RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

Inria

~~(Don't you) Forget about me:~~
~~On the importance of time in data~~
Smart compression is bad

~~Too much data to handle? Let's see what we can do!~~
Don't overthink it, truly

Rémy Raes

01
Context

Time series

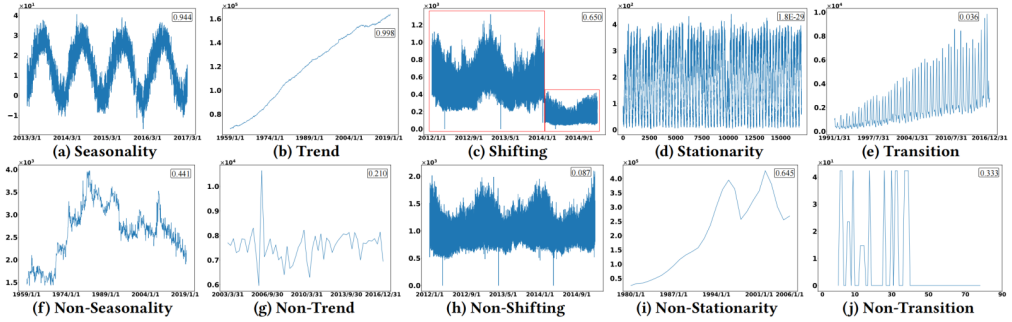
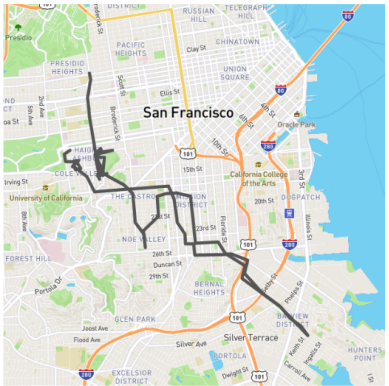
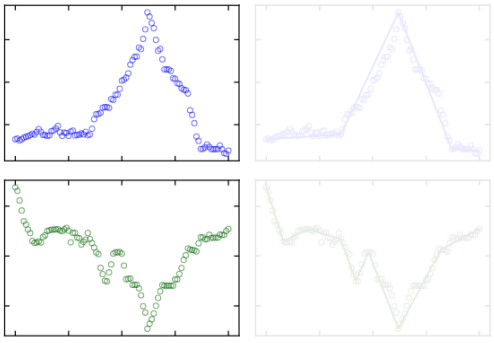


Figure 1: Visualization of data with different characteristics.

Qiu et al. TFB: Towards Comprehensive and Fair Benchmarking of Time Series Forecasting Methods. *Proceedings of the VLDB Endowment, Vol. 17, No. 9 ISSN 2150-809* **(10.14778/3665844.3665863)**

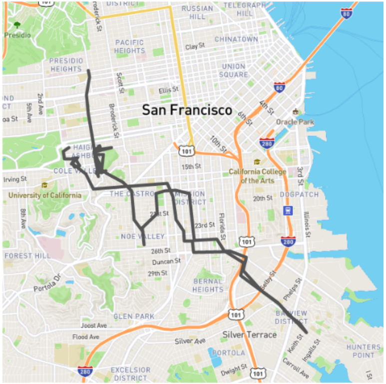
Fast linear interpolation (FLI)

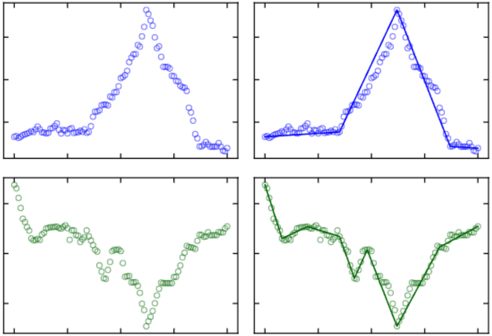




Rémy Raes, Olivier Ruas, Adrien Luxey-Bitri, Romain Rouvoy. Compact Storage of Data Streams in Mobile Devices. *DAIS'24 - 24th International Conference on Distributed Applications and Interoperable Systems, Jun 2024, Groningen, Netherlands.* **(hal-04535716v3)**

Fast linear interpolation (FLI)





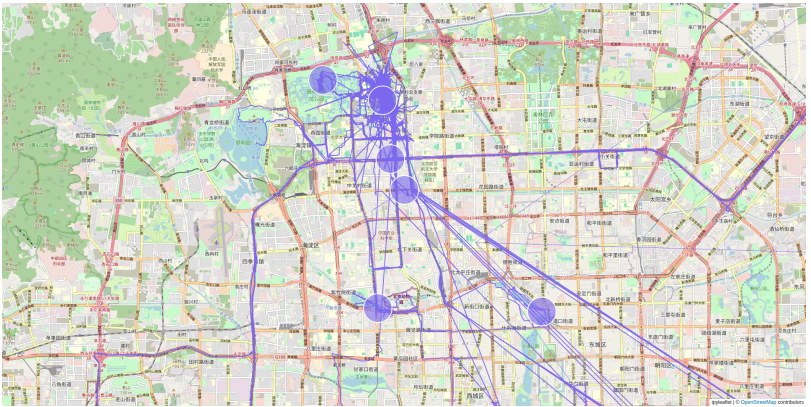
Rémy Raes, Olivier Ruas, Adrien Luxey-Bitri, Romain Rouvoy. Compact Storage of Data Streams in Mobile Devices. *DAIS'24 - 24th International Conference on Distributed Applications and Interoperable Systems, Jun 2024, Groningen, Netherlands.* **(hal-04535716v3)**

Location privacy

- Location data is sensitive
- Can be mined to extract **Points of Interest** *(POIs)*

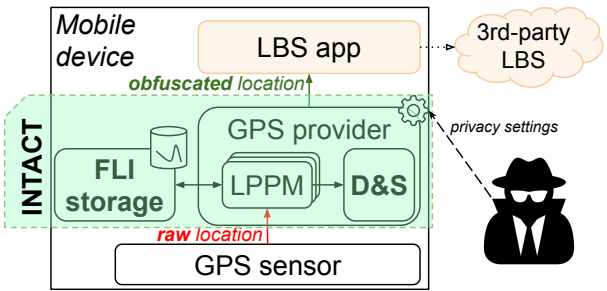
Location privacy

- Location data is sensitive
- Can be mined to extract **Points of Interest** *(POIs)*



Embedded, mobile privacy framework

- INTACT: *in situ* location protection

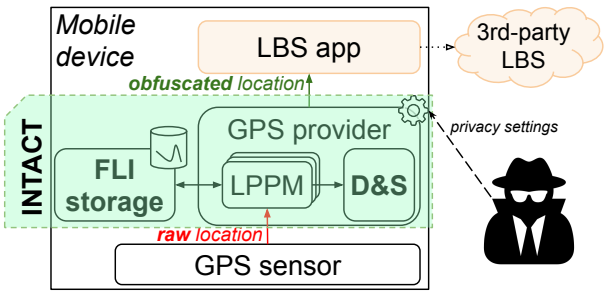


- Local storage of private data
- Local attack and protection mechanisms
- Locally ensure data is safe before sharing it

Rémy Raes, Olivier Ruas, Adrien Luxey-Bitri, Romain Rouvoy. INTACT: Compact Storage of Data Streams in Mobile Devices to Unlock User Privacy at the Edge. *Journal of Internet Services and Applications* **(hal-05126130v1)**

Embedded, mobile privacy framework

- INTACT: *in situ* location protection

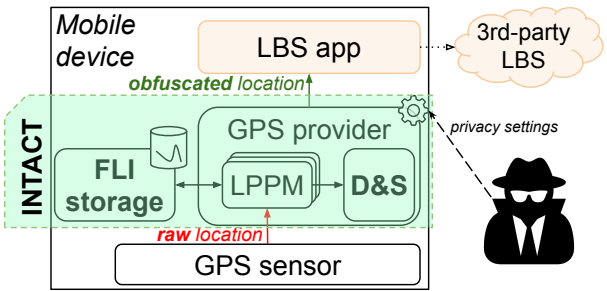


- Local storage of private data
- Local attack and protection mechanisms
- Locally ensure data is safe before sharing it

Rémy Raes, Olivier Ruas, Adrien Luxey-Bitri, Romain Rouvoy. INTACT: Compact Storage of Data Streams in Mobile Devices to Unlock User Privacy at the Edge. *Journal of Internet Services and Applications* **(hal-05126130v1)**

Embedded, mobile privacy framework

- INTACT: *in situ* location protection

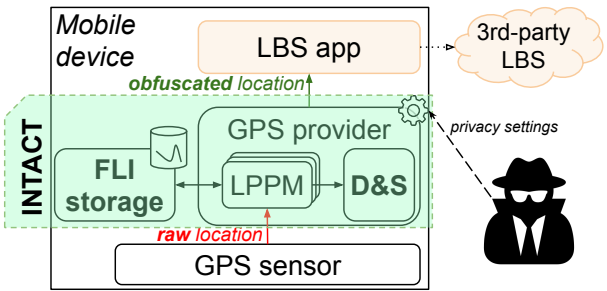


- Local storage of private data
- Local attack and protection mechanisms
- Locally ensure data is safe before sharing it

Rémy Raes, Olivier Ruas, Adrien Luxey-Bitri, Romain Rouvoy. INTACT: Compact Storage of Data Streams in Mobile Devices to Unlock User Privacy at the Edge. *Journal of Internet Services and Applications* **(hal-05126130v1)**

Embedded, mobile privacy framework

- INTACT: *in situ* location protection



- Local storage of private data
- Local attack and protection mechanisms
- Locally ensure data is safe before sharing it

Rémy Raes, Olivier Ruas, Adrien Luxey-Bitri, Romain Rouvoy. INTACT: Compact Storage of Data Streams in Mobile Devices to Unlock User Privacy at the Edge. *Journal of Internet Services and Applications* **(hal-05126130v1)**

Error selection

- Use a function to pick tolerated error
- Function is time-indexed
- Different behaviours:
  - Constant value (FLI)
  - Decreasing value:
    - Linear
    - by step
    - with power function

Error selection

- Use a function to pick tolerated error
- Function is time-indexed
- Different behaviours:
  - Constant value (FLI)
  - Decreasing value:
    - Linear
    - by step
    - with power function

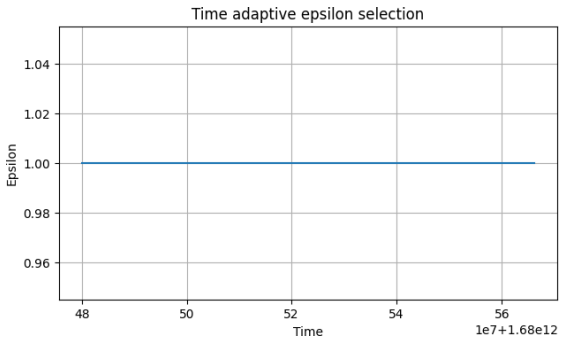
Error selection

- Use a function to pick tolerated error
- Function is time-indexed
- Different behaviours:
  - Constant value (FLI)
  - Decreasing value:
    - Linear
    - by step
    - with power function

Error selection

- Use a function to pick tolerated error
- Function is time-indexed
- Different behaviours:
  - Constant value (FLI)
  - Decreasing value:
    - Linear
    - by step
    - with power function

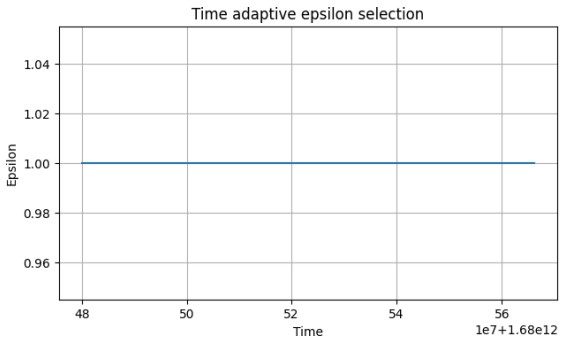
Error selection



- Use a function to pick tolerated error
- Function is time-indexed
- Different behaviours:
  - Constant value (FLI)
  - Decreasing value:
    - Linear
    - by step
    - with power function

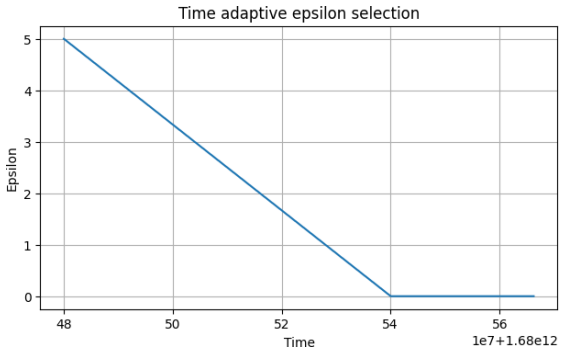
02
Works

Error selection



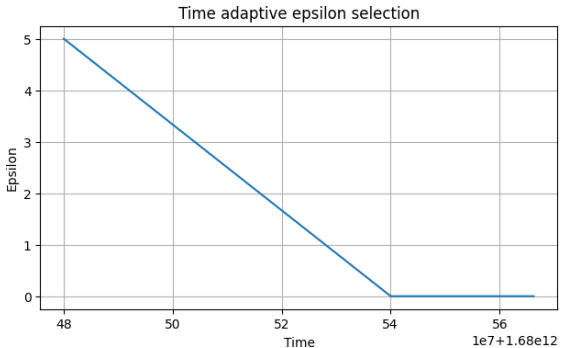
- Use a function to pick tolerated error
- Function is time-indexed
- Different behaviours:
  - Constant value (FLI)
  - Decreasing value:
    - Linear
    - by step
    - with power function

Error selection



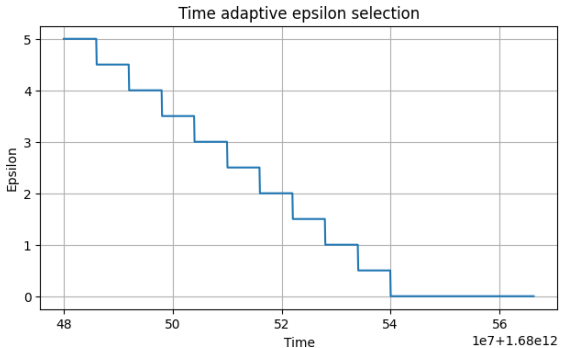
- Use a function to pick tolerated error
- Function is time-indexed
- Different behaviours:
  - Constant value (FLI)
  - Decreasing value:
    - Linear
    - by step
    - with power function

Error selection



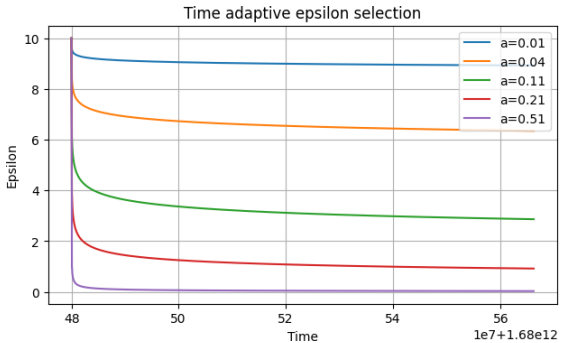
- Use a function to pick tolerated error
- Function is time-indexed
- Different behaviours:
  - Constant value (FLI)
  - Decreasing value:
    - Linear
    - by step
    - with power function

Error selection



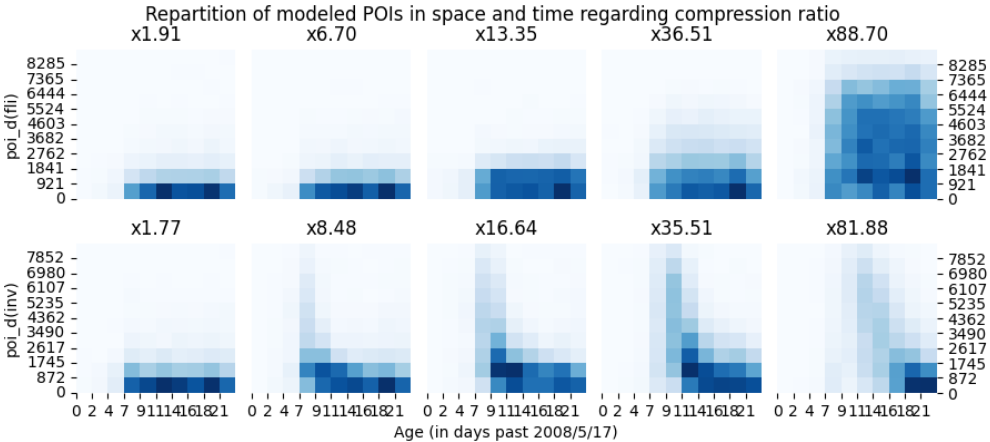
- Use a function to pick tolerated error
- Function is time-indexed
- Different behaviours:
  - Constant value (FLI)
  - Decreasing value:
    - Linear
    - by step
    - with power function

Error selection



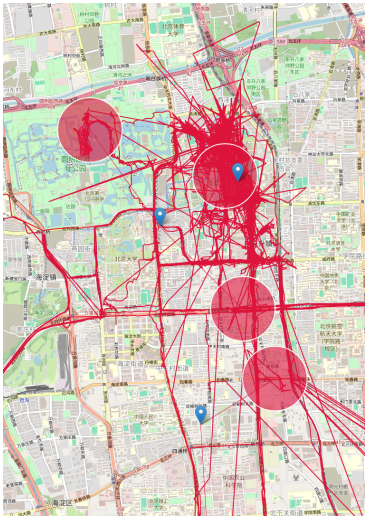
- Use a function to pick tolerated error
- Function is time-indexed
- Different behaviours:
  - Constant value (FLI)
  - Decreasing value:
    - Linear
    - by step
    - with power function

Old data degradation hint



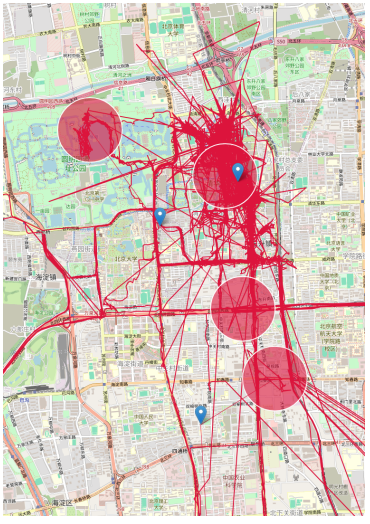
Benchmark

- POI prediction
  - Split traces in *train* and *test* datasets
  - Extract POIs from *train*
  - Extract POIs from compressed *test*
  - Compare both POI sets
- Several compression techniques
  - Modelling (with constant and varying error)
  - Removal (of newer or older data)

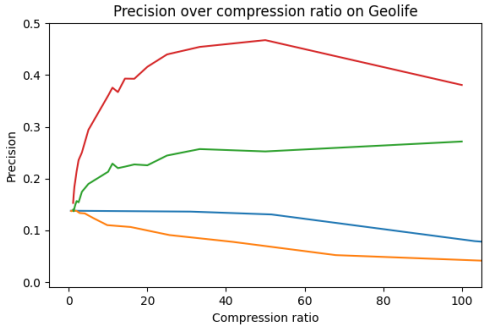


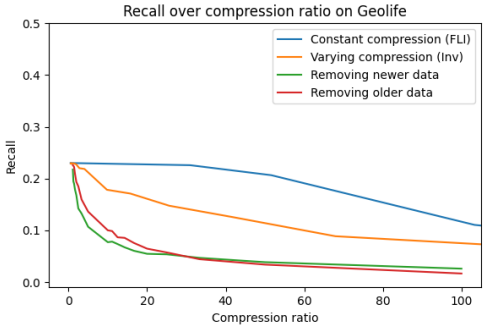
Benchmark

- POI prediction
  - Split traces in *train* and *test* datasets
  - Extract POIs from *train*
  - Extract POIs from compressed *test*
  - Compare both POI sets
- Several compression techniques
  - Modelling (with constant and varying error)
  - Removal (of newer or older data)

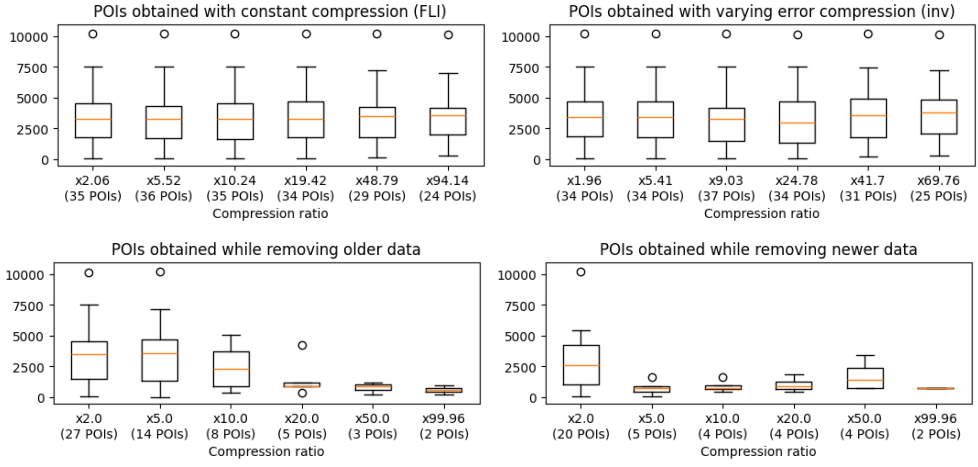


Results (1/2)


Precision over compression ratio on Geolife
Precision
Compression ratio
0.0
0.1
0.2
0.3
0.4
0.5
0
20
40
60
80
100


Recall over compression ratio on Geolife
Recall
Compression ratio
Constant compression (FLI)
Varying compression (Inv)
Removing newer data
Removing older data
0.0
0.1
0.2
0.3
0.4
0.5
0
20
40
60
80
100

Results (2/2): Distance of modeled POIs to closest raw POI



03
Future works

What's next

- Looks like smart compression is not good
- Is there a use case requiring keeping a highly degraded version of the data rather than its removal?
  - "Right to be forgotten"?

What's next

- Looks like smart compression is not good
- Is there a use case requiring keeping a highly degraded version of the data rather than its removal?
  - "Right to be forgotten"?

What’s next

- Looks like smart compression is not good
- Is there a use case requiring keeping a highly degraded version of the data rather than its removal?
  - ”Right to be forgotten”?

”Right to be forgotten” hint

Data days old

Data months old

Data years old

”Right to be forgotten” hint

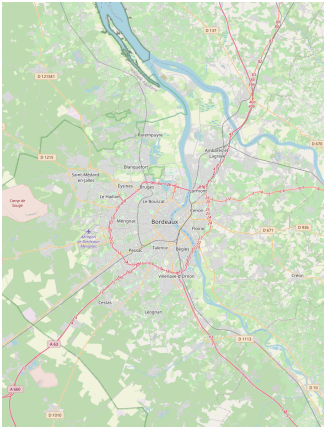
Data days old

Data months old

Data years old

”Right to be forgotten” hint

Data days old



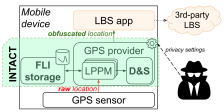
Data months old

Data years old

Take away

Embedded, mobile privacy framework

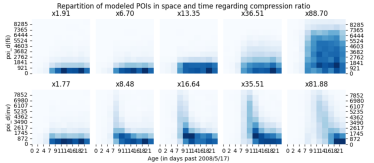
- INTACT: *in situ* location protection



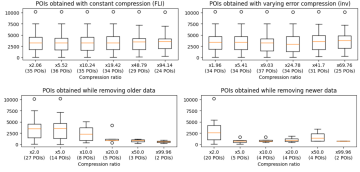
- Local storage of private data
- Local attack and protection mechanisms
- Locally ensure data is safe before sharing it

Rémy Raes, Olivier Ruas, Adrien Luxey-Bitri, Romain Rouvoy. INTACT: Compact Storage of Data Streams in Mobile Devices to Unlock User Privacy at the Edge. *Journal of Internet Services and Applications (to be published soon*[TM]*)*

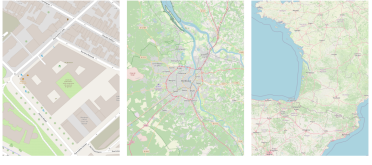
Old data degradation hint



Results (2/2): Distance of modeled POIs to closest raw POI



"Right to be forgotten" hint



Merci.

Research questions

Distributed Machine Learning in Ubiquitous Environments using Location-dependent Models

- How to store unbounded data streams on constrained mobile devices?
- How to exchange relevant model samples among nearby devices?
- How to program DML algorithms for the masses?

About the *epsilon* value

- Selecting a good ϵ value requires data domain knowledge
- Drift between consecutive values (x_1, y_1) and (x_2, y_2): $|y_2 - y_1|/|x_2 - x_1|$.

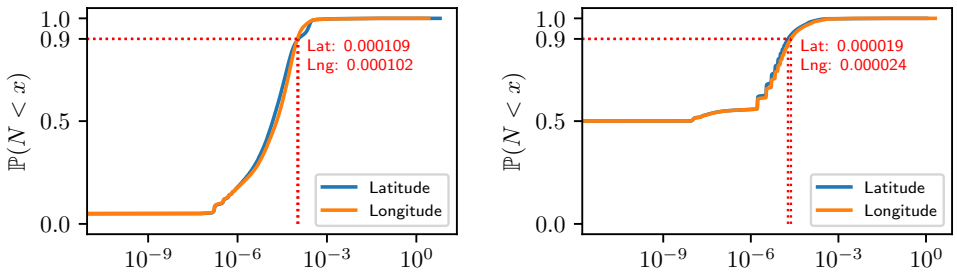


Figure – CDF of latitude and longitude variations of successive locations in CABSPOTTING and PRIVAMOV.

- We used $\epsilon = 10^{-3}$ as a baseline value in the FLI paper

Data utility with location data

Size results

- $\epsilon = 10^{-3}$
- From 7.2 GB to 25 MB

Data utility

- Latitude
  - Tolerated error: 10^{-3} *deg* ≈ 111 *m*
  - Median error: 5.33×10^{-5}
  - RMSE: 3.72×10^{-4}
- Longitude
  - Tolerated error: 10^{-3} *deg* ≈ 88 *m*
  - Median error: 2.81×10^{-5}
  - RMSE: 3.44×10^{-4}
- Privacy utility

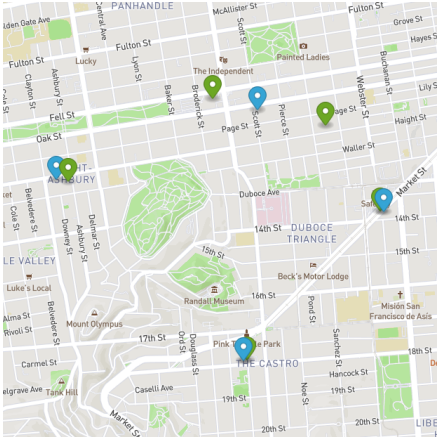


Figure – Points of Interest computed using raw data and FLI-modeled data.

Data utility with location data

Size results

- $\epsilon = 10^{-3}$
- From 7.2 GB to 25 MB

Data utility

- Latitude
  - Tolerated error: 10^{-3} *deg* ≈ 111 *m*
  - Median error: 5.33×10^{-5}
  - RMSE: 3.72×10^{-4}
- Longitude
  - Tolerated error: 10^{-3} *deg* ≈ 88 *m*
  - Median error: 2.81×10^{-5}
  - RMSE: 3.44×10^{-4}
- Privacy utility

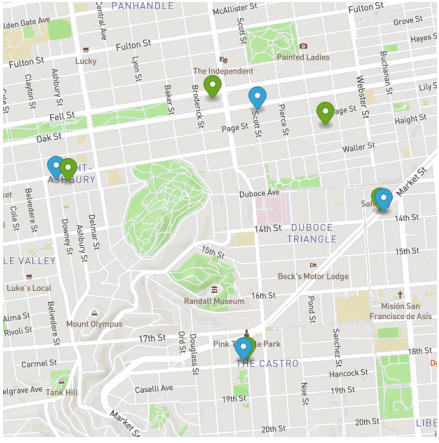


Figure – Points of Interest computed using raw data and FLI-modeled data.

Data utility with location data

Size results

- $\epsilon = 10^{-3}$
- From 7.2 GB to 25 MB

Data utility

- Latitude
  - Tolerated error: 10^{-3} *deg* ≈ 111 *m*
  - Median error: 5.33×10^{-5}
  - RMSE: 3.72×10^{-4}
- Longitude
  - Tolerated error: 10^{-3} *deg* ≈ 88 *m*
  - Median error: 2.81×10^{-5}
  - RMSE: 3.44×10^{-4}
- Privacy utility

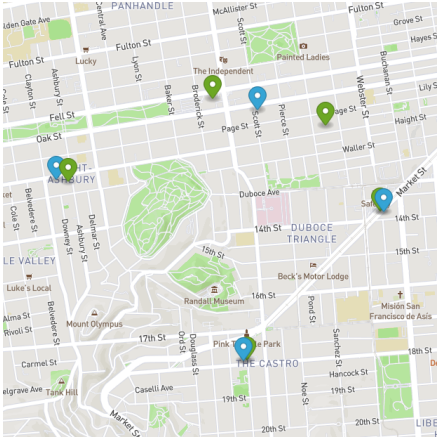


Figure – Points of Interest computed using raw data and FLI-modeled data.

Data utility with location data

Size results

- $\epsilon = 10^{-3}$
- From 7.2 GB to 25 MB

Data utility

- Latitude
  - Tolerated error: 10^{-3} *deg* ≈ 111 *m*
  - Median error: 5.33×10^{-5}
  - RMSE: 3.72×10^{-4}
- Longitude
  - Tolerated error: 10^{-3} *deg* ≈ 88 *m*
  - Median error: 2.81×10^{-5}
  - RMSE: 3.44×10^{-4}
- Privacy utility

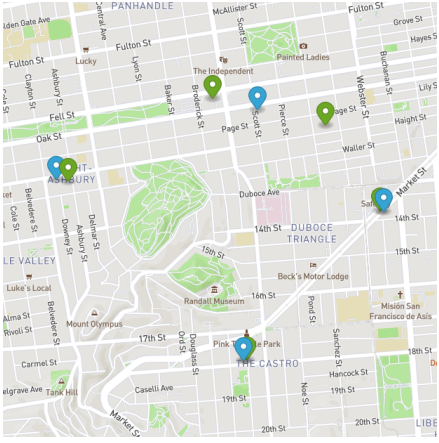


Figure – Points of Interest computed using raw data and FLI-modeled data.