Journal of Network and Computer Applications 168 (2020) 102778

Contents lists available at ScienceDirect

Journal of Network and Computer Applications

journal homepage: www.elsevier.com/locate/jnca

# A deep stochastical and predictive analysis of users mobility based on Auto-Regressive processes and pairing functions

Peppino Fazio [a,b,*], Miralem Mehic [c,a], Miroslav Voznak [a]

[a] VSB – Technical University of Ostrava, 17. Listopadu 2172/15, 708 33 Ostrava, Czech Republic
[b] DIMES, University of Calabria, Cube 39c, 87036 Arcavacata di Rende, Italy
[c] Department of Telecommunications, Faculty of Electrical Engineering, University of Sarajevo, Zmaja Od Bosne Bb, 71000, Sarajevo, Bosnia and Herzegovina

ARTICLE INFO

*Index Terms:*
Mobile networking
Mobility
Prediction
Quality of service
Stability
Correlation function
Pairing functions

ABSTRACT

With the proliferation of connected vehicles, new coverage technologies and colossal bandwidth availability, the quality of service and experience in mobile computing play an important role for user satisfaction (in terms of comfort, security and overall performance). Unfortunately, in mobile environments, signal degradations very often affect the perceived service quality, and predictive approaches become necessary or helpful, to handle, for example, future node locations, future network topology or future system performance. In this paper, our attention is focused on an in-depth stochastic micro-mobility analysis in terms of nodes coordinates. Many existing works focused on different approaches for realizing accurate mobility predictions. Still, none of them analyzed the way mobility should be collected and/or observed, how the granularity of mobility samples collection should be set and/or how to interpret the collected samples to derive some stochastic properties based on the mobility type (pedestrian, vehicular, etc.). The main work has been carried out by observing the characteristics of vehicular mobility, from real traces. At the same time, other environments have also been considered to compare the changes in the collected statistics. Several analyses and simulation campaigns have been carried out and proposed, verifying the effectiveness of the introduced concepts.

## 1. Introduction

Pattern prediction, location tracking and micro/macro mobility analysis represent several research topics that attracted researchers' attention from decades (Pirozmand et al., 2014). The ability to predict users' movements benefits a wide range of mobile wireless systems, such as location-based applications, mobile access control, Quality of Service (QoS) provisioning, resource allocation/management, urban planning, epidemic control, location-based services, and intelligent transportation management (Fazio et al., 2017). For example, when referring to ad-hoc networks, node mobility is one of the critical aspects that have been investigated and predicted, given that it is crucial for Mobile Ad-hoc NETworks (MANETs). Additionally, an Information-Centric Network (ICN), dedicated to offering some features such as distributed storage, caching and content relocation, could use mobility prediction to optimize the content distribution, according to user's future locations: the user content will be available in the area it will visit in the future, with a considerable reduction of content access delay. The overall performance of the considered system, for which mobility prediction is exploited, depends on the accuracy of the proposed approach, as well as on the intrinsic traffic/mobility dynamics (Satria et al., 2014).

In this paper, we do not propose only a new predictive scheme, but also an in-depth analysis of the way mobility samples should be considered and collected to build the historical records, which represent the starting point for any conventional/unconventional approach (neural networks, machine learning, Markov chains, Kalman's filtering and other). The prediction process is always based on a preliminary training phase, represented by the history of past user movements. To the best of our knowledge, no works until now have considered the following issues:

- Which should be the frequency of collecting mobility samples (represented by some form of locations or, by the coordinates)?
- How far should the prediction approach go when obtaining the future mobility sample? Is it useful, for a mobile system to predict the next user position after several seconds or the system is inter-

* Corresponding author. VSB – Technical University of Ostrava, 17. Listopadu 2172/15, 708 33 Ostrava, Czech Republic.
*E-mail addresses:* peppino.fazio@vsb.cz (P. Fazio), miralem.mehic@ieee.org (M. Mehic), miroslav.voznak@vsb.cz (M. Voznak).

https://doi.org/10.1016/j.jnca.2020.102778
Received 14 February 2020; Received in revised form 1 July 2020; Accepted 10 July 2020
Available online 9 August 2020

ested in knowing the future location after a more substantial period of time?

- Assuming that its spatial coordinates represent the location of a moving node (such as *x,y* in a 2D space), is there a way to make only one evaluation of the next position, without evaluating the variations of *x* and *y* separately?

In this paper we address listed questions, both from a theoretical and practical point of view, capturing the main concepts, dynamics and aspects strictly related to the mobility prediction analysis, giving the reader the needed knowledge about the opportunities of making a right choice during the exploitation of predictive approaches. Without loss of generality, we carry out our analysis on mobility records composed only by GPS coordinates (or some representations of them), without considering any additional feature (such as social relations, PoIs, PSIs, LBSNs, etc.). In this way, the proposed approach is completely general and can be enhanced, if needed, by considering the relevant issue, based on the considered scenario.

The paper is organized as follows: Section 2 introduces some recent works about mobility prediction models and recent advances, while Section 3 gives a deeper description of the proposed idea, under a theoretical point of view. Section 4 gives a deep description of the main reachable results, and section 5 concludes the paper.

## 2. State of the art and main contributions

There are many works in the literature about predictive approaches in mobile networks (Zhang and Dai, 2019): in many mobile network architectures, it is desired to a priori know (with a possibly low prediction error) the future movements of mobile nodes, to enhance the overall performance of the considered system (Zareei et al., 2018).

In (Zhao et al., 2015), the authors emphasize the concept of location prediction, underlining its effects on mobility and bandwidth evaluation. The usage of the knowledge about the future node locations in LTE networks for self-adaptation procedures and optimal network configuration during run-time operations is illustrated.

In (Yamada et al., 2018), a novel prediction scheme is proposed, based on the management of smartphones data (location, schedule, e-mail information, etc.); the authors, after the illustration of the importance of big data management, show the way the data over more than one year has been collected and, based on it, they demonstrated that the proposed scheme can predict user location precisely, giving to mobile users some enhanced services (about location, torrential rain, train delays, traffic jams, etc.).

The article in (Chon et al., 2012) evaluates several mobility models (Markov and Next-Place of different orders) for regularity and predictability purposes. The carried-out empirical studies show the hegemony of location-dependent predictors; their performance can be enhanced with the exploitation of the adaptive use of mobility models and high-granularity data.

In (Cao et al., 2017; Yu et al., 2017), the concept of Location-Based Social Networks (LBSNs) and Next-Place is considered. In the former paper, the authors take into account the prediction of future check-in locations of mobile users by analyzing user mobility patterns (in terms of time periodicity, global popularity and user preference). The proposed algorithm can extract a set of predictive features, which are then classified, to predict fine-grained future movements. The prediction approach is based on the construction of a heterogeneous social network model. In the latter, instead, the authors propose a novel approach based on the activity pattern for location prediction: their idea is composed of two features. Firstly, the individual's next activity is determined by modelling user activity patterns; then, the next-place is predicted based on the next activity. Also, in this case, the authors demonstrated that their approach represents a robust scheme for predicting future places.

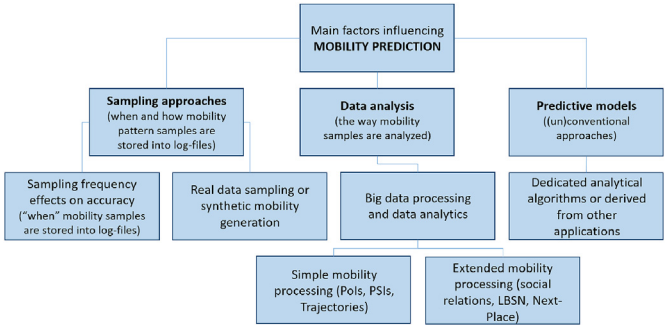


**Fig. 1.** Taxonomy of the main elements contributing to the accuracy/complexity of mobility prediction performance in mobile networks.

The work in (Wu et al., 2018) is strictly related to location prediction, social interactions and Points of Interest (POIs). The authors compared the last two aspects together with a two steps PSI model and a two-stage POIs clustering approach, to reduce the effects of randomness and to improve the overall performance of the prediction scheme. The paper illustrates several results, by which it can be understood how the PSI approach outperforms other predictive algorithms.

In (Li et al., 2019), the authors, face the issue of sparse individual trajectory data, which often results in a high error of prediction results. The proposed scheme is called Individual Trajectory-Group Trajectory (ITGT), and it is based on the pattern created by group travels. Different stages are considered, starting from a stay point extraction with spatial clustering, and different Markov models (PPM and PST) are then exploited to predict the clustering link. A massive amount of real data points have been used, and the obtained results confirmed authors expectations, with an accuracy of almost 90%.

The paper in (Katsikouli et al., 2017) describes an experimental analysis of the way a continuous human mobility pattern can be reconstructed after being sampled. The authors show the committed reconstruction error based on the sampling frequency, claiming that the inaccuracy grows of 1–4 m for each minute added to the sampling interval. The authors did not consider the effects of changing the sampling frequency on any predictive approach and the existing relationship among the collected samples.

The work in (Wu, 2018) represents a recent overview of different methods and approaches for predicting mobile trajectories, basing the choice of next places on mobility data. The paper, after an interesting introduction, describes the basic concepts of location prediction, including the different sources of trajectory data, the general prediction framework, challenges in location prediction, and common trajectory data preprocessing methods.

In the paper in (Suraj et al., 2016) the authors underline the importance of mobility prediction in routing operations for mobile networks. In particular the authors base their approach on the genetic theory, able to remove outliers on the basis of heuristics and parent selection. Numerical analysis demonstrates that, in general, a good accuracy level can be reached.

Fig. 1 shows the classification of the above mentioned works, identifying the main elements which contribute to the acccuracy of a prediction approach (model, algorithm, etc.). To the best of our knowledge and from the reading of the most recent papers on mobility prediction (as the ones described before), no works are focusing on giving a detailed analysis of the way the mobility process should be sampled for prediction purposes, neither from a temporal point of view (the sampling period/frequency, the time at which the next location/place/sample is needed) nor from a computational/space complexity point of view (the number of values or features to be stored and needed to build/train the prediction model). Referring to Fig. 1, our proposed idea belongs to the "Sampling Approaches", and the "Sampling effects on accuracy" are deeply studied.

Table 1 shows a comparison between the main features proposed in this paper and the ones of the reviewed articles (for (Wu, 2018) round brackets are used given that it represents a survey).

**Table 1**
A comparison of the key aspects of the various surveyed articles respect to our proposal.

| Features | Our Proposal | Zhao et al. (2015) | Yamada et al. (2018) | Chon et al. (2012) | Cao et al. (2017) | Yu et al. (2017) | Wu (2018) | Wu (2018) | Li et al. (2019) | Katsikouli et al. (2017) | Suraj et al. (2016) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Sampling Frequency Analysis | yes | | | | | | | | | yes | |
| Samples Correlation Analysis | yes | | | | | | | | | | |
| Fine-Granularity Prediction | yes | | | Yes | yes | yes | | (yes) | yes | yes | |
| Prediction Scheme Proposal | yes | yes | yes | Yes | yes | yes | yes | (yes) | yes | yes | yes |
| Mobility Trace Encoding | yes | | | | | | | | | | |
| Data Analytics Approach | yes | | yes | | yes | yes | | (yes) | yes | | |

## 3. Mobility prediction and related issues

First of all, the considered issue is introduced and, then, an in-depth analysis is carried out, to characterize each mobility process from a stochastic point of view. We start the description of our idea by considering a mono-dimensional moving space (finding only one mobility coordinate, i.e. a 1D space) and, then, the approach is generalized to 2D and 3D spaces, enhancing the approach by reducing the needed storing space (as explained in next sections). So, we will consider a generic mobile network, in which all the nodes are free to move among a given geographical region (e.g. Vehicular Ad-hoc NETworks - VANETs, Wireless Sensor Networks - WSNs, Mobile Ad-hoc NETworks - MANETs, etc.).

### 3.1. Mobility as an auto-regressive discrete process

As introduced in (Wu, 2018; Chaudhari and Biradar, 2016; Wang et al., 2019), each time a mobile network needs to know the future positions of mobile users, mobility patterns are stored in a centralized/distributed database. Then, a predictive model is built-up by analyzing the structure, and features of the stored patterns and as well as the model are exploited to obtain the so-called "next-places". This is the generic process of mobility prediction, and it is based on the sampling of the mobility pattern of each node (continuous in time), obtaining a sequence of coordinates. Here, we do not consider whether the observed mobility belongs to individual or group mobility. Consider a moving node $n$ on a 1D linear map, as depicted in Fig. 2. In this sub-section we will refer only to one mobility coordinate of the mobile nodes, then the approach is generalized for a 2D/3D environment.

Let us define $x_n(t)$ as the value of the $x$ coordinate of user $n$ at time $t$. It is a continuous function of time. We assume that the mobile network (or, directly, the mobile node $n$) is able to store $x_n(t)$ each $T$ seconds (sampling period): we indicate with $X_n(kT)$ the discretized version of $x_n(t)$, where $k$ is a positive and integer value (for $k = 0$ the sampling operation is started). The concept of discretization, here, is referred to the effects of sampling mobility trajectories. In this sense, the term $X_n$ can be considered as a random variable, defined on the space $\Omega \equiv \mathbb{R}$. After the collection of mobility samples, the vector $\vec{X}_n(T)$ is obtained, with $\|\vec{X}_n(T)\| = N$.

At this point, the content of $\vec{X}_n(T)$ should be analyzed to evaluate the potential relationship between $X_n(kT)$ and $X_n[(k - j)T]$, where $j$ is called *lag* and it is a non-zero integer value.

To this aim, we assume that $X_n(kT)$ is an Auto-Regressive process of order $j$, $AR(j)$. Additionally, we consider the AutoCorrelation Function (ACF) and the Partial ACF (PACF) (Bisgaard and Ankenman, 1996), as indexes of the correlation between two values of the process (Borrego et al., 2019). Thus, for the process $X_n(kT)$ the autocovariance (Cochrane, 1997; Hornik et al., 1989; Lee) at lag $j$ is defined as:

$$\gamma_j^{X_n} = \mathrm{Cov}(X_n(kT), X_n[(k-j)T]) = \ldots$$
$$\ldots = E[(X_n(kT) - \mu) \cdot (X_n[(k-j)T] - \mu)] \quad (1)$$

where $\mu$ is the mean of the process, i.e. $\mu = E[X_n(kT)]$, and the autocorrelation coeffcient at lag $j$ is:

$$\rho_j^{X_n} = \frac{\gamma_j^{X_n}}{\gamma_0^{X_n}} \quad (2)$$

where the autocovariance at lag zero $\gamma_0^{X_n}$ is the variance of the process. In the rest of the paper, the notations $X_n(kT)$ and $X_n$ are used as equivalent, as well as $X_n[(k - j)T]$ and $X_n(j)$. It is clear that, from

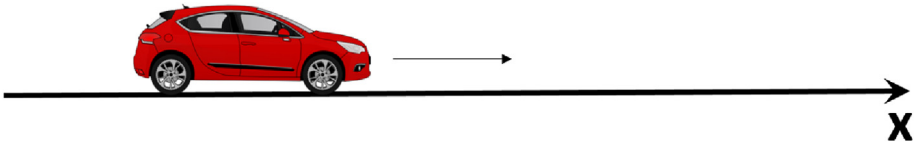


**Fig. 2.** An example of a vehicle moving along a linear path, with its position defined by $x(t)$.

**Table 2**
The average values of PACF for different pedestrian datasets of (Rhee, 2009).

| j | KAIST | NCSU | NY | ORLANDO | N.CAR. |
|---|---|---|---|---|---|
| 1 | −0.9933987 | −0.997281 | −0.99201195 | −0.96613169 | −0.98492919 |
| 2 | 0.15348976 | 0.150826 | 0.10655107 | 0.07390794 | 0.041613918 |
| 3 | 0.06507763 | 0.0854651 | 0.05102174 | −0.12090218 | 0.020630298 |
| 4 | 0.04794046 | 0.0567063 | −0.00138689 | −0.05234692 | 0.03178502 |
| 5 | −0.0734971 | 0.0330214 | −0.0094392 | −0.1179648 | 0.032317315 |
| 6 | 0.01185847 | 0.0418836 | −0.00484217 | −0.11924598 | 0.015642719 |
| 7 | −0.0688321 | 0.0233922 | −0.02560779 | −0.13593025 | 0.003799786 |
| 8 | −0.0316733 | −0.009589 | −0.04205781 | 0.04299811 | −0.00494047 |
| 9 | −0.0201564 | −0.007314 | −0.03827792 | −0.09701921 | −0.00490907 |
| 10 | −0.0369393 | −0.009194 | −0.03666558 | −0.11053815 | −0.01280945 |

the definition, the autocorrelation coefficient $\rho_K^{X_n}$ is dimensionless, so independent from the measurement scale, and it belongs to the interval $[-1, 1]$. From (Box et al., 2008), it is known that the term in eq. (2) is the theoretical ACF. A lag $j$ autocorrelation represents the relation between mobility values that are $j$ time periods apart. So, the ACF is a way to consider the linear relationship between a time instant $kT$ and all the previous process observations. As said before, in our work, we assume that node mobility can be modelled as a $AR(j)$ process. Still, we want to know which is the relation among $X_n(kT)$ and $X_n[(k-j)T]$, without considering the contributions of any intermediate terms $X_n[(k-1)T], \ldots, X_n[(k-j+1)T]$. Clearly, at lag 1, PACF(1) is the same as ACF(1). Following the theory in (Yule, 1927), to describe the expression for the PACF, we have to consider $j$ Yule-Walker equations (Yule, 1927) written for the $AR(j)$ process, and solve them for the $j$ variables $\varphi_{j1}, \ldots, \varphi_{jj}$. Typically they are written in a matrix form as follows:

$$\begin{bmatrix} 1 & X_n(1) & \ldots & X_n(j-1) \\ X_n(1) & 1 & \ldots & X_n(j-2) \\ X_n(2) & X_n(3) & \ldots & X_n(j-3) \\ \ldots & \ldots & \ldots & \ldots \\ X_n(j-1) & X_n(j-2) & \ldots & 1 \end{bmatrix} \cdot \begin{bmatrix} \varphi_{j1} \\ \varphi_{j2} \\ \varphi_{j2} \\ \ldots \\ \varphi_{jj} \end{bmatrix} = \begin{bmatrix} X_n(1) \\ X_n(2) \\ X_n(3) \\ \ldots \\ X_n(j) \end{bmatrix} \tag{3}$$

and the PACF will be represented by the $j-th$ solution $\varphi_{jj}$, a function of lag $j$.

Based on the discussion above, and as shown in the next sections, we can conclude that using the PACF instead of the ACF will lead us to obtain the values of $j$ for which the mobility samples are strictly correlated. So, the ACF and PACF are statistical measures that reflect how the observations of a process evolution are related to each other.

Based on the value of $j$, solving the system in eq. (3), by inverting the matrix, could result in $O(j^3)$ operations with many numerical problems (due to the limited accuracy of the processing unit). So we based our analysis, instead, on the Levinson-Durbin Recursion (LDR) approach (Hänsler, 2001), which represents a solving method able to exploit some properties of the matrix (such as the Toeplitz structure). It has been demonstrated that this algorithm has a computational complexity of $O(j^2)$ (Hänsler, 2001). An example of the results obtained by applying Eq. (3) and LDR to real traces is illustrated in Table 2.

In particular, we considered the datasets in (Rhee, 2009), consisting of human mobility traces in GPS format from five different sites: two university campuses (NCSU and KAIST), New York City, Disney World (Orlando), and North Carolina Raleigh (during the state fair event). As illustrated in the figure, for each column, the decay of PACF for higher values of $j$ (i.e. the higher distance between $X_n(kT)$ and $X_n[(k-j)T]$) becomes evident (only the absolute value is taken into account), so the influence of farther samples on the considered one becomes negligible. We will deep this aspect in the next sections, when the proper value of $j$ is discussed, considering different types of mobility, instead of the pedestrian one.

Foremost, we preliminarily provided to verify the results obtained in (Katsikouli et al., 2017) regarding the spectral content of $X_n(kT)$. In particular, considering $X_n(kT)$ as a discrete unidimensional signal, its spectrum can be easily obtained as follows (by applying the Discrete Fourier Transform - DFT):

$$X_n(f) = \sum_{k=0}^{N-1} X_n(kT) \cdot e^{-j\omega kT} \tag{4}$$

and the authors of (Katsikouli et al., 2017) affirm that the sampling frequency does not affect the spectral counterpart and the Power Spectral Density (PSD) of the collected samples. We also recall that:

$$PSD_{X_n}(f) = \mathcal{F}[\rho_j^{X_n}] \tag{5}$$

where $\rho_j^{X_n}$ is defined in 2 (Wiener–Khintchine theorem).

We referred to the previous traces, for which an observation window of the 30s has been considered, that is to say, each sample collection activity has a global duration of 30s. For the KAIST traces, the average number of samples $\overline{N}$ is 1608, for the NCSU traces $\overline{N} = 1431$, for the NY traces $\overline{N} = 1600$, for the Orlando, traces $\overline{N} = 1284$ and for the North Carolina traces $\overline{N} = 415$. We can conclude that the average sampling periods $\overline{T}$ are 18.65ms, 21ms, 18.75ms, 23.36ms and 72.29ms, respectively. We evaluated the spectral content for the whole datasets but, due to space limitations, we cannot show all the obtained PSD shapes: just, for example, Fig. 3 gives an idea of the PSD trend based on the sampling period. In particular one trace from KAIST dataset is shown for sampling frequencies $F_1 = 1/T$, $F_2 = 1/4T$ and $F_3 = 1/8T$. We observed the same trend for most of the traces, without noticing a considerable change in the PSD shape, confirming the results of the previous studies (Katsikouli et al., 2017).

At this point, we have to discuss the main aim of this sub-section: what happens to the PACF if we change the sampling period $T$ (or sampling frequency $F = 1/T$) for $\vec{X}_n(T)$. To this aim, we provided to extend the spectral analysis by investigating the effects of the sampling frequency on the PACF. In other words, the question is: to make a next location prediction with reasonable accuracy, is it necessary to sample

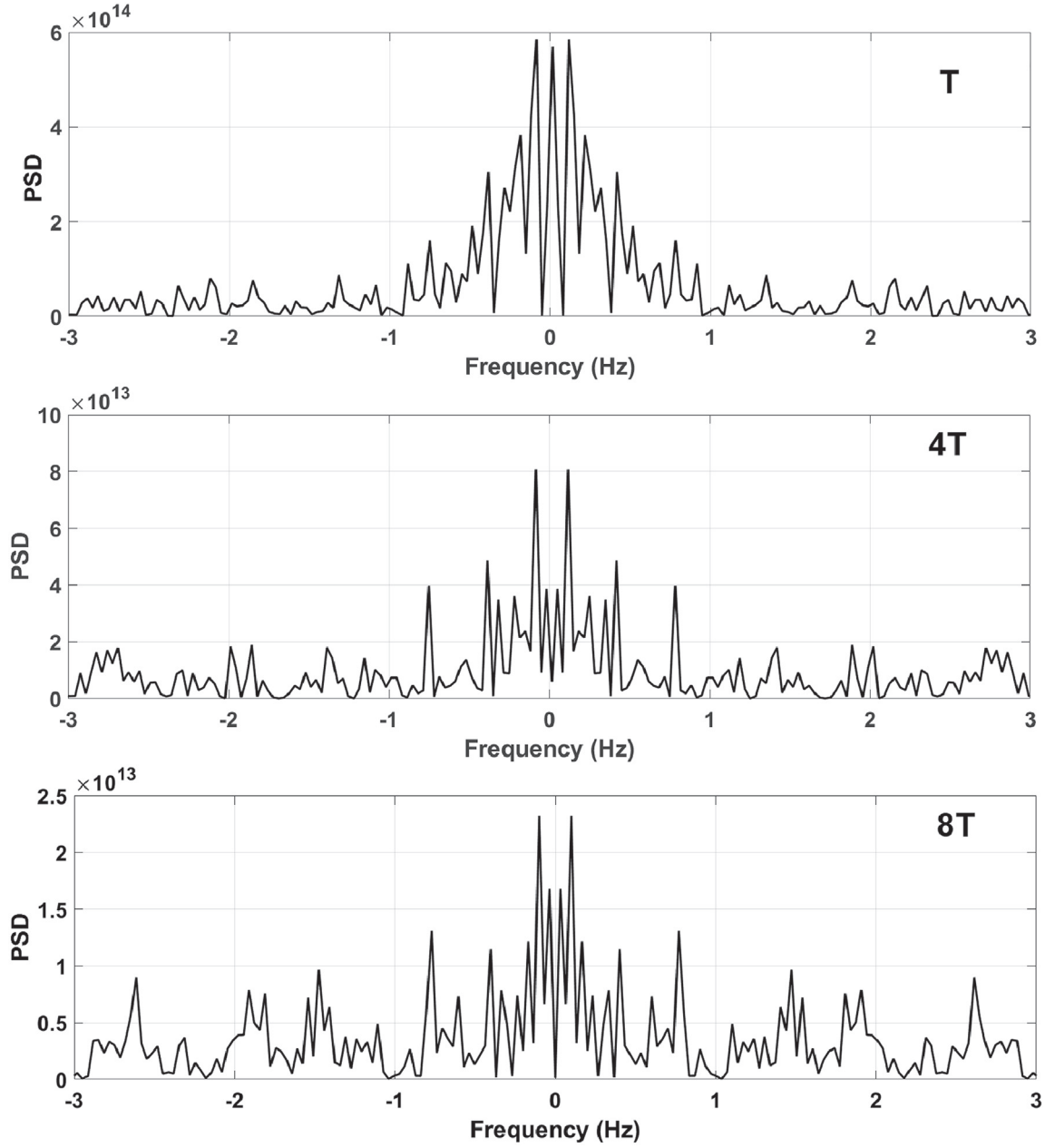


**Fig. 3.** Double-sided Power Spectral Density (PSD) of one trace from KAIST dataset with sampling periods T = 18.65ms, 4T = 74.60ms and 8T = 149.2ms.

mobility patterns so frequently? Table 3 illustrates the obtained PACF for different values of $j$ (ranging from 1 to 3) and different sampling periods from $T$ to $16T$ (considering only the x coordinate). The last column is the PACF averaged on the different traces, and the absolute values have been considered. The main experimental result which can be inferred is that the sampling frequency does not affect the nature of the AR process: for $j = 1$, the absolute value of PACF is always near to 1, while from $j = 2$ the magnitude of PACF becomes negligible.

Concluding this sub-section, we can state that the choice of a sampling frequency for storing a discretized mobility pattern affects neither the spectral content nor the correlation between any sample of the trace. In Section 4, we will deeply analyze these preliminary results on several mobility traces.

### *3.2. A possible criterion for choosing the sampling frequency aimed at prediction purposes*

After the description of the previous section, it is clear that a way for choosing a proper value of sampling frequency should be found. From one side, high sampling rates could be desired, due to the possibility of exactly capturing the dynamics of the system and create precise historical traces; from the other side, the amount of data to be stored should be minimized. Some examples of possible scenario for which sampling frequency becomes critical are the following ones:

- Energy-based computing: higher is the sampling frequency, higher will be the energy consumption;
- Protocol optimization: higher is the sampling rate, higher will be the overhead in a communication system, because of the huge amount of needed signalling messages;
- Sending probing messages for mobile polling devices: nodes are localized periodically for different purposes of network management and, also, in this case, the polling period becomes fundamental;
- Pattern compression: when mobility data is stored, it needs to be compressed to reduce the needed space; the number of collected samples will impact on the computational performance and the needed space;

**Table 3**
The average values of PACF for different sampling periods referred to the traces of (Rhee, 2009).

| j = 1 | KAIST | NCSU | NY | ORLANDO | N.CAR. | ABS-AVG |
|---|---|---|---|---|---|---|
| T | −0.99339 | −0.99728 | −0.992011 | −0.966 | −0.9849 | 0.98672 |
| 2T | −0.9972 | −0.9987 | −0.9963 | −0.9853 | −0.969 | 0.9893 |
| 4T | −0.9934 | −0.9973 | −0.992 | −0.9661 | −0.9359 | 0.97694 |
| 8T | −0.9842 | −0.9932 | −0.9825 | −0.9397 | −0.863 | 0.95252 |
| 16T | −0.9644 | −0.9851 | −0.9588 | −0.8724 | −0.7655 | 0.90924 |
| j = 2 | KAIST | NCSU | NY | ORLANDO | N.CAR. | ABS-AVG |
| T | 0.15348 | 0.150826 | 0.10655 | 0.0739 | 0.04161 | 0.10527 |
| 2T | 0.1475 | 0.0785 | 0.0749 | 0.3532 | 0.0544 | 0.1417 |
| 4T | 0.1535 | 0.1508 | 0.1066 | 0.0739 | 0.0921 | 0.11538 |
| 8T | 0.1093 | 0.1353 | 0.1083 | −0.1078 | −0.0125 | 0.09464 |
| 16T | −0.0254 | 0.2112 | −0.0452 | −0.3273 | −0.443 | 0.21042 |
| j = 3 | KAIST | NCSU | NY | ORLANDO | N.CAR. | ABS-AVG |
| T | 0.06507 | 0.085465 | 0.051021 | −0.1209 | 0.02063 | 0.06862 |
| 2T | 0.0659 | 0.0472 | 0.052 | −0.3091 | 0.0612 | 0.10708 |
| 4T | 0.0651 | 0.0855 | 0.051 | −0.1209 | −0.09 | 0.0825 |
| 8T | −0.007 | 0.0897 | −0.0267 | −0.175 | −0.0739 | 0.0746 |
| 16T | −0.0963 | 0.0441 | −0.1571 | 0.2714 | −0.2028 | 0.15434 |

- Trajectory prediction in dynamic networks: mobile nodes frequently cause changes to the network topology; it is often desired to know which will be the future position of a node (a relay node for example), to take into account its stability or its contribution to the overall path duration. Based on node speeds, the sampling frequency should be tuned, to make adequate in-advance location predictions.

The concept of sampling period/frequency can be applied in different scenarios, and each one has its features and needings. In this subsection, then, we propose a general approach for choosing a proper value of the sampling period, which can be particularized for the desired scenario.

So, let us consider two sampling periods $T_1$ and $T_2$, with $T_1 < T_2$ and, without loss of generality we assume that $T_2 = l \cdot T_1$, with $l$ a positive integer. We can define $T_1$ as a fine-grained sampling period, while $T_2$ as a coarse-grained sampling period. Without considering any particular prediction scheme (see, for example (Fazio et al., 2017), for a complete description of the main algorithms for mobility prediction in cellular networks), let $P(j)$ be a generic j-th order predictor. So, we can write that, in the case of fine-grained sampling, the next process value at $(k+1)T_1$ is:

$$X_n^*[(k+1)T_1] = g_{P(j)}\{X_n[(kT_1)], X_n[(k-1)T_1], \ldots \\ \ldots, X_n[(k-j)T_1]\} \tag{6}$$

that is to say, after the collection of the last $k-th$ sample, the next process value at $(k+1)T_1$ is a function $g$ (depending on the predictor $P$) of the previous $j$ samples, each one collected every $T_1$ seconds. We use the notation $X_n^*$ for indicating a predicted sample. If another next value is needed, then for the $(k+2)-th$ sample we will have:

$$X_n^*[(k+2)T_1] = g_{P(j)}\left\{X_n^*[(k+1)T_1], X_n[(kT_1)], \ldots \\ \ldots, X_n[(k-j+1)T_1]\right\} \tag{7}$$

where the previously predicted $(k+1)$-th sample is needed to collect the $j$ samples (as the order of the exploited predictor). So, in general, if we need to predict $h$ next samples (with $h < j$) starting from the $k$-th one, we can write:

$$X_n^*[(k+h)T_1] = g_{P(j)}\left\{X_n^*[(k+h-1)T_1], \ldots \right. \\ X_n^*[(k+h-2)T_1)], \ldots, X_n[(kT_1)], \ldots, X_n[(k-1)]T, \ldots \\ \left. \ldots X_n[(k-j+h-1)T_1]\right\} \tag{8}$$

where the samples from the $(k-j+h-1)$-th to the $k$-th represent the real history, while the samples from the $(k+1)$-th to the $(k+h-1)$-th have to be predicted previously. This is the only way

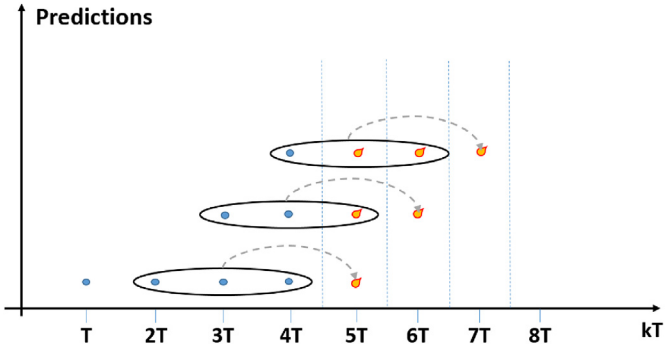


**Fig. 4.** An example of the application of equations (6) and (7), with $j = 3$, $T_1 = T = 1s$. Starting from the bottom row, it can be seen that, at $t = 4s$, for the sample at $t = 5s$ the previous three samples are used. In the middle row, instead, at $t = 4s$ for predicting the sample at $t = 6s$ we need to consider the previously predicted sample at $t = 5s$, because we still do to know the real sample at $t = 5s$. The same if we need to know the sample at $t = 7s$ (top row).

a $j$-th order predictor can be exploited when more than one prediction is needed. Numerically, if $T_1 = 1s$ and the last $k$-th sample has been collected at $t = 4s$, then the next $(k+1)$-th sample will be predicted for $t = 5s$. If we need at $t = 4s$ to know the process value at $t = 6s$, then we have to apply equations (6) and (7) iteratively as in 8, using some already predicted samples to predict the 6-th one (the number of already predicted samples depend on $j$). Fig. 4 illustrates this concept graphically, for a predictor P(3), hence $j = 3$.

Clearly, fine-grained sampling leads to a prediction error:

$$\Delta e(k+1) = |X_n^*[(k+1)T_1] - \overline{X}_n[(k+1)T_1]|^2 \tag{9}$$

where $\overline{X}_n[(k+1)T_1]$ is the real process value at time $(k+1)T_1$. In the case of multiple predictions we have:

$$\Delta e(k+h) = \sum_{i=1}^{h} \Delta e(k+i) = \sum_{i=1}^{h} |X_n^*[(k+i)T_1] + \ldots \\ \ldots - \overline{X}_n[(k+i)T_1]|^2. \tag{10}$$

In the case of coarse-grained sampling, we use the period $T_2 = l \cdot T_1$ and equations (6)–(10) remain the same (just substituting $T_1$ with $T_2$). At this point, if we want to know the prediction error after $h \cdot T_1$ amount

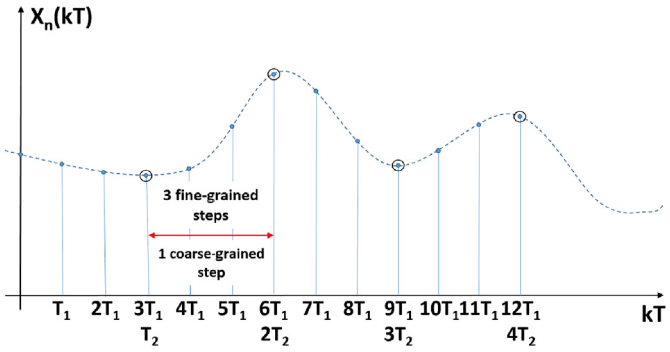


**Fig. 5.** An example of fine-grained sampling (period $T_1$) and coarse-grained sampling (period $T_2$), with $l = 3$, $T_2 = 3 \cdot T_1$.

of time, the two expressions for fine and coarse samplings are:

$$\Delta e_{fine}(k+h) = \sum_{i=1}^{h} \Delta e_{fine}(k+i)$$
$$\Delta e_{coarse}(k+1) = |X_n^*[(k+1)T_2] - \overline{X}_n[(k+1)T_2]|^2 \quad (11)$$

that is to say for knowing the predicted value of the process after $h \cdot T_1$ amount of time, in the case of fine sampling $h$ steps are necessary (and the global error will be the sum of the $h$ error terms), while for the coarse sampling only one step is enough (and the global error will be equal to the single prediction error). Fig. 5 better illustrates the relationship between fine and coarse sampling (with $l = 3$).

So, the choice of $l$ (and, hence, of the sampling period) strictly depends on the accuracy of $P(j)$ and, also, on the intrinsic dynamic of the considered system. If we are dealing with a periodic system with period $T_i$ (let us think, for example, to the triggered updates in a routing process, or the beaconing signalling in a mobile network, etc.), then it is desired to have next predicted values at a time that is at least $T_i$ far away. In the following sections, we will deeply analyze what happens to the prediction error when we deal with mobile nodes, and the proper value of $l$ will be discussed.

### 3.3. Correlating the spatial coordinates with one value: the pairing functions (PFs)

In this sub-section, we generalize our approach (until now it has been referred only to one coordinate), introducing a correlation between the two spatial coordinates $x$ and $y$ (it is also possible to extend the analysis to the third variable $z$, as explained later). As already mentioned, most of the existing works do not account for the intrinsic correlation between the spatial component of a 2D space (we consider the dimensions up-to $\mathbb{R}^2$). To introduce the two spatial components, we will use the notations $X_n(kT)$ and $Y_n(kT)$ to indicate the $x$ and $y$ coordinates respectively. All the equations defined before are still valid for the other mobility components.

One way to proceed is represented by studying the two components separately, but there are many disadvantages to this kind of approach:

- One point on a surface is characterized by $m$ components (with $m = 2$), and each node in a mobile network moves in the considered geographical region by coordinating all the spatial variables; besides, based on the considered scenario (a street, a road, a highway, a free-space, etc.) a node moves by respecting the environmental constraints; so, analyzing the individual process, independently from the other ones, leads to the definition of some models which may leak some precious information;
- If we deal with $m$ separate processes, it means that we have to consider $m$ different historical traces, $k$ distinct predictors ($P_x(i)$ and $P_y(j)$ in the case of $\mathbb{R}^2$) and we have to make $m$ independent predictions, each time a next-place is needed. It is evident that this kind of approach is not efficient from a computational point of view;
- The available space needed to store $m$ traces may be scarce.

For the reasons above, we studied a way to encode the sequence of the $m$ samples into only one value: in this way, only one trace is needed, only one prediction needs to be made, and once the next-value is obtained, it can be decoded back into the $m$ components. To do this, we based our approach on the Pairing Functions (PFs) (Krishna et al., 2016; Szudzik, 2017). The concept of PF is briefly explained below and some PFs are introduced for encoding the content of the trace files.

A PF defined on a set A relates each pair of members from A with a single member of A, and any two distinct pairs are associated with two distinct members. In this way, it is possible to encode a couple of values with a single one and, then, decode the original values when needed. A PF is generally indicated as a function $pf : A^m \rightarrow A$, and they are used in a wide variety of applications (renderers, shaders, theoretical computer science, etc.). We will indicate with $pf^{-1} : A \rightarrow A^m$ the inverse PF function to decode back the $m$ values (it is also called unpairing function). Many PFs have been defined in literature (Wolfram and Gad-el-Hak, 2003): their study and evaluation are out of the scope of this paper, while the main aim of this sub-section consists in the application of some PFs to encode mobility traces. Cantor's PF is the most known (Wolfram and Gad-el-Hak, 2003), defined as a bijection $\mathbb{N}^2 \rightarrow \mathbb{N}$:

$$pf_{Cantor}(x,y) = \frac{(x+y)\cdot(x+y+1)}{2} + y \quad (12)$$

but it has been demonstrated that it has limitations in terms of value packing efficiency. For example, if we set $x = 8$ and $y = 8$ we would expect to obtain a maximum of 81 as a result (given that two digits 0–8 and 0–8 can create only 81 combinations), but $pf_{Cantor}(8,8) = 144$, with an efficiency of only 56%. This result can be significantly improved by Szudzik's PF, also defined as a bijection $\mathbb{N}^2 \rightarrow \mathbb{N}$:

$$pf_{Szudzik}(x,y) = \begin{cases} x + y^2 & x < y \\ x^2 + x + y & x \geq y \end{cases} \quad (13)$$

with $pf_{Szudzik}(8,8) = 80$. Among the wide variety of existing PFs, we based our approach on the one defined in eq. (13). It is just an example of the way a PF can be exploited to optimize the analysis of mobility traces.

At this point, we have to see if and how a PF can be adapted to our scopes. There are some drawbacks which have to be taken into account:

- Natural numbers: PFs are defined from $\mathbb{N}^2$ to $\mathbb{N}$, so the real numbers are not considered (PFs are polynomial functions, and no continuous bijections are possible for $\mathbb{R}^2$ and $\mathbb{R}$ (Brouwer, 1912)). To overcome this issue, different ways can be chosen.
  - Mobility trace values can be quantized: given a geographical region in which nodes are moving, it is easy to derive the minimum and maximum extension of $x_n(t)$ and $y_n(t)$ and, after the sampling operation, values can be quantized, after setting a proper resolution;
  - A second possible solution is represented by the approximation of mobility values. In some cases, depending on the used mobility format, the decimal part can be neglected (e.g. planar coordinates), or any value can be transformed into an integer one by a multiplication factor (e.g. GPS coordinates).
- Non-negative numbers: mobility traces often contain negative values which do not belong to $\mathbb{N}$; they can be converted into integers, but no operations can be made on the sign. Also, in this case, we have some solutions:
  - x and y values can be translated to move the origin of the reference system;
  - PF functions can be transformed to account for negative integers.
- Extension to the third coordinate: with the proliferation of the Unmanned Aerial Vehicles (UAVs), for example, the third coordinate assumes critical importance (ur Rahman et al., 2018), differ-

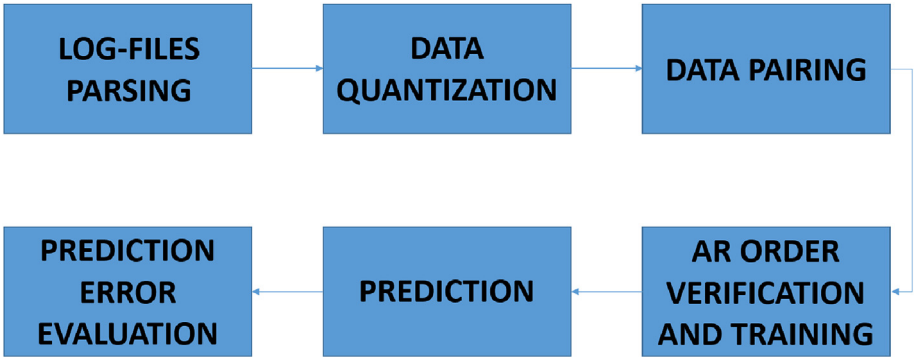


**Fig. 6.** The logical structure of the MATLAB testbed.

ently from classical approaches for which the mobile nodes belong to a planar geographical region. In this paper, we consider mainly vehicular mobility (so a 2D region is enough), but the approach can be extended to UAVs or other technologies.

- The first greedy approach is to consider a pairing of pairing: given $x$, $y$ and $z$ coordinates, we can evaluate $a = pf(x, y)$ and $b = pf(a, z)$ so the prediction analysis will be made only on $b$;
- A second alternative is represented by the exploitation of encoding techniques different from PFs (e.g. bit-interleaving, etc.). We will consider this research topic in future works.

For the Szudzik PF in eq. (13), the unpairing function is defined as follows (for Cantor's unpairing function please refer to (Wolfram and Gad-el-Hak, 2003) and (Cantor, 1878)):

$$pf^{-1}_{Szudzik}(a) = \begin{cases} a - \lfloor\sqrt{a}\rfloor^2 & x \\ \lfloor\sqrt{a}\rfloor & y \end{cases} \tag{14}$$

if $x < y$, or

$$pf^{-1}_{Szudzik}(a) = \begin{cases} \lfloor\sqrt{a}\rfloor & x \\ a - \lfloor\sqrt{a}\rfloor^2 - \lfloor\sqrt{a}\rfloor & y \end{cases} \tag{15}$$

else. Given its higher efficiency, we will consider Szudzik's PF in the following, with the assumptions:

- Trace values are firstly rounded to integers, and the next section gives the details about this approach;
- Negative numbers are taken into account by applying the following transformation to the expression in Eq. (13):

$$c = \begin{cases} -2x - 1 & x < 0 \\ 2x & x \geq 0 \end{cases} \tag{16}$$

$$d = \begin{cases} -2y - 1 & y < 0 \\ 2y & y \geq 0 \end{cases} \tag{17}$$

and evaluating $pf_{Szudzik}(c, d)$.

In the next section, some numerical results are obtained, showing the possible reachable results which can be reached by considering the approaches proposed in this section.

## 4. Simulation results and analysis

To test and verify concepts illustrated before, a MATLAB testbed has been setup. Different functions have been defined for analyzing and characterizing the downloaded data in terms of quantization, AR processes, pairing functions, etc. The considered traces are the ones linked in (Dias, 2018) for the buses mobility and (Bracciale, 2014) for taxis mobility. Pedestrian log files (Rhee, 2009) are also found for comparison purposes. The main steps of our testbed are illustrated in Fig. 6 and the following subsections will describe them in detail.

### 4.1. Integer values of mobility trace files

Foremost, for applying the concepts related to PFs as in subsection 3.3, the values of the trace-files should belong to $\mathbb{N}$ (or to $\mathbb{Z}$ if negative values will be taken into account). In general, the content of the downloaded files contains real values, so we decided to transform them into integer values by finding a proper multiplying factor (integer values can be obtained also by different transformations approaches, as in (Hernández-Orallo et al., 2018), where the authors consider the space-time discretization in opportunistic mobile networks). For the BUS and TAXI cases, all the coordinates format, for each processed log-file, follow the Decimal Degree (DD) representation, based on the WGS84 standard. For the PEDESTRIAN case, they are just cartesian values referred to a particular reference point and expressed in meters.

In the case of bus traces (Dias, 2018), each row is formatted as *date*, *time*(24*hformat*), *busID*, *busline*, *latitude*, *longitude*, *speed* and $T_{buses} = 1s$. An example of entry extracted from a BUS downloaded log-file is the following one: [10-01-2014, 00:00:01, A48177, 0, −22.924088, −43.255466, 0.19], which corresponds to a point on R. Barão de Mesquita, 916–928 - Tijuca, Rio de Janeiro - RJ, 20540-004, Brasil.

In particular, the *latitude* and *longitude* values belong to $\mathbb{R}$ with four decimal digits, so we converted them to $\mathbb{Z}$ values with a factor of $10^4$. Fig. 7 illustrates an example of bus pattern: in the upper part the trends of $x$ and $y$ in the function of time are shown ($T = 1s$) and the typical "bus periodical loop" trend can be observed. The almost-same pattern is repeated (every 75 s) and, in the end, the bus goes to a dedicated parking lot. The complete pattern in a 2D space can be observed at the bottom of the figure.

When dealing with taxi traces (Bracciale, 2014), each row is formatted as *taxiID*, *date*, *time*(24*hformat*), *latitude*, *longitude*, $speed_x$, $speed_y$, $speed_z$ and $T_{taxi} = 7s$. An example of entry extracted from a TAXI downloaded log-file is the following one: [86349, 2007-02-20, 00:02:48, 41.9057, 12.4821, 56, 67,0], which corresponds to a point on Via dei Condotti, Rome, Italy.

Also in this case, a conversion factor of $10^4$ is enough to obtain integer values. Fig. 8has been added just to give a qualitative illustration for the trend of the $x$ and $y$ coordinates.

In the case of pedestrian traces (Rhee, 2009), each row of the trace files is simply formatted as *ID*, *time*(24*hformat*), *coordX*, *coordY* and $T_{pedestrian} = 50$ ms. This time, as indicated in the specifications of (Rhee, 2009), the coordinates are not expressed as GPS or DD/WGS84 values, but they are simply cartesian values referred to a reference point. We assumed, also in this case, that a factor of $10^4$ is enough to represent the cartesian values as integer ones. Fig. 8 illustrates an example of taxi pattern: in the upper part, the trends of $x$ and $y$ in the function of time are shown ($T = 7s$). The complete pattern in a 2D space can be observed at the bottom of the figure. The same illustration is shown in Fig. 9 for a pedestrian trace.

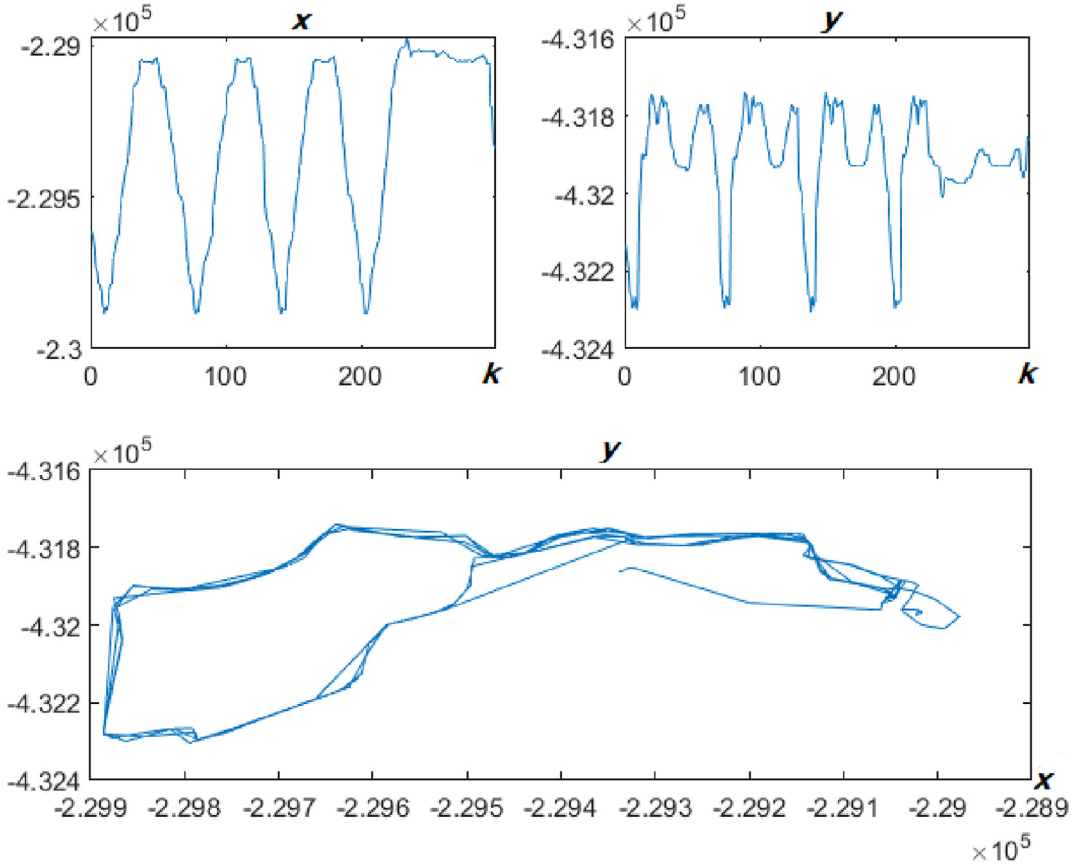


**Fig. 7.** The trend of x and y coordinates in the case of bus traces. The bottom plot represents the complete pattern in a 2D space.

### 4.2. The PACF in the function of the sampling period and PF

In this subsection, the primary analysis in terms of ACF, PF and $T$ is carried out. First of all, both coordinates are analyzed separately; then the PF is used. A comparison is made to see if some differences are encountered in the obtained stochastic properties.

The PACF has been derived for the three types of mobility traces, by applying the LDR approach (Hänsler, 2001) on the expression of eq. (3) to $x$ and $y$ coordinates separately. Assuming that the mobility process can be considered as an $AR(j)$ process, the trend of the PACF has been observed for different cases, as illustrated in the following figures.

Different values of lag $j$ have been considered, from $j = 1$ to $j = 60$ in the case of bus and taxi mobility, and from $j = 1$ to $j = 100$ for the pedestrian mobility. In fact, from Fig. 10 and Fig. 11 it can be seen that for $j = 1$ the absolute value of PACF assumes the maximum value for all three cases (both for $x$ and $y$), while for $j > 1$ it has a flat trend (zero) for buses and taxis, and near-to-zero for pedestrian mobility.

Given that the previous figures show the result of only three traces (one for a taxi, one for bus and one for pedestrian mobility), we provided to carry on the same analysis on a massive set of mobility traces, obtaining some interesting results. In particular, we considered 4300 Taxi traces, 19000 Bus traces and 700 Pedestrian traces (from NCSU and KAIST campuses, New York, Raleigh and Orlando cities). After a preliminary parsing stage in MATLAB, we provided to observe the trend of PACF, obtaining the results illustrated in Table 4.

In particular, we considered the first values of $j$ (from 1 to 5) to see that $PACF(1)$ always has a "near-to-one" value. From $j = 2$ to $j = 5$ the obtained PACF values are negligible for "high speed" mobilities (Taxi and Buses), while for pedestrian mobility for $j = 2$ the PACF has a still comparable value with $PACF(1)$. For larger lags ($j = 3$, 4 or 5) the memory effect goes vanishing, with negligible values of PACF. These results are valid for both $x$ and $y$ coordinates. At the moment, these results suggest choosing an $AR(1)$ model for "high-mobility" environments, while for pedestrian scenarios at least an $AR(2)$ model should be considered.

It should be underlined that, until now, we considered the "native" sampling period $T$ of the trace files, that is $T_{Taxi} = 7$s, $T_{Bus} = 1$s, and $T_{Pedestrian} = 50$ms in the average. Let us consider, now, what happens to the studied parameters (PACF values) when $T$ is changed, by choosing a new value of sampling period $T^{new} = l \cdot T$.

In particular, for Taxi traces we considered $l = 2, \ldots, 10$ so a new sampling period $T^{new}_{Taxi}$ ranging from 14$s$ to 70$s$ (larger values are not useful because vehicle movements would be observed very rarely), for Bus traces $l = 2, \ldots, 7$ with a new sampling period $T^{new}_{Bus}$ ranging from 2$s$ to 7$s$ (the number of samples contained into the trace files did not permit us to extend the sampling period further). For Pedestrian mobility, we first considered $l = 2, \ldots, 10$ and, then, additional analysis would be shown.

Fig. 12 shows the first significant result derived from our analysis. Increasing $T$ will influence the $lag = 1$ correlation between samples: there is a decreasing trend (almost linear for "high speed" mobility) for the average PACF(1). For those traces, the trend is more regular, while for Pedestrian curves, the effects of a higher sampling frequency are noticeable (from 10 Hz to 2 Hz). Besides, when sampling more rarely, the dynamics of the coarse-grained mobility process are reflected in each period, creating an additional correlation between farther times. We do not illustrate the variance of $PACF(j)$ trends, given that it assumes always near-to-zero values (typical variance values range from $5 \cdot 10e^{-8}$ to a maximum of $8 \cdot 10e^{-4}$ for "high-mobility" traces. For Pedestrian samples, we noticed higher variance values (with an average of $4 \cdot 10e^{-2}$). Concluding the analysis of the curves, we can say that the decreasing is almost negligible for "high mobility" pro-

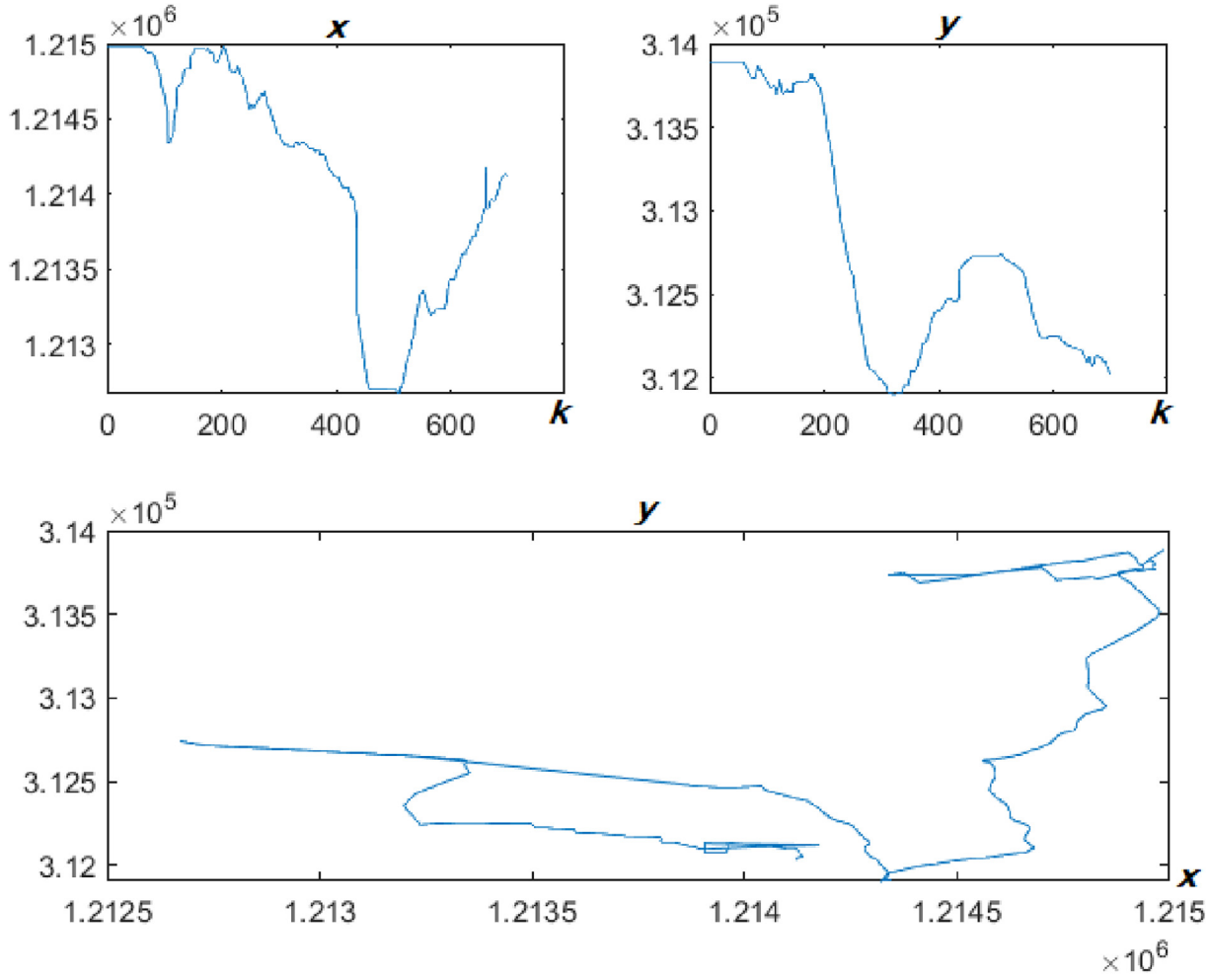


**Fig. 8.** The trend of x and y coordinates in the case of taxi traces. The bottom plot represents the complete pattern in a 2D space.

cesses, while for "slow mobility" traces (Pedestrian ones) it is more evident: in the last case, the order of any predictor should be increased if the phenomenon is observed more rarely.

To show the other results, obtained for different values of $j$, for space limitations, we illustrate only the behaviour of $PACF(3)$ in Fig. 13, given that the other curves have a very similar trend. The trend is exactly the opposite of the previous one: it is increasing, but almost negligible for "high-mobility" traces.

From the analysis above, we can conclude that larger is $T$, more substantial is the correlation lag $j$, which should be chosen for any predictor: this effect depends on the moving speed. For "low-mobility" samples, the influence is stronger than in the case of "high-mobility" traces.

### 4.3. Pairing results: numerical analysis

At this point, the next step is represented by the approach for avoiding a separate analysis of $x$ and $y$ coordinates, i.e. the use of PFs. First of all, we need to verify if the PACF trend of the paired process is similar to one of the independent variables.

For example, Fig. 14 shows the trend of 1800 samples of a Taxi mobility trace: in the upper part, the separate trend of $x$ and $y$ coordinates is visible, while on the bottom, the *Szudzik′spaired* trace is obtained (equations (13), (16) and (17) have been used). At this point, all the mobility traces were paired, and the obtained results are illustrated in the next figures. In particular, in Table 5 the obtained values of PACF for different lags $j$ are shown in terms of mean and variance. It shows that for $j = 1$ the correlation is always near to 1, while for $j > 1$ the other PACF values are negligible, except for Pedestrian mobility (in this case also the variance is not near-to-zero).

Fig. 15 depicts effects of the sampling period $T^{new}$ on the PACF(1) of the paired traces. Comparing it with the PACF(1) of the unpaired traces (Fig. 12), it can be seen that the trend is the same, with the advantage of analyzing only one variable (instead of two, or three in the case of 3D traces).

Just for completeness, PACF(3) is also shown in Fig. 16, showing the same trend of the previous case (Fig. 13), that is a negligible increase for Taxi and Bus mobility and noticeable improvement for the Pedestrian case. So, after the pairing of the mobility trace samples, we can confirm the same trend of the unpaired case: for larger values of $T^{new}$, the partial autocorrelation decreases for lag 1 ($j = 1$), while for the other values of $j$ it increases (in a negligible way for Taxi and Bus traces).

We can conclude that PFs are useful for storing mobility samples, they preserve the order of the $x$ and $y$ processes (in general also for the $z$ coordinate), reducing the time complexity by giving the possibility to analyze only one process instead of two or three. The price is paid in terms of the needed space to store the numbers (more bits are necessary); from Fig. 14 we can see that, for example, the samples go from the order of $10^4$ to the order of $10^8$, due to the power functions introduced by the chosen PF.

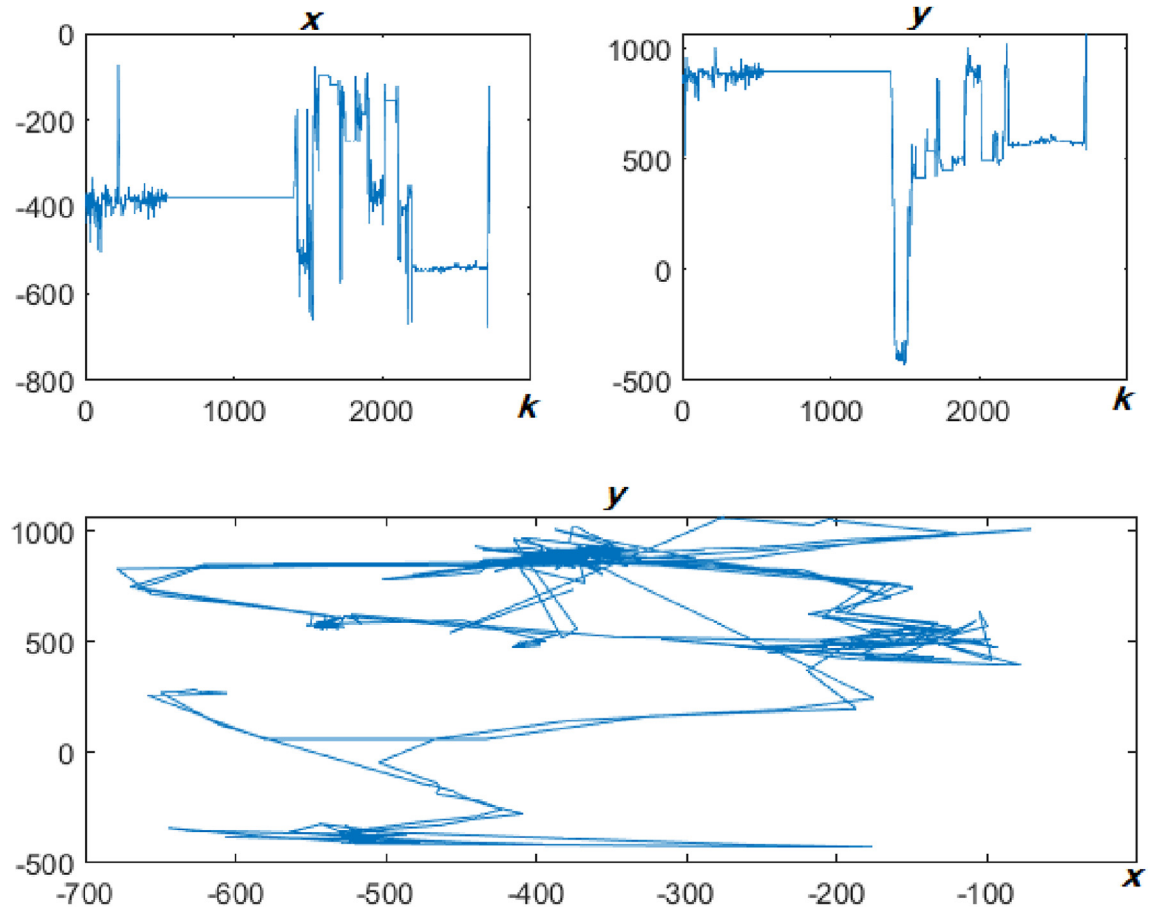


**Fig. 9.** The trend of x and y coordinates in the case of pedestrian traces. The bottom plot represents the complete pattern in a 2D space.

### 4.4. Discussion, use-cases and performance

The elements collected until now from our discussion are the following one:

- It is possible to pair each couple (or triplet) of coordinates and encode them into only one value;
- By considering PF equations, the stochastical properties of the paired processes are the same as the original uncoupled ones;
- The order 1 ($j = 1$) relationship between samples decreases by increasing the sampling period (or, equivalently, by decreasing the sampling frequency);
- The order $n$ ($j$ = n, with $n > 1$) relationship between samples is directly proportional to the sampling period.

The main question is: how can we choose the sampling period with the knowledge above?

It depends on the *observation granularity* we are interested in, that is to say (let us consider some particular scenario as examples):

- In a distributed network in which a proactive routing protocol is carrying on relay operations, the update interval is set to $T_{upd}$, generally its value is around 15s; so, each $T_{upd}$ amount of time, routing table entries are exchanged between nodes, leading to the building of the optimal routes from source to destination nodes. If a predictive approach is integrated with the routing layer, for example by considering future link stability (Fazio, 2016) and/or future node positions (for future topology evaluation) at $t = T_{upd}$, it is important to know in advance what will be the situation at $t = n^*T_{upd}$ (with $n > 1$). To this aim, we have to choose the sampling period $T$ to have enough samples to predict the future values at the next step. We can set $T < T_{upd}$, for example, $T = T_{upd}/2$ or $T = T_{upd}/4$, based on the needed lag $j$, that is the number of needed samples to perform a correct prediction. So it depends on "how far" we have to predict future mobility samples;
- In a cellular network, mobile nodes move among different coverage areas, each one served by a centralized device (access points, base stations, etc.). It is beneficial to a-priori know which cell a mobile node will visit during the call lifetime (Fazio, 2016) to be able to reserve the right amount of bandwidth when needed. Also, in this case, it is important to know the time after which next positions should be predicted, generally in correspondence of the hand-over events. The considered parameter is the so-called, Cell Stay Time (CST) or Cell Residence Time (CRT), which indicates how often the coverage between two cells is exchanged, and the needed bandwidth is requested. From our previous studies (Rango, 2009), it is known that the CST/CRT depends on different parameters (e.g. coverage radius, average speed, mobility model, etc.), ranging from few tens of seconds up to few hundreds of seconds.

In both cases, if we need, for example, to know the future sample every 100 s, then is it useful to have one sample every 100 s? Or is it better to have one sample each $j$ seconds and make $100/j$ predictions to obtain the 101-th value (where $j$ is the predictor order)? We will now illustrate which kind of results can be reached by following different ways. We would like only to underline that the proposal of a new prediction approach is out of the scope of this paper. So, we are not focusing on a particular prediction algorithm or model (markovian, neural, Bayesian, Kalman, etc.): we use an $AR(j)$ model, defined on the collected mobility samples (Taxi, Bus and Pedestrian) as a valid way to obtain some information about the prediction error. From the previ-

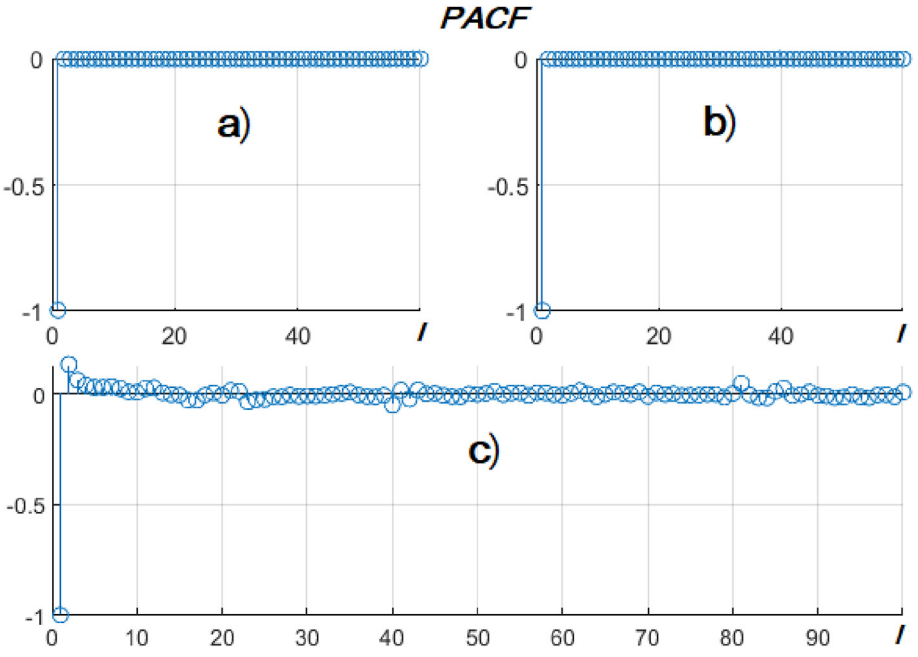


**Fig. 10.** PACF values of the $x$ coordinate for the a) Bus, b) Taxi, c) Pedestrian mobility patterns.

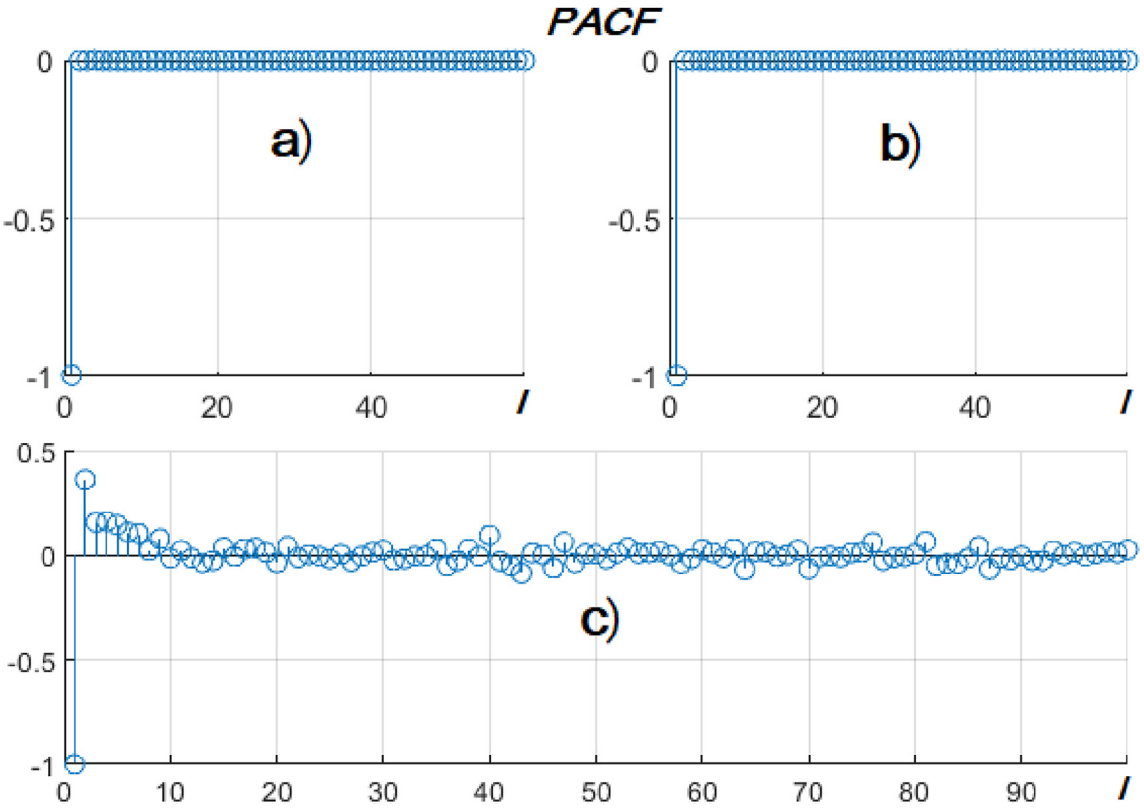


**Fig. 11.** PACF values of the $y$ coordinate for the a) Bus, b) Taxi, c) Pedestrian mobility patterns.

ous sub-section, we know that $j = 1$ is enough for Taxi and Bus traces (independently from the sampling period), while for Pedestrian mobility $j > 1$ should be considered (especially for larger sampling periods). However, we show general trends to make some comparisons. In particular, we paired the coordinates with the Szudzik's function; we applied an $AR(j)$ predictor to the paired coordinates. We unpaired the predicted samples $x^*$ and $y^*$ and compared them with the original ones $x$ and $y$, obtaining the error percentage with the following equation:

$$e = 0.5 \cdot [|((x^* - x)/x)| + |((y^* - y)/y)|] \tag{18}$$

In Fig. 17, each point represents the average error committed to predicting the next mobility point (1-step), given the knowledge of $j$ previous samples. A sequence of 50 predictions has been considered for different Taxi traces and different $AR(j)$ predictors, with $j = 1..15$. It

**Table 4**
PACF statistical parameters (mean value $\mu$ and variance $\sigma^2$) for $x$ and $y$ coordinates for Taxi, Bus and Pedestrian (PADE) trace files.

| | x | TAXI | y | |
|---|---|---|---|---|
| j | $\mu$ | $\sigma^2$ | $\mu$ | $\sigma^2$ |
| 1 | 9.99e-01 | 534e-07 | 9.99e-01 | 5.34e-07 |
| 2 | 4.78e-04 | 2.37e-07 | 6.94e-04 | 1.32e-05 |
| 3 | 4.73e-04 | 1.85e-07 | 5.88e-04 | 4.54e-06 |
| 4 | 4.47e-04 | 1.83e-07 | 3.70e-04 | 2.75e-06 |
| 5 | 4.68e-04 | 1.59e-07 | 5.36e-04 | 1.62e-06 |
| | x | BUS | y | |
| j | $\mu$ | $\sigma^2$ | $\mu$ | $\sigma^2$ |
| 1 | 9.41e-01 | 5.22e-06 | 9.41e-01 | 6.72e-06 |
| 2 | 3.00e-03 | 2.52e-05 | 3.10e-03 | 4.86e-05 |
| 3 | 2.90e-03 | 7.46e-06 | 2.92e-03 | 1.51e-05 |
| 4 | 2.90e-03 | 3.09e-06 | 2.90e-03 | 7.46e-06 |
| 5 | 3.00e-03 | 1.96e-06 | 2.94e-03 | 4.53e-06 |
| | x | PEDE | y | |
| j | $\mu$ | $\sigma^2$ | $\mu$ | $\sigma^2$ |
| 1 | 9.94e-01 | 2.24e-04 | 9.98e-01 | 7.32e-06 |
| 2 | 2.22e-01 | 4.24e-02 | 1.83e-01 | 2.31e-02 |
| 3 | 2.87e-02 | 1.41e-02 | 3.52e-02 | 8.50e-03 |
| 4 | 1.15e-02 | 6.60e-03 | 3.97e-02 | 4.60e-03 |
| 5 | 2.20e-03 | 5.10e-03 | 3.29e-02 | 3.40e-03 |

can be seen that the average error is always below the 1.4% threshold, and goes diminishing for a low order predictor (from a maximum of 1.36% to a maximum of 0.395%).

Fig. 18, instead, represents the average error committed to predicting the second next mobility point (2-step), given the knowledge of $j$ previous samples (the prediction is made following the scheme illustrated in Fig. 5). Also, in this case, a sequence of 50 predictions has been considered for different Taxi traces and different $AR(j)$ predictors, with $j = 1..15$. The error is negligible (under 1%) and it is minimized for $j = 1$.

Fig. 19 represents the average error committed to predicting the third next mobility point (3-step), given the knowledge of $j$ previous samples (the prediction is made following the scheme illustrated in Fig. 5). Also, in this case, a sequence of 50 predictions has been considered for different Taxi traces and different $AR(j)$ predictors, with $j = 1..15$. In general, the error is not negligible as in the previous cases: it maintains below 2% or 3% in the general case (it is not clear from the figure due to different scales) but, in many cases, there are also error spikes, up to 15.2%, indicating that predicting farther samples is more difficult.

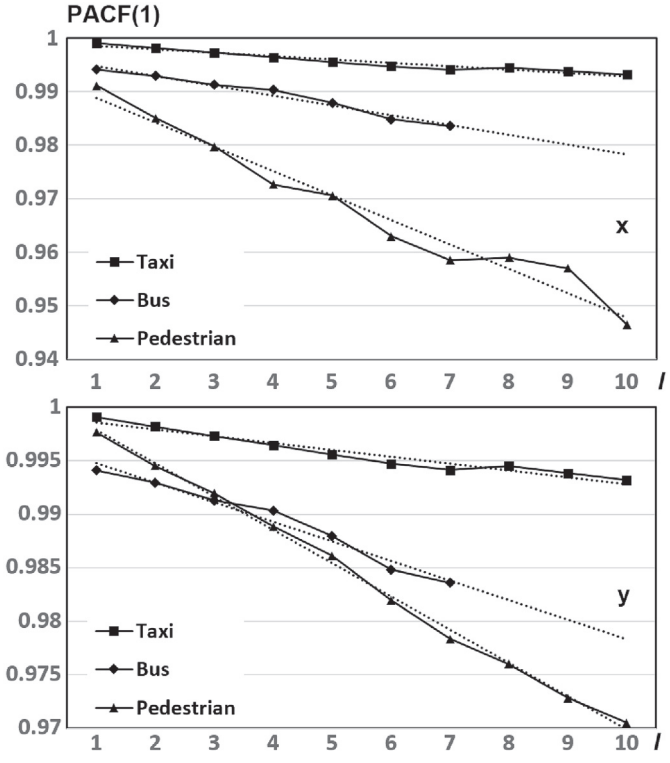


**Fig. 12.** PACF(1) trend versus sampling factor $l$ for Taxi, Bus and Pedestrian traces (x and y coordinate); dotted lines represent the average trend.

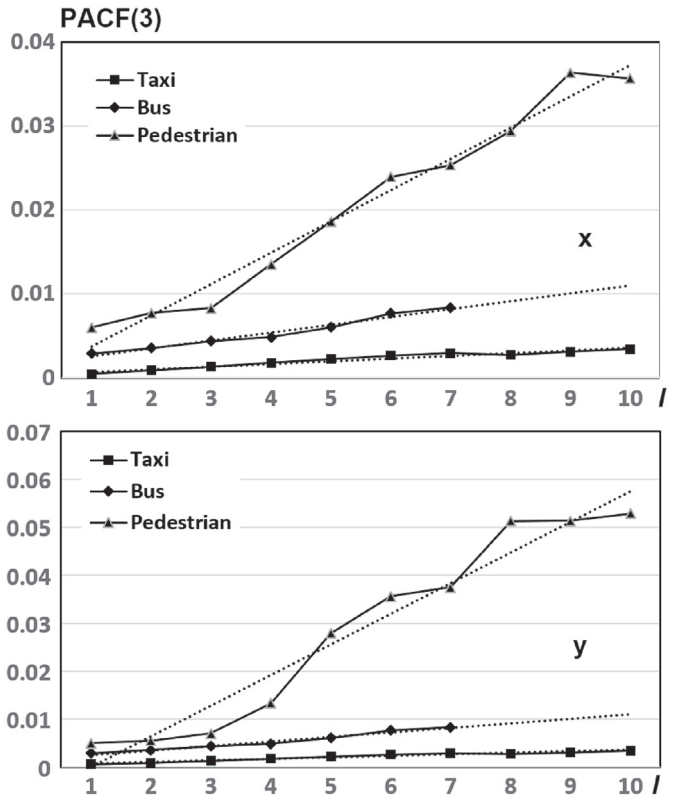


**Fig. 13.** PACF(3) trend versus sampling factor $l$ for Taxi, Bus and Pedestrian traces (x and y coordinate); dotted lines represent the average trend.

Just for completeness, we show the same results for the Bus traces. As it can be seen from Fig. 20, Fig. 21, Fig. 22, the trend is the same as the Taxi case: for higher $j$ values, the average prediction error increases. For the prediction of more than two future samples (3-step prediction) the predictor becomes unreliable, with error spikes up to 7.95%.

With the results above, we conclude that the prediction error increases for higher $j$ because the mobility process for Taxi and Buses is a $j = 1$ process (1st order), so with higher-order predictors, the relationship between the future sample and the $j - th$ previous ones cannot be found adequately, leading the predictor to obtain a not stable set of model coefficients. So, the predictor over-dimensioning is not suitable for these purposes. Besides, considering the history for predicting more than two future samples leads to undesirable and unreliable results. The sentences above are confirmed if we find a Pedestrian trace as in Fig. 23.

In this case, indeed, as stated in subsection 4.2 an $AR(1)$ predictor is not enough (unacceptable prediction error values are obtained, with spikes of 30% and 50%). It can be seen how, for higher $j$, the prediction error is minimized (around 10% for $j = 15$).

As regards the effects of the sampling frequency, the following Fig. 24 illustrates the impact of changing $T$ on the prediction error. For space limitations, we illustrate only one figure (for the Taxi traces), considering only one-step predictions (for Bus patterns the error trend is the same, while for Pedestrian traces it strictly depends on the chosen

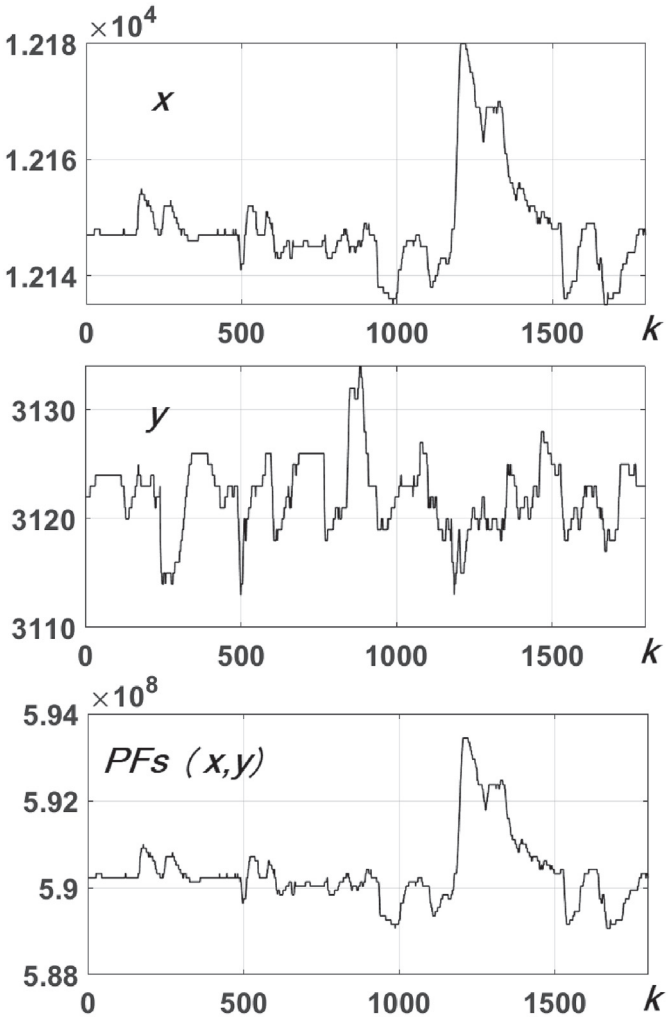


**Fig. 14.** Szudzik's representation of a Taxi mobility trace over 1800 samples: the third curve, on the bottom, represents the paired version of the above $x$ and $y$ coordinates.

**Table 5**
Mean and variance of the PACF for the Szudzik paired traces, for different values of $j$.

| | j = 1 | j = 2 | j = 3 | j = 4 | j = 5 |
|---|---|---|---|---|---|
| TAXI | 0.99909<br>1.00e-06 | 0.00055<br>2.00e-06 | 0.00051<br>1.00e-06 | 0.00042<br>1.00e-06 | 0.00049<br>0.00e+0 |
| BUS | 0.9907<br>3.8e-05 | 0.00306<br>2.6e-05 | 0.0029<br>1.1e-05 | 0.00293<br>7e-06 | 0.00295<br>5e-06 |
| PADE | 0.98957<br>0.00113 | 0.23502<br>0.0541 | 0.00951<br>0.01594 | 0.03526<br>0.00944 | 0.01534<br>0.00708 |

order of the predictor), with an $AR(1)$ process.

In particular, we can observe how the error increases for larger $l$: that is to say, when $T^{new}_{Taxi} = l \cdot T$ becomes higher, the order of the process tends to be higher too, and the $1 - lag$ relationship among samples becomes weaker, losing the ability to predict them. We recall that, for the considered Taxi traces, we have $T^{new}_{Taxi}$ belonging to the interval $[14, 84]s$. After the overall analysis carried out in this paper, we can conclude that:

- No matter the prediction purpose (Cell Stay Time evaluation, predictive resource reservation, routing optimization, etc.), increasing the sampling interval (decreasing the sampling frequency) leads to the needing of increasing the process order, with a related prediction error;
- Samples collection depends on the frequency of the predictions: if we consider the case of Taxi mobility, we have to focus on how frequently we need to know the future node positions. With the collected data, we can predict next positions at least each $T = 7$ s, with the lowest error and simplest predictor (in terms of order). If we need to know node positions after much more time ($l > 1$), higher-order predictor should be considered, with a gain in terms of sampling activity.

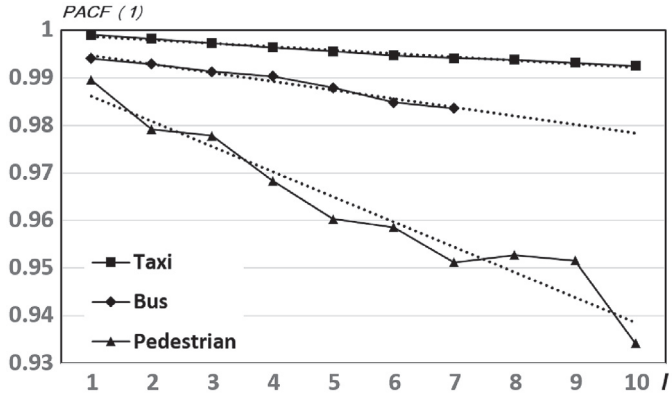


**Fig. 15.** PACF(1) trend versus sampling factor $l$ for Taxi, Bus and Pedestrian paired traces; dotted lines represent the average trend.

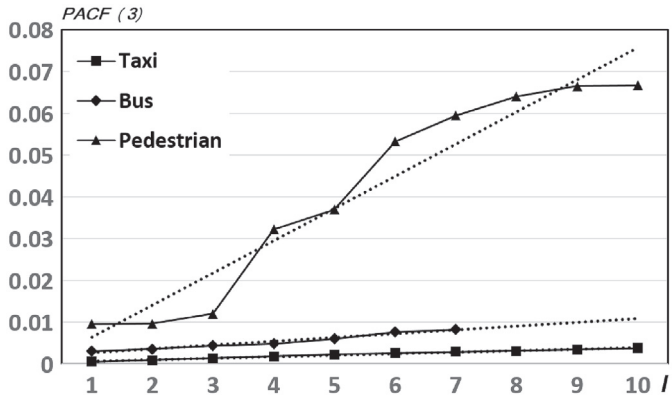


**Fig. 16.** PACF(3) trend versus sampling factor $l$ for Taxi, Bus and Pedestrian paired traces; dotted lines represent the average trend.

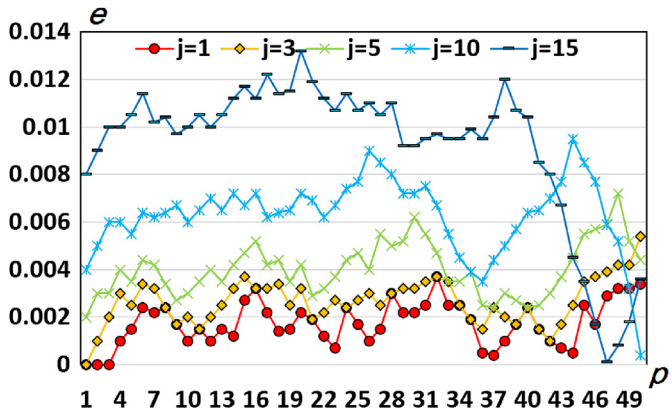


**Fig. 17.** Average error $e$ in predicting 50 1-step coordinates, for Taxi traces.

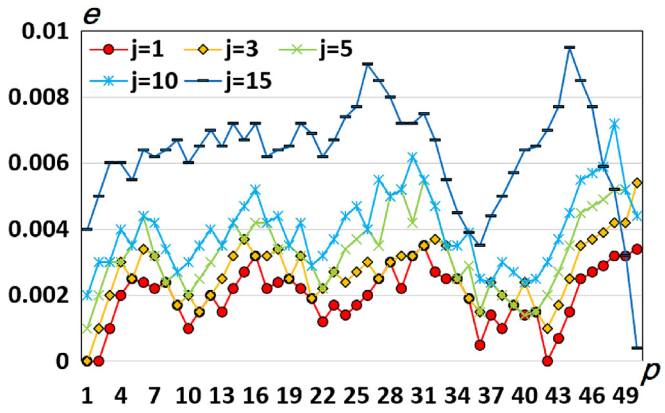


**Fig. 18.** Average error $e$ in predicting 50 2-step coordinates, for Taxi traces.

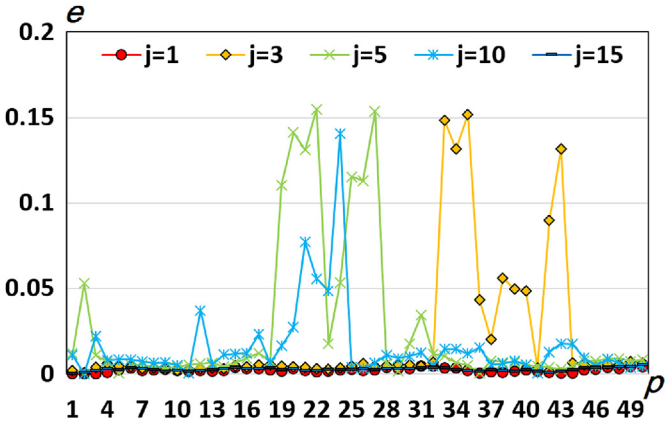


**Fig. 19.** Average error $e$ in predicting 50 3-step coordinates, for Taxi traces.

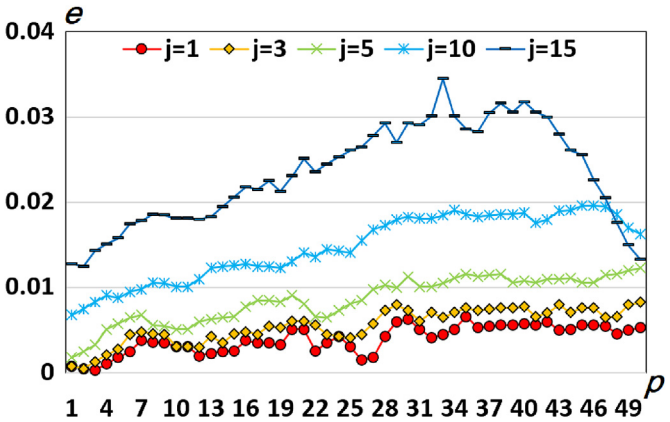


**Fig. 20.** Average error $e$ in predicting 50 1-step coordinates, for Bus traces.

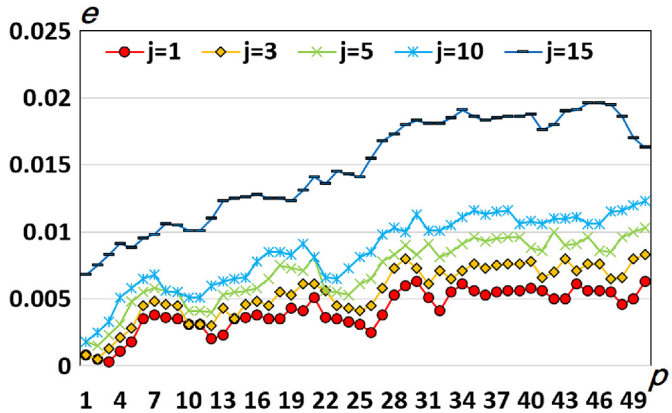


**Fig. 21.** Average error $e$ in predicting 50 2-step coordinates, for Bus traces.

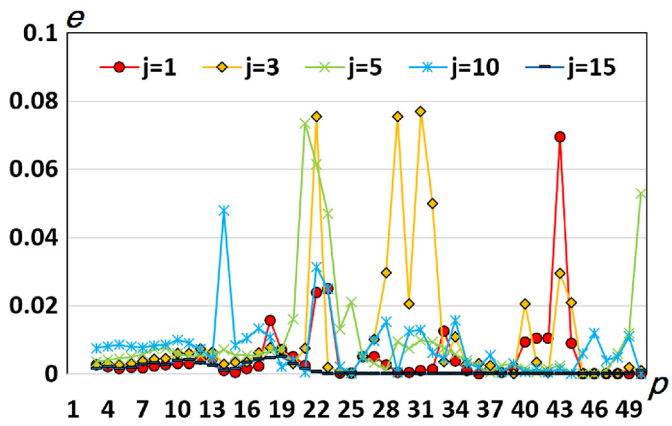


**Fig. 22.** Average error $e$ in predicting 50 3-step coordinates, for Bus traces.

## 5. Conclusion and future works

In this paper, an in-depth stochastical analysis of nodes mobility has been carried out: in particular, we focused on the study of the effects of changing the sampling frequency of nodes mobility, used to create historical patterns for future mobility prediction activities. We did not consider any particular predictor (Markov chain, neural network, cellular automata, etc.). Still, we verified what happens when we collect mobility samples more/less frequently and why we should adjust the sampling frequency. In particular, we noticed that increasing the sampling period, the number of collected samples decreases, of course, but the correlation among consecutive samples decreases, adding more dependence from older samples. This implies that more complicated predictors need to be implemented. However, the choice of the sampling frequency strictly depends on the considered prediction scenario: on the basis of the time prediction horizon (how far the future sample needs to be predicted), the sampling frequency needs to be set accordingly to desirably make a few number of prediction. We are continuing our research activity about this topic, and the next step is to establish a clear relationship between sampling frequency and samples correlation, adding a more profound analysis in the frequency/Laplace/Wavelet

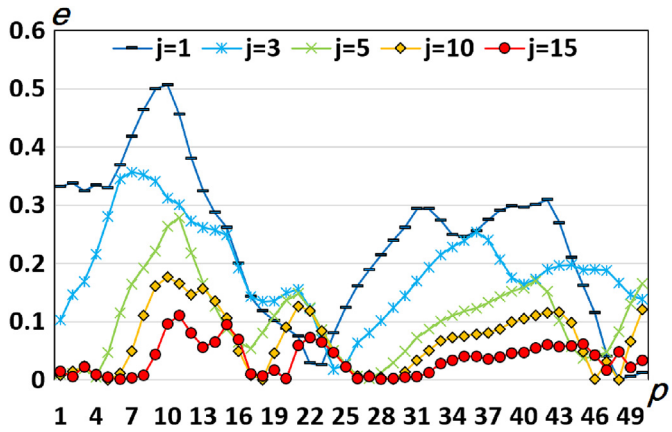


**Fig. 23.** Average error $e$ in predicting 50 1-step coordinates, for pedestrian traces.

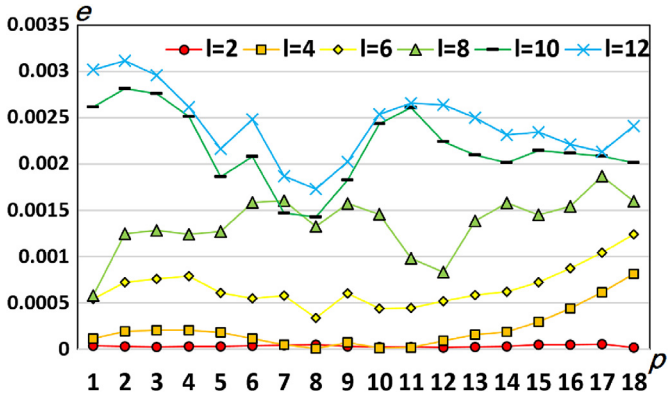


**Fig. 24.** Average error *e* in predicting 18 1-step coordinates, for Taxi traces.

domains. The obtained results confirmed the theoretical expectations of our study.

## CRediT authorship contribution statement

**Peppino Fazio:** Conceptualization, Methodology, Formal analysis, Software, Investigation. **Miralem Mehic:** Data curation, Investigation, Writing - original draft, Writing - review & editing, Supervision. **Miroslav Voznak:** Supervision, Funding acquisition, Project administration, Writing - review & editing, Supervision.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgment

This work was supported by by the Czech Ministry of Education, Youth and Sports within the institutional grants SGS reg. no. SP2020/65 and SP2019/41 conducted at VSB - Technical University of Ostrava and also within the ESF project “Science without borders” reg. no.CZ.02.2.69/0.0/0.0/16027/0008463 of the operational programme “Research, development and education in the Czech Republic”.

## References

Bisgaard, S., Ankenman, B., aug 1996. Standard Errors for the Eigenvalues in Second-Order Response Surface Models. ser. (Chapter 3). plus 0.5em minus 0.4emWiley Series in Probability and Statistics. vol. 38. john wiley. no. 3. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/00401706.1996.10484503.

Borrego, C., Hernndez-Orallo, E., Magaia, N., oct 2019. General and Mixed Linear Regressions to Estimate Inter-contact Times and Contact Duration in Opportunistic Networks, vol. 93, [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1570870518309314.

Box, G.E.P., Jenkins, G.M., Reinsel, G.C., jun 2008. *Time Series Analysis*, Ser. Wiley Series in Probability and Statistics. plus 0.5em minus 0.4emHolden-Day. Wiley, San Francisco. ISBN: 9780470272848. [Online]. Available: https://onlinelibrary.wiley.com/doi/book/10.1002/9781118619193.

Bracciale, L., jul 2014. (v. 2014-07-17), , no. 15783. CRAWDAD Dataset Roma/taxi, vol. 10, pp. 2007–2014 [Online]. Available: https://crawdad.org/roma/taxi/20140717.

Brouwer, L.E.J., 1912. Beweis der Invarianz des n-dimensionalen Gebiets. Math. Ann. 71, 305–315.

Cantor, G., 1878. Ein beitrag zur mannigfaltigkeitslehre. J. fr die Reine Angewandte Math. (Crelle’s J.) 84, 242–258.

Cao, J., Xu, S., Zhu, X., Lv, R., Liu, B., aug 2017. Efficient fine-grained location prediction based on user mobility pattern in LBSNs. In: 2017 Fifth International Conference on Advanced Cloud and Big Data (CBD). Plus 0.5em Minus 0.4em13th-16th August Shangai. IEEE, China. ISBN: 978-1-5386-1072-5, pp. 238–243, https://doi.org/10.1109/CBD.2017.48. [Online]. Available: http://ieeexplore.ieee.org/document/8026943/.

Chaudhari, S.S., Biradar, R.C., sep 2016. Traffic and mobility aware resource prediction using cognitive agent in mobile ad hoc networks. J. Netw. Comput. Appl. 72, 87–103, https://doi.org/10.1016/j.jnca.2016.06.010. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S108480451630131X.

Chon, Y., Shin, H., Talipov, E., Cha, H., mar 2012. Evaluating mobility models for temporal prediction with high-granularity mobility data. In: 2012 IEEE International Conference on Pervasive Computing and Communications. Plus 0.5em Minus 0.4em. IEEE. ISBN: 978-1-4673-0258-6, pp. 206–212, https://doi.org/10.1109/PerCom.2012.6199868. [Online]. Available: http://ieeexplore.ieee.org/document/6199868/.

Cochrane, J., 1997. Time Series for Macroeconomics and Finance. vol. 1997. *Graduate School of Business, University of Chicago*.

Dias, D., mar 2018. (v. 2018-03-19), , no. 15783. CRAWDAD Dataset Coppe-ufrj/RioBuses, vol. 10, pp. 2003–2018 [Online]. Available: https://crawdad.org/coppe-ufrj/RioBuses/20180319.

Fazio, P., 2016. A predictive cross-layered interference management in a multichannel MAC with reactive routing in VANET. IEEE Trans. Mobile Comput. 15 (8), 1850–1862.

Fazio, P., De Rango, F., Tropea, M., 2017. Prediction and QoS enhancement in new generation cellular networks with mobile hosts: a survey on different protocols and conventional/unconventional approaches. IEEE Commun. Surv. Tutor. 19 (3), 1822–1841, https://doi.org/10.1109/COMST.2017.2684778. [Online]. Available: http://ieeexplore.ieee.org/document/7882671/.

Hnsler, E., 2001. Statistische Signale, vol. 2001,, https://doi.org/10.1007/978-3-642-56674-5 [Online]. Available: http://link.springer.com/10.1007/978-3-642-56674-5.

Hernndez-Orallo, E., Carlos Cano, J., Calafate, C.T., Manzoni, P., 2018. FALCON: a new approach for the evaluation of opportunistic networks. Ad Hoc Netw., https://doi.org/10.1016/j.adhoc.2018.07.004.

Hornik, K., Stinchcombe, M., White, H., jan 1989. Multilayer feedforward networks are universal approximators. Neural Network. 2 (5), 359–366, https://doi.org/10.1016/0893-6080(89)90020-8. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/0893608089900208.

Katsikouli, P., Viana, A.C., Fiore, M., Tarable, A., dec 2017. On the sampling frequency of human mobility. In: GLOBECOM 2017 - 2017 IEEE Global Communications Conference, vol. 2017, pp. 1–6, https://doi.org/10.1109/GLOCOM.2017.8254476 [Online]. Available: http://ieeexplore.ieee.org/document/8254476/.

Krishna, B.H., Reddy, I.R.S., Kiran, S., Reddy, R.P.K., mar 2016. Multiple text encryption, key entrenched, distributed cipher using pairing functions and transposition ciphers. In: 2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET). Plus 0.5em Minus 0.4em. IEEE. ISBN: 978-1-4673-9338-6, pp. 1059–1061, https://doi.org/10.1109/WiSPNET.2016.7566299. [Online]. Available: http://ieeexplore.ieee.org/document/7566299/.

J. Lee, Univariate time series modeling and forecasting (Box-Jenkins Method), Econ. Times 413, vol. 4.

Li, F., Li, Q., Li, Z., Huang, Z., Chang, X., Xia, J., 2019. A personal location prediction method based on individual trajectory and group trajectory. IEEE Access 7, 92850–92860, https://doi.org/10.1109/ACCESS.2019.2927888. [Online]. Available: https://ieeexplore.ieee.org/document/8758816/.

Pirozmand, P., Wu, G., Jedari, B., Xia, F., jun 2014. Human mobility in opportunistic networks: characteristics, models and prediction methods. J. Netw. Comput. Appl. 42, 45–58, https://doi.org/10.1016/j.jnca.2014.03.007. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1084804514000587.

Rango, F.D., mar 2009. Utility-based predictive services for adaptive wireless networks with mobile hosts. IEEE Trans. Veh. Technol. 58 (3), 1415–1428.

Rhee, I., jul 2009. (v. 2009-07-23), , no. 15783. CRAWDAD Dataset Ncsu/mobilitymodels, vol. 10, pp. 2007–2009 [Online]. Available: https://crawdad.org/ncsu/mobilitymodels/20090723.

Satria, T.A., Karimzadeh, M., Karagiannis, G., oct 2014. Performance evaluation of ICN/CCN based service migration approach in virtualized LTE systems. In: 2014 IEEE 3rd International Conference on Cloud Networking (CloudNet), 2014, pp. 461–467, https://doi.org/10.1109/CloudNet.2014.6969038 [Online]. Available: http://ieeexplore.ieee.org/document/6969038/.

Suraj, R., Tapaswi, S., Yousef, S., Pattanaik, K.K., Cole, M., aug 2016. Mobility prediction in mobile ad hoc networks using a lightweight genetic algorithm. Wireless Network 22 (6), 1797–1806, https://doi.org/10.1007/s11276-015-1059-0. [Online]. Available: http://link.springer.com/10.1007/s11276-015-1059-0.

Szudzik, M.P., jun 2017. The Rosenberg-Strong Pairing Function. vol. 2019, [Online]. Available: https://arxiv.org/abs/1706.04129 http://arxiv.org/abs/1706.04129.

ur Rahman, S., Kim, G.-H., Cho, Y.-Z., Khan, A., oct 2018. Positioning of UAVs for throughput maximization in software-defined disaster area UAV communication networks. J. Commun. Network. 20 (5), 452–463, https://doi.org/10.1109/JCN.2018.000070. [Online]. Available: https://ieeexplore.ieee.org/document/8533581/.

Wang, J., Kong, X., Xia, F., Sun, L., may 2019. Urban human mobility. ACM SIGKDD Explor. Newslett. 21 (1), 1–19, https://doi.org/10.1145/3331651.3331653. [Online]. Available: https://dl.acm.org/doi/10.1145/3331651.3331653.

Wolfram, S., Gad-el-Hak, M., mar 2003. A new kind of science. Appl. Mech. Rev. 56 (2), B18–B19, https://doi.org/10.1115/1.1553433. [Online]. Available: https://asmedigitalcollection.asme.org/appliedmechanicsreviews/article/56/2/B18/458836/A-New-Kind-of-Science.

Wu, R., jun 2018. Location prediction on trajectory data: a review. Big Data Min. Anal. 1 (2), 108–127, https://doi.org/10.26599/BDMA.2018.9020010. [Online]. Available: https://ieeexplore.ieee.org/document/8336847/.

Wu, R., Luo, G., Yang, Q., Shao, J., 2018. Learning individual moving preference and social interaction for location prediction. IEEE Access 6, 10675–10687, https://doi.org/10.1109/ACCESS.2018.2805831. [Online]. Available: http://ieeexplore.ieee.org/document/8290840/.

Yamada, N., Katsumaru, N., Nishijima, H., Kimoto, M., oct 2018. Location prediction based on smartphone multimodal personal data for proactive support services. In: 2018 Eleventh International Conference on Mobile Computing and Ubiquitous Network (ICMU). Plus 0.5em Minus 0.4em. IEEE. ISBN: 978-4-907626-34-1, pp. 1–2, https://doi.org/10.23919/ICMU.2018.8653598. [Online]. Available: https://ieeexplore.ieee.org/document/8653598/.

Yu, C., Liu, Y., Yao, D., Yang, L.T., Jin, H., Chen, H., Ding, Q., jun 2017. Modeling user activity patterns for next-place prediction. IEEE Syst. J. 11 (2), 1060–1071, https://doi.org/10.1109/JSYST.2015.2445919. [Online]. Available: http://ieeexplore.ieee.org/document/7150318/.

Yule, G.U., 1927. On a method of investigating periodicities in disturbed series, with special reference to wolfer's sunspot numbers. Phil. Trans. Roy. Soc. 226, 267–298.

Zareei, M., Islam, A.M., Vargas-Rosales, C., Mansoor, N., Goudarzi, S., Rehmani, M.H., feb 2018. Mobility-aware medium access control protocols for wireless sensor networks: a survey. J. Netw. Comput. Appl. 104, 21–37, https://doi.org/10.1016/j.jnca.2017.12.009. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1084804517304216.

Zhang, H., Dai, L., 2019. Mobility prediction: a survey on state-of-the-art schemes and future applications. IEEE Access 7, 802–822, https://doi.org/10.1109/ACCESS.2018.2885821. [Online]. Available: https://ieeexplore.ieee.org/document/8570749/.

Zhao, Z., Karimzadeh, M., Braun, T., Pras, A., van den Berg, H., oct 2015. A demonstration of mobility prediction as a service in cloudified LTE networks. In: 2015 IEEE 4th International Conference on Cloud Networking (CloudNet). Plus 0.5em Minus 0.4em. IEEE. ISBN: 978-1-4673-9501-4, pp. 78–80, https://doi.org/10.1109/CloudNet.2015.7335285. [Online]. Available: http://ieeexplore.ieee.org/document/7335285/.

**Peppino Fazio** was born 1977 in Italy, he received the PhD. degree in Electronics and Communications Engineering, University of Calabria (UNICAL - Italy) in 2008 and completed his habilitation as Associate Professor in 2017, after being an Assistant Professor at DIMES Dept. (UNICAL) until 2016. He is co-author of 98 papers (32 in International Journals, 63 in Conference Proceedings and 3 Book Chapters), all indexed in Scopus and/or WoS. His reputation in the open community network Research Gate, measured by RGScore (25.18), is higher than 80% of ResearchGate members. His research interests include mobile communication networks, QoS architectures and interworking, wireless and wired networks, mobility modelling for WLAN environments, mobility analysis for prediction purposes, routing, vehicular networking, MANET and VANET. He is peer reviewer and TPC member of different international conferences, as well as for many international journals, such as IEEE TVT, COMMLETT, VTM, SPRINGER TELS, MONET, ELSEVIER VEHCOMM, COMNET, and many others.

**Miralem Mehic** was born in 1988 in Bosnia and Herzegovina. He received the Ph.D. degree in telecommunications from the VSB-Technical University of Ostrava in 2017 where he works as a postdoctoral researcher. Also, he studied at the AGH University of Science and Technology, Krakow, Poland, Alpen-Adria-Universitat Klagenfurt, Austria and Austrian Institute of Technology (AIT) in Department of Digital Safety & Security Business Units - Optical Quantum Technology, Vienna & Klagenfurt, Austria. Miralem is the author of the unique QKD network simulator QKDNETSIM and participated in the successful realization of the virtual VSB-AIT QKD link. His field of research is related to the quality of service and management of QKD networks with a focus on real-time traffic and the utilization of network resources. His works have been published in leading journals ranked in Q1 and Q2 quarters (IEEE Access, Quantum Information Processing, IEEE Journal of Quantum Electronics) while his work on QKD network signalling protocols was awarded as the best paper at the MCSS 2017 conference in Krakow, Poland. Since 2019, prof. Mehic is a Head of Department of Telecommunications at the University of Sarajevo, Bosnia and Herzegovina.

**Miroslav Voznak** was born 1971 in Czechia, he received the PhD. degree in telecommunications from the Faculty of Electrical Engineering and Computer Science, VSB - Technical University of Ostrava and completed his habilitation in 2002 and 2009, respectively. He was appointed Full Professor in 2017 in Electronics and Communication technologies. He is an IEEE senior member, author and co-author of 185/67 (overall/articles) indexed results in WoS and 259/115 in Scopus and his reputation in the open community network Research Gate, measured by RGScore, is higher than 90% of ResearchGate members. He participated in more than twenty european and national research projects. He was doing a part-time job as a researcher for CESNET, serving in two task-forces of TERENA and entire professional life has been employed in the VSB-Technical University of Ostrava at positions junior researcher, senior researcher, assistant professor, associate professor, professor, since 2013 department chair in Dept. of Telecommunications in Faculty of Electrical Engineering and Computer Science and since 2017 also as a head of Laboratory for Big Data Analysis in National Supercomputing Centre IT4Innovations.