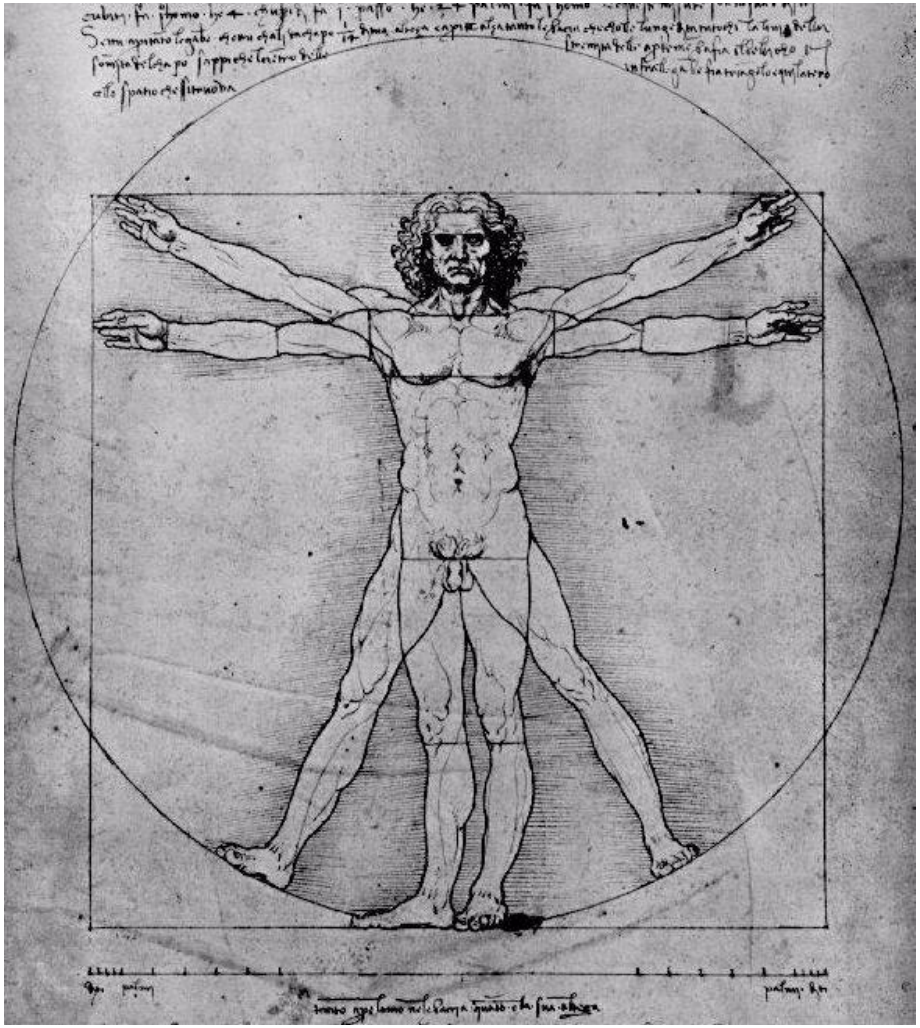
Smart Route Planning for Public Transit

Giuseppe F. Italiano

Univ. of Rome “Tor Vergata”

giuseppe.italiano@uniroma2.it

Theory

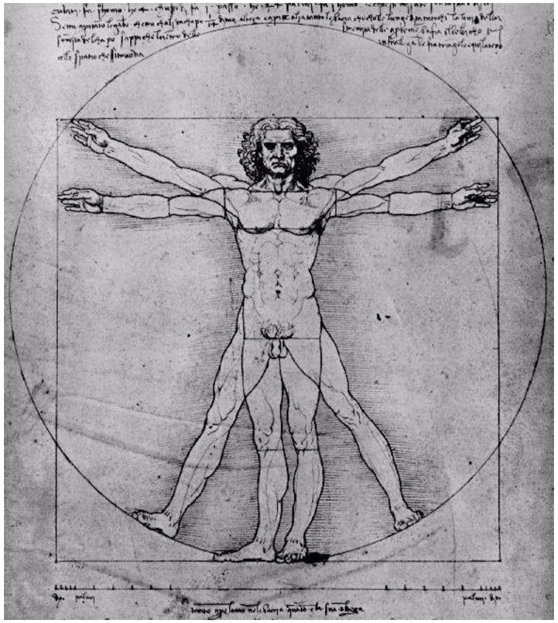


Theory is when you know something, but it doesn't work.

The real world out there…

Practice is when something works, but you don't know why.

Bridging the Gap between Theory and Practice?



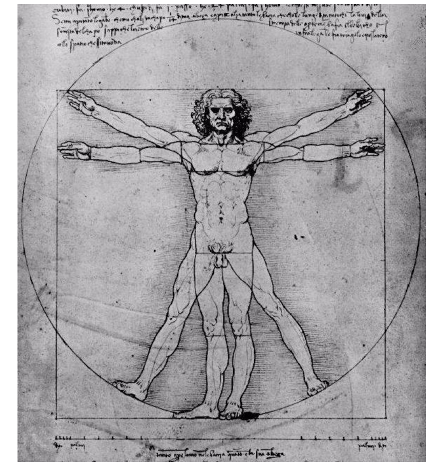
Big challenge: combine theory and practice…

Practice is when something works, but you don't know why.

Theory is when you know something, but it doesn't work.

…i.e., nothing works and you don't know why.

Public Transit Route Planners



Vast majority based on timetabling

Theory: all transit vehicles start their trip at the planned time, no delays throughout their journey

Practice: buses often run behind schedule, and for many unplanned reasons, such as:

- traffic
- road closures
- inclement weather
- sometimes even unrealistic scheduling

Outline

1. Are timetabled route planners good enough for public transit?

2. Investigate impact of GPS data: run same algorithm with timetabled and GPS data

3. Design of GPS-aware algorithms

Outline

1. ***Are timetabled route planners good enough for public transit?***

2. Investigate impact of GPS data: run same algorithm with timetabled and GPS data

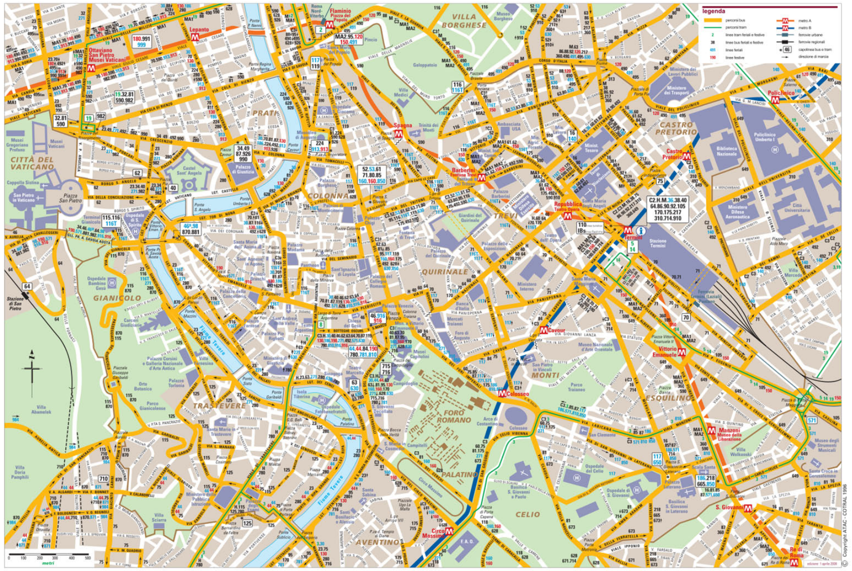
3. Design of GPS-aware algorithms

[Firmani, I., Laura, Santaroni, ATMOS 2013]

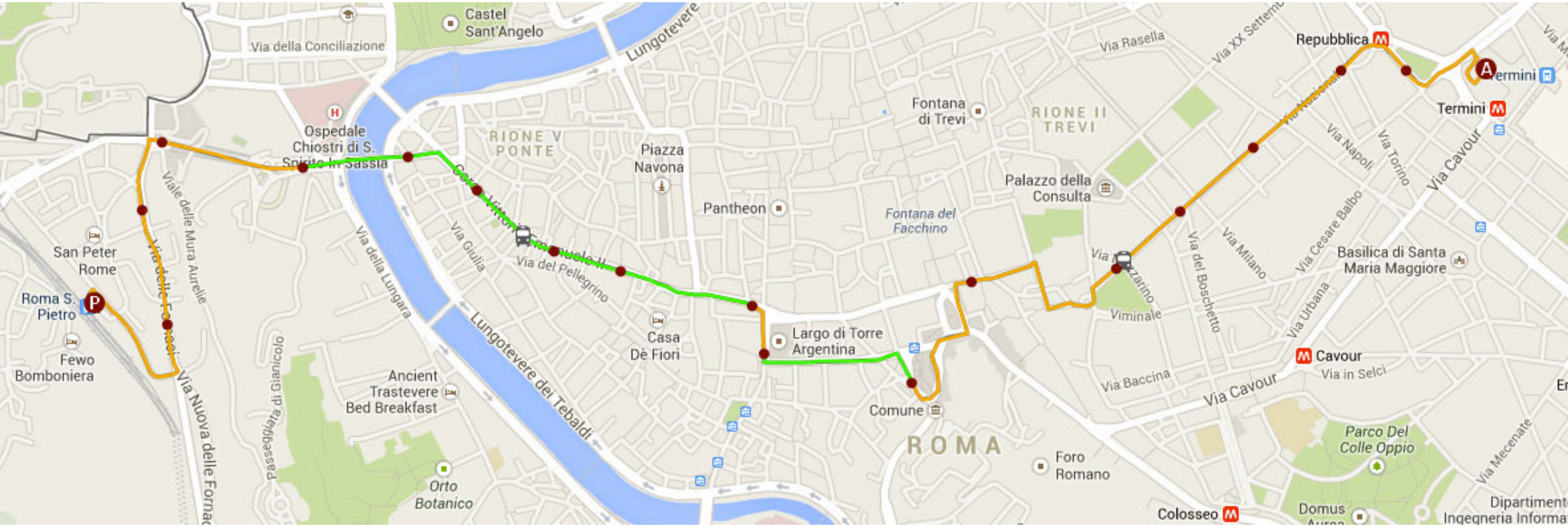
Timetabled Route Planners

- Are **timetabling-based routing methods** able to deliver “good” solutions in practice?

- Try to **measure the quality** of timetabling-based solutions in the metropolitan area of a big city.



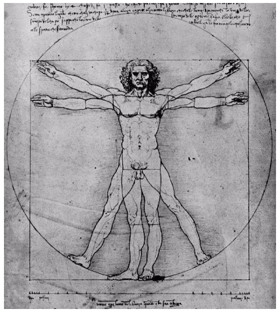
Analysis of bus trips

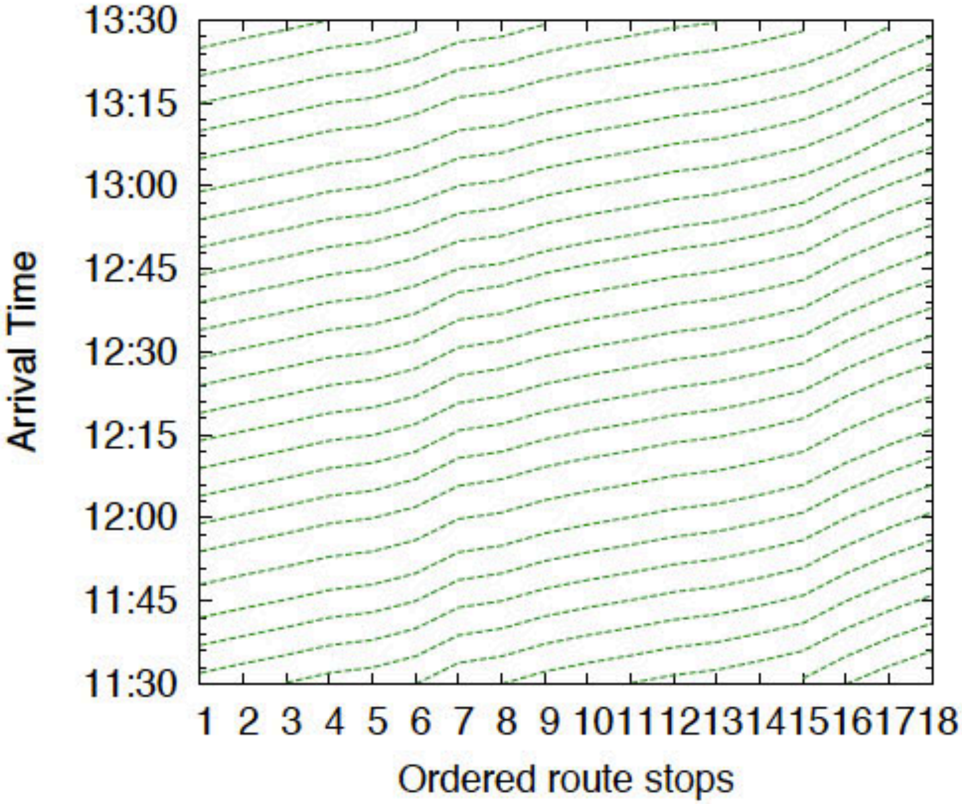


Bus 64, route between S. Pietro to main train station in Rome (Termini)

Connects many important touristic attractions in the city center of Rome

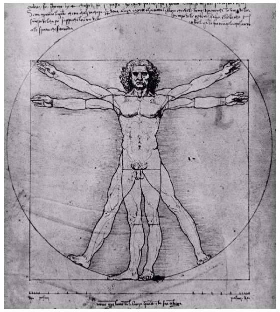
Analysis of bus trips

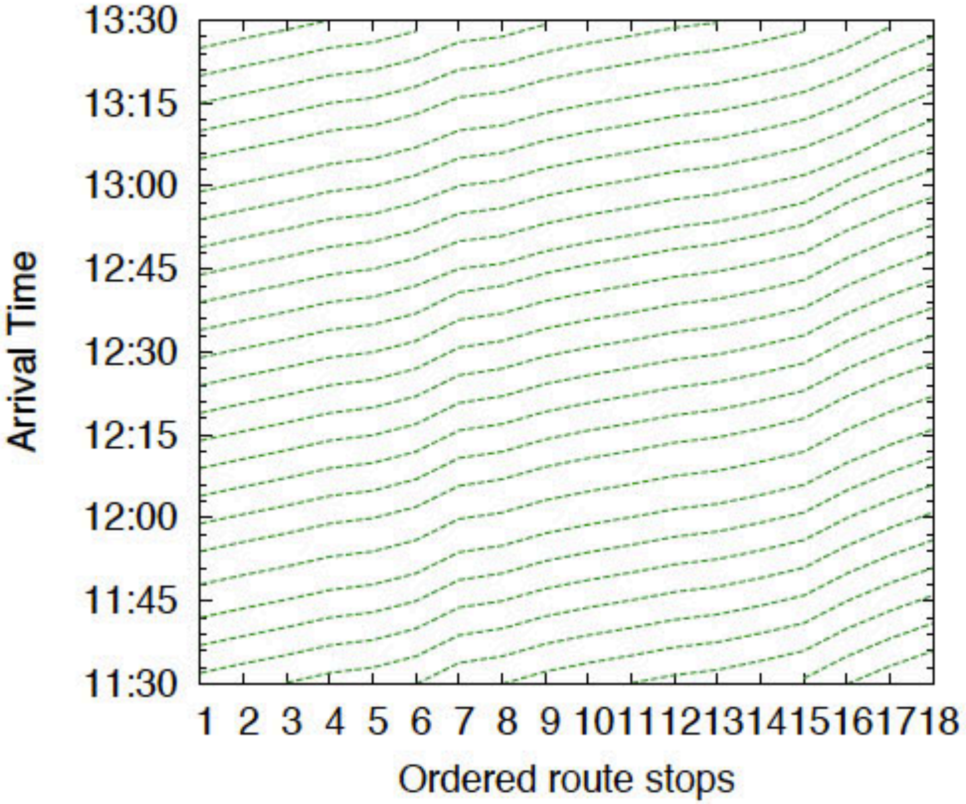




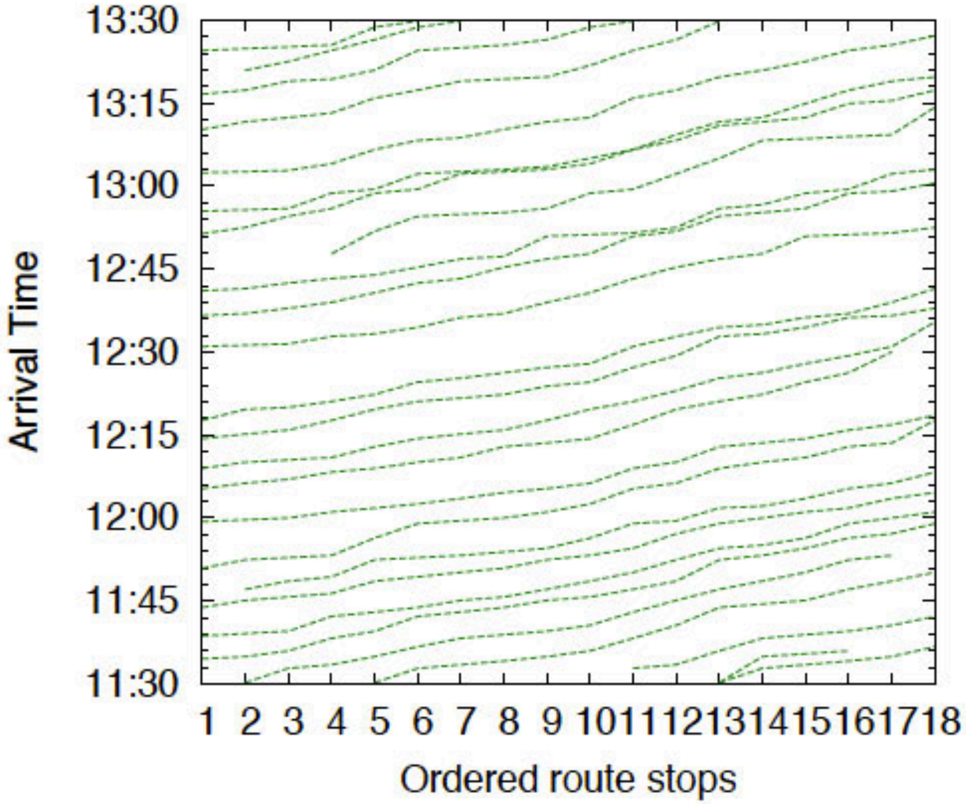
(a) Official timetable (11:30am–1:30pm)

Analysis of bus trips



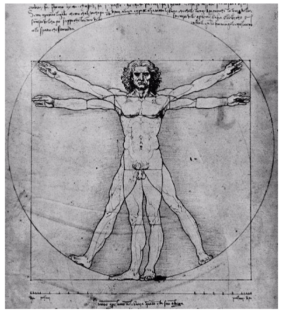


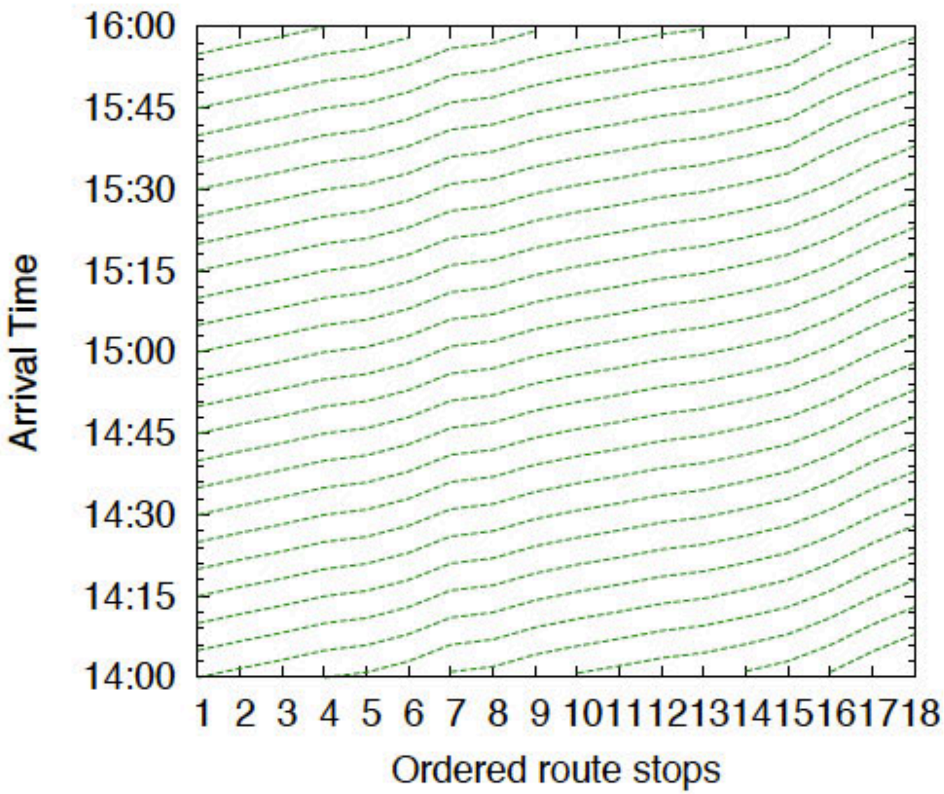
(a) Official timetable (11:30am–1:30pm)



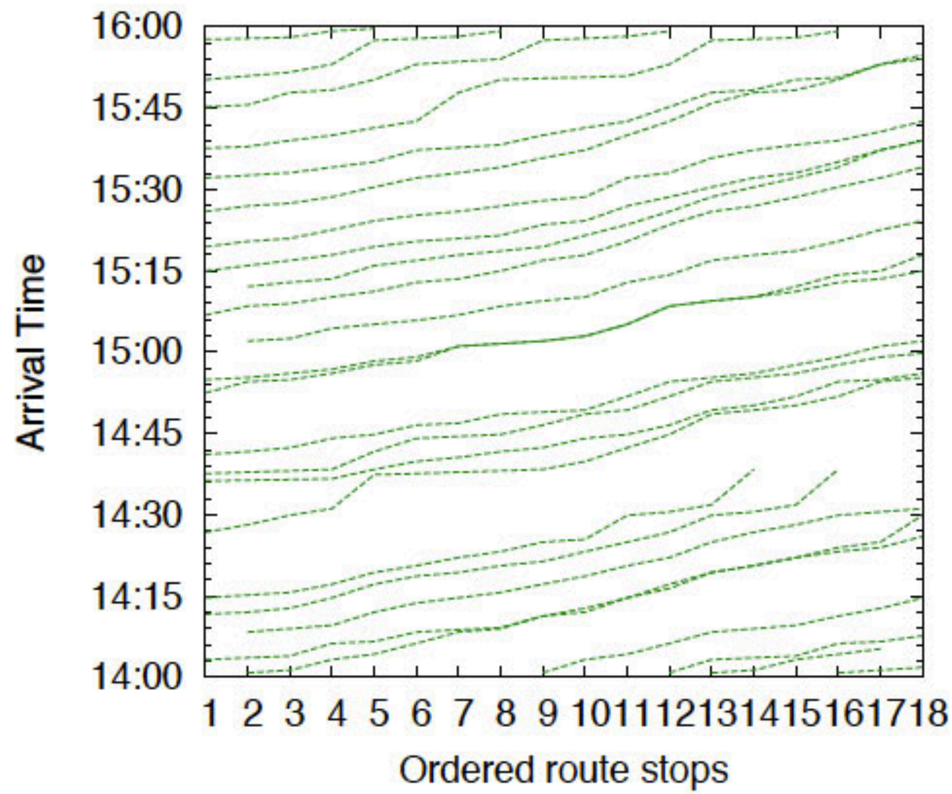
(b) GPS stream (11:30am–1:30pm)

Analysis of bus trips



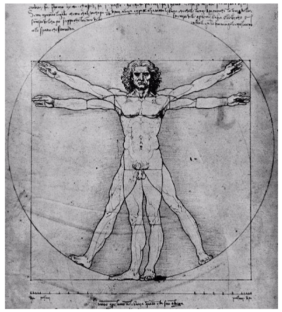


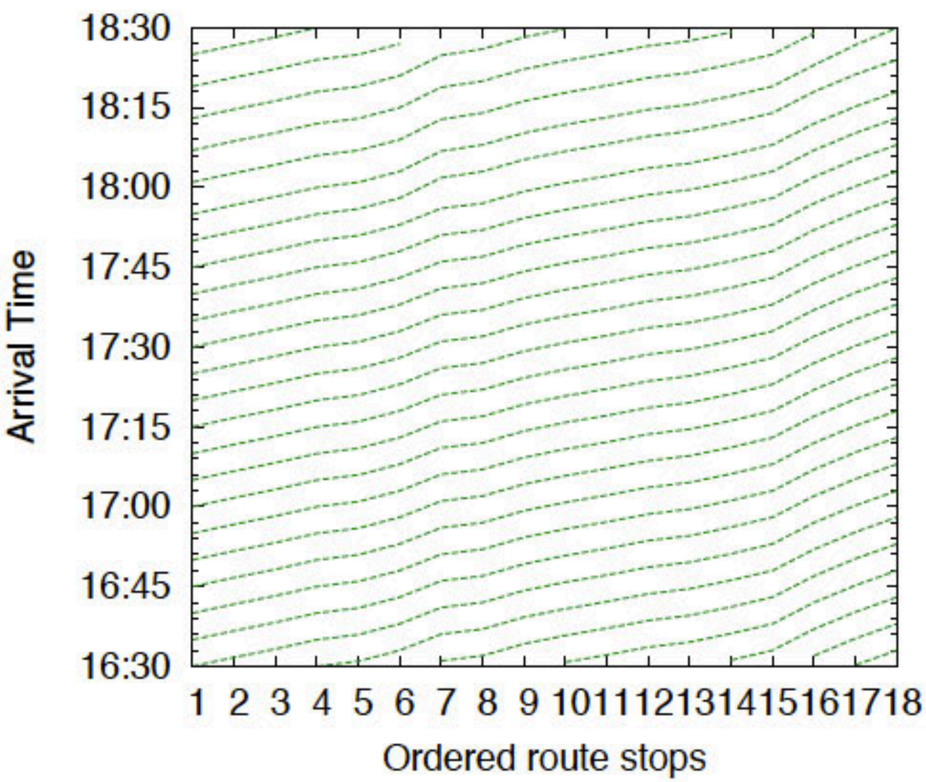
(c) Official timetable (2:00pm–4:00pm)



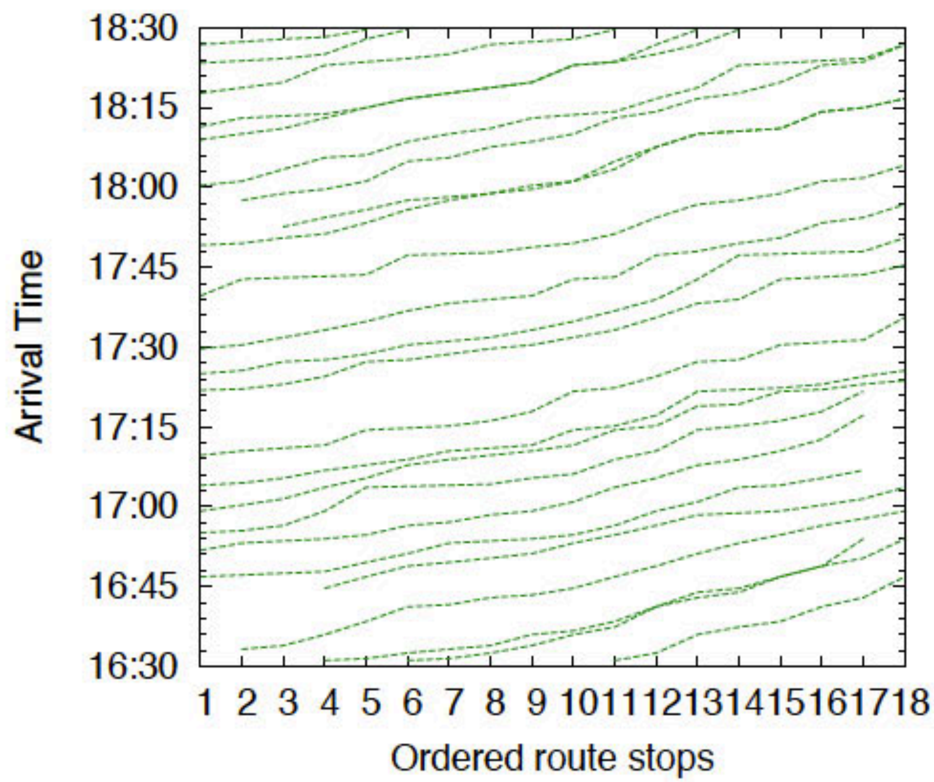
(d) GPS stream (2:00pm–4:00pm)

Analysis of bus trips





(e) Official timetable (4:30pm–6:30pm)



(f) GPS stream (4:30pm–6:30pm)

Experiment (1/2)

How does this affect a route planner?

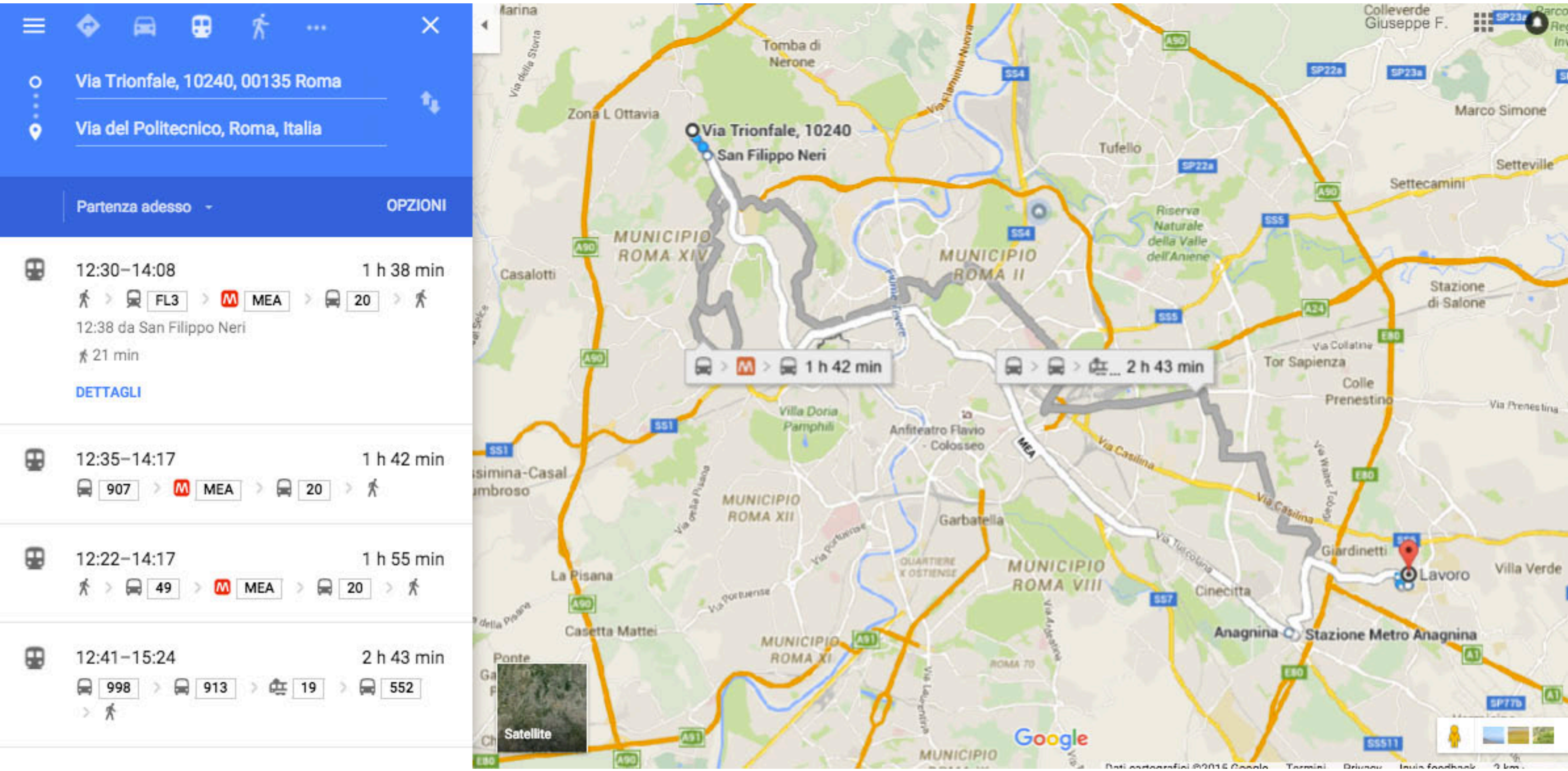
Submitted queries to Google Transit

- *Origin* and *destination* in metropolitan area of Rome
- Each query has a *departure time* from origin

Collected suggested *journeys* and their **estimated travel times**

- Journeys provided by Google Transit based on **timetable data**

Google Transit



Experiment (2/2)

“Follow” each journey through GPS data

- GPS data → real-time position of transit vehicles (provided by the very same transit agency)
- Simulate the experience of a user traveling according to a given journey, after leaving the origin at the corresponding time

Compare **actual travel times** against their original estimates

Experimental Setup

- Public Transport Network in Rome
  - 309 bus lines and 2 subway lines
  - 7,089 stops (7,037 bus stops and 52 subway stops)
- More than 1,000 queries $q_i = \langle s_i, t_i, \tau_i \rangle$
  - s_i, t_i uniformly at random in Rome's metropolitan area
  - τ_i uniformly at random in [7:00am; 9:00pm]
  - On Wed January 8, 2014 (just a typical week day, no major disruptions, constructions works or extreme weather conditions)
- Collected **4,018** journeys
  - For each query, Google Transit returns 4 journeys
  - Possibly less: e.g, one of the journeys might be footpath

Our Simulation

On same day (January 8) submitted queries every minute to the Mobility Agency of Rome

- Obtained (from GPS data) the instantaneous geo-location of all transit vehicles in the network

Simulation

- For each *trip* in the journey, pick-up and drop-off times computed according to the position of transit vehicles
- If a delayed transit vehicle misses a connection then the next trip of the same line was chosen (similarly, could catch earlier connections)
- Footpath times computed with Google Maps
- No GPS data for subway lines (two lines, usually on schedule)

Experimental Results

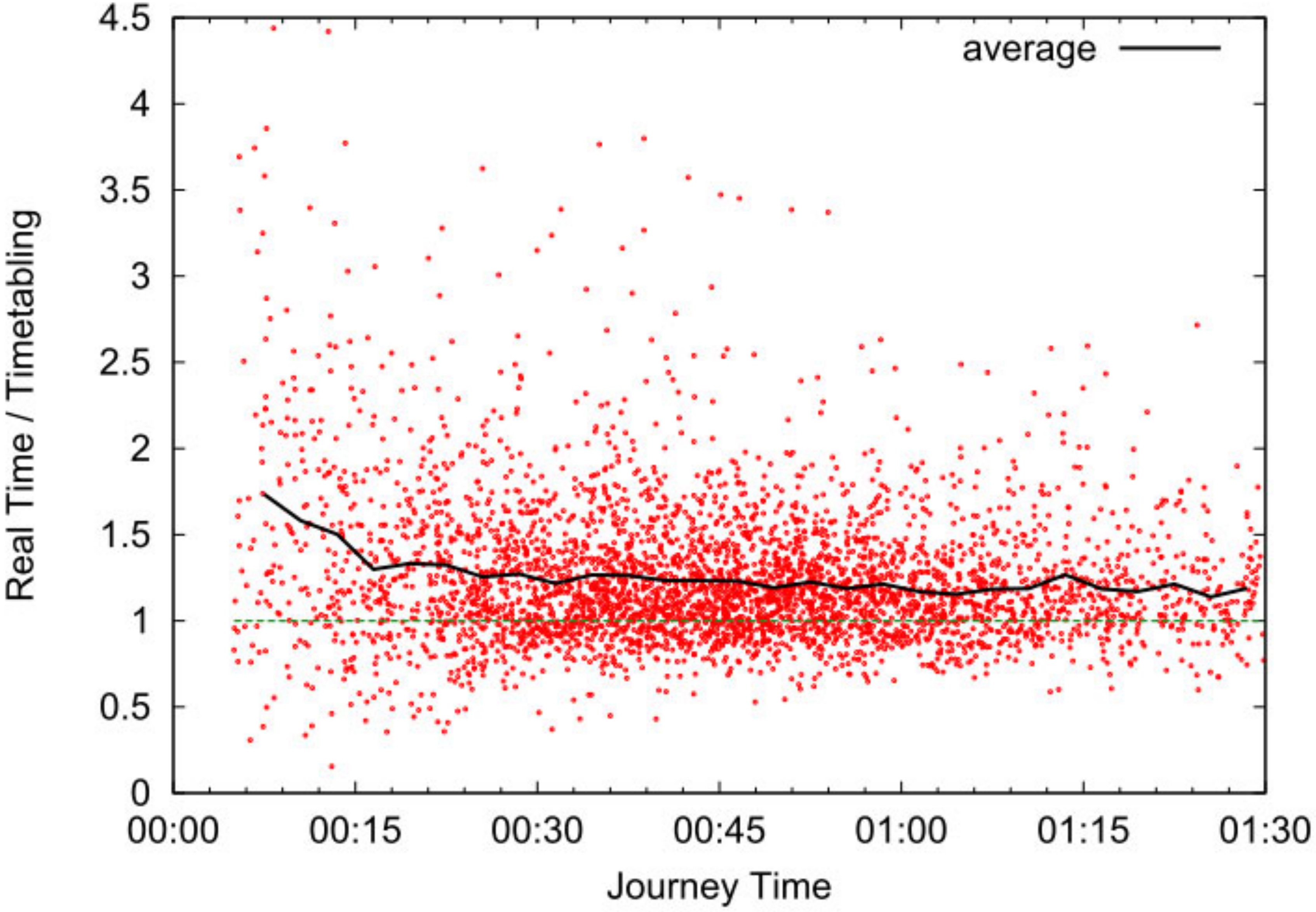
For each **journey** *j*, compare $t_a(j)$ and $t_e(j)$

- $t_e(j)$: estimated (planned) time (timetable ← Google Transit)
- $t_a(j)$: actual time (vehicle real-time positions ← simulator)

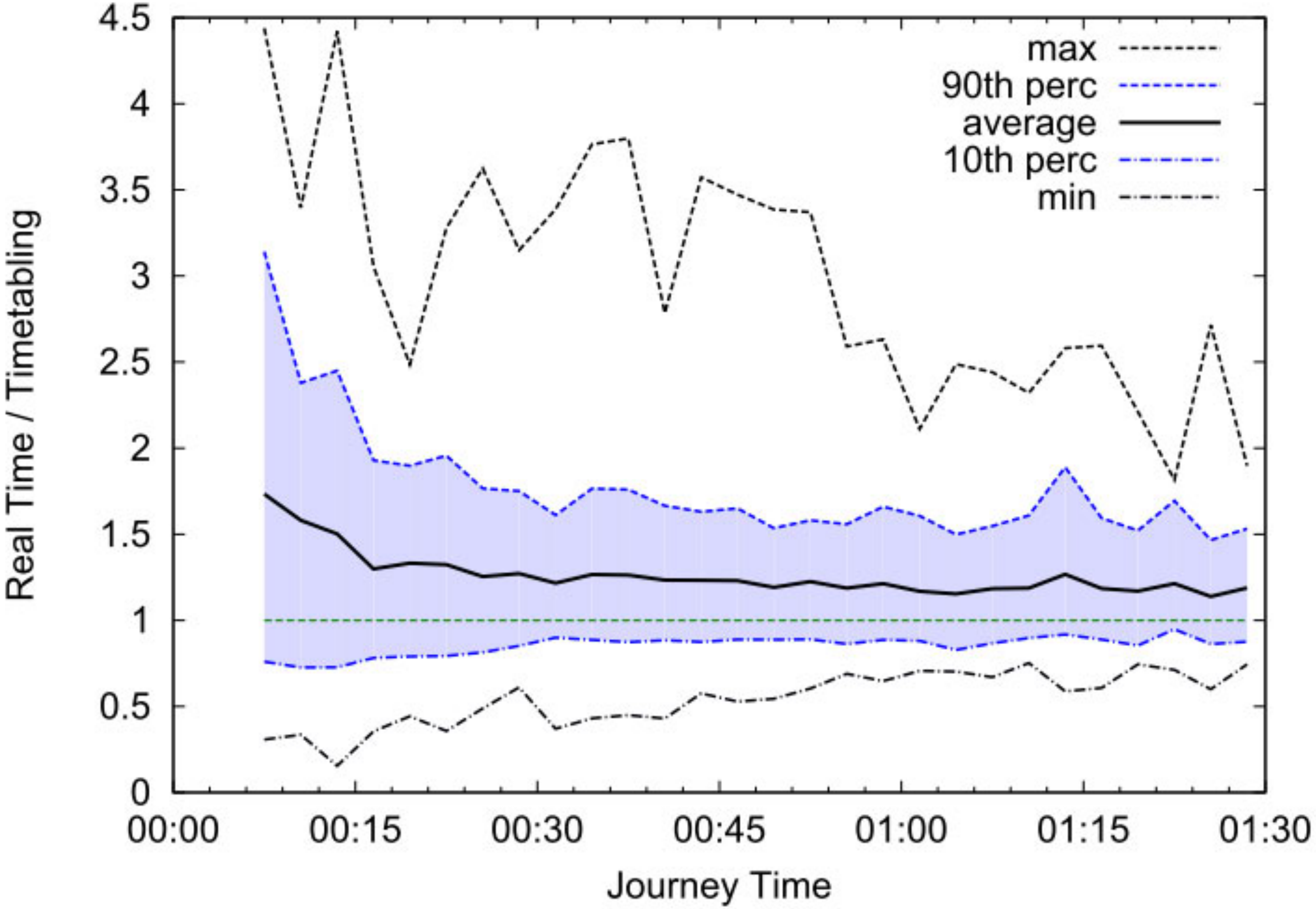
Define **error coefficient** $t_a(j)/t_e(j)$

- Plot $t_a(j)/t_e(j)$ as $t_e(j)$ increases:
  - **short distance** journeys : $t_e(j) \leq 30$min
  - **medium distance** journeys : 30min $< t_e(j) \leq 60$min
  - **long distance** journeys : $t_e(j) > 60$min
- Group journeys into **time slots** within a 3-minute resolution and plot statistics as $t_e(j)$ increases

Error Coefficient (1/2)

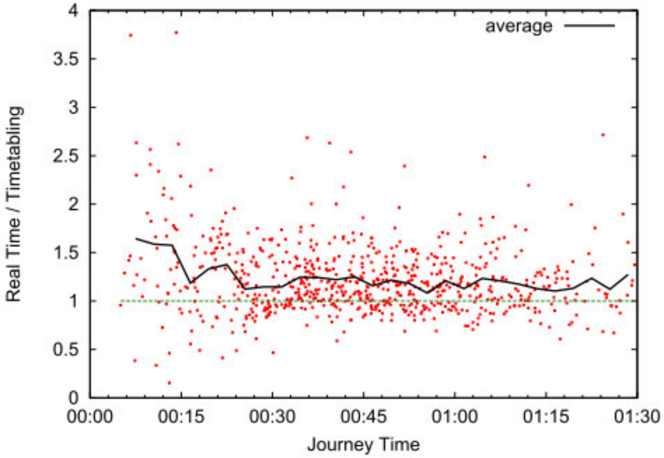


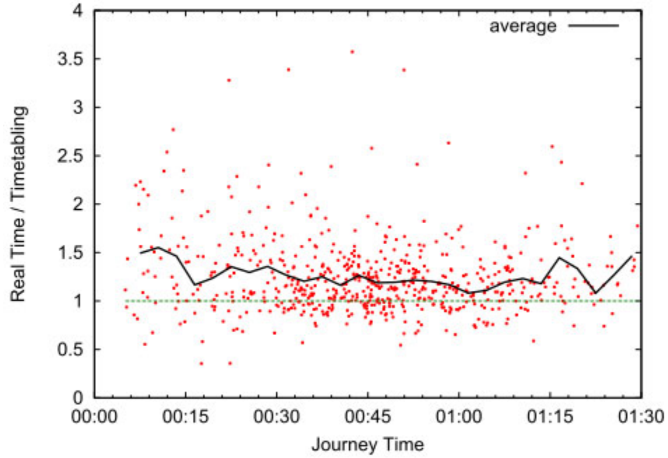
Error Coefficient (2/2)

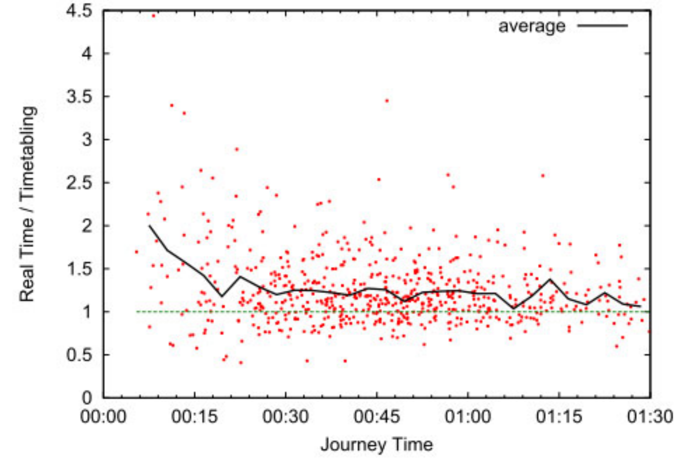


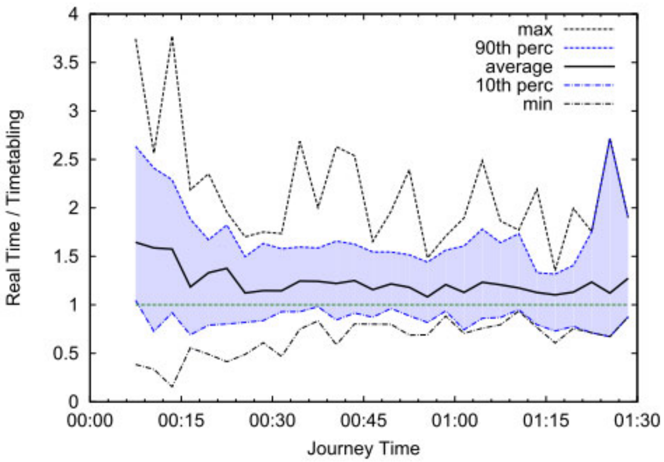
Different times of day

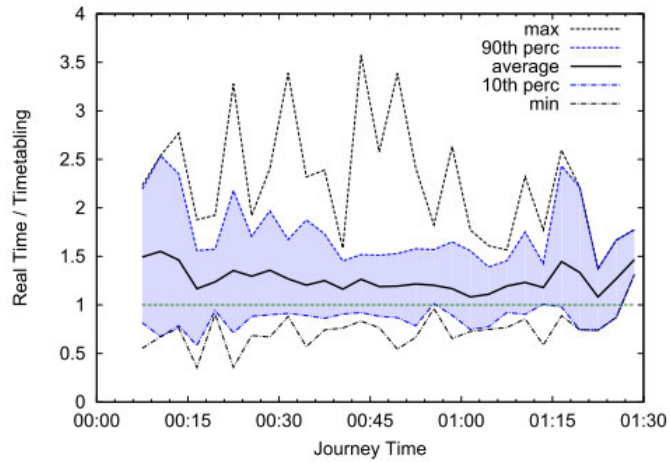
7:30am - 9:30am	11:30am - 1:30pm	5:00pm - 7:00pm

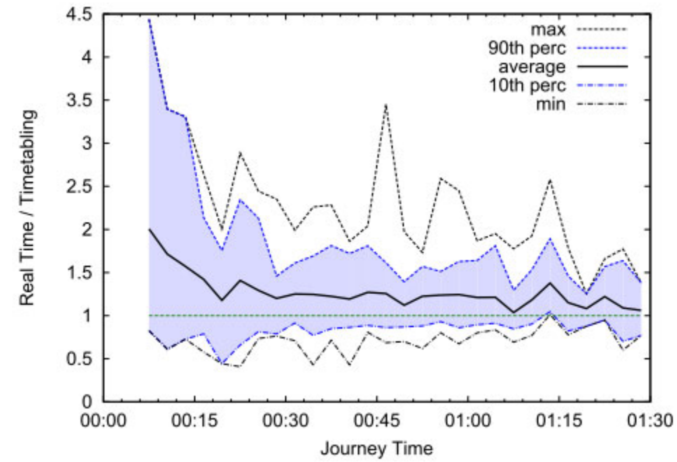












Correlations in ranking

For each **query** q_i, take four journeys suggested and compare:

- Relative rankings by **estimated** times $t_e(j)$
- Relative rankings by **actual** times $t_a(j)$

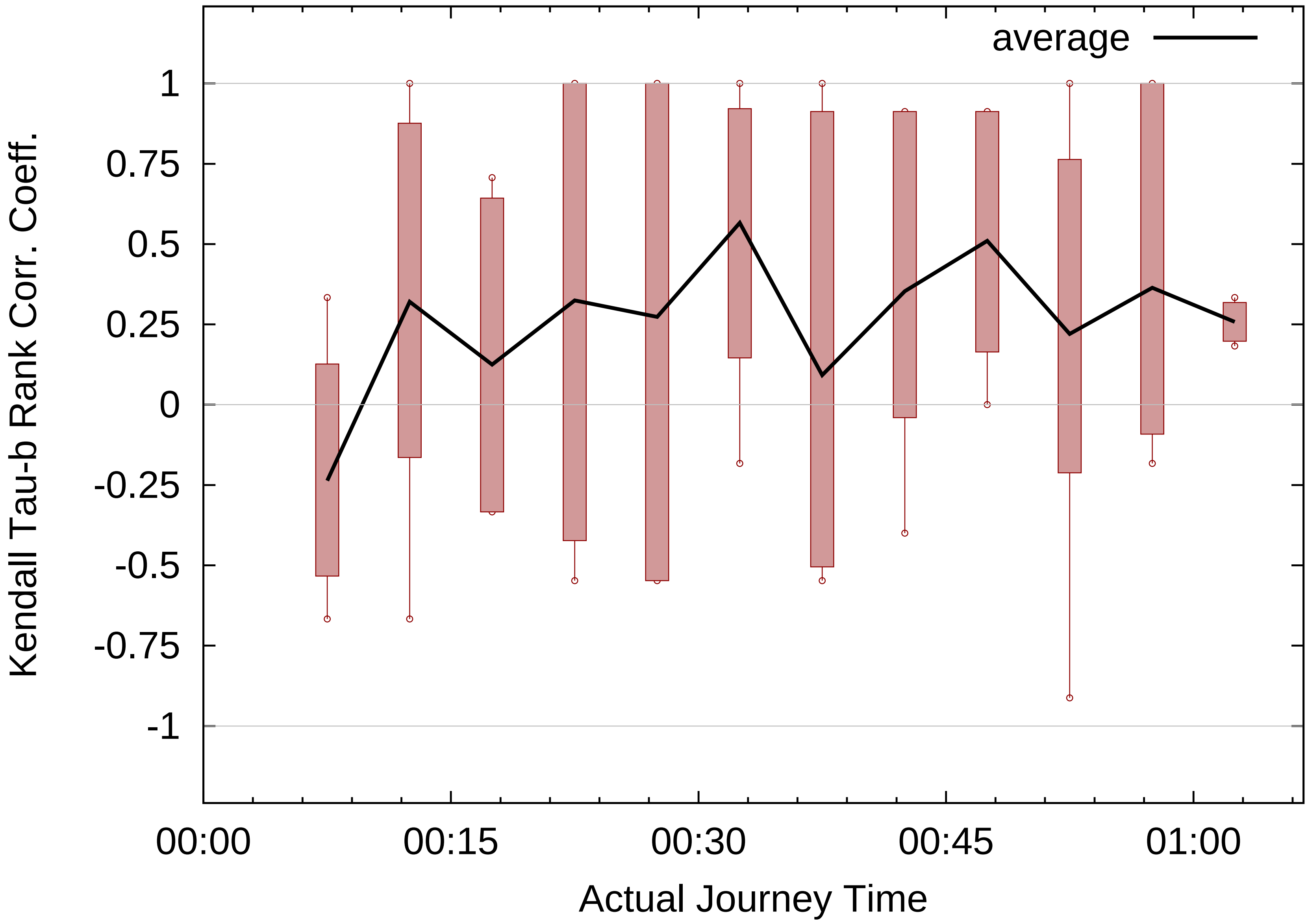
Use **Kendall tau coefficient**

$$\frac{(\#\ \text{concordant pairs}) - (\#\ \text{discordant pairs})}{\frac{1}{2}\, n\, (n-1)}$$

- From -1 (perfect disagreement) to +1 (perfect agreement)
- 0 if independent (uncorrelation)

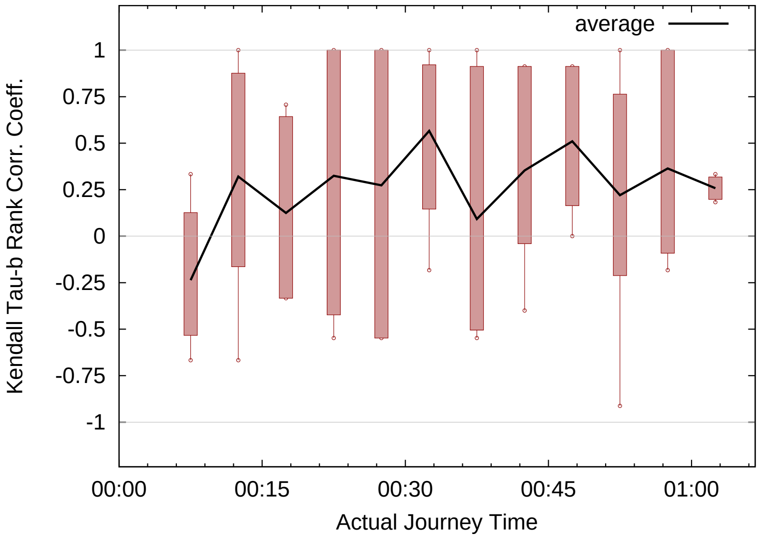
Actually used Kendall tau b (to adjust for ties)

Kendall Tau Coefficient



Kendall Tau Coefficient

Only very mild correlation between *estimated* and *actual* duration of a journey.



Smaller Kendall values for short journeys: they appear more vulnerable to fluctuations in the schedule (thus larger errors on time estimates provided by timetabled routing).

Overall, strong evidence that timetabling fails to deliver good solutions in this scenario

Is this just for Rome?

Is this just for Rome?

Is this just for Rome?

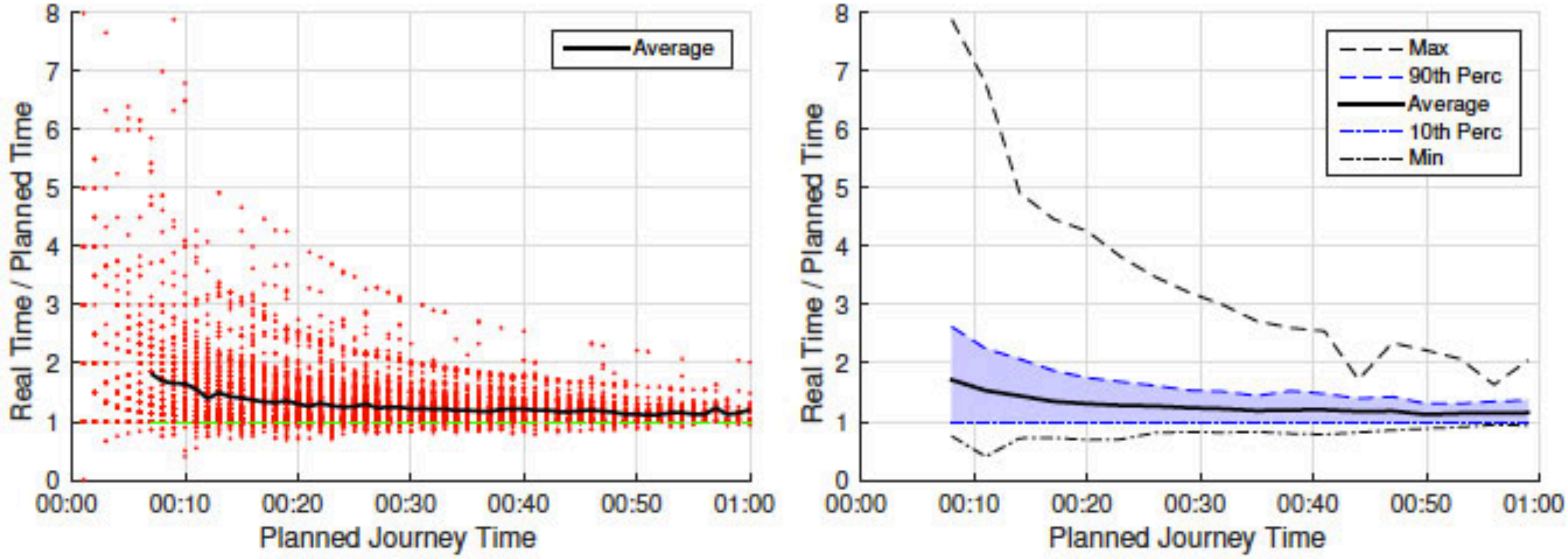
Is this just for Rome?

Verbatim from [Böhmová, Mihalák, Neubert, Pröger, Widmayer, ATMOS 2015]:

``Firmani et al. [9] observed in an experimental study on the transportation network of Rome that the timetable information and the real movement of the vehicles (based on GPS data) are only mildly correlated.”

Is this just related to Rome?

Verbatim from [Böhmová, Mihalák, Neubert, Pröger, Widmayer, ATMOS 2015]:



(a) (b)

``In overall, we observed that the behaviour in Zürich is comparable to the one in Rome."

Takehome message

In several cases, timetable information and real movement of vehicles are only mildly correlated: it's not only a small input perturbation (not only in Italy but also in Switzerland!)

Thus, widely used timetable routing algorithms suffer from many inaccuracies, independently of their own merits. Due to incorrect estimations of the waiting/ transfer times at transit stops, might fail to deliver a good solution.

Need smarter data: GPS feeds of transit vehicles?

Outline

1. Are timetabled route planners good enough for public transit?

2. ***Investigate impact of GPS data: run same algorithm with timetabled and GPS data***

3. Design of GPS-aware algorithms

[Delling, I., Pajor, Santaroni, IWCTS@SIGSPATIAL 2014]

Routing Algorithms

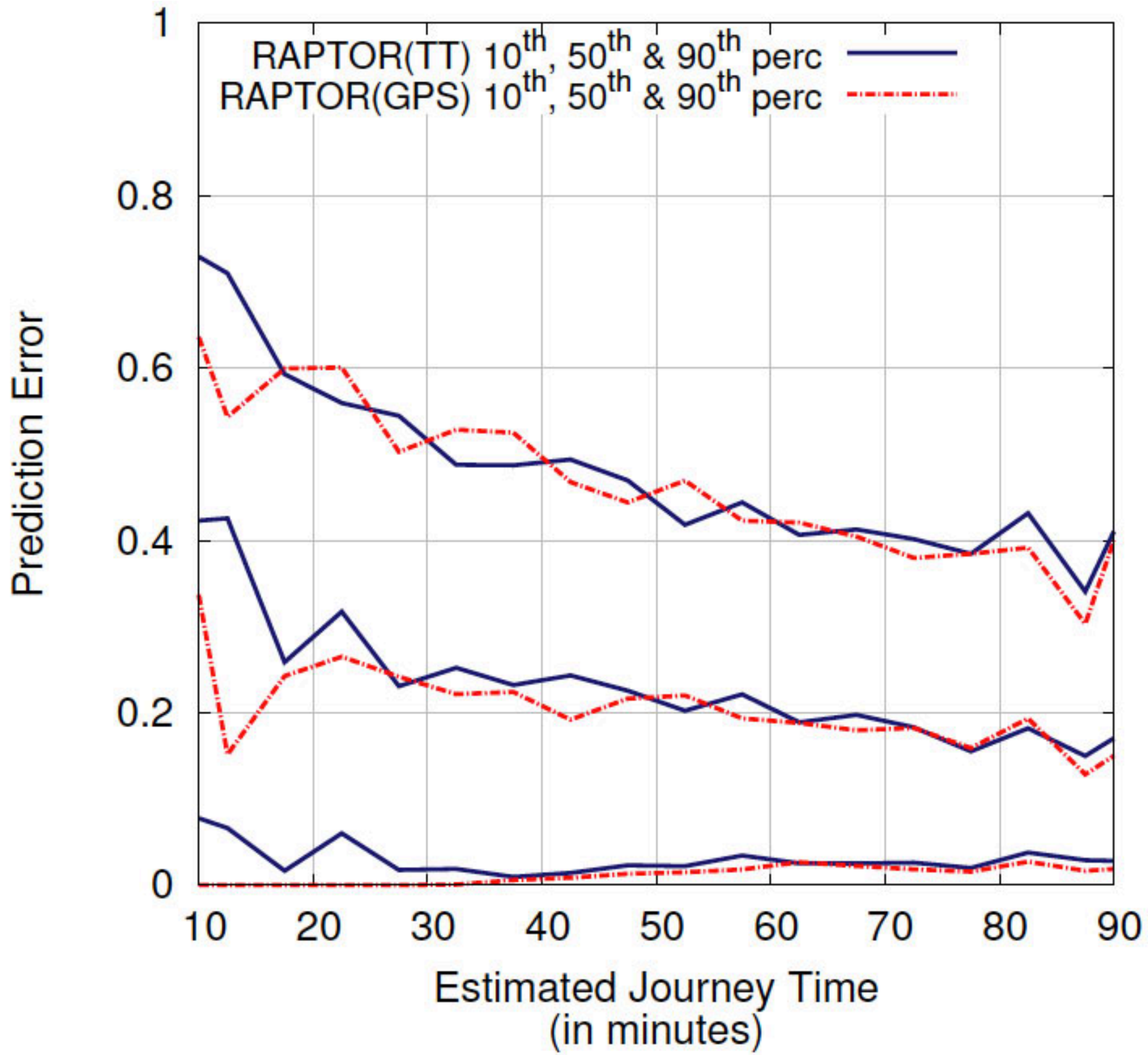
Few algorithms that work with timetabled data could be adapted to work with GPS data

- RAPTOR [Delling, Pajor, Werneck. Transportation Science, 2014].
- CSA [Dibbelt, Pajor, Strasser, Wagner. SEA 2013]
- …

RAPTOR

- Efficiently computes Pareto sets of journeys optimizing arrival time and number of transfers.
- It works in rounds, one per trip, computing in round i earliest arrival times for all stops that can be reached with i trips (i − 1 transfers).
- Our adaptation accepts GPS feeds and timetabled data

Relative Prediction Error



Quality of solutions?

Defining the “best journey” is a very complicated issue

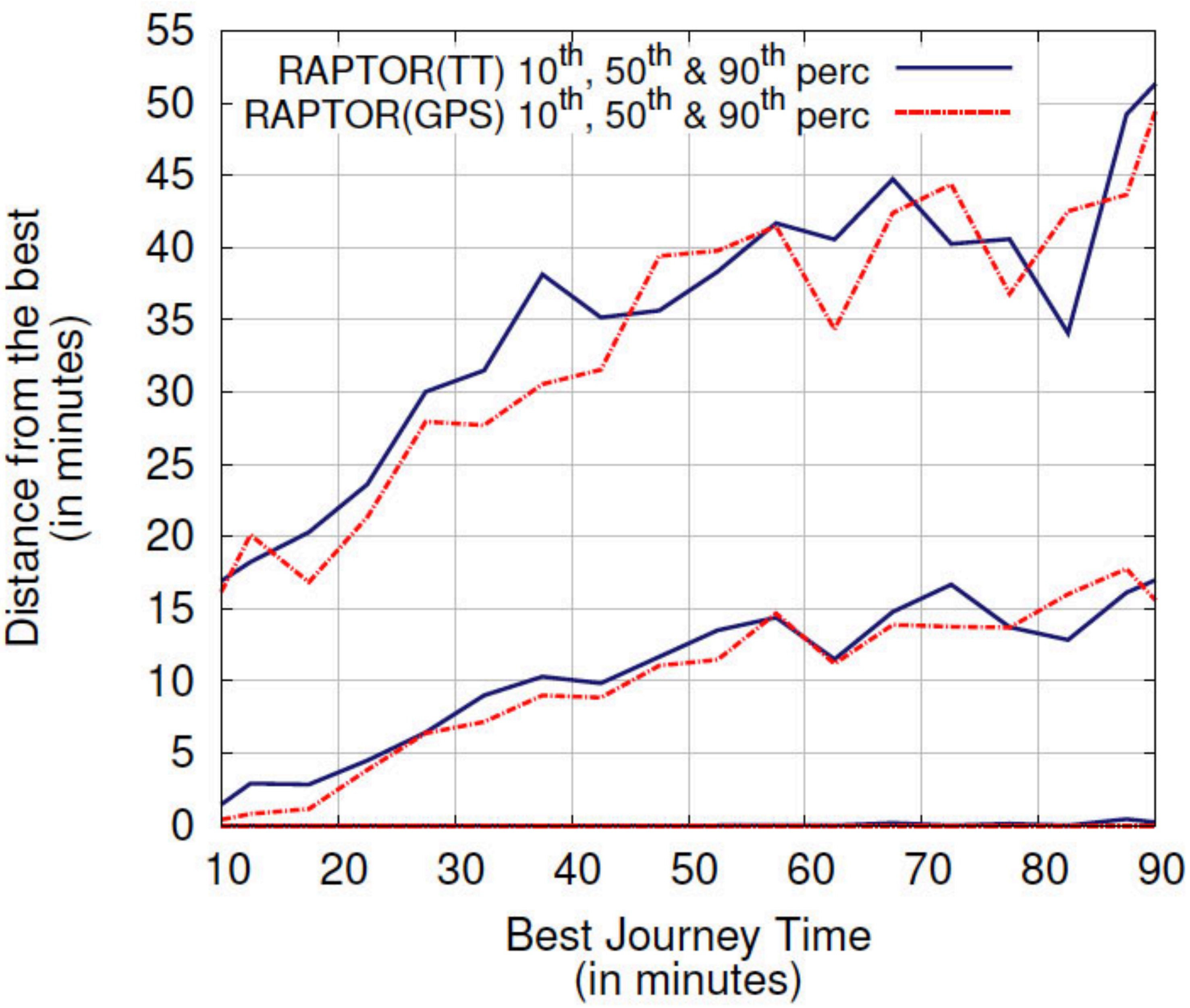
Journey planners try to optimize multiple cost criteria…

With all the data available we computed a rough proxy: fastest journey (earliest arrival time)

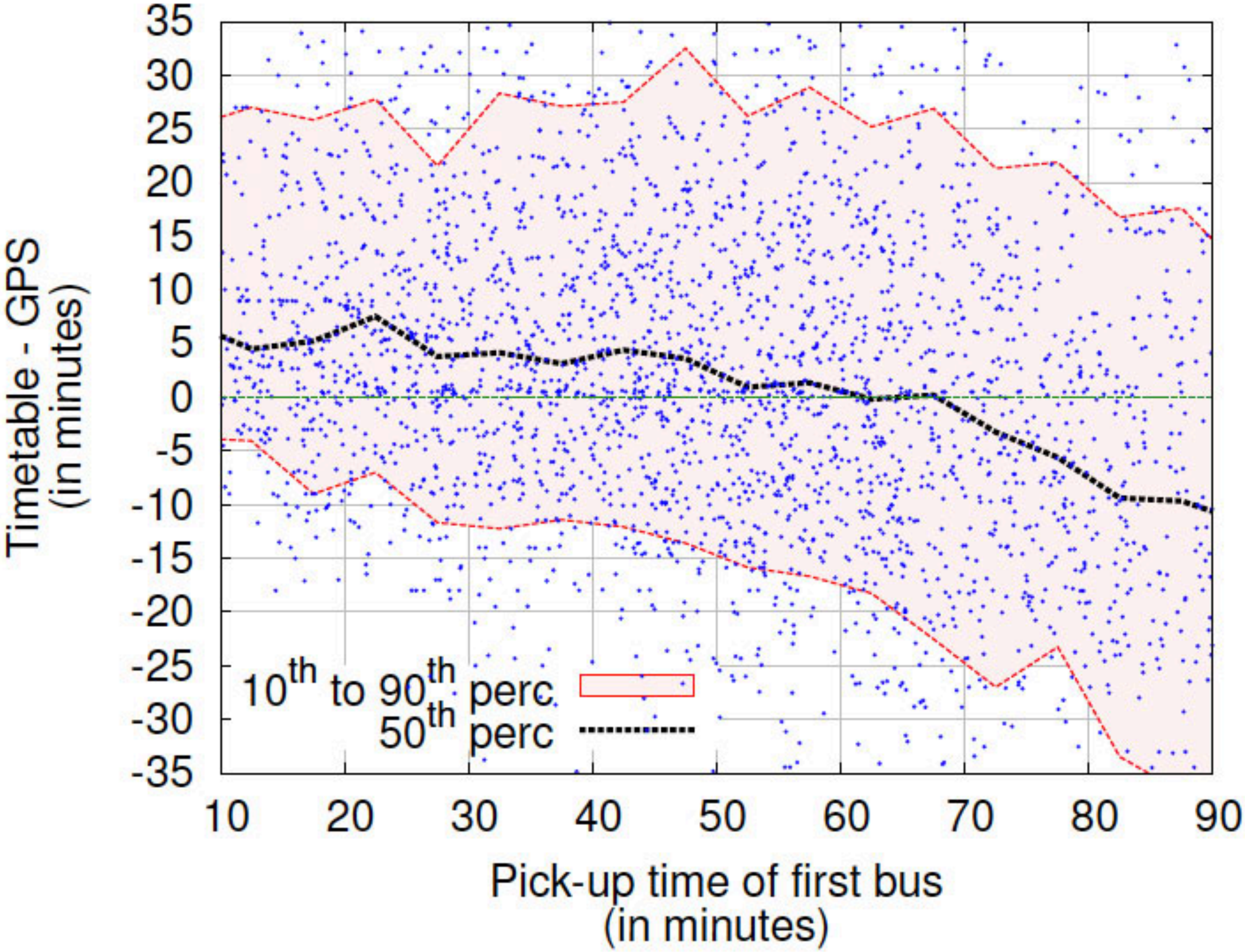
That’s our “best journey”

Measured “distance from the best” (slack)

Distance from the best (Slack)



GPS data useful in short term



Takehome message

Effective exploitation of GPS data seems important to make public transport journey planners smarter.

GPS data helps in short and medium range. Not very useful in the long range.

Main impact of GPS data depends on pick-up time of first bus

Smart data not enough: need smart (i.e., GPS-aware) algorithms

Outline

1. Are timetabled route planners good enough for public transit?

2. Investigate impact of GPS data: run same algorithm with timetabled and GPS data

3. ***Design of GPS-aware algorithms***

[Allulli, I., Santaroni, SEA 2014]

Muovi Roma

http://muovi.roma.it/

Journey planner available for the metropolitan area of Rome

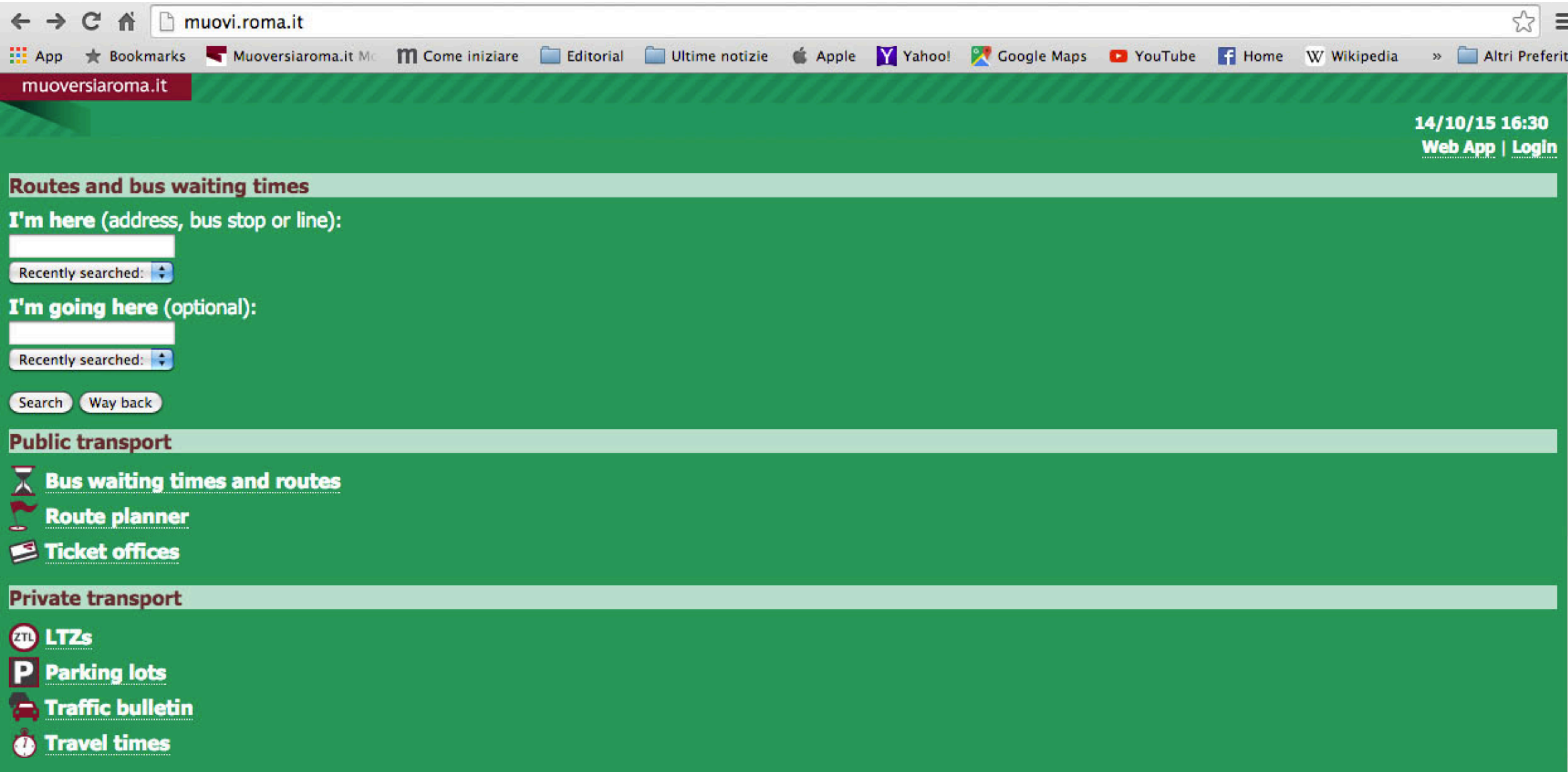
Uses GPS data and historical data to improve the accuracy of time estimates and to predict position of buses in the near future

Based on dynamic shortest paths on a time-dependent network

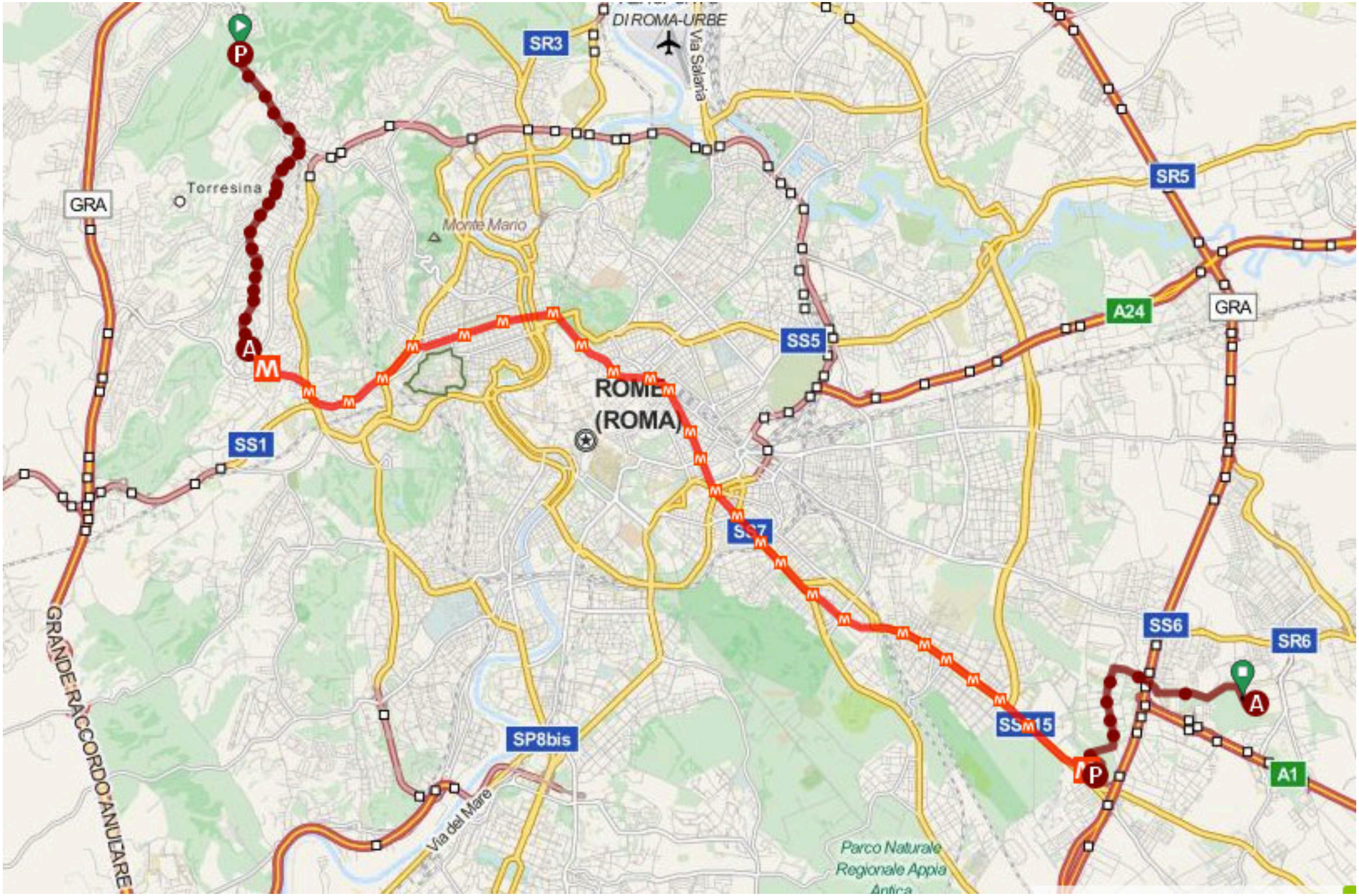
Incorporates some of our own algorithmic ideas

Developed and maintained by the Mobility Agency of Rome (open source / open data)

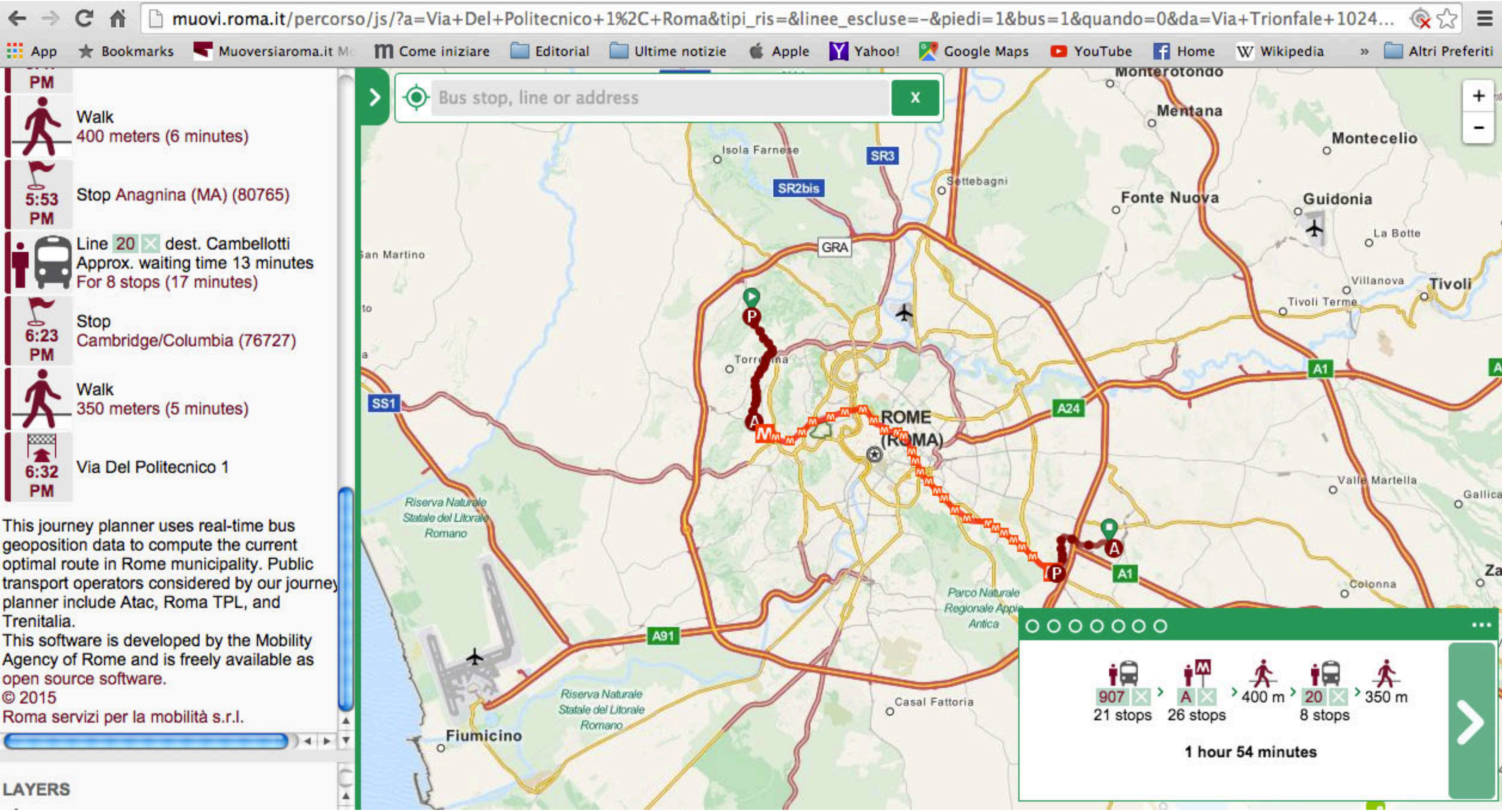
Muovi Roma


muovi.roma.it
App
Bookmarks
Muoversiaroma.it
Come iniziare
Editorial
Ultime notizie
Apple
Yahoo!
Google Maps
YouTube
Home
Wikipedia
Altri Preferit
muoversiaroma.it
14/10/15 16:30
Web App | Login
Routes and bus waiting times
I'm here (address, bus stop or line):
Recently searched:
I'm going here (optional):
Recently searched:
Search
Way back
Public transport
Bus waiting times and routes
Route planner
Ticket offices
Private transport
LTZs
Parking lots
Traffic bulletin
Travel times

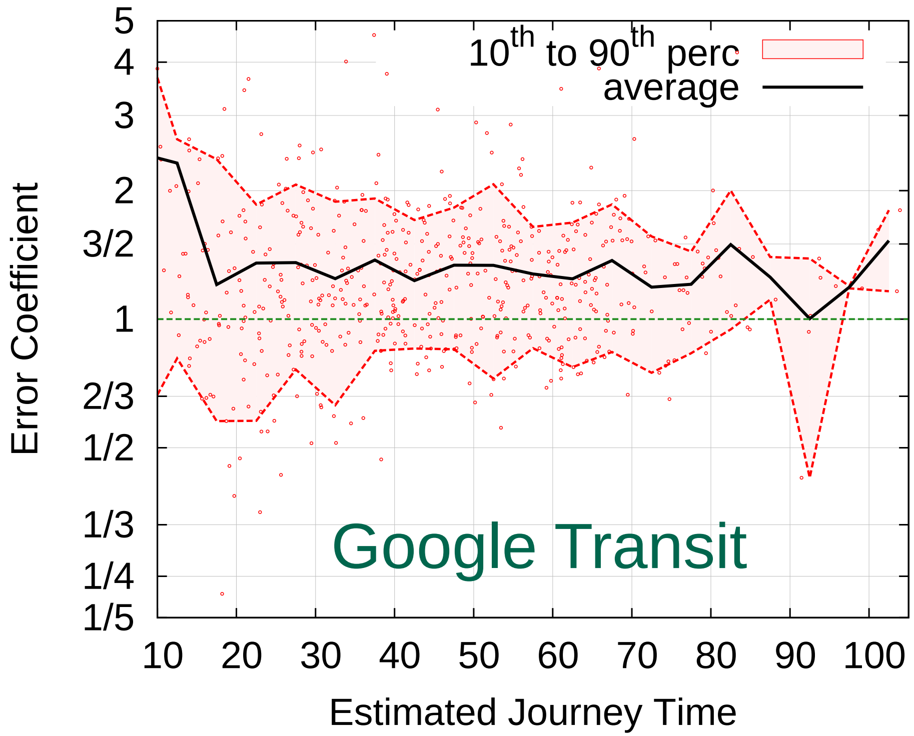
Muovi Roma

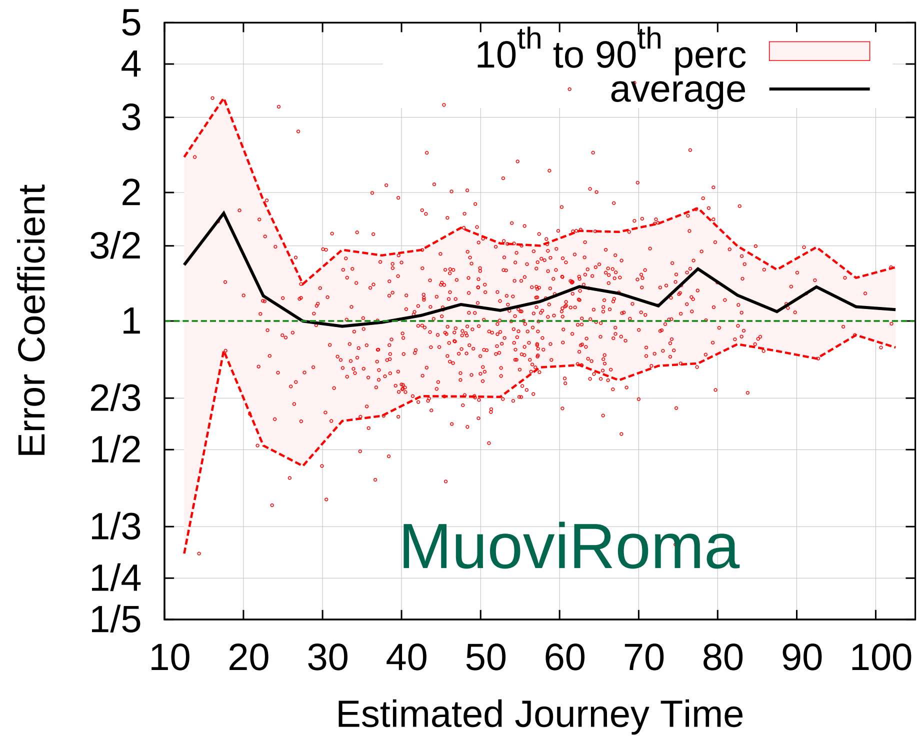

DI ROMA-URBE
Via Salaria
SR3
GRA
Torresina
Monte Mario
SS5
SR5
A24
ROME
(ROMA)
SS1
SS7
SS6
SR6
SS515
SP8bis
A1
Via del Mare
GRANDE RACCORDO ANULARE
Parco Naturale
Regionale Appia
Antica

Muovi Roma


muovi.roma.it/percorso/js/?a=Via+Del+Politecnico+1%2C+Roma&tipi_ris=&linee_escluse=-&piedi=1&bus=1&quando=0&da=Via+Trionfale+1024...
App
Bookmarks
Muoversiaroma.it
Come iniziare
Editorial
Ultime notizie
Apple
Yahoo!
Google Maps
YouTube
Home
Wikipedia
Altri Preferiti
PM
Walk
400 meters (6 minutes)
5:53 PM
Stop Anagnina (MA) (80765)
Line 20 dest. Cambellotti
Approx. waiting time 13 minutes
For 8 stops (17 minutes)
6:23 PM
Stop
Cambridge/Columbia (76727)
Walk
350 meters (5 minutes)
6:32 PM
Via Del Politecnico 1
This journey planner uses real-time bus geoposition data to compute the current optimal route in Rome municipality. Public transport operators considered by our journey planner include Atac, Roma TPL, and Trenitalia.
This software is developed by the Mobility Agency of Rome and is freely available as open source software.
© 2015
Roma servizi per la mobilità s.r.l.
LAYERS
Bus stop, line or address
X
ROME (ROMA)
Fiumicino
Tivoli
Guidonia
Mentana
Monterotondo
Montecelio
Fonte Nuova
907
21 stops
A
26 stops
400 m
20
8 stops
350 m
1 hour 54 minutes

Error Coefficient





Error Coefficient = Actual Time / Estimated Time

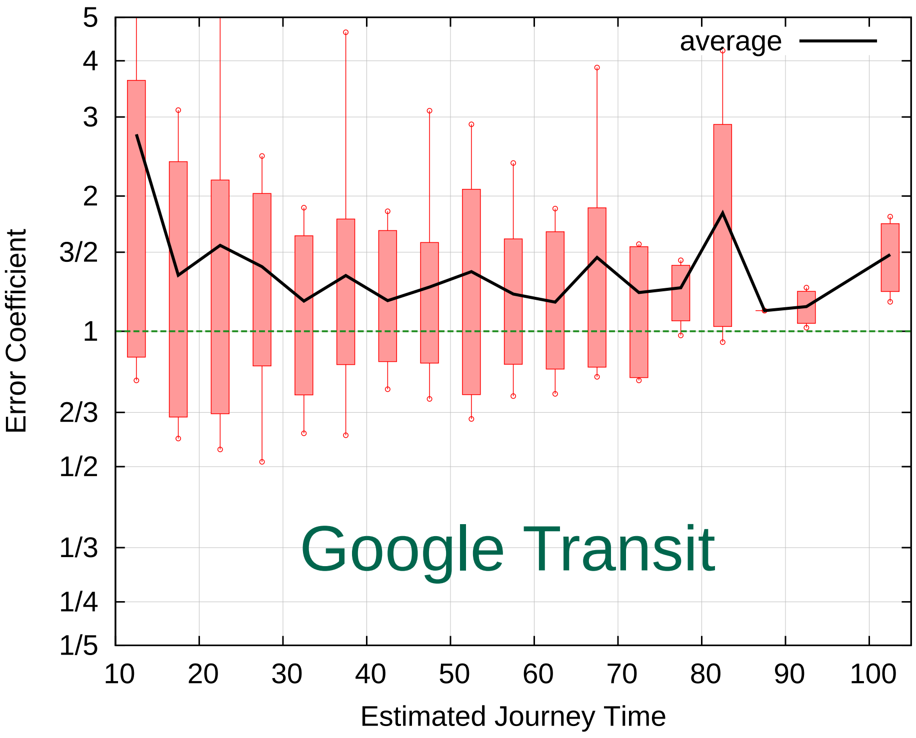
Short journeys affected more (MR estimates fewer short journeys)

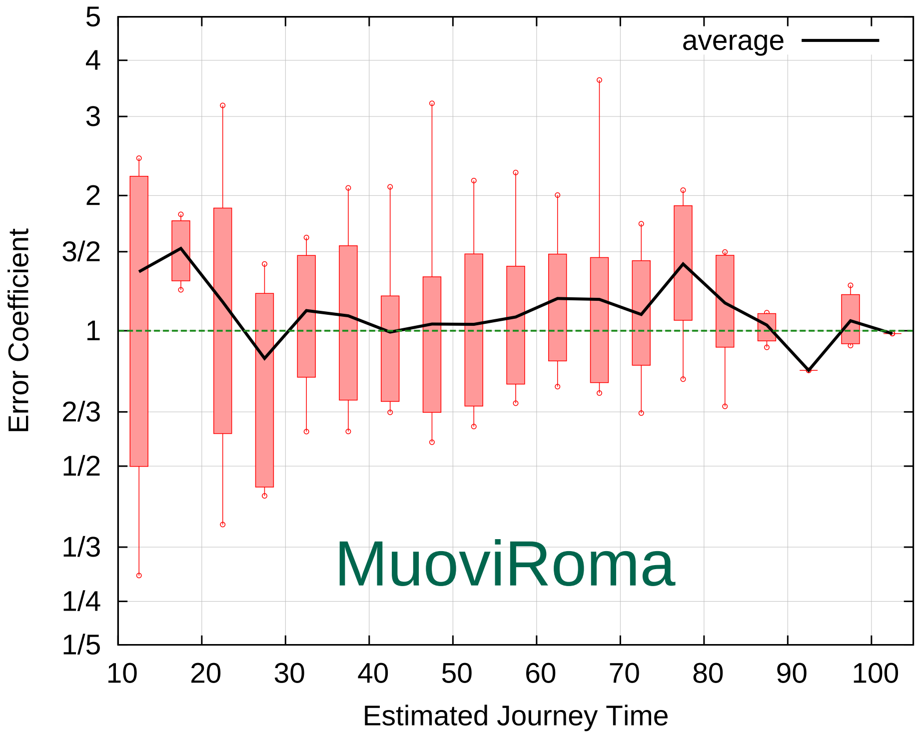
Both err on the optimistic side (actual journeys take longer on avg)

MR much better for medium range journeys

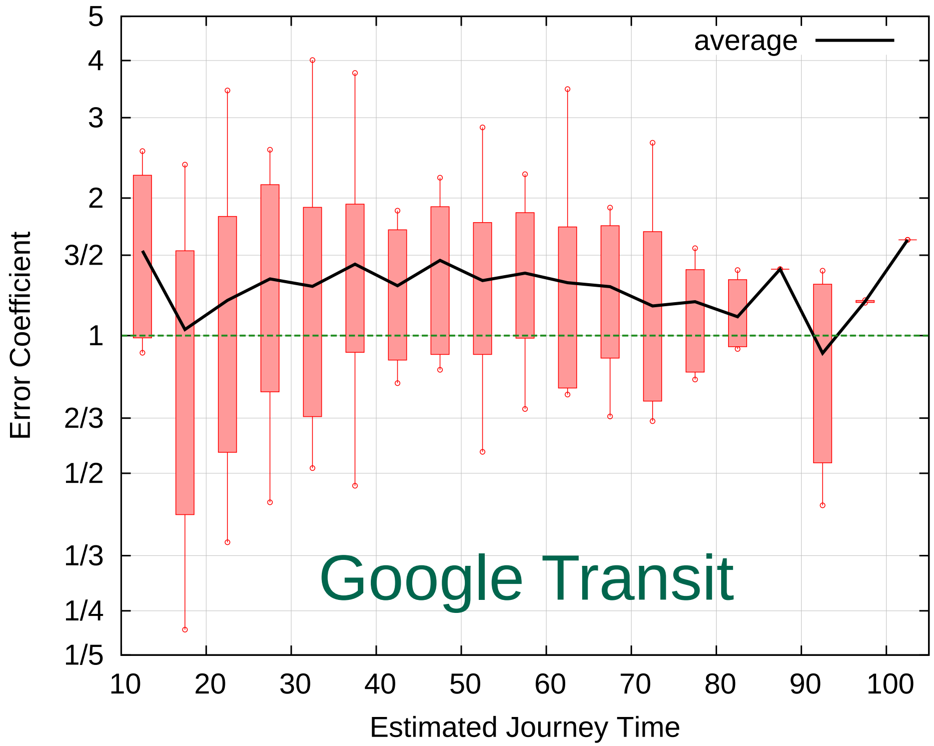
MR slightly better for long range journeys (GPS data not very useful in a hour from now?)

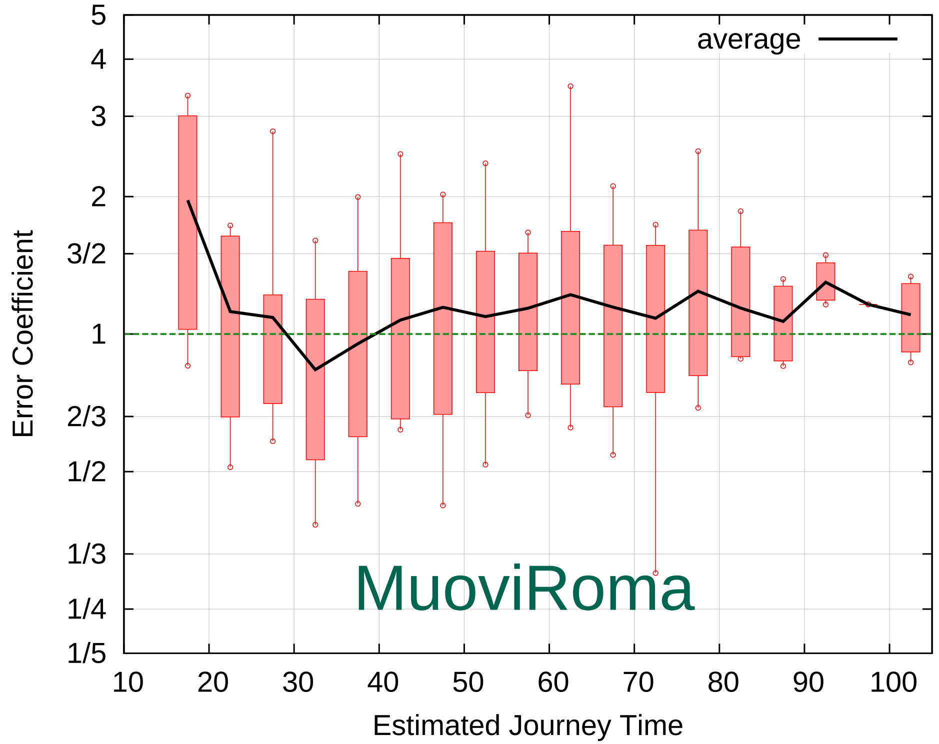
Morning (7:00am-1:30pm)





Afternoon (1:30pm-8:00pm)





Quality of solutions

From previous experiments can only conclude that GPS-aware planner (MR) estimates its journeys better than what timetabled planner (GT) does

Accuracy does not say much about quality of solutions provided (fast/best journeys)

MR could estimate better its journeys but could still provide low-quality solutions (slow journeys)

Quality of solutions

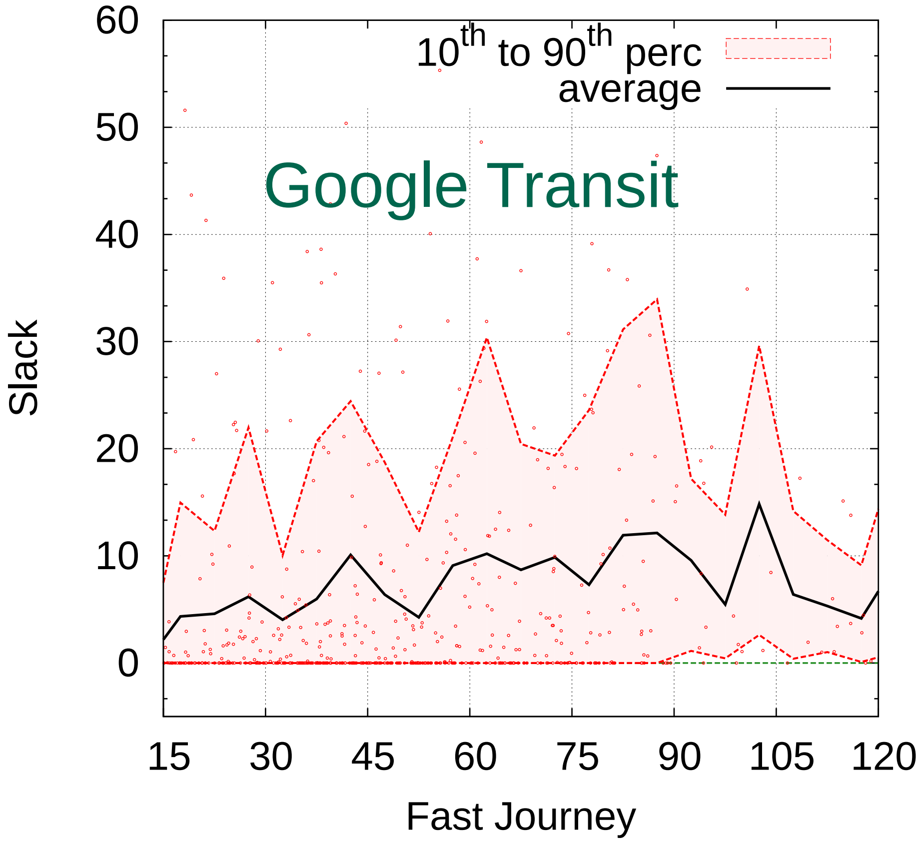
Again, computed *best journey* for query q

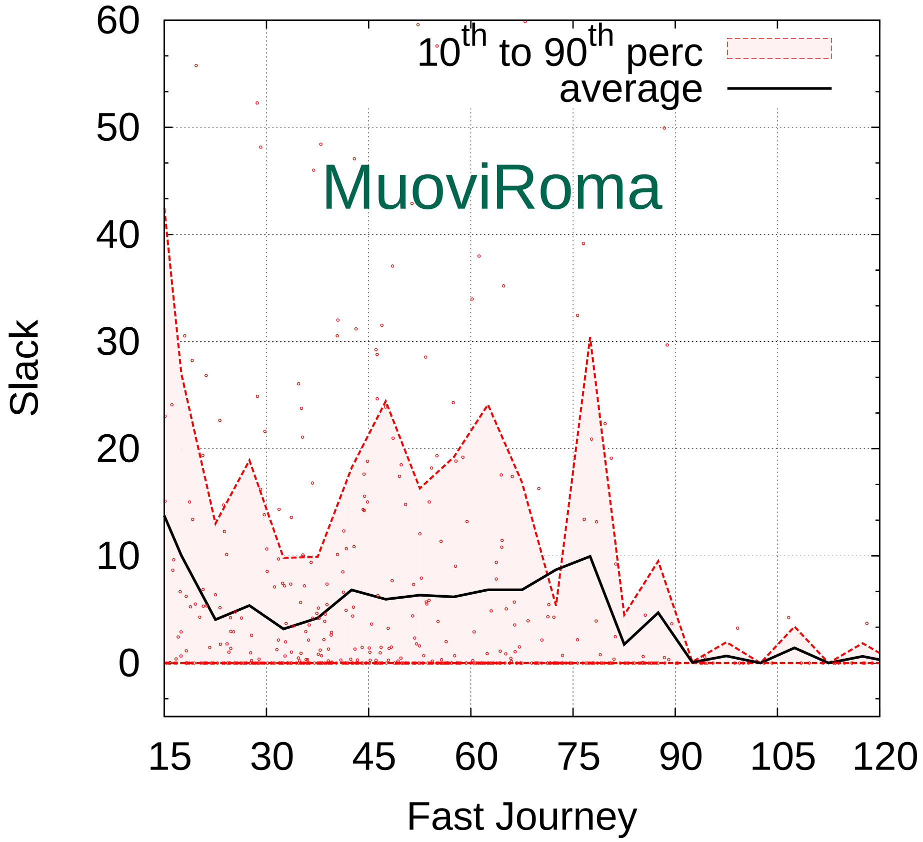
Distance from best (Slack) is distance (actual time) from best journey

Best journey has slack equal to 0

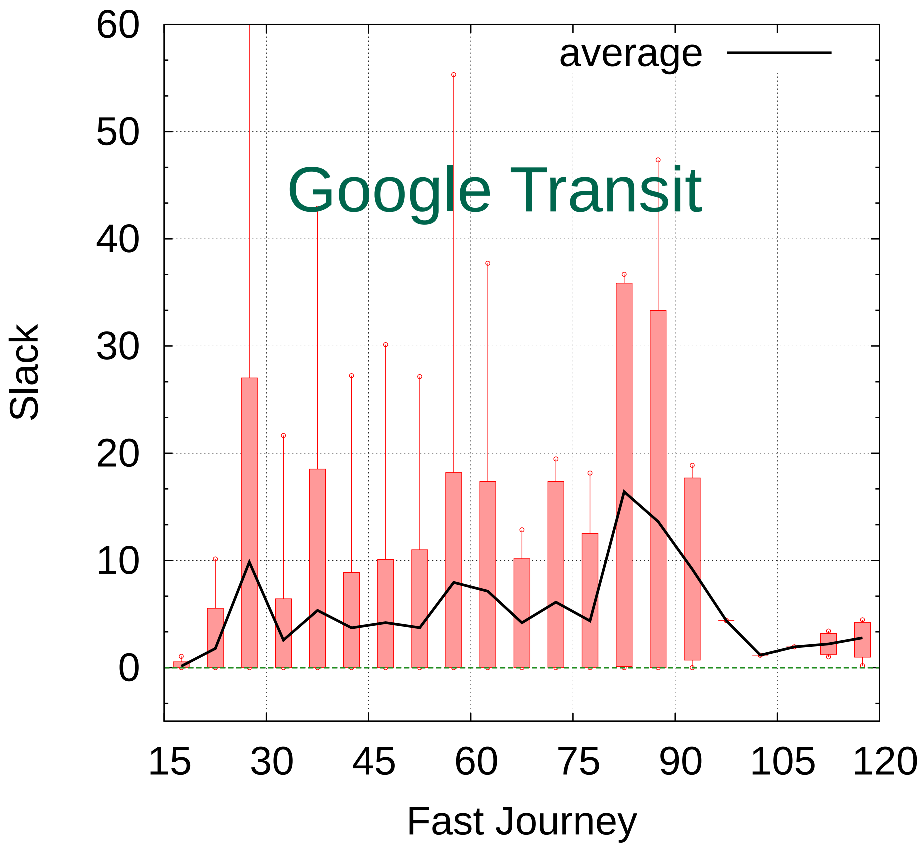
The smaller the slacks, the better the solutions provided by the route planner

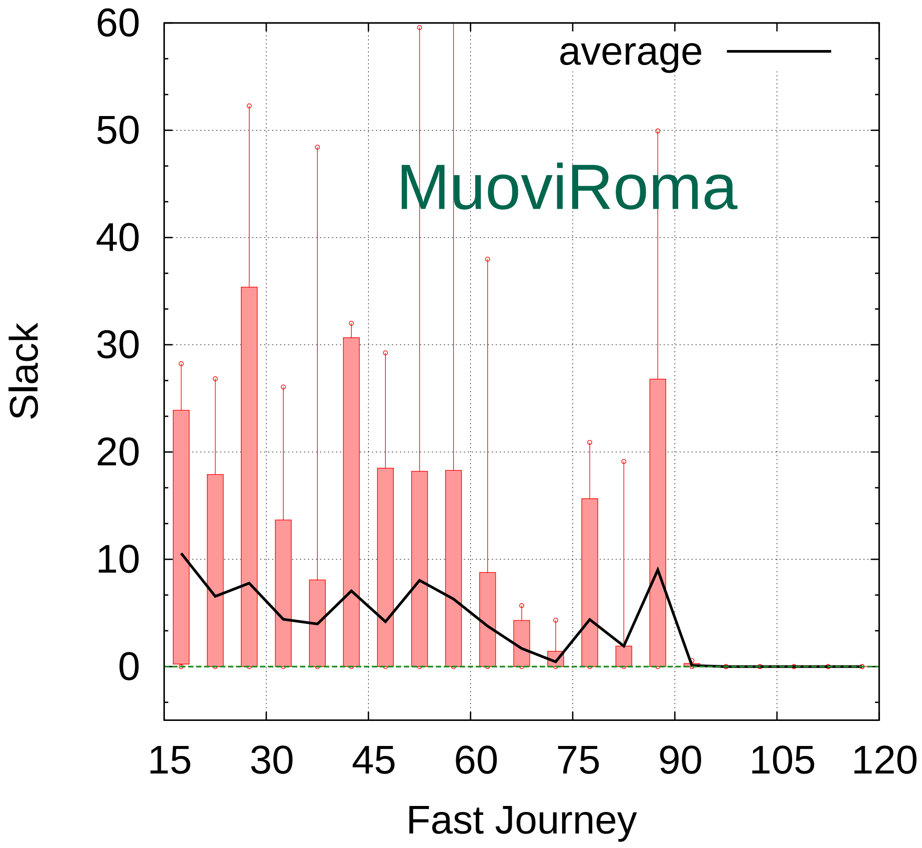
Quality of solutions (slack)



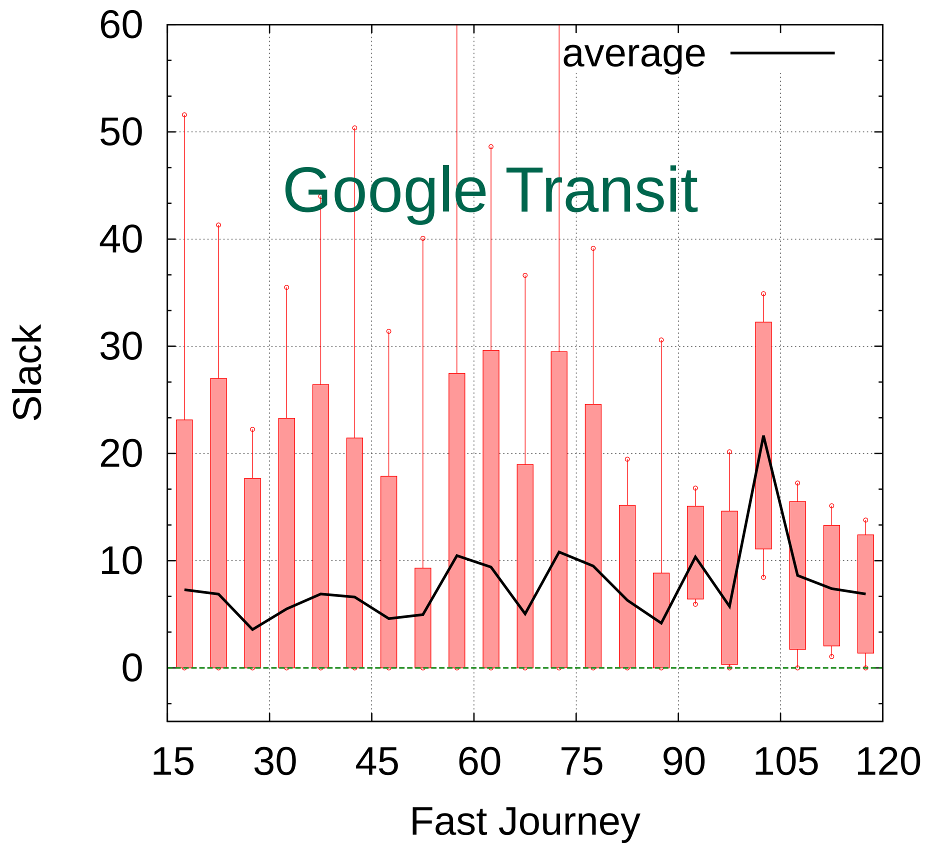


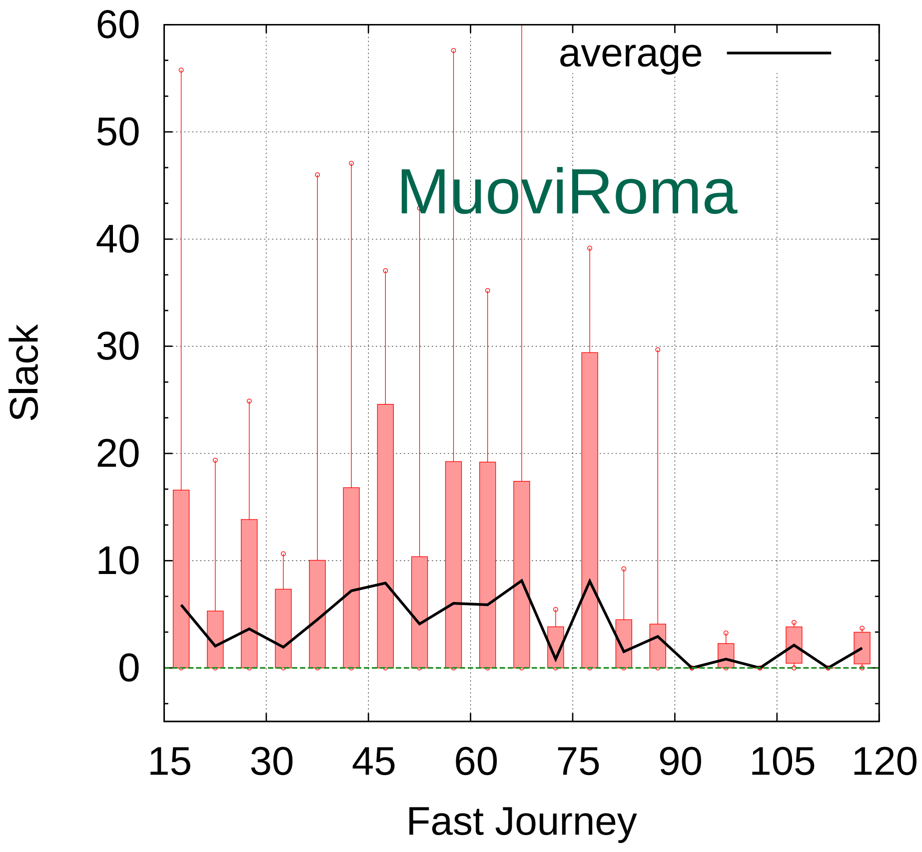
Slacks - Mornings





Slacks - Afternoons





Quality of solutions

GPS-aware (MR) able to achieve small slacks and fast journeys, especially for long-range journeys (>60’)

How could GPS data be still useful in a hour from now?

Prediction on historical data seems to works well

Long-range journeys likely to involve trips on (less frequent) suburban lines, where missing a connection will cause significant delays

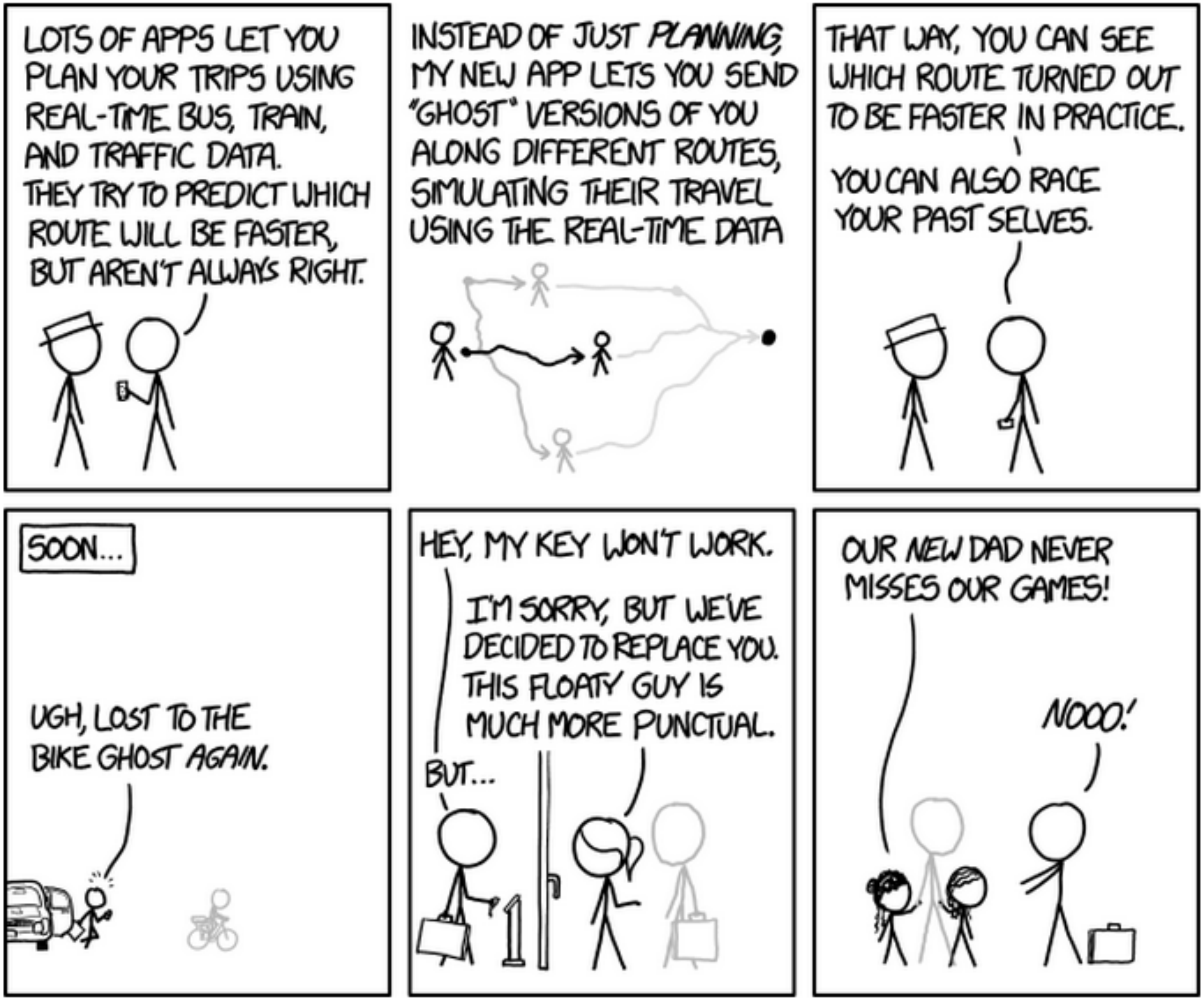
Long-range journeys based on unreliable timetables likely to incur in discrepancies with higher impact on overall travel time.

Conclusions

- Tried to assess empirically impact of GPS data in real public transport network (Rome)
- With significant fluctuations in schedule, timetabling inherently affected. GPS data able to provide better accuracy
- “Traditional” algorithms do not seem to be able to fully exploit GPS data
- We have smart data. Need to work more on smarter algorithms!
- Prediction models are critical issue (what will be the state of the network in 60 minutes?)

Future Work

Future Work



LOTS OF APPS LET YOU PLAN YOUR TRIPS USING REAL-TIME BUS, TRAIN, AND TRAFFIC DATA.
THEY TRY TO PREDICT WHICH ROUTE WILL BE FASTER, BUT AREN'T ALWAYS RIGHT.
INSTEAD OF JUST *PLANNING*, MY NEW APP LETS YOU SEND "GHOST" VERSIONS OF YOU ALONG DIFFERENT ROUTES, SIMULATING THEIR TRAVEL USING THE REAL-TIME DATA
THAT WAY, YOU CAN SEE WHICH ROUTE TURNED OUT TO BE FASTER IN PRACTICE.
YOU CAN ALSO RACE YOUR PAST SELVES.
SOON...
UGH, LOST TO THE BIKE GHOST *AGAIN*.
HEY, MY KEY WON'T WORK.
I'M SORRY, BUT WE'VE DECIDED TO REPLACE YOU. THIS FLOATY GUY IS MUCH MORE PUNCTUAL.
BUT...
OUR *NEW* DAD NEVER MISSES OUR GAMES!
NOOO!