Modeling Route Choice Behavior From Smartphone GPS data

Michel Bierlaire

TRANSP-OR, EPFL

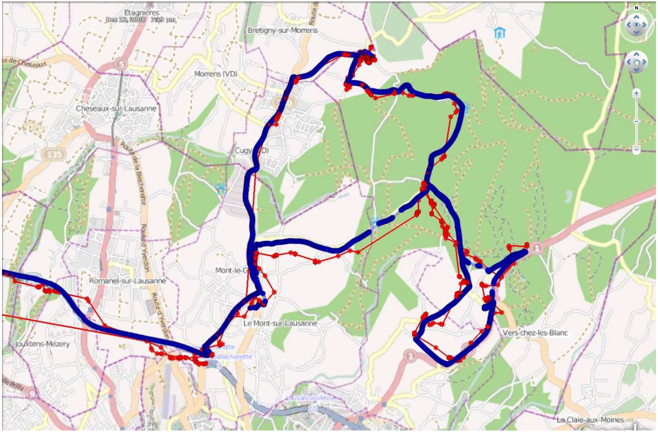
August 19, 2010

Introduction

- Objective: estimate route choice models from GPS data
- method: GPS data => path observations => route choice behavior
- challenge:
  - sparsity and inaccuracy of GPS data
  - ambiguity with respect to the real path

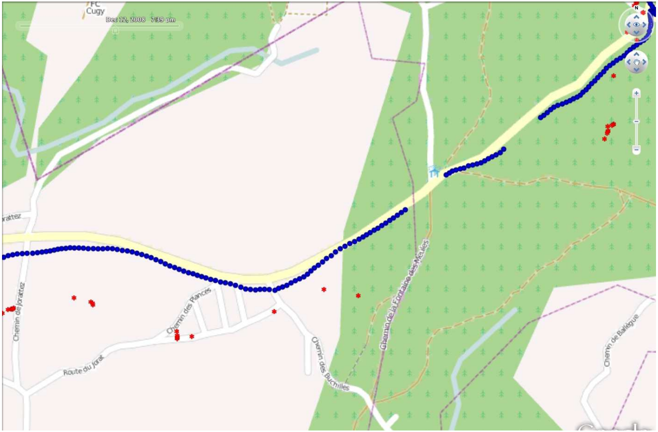
Sparsity and inaccuracy of the smartphone GPS data

Blue: GPS device Red: Nokia N95

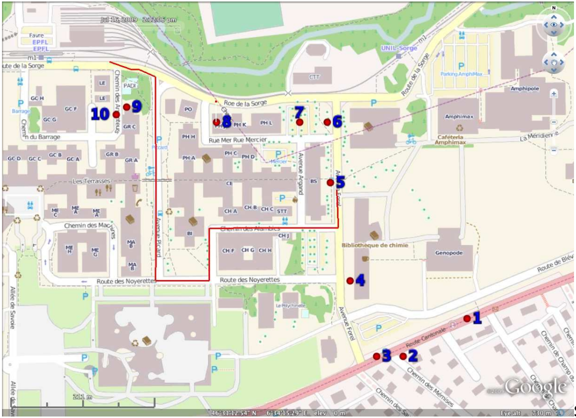


Sparsity and inaccuracy of the smartphone GPS data

Blue: GPS device Red: Nokia N95



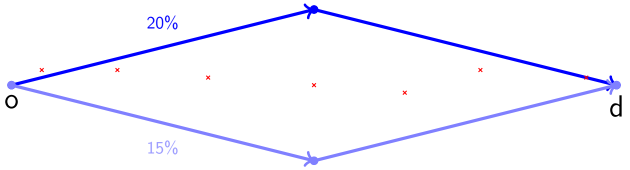
Deterministic map matching introduces biases



Schuessler and Axhausen (2009)

Network-free data

Bierlaire and Frejinger (2008)



Contribution to the likelihood:

$$\Pr(\widehat{x}_1, \ldots, \widehat{x}_T | S) = \sum_{s \in S} \Pr(s|S) \sum_{p \in C(s)} \Pr(\widehat{x}_1, \ldots, \widehat{x}_T | p) \Pr(p | C(s); \beta),$$

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

Probability that a vehicle traveling along p generates the GPS trace

Decomposition:

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

The recursion starts with:

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

where

- $\Pr(x_1 | p) = 1/L_p$: prior on the location
- $\Pr(\widehat{x}_1 | x_1, p) = \Pr(\widehat{x}_1 | x_1)$: measurement error of the device.

Measurement error of the device: $\Pr(\widehat{x}|x)$

- Assumption: the latitudinal and longitudinal errors are i.i.d. normal with variance σ^2.
- Therefore, the measurement error follows a Rayleigh distribution

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

- σ accounts also for errors in network coding

Iteration k

$$\Pr(\widehat{x}_k|\widehat{x}_1,\ldots,\widehat{x}_{k-1},p) =$$
$$\int_{x_k \in p} \Pr(\widehat{x}_k|x_k,\widehat{x}_1,\ldots,\widehat{x}_{k-1},p)\Pr(x_k|\widehat{x}_1,\ldots,\widehat{x}_{k-1},p)dx_k.$$

- $\Pr(\widehat{x}_k|x_k,\widehat{x}_1,\ldots,\widehat{x}_{k-1},p) = \Pr(\widehat{x}_k|x_k)$: measurement error.
- $\Pr(x_k|\widehat{x}_1,\ldots,\widehat{x}_{k-1},p) = \Pr(x_k|\widehat{x}_{k-1},p)$ predicts the position at time $\widehat{t}_k$ of the traveler.

Position of the traveler

$$\Pr(x_k | \widehat{x}_{k-1}, p)$$

$$= \int_{x_{k-1} \in p} \Pr(x_k | x_{k-1}, \widehat{x}_{k-1}, p) \Pr(x_{k-1} | \widehat{x}_{k-1}, p) dx_{k-1}.$$

$$= \int_{x_{k-1} \in p} \Pr(x_k | x_{k-1}, \widehat{t}_{k-1}, \widehat{t}_k, p) \frac{\Pr(\widehat{x}_{k-1} | x_{k-1}, p)}{\int_{x_{k-1} \in p} \Pr(\widehat{x}_{k-1} | x_{k-1}, p)} dx_{k-1}.$$

- $\Pr(\widehat{x}_{k-1} | x_{k-1}, p)$: measurement error
- $\Pr(x_k | x_{k-1}, \widehat{t}_{k-1}, \widehat{t}_k, p)$: movement model

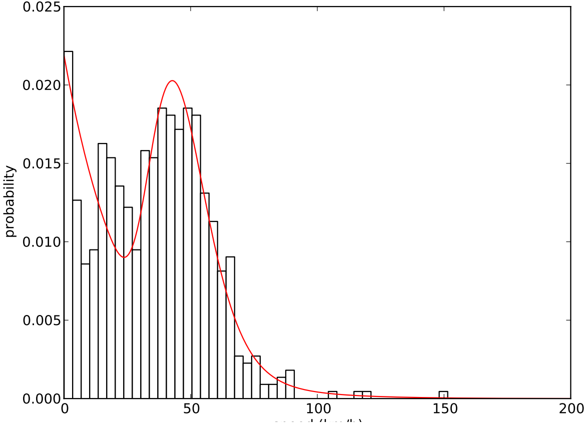
Movement model $\Pr(x|x^-, t^-, t, p)$

x is the position of the device at time t if the position at time t^- is x^-, and the device is traveling along path p.
Random variable with pdf:

$$f_x(x|x^-, t^-, t, p) = f_v\left(\frac{d_p(x^-, x)}{t - t^-}\right)$$

- v: traveling speed of the device.
- proposed distribution:
  $w *$ Negative Exponential $+ (1 - w) *$ Lognormal
- Mixture of two regimes: stop and go
- distribution of v estimated from observed GPS data.

Model fitting



Model estimates

$$f_v(v) = w\lambda \exp^{-\lambda v} + (1-w)\frac{1}{v\sqrt{\pi\sigma^2}} \exp^{-\frac{(\ln v - \mu)^2}{2\sigma^2}},$$

Speed records: 658.

parameter	estimate	standard error
w	0.528	0.0362
λ	0.041	0.0032
μ	3.843	0.0206
σ	0.250	0.0200
Parameters estimated by R.		

Summary

At this point,

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

can be computed combining

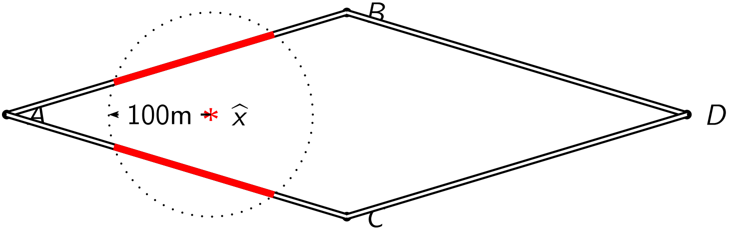
- a device measurement error model
- a movement model

But it involves complex integrals.

Computing integrals $I = \int_{x \in p} f(x)dx$

Truncate the domain of the integrals in each GPS point's DDR:

$$\exp\left(-\frac{||\widehat{x} - x||_2^2}{2\widehat{\sigma}^2}\right) \geq \theta$$



$$\theta = 0.65 \text{ and } \widehat{\sigma} = 104.4 = \sqrt{100^2 + 30^2}$$

Illustration: the actual path (0.08)

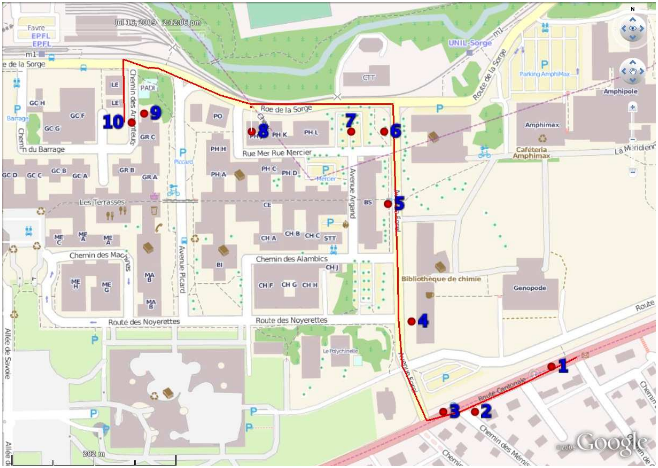


Illustration: another path (0.04)

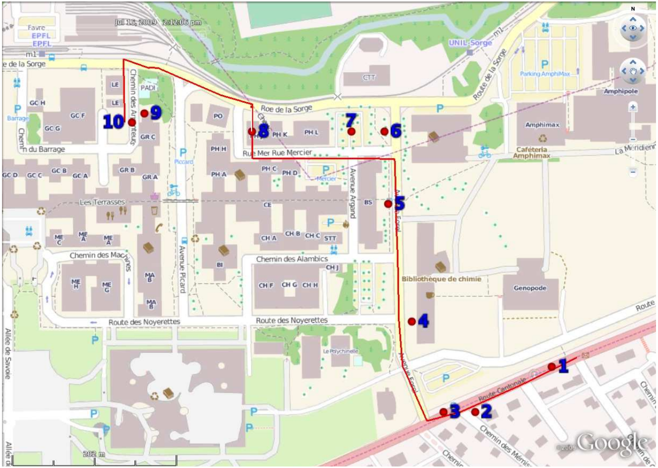
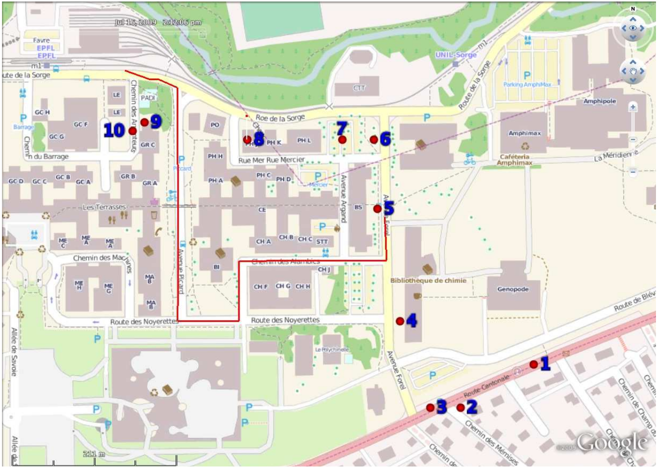


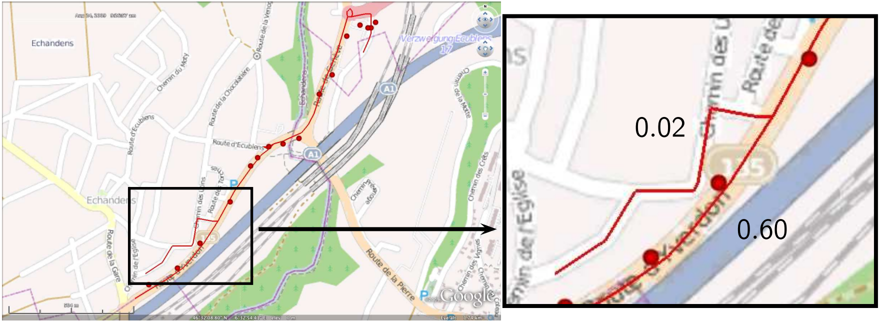
Illustration: map matching result (0)



Path generation algorithm

- Find DDR D_1 and associated link set L_1 for the first GPS point.
- generate P_1 from L_1.
- At each iteration k:
  - Build bounded shortest path trees from end nodes of P_{k-1}
  - Generate L_k.
  - Combine P_{k-1}, L_k and the shortest path trees to form P_k
  - Eliminate unlikely paths from P_k

Illustration: path generation algorithm



Details of a trip

path id	origin id	destination id	likelihood	normalized likeli-hood	length (km)	path size
1	289488138	287096130	0.01	0.02	1.65	0.03
2	289488138	287095960	0.01	0.01	1.62	0.03
3	313144339	287095960	0.07	0.12	1.66	0.04
4	313144339	287095960	0.06	0.10	1.70	0.04
5	313144339	287096130	0.09	0.14	1.74	0.04
6	313144339	287096130	0.07	0.12	1.78	0.05
7	313144339	287095960	0.06	0.10	1.70	0.05
8	313144339	287096130	0.10	0.16	1.69	0.03
9	313144339	287096130	0.08	0.14	1.73	0.03
10	313144339	287095960	0.05	0.09	1.74	0.06
Number of GPS points: 16, travel time: 3min						

Route choice model with probabilistic observations

$$\Pr(\widehat{x}_1, \ldots, \widehat{x}_T | S) = \sum_{s \in S} \Pr(s|S) \sum_{p \in C(s)} \Pr(\widehat{x}_1, \ldots, \widehat{x}_T | p) \Pr(p|C(s); \beta),$$

- $\Pr(s|S) = 1/\#S$;
- $\Pr(\widehat{x}_1, \ldots, \widehat{x}_T | p)$: measurement model
- $\Pr(p|C(s); \beta)$: route choice model.

Model specification

- Path Size Logit (PSL):

$$V_p = \beta_{PS} \ln PS_p + \beta_\ell \ln Length_p \quad (1)$$

$$PS_p = \sum_{a \in p} \frac{l_a}{l_p} \frac{1}{\sum_{q \in C(s)} \delta_{aq}} \quad (2)$$

- Alternative sampling: biased random walk: 50 draws, Kumaraswamy parameters $b_1 = 30$ and $b_2 = 0.4$
- Data: 17 car trips of one smartphone user

Model estimates

Coefficient	Value	Rob. Std. Error	Rob. t-test	p value
β_{PS}	3.95	1.40	2.81	0.00
β_I	-76.2	36.8	-2.07	0.04

Number of observations: 17
Null log-likelihood: -66.008
Final log-likelihood: -33.975
Adjusted rho-square: 0.455
Model estimated by BIOGEME

Conclusions

- The measurements likelihood model is meaningful.
- The path generation algorithm is suitable for smartphone data.
- The estimated route choice behavior is reasonable.