

# A Bayesian approach to detect pedestrian destination-sequences from WiFi signatures

Antonin Danalet, Bilal Farooq, Michel Bierlaire  
Transportation Research Part C 44, pp. 146-170, 2014

発表者：福山 祥代

# Introduction

## 歩行者の活動選択モデル

### ■活動スケジュールの意思決定に関するモデルの必要性

- ・ 時間帯による活動施設の制約 ・ 経路上にある目的地の選択などの考慮が必要
- ・ 新しい施設の効率的なデザイン, 移動案内・情報システムの構築のために有効

現状： activity-based モデルの提案は多いが歩行者行動を扱うものは少ない

└── データがないことが主な原因

### ■データの利用可能性の向上と、データに対応するモデルの必要性

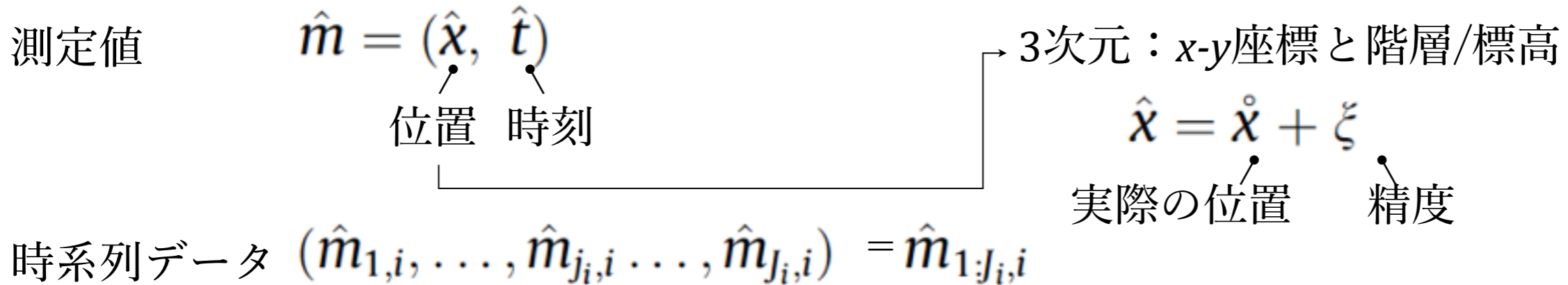
- ・ 携帯電話やWiFiなどの普及により利用可能に
- ・ 位置データの密度や精度が不十分

データの誤差を明確に扱う モデルの提案

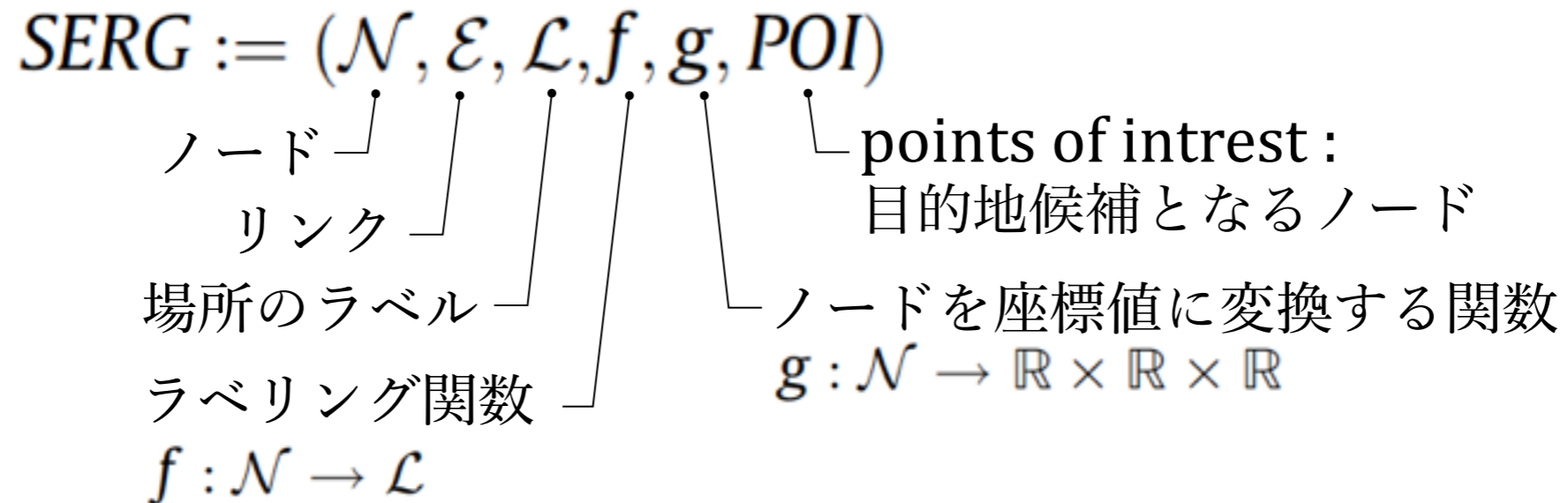
施設の事前情報を利用

# Data requirement

Network traces :data from communication network infrastructure



Semantically-enriched routing graph (*SEREG*)



# Methodology

測定されたNetwork traces + *SERG*

活動系列の生成と尤度の計算

activity episode  
一つの活動

$$a = (x, t^-, t^+)$$

位置      開始時刻      終了時刻

$$x \in POI \quad t^+ - t^- \geq T_{min}$$
$$f(x) \in \mathcal{L}$$

output:  $L$ 個の活動系列候補

$$(a_1, \dots, a_{K_i}, \dots, a_{K_i}) = a_{1:K_i}$$

$K_i$  活動の数  
個人ごとに異なり  
分析者には不明

## 1. Activity probability

活動の尤度の計算方法の提示

- 1) Measurement likelihood
- 2) prior 事前確率

## 2. Activity-episode sequences

活動系列の生成方法の提示

- 1) Episode location
- 2) Episode start and end times

## 3. Intermediary measurements

中間測定値（移動中）の特定

## 4. Sequence elimination procedure

活動系列候補数の限定

# Methodology 1. Probabilistic measurement model : a Bayesian approach

## Activity probability 活動の事後確率

: 測定値が  $\hat{m}_{1:J}$  であるとき実際の活動が  $a_{1:K}$  である確率

$$P(a_{1:K} | \hat{m}_{1:J}) \propto P(\hat{m}_{1:J} | a_{1:K}) \cdot P(a_{1:K}) \quad [\text{ベイズの定理}]$$

活動の事前確率

## Measurement likelihood 測定尤度

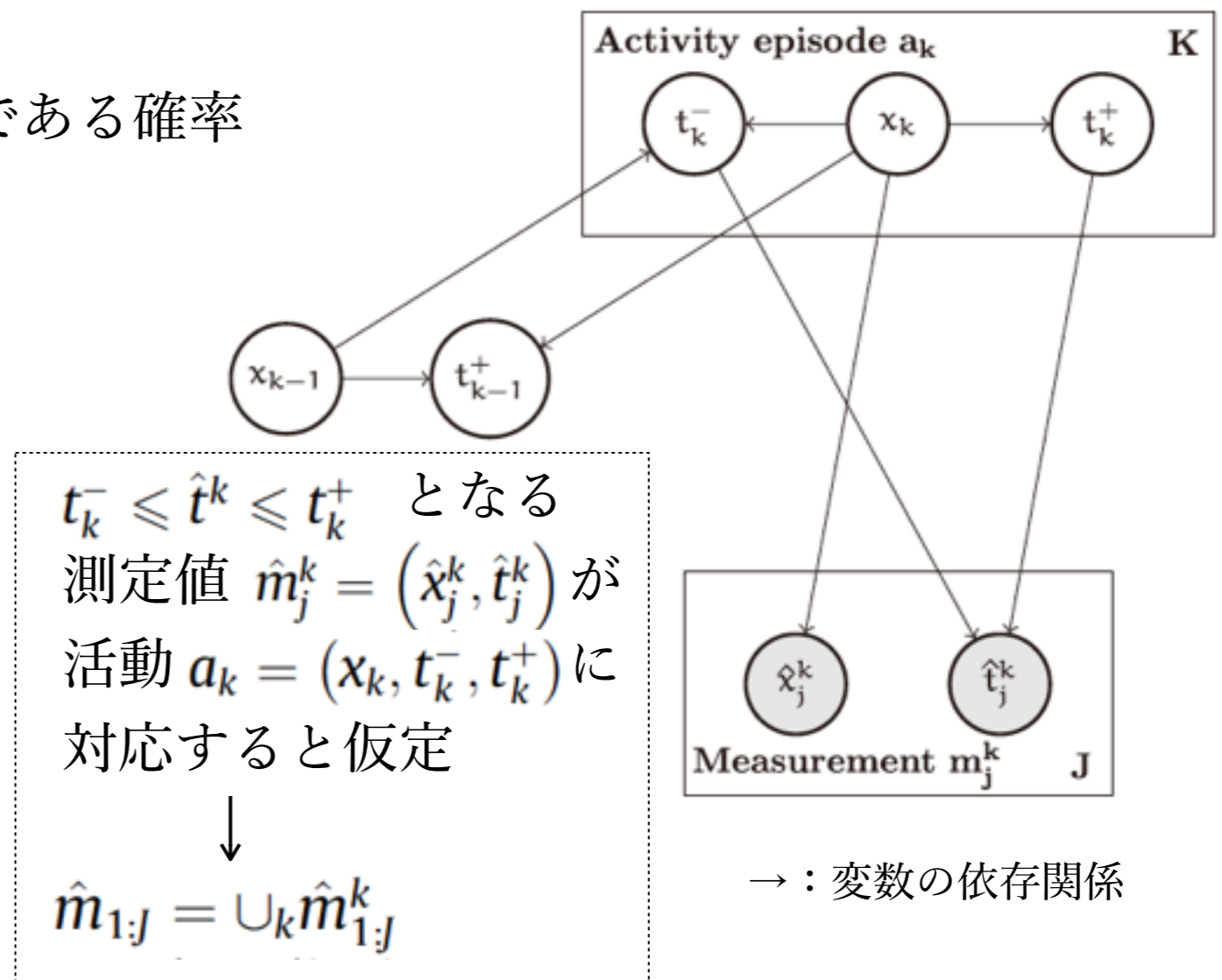
: 活動が  $a_{1:K}$  であるとき測定値が  $\hat{m}_{1:J}$  である確率

$$P(\hat{m}_{1:J} | a_{1:K}) = \prod_{k=1}^K P(\hat{m}_{1:J}^k | a_k) \quad (1)$$

$$= \prod_{k=1}^K \prod_{j=1}^J P(\hat{m}_j^k | a_k) \quad (2)$$

$$= \prod_{k=1}^K \prod_{j=1}^J P(\hat{x}_j^k | x_k) \quad (3)$$

- (1) 活動ごとの測定値の独立性を仮定
- (2) 測定値間の独立性を仮定
- (3) 測定時刻に誤差はないと仮定



# Methodology 1. Probabilistic measurement model : a Bayesian approach

## 活動の事前確率

### Potential attractivity measure

[時間帯ごとの場所の魅力度] により設定

$$\begin{aligned} P(a_{1:K}) &= \prod_{k=1}^K P(a_k) \\ &= \prod_{k=1}^K P(x_k, t_k^-, t_k^+) \\ &= \prod_{k=1}^K \frac{S_{x_k, i}(t_k^-, t_k^+)}{\sum_{x \in POI} S_{x, i}(t_k^-, t_k^+)} \end{aligned}$$

### Instantaneous potential attractivity measure

$$S_{x, i}(t^-, t^+) = \int_{t=t^-}^{t^+} S_{x, i}(t) dt$$

$$S_{x, i}(t) = \delta_{x, i}(t) \cdot \mathcal{A}_i(x, t)$$

時間制約に関する  
ダミー変数  
POIが開いていれば1,  
閉まっていれば0  
店の営業時間,  
会議や大学の時間割,  
公共交通の営業時間など

スーパーマーケットの  
勘定, メトロカードの  
利用データ, レストランの  
席数, オフィスの  
雇用者数など

### 事前確率特定上利用可能なデータの分類

- uniform 利用可能な情報がない場合
- aggregate 個人ごとの情報がない場合
- disaggregate 個人の特定されないスケジュール情報がある場合
- diary 個人ごとの情報がある場合

# Methodology 2. Generation of activity-episode sequences

## 1) Generating episode location

**DDR (domain of data relevance)**

: 測定した位置データが関連する  
空間範囲

Bierlaire and Frejinger(2008)

$P(\hat{x}|x) \geq \theta$  のとき

$x \in POI$  は測定値  $\hat{x}$  のDDRに

含まれるとする ( $\theta$ : 設定した閾値)

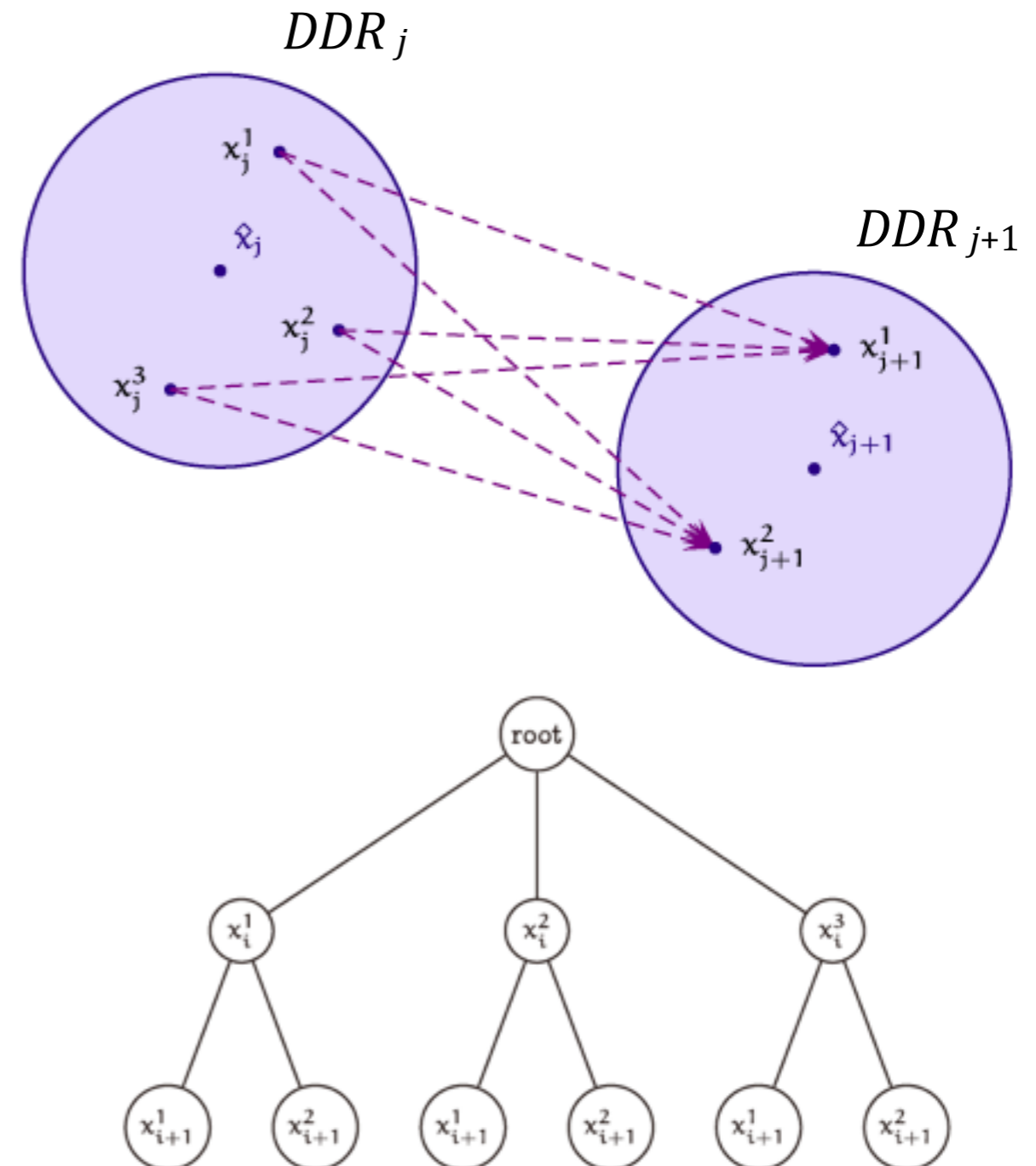
$$\text{測定尤度 } P(\hat{m}_{1:J}|a_{1:K}) = \prod_{k=1}^K \prod_{j=1}^J P(\hat{x}_j^k|x_k)$$

測定値  $\hat{m}_1, \hat{m}_2, \dots, \hat{m}_J$  に対して

DDR内の $x$ を抽出し時系列で関連づけ

→経路  $x_{1:J}$  を要素とする

長さ $J$ のネットワークを生成



# Methodology 2. Generation of activity-episode sequences

## 2) Generating episode start and end times

$tt_{x_j, x_{j+1}}$  ( $x_j$ から $x_{j+1}$ までの移動時間)  
 :  $x_j$ から $x_{j+1}$ 間の最短距離と  
 速度1.34m/sにより算出

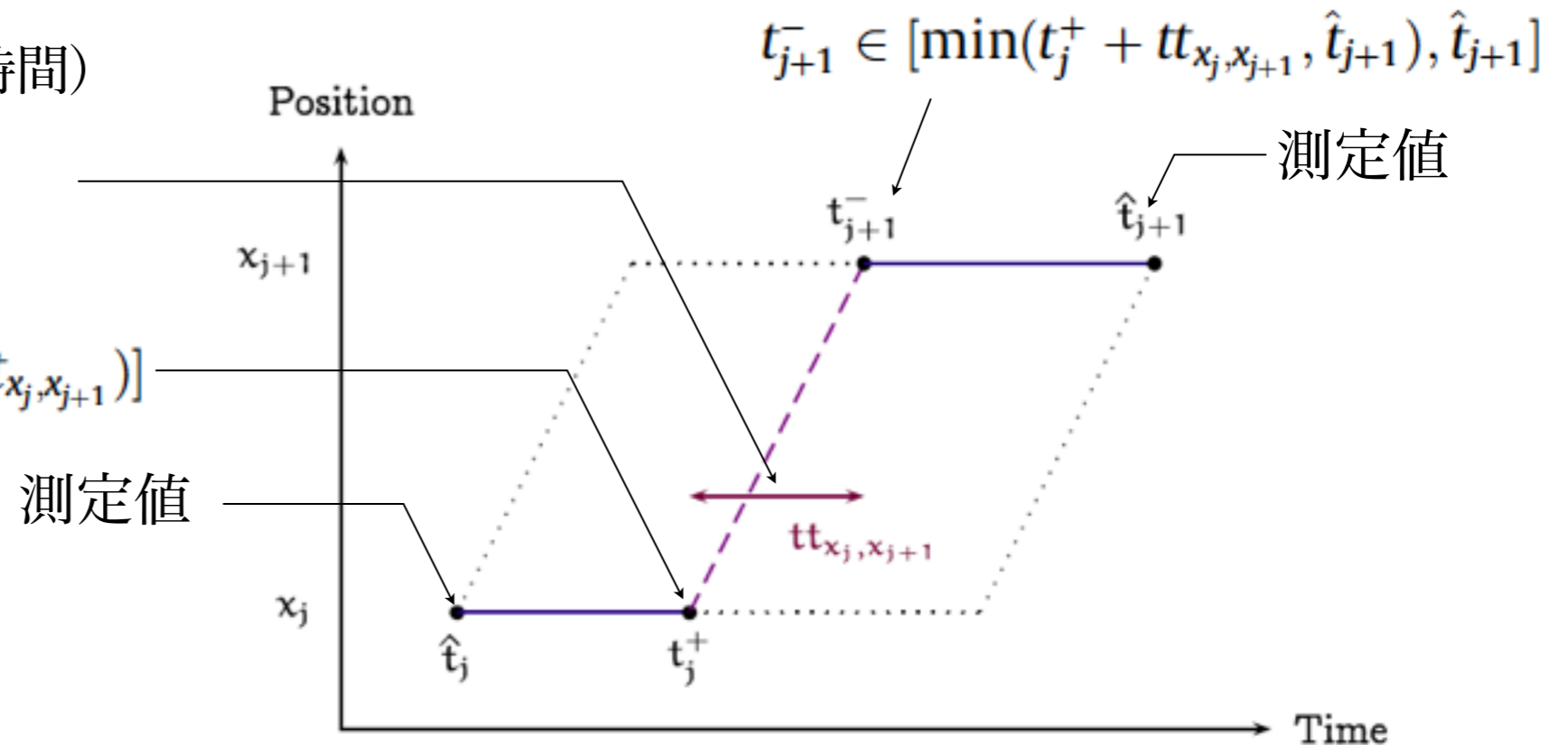
$$t_j^+ \in [\hat{t}_j, \max(\hat{t}_j, \hat{t}_{j+1} - tt_{x_j, x_{j+1}})]$$

$t_j^+ \sim U(\hat{t}_j, \hat{t}_{j+1} - tt_{x_j, x_{j+1}})$   
 一様分布を仮定

密度関数 :  $f(x) = \frac{1}{\hat{t}_{j+1} - tt_{x_j, x_{j+1}} - \hat{t}_j}$

$t_{j+1}^-$  :  $t_j^+ + tt_{x_j, x_{j+1}}$  と  $\hat{t}_{j+1}$  の間で一様分布を仮定

密度関数 :  $f(t_{j+1}^-) = \frac{1}{\hat{t}_{j+1} - tt_{x_j, x_{j+1}} - \hat{t}_j} \ln \frac{\hat{t}_{j+1} - tt_{x_j, x_{j+1}} - \hat{t}_j}{\hat{t}_{j+1} - t_{j+1}^-}$        $E(t_{j+1}^-) = \frac{\hat{t}_j + tt_{x_j, x_{j+1}} + 3 \cdot \hat{t}_{j+1}}{4}$





# Methodology 2. Generation of activity-episode sequences

参考  $f(t_{j+1}^-)$  の導出：

$t_i^+ \sim U(\hat{t}_i, \hat{t}_{i+1} - tt_{x_i, x_{i+1}})$ , with density function  $f_{t_i^+}(x) = \frac{1}{\hat{t}_{i+1} - tt_{x_i, x_{i+1}} - \hat{t}_i}$  for  $x \in [\hat{t}_i, \hat{t}_{i+1} - tt_{x_i, x_{i+1}}]$  and 0 otherwise

$t_{i+1}^-$  is uniformly distributed between  $t_i^+ + tt_{x_i, x_{i+1}}$  and  $\hat{t}_{i+1}$  with density function  $f_{t_{i+1}^- | t_i^+ = x}(y) = \frac{1}{\hat{t}_{i+1} - t_i^+ - tt_{x_i, x_{i+1}}}$  for  $y \in [x + tt_{x_i, x_{i+1}}, \hat{t}_{i+1}]$  and 0 otherwise

$$\begin{aligned}
 f_{t_{i+1}^-}(y) &= \int_{x=\hat{t}_i}^{\hat{t}_{i+1} - tt_{x_i, x_{i+1}}} f_{t_{i+1}^- | t_i^+ = x}(y) \cdot f_{t_i^+}(x) dx \\
 &= \int_{x=\hat{t}_i}^{y - tt_{x_i, x_{i+1}}} f_{t_{i+1}^- | t_i^+ = x}(y) \cdot f_{t_i^+}(x) dx \quad \left[ \begin{array}{l} y : y \in [x + tt_{x_i, x_{i+1}}, \hat{t}_{i+1}], \text{ i.e., } x + tt_{x_i, x_{i+1}} \leq y \\ \text{Note that } x \leq y - tt_{x_i, x_{i+1}} \leq \hat{t}_{i+1} - tt_{x_i, x_{i+1}} \end{array} \right. \\
 &= \int_{x=\hat{t}_i}^{y - tt_{x_i, x_{i+1}}} \frac{1}{\hat{t}_{i+1} - x - tt_{x_i, x_{i+1}}} \cdot \frac{1}{\hat{t}_{i+1} - tt_{x_i, x_{i+1}} - \hat{t}_i} dx \\
 &= \frac{1}{\hat{t}_{i+1} - tt_{x_i, x_{i+1}} - \hat{t}_i} \ln \left( \frac{\hat{t}_{i+1} - tt_{x_i, x_{i+1}} - \hat{t}_i}{\hat{t}_{i+1} - t_{i+1}^-} \right)
 \end{aligned}$$

## Intermediary measurements

活動継続時間の最低限度  $T_{min}$  を仮定  $\rightarrow$  継続時間が  $T_{min}$  未満のものは候補から削除  
 : 移動中と想定される測定値の削除

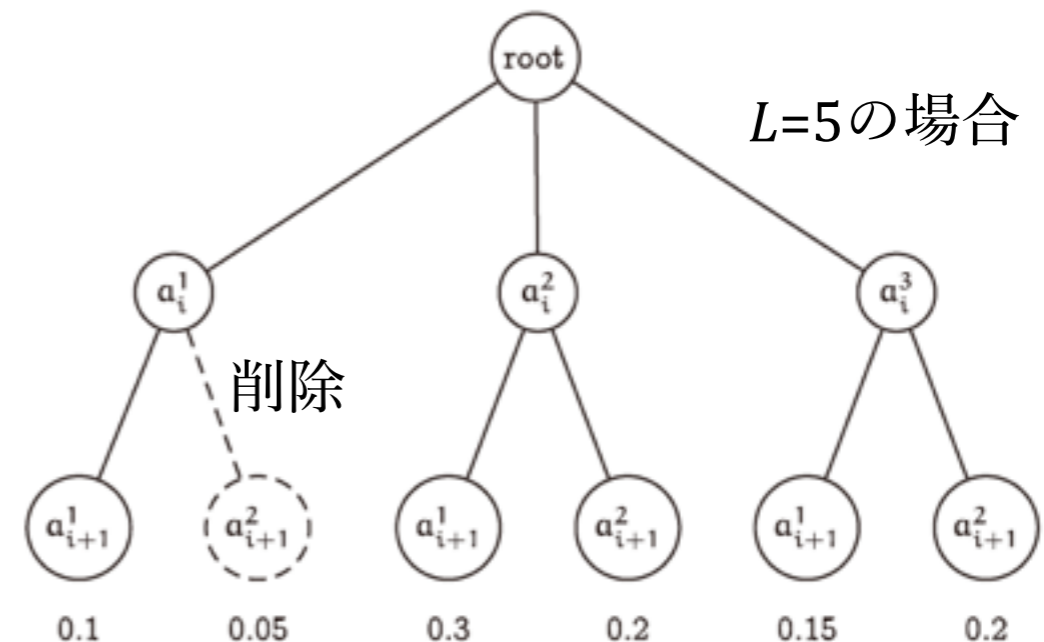
## Sequence elimination procedure

$\prod_{j=1}^J |DDR_j| \rightarrow$  候補内の全ての組み合わせを想定することは現実的でない

$\rightarrow$  候補数の上限  $L$  を設定

$$P(a_{1:K} | \hat{m}_{1:J}) \propto P(\hat{m}_{1:J} | a_{1:K}) \cdot P(a_{1:K})$$

に基づき最も確率の低いものから削除



# Methodology Algorithm

```

for each ID do
  for measurement  $\hat{m}_j = (\hat{x}_j, \hat{t}_j)$ ,  $j = 1, \dots, J$   $O(J)$  do
    Define the corresponding Domain of Data Relevance,  $DDR_j$   $O(|DDR| \log(|N|))$ ;
    for each  $x \in DDR_j$   $O(|DDR|)$  do
      Compute the measurement likelihood ;
      if  $T$  empty then
        Initialize the network structure for activity-episodes sequences  $T$  with  $x_1 = x$ ,
         $t_1^- = t_1^+ = \hat{t}_1$  ;
        Update the activity probability with the measurement likelihood and the prior ;
      else
        for each  $a_{1:k}$  of  $T$   $O(L)$  do
          if  $x_k = x$  then
            Update the definition of the episode end time:  $t_k^+ = \hat{t}_j$  ;
            Update the prior for  $a_k$  ;
            Update the activity probability of  $a_{1:k}$  with the measurement
            likelihood and the prior ;
          else
            Compute the shortest path between  $x_k$  and  $x$ , and the travel time
             $tt_{x_k, x}$   $O((|E| + |N|) \log(|N|))$ ;
            Define the last episode end time:  $t_k^+ \sim U(\hat{t}_{j-1}, \hat{t}_j - tt_{x_k, x})$  ;
            Define the new episode start time:  $t_{k+1}^- \sim U(\hat{t}_{j-1} + tt_{x_k, x}, \hat{t}_j)$  ;
             $a_{k+1} = (x, t_{k+1}^-, \hat{t}_j)$  ;
            if  $a_k$  is an intermediary measurement then
              Connect  $a_{k-1}$  with  $a_{k+1}$  in  $T$  ;
              Compute the prior for  $a_{k+1}$  ;
              Update the activity probability of  $a_{1:k+1}$  with the new
              measurement likelihood and prior, but without the prior for  $a_k$  ;
            else
              Connect  $a_k$  with  $a_{k+1}$  in  $T$  ;
              Update the prior for  $a_k$  and compute it for  $a_{k+1}$  ;
              Update the activity probability of  $a_{1:k+1}$  with the new
              measurement likelihood and priors ;
        Sequence elimination procedure: keep the  $L$  most likely paths of the network  $T$ 
         $O(L|DDR| \log(L|DDR|))$ 

```

測定値  $\hat{x}_j$  のDDRに  
 含まれる  $x$  の抽出  
 ↓  
 各  $x$  に対する測定尤度の計算  
 ↓  
 各  $a_{1:k}$  に対して  
 $x_k = x$  のとき  $a_k$  に対する  $t_k^+$ ,  
 事前確率, 活動の尤度を更新  
  
 $x_k \neq x$  のとき  $x_k - x$  間の最短経路,  
 移動時間  $tt_{x_k, x}$  の計算  
 ↓  
 前回活動終了時刻  $t_k^+$  と  
 今回活動開始時刻  $t_{k+1}^-$  の設定  
 ↓  
 $a_{k+1} = (x, t_{k+1}^-, \hat{t}_j)$  の設定  
 ↓  
 $a_{1:k}$  に関する intermediary  
 measurement のチェック  
 ↓  
 活動間のリンクを設定  
 ↓  
 事前確率の更新  
 活動の尤度の更新  
 ↓  
 Sequence elimination  
 procedure:  $L$  個に絞り込み

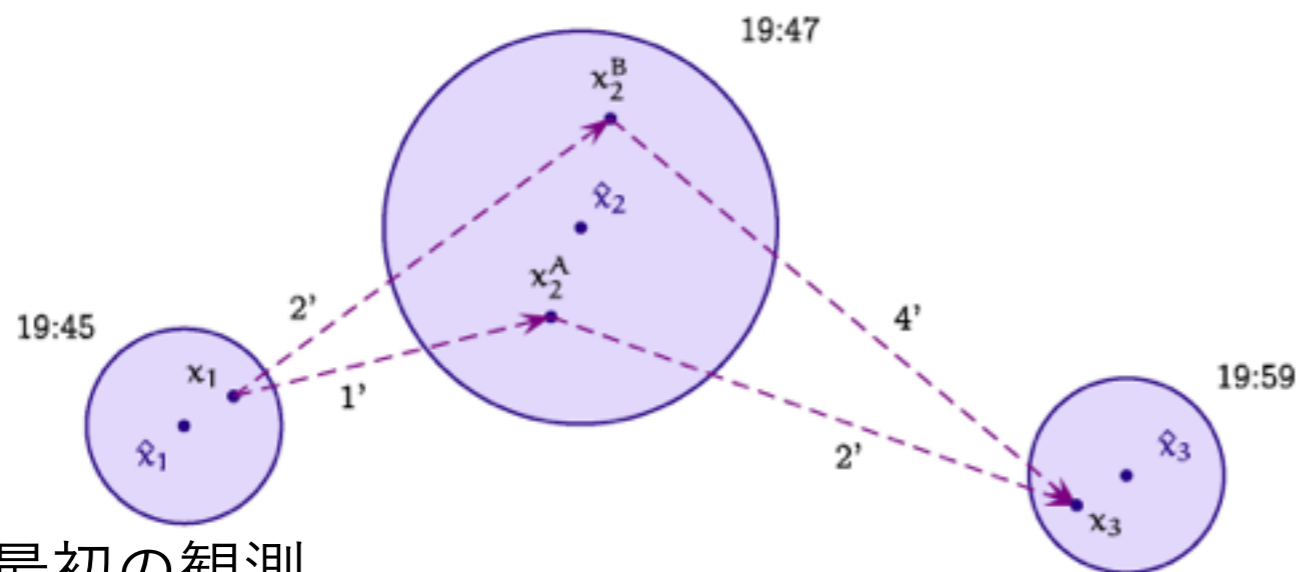
# Methodology

## 2番めの観測

DDRに2つの候補 $(a_1, a^{A_2}), (a_1, a^{B_2})$

- $a^{A_2}$ の魅力度が $a^{B_2}$ の2倍
- $x^{A_2}$ と $x^{B_2}$ から等距離とする

→ $(a_1, a^{A_2})$ の尤度は $(a_1, a^{B_2})$ の2倍



## 最初の観測

DDRに  
1個のPOIを含む

## 3番めの観測

DDRに  
1個のPOIを含む

L=1の場合

$(a_1, a^{A_2})$ のみ保持

$a^{A_2}$ での活動開始, 終了時刻

$$t_2^- \sim U(19:45 + 1', 19:47)$$

$$t_2^+ \sim U(19:47, 19:59 - 2')$$

→ $a^{A_2}$ での期待滞在時間 5'30

L=2の場合

$(a_1, a^{A_2}), (a_1, a^{B_2})$ とも保持

$a^{B_2}$ での活動開始, 終了時刻

$$t_2^- \sim U(19:45 + 2', 19:47)$$

$$t_2^+ \sim U(19:47, 19:59 - 4')$$

→ $a^{B_2}$ での期待滞在時間 4'

→5分( $T_{min}$ )未満のため削除

↳ 候補  $(a_1, a^{A_2}, a_3), (a_1, a_3)$

測定尤度が同等で

priorが  $P(a_1, a_3) > P(a_1, a^{A_2}, a_3)$

→ $(a_1, a_3)$ の確率が最も高い

# Case study on EPFL campus

## EPFL WiFi data

- 緯度方向，経度方向の誤差はそれぞれ独立な正規分布であると仮定
- 緯度方向と経度方向の誤差は独立であると仮定

$$P(\hat{x}|x) = P(\hat{x}_{lat}|x_{lat}) \cdot P(\hat{x}_{long}|x_{long})$$

$$P(\hat{x}_{lat}|x_{lat}) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(\hat{x}_{lat} - x_{lat})^2}{2\sigma^2}\right)$$

$$P(\hat{x}_{long}|x_{long}) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(\hat{x}_{long} - x_{long})^2}{2\sigma^2}\right)$$

$$\sigma = \frac{cF}{2}$$

→これをもとにDDRを計算

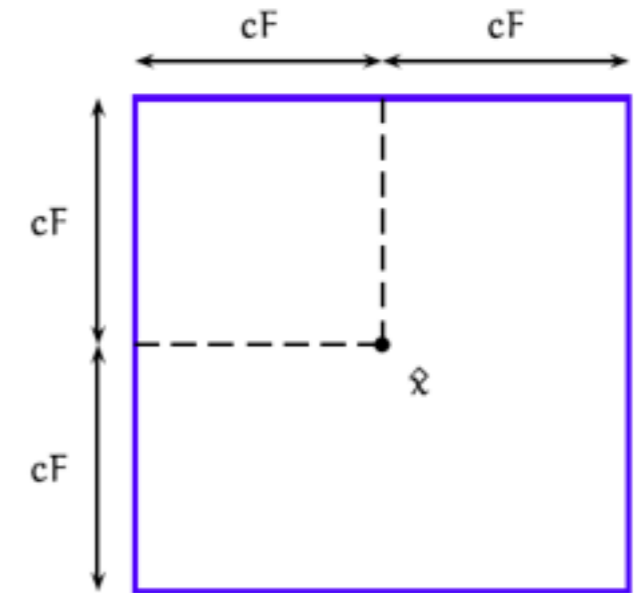
- 階層方向の測定精度

F：検知された階にいる確率

(1-F)/2：検知された階の上階または下階の

2r×2rの正方形の範囲内にいる確率

※r,Fは分析者が設定



cF: confidence factor  
95% probability

# Case study on EPFL campus

## EPFL pedestrian network

$x \in POI$

カテゴリ

- offices
- classrooms
- restaurants
- shops
- library
- その他



# Case study on EPFL campus

## EPFL pedestrian network

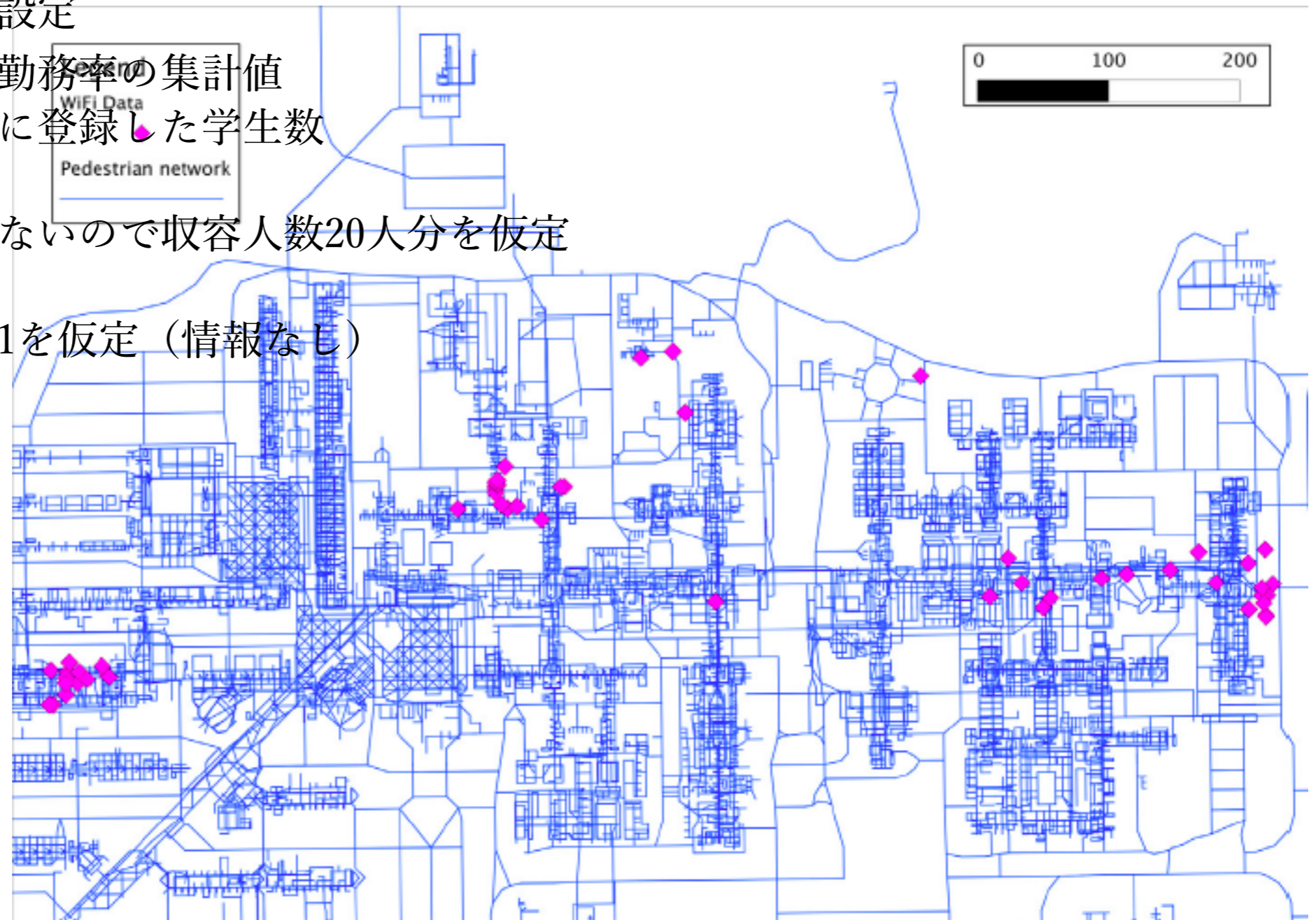
$x \in POI$

$A(x, t)$

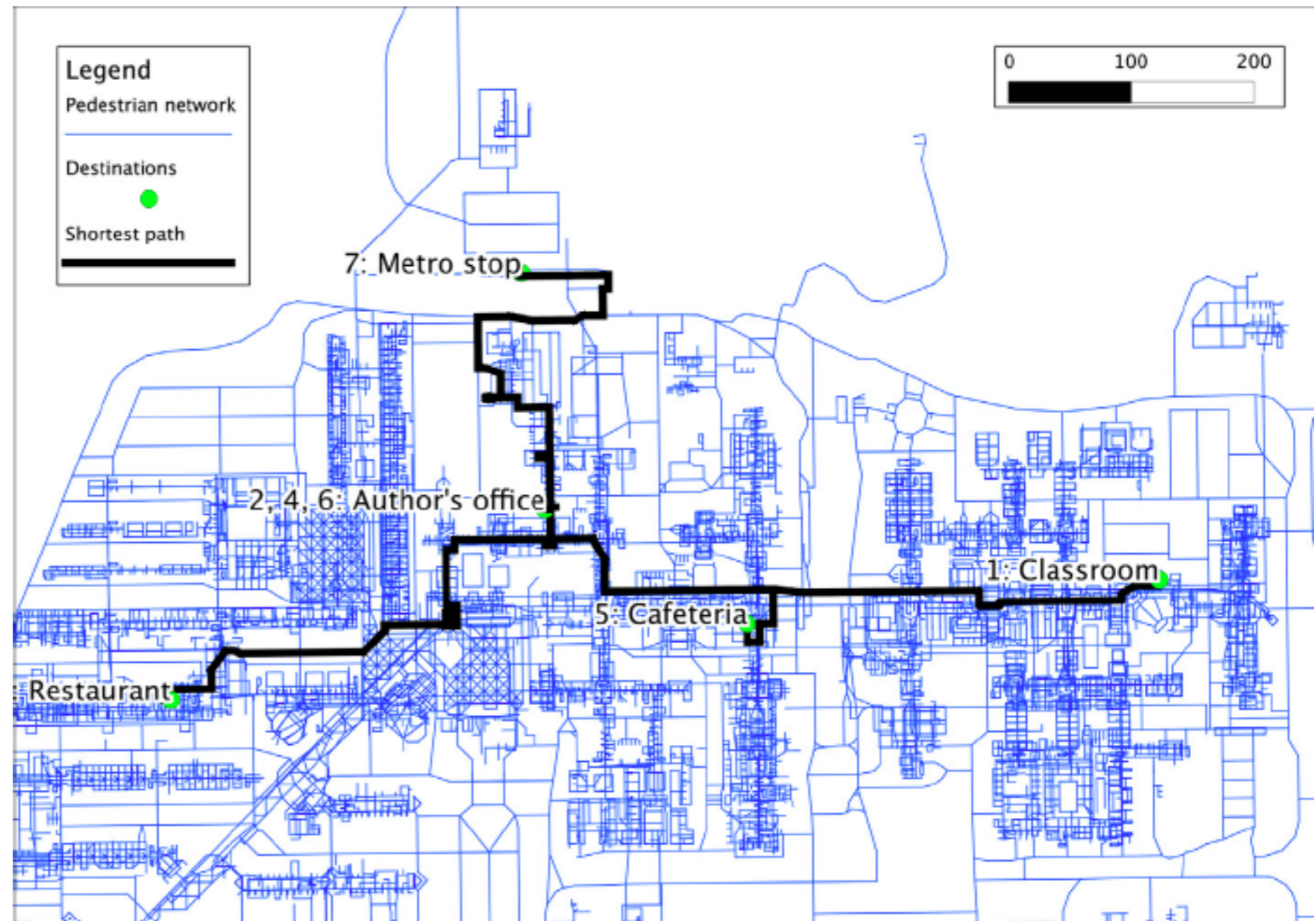
カテゴリ

魅力度の設定

- offices : 職員の勤務率の集計値
- classrooms : コースに登録した学生数
- restaurants : 席数
- shops : 情報がないので収容人数20人分を仮定
- library : 席数
- その他 : 魅力度1を仮定 (情報なし)



# Case study on EPFL campus



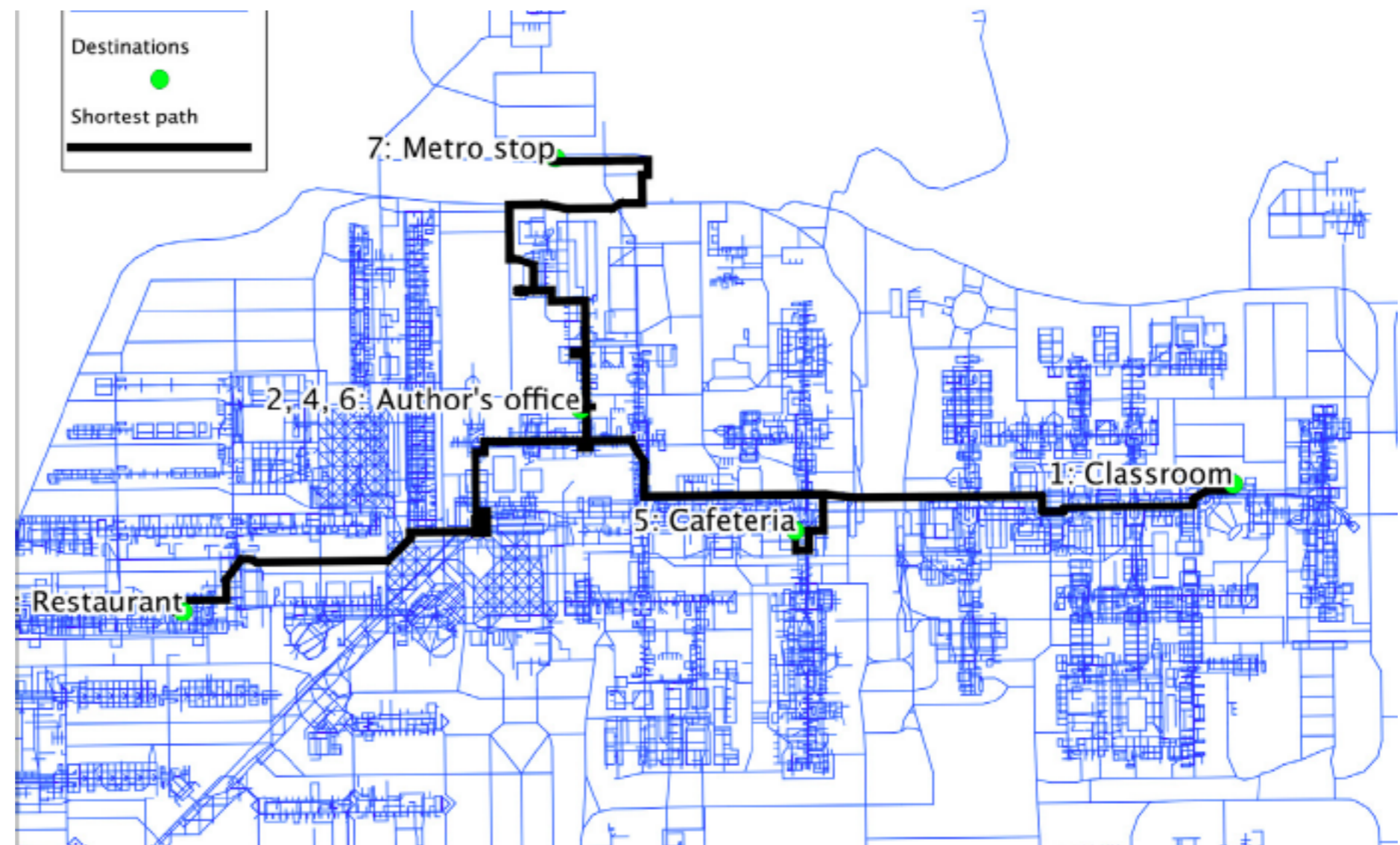
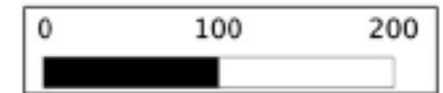


# Case study on EPFL campus

**Table 1**

Sequence of activity episodes as reported by one author. It contains 6 activity episodes.

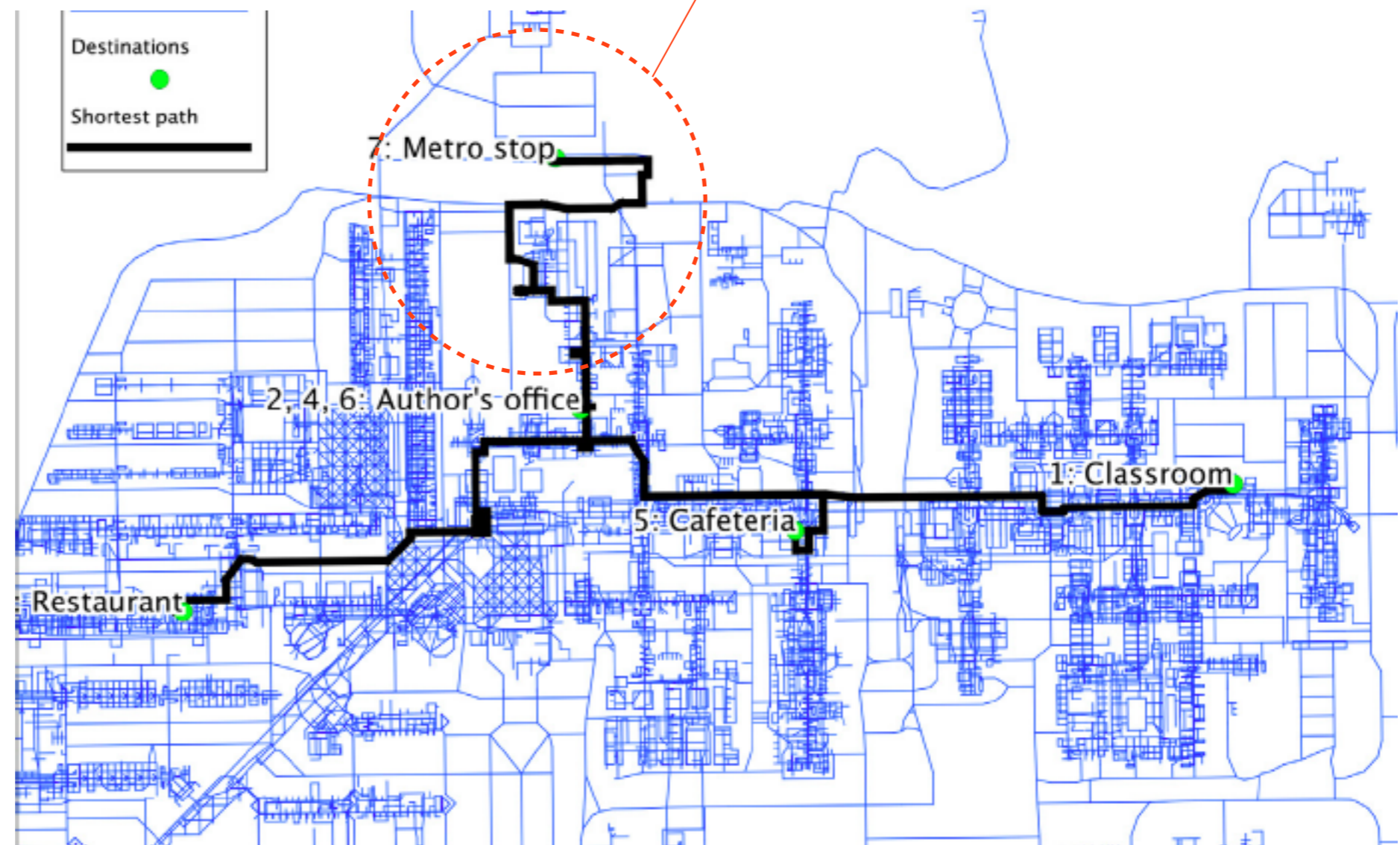
Time spent	Floor	Location
<i>Activity log</i>		
8.32 am–10.30 am	1	Classroom
Until 11.47 am	3	Author's office
From 11.55 am	1	Restaurant
Around 1 pm	3	Author's office
Around 2 pm	2	Cafeteria
Until around 7.45 pm	3	Author's office



# Case study on EPFL campus

**Table 1**  
Sequence of activity episodes as reported by one author. It contains 6 activity episodes.

Time spent	Floor	Location
<i>Activity log</i>		
8.32 am–10.30 am	1	Classroom
Until 11.47 am	3	Author's office
From 11.55 am	1	Restaurant
Around 1 pm	3	Author's office
Around 2 pm	2	Cafeteria
Until around 7.45 pm	3	Author's office

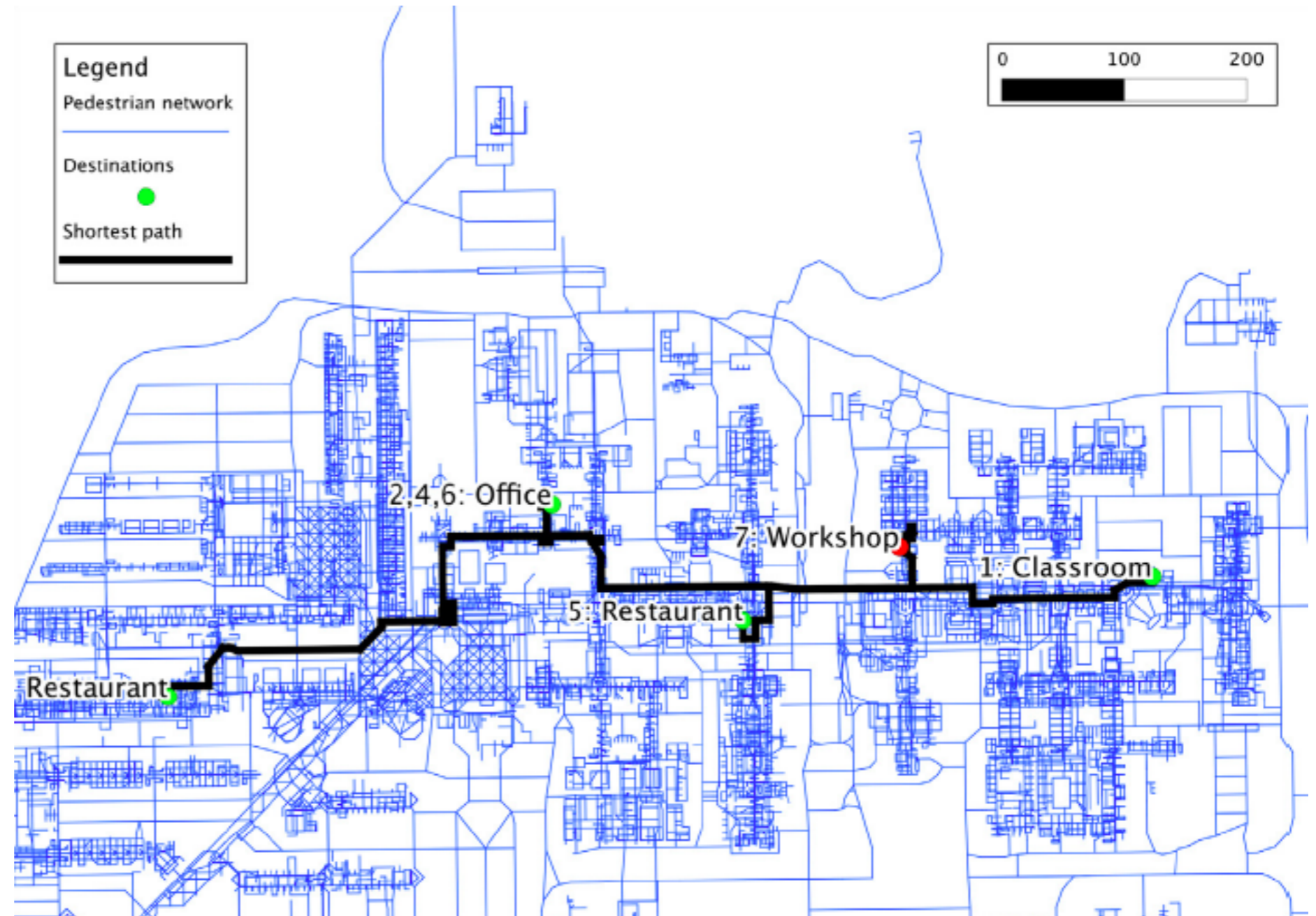


# Case study on EPFL campus

モデルと実行動の差(最短経路長さ)



$L=1$ の場合



# Case study on EPFL campus

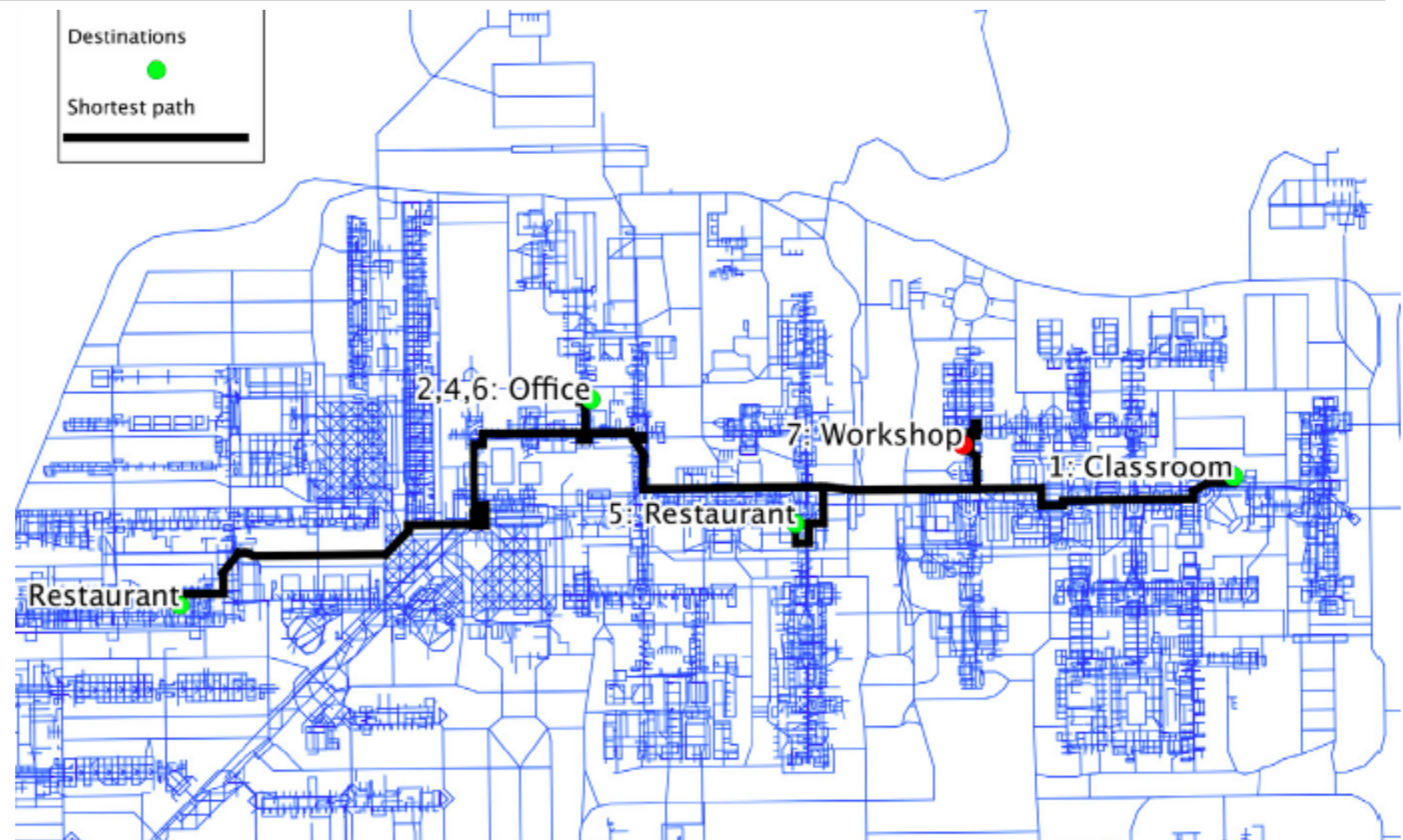
**Table 2**

Comparison between the most likely output of the model and the activity log as reported by one author.

モデルと実行動の差(最短経路長さ)

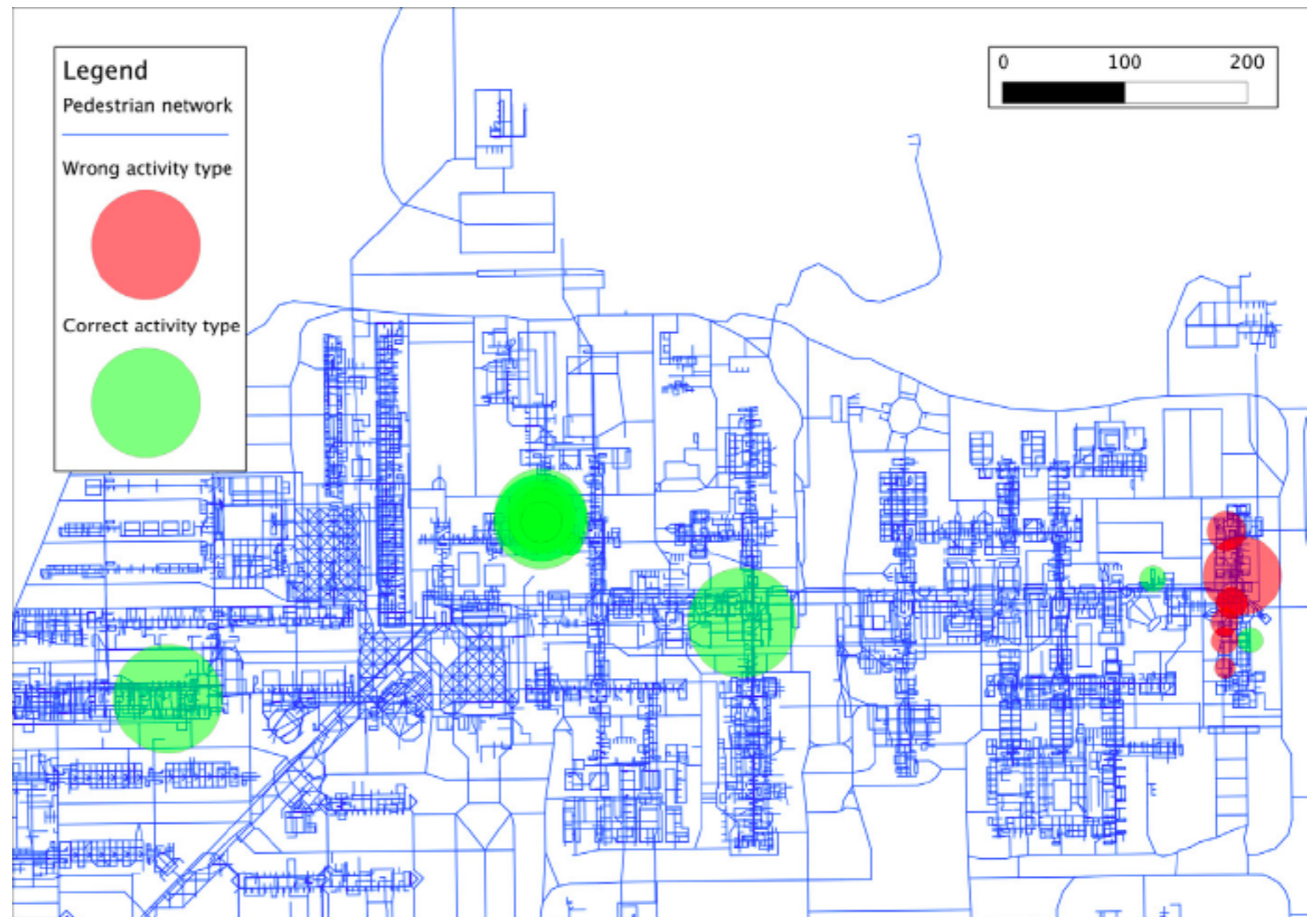
Model with disaggregate prior				Activity log			$\Delta x$
Arrival time	Departure time	Floor	Location	Time spent	Floor	Location	(in m)
8:35–8:35	10:38–10:38	1	Classroom	8.32 am–10.30 am	1	Classroom	0
10:40–10:40	11:51–11:51	3	Office	Until 11.47 am	3	Author's office	9
11:54–11:54	12:47–12:53	1	Restaurant	From 11.55 am	1	Restaurant	0
12:51–12:58	13:03–13:44	3	Office	Around 1 pm	3	Author's office	9
13:06–13:47	13:53–14:02	2	Cafeteria	Around 2 pm	2	Cafeteria	0
13:55–14:04	19:40–19:44	3	Office	Until around 7.45 pm	3	Author's office	9

$L=1$ の場合



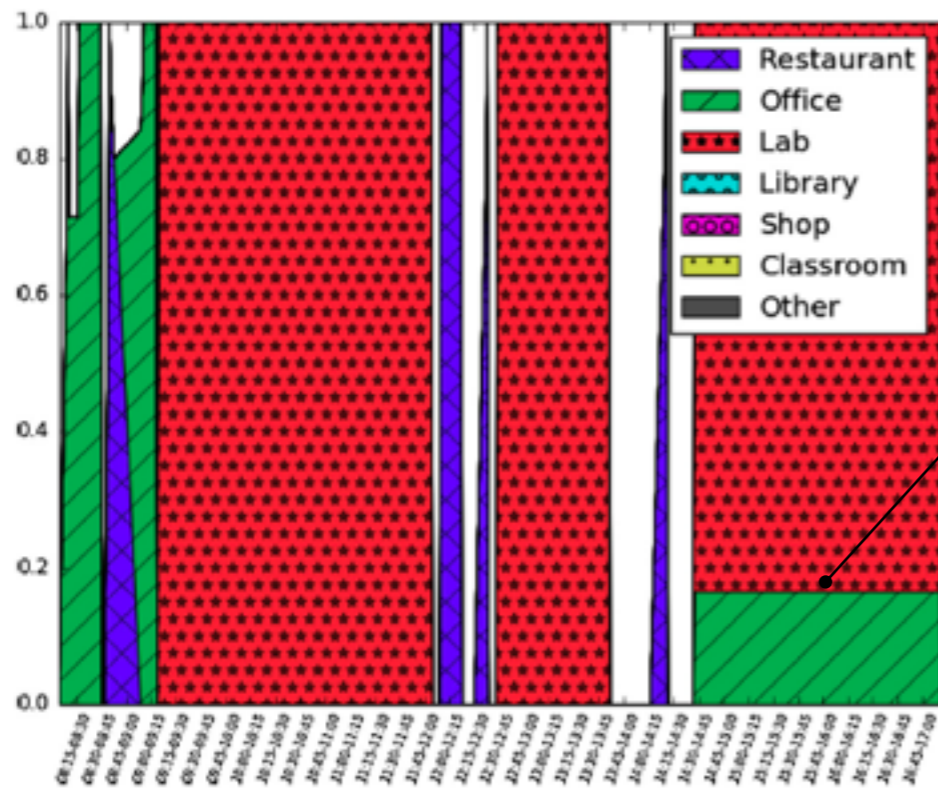
# Case study on EPFL campus

$L=100$ の場合



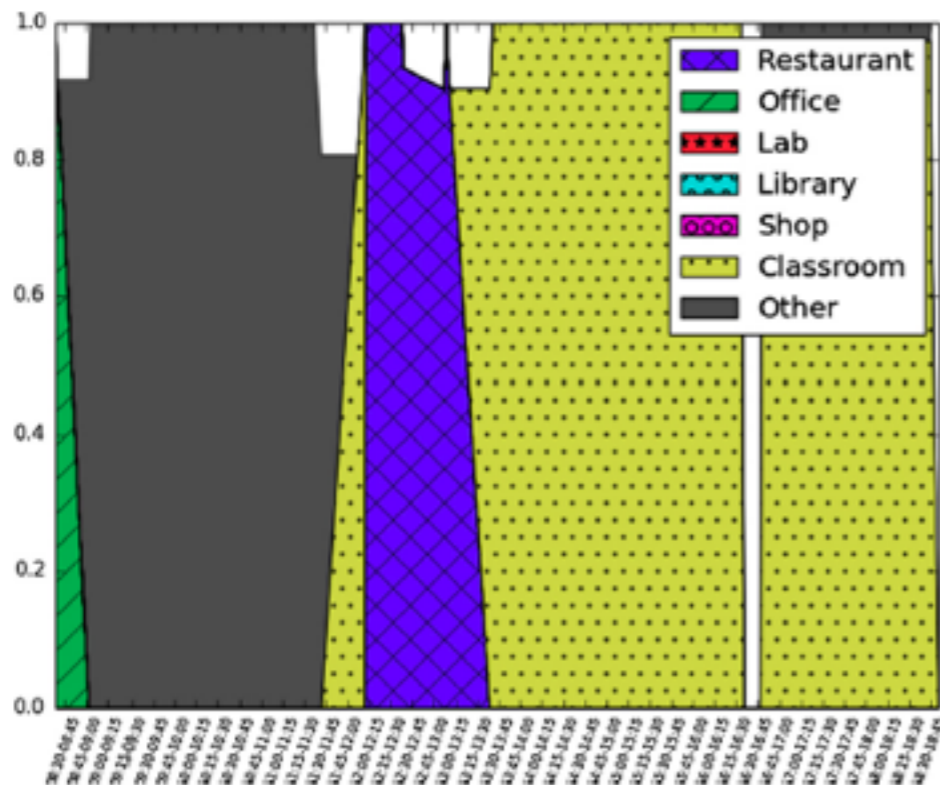
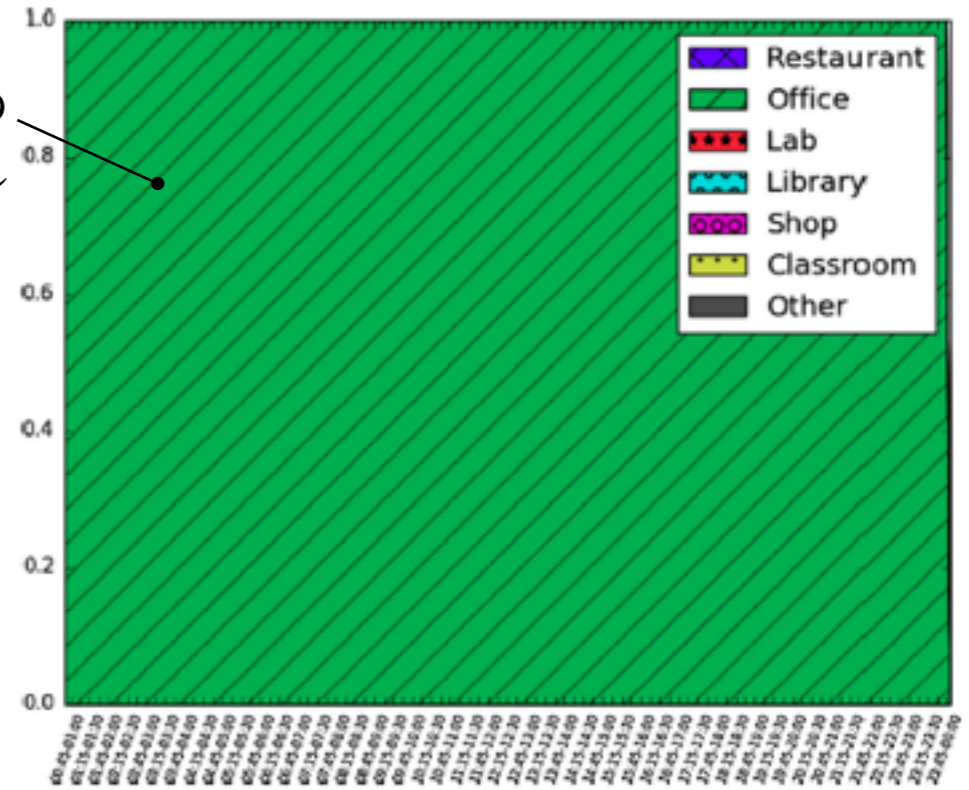
# Case study on EPFL campus

職員3490人+学生767人に適用  $L=20, F=0.99$  5平日

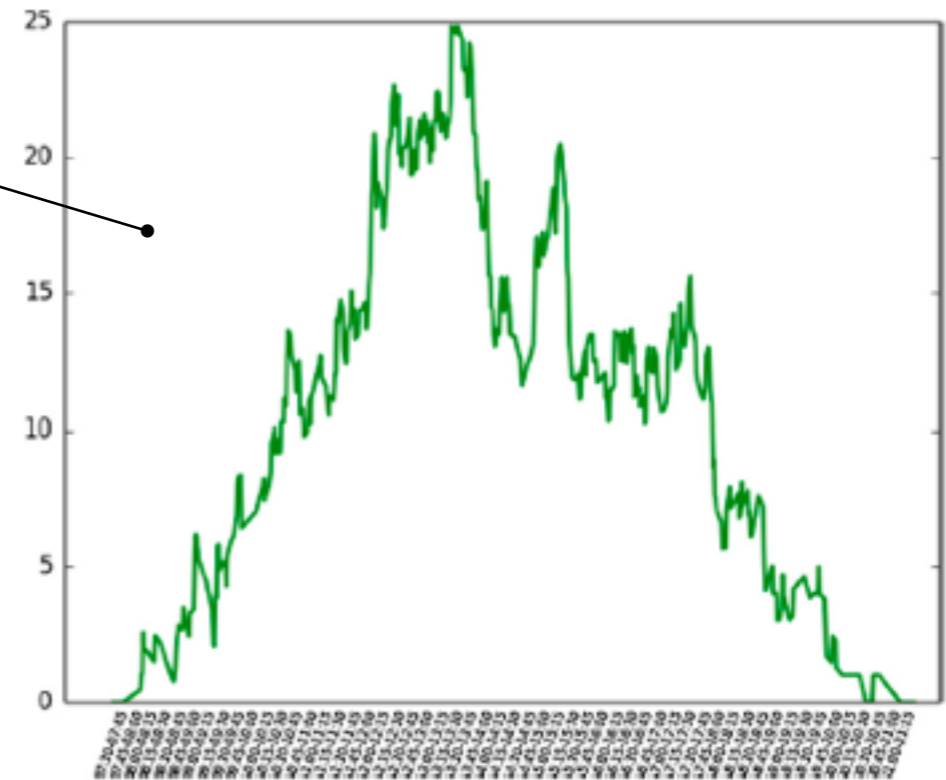


Lab 80%,  
office 20%  
の確率

device の  
移動なし



レストランで  
検出された  
時間ごとの  
累積人数



# Case study on EPFL campus

## Sensitivity to the parameters

設定すべきパラメータ

$R=80\text{m}$  :DDRの最大半径

$F=1.0$  :検出された階層にいる確率

$T_{min}=5\text{min}$  :目的地での最小滞在時間

$L=40$  :最大候補数

### 最大候補数 $L$

$L=1$ は避けるべき

:最も確度の高いものしか残さず測定のあいまいさが反映されない



Fig. 13. Sensitivity to the number  $L$  of candidates kept between each measurement,  $L = 1, 2, 3, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100$ .

# Case study on EPFL campus

## 設定すべきパラメータ

$R=80\text{m}$  :DDRの最大半径  
 $F=1.0$  :検出された階層にいる確率  
 $T_{min}=5\text{min}$  :目的地での最小滞在時間  
 $L=40$  :最大候補数

## 最小滞在時間 $T_{min}$

設定時間が長いと  
検出されないものが  
生じる

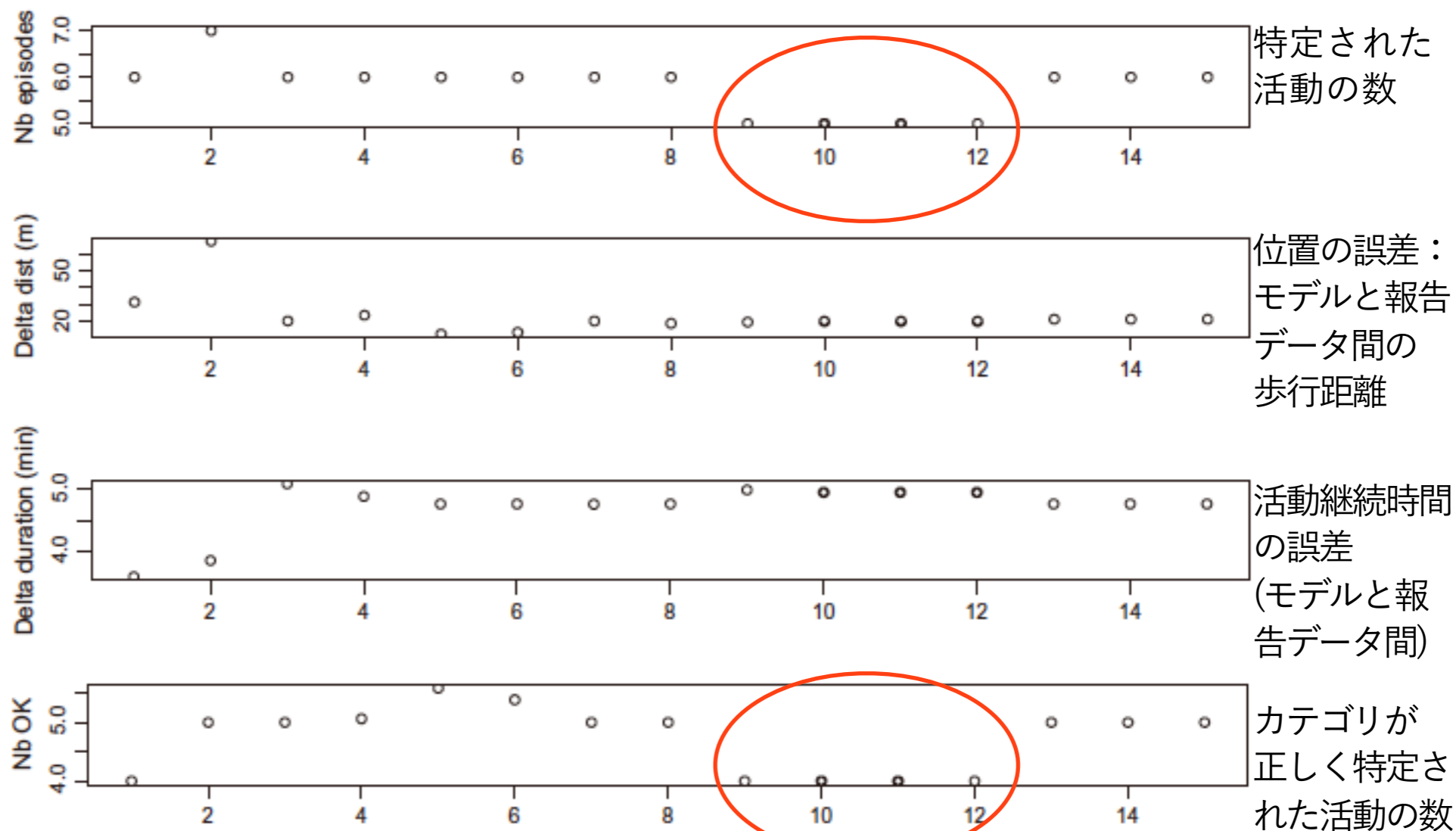
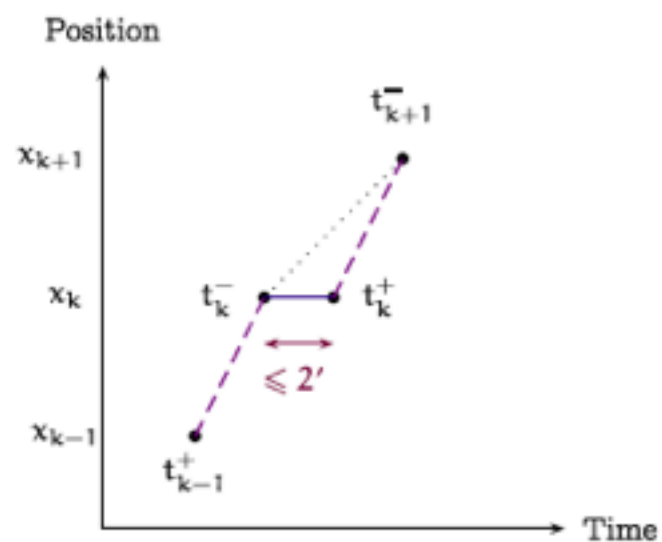


Fig. 14. Sensitivity to the minimum time spent at destination  $T_{min}$ ,  $T_{min} = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15$  min.



# Case study on EPFL campus

## 設定すべきパラメータ

$R=80\text{m}$  :DDRの最大半径  
 $F=1.0$  :検出された階層にいる確率  
 $T_{min}=5\text{min}$  :目的地での最小滞在時間  
 $L=40$  :最大候補数

## DDRの最小半径 $R$

- 一般的には $R$ が小さいと目的地が検出されない
- $R$ が大きいとWiFiから提供される地理情報がフラットになり事前確率の影響が非常に大きくなる
- $R=60,70,80$ 付近が安定

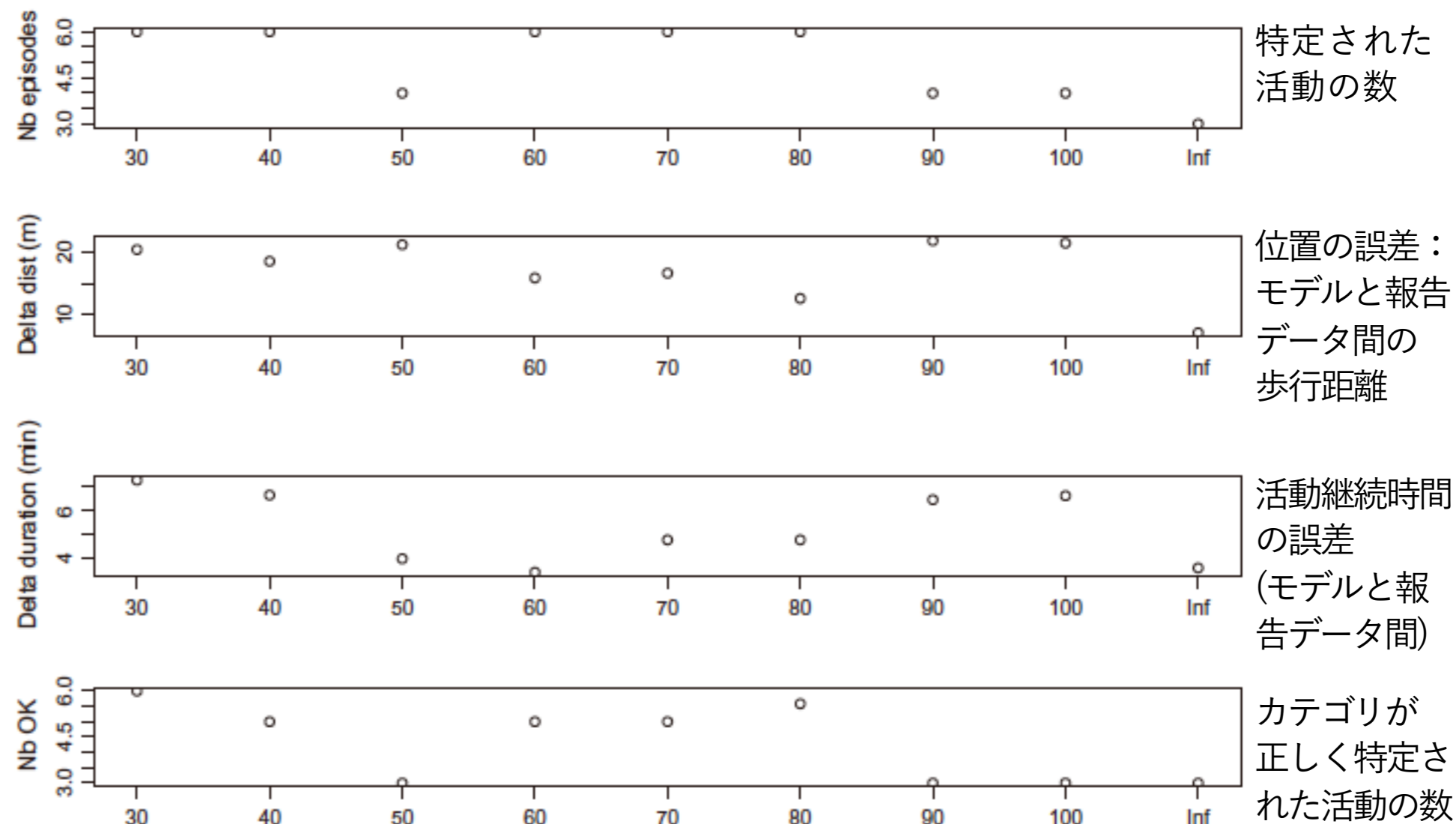


Fig. 16. Sensitivity to the maximum radius  $R$  of the DDR,  $R = 30, 40, 50, 60, 70, 80, 90, 100$  m and  $R = \infty$ .

# Case study on EPFL campus

設定すべきパラメータ

$R=80\text{m}$  :DDRの最大半径

$F=1.0$  :検出された階層にいる確率

$T_{min}=5\text{min}$  :目的地での最小滞在時間

$L=40$  :最大候補数

検出された階層に  
いる確率  $F$

- このケースでは0.9以上の  
場合には  
影響が少ない

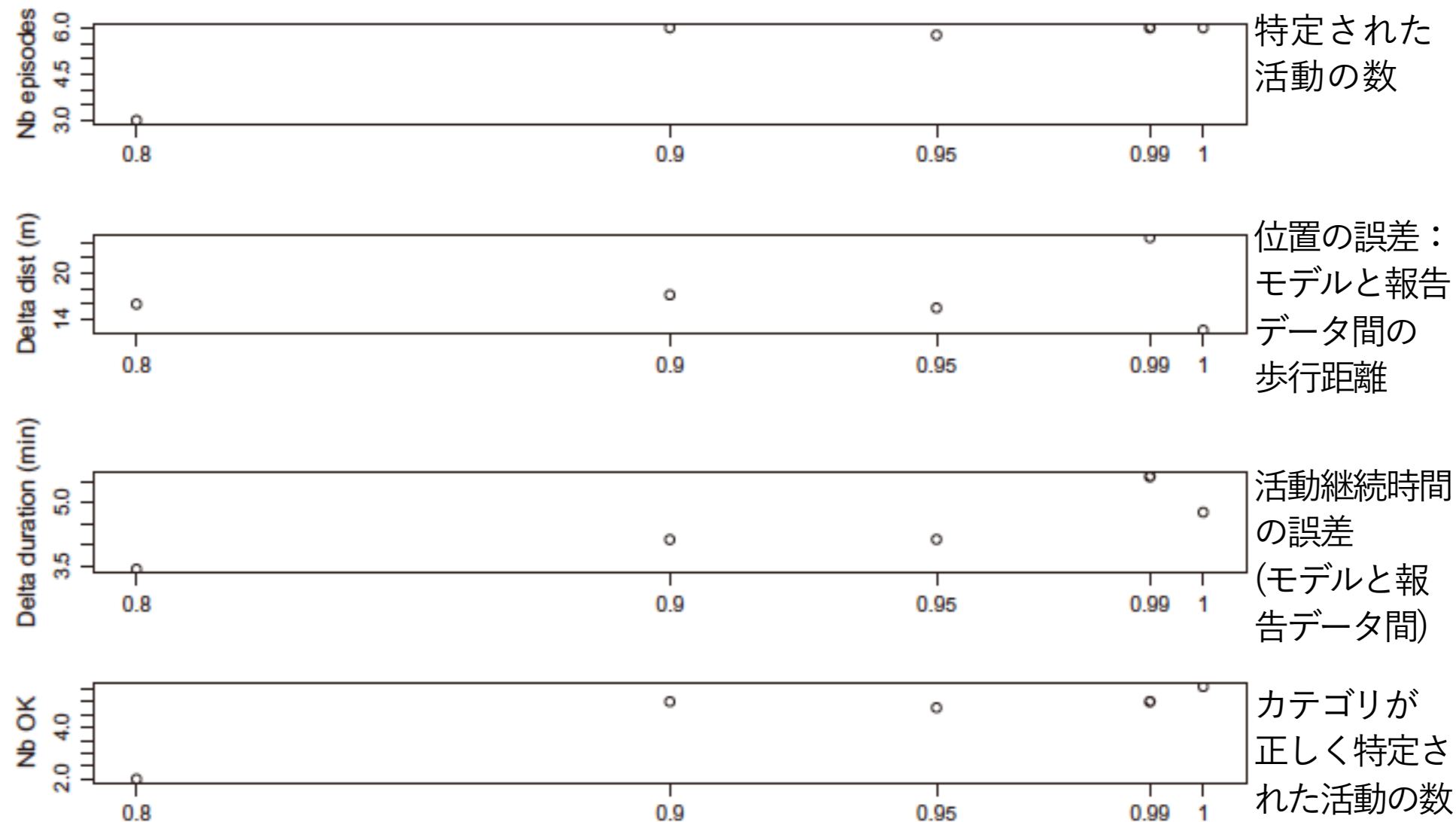


Fig. 17. Sensitivity to the probability of being in the detected floor  $F$ , and not in the upper or lower floor,  $F = 0.8, 0.9, 0.95, 0.99, 1.0$ .  $F = 1.0$  means that only the detected floor is considered.

# Case study on EPFL campus

## Sensitivity to the prior

- uniform 利用可能な情報がない場合
- aggregate 個人ごとの情報がない場合
- disaggregate 個人の特定されないスケジュール情報がある場合
- diary 個人ごとの情報がある場合

- WiFiのデータのみで活動数は正しく特定 (uniform priorでも特定)
- aggregate priorでは情報として不十分 : 対象人数の多い目的地 (教室など) の影響が強く現れバイアスを生む
- 個人を特定しなくてもスケジュールが特定できれば精度の高い結果が得られる

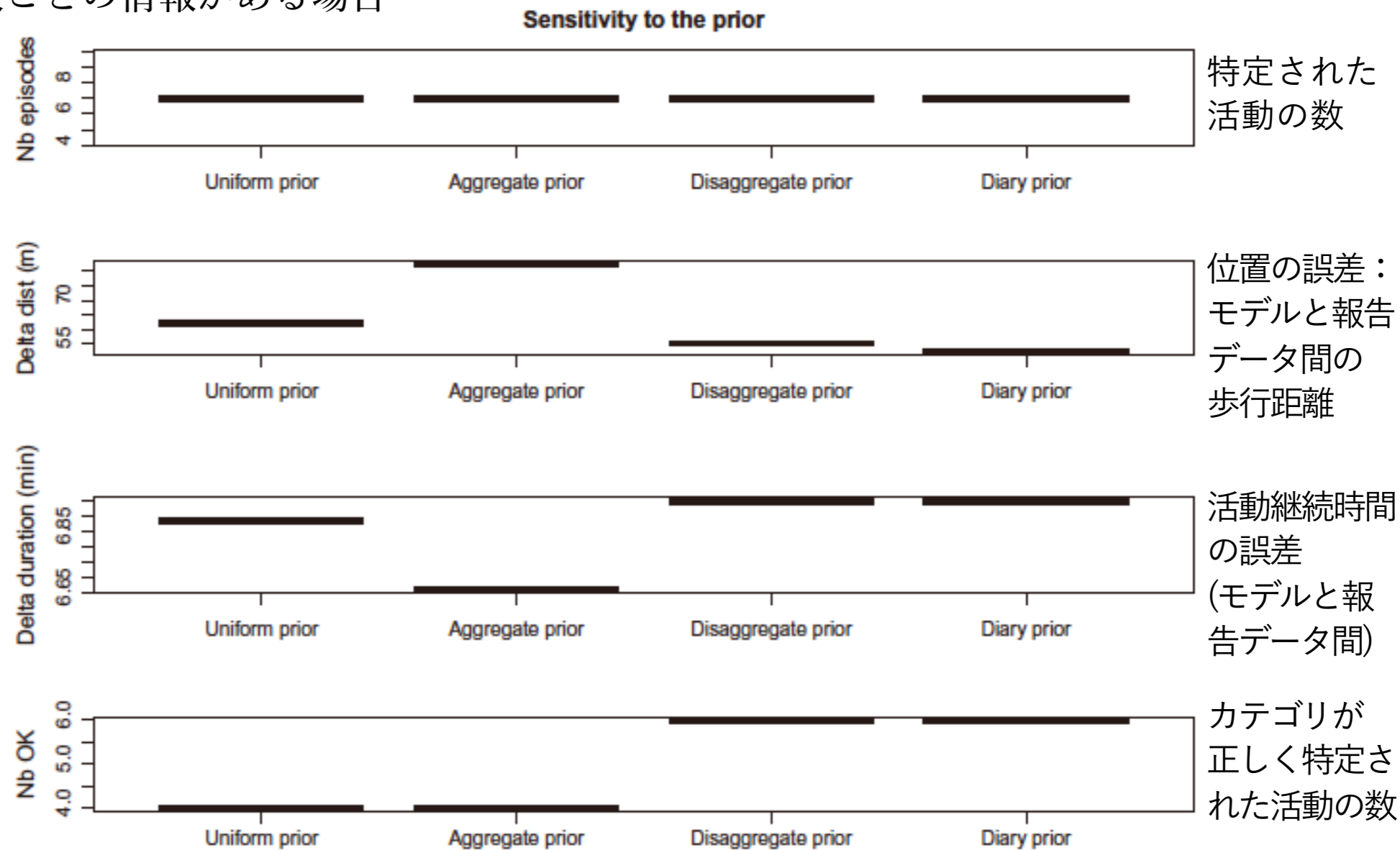


Fig. 18. Sensitivity to the prior, with uniform, aggregate, disaggregate and diary prior.

# Case study on EPFL campus

## Sensitivity to the density of measurements

- データの一部分をランダムに削除  
(各ケースについて100回試行)

- 15%までは安定

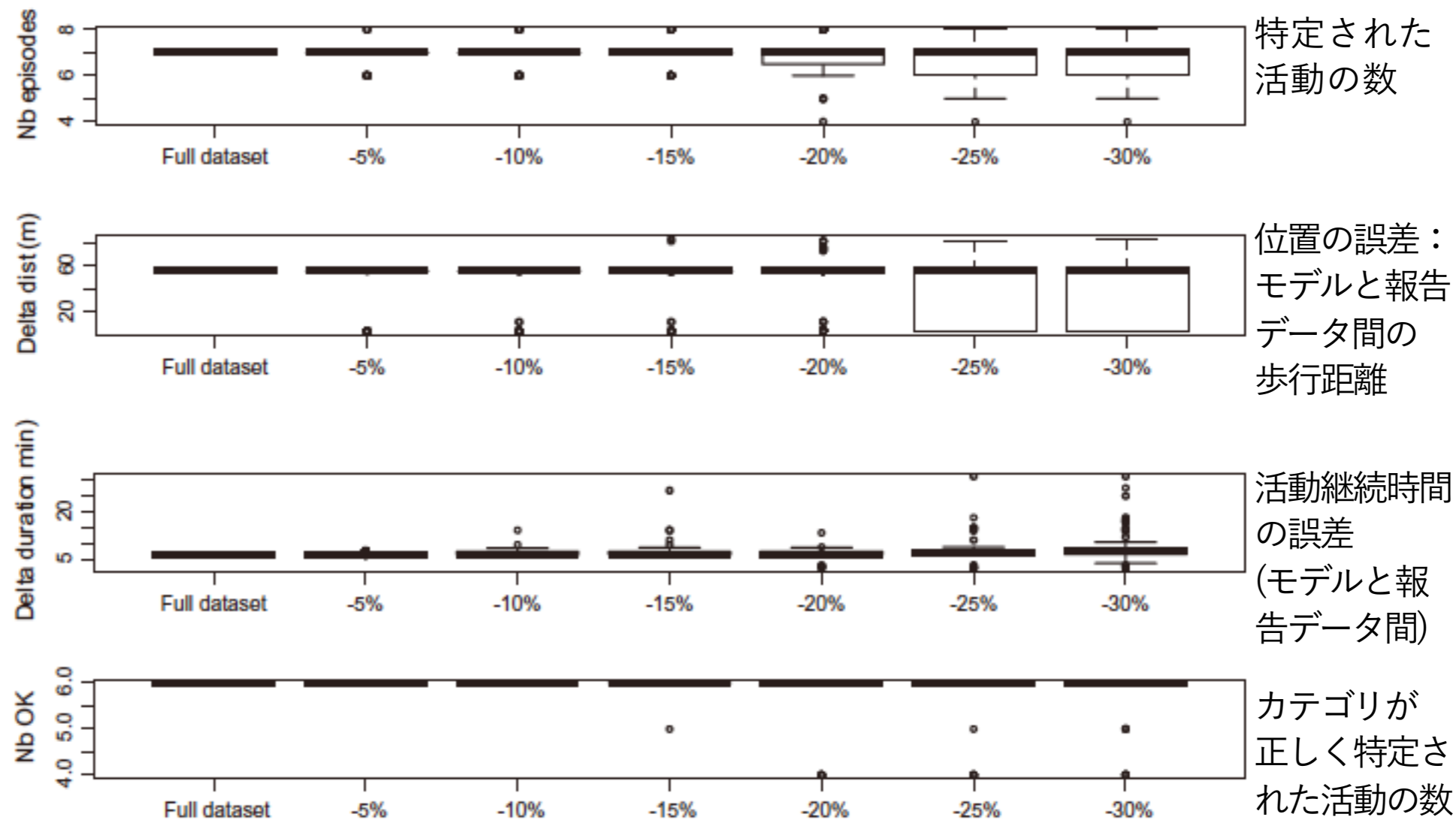


Fig. 19. Sensitivity to datasets with less data, with the full dataset as a base case.

# Conclusion

## 測定データから活動場所を検出する手法を提案

- ・ ネットワークが密で位置データが疎な場合が多い歩行者行動を考慮し、柔軟で調整しやすい方法を提案
- ・ 活動に関する事前情報と歩行者の地図情報を利用
- ・ 時間に関する制約条件を考慮
- ・ DDRを用いてpingpong effectピンポン効果を回避→疎な測定結果に有効
- ・ 測定のあいまいさを個々の行動系列の尤度により明確に記述

## 提案手法の限界

- ・ 歩行者行動に限定しており交通手段の検出はできない
- ・ 地図情報の重要性
- ・ 事前情報の量が必ずしもよい結果とはならず、魅力度や時間制約条件の慎重な設定が必要
- ・ DDRの半径 $R$ を分析者が恣意的に設定：検出結果に影響