Route choice with smartphones GPS data

Michel Bierlaire

transp-or.epfl.ch

Transport and Mobility Laboratory, EPFL

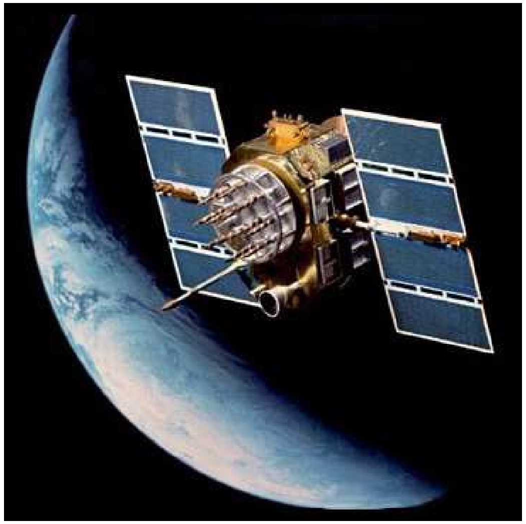
Motivation

- Travelers are equipped with smartphones
- Smartphones are equipped with sensors
- Can we learn mobility patterns from them?

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Objectives

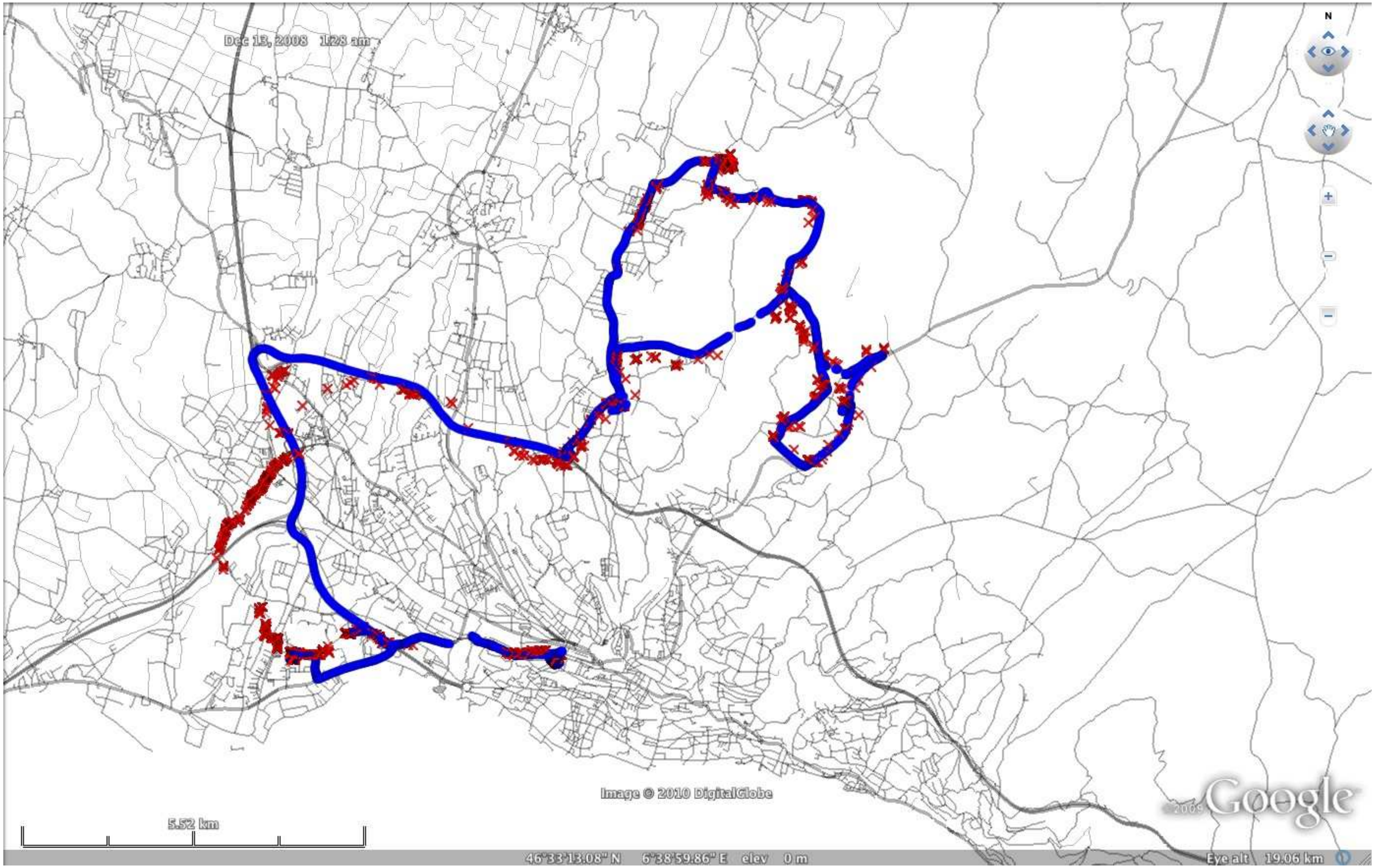


- Focus on GPS data from smartphone
- Reconstruct actual paths
- Model route choice behavior

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

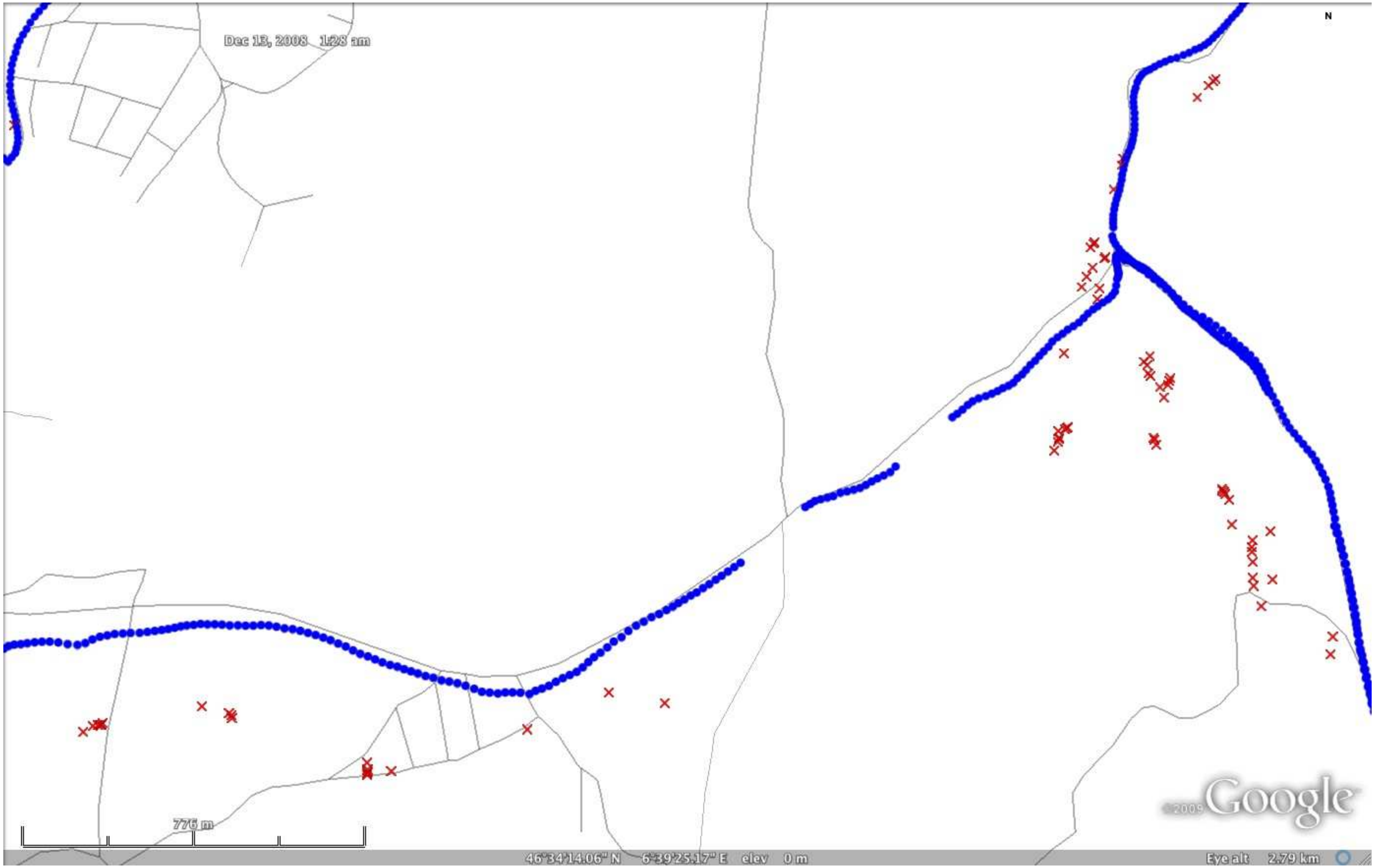
Issues



TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Issues



Dec 13, 2008 1:28 am
N
776 m
46°34'14.06" N 6°39'25.17" E elev 0 m
Eye alt 2.79 km
©2009 Google

Issues

- Low data collection rate to save battery
- Inaccuracy sue to technological constraints
- Smartphone carried in bags, pockets: weaker signal
- Map matching algorithms do not work with this data

Context

- Network: $G = (N, A)$
- Node coordinates: $x_n = \{\text{lat}, \text{lon}\}$
- Arc geometry:

$$\mathcal{L}_a : [0, 1] \to \mathbb{R}^2.$$

  Example: straight line

$$\mathcal{L}_a(\ell) = (1 - \ell)\, x_u + \ell x_d.$$

- Model for the movement of the mobile phone:

$$x = S(x^-, t^-, t, p)$$

  - Ideally a traffic simulator
  - Simpler models are used in practice
  - Random variable with density $f_x(x|x^-, t^-, t, p)$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Data

One measurement: $\widehat{g} = \left(\widehat{t}, \widehat{x}, \widehat{\sigma}^x, \widehat{v}, \widehat{\sigma}^v, \widehat{h}\right)$,

- $\widehat{t}$, a time stamp ;
- $\widehat{x} = (\widehat{x}_{\text{lat}}, \widehat{x}_{\text{lon}})$, a pair of coordinates;
- $\widehat{\sigma}^x$, the standard deviation of the horizontal error in the location measurement;
- $\widehat{v}$, a speed measurement (km/h) and,
- $\widehat{\sigma}^v$, the standard deviation of the error in that measurement;
- $\widehat{h}$, a heading measurement, that is the angle to the north direction, from 0 to 359, clockwise.

Sequence: $(\widehat{g}_1, \ldots, \widehat{g}_T)$

Measurement equations

Objective:

- Given a path p
- Given a sequence $(\widehat{g}_1, \ldots, \widehat{g}_T)$
- What is the likelihood that the sequence has been generated by a smartphone moving along path p?
- Note: different approach from map matching, which is essentially a projection procedure.
- We focus on the position only
- We derive

$$\Pr(\widehat{x}_1, \ldots, \widehat{x}_T | p),$$

- ... recursively

$$\Pr(\widehat{x}_1, \ldots, \widehat{x}_T | p) = \Pr(\widehat{x}_T | \widehat{x}_1, \ldots, \widehat{x}_{T-1}, p) \Pr(\widehat{x}_1, \ldots, \widehat{x}_{T-1} | p).$$

Recursion: first step

$$\Pr(\widehat{x}_1|p) = \int_{x_1 \in p} \Pr(\widehat{x}_1|x_1, p) \Pr(x_1|p) dx_1,$$

- integral spans all locations x_1 on path p
- no prior information on x_1

$$\Pr(x_1|p) = 1/L_p$$

- a smarter way would be to assign more probability in the begining of the path
- measurement error of the device:

$$\Pr(\widehat{x}_1|x_1, p) = \Pr(\widehat{x}_1|x_1)$$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Measurement error of the device

- Assume that latitudinal and longitudinal errors are i.i.d. normal with variance σ^2
- Measurement error is Rayleigh
- σ^2 unknown, estimate:

$$\widehat{\sigma}^2 = \sigma^2_{\text{network}} + (\widehat{\sigma}_1^x)^2$$

where

  - $\sigma^2_{\text{network}}$: network coding errors
  - $(\widehat{\sigma}_1^x)^2$: GPS errors.

$$\Pr(\widehat{x}_1 | x_1) = \exp\left(-\frac{\|\widehat{x}_1 - x_1\|_2^2}{2\widehat{\sigma}^2}\right).$$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Recursion: first step

$$\Pr(\widehat{x}_1|p) = \frac{1}{L_p} \int_{x_1} \exp\left(-\frac{\|\widehat{x}_1 - x_1\|_2^2}{2\widehat{\sigma}^2}\right) dx_1.$$

- Integral may be cumbersome for long paths
- Can be simplified using the concept of Domain of Data Relevance
- See Bierlaire & Frejinger (2008) and Bierlaire, Chen and Newman (2010)

Recursion: second step

$$\Pr(\widehat{x}_1, \widehat{x}_2 | p) = \Pr(\widehat{x}_2 | \widehat{x}_1, p) \Pr(\widehat{x}_1 | p),$$

Focus now on

$$\Pr(\widehat{x}_2 | \widehat{x}_1, p) = \int_{x_2 \in p} \Pr(\widehat{x}_2 | x_2, \widehat{x}_1, p) \Pr(x_2 | \widehat{x}_1, p) dx_2.$$

- first term = $\Pr(\widehat{x}_2 | x_2)$ measurement error, same as before
- second term: predicts the position at time $\widehat{t}_2$ of the traveler

$$\Pr(x_2 | \widehat{x}_1, p) = \int_{x_1 \in p} \Pr(x_2 | x_1, \widehat{x}_1, p) \Pr(x_1 | \widehat{x}_1, p) dx_1.$$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Position predictor

$$\Pr(x_2|\widehat{x}_1,p) = \int_{x_1 \in p} \Pr(x_2|x_1,\widehat{x}_1,p) \Pr(x_1|\widehat{x}_1,p) dx_1.$$

- First term: movement model

$$\Pr(x_2|x_1,\widehat{x}_1,p) = f_x(x_2|x_1,\widehat{t}_1,\widehat{t}_2,p),$$

- Second term: Bayes rule

$$\Pr(x_1|\widehat{x}_1,p) = \frac{\Pr(\widehat{x}_1|x_1,p)\Pr(x_1|p)}{\int_{x_1} \Pr(\widehat{x}_1|x_1,p)\Pr(x_1|p)dx_1}.$$

simplifies to

$$\Pr(x_1|\widehat{x}_1,p) = \frac{\Pr(\widehat{x}_1|x_1,p)}{\int_{x_1} \Pr(\widehat{x}_1|x_1,p)dx_1}$$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Measurement equations

- Step k of the recursion based on same principles
- but requires some technical simplifications

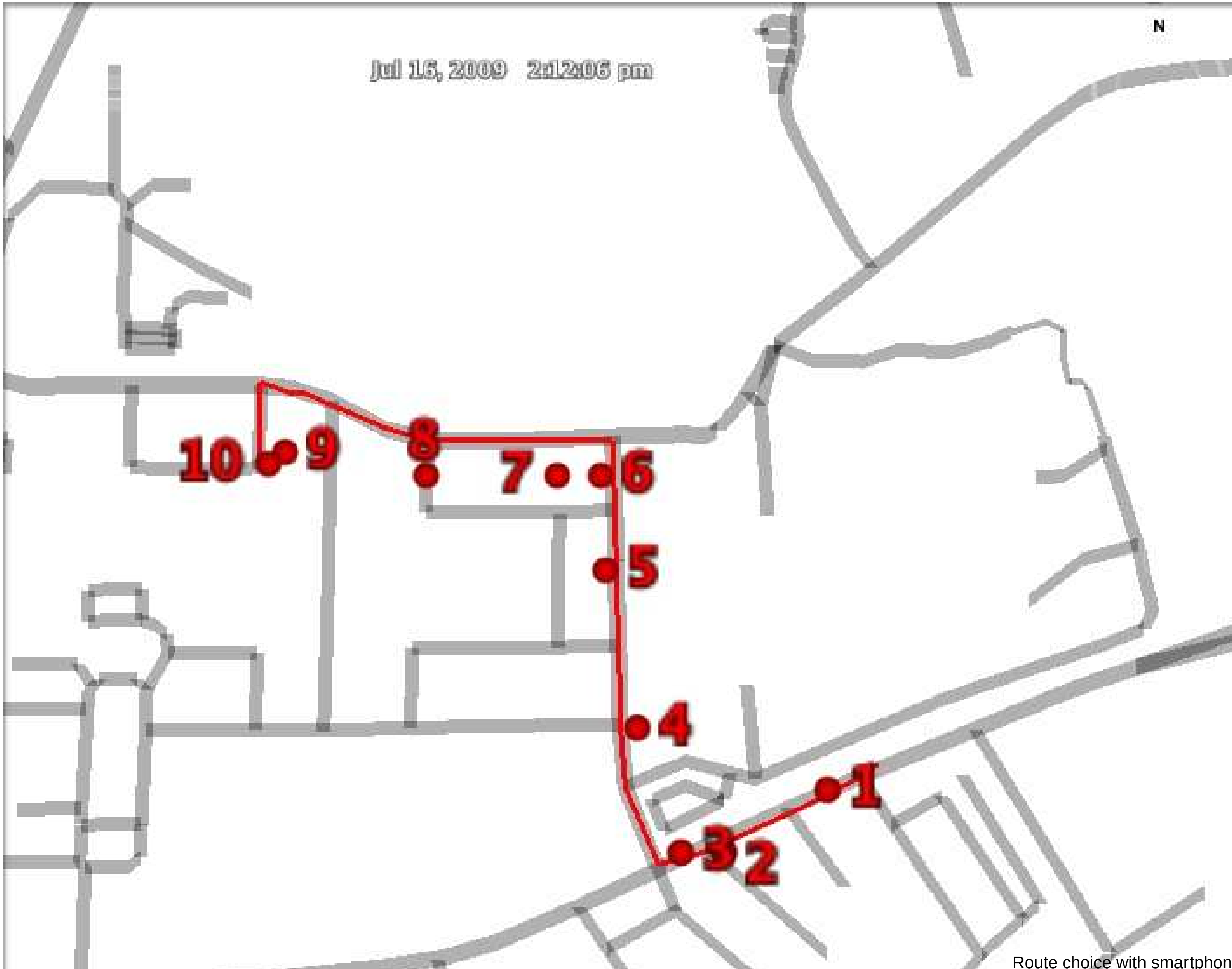
$$\Pr(x_{k-1}|\widehat{x}_{k-1},p) = \frac{\Pr(\widehat{x}_{k-1}|x_{k-1},p)}{\int_x \Pr(\widehat{x}_{k-1}|x,p)dx}.$$

- Integrals can be simplified using the DDR

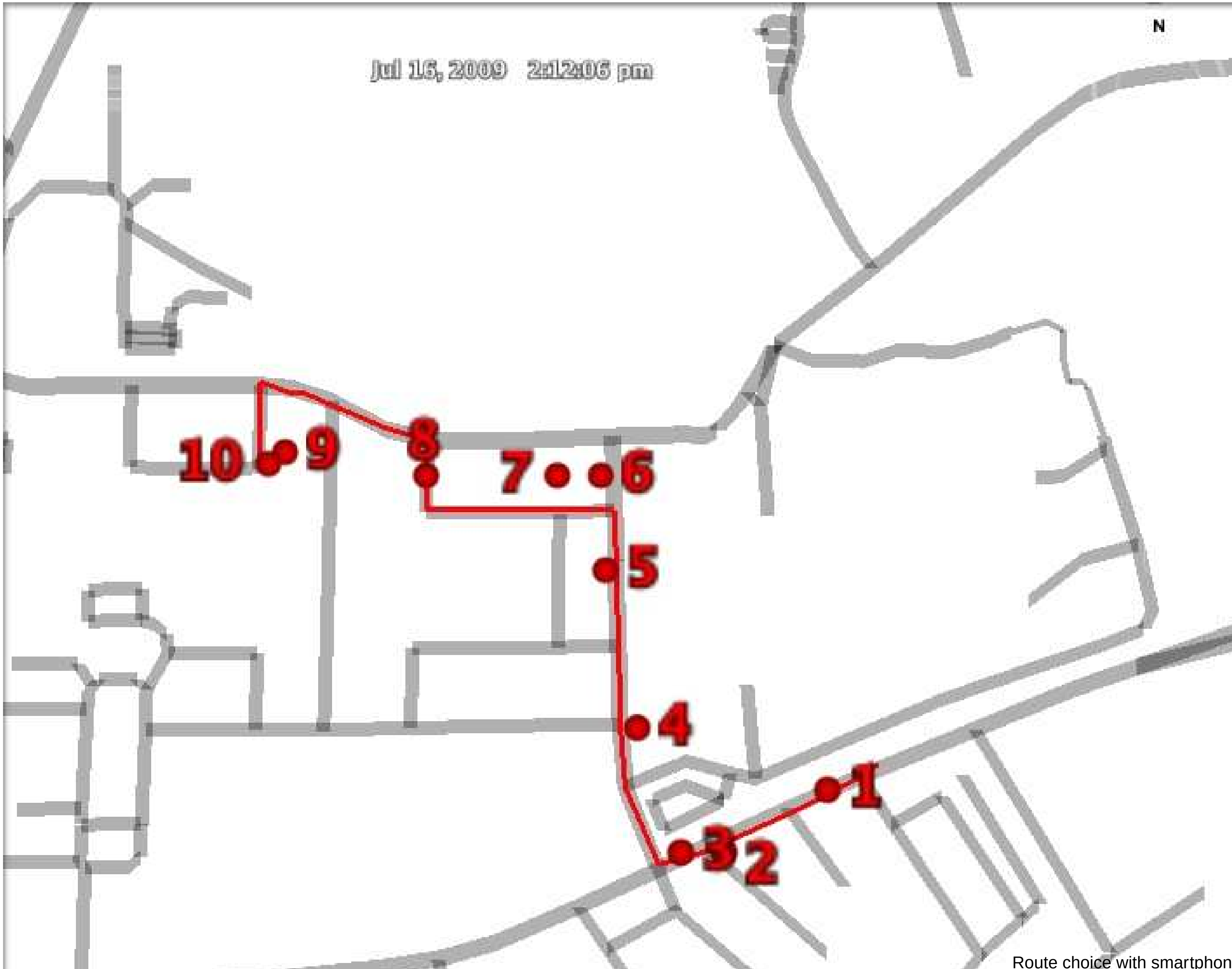
TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

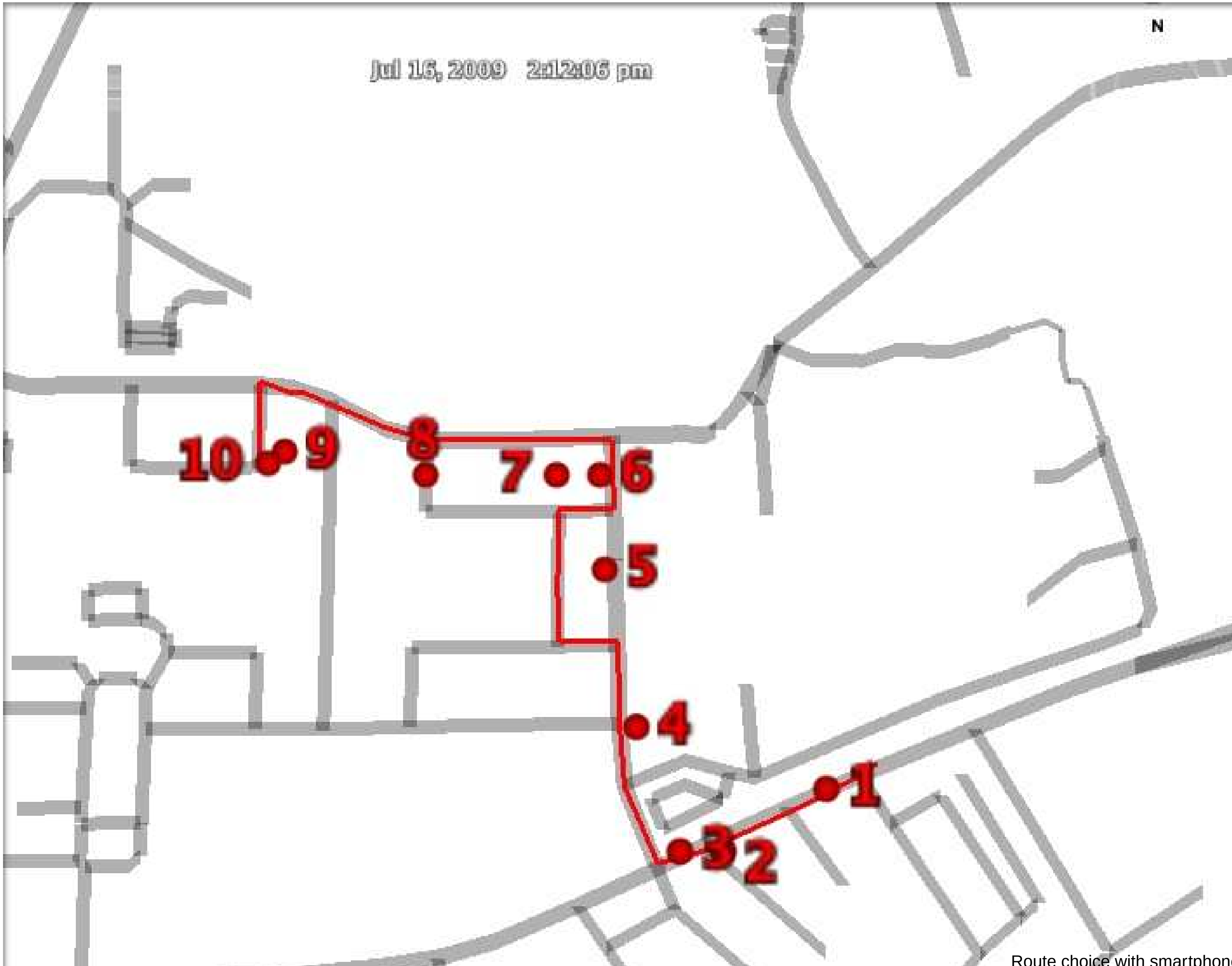
Case study: true path



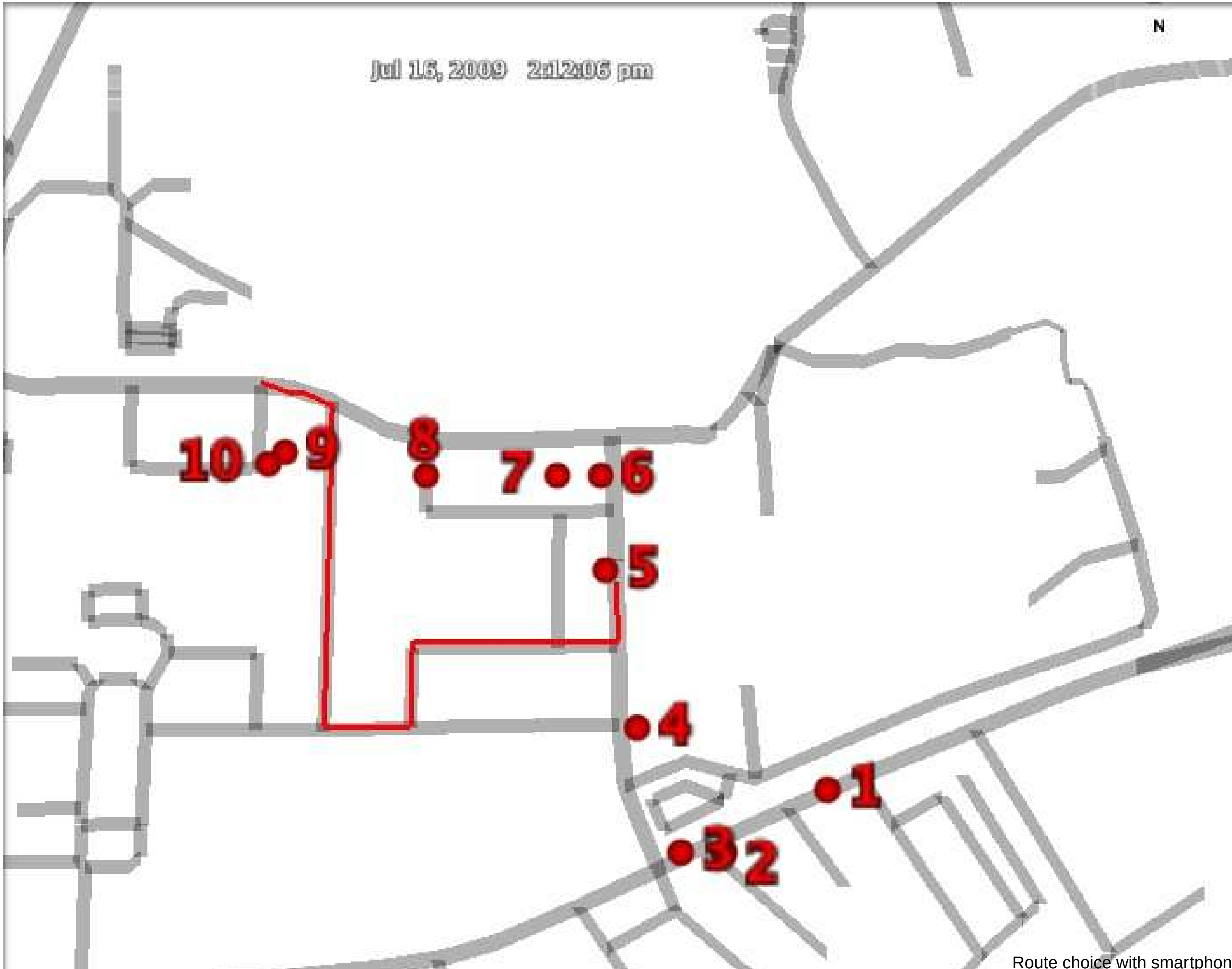
Case study: path with a deviation (1)



Case study: path with a deviation (2)



Case study: path from map matching algo



Case study: log likelihood from measurement equati

True path	-11.3
Deviation 1	-12.9
Deviation 2	-13.2
Map matching	$-\infty$

- DDR simplifications assigns 0 probability on irrealistic paths
- Results are consistent with intuition

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Stochastic map matching algorithm

- Generate a set of paths candidates
- Compute for each of them the likelihood of the GPS data
- Select a subset based on this likelihood

Summary

- We have designed a procedure that account for
  - the error of the GPS device
  - the error of the network coding
  - the movement of the smartphone
- It can involve complex models
- Technical simplifications are possible to make it operational on real data
- We have also designed
  - a path generation algorithm (Bierlaire et al. 2010)
  - a procedure for the estimation of route choice models (Bierlaire & FRejinger, 2008)

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

References

- Bierlaire, M., and Frejinger, E. (2008). Route choice modeling with network-free data, Transportation Research Part C: Emerging Technologies 16(2):187-198.

  http://dx.doi.org/10.1016/j.trc.2007.07.007

- Bierlaire, M., Chen, J., and Newman, J. P (2010). Modeling Route Choice Behavior From Smartphone GPS data. Technical report TRANSP-OR 101016. Transport and Mobility Laboratory, ENAC, EPFL.

  http://transp-or.epfl.ch/php/abstract.php?type=1&id=BierChenNewm10

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE