

TransitData 2019, Paris - La Sorbonne, July 10th

A Probabilistic Traffic Model to Estimate Fail-to-Board Probabilities in Transit Lines on the basis of AFC and AVL data

Fabien Leurent and Thomas Jasmin

*Ecole des Ponts ParisTech, Laboratoire Ville Mobilité Transport,
Urban Mobility Chair in partnership with IDFM*



Motivation

Train platform may be loaded with user flows larger than residual capacity in incoming train



Resulting in Failure-to-Board (F2B) for some users (Schmöcker, 2006) that are “left behind” (Zhu et al, 2017)

Objective

Research question

To define, model and estimate the Fail-to-Board Probabilities

Contribution

- + A Probabilistic Model involving F2B within “Transition Probabilities”
- + Maximum Likelihood (ML) estimation method
- + Application to Paris line A of RER (Regional Express Railway)

F2B probabilities: definition and measurement

Definition

Among users waiting for an incoming train, the probability of failing to board it

Thus it is particular to train run i and station r : denote it as $\bar{\rho}_{i,r}$

Measurement issues

Direct measurement:

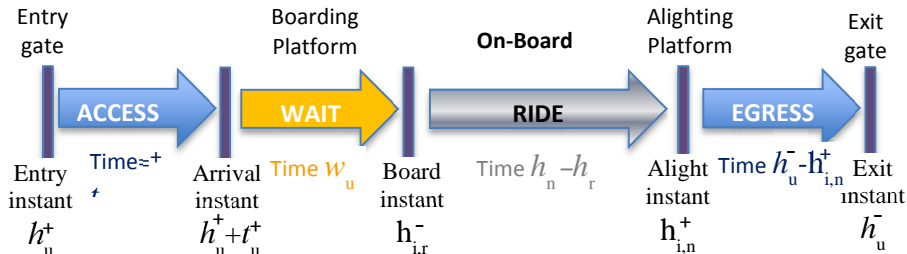
get numbers $B_{i,r}$ of boarded passengers and $N_{i,r}$ of users remaining on platform, to derive $\bar{\rho}_{i,r} = N_{i,r} / (B_{i,r} + N_{i,r})$

Indirect measurement: based on AFC data for users' trips and AVL data for train runs



SITUATION & NOTATION

Situation



Notation

Random variable $t_u^+ \Rightarrow$ CDF denoted as $T_u^+(x) \equiv \Pr\{t_u^+ \leq x\}$

PDF denoted as $\dot{T}_u^+(x) \equiv \partial T_u^+(x) / \partial x$

Similarly: for an instant h_u^- , CDF $H_u^-(x)$ and PDF $\dot{H}_u^-(x)$

Physical and probabilistic model: 4 legs in u 's trip

1/ Access: from entry validation gate at h_u^+ to platform, in time t_u^+ to arrive during the HEADWAY of run k such that $h_u^+ + t_u^+ \in]h_{k-1,r}^-, h_{k,r}^-]$

$$\Pr\{k | h_u^+\} = T_r^+(h_{k,r}^- - h_u^+) - T_r^+(h_{k-1,r}^- - h_u^+)$$

2/ Wait: during w_u up to boarding run i and departing at $\hat{h}_u = h_{i,r}^-$, implying that $w_u = \hat{h}_u - (h_u^+ + t_u^+)$

Transition probability $\pi_{i|k}^{(r)}$ between two runs $k \leq i$

3/ Ride: from station r to station n , hence in time $h_{i,n}^+ - h_{i,r}^-$

4/ Egress: from alighting platform n to exit validation gate at h_u^- , in time $t_u^- \equiv h_u^- - h_{i,n}^+$. As $\{h_u^- \leq x | i\} = \{h_{i,n}^+ + t_u^- \leq x | i\}$, it holds that

$$\Pr\{h_u^- \leq x | i\} = \Pr\{t_u^- \leq x - h_{i,n}^+\} = T_n^-(x - h_{i,n}^+)$$

From F2B to Transition probabilities

The MINGLING postulate

All users waiting on the boarding platform at a particular instant have equal chance of boarding an incoming vehicle

Induction algorithm

Denote $\rho_{ir} \equiv 1 - \bar{\rho}_{i,r}$ the success probability,

and $\tilde{\pi}_{i|k}^{(r)}$ the probability of failure to board from k to i included

Initialization: $\pi_{k|k}^{(r)} = \rho_{kr}$

Induction step: calculate $\tilde{\pi}_{i+1|k}^{(r)} = \bar{\rho}_{i+1,r} \tilde{\pi}_{i|k}^{(r)}$ and $\pi_{i+1|k}^{(r)} = \rho_{i+1,r} \tilde{\pi}_{i|k}^{(r)}$

It's easy!

Exit time conditionally to entry time

Principle

Disaggregation according to arrival headway k and boarded run i :

$$\{h_u^- \leq x \mid h_u^+\} = \bigcup_{k,i} \{h_u^- \leq x \cap i \cap k \mid h_u^+\}$$

$$\begin{aligned} \Pr\{h_u^- \leq x \mid h_u^+\} &= \sum_{k \leq i} \Pr\{h_u^- \leq x \cap i \cap k \mid h_u^+\} \\ &= \sum_{k \leq i} \Pr\{h_u^- \leq x \mid h_u^+, k, i\} \Pr\{i \mid k\} \Pr\{k \mid h_u^+\} \\ &= \sum_{k \leq i} T_n^-(x - h_{i,n}^+) \cdot \pi_{i|k}^{(r)} \cdot [T_r^+(h_{k,r}^- - h_u^+) - T_r^+(h_{k-1,r}^- - h_u^+)] \end{aligned}$$

Characterization of random variable h_u^- conditionally to h_u^+

Conditional CDF is obtained as $H_u^-(x) \equiv \Pr\{h_u^- \leq x \mid h_u^+\}$ above

$$\dot{H}_u^-(x) \equiv \Pr\{h_u^- = x \mid h_u^+\} \text{ by partial derivation}$$

$$\begin{aligned} \text{PDF:} & \\ &= \sum_{k \leq i} \pi_{i|k}^{(r)} \dot{T}_n^-(x - h_{i,n}^+) \cdot (T_r^+(h_{k,r}^- - h_u^+) - T_r^+(h_{k-1,r}^- - h_u^+)) \end{aligned}$$

Estimation scheme

Likelihood function

$$\dot{H}_u^-(h_u^-) = \sum_{k \leq i} \pi_{i|k}^{(r)} \dot{T}_n^-(h_u^- - h_{i,n}^+) \cdot (T_r^+(h_{k,r}^- - h_u^+) - T_r^+(h_{k-1,r}^- - h_u^+))$$
$$\Leftrightarrow L(\Theta \text{ in } T_r^+, \dot{T}_n^-, \pi_{i|k}^{(r)} | u, k, i)$$

Maximum Likelihood Estimation

Log-likelihood function of one observation: $\Lambda_u(\Theta) \equiv \ln L_u(\Theta)$

Parameter Θ : distributions T_r^+ and T_n^- and transition probabilities $\pi_{i|k}^{(r)}$

Maximize total log-likelihood function

$$\Lambda_U(\Theta) \equiv \ln L_U(\Theta) = \sum_{u \in U} \Lambda_u(\Theta)$$

w.r.t. Success-to-Board Probabilities ρ_{kr} , s.t. $0 \leq \rho_{kr} \leq 1$

Comparison to Model by Zhu et al (2017)

MIT

ACCESS proba & EGRESS proba
 No JOINT integration w.r.t. speed
 Notion of “*Left behind probability*”:
 from run k to run i as P_{i-k} :
 it depends neither on the boarding
 station nor on the individual runs as it
 has one value only for all run pairs of
 given value i-k

ESTIMATION METHOD

ML with Bayesian framework

ENPC

ACCESS proba & EGRESS proba
 WITH JOINT integration w.r.t. speed
Transition probability from run k to run
 i as $\pi_{i|k}^{(r)}$: it depends on the individual
 runs and also on the boarding station
 There is a PHYSICAL model behind the
 Stochastic model, from F2B proba to
 Transition proba

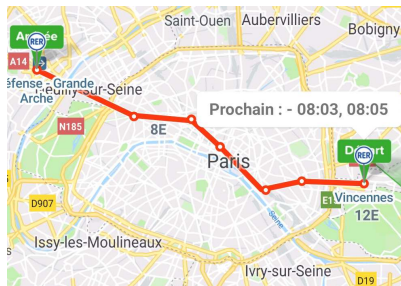
ESTIMATION METHOD

ML on F2B proba

Train line and O-D pair

Train line

RER A => focus on common trunk



(Source: CityMapper)

Origin-Destination pair

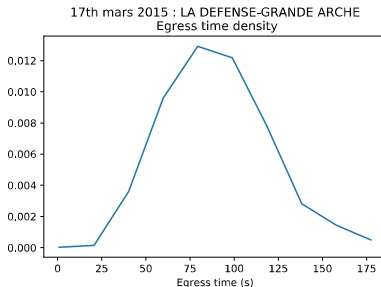
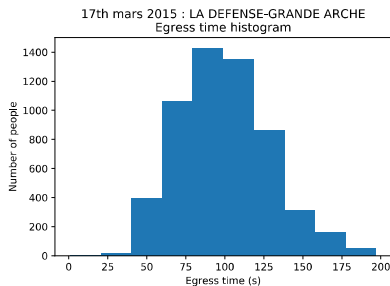
- From Vincennes to La Défense
- 27 trains / hour at morning peak
- About 20' of ride time
- Vincennes: simple platform at level -1 with 2 access points
- La Défense: complex platform at level -3 with 3 access points and many validation gates at different places and levels

Determination of access and egress time distributions T_r^+ and T_n^-

Direct estimation from AFC and AVL data on March 17th, 2015

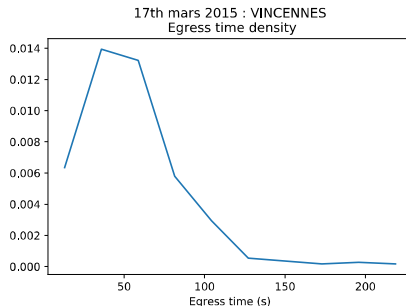
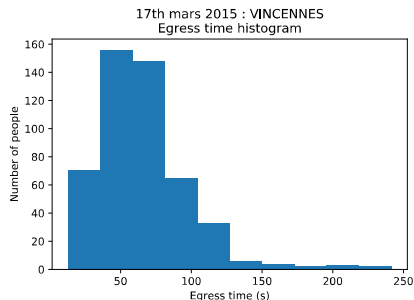
Selection of users that can physically board in only one train

Egress time at a station: difference between tap out time and train arrival at this station



Concerning the Vincennes station

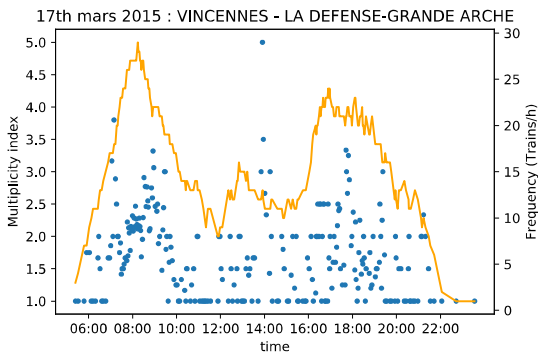
Hypothesis: access times and egress times at Vincennes have identical statistical distributions



Combining AFC & AVL datasets: determination of train multiplicity index

Multiplicity index of train k : Mean of the numbers of available trains for travellers that could board into train k on the basis of respective times

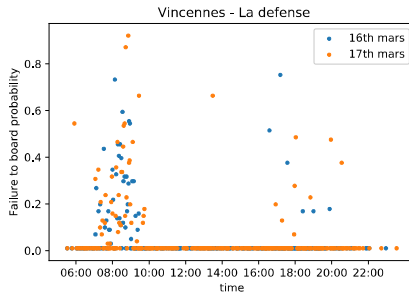
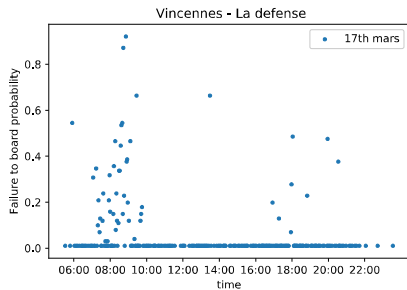
First idea of the panel of train possibilities



Depend on:

- Train frequencies
- Failure to board probabilities

Failure to board: statistical determination by Maximum Likelihood estimation

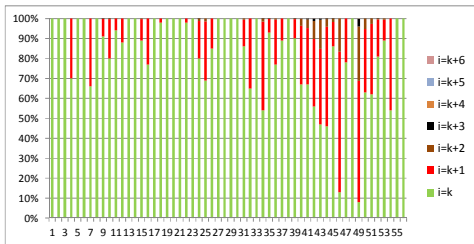


Comments

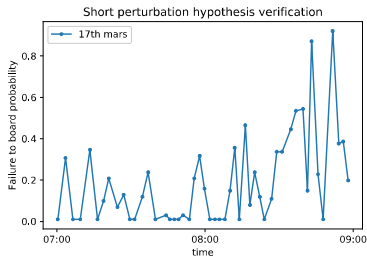
Null F2B value for off peak trains and most of trains at peak

Yet, at peak, several trains have positive F2B values, with quite large value for some of them (up to 0.9)

Contribution to Traffic analysis



*Transition probabilities
between 7 and 9am*

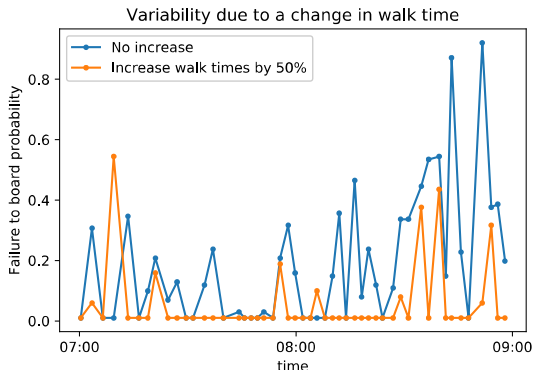


Saturated intervals are short

- Presence of consecutive non zero values
- 2 or more missed trains: quite rare
- Next step: solve the full estimation problem

Discussion: walk times

Results depend on the choice of walk time distribution (hypothesis of one single distribution for the entire day, some passengers are not taken into account in the determination of distributions, possible dependence between failure to board probability and walk time)



Conclusion

Contribution

Probabilistic and physical model that captures F2B probabilities

Maximum Likelihood Estimation scheme

Computation scheme

Next steps

Focus on Metro line 13 along North-South axis

Acknowledgment

IDFM for provision of AFC data and its support to Urban Mobility Chair

RATP for provision of AVL data