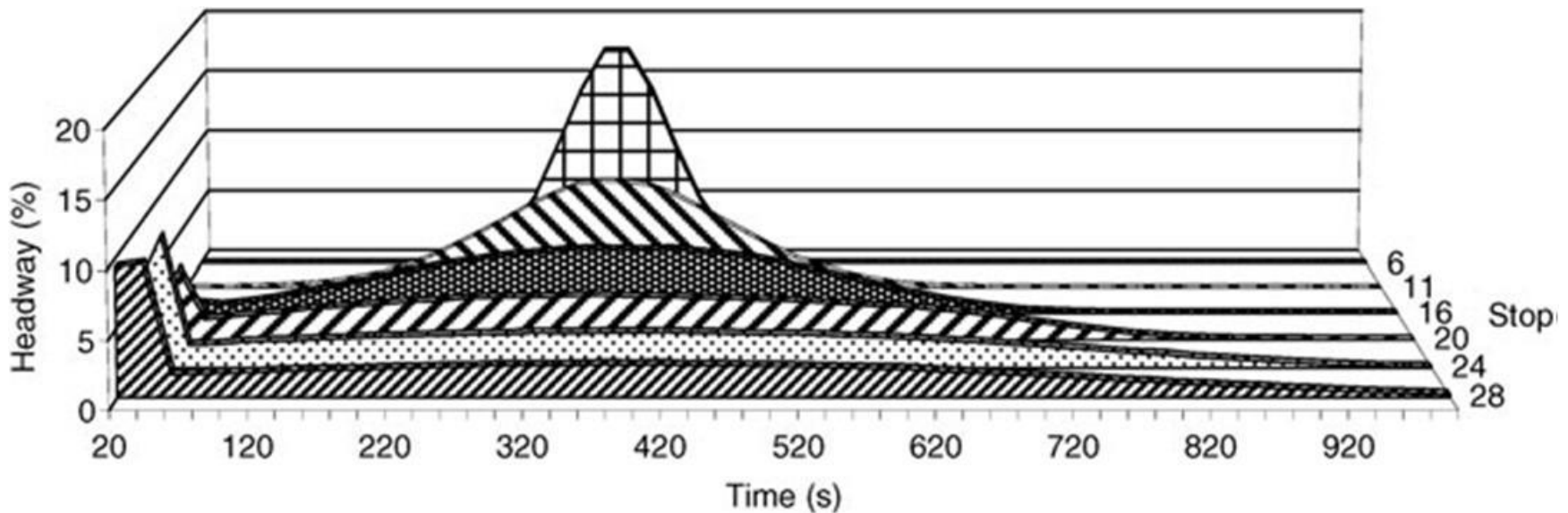
$$t_{i,s_2|s_1}^{(k)} = \bar{t}_{i,s_1} + \sum_{s=s_1}^{s_2} (d_{i,s}^{(k)} + \pi_{i,s}^{(k)})$$

$$t_{i,s_2|s_1} = \text{median}_{k \in \{1, \dots, K\}} (t_{i,s_2|s_1}^{(k)})$$



# Headways and dwell times

- Headway distribution could be bimodal



- Typical dwell time distributions : Burr, log-logistic, log-normal

## Stochastic processes and models (2)

- By means of (Daganzo, 2009)
  - $u_{n,s} = a_{n,s+1} - a_{n,s}$  global travel time for segment  $s, s+1$
  - $h_{n,s} = a_{n,s} - a_{n-1,s}$  inter-arrival headway between buses  $n-1$  and  $n$  at stop  $s$

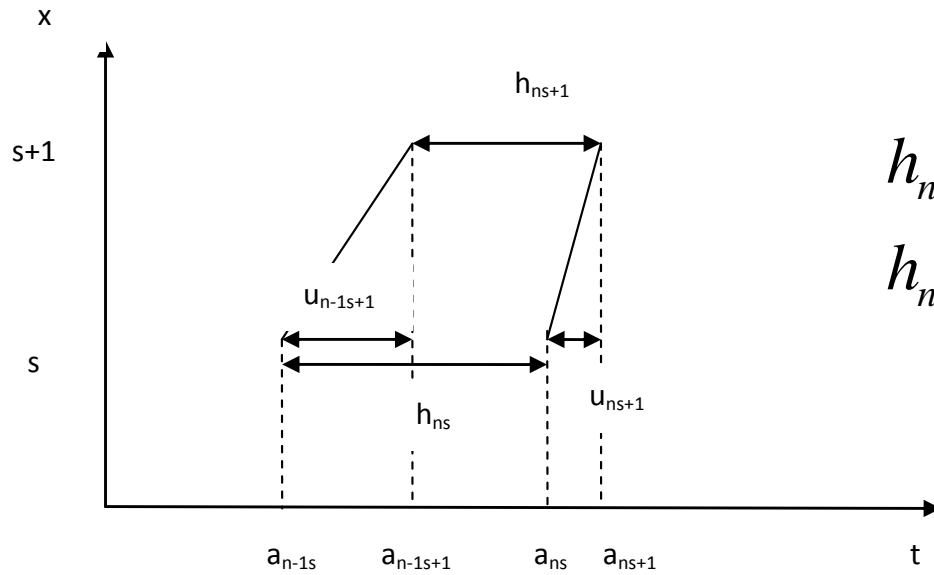
$$u_{n,s} = c_s + \beta_s (h_{n,s} - H) + v_{n,s+1}$$

$$a_{n,s+1} = a_{n,s} + c_s = a_{n,s} + \beta_s (a_{n,s} - a_{n-1,s} - H) + v_{n,s+1}$$

- With  $\beta_s$  marginal increase in expected bus delay from a unit increase in headway
- $v_{n,s+1}$  Random noise detected at  $s+1$
- Origin of bus bunching (instability of the dynamical system)

$$u_{n,s} = c_s + \beta_s (h_{n,s} - H) + v_{n,s+1}$$

$$a_{n,s+1} = a_{n,s} + c_s = a_{n,s} + \beta_s (a_{n,s} - a_{n-1,s} - H) + v_{n,s+1}$$



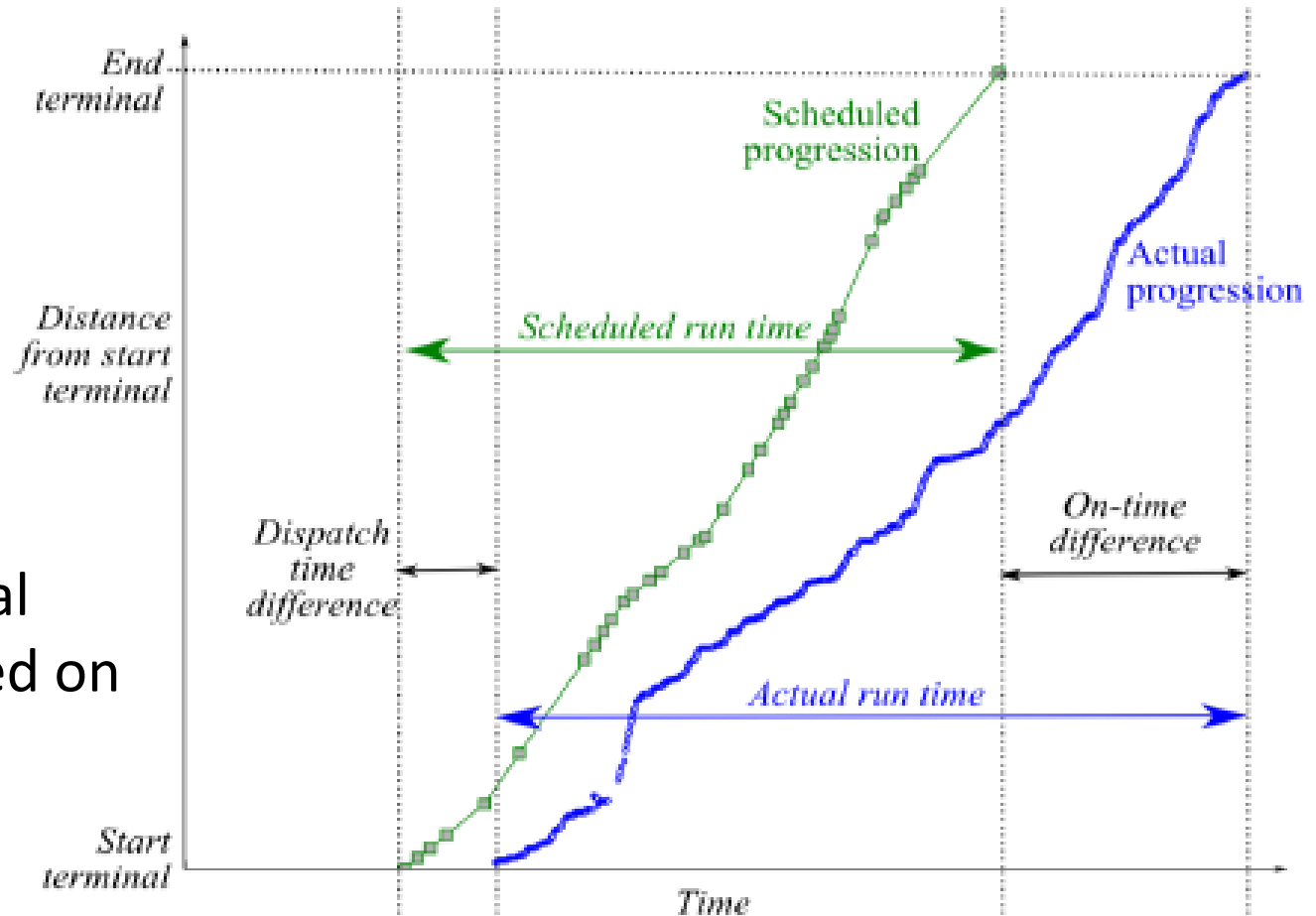
$$h_{n,s+1} = h_{n,s} - u_{n-1,s+1} + u_{n,s+1}$$

$$h_{n,s+1} - h_{n,s} = u_{n,s+1} - u_{n-1,s+1}$$

# Observed trajectories

Time sampling  
Frequency 10s.

Departure and arrival  
times to be estimated on  
observed data



Dwell times missing on scheduled trajectory

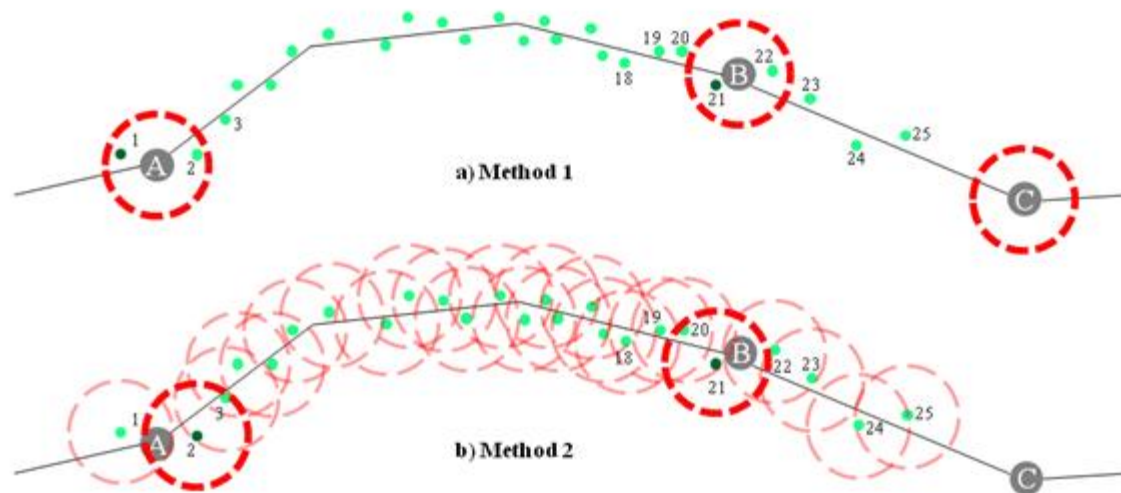
# Missing data

- Checking data to face the following anomalies
  - **Bus overtaking (BO)**, succeeding bus passes the predecessor
  - **Technical failures (TF)** in the recording of GPS data. Results in missing data and temporal technical gaps
  - **Incorrect Operations in the Service (IOS)** consequences of unpredicted missed trips and unexpected breakdowns. Results in missing data and temporal technical gaps



# Estimation of arrival times from the bus trajectories (Lakhotia et al., 2017)

- Bus stop as the unit of reference. The bus is defined to be at a bus stop if the distance between the stop and the bus is less than a certain value, which is called buffer radius (30m. radius).
- Bus as the unit of reference with route matching



*Note:*  
A, B, C represent 3 bus stops  
Green dots represent GPS points of a bus moving in the direction A to C  
Red dashed circles represent buffers

- Select a timestamp only if the distance to nearest station is less than the buffer radius.
  - Since the GPS data was at a frequency of 10 seconds, it is possible that we have multiple timestamps for a particular stop. In this case, the first time stamp is designated as arrival time and the last is departure time. It is also to be noted that it is not uncommon for drivers to stop at bus stops for less than 10 seconds in case there is no demand, in which case we have no timestamp for arrival (Missing values).
- Order of arrival of buses at intermediate stations may not be the same as overtaking is possible
  - Complicate the calculation of headways

# Estimation of statistical models

- The model is a system of equations ( $n=1$  to  $N$  bus) on the first differences between the arrival times on a sample of days ( $j = 1, l$ )

$$\Delta^s Y_{nsj} = c_s + \beta_s (Y_{ns-1j} - Y_{n-1s-1j} - H) + v_{nsj}$$

$$\Delta^s Y_{nsj} = c_s + \beta_s (\Delta^n Y_{ns-1j} - H) + v_{nsj}$$

- which can be written

$$U_{nsj} = c_s + \beta_s (h_{ns-1j} - H) + v_{nsj}$$

- The first statistical approach is to use a gamma distribution for the global travel time with two parameters shape and scale  $G(v_s, \lambda_s)$  and a gamma regression knowing the departure times of the previous bus.

- Then we have to change the formulation of the model to adapt to the form of a generalized mixed linear model with a link function which could be a logarithm (or the inverse)

$$\text{Log}E(U_{nsj}) = c_s + \beta_s (h_{ns-1j} - H)$$

$$E(U_{nsj}) = e^{c_s} e^{\beta_s (h_{ns-1j} - H)}$$

- It is an accelerated failure model. Or

$$\text{Log}E(U_{nsj}) = c_s + \beta_s \log(h_{ns-1j} - H)$$

$$E(U_{nsj}) = e^{c_s} (h_{ns-1j} - H)^{\beta_s}$$

for a multiplicative model

- We can also use the canonical link which is the inverse function (so the travel speed)

$$1/E(U_{nsj}) = c_s + \beta_s (h_{ns-1j} - H)$$

$$E(U_{nsj}) = 1/(c_s + \beta_s (h_{ns-1j} - H))$$

# Spatio-temporal processes for passengers

- Bulk queue models (serving more than one customer at a time) and Markov chain

- Stock waiting at stop  $s$  at run  $n$

$$S_s^n = S_s^{n-1} - B_s^{n-1} + \beta_s (a_s^n - a_s^{n-1})$$

- Hypothetical number of passengers boarding

- Hypothetical departure time

$$\bar{B}_s^n = S_s^n + \beta_s (\bar{d}_s^n - a_s^n)$$

$$\bar{d}_s^n = a_s^n + ad_s L_s^n + b \bar{B}_s^n$$

$L$  is the number of passengers in the bus

$$\bar{d}_s^n = a_s^n + \frac{1}{1 - b\beta_s} (ad_s L_s^n + bS_s^n) \quad L_{n,s+1} = L_{n,s} - dL_{n,s} + B_{n,s}$$

$$\bar{B}_s^n = S_s^n + \frac{\beta_s}{1 - b\beta_s} (ad_s L_s^n + bS_s^n) \quad \text{hypothesis: all passengers can board}$$

- Then look at capacity
  - Bus can accommodate the hypothetical number

If residual capacity + alighting > waiting, then

Check that the departure time is not earlier than the scheduled departure (if not hold)

  - Bus cannot accommodate the hypothetical number then boarding=residual capacity and departure time= min (theoretical , feasible)

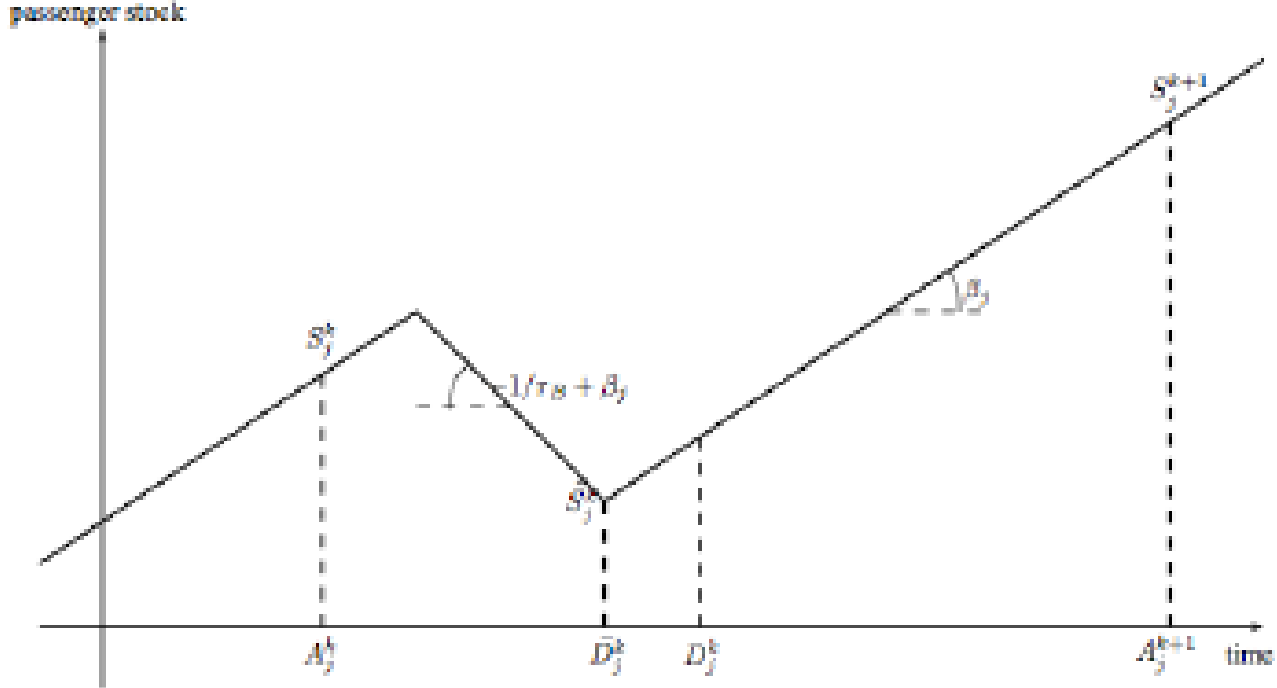


Figure 1: number of passengers waiting at a station, function of time

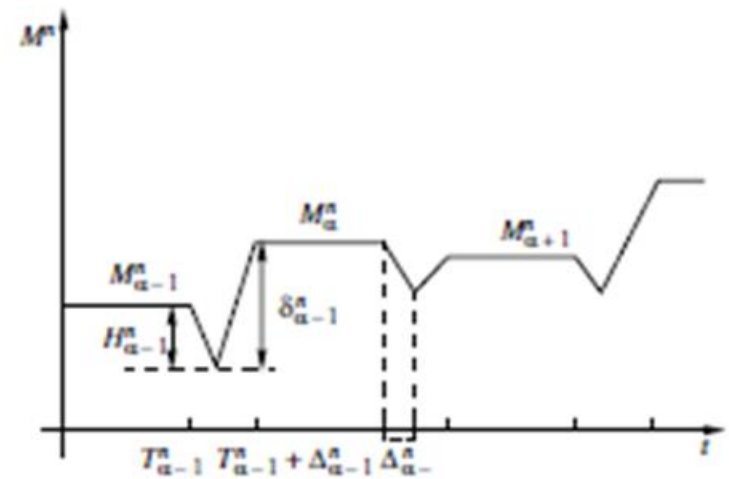


Fig. 3. Scheme of the change of bus filling with time.

# Performance indicators

- Total passenger travel time

$$\sum_{s=1}^{S-1} (a_s^n - d_{s-1}^n) L_s^n + \sum_{s=1}^{S-2} (L_{s+1}^n - B_s^n) (d_s^n - a_s^n)$$

- Waiting time

- Delay

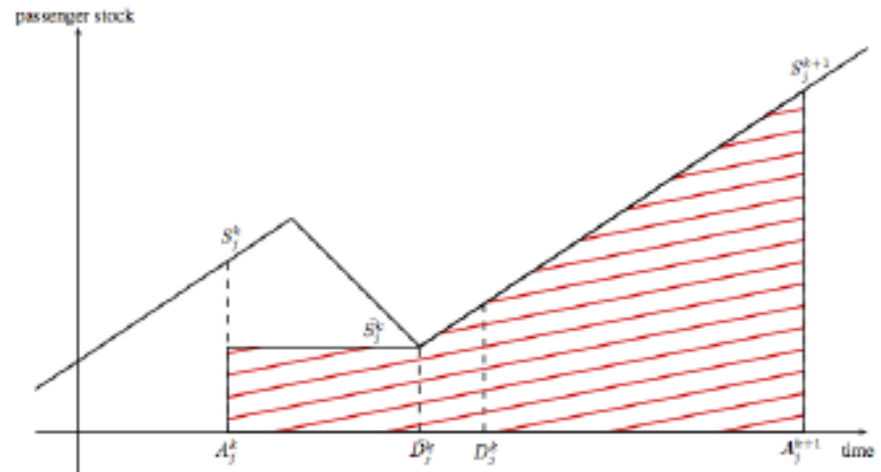


Figure 2: Area representing total waiting time at a station



# User cost component between i and j

- $C(t)$  travel disutility when leaving at time  $t$  with schedule delay cost or head start (time allocated to waiting)

$t^*$  preferred arrival time,  $T$  travel time

$$C(t) = \alpha_w T_w + \alpha_v T_v + \beta(t^* - t - T_w)^+ + \gamma(t + T_w - t^*)^+$$

- Excess waiting time at the origin stop  $i$  (probability to miss the targeted departure)
- Potential travel time called the travel time buffer  $A_j^{95th} - E(A_j)$
- Mean riding time  $E(A_j) - E(D_i)$

- Excess waiting time :  $On(s)$  or boarding
- Potential travel time :  $OFF(s)$  or alighting
- Mean riding :  $OD(i,j)$  but in fact  $ON(i)$  and  $OFF(j)$

# Conclusion

- Well known stochastic processes for bus and passenger movements. Interesting non stable dynamics with bus bunching
- Fit of Statistical distributions for travel and dwell times
  - Data problems to solve with GPS (missing data)
  - More than one candidate for travel times, dwell times, headways: non standard distributions (very skewed)
- Use of statistical models to fit Daganzo's models
- Comprehensive transition to performance indicators

# Bibliography

- Daganzo, C.F., 2009. A headway-based approach to eliminate bus bunching: Systematic analysis and comparisons. *Transportation Research Part B: Methodological*, 43(10), pp.913-921.
- Nair, R., Bouillet, E., Gkoufas, Y., Verscheure, O., Mourad, M., Yashar, F., Perez, R., Perez, J., Bryant, G. and Transit, M.D., 2014. Data as a resource: real-time predictive analytics for bus bunching.
- Hans E. et al., 2015. Real-time bus route state forecasting using particle filter and mesoscopic modeling. *Transportation Research Part B*, 61, pp.121-140.
- Hans E. et al., 2015. Investigating the irregularity of bus routes: highlighting how underlying assumptions of bus models impact the regularity results. *J. Adv. Transp.* ; 49:358–370
- Combes F., Benezech V., 2014. Quality of service and congestion in urban public transport : a simulation approach. Conference AET.
- Islam K. et al., 2014. A bulk queue model for the evaluation of impact of headway variations and passenger waiting behavior on public transit performance. *IEEE transactions on intelligent transport systems*, 15(6).
- Furth P., Muller T., 2007. Service reliability and optimal running time schedules. *Transportation Research Record*, 2034, pp 55-61.