Structural estimation for a route choice model with uncertain measurement

Yuki Oyama* & Eiji Hato

*PhD student
Behavior in Networks studies unit
Department of Urban Engineering
School of Engineering, The University of Tokyo

September 25, 2016

Outline

1. Introduction
2. Link-based route measurement model
3. Structural estimation method
4. Numerical examples
5. Conclusions

Outline

1. **Introduction**
2. Link-based route measurement model
3. Structural estimation method
4. Numerical examples
5. Conclusions

Route choice analysis

*Parameter **estimation** results largely **depend on accuracy of route measurement***

GPS data

↓

Route measurement model

↓

Route choice data

↓

Route choice model

↓

Estimated parameters

Route choice analysis

*Parameter **estimation** results largely **depend on accuracy of route measurement***

GPS data

↓

Route measurement model

↓

Route choice data

↓

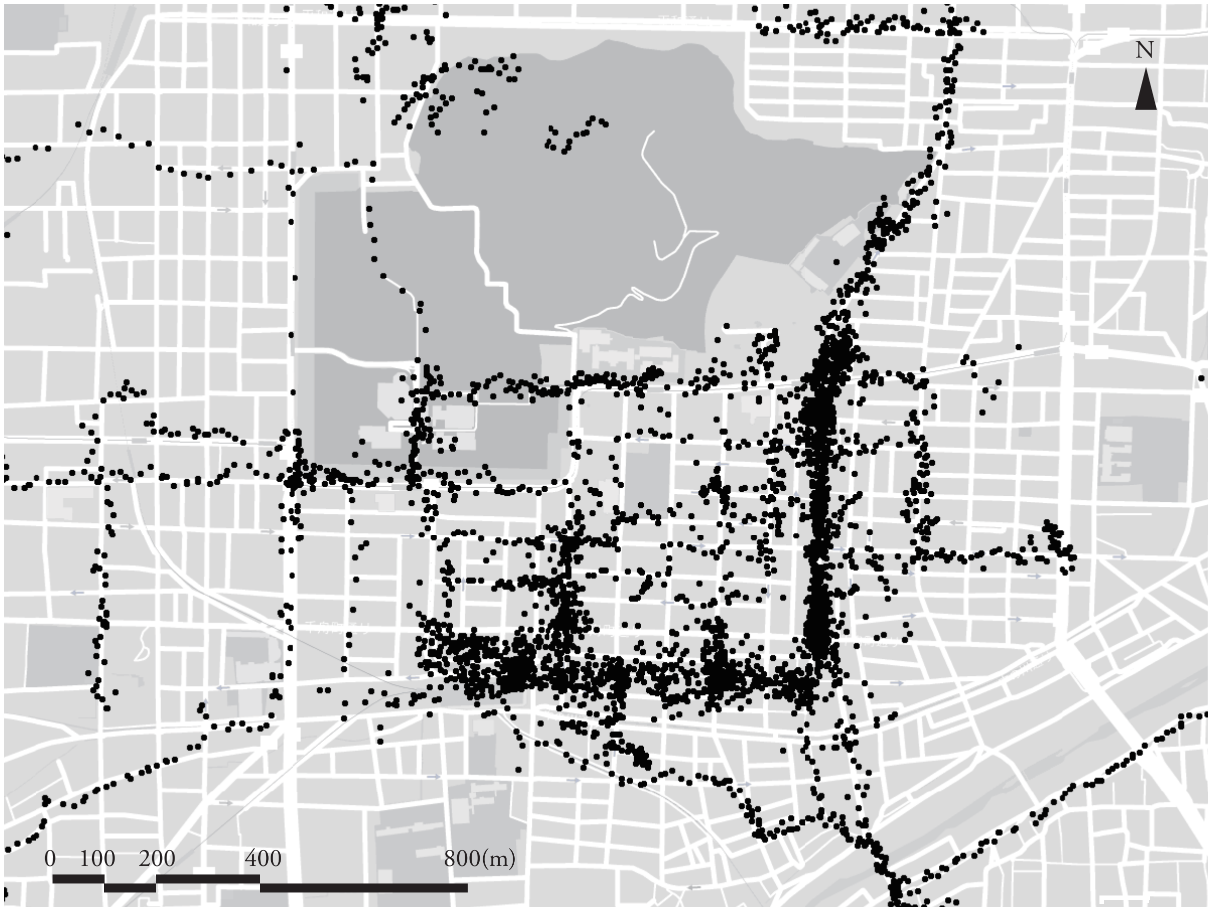
Route choice model

↓

Estimated parameters

Pedestrian route choice analysis

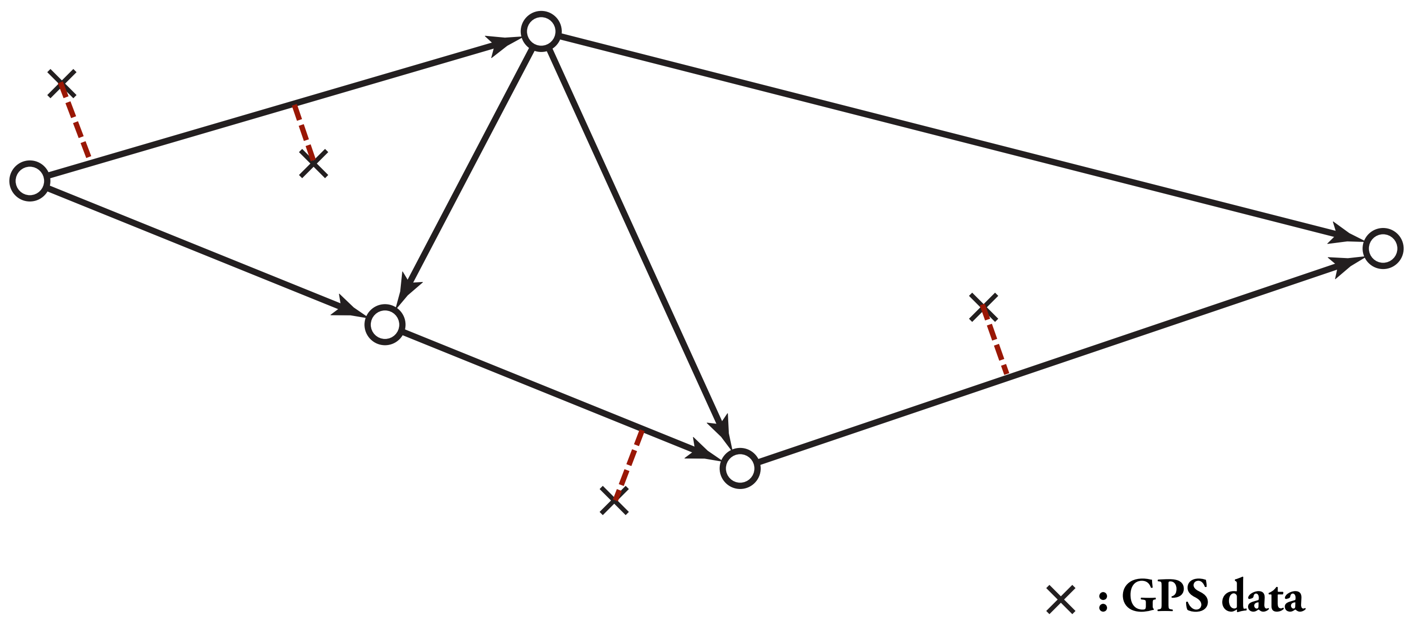
Measurement uncertainty; Dense and high-resolution network



- **Dense network**
- **Spatial dependence of Measurement errors**
  - ✓ Along river
  - ✓ Wide street
  - ✓ Narrow street
  - ✓ With arcade

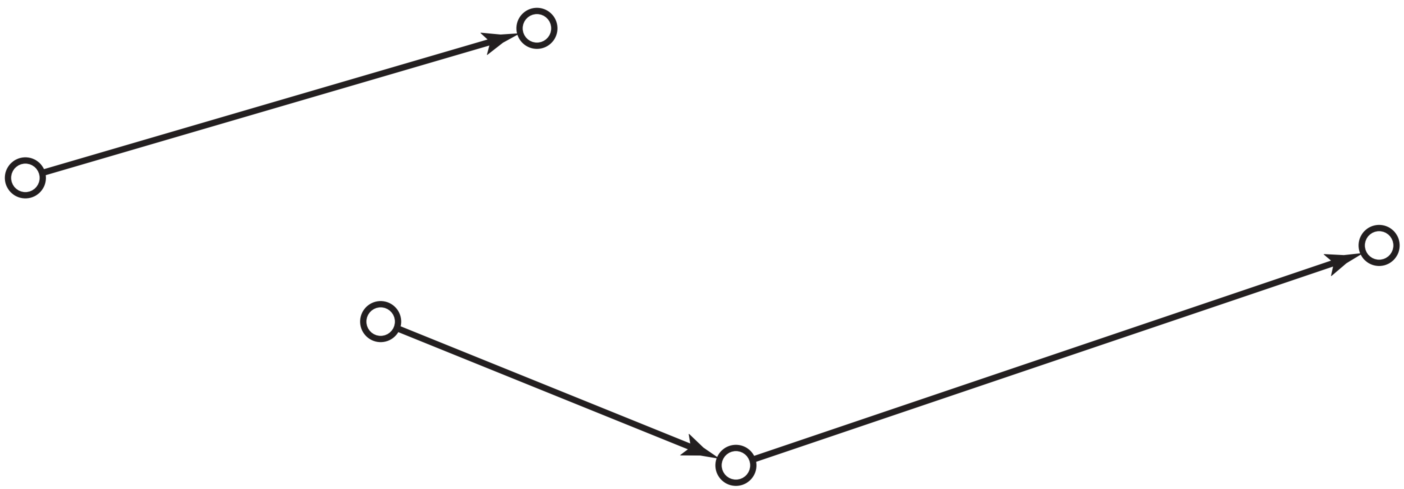
Route measurement models (1)

Sequential approach *infers the* ***true location at each data*** *in chorological order*



Route measurement models (1)

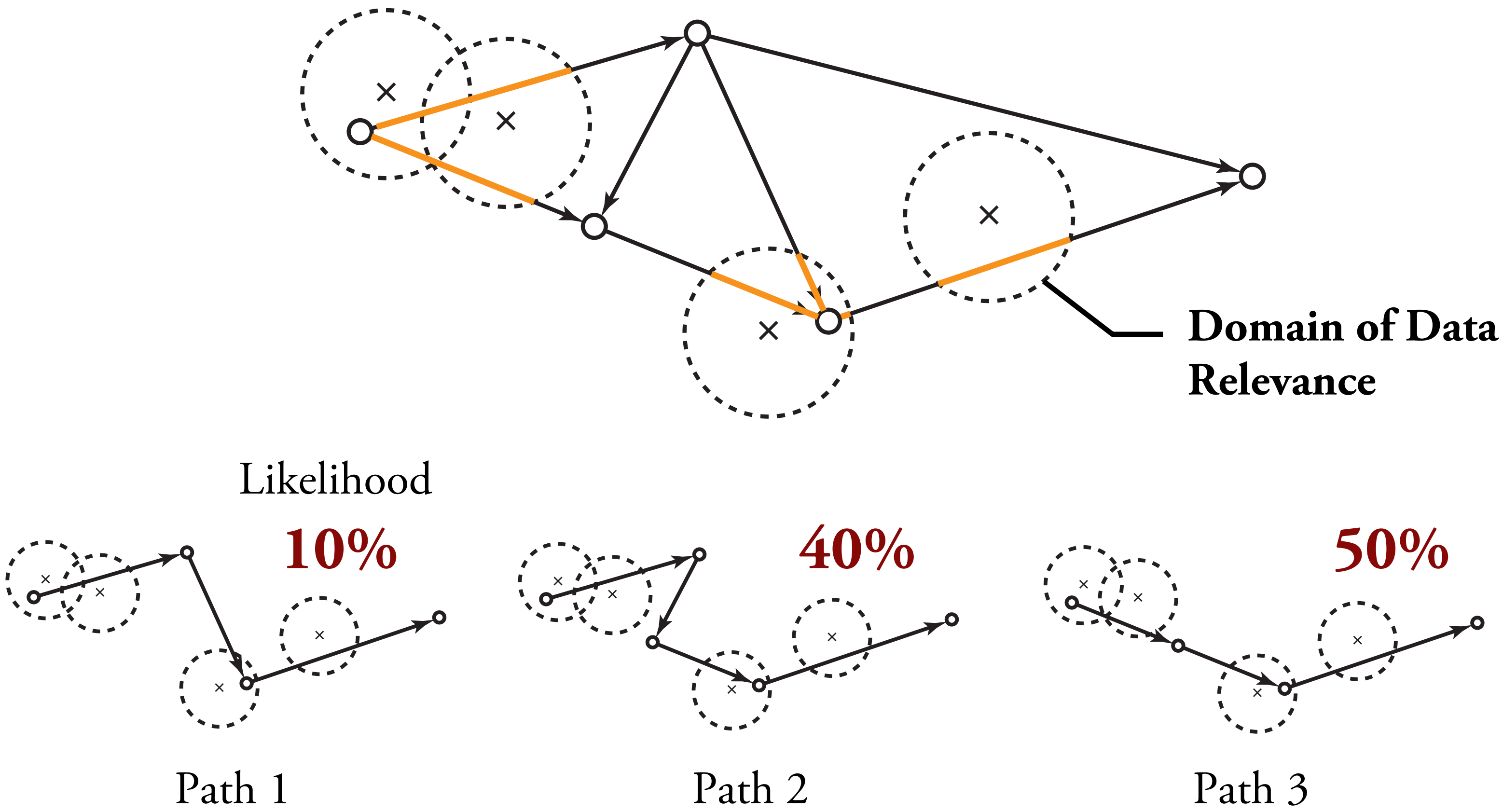
<u>Sequential approach</u> *can output* ***meaningless paths***



Route measurement models (2)

Path-based probabilistic approach *evaluates* ***path likelihood regarding all GPS data*** *included in a trip*

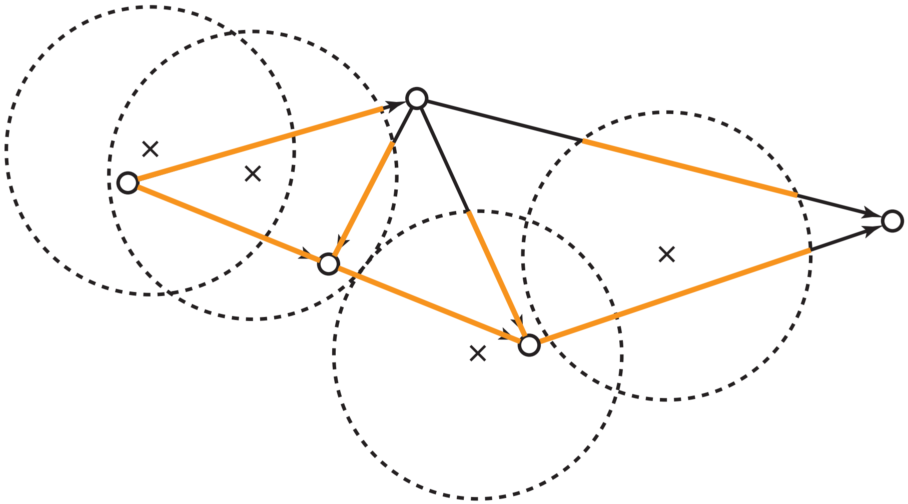
Pyo et al. (2001); Bierlaire et al. (2013)

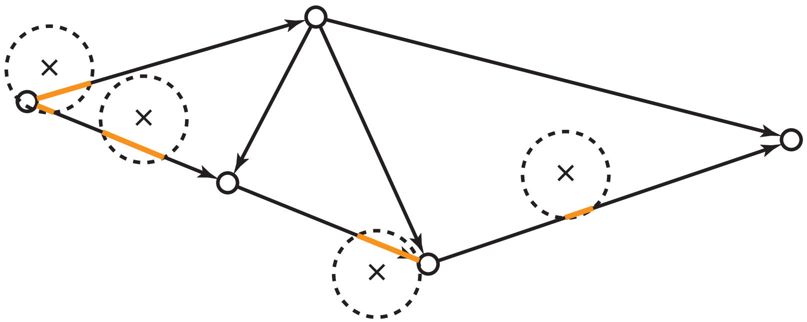


Route measurement models (2)

Path-based probabilistic approach *suffers with* ***trade-off*** *between computational efficiency and measurement accuracy*

Pyo et al. (2001); Bierlaire et al. (2013)





Large **Small**

Path candidate set

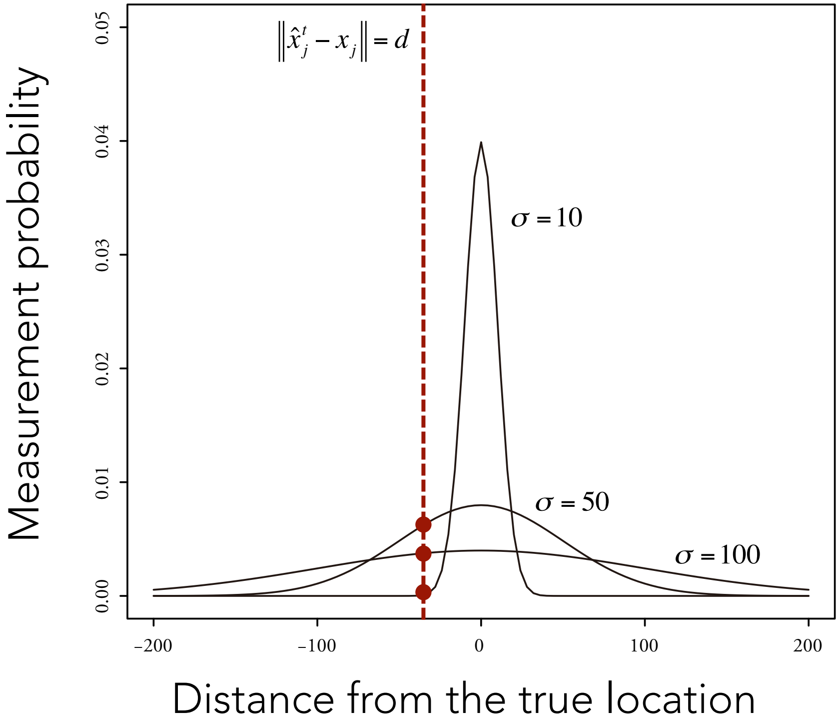
Route measurement models (2)

Assuming error variance ***constant*** *on network* ***distort measurement probabilities***

PDF of GPS measurement error: $$p(\hat{x}_j \mid x_j;\sigma) = \frac{\left\|\hat{x}_j - x_j\right\|}{\sigma^2}\exp\left(-\frac{\left\|\hat{x}_j - x_j\right\|^2}{2\sigma^2}\right)$$

σ : Large

σ : Small

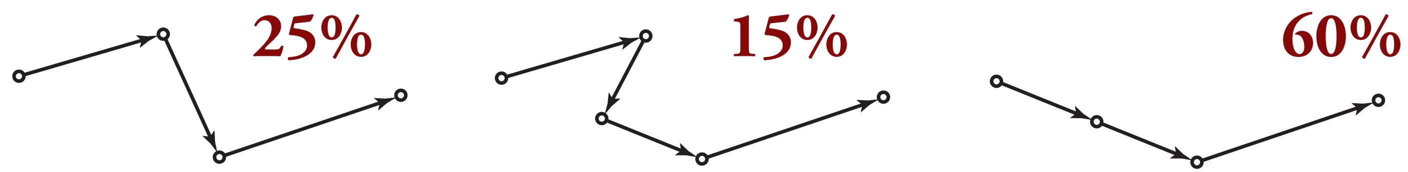


Route measurement models (3)

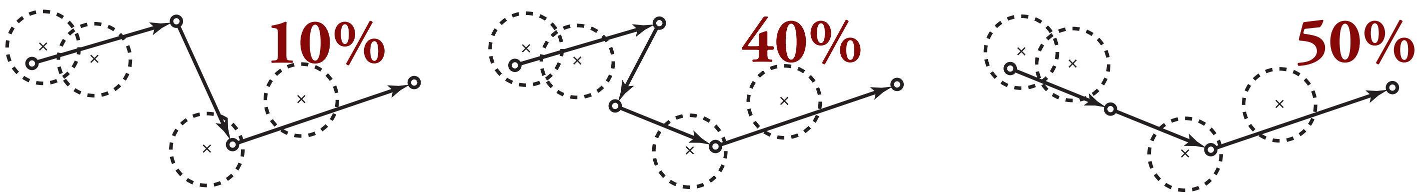
Bayesian approach ***incorporates behavioral models*** *into measurement models*

Chen et al. (2013); Danalet et al. (2014)

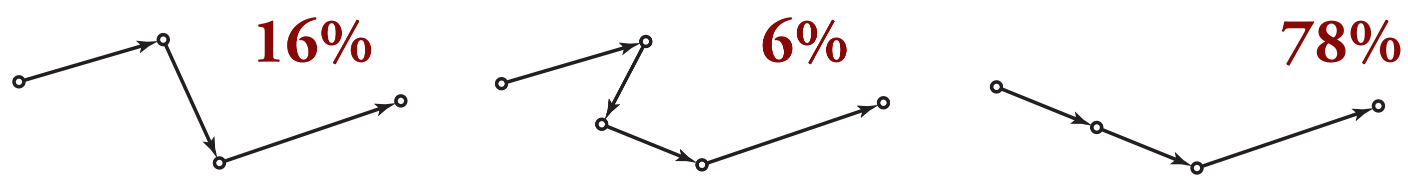
Prior (Route choice model)



Measurement

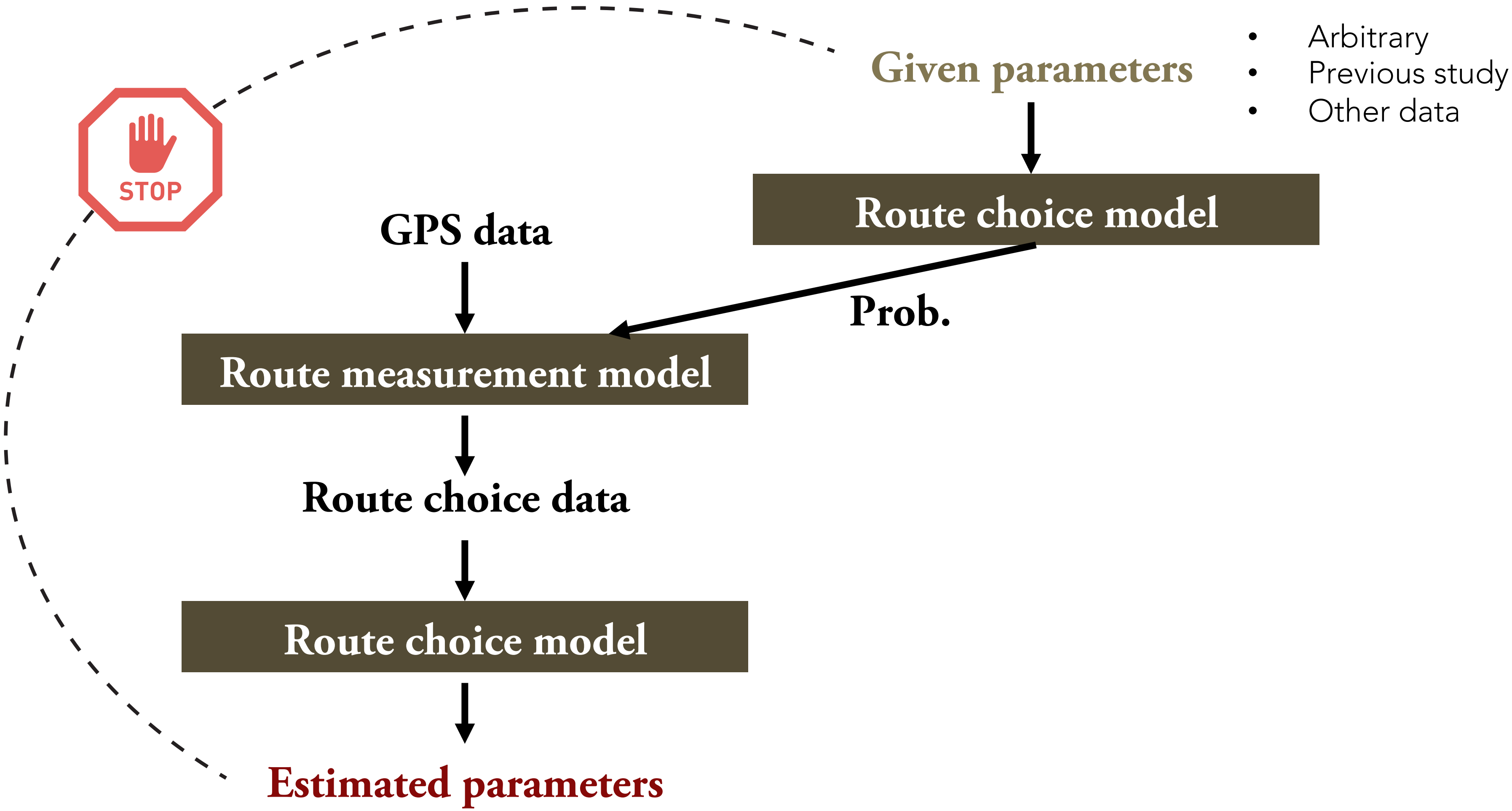


Posterior



Route measurement models (3)

Bayesian approach *has a problem regarding* ***parameter inconsistency***



Route measurement models

- **Challenges:**
  - *Disconnected path*
    - *Not suitable for route choice models*
  - *Setting of the measurement parameter*
    - *Possible to miss the true path*
    - Ignorance of spatial difference distorts path likelihood
  - *Parameter inconsistency of route choice model*
    - Estimated parameter includes biases regarding initial parameter

Framework

1. ***Link-based route measurement model***
   - Matching each decomposed trip data to a link
   - **Estimating a measurement parameter** for each link
   - Incorporating a **link-based route choice model** as prior

2. ***Structural estimation method***
   - Parameters at convergence **satisfy the parameter consistency** of the route choice model

Outline

1. Introduction
2. **Link-based route measurement model**
3. Structural estimation method
4. Numerical examples
5. Conclusions

Problem & Notation

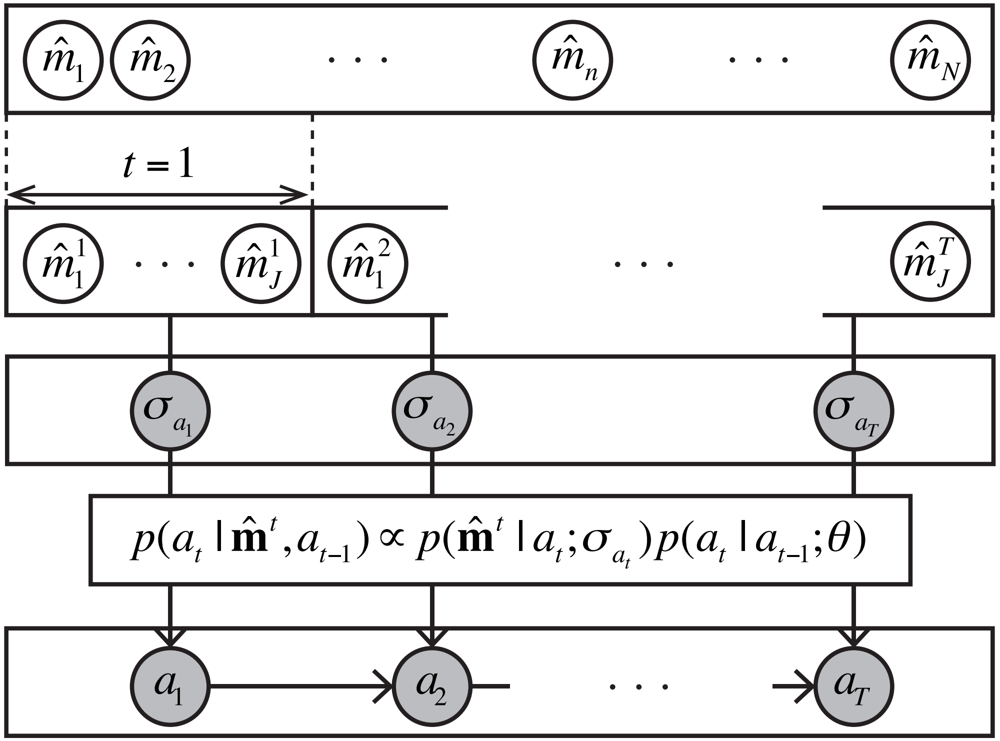
Matching row GPS data $\hat{\mathbf{m}}$ *to the transportation network* G

- **GPS data** $\hat{m} = (\hat{x}, \hat{\tau})$
  - Pair of coordinates $\hat{x} = (\hat{x}_{lat}, \hat{x}_{lon})$ with error variance σ
  - Timestamp $\hat{\tau}$
  - A given trip $\hat{\mathbf{m}} = (\hat{m}_1, \ldots, \hat{m}_n, \ldots, \hat{m}_N)$

- **Network** $G = (V, A)$
  - Node $v \in V$: the horizontal position $x_v = \{x_{lat}, x_{lon}\}$
  - Link $a = (v_u, v_d) \in A$: the vector of spatial attributes y_a
  - Network connection $\delta(a' | a)$: 1/0

Link-based route measurement

*Matching **all data observed within a period** to **the same link***



GPS data of a trip

$$\hat{\mathbf{m}} = (\hat{m}_1, \ldots, \hat{m}_n, \ldots, \hat{m}_N)$$
$$\hat{m}_n = (\hat{x}_n, \hat{\tau}_n)$$

Data decompopsition

$$\hat{\mathbf{m}} = (\hat{\mathbf{m}}^1, \ldots, \hat{\mathbf{m}}^t, \ldots, \hat{\mathbf{m}}^T)$$
$$\hat{\mathbf{m}}^t = (\hat{m}_1^t, \ldots, \hat{m}_j^t, \ldots, \hat{m}_J^t)$$

Estimation: σ_{a_t}

Link probabilities

Path inference:

$$\psi = [a_1, \ldots, a_t, \ldots, a_T]$$

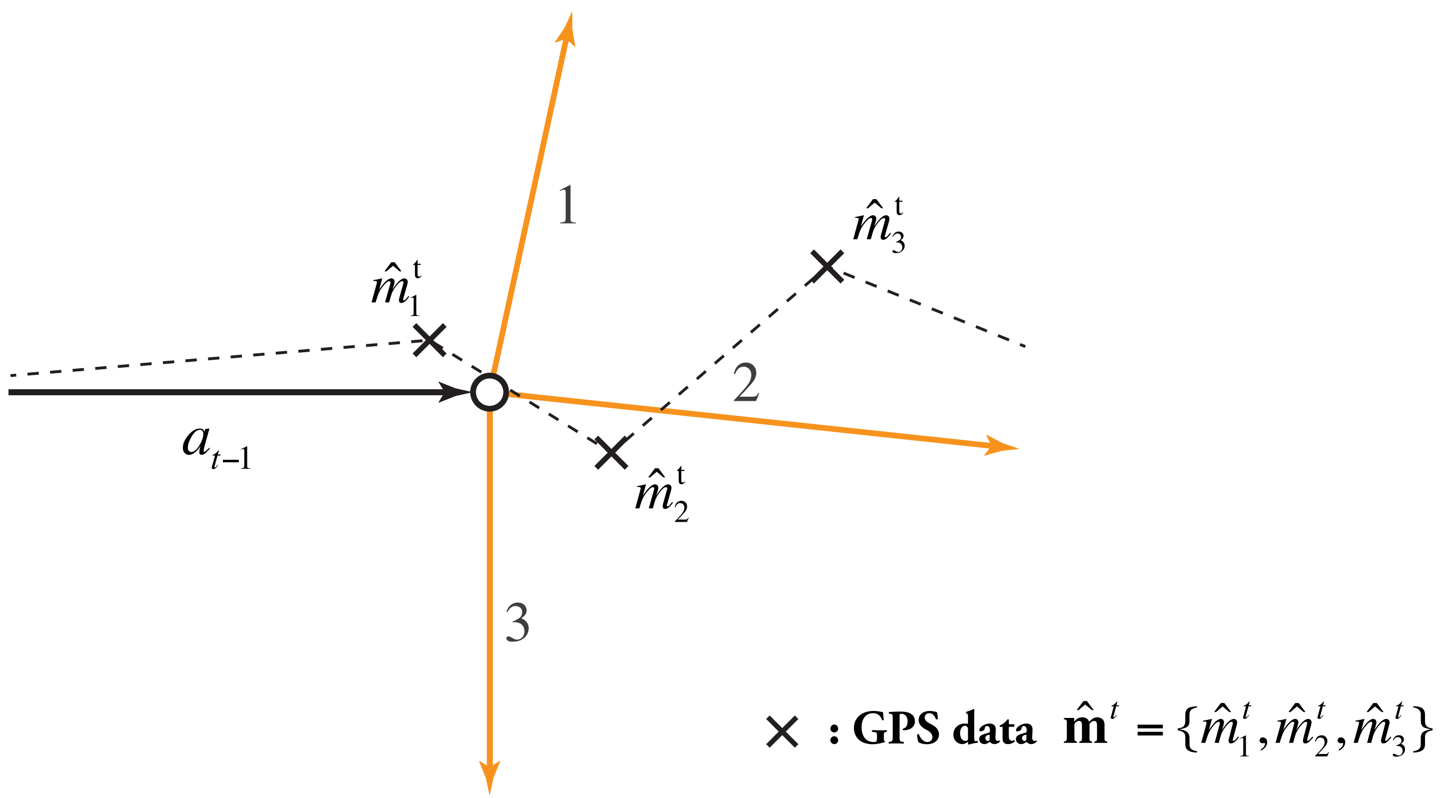
$p(\hat{\mathbf{m}}^t \mid a_t)$: Measurement probability of $\hat{\mathbf{m}}^t$ given a_t ; *measurement equation*

$p(a_t \mid a_{t-1})$: Prior probability of a_t given a_{t-1} ; *system equation*

Link probability $p(a_t \mid \hat{\mathbf{m}}^t, a_{t-1})$

The probability of a_t *given measurements* $\hat{\mathbf{m}}^t$ *and state* a_{t-1}

- **Candidate set**: $A(a_{t-1}) = \{a_t \mid \delta(a_t \mid a_{t-1}) = 1\}$
  - Calculate link probabilities for all $a_t \in A(a_{t-1})$



Measurement equation $p(\hat{\mathbf{m}}^t \mid a_t; \sigma_{a_t})$

The probability of measurements $\hat{\mathbf{m}}^t$ *given* a_t

- **Assumption**:
  - Timestamp $\hat{\boldsymbol{\tau}}$ has no measurement error; $p(\hat{\mathbf{m}}^t \mid a_t) = p(\hat{\mathbf{x}}^t \mid a_t)$
  - Measurement probability of data is independent from each other
  - Traveler moves at the constant speed on the same link

$$p(\hat{x}_1^t, \ldots, \hat{x}_J^t \mid a_t; \sigma_{a_t}) = \prod_{j=1}^{J} p(\hat{x}_j^t \mid a_t; \sigma_{a_t})$$

$$= \prod_{j=1}^{J} \int_{x_j \in a_t} p(\hat{x}_j^t \mid x_j^t, a_t; \sigma_{a_t}) p(x_j \mid a_t) dx_j$$

PDF of GPS measurement error: *Rayleigh distribution* (van Diggelen, 2007)

$$p(\hat{x}_j^t \mid x_j^t, a_t; \sigma_{a_t}) = \frac{\left\| \hat{x}_j^t - x_j \right\|}{\sigma_{a_t}^2} \exp\left(-\frac{\left\| \hat{x}_j^t - x_j \right\|^2}{2\sigma_{a_t}^2} \right)$$

Estimation of measurement parameter σ_{a_t}

Link-based map matching can regard error variance as a ***link peculiar variable***

Measurement likelihood maximization

$$\sigma_{a_t} = \underset{\sigma}{\operatorname{argmax}}\, p(\hat{\mathbf{m}}^t \mid a_t; \sigma)$$

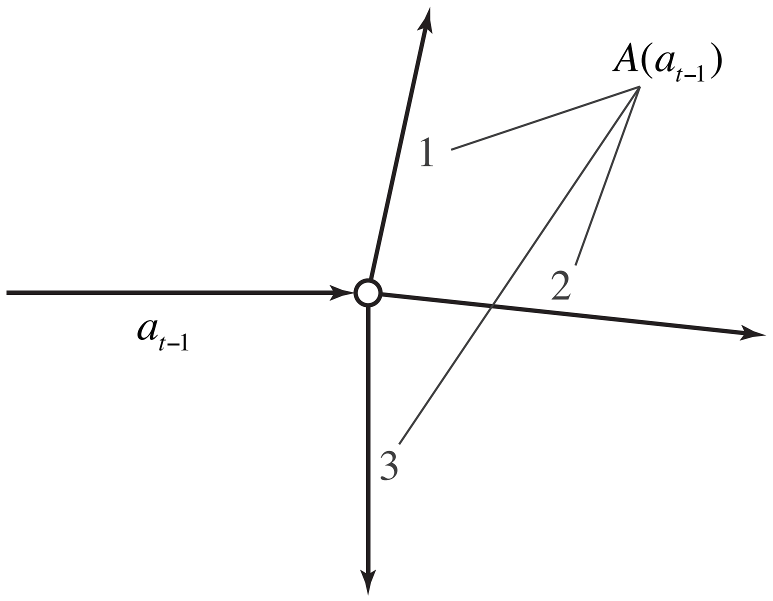
Where,

$$p(\hat{\mathbf{m}}^t \mid a_t; \sigma_{a_t}) = \prod_{j=1}^{J} \int_{x_j \in a_t} \left\{ \frac{\left\| \hat{x}_j^t - x_j \right\|}{\sigma_{a_t}^2} \exp\left(-\frac{\left\| \hat{x}_j^t - x_j \right\|^2}{2\sigma_{a_t}^2} \right) \cdot p(x_j \mid a_t) \right\} dx_j$$

System equation $p(a_t \mid a_{t-1};\theta)$

The prior probability of a_t *given a state* a_{t-1}

Link-based route choice model



Utility function:

$$u(a_t \mid a_{t-1}) = v(a_t \mid a_{t-1}) + \varepsilon(a_t) = \theta \mathbf{y}_{a_t \mid a_{t-1}} + \varepsilon(a_t)$$

$v(\cdot)$: Deterministic component of utility

$\varepsilon(\cdot)$: Probabilistic component of utility (i.i.d. gumbel distribution)

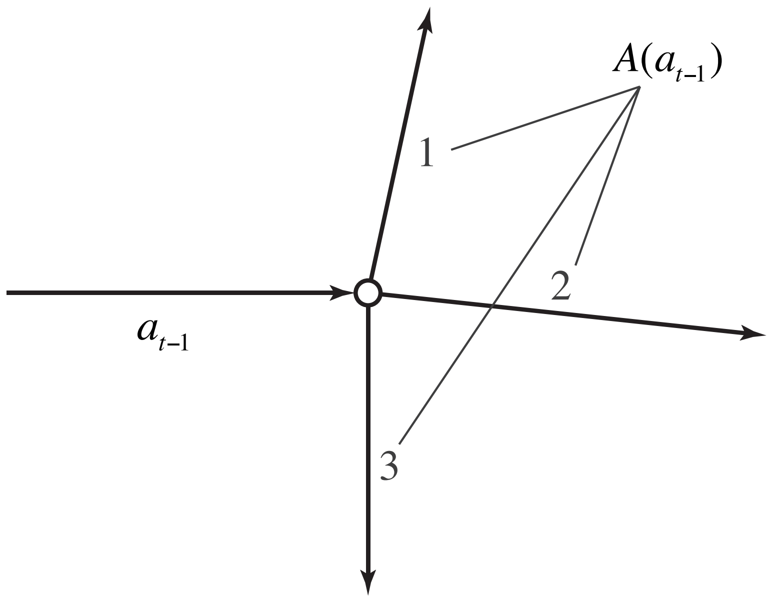
$\mathbf{y}_{a_t \mid a_{t-1}}$: Vector of explanatory variables

θ : Vector of parameters

System equation $p(a_t \mid a_{t-1};\theta)$

The prior probability of a_t *given a state* a_{t-1}

Link-based route choice model



Choice probability option:

Markov model

$$p(a_t \mid a_{t-1}) = \frac{e^{v(a_t|a_{t-1})}}{\sum_{a_t \in A(a_{t-1})} e^{v(a_t|a_{t-1})}}$$

Recursive logit model *Fosgerau et al. (2013)*

$$p(a_t \mid a_{t-1}) = \frac{e^{v(a_t|a_{t-1})+V^d(a_t)}}{\sum_{a_t \in A(a_{t-1})} e^{v(a_t|a_{t-1})+V^d(a_t)}}$$

And others: e.g.,

Mai et al. (2015); Mai (2016); Oyama et al. (2016)

Link inference

- **Link (posterior) probability:**
  - *The probability of* a_t *given measurements* $\hat{\mathbf{m}}^t$ *and a state* a_{t-1}

$$p(a_t \mid \hat{\mathbf{m}}^t, a_{t-1}) \propto p(\hat{\mathbf{m}}^t \mid a_t; \sigma_{a_t}) p(a_t \mid a_{t-1}; \theta)$$

- **Link inference:**
  - *Link likelihood maximization subject to switching condition*

$$a_t = \underset{a_t \in A(a_{t-1})}{\operatorname{argmax}} \, p(a_t \mid \hat{\mathbf{m}}^t, a_{t-1})$$

$$\text{s.t.,} \quad \max_{a \in A(a_t)} p(\hat{\mathbf{m}}^{t+1} \mid a; \sigma_a) > \gamma$$

Outline

1. Introduction
2. Link-based map matching algorithm
3. **Structural estimation method**
4. Numerical examples
5. Conclusions

A fixed point problem

Need to solve a fixed point problem *of route measurement and estimation*

Route measurement

Route choice probability

$p(a_t \mid a_{t-1}; \boxed{\tilde{\theta}})$

$\downarrow$

Path ψ

$\uparrow$

Measurement probability

$p(\hat{\mathbf{m}}^t \mid a_t; \sigma_{a_t})$

Route choice model

$$\boxed{\theta =} \operatorname*{argmax}_{\theta} LL(\theta; \psi)$$

s.t., $|\theta - \tilde{\theta}| < \zeta$

Where,

$$LL(\theta;\psi) = \ln\left(\prod_i \prod_{t=2}^{T} p(a_t \mid a_{t-1}; \theta)^{\delta^i_{a_t}}\right)$$

$$\delta^i_{a_t} = \begin{cases} 1, \text{ if } a_{i,t} \in \psi_i \\ 0, \text{ otherwise} \end{cases}$$

Structural estimation

A method for parameter estimation of models with fixed point problem

- *NFXP* (Nested Fixed Point) *Rust (1987)*
- *NPL* (Nested Pseudo Likelihood) *Aguirregabiria and Mira (2002)*
- *MPEC* (Mathematical Programming with Equilibrium Constraint) *Su and Judd (2012)*
- …

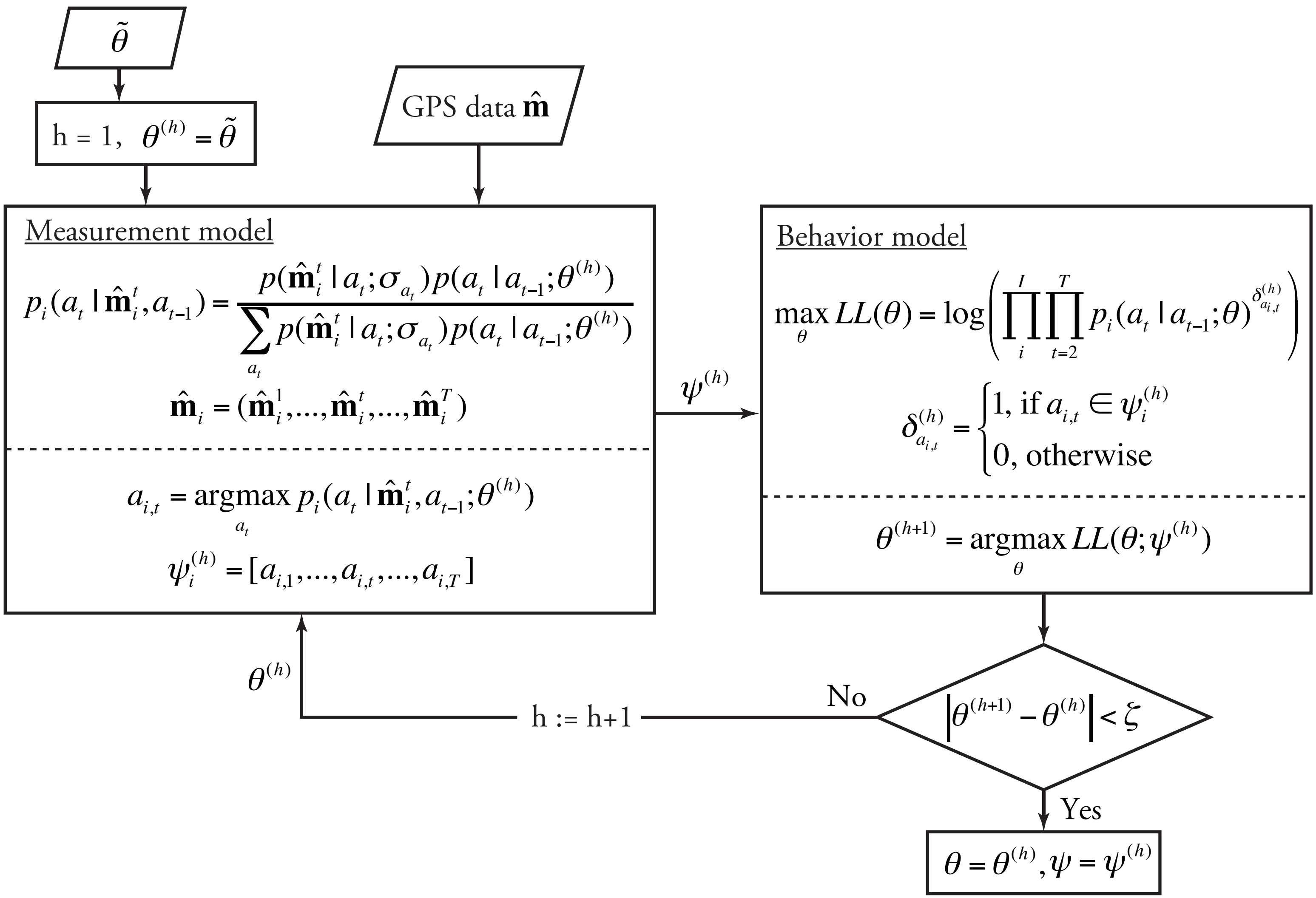
Structural estimation for route choice model with uncertain data

- Solving a fixed problem regarding parameter of route choice model

- Inner problem: **Route measurement model**
- Outer problem: **Parameter estimation of route choice model**

Structural estimation

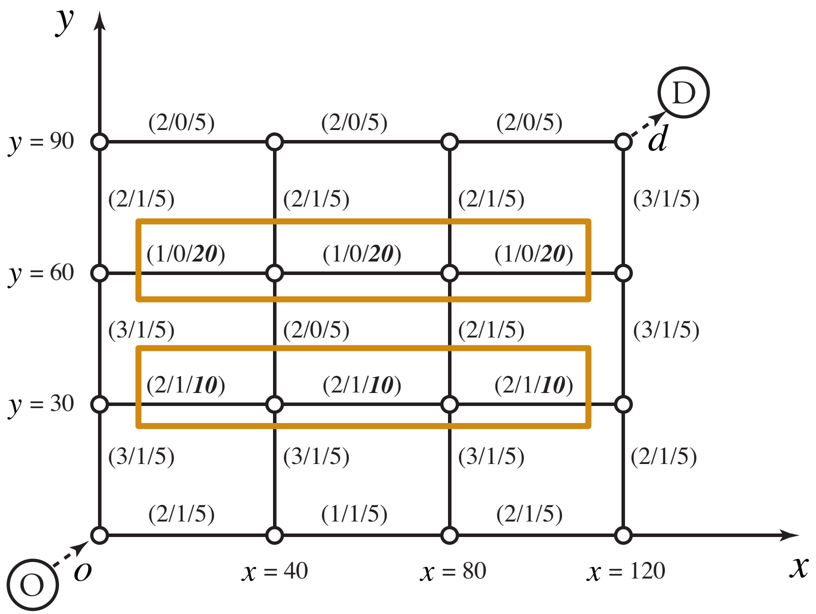
A estimation method for solving a fixed point problem *of route choice parameter*



Outline

1. Introduction
2. Link-based map matching algorithm
3. Structural estimation method
4. **Numerical examples**
5. Conclusions

Twins experiments | Simulation



*(continuous cost: CC_a / discrete cost: DC_a / variance: σ_a)

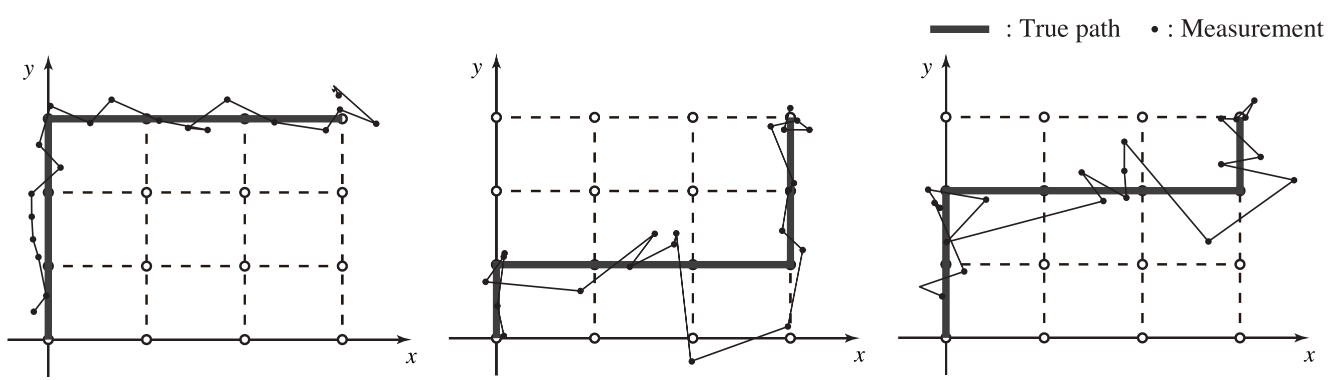
Settings

$$v(a\,|\,k) = \theta_1 TT_a + \theta_2 CC_a + \theta_3 DC_a + \theta_4 UT_{a|k}$$

True parameter: $\tilde{\theta} = [-0.1, -2, -1.5, -4]$

Period interval: $\bar{t} = 30s$

Data generation: $\hat{\tau}_j - \hat{\tau}_{j-1} = 10s$



(a) (b) (c)

Twins experiments | Measurement results

Which model improves the route measurement accuracy ?

Table: Measurement accuracy and the difference of the parameter from the true value

				accuracy(%)		Ave. $\lvert\sigma_{\text{est}} - \sigma_{\text{true}}\rvert$	
	Model	σ	$\tilde{\theta}$	-	Switching	-	Switching
1	MEQ	given	-	54.571	68.857	-	-
2	MEQ	estimated	-	76.857	82.857	5.848	4.397
3	MEQ+SEQ	estimated	$[0, 0, 0, 0]$	76.857	82.857	5.848	4.397
4	MEQ+SEQ	estimated	$[-1.5, -0.1, -2, -10]$	4.857	38.286	41.992	21.206
5	MEQ+SEQ	estimated	$[-0.1, -2, -1.5, -4]$	76.857	91.714	7.579	4.056

*MEQ: Measurement Equation

*SEQ: System Equation

Twins experiments | Estimation results

*Does **structural estimation method** improve the parameter estimation results ?*

Input: $\tilde{\theta} = [0, 0, 0, 0]$ (No information)

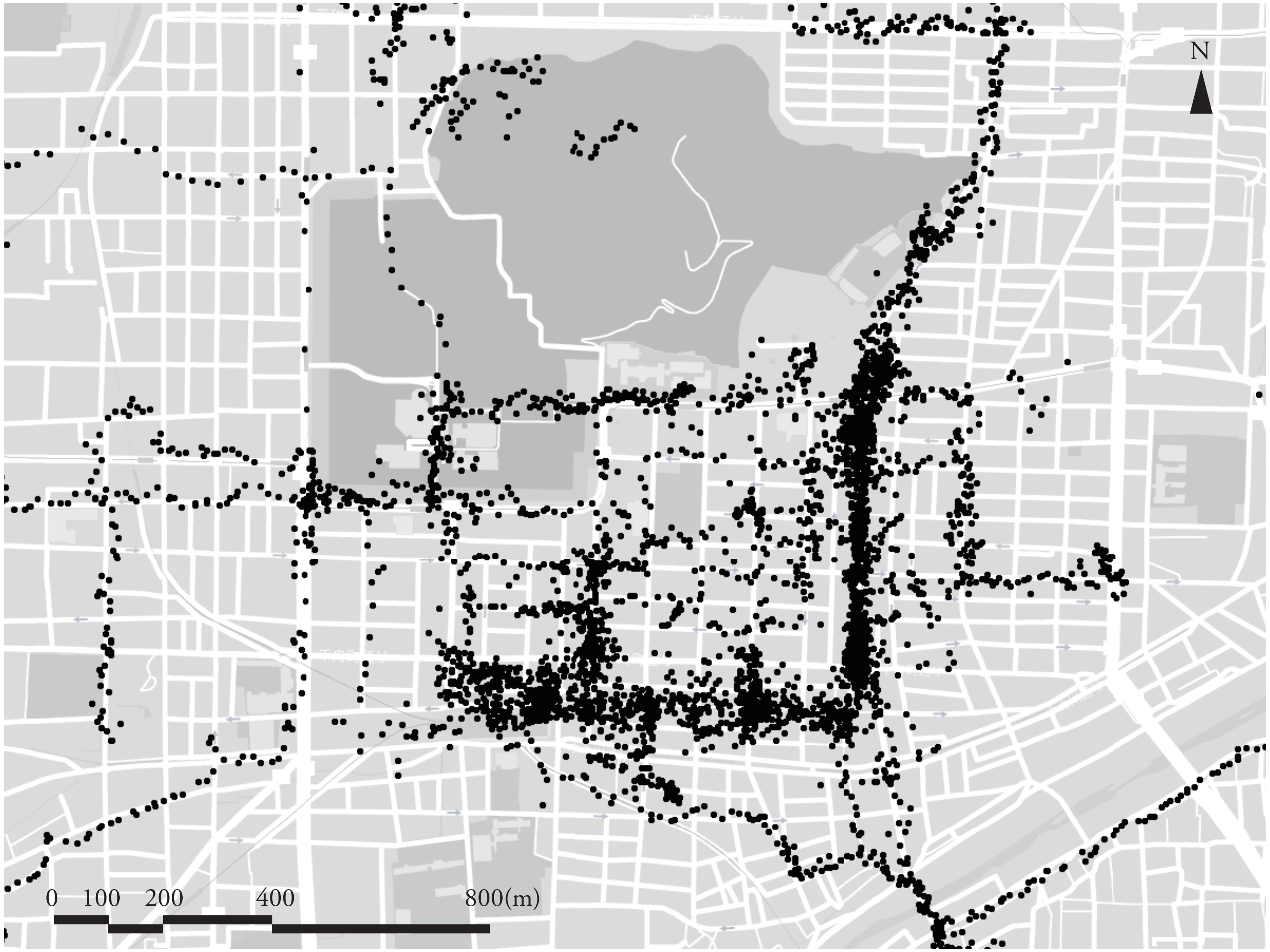
	One-way				Structural Estimation		
	TRUE	Estimates	abs(diff.*)	t-value	Estimates	abs(diff.)	t-value
θ_1	-0.1	0.002	0.102	0.101	-0.064	0.036	-2.562
θ_2	-2	-0.755	1.245	-4.164	-1.727	0.273	-6.882
θ_3	-1.5	-1.312	0.188	-4.772	-1.046	0.454	-3.519
θ_4	-4	-1.892	2.108	-8.864	-3.519	0.481	-9.739
total error			3.643			1.244	
sample				350			350
L0				-373.221			-371.887
LL				-269.872			-211.308
ρ^2				0.266			0.421
iteration							6

Input: $\tilde{\theta} = [-1.5, -0.1, -2, -10]$ (Wrong values)

	One-way				Structural Estimation		
	TRUE	Estimates	abs(diff.)	t-value	Estimates	abs(diff.)	t-value
θ_1	-0.1	-0.097	0.003	-5.312	-0.064	0.036	-2.562
θ_2	-2	-0.419	1.581	-2.710	-1.727	0.273	-6.882
θ_3	-1.5	0.178	1.678	0.963	-1.046	0.454	-3.519
θ_4	-4	-1.204	2.796	-6.774	-3.519	0.481	-9.739
total error			6.058			1.244	
sample				350			350
L0				-373.560			-371.887
LL				-328.587			-211.308
ρ^2				0.110			0.421
iteration							8

Real data

Matsuyama Probe Person data in 2007, 30 pedestrians, 729 locations



Real data | Model specification

- **Route choice model**: (static) Markov model
- **Target:** Pedestrian trip in city center

- **Utility function:**

$$v(a \mid k) = \theta_1 TT_a + \theta_2 CU_a + \theta_3 DU_a + \theta_4 UT_{alk}$$

TT : Travel time (min.)

CU : Sidewalk width (m)

DU : Arcade dummy variable

UT : U-turn dummy variable

Real data | Parameter estimation results

- Travel time (θ_1) seems to be significant from the result of ***one-way model***, however,
- ***Structural estimation*** results show that links with arcade (θ_3) are the most likely to be passed by pedestrians; travel time (θ_1) is not significant
- Other t-values and rho-square (ρ^2) indicate that the ***structural estimation*** improves parameter estimation results

Input: $\tilde{\theta} = [0, 0, 0, 0]$ (No information)

		One-way		Structural Estimation	
		Estimates	t-value	Estimates	t-value
Travel time (min.)	θ_1	-0.007	-2.473	-0.001	-0.428
Sidewalk width (m)	θ_2	0.088	1.497	0.134	1.582
With arcade	θ_3	-0.004	-0.011	2.760	4.288
U-turn	θ_4	0.774	0.532	0.469	3.344
	sample		270		270
	L0		-307.608		-309.066
	LL		-302.174		-225.162
	ρ^2		0.005		0.259
	iteration				11

Real data | Estimation results of error variance

Estimated measurement error variance is dependent on each link

	Average	Variance
σ_a	31.622	941.021

Estimation result of σ

- *~ 20 (or no measurement)*
- *20 ~ 30*
- *30 ~ 50*
- *50 ~ 100*
- *100 ~*

Outline

1. Introduction
2. Link-based map matching algorithm
3. Structural estimation method
4. Numerical examples
5. **Conclusions**

Conclusions and Future work

- **Conclusions**
  - *A link-based measurement model* with route choice model
  - Estimation of measurement parameter for each link
  - *Structural estimation method* for solving a fixed point problem regarding route choice parameters
- **Future work**
  - Comparison of computational efficiency with previous measurement models
  - Alternatives and utility of pedestrian link choice
  - Characteristics of the fixed point problem

Thank you for attention!
Questions?

oyama@bin.t.u-tokyo.ac.jp

References

1. Bierlaire, M., Chen, J., Newman, J., 2013. A probabilistic map matching method for smartphone GPS data. Transportation Research Part C: Emerging Technologies 26, 78-98.
2. Bierlaire, M., Frejinger, E., 2008. Route choice modeling with network-free data. Transportation Research Part C: Emerging Technologies 16(2), 187-198.
3. Chen, J., Bierlaire, M., 2015. Probabilistic multimodal map matching with rich smartphone data. Journal of Intelligent Transportation System 19(2), 134-148.
4. Danalet, A., Farooq, B., Bierlaire, M., 2014. A bayesian approach to detect pedestrian destination-sequences from WiFi signatures. Transportation Research Part C: Emerging Technologies 44, 146-170.
5. Fuse, T., Nakanishi, W., 2012. A study on multiple human tracking by integrating pedestrian behavior model (in Japanese). Journal of JSCE Series D3: Infrastructure Planning and Management 68(2), 92-104.
6. Fosgerau, M., Frejinger, E. and Karlstrm A., 2013. A link-based network route choice model with unrestricted choice set. Transportation Research Part B, 56, pp.70-80.
7. Hunter, T., Abbeel, P., Bayen, A., 2014. The path inference filter: model-based low-latency map matching of probe vehicle data. IEEE Transactions on Intelligent Transportation Systems 15(2), 507-529.
8. Pyo, J.S., Shin, D.H., Sung, T.K., 2001. Development of a map matching method using the multiple hypothesis technique. IEEE Proceedings on Intelligent Transportation Systems, 23-27.
9. Quddus, M.A., Noland, R.B., Ochieng, W.Y., 2005. Validation of map matching algorithms using high precision positioning with GPS. The Journal of Navigation 58(2), 257-271.
10. Quddus, M.A., Ochieng, W.Y., Noland, R.B., 2007. Current map-matching algorithms for transport applications: State-of-the art and future research directions. Transportation Research Part C: Emerging Technologies 15(5), 312-328.
11. Quddus, M.A., Ochieng, W.Y., Zhao, L., Noland, R.B., 2003. A general mao matching algorithm for transport telematics applications. GPS Solutions 7(3), 157-167.
12. Quddus, M.A., Washington, S., 2015. Shortest path and vehicle trajectory aided mapmatching for low frequency GPS data. Transportation Research Part C: Emerging Technologies 55, 328-339.
13. van Diggelen, F., 2007. GNSS accuracy: lies, damn lies, and statistics. GPS World 18(1), 26-32.
14. Velaga, N.R., Quddus, M.A., Bistow, A.L., 2009. Developing an enhanced weight-based topological map-matching algorithm for intelligent transport systems. Transportation Research Part C: Emerging Technologies 17, 672-683.
15. White, C.E., Bernstein, D., Kornhauser, A.L., 2000. Some map matching algorithms for personal navigation assistants. Transportation Research Part C: Emerging Technologies 8, 91-108.

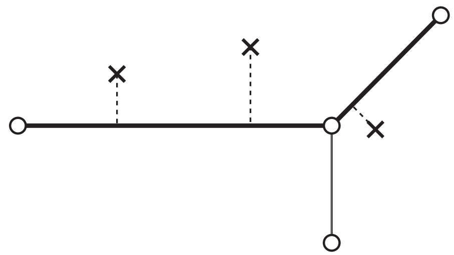
Appendix

Previous route measurement models

Geometric

White et al. (2000)

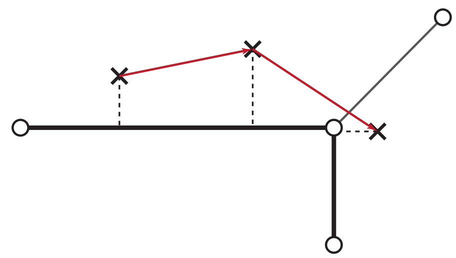
- Point-to-point
- Point-to-curve
- Curve-to-curve



Topological

Greenfield (2002)

- Adjacency
- Connectivity
- Vehicle heading

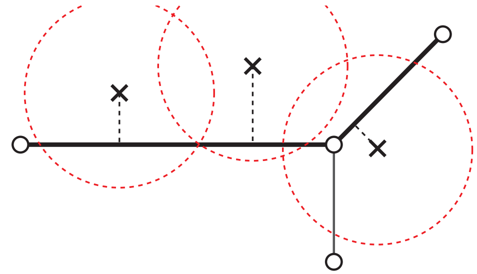


Probabilistic

Ouchieng et al. (2004)

Quddus et al. (2006)

- **Error region**
- Fuzzy logic

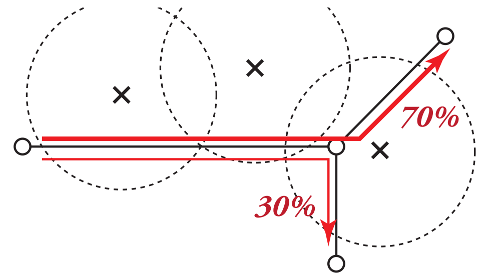


Path-based

Pyo et al. (2006)

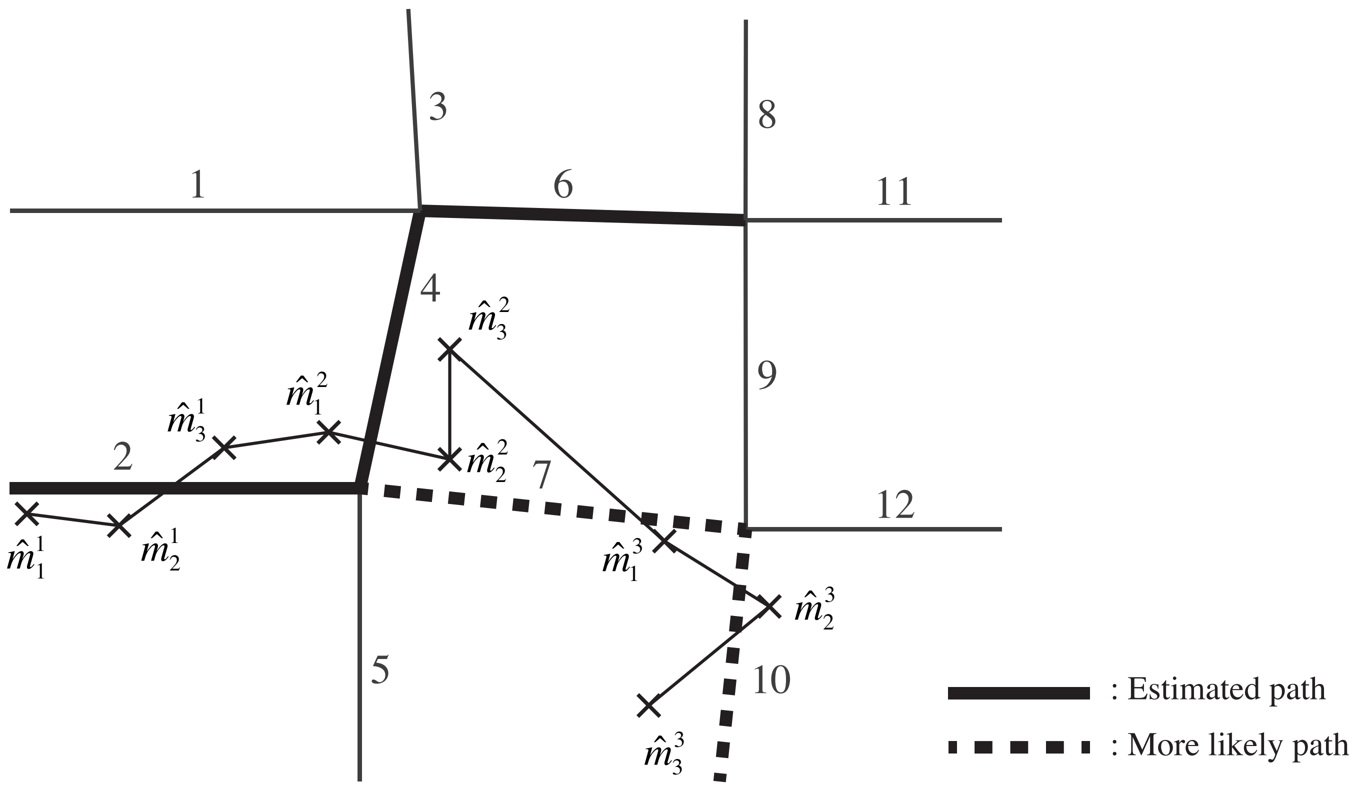
Bielraire et al. (2013)

- MHT
- **Measurement equation**

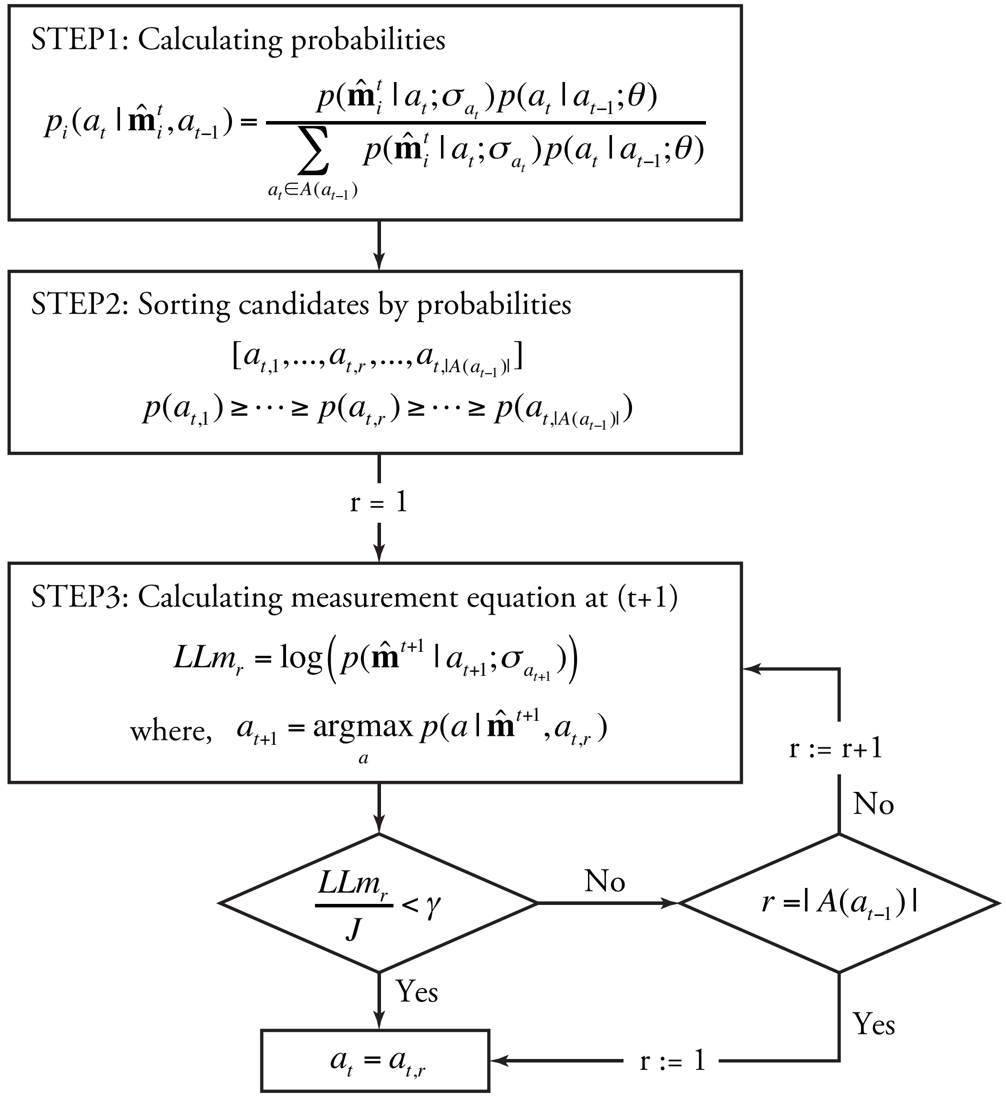


Link switching | errors

Difficulties regarding link connectivity because of myopic optimization



Link switching | algorithm



Twins experiments | iterations