Interdependence and Predictability of Human Mobility and Social Interactions

Antonio Lima
Joint work with Manlio De Domenico and Mirco Musolesi

Nokia MDC Workshop
Newcastle - June 18, 2012

UNIVERSITY OF BIRMINGHAM

Courtesy of Indiana University

Human Mobility.
Can we predict it?

We can, to a certain extent and at different geographic scales.

UNIVERSITY OF BIRMINGHAM

Research Questions

- Is it possible to improve the accuracy of the prediction by considering traces of multiple users?
- If yes, who should we select for improving the prediction of the movements of a given user?
- Can mobility correlation be considered as a cue for inferring social ties?

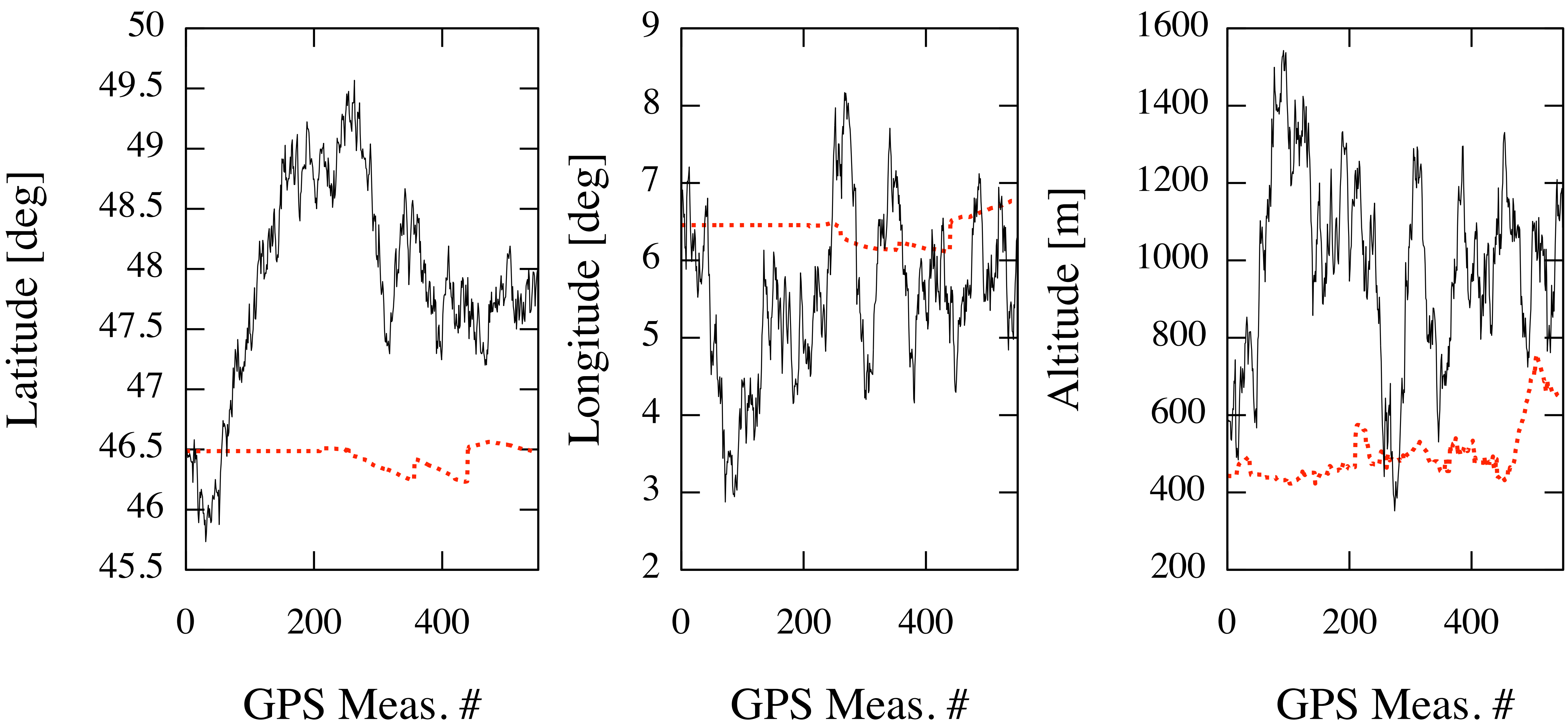
flickr: cubagallery

The Nokia MDC Dataset

- The complete dataset contains information from 152 smartphones (Nokia N95) for a year: address book, GPS, WLAN and Bluetooth traces, calls and SMS logs.

- We received data from 39 devices, 14 phone numbers were missing. We analysed a subset of the data related to 25 devices.

UNIVERSITY OF BIRMINGHAM



600 GPS (~60 hours) measurements (red) against forecast (black) for user 129

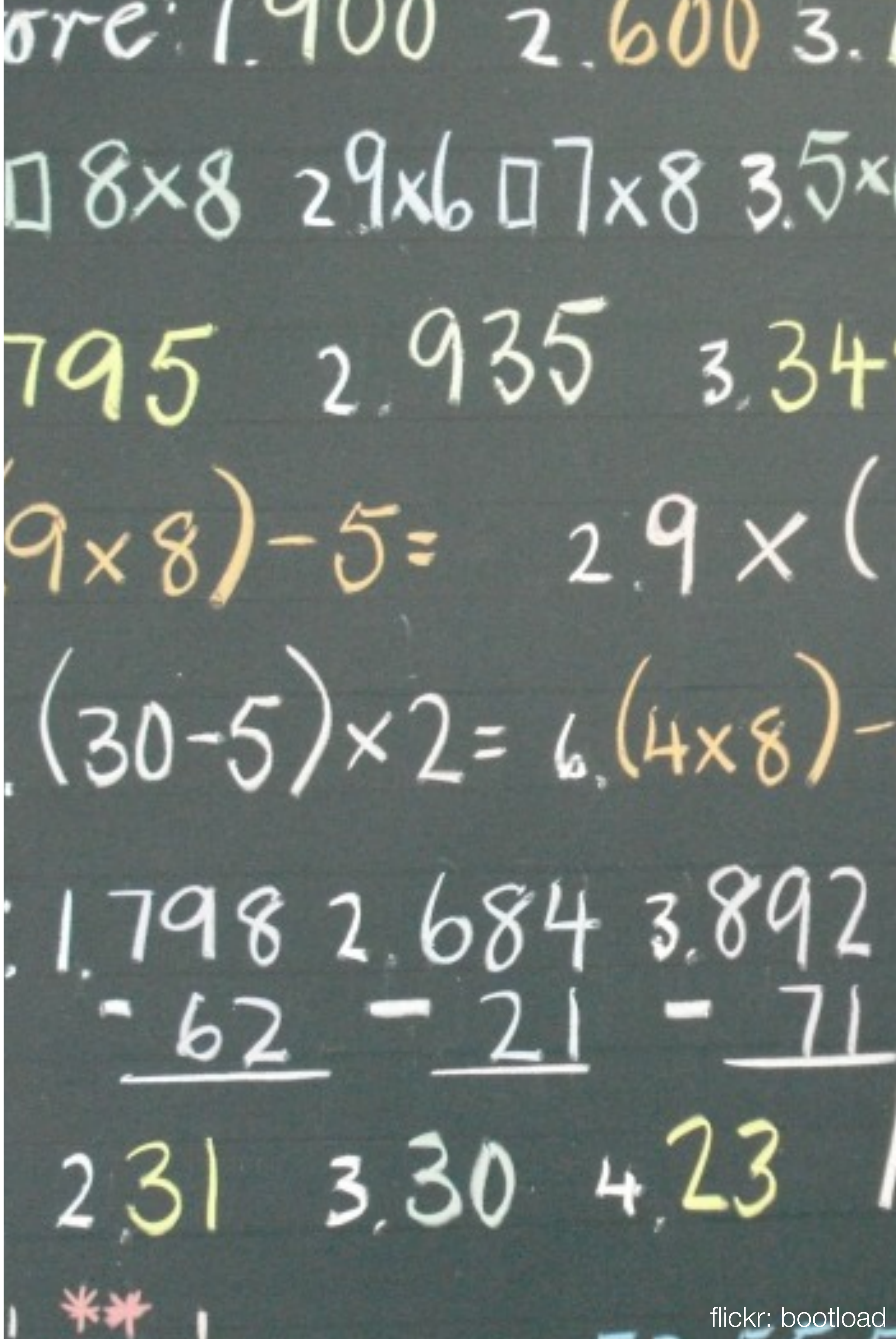
Linear Predictor

Error is of the order of 3 deg for lat-lng and 600 m for altitude.

UNIVERSITY OF BIRMINGHAM

Our Approach

- Multivariate nonlinear time series prediction.
- Extension to the case of multiple users. In particular pairs of users:
  - connected by a social link; and/or
  - with correlated mobility patterns.



flickr: bootload

UNIVERSITY OF BIRMINGHAM

The Mobility Model

$$\dot{\mathbf{x}}(t) = f(x, t) + \eta(t)$$

We consider the mobility model as a nonlinear dynamical system of a deterministic signal with a stochastic noise.

$$\mathbf{x_n} = (h_n, \phi_n, \lambda_n, \xi_n)$$

The position of a user is modeled with a 4-dim state vector.

We cannot analyse the d-dimensional phase space of the system directly.

The Mobility Model

User movements

$$\dot{\mathbf{x}}(t) = f(x,t) + \eta(t)$$

Noise

$$\mathbf{x_n} = (h_n, \phi_n, \lambda_n, \xi_n)$$

We consider the mobility model as a nonlinear dynamical system of a deterministic signal with a stochastic noise.

The position of a user is modeled with a 4-dim state vector.

We cannot analyse the d-dimensional phase space of the system directly.

The Mobility Model

User movements

$$\dot{\mathbf{x}}(t) = f(x,t) + \eta(t)$$

Noise

Hour of the day Longitude

$$\mathbf{x_n} = (h_n, \phi_n, \lambda_n, \xi_n)$$

Latitude Altitude

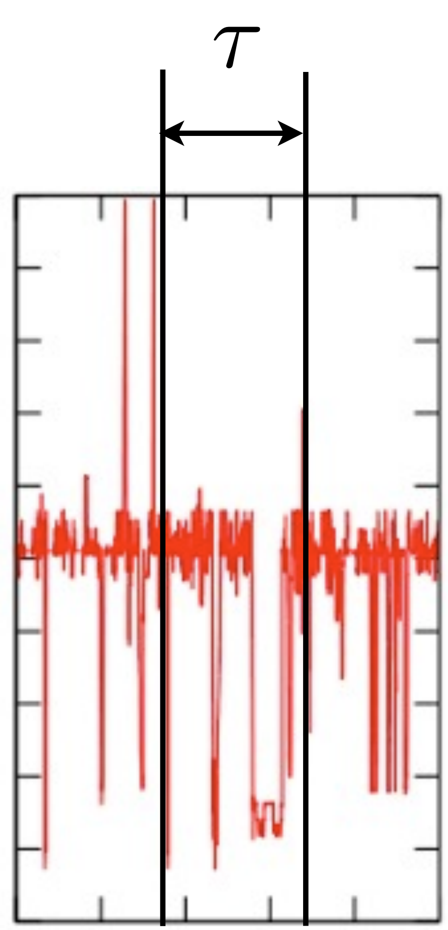
We consider the mobility model as a nonlinear dynamical system of a deterministic signal with a stochastic noise.

The position of a user is modeled with a 4-dim state vector.

We cannot analyse the d-dimensional phase space of the system directly.

Takens' Embedding Theorem

$$\mathbf{x}_n \equiv (x_{n-(m-1)\tau}, \ldots, x_{n-\tau}, x_n)$$



We can construct a space which preserves the dynamic properties of the system by using delayed measurements of the time-series.

The theorem holds for noiseless time series of infinite length.
We need a **multivariate** analysis to have a good precision on real-world time-limited noisy data.

Takens' Embedding Theorem

$$\mathbf{x}_n \equiv (x_{n-(m-1)\tau}, \ldots, x_{n-\tau}, x_n)$$

We can construct a space which preserves the dynamic properties of the system by using delayed measurements of the time-series.

$$\begin{aligned}\mathbf{v}_n \equiv (&y_{1,n-(m_1-1)\tau_1}, \ldots, y_{1,n}, \\ &y_{2,n-(m_2-1)\tau_2}, \ldots, y_{2,n}, \\ &\ldots \\ &y_{M,n-(m_M-1)\tau_M}, \ldots, y_{M,n})\end{aligned}$$

The theorem holds for noiseless time series of infinite length.
We need a **multivariate** analysis to have a good precision on real-world time-limited noisy data.

Takens' Embedding Theorem

Embedding dimension = 8

$$\mathbf{x}_n \equiv (x_{n-(m-1)\tau}, \ldots, x_{n-\tau}, x_n)$$

Delay time ~= 1000

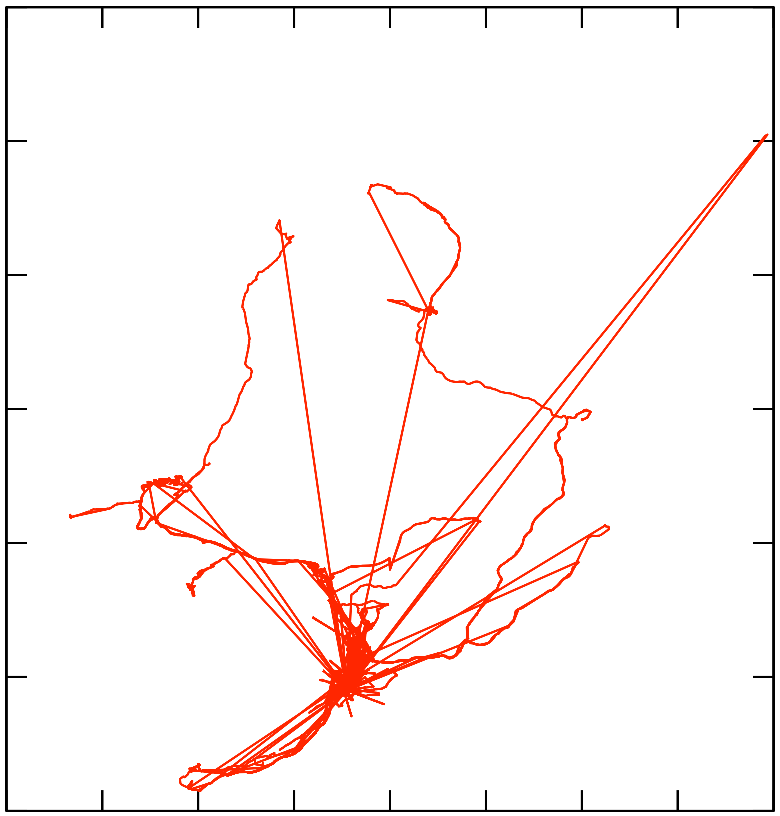
We can construct a space which preserves the dynamic properties of the system by using delayed measurements of the time-series.

$$\begin{aligned}\mathbf{v}_n \equiv (&y_{1,n-(m_1-1)\tau_1}, \ldots, y_{1,n}, \\ &y_{2,n-(m_2-1)\tau_2}, \ldots, y_{2,n}, \\ &\ldots \\ &y_{M,n-(m_M-1)\tau_M}, \ldots, y_{M,n})\end{aligned}$$

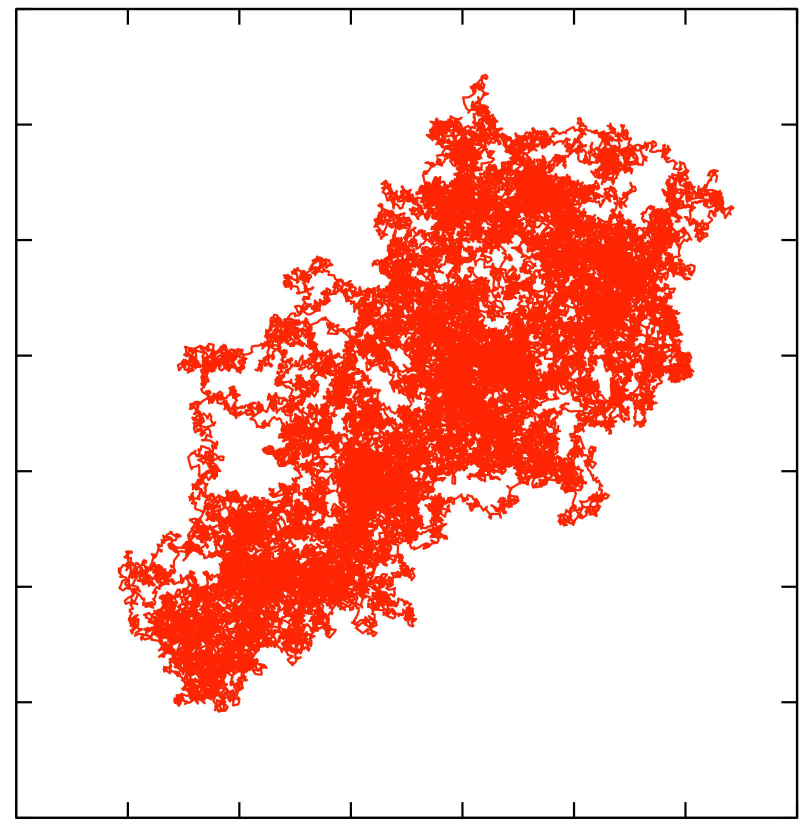
The theorem holds for noiseless time series of infinite length.
We need a **multivariate** analysis to have a good precision on real-world time-limited noisy data.

UNIVERSITY OF BIRMINGHAM

Delay Embedding Reconstructions



Reconstruction for
user 179

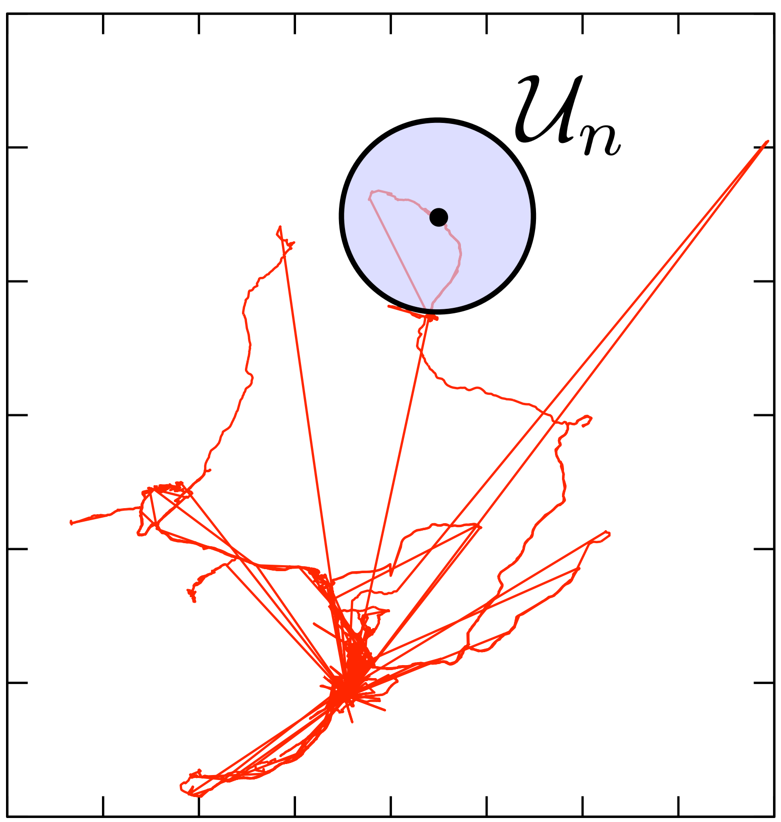


Reconstruction for
a Brownian motion

UNIVERSITY OF
BIRMINGHAM

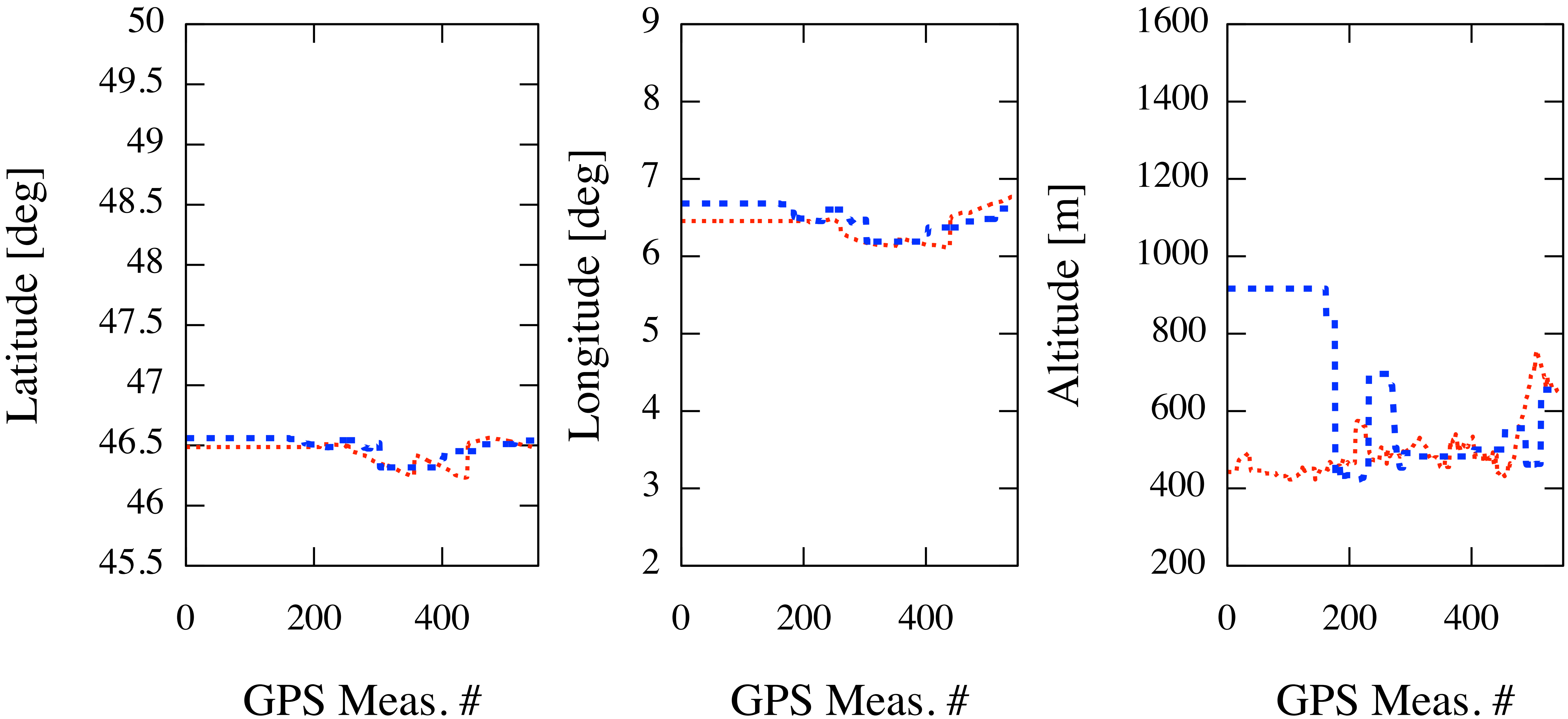
Multivariate Nonlinear Prediction

$$\hat{\mathbf{v}}_{n+k} = \frac{1}{|\mathcal{U}_n|} \sum_{\mathbf{v}_j \in \mathcal{U}_n} \mathbf{v}_{j+k}$$



The prediction is performed considering the average over the states which are *k* steps ahead of the neighbours states.

In the reconstruction represented here for m=2, neighbours are inside the azure area.

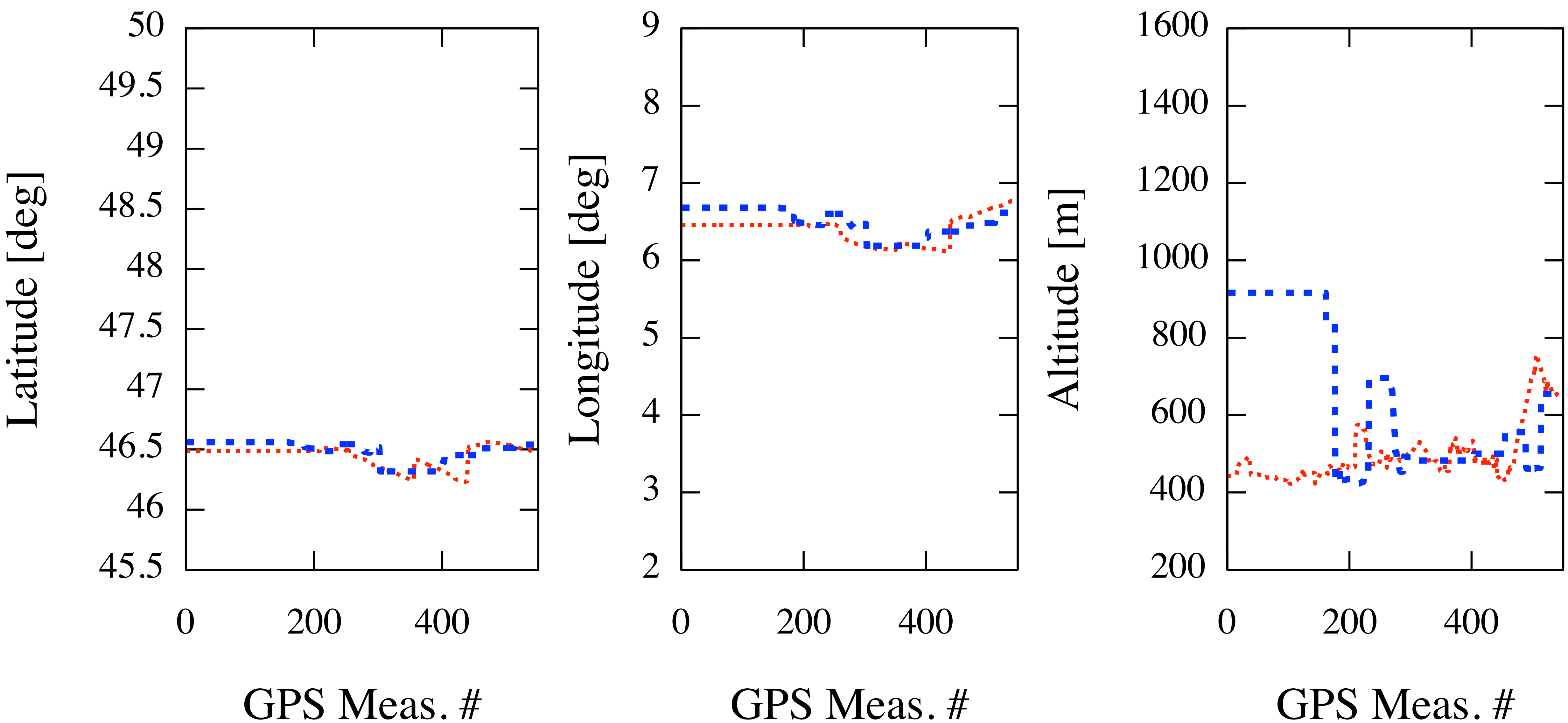


600 GPS (~60 hours) measurements (red)
against forecast (black) for user 129

Multivariate One-user Prediction

Much better, global prediction error of 0.19 deg for lat/lng and 219.43 m for the altitude.

UNIVERSITY OF BIRMINGHAM



600 GPS (~60 hours) measurements (red)
against forecast (black) for user 129

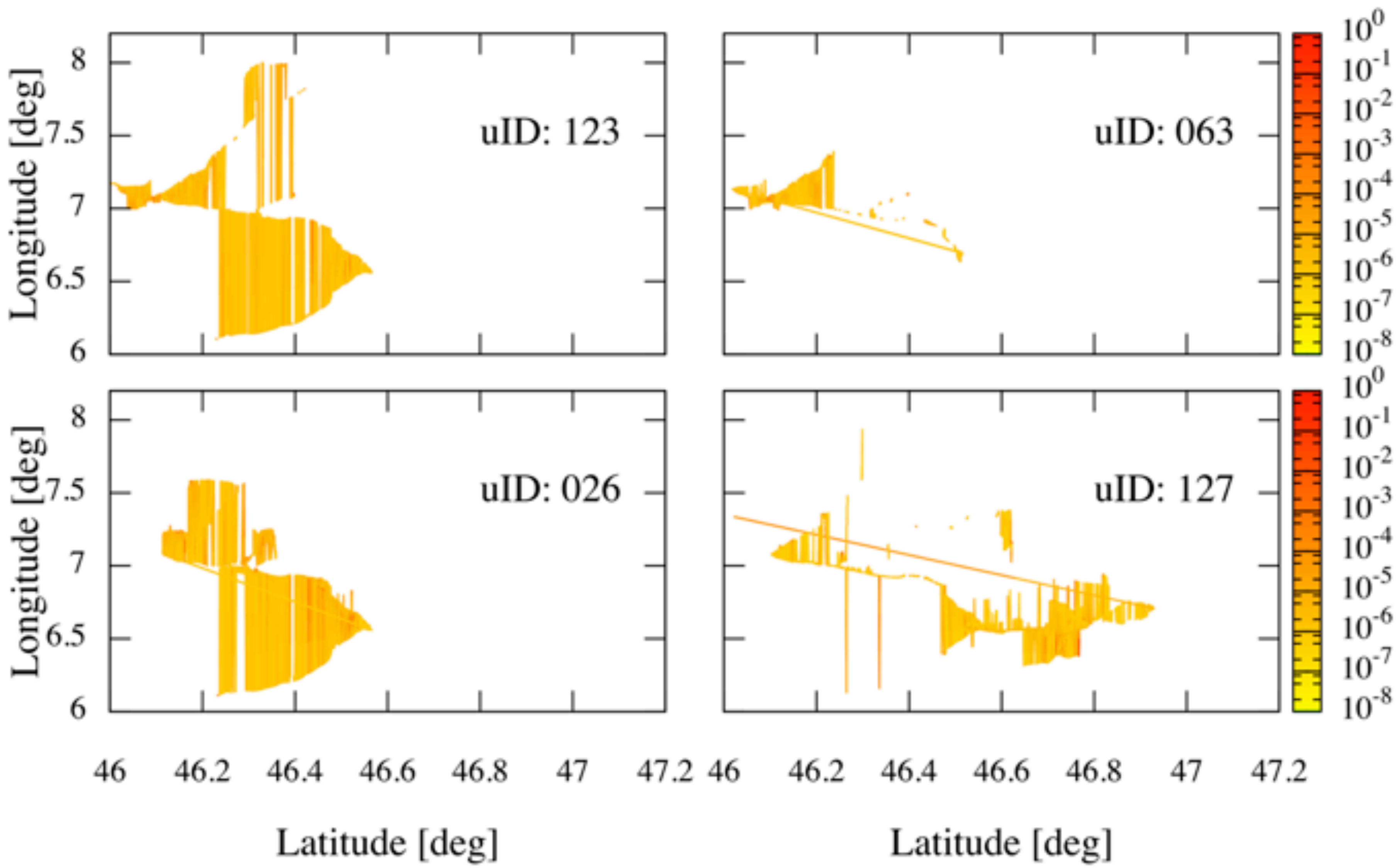
Multivariate One-user Prediction

Much better, global prediction error of 0.19 deg for lat/lng and 219.43 m for the altitude.

Let's do even better!

UNIVERSITY OF BIRMINGHAM

Mobility Probability Density Function



PDF of positions of users who are friends (top) and who are not friends (bottom)

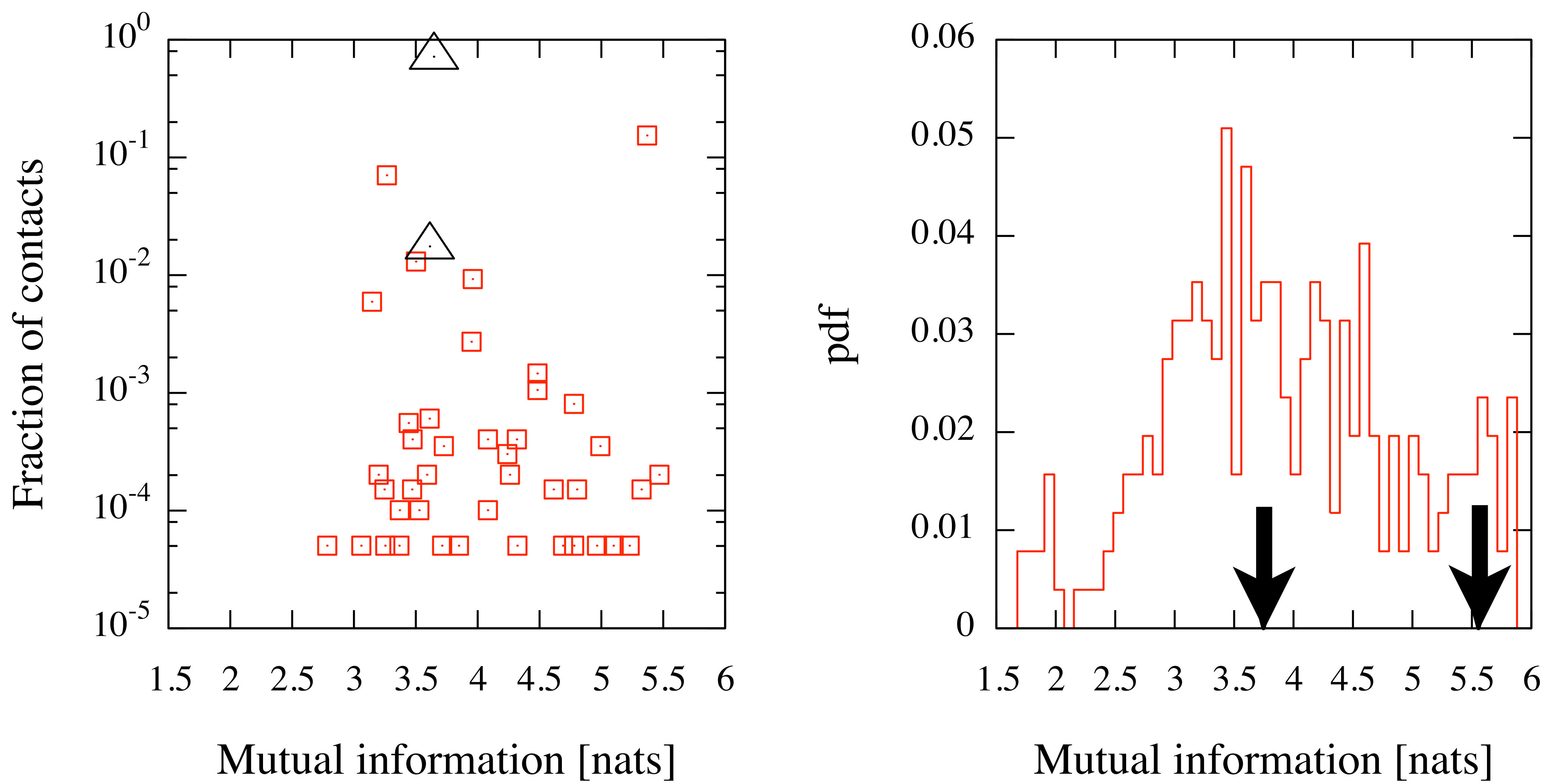
Mutual Information

$$\mathcal{I}(\mathbf{X},\mathbf{Y}) = \sum_{\mathbf{x}\in\mathbf{X}}\sum_{\mathbf{y}\in\mathbf{Y}} P_{\mathbf{XY}}(\mathbf{x},\mathbf{y}) \log \frac{P_{\mathbf{XY}}(\mathbf{x},\mathbf{y})}{P_{\mathbf{X}}(\mathbf{x})P_{\mathbf{Y}}(\mathbf{y})}$$

The mutual information quantifies how much information a stochastic variable can provide about another stochastic variable. It can be used as an estimator of the amount of correlation between them. If they are uncorrelated, it is null.

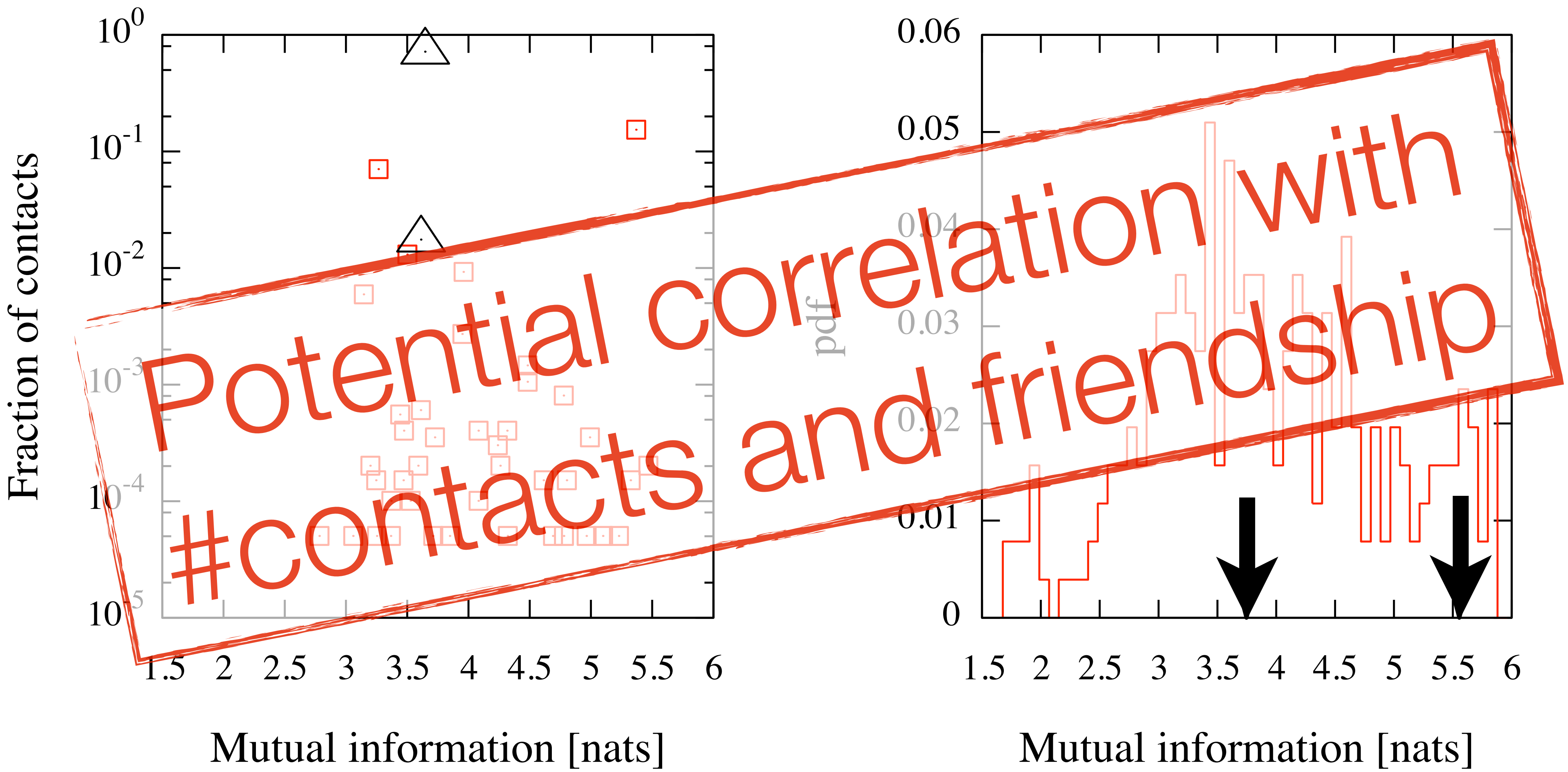
We use it to quantify how much the motion of a user can give us information about the motion of another.

Mutual Information, Contacts, Friendship



M.I. for people with at least one contact (left) and for people with no contacts at all (right).

Mutual Information, Contacts, Friendship



M.I. for people with at least one contact (left) and for people with no contacts at all (right).

Nodes	Social Link	Pos. Error [deg]	Alt. error [m]
026, 127	None	0.167	66.33
063, 123	Present	0.011	20.95
094, 009	Present	0.003	5.57

Multivariate two-users prediction

The accuracy of the prediction improves by at least one order of magnitude (often two).

Take-away Messages

Human mobility traces are sometimes correlated.

Correlated traces improve forecasting accuracy.

Correlation can be a signal of social interaction.

Thanks!

Questions?

Antonio Lima

a.lima@cs.bham.ac.uk
http://cs.bham.ac.uk/axl162
@themiurgo

Thanks to flickr authors of
CC-licenced images.
http://urli.st/M39

UNIVERSITY OF BIRMINGHAM