From movement purpose to perceptive spatial mobility prediction

Licia Amichi, Aline Carneiro Viana, Mark Crovella, Antonio a F Loureiro

FROM MOVEMENT PURPOSE TO PERCEPTIVE SPATIAL MOBILITY PREDICTION

A PREPRINT

Licia Amichi
Ecole Polytechnique (IPP) and Inria, France
licia.amichi@inria.fr

Aline Carneiro Viana
Inria, France
aline.viana@inria.fr

Mark Crovella
Boston University, USA
crovella@bu.edu

Antonio A.F. Loureiro
Federal University of Minas Gerais, Brazil
loureiro@dcc.ufmg.br

November 23, 2021

ABSTRACT

A major limiting factor for prediction algorithms is the forecast of new or never before-visited locations. Conventional personal models utterly relying on personal location data perform poorly when it comes to discoveries of new regions. The reason is explained by the prediction relying only on previously visited/seen (or known) locations. As a side effect, locations that were never visited before (or explorations) by a user cause disturbance to known location's prediction. Besides, such explorations cannot be accurately predicted. We claim the tackling of such limitation first requires identifying the purpose of the next probable movement. In this context, we propose a novel framework for adjusting prediction resolution when probable explorations are going to happen. As recently demonstrated [1, 2], there exist regularities in returning and exploring visits. Moreover, the geographical occurrences of explorations are far from being random in a coarser-grained spatial resolution. Exploiting these properties, instead of directly predicting a user's next location, we design a two-step predictive framework. First, we infer an individual's next type of transition: (i) a *return*, i.e., a visit to a previously known location, or (ii) an *exploration*, i.e., a discovery of a new place. Next, we predict the next location or the next coarse-grained zone depending on the inferred type of movement. We conduct extensive experiments on three real-world GPS mobility traces. The results demonstrate substantial improvements in the accuracy of prediction by dint of fruitfully forecasting coarse-grained zones used for exploration activities. To the best of our knowledge, we are the first to propose a framework solely based on personal location data to tackle the prediction of visits to new places.

***K*eywords** human mobility · predictability · entropy

1 Introduction

Accurately predicting human trajectories is relevant to many domains and applications such as targeted advertising, epidemic prevention, or smooth resource and handover management for mobile networks [3, 4, 5]. Due to predictors' invaluable contributions, the research community has witnessed a plethora of mobility prediction methods and techniques becoming more and more robust and accurate [6, 7, 8, 9].

Prediction tasks can be classified into two categories [4]: (i) *next-place* formulation that aims at predicting transitions between places (ii) *next-cell* formulation that seeks to forecast the location of a user within the next time bin. In the next-place prediction task, the stationary behavior anchored in human movements is omitted. Therefore, it is more sensitive to irregular visits, and in particular, to *discoveries of new or never before-visited places* [4, 5]. In this paper, we focus on the next-place task, which performs dramatically lower in terms of prediction accuracy than the next-cell formulation and represents a more challenging problem [4, 3].

Discoveries of new locations commonly referred to as *explorations* are a major limiting factor for prediction tasks [4]. Indeed, forecasting discoveries of new places is ambitious and difficult to tackle as it is about predicting the unknown. Conventional *personal* predictors such as Markov-based models [7, 10, 8] or Hidden Markov Models (HMM) [11] utterly rely on historic personal location data to predict future locations. Moreover, they predict a user's next location on the assumption that it belongs to the set of her known places [3]. This engenders erroneous forecasts at each occurrence of an exploration event, which is worsened by the fact that such events are numerous and largely present in the daily lives of users: on average 70% of visits happen only once [4]. This representative rate highlights how impacting exploration-intended visits are for conventional personal predictors and puts in evidence the need for detecting such types of movements.

Advanced contextual information has recently been jointly used with mobility data to better tackle exploration visits in predictions [3, 11]. Examples are the semantic of the visited location, the activity performed within the location, the personality traits of the user [12], or her social circle [4, 3]. Such contexts request massive data collection and bring privacy concerns [13, 14]. Although possibility enhancing prediction, we let context-aware prediction to future work and focus our investigations uniquely on individuals' mobility data.

In this paper, we propose a newly tailored *mobility prediction framework* that tackles the exploration problem by leveraging the purpose of movements at prediction decisions, only using location data. Several works demonstrate the inherent temporal periodicity and spatial regularity of human return visits [1]. Furthermore, in our previous work [2], we show that, though the apparent randomness of exploration visits, their temporal and geographical occurrences are far from being random when considering coarser-grained spatial scopes. Exploiting these properties – instead of conventionally predicting a user's next location based on the history of known visited places –, we design a two-step prediction framework that encloses two modules: (i) the *purpose of movement predictor* and (ii) the *spatial predictor*. By doing so, we claim spatial prediction of users' mobility is enhanced by a better perception of the motivation behind movement decisions.

First, we add movement semantic to the mobility traces. To do this, we split the visits into two purposes of movements: (i) *explorations* and (ii) *returns*. Such two purposes hereafter referred to as *types"*, and the related transitions among the two movement types are then leveraged in the second step of our framework. In accordance with the resulting sequences of types of movements, we propose two distinct movement predictors: (i) Successive Types of Movements Predictor (STMP) and (ii) Inter Exploration Interval based Predictor (IEIP) to infer whether the next transition is an *exploration* or a *return*. *Finally*, given the inferred type of visit, we propose two spatial predictors: (i) Personal Spatial Predictor (PSP) and (ii) Joint Spatial Predictor (JSP). If the inferred type is a *return*, both spatial predictors employ a first-order Markov Chain (MC) predictor to forecast the next visited location (small cells). On the contrary, if the next movement is an *exploration*, PSP and JSP adapt their prediction to forecast a zone (large cells) instead of a location.

Although predicting explorations in a coarse-grained spatial resolution, we believe *our framework brings individual-perception capabilities to mobility predictors.* First, *it tackles the conventional predictors' limitation of being oblivious to users' explorations*, which results in low accuracy forecasts, especially for highly exploratory users. Second, *it provides a way for predictors (and for entities taking benefits from their results) to identify how trustable predictors results are in terms of accuracy, and accordingly, adapt their prediction to users' intention behind their movement.* Our contributions are:

- To the best of our knowledge, we are the first to propose an exploration-aware mobility prediction framework that solely relies on timestamped location data. The proposed framework splits the location prediction problem into two main steps: (i) *predicting the next type of movement* (exploration or return) (ii) *inferring the spatial location of the visit*.
- We design a first movement predictor STMP that exploits the regularity of *return-like* visits, as well as the *exploration-like* visits reasoning human mobility, to infer the type of the next movement. We propose a second predictor IEIP that focuses on explorations habits in time to forecast their occurrence (Section 4.1). And evaluate the performance of the movement predictors with regard to each type of movement (Section 6).
- Next, we propose two spatial predictors that take as input the outcomes of the movement type's predictors. If the input is a *return*, the spatial predictors PSP and JSP employ an MC predictor to predict the *next location* of size $(200m)^2$. If the input is an *exploration*, the PSP looks at the user's past history to infer the next *zone*. Instead, the JSP exploits the collective exploratory behavior besides the user's past history for its forecasts. We consider zones of different size in our study: $(800m)^2$, $(1km)^2$, $(2km)^2$, and $(4km)^2$ (Section 4.2).
- Using three real-world GPS datasets, we evaluate the performance of our proposed framework. We show our framework allows increasing the prediction performance in general, but it can also be tuned according to the needs and requirements of the using applications or services. For instance, exploration forecasts decrease uncertainty in population mobility anticipation, which directly enhances resource allocation in network planning (as for the placement of Mobile Edge Computing (MEC) by telecom operators) and recommendation systems performance.

In these cases, the proposed framework allows tuning the accuracy of exploration forecasts (from low to high) according to the intended spatial resolution or aimed effectiveness of services: e.g., adjusting coverage areas from $1km^2$ to $(4km)^2$ in MEC deployment according to population density or still, the spatial precision of target areas from $(200m)^2$ or $1km^2$ in recommendation systems).

The rest of the paper is organized as follows. In Section 2, we review the related work. In Section 3, we formulate the problem and introduce the general structure of the proposed framework. In Section 4, we detail our proposed prediction framework. After the framework description in Section 5, we present the three real-world GPS traces that we employ to evaluate the performance of the proposed framework (Sections 6 and 7). Finally, we conclude our paper in Section 8.

2 Related Work

Accurately predicting human mobility can benefit various applications particularly in the field of ubiquitous computing and urban planning. Existing individual-level predictors seek to predict the future location of a given user [7, 10, 6, 11] and can be classified into two categories: *personal*, i.e., the forecasts are solely based on the user's personal data [7, 10] or *joint*, i.e., that in addition to the personal data of the user, common mobility patterns and behavior are exploited to infer the user's future locations [6, 11, 15].

Markov-based models are common models for personal predictors. Gambs et al. [7] propose an MC model that exploits the n previously visited locations to predict future visits. Using GPS mobility traces they reveal that the next location can be predicted with a markedly high accuracy of 70-95%. Likewise, Lu et al. [10] and Song et al. [16] use Markov-based models. The former, use a varying order MC model, where the order is high when the historical information is limited and becomes smaller when the historical trajectory is over 100 points. The achieved accuracy of prediction with a first-order MC model on a CDR dataset surpasses 90%. The latter exploit several methods 0-order MC, LeZi, Prediction by Partial Matching (PPM), and Sampled Pattern Matching (SPM) to infer future locations based on the past history. They show that Markov predictors work as well or better than more complex compression-based predictors and report an accuracy of prediction between 65% and 72% on Dartmouth's campus-wide Wi-Fi wireless network.

In contrast to the aforementioned works that tackle the next-cell prediction problem, Gidofalvi et al. [8] consider the next-place prediction task i.e., predicting transitions between places (*as in our case of study*). The authors propose an inhomogeneous continuous-time Markov model to predict when a user will leave her current location and where she will move next. The evaluation of the predictive performance with a GPS dataset reports that the accuracy of prediction is above 67%. Other more complex methods are employed for personal predictions. For instance, Mathew et al. [11] propose a hybrid approach that first clusters the locations according to their characteristics (temporal period in which they occurred), then trains HMM on each cluster. To predict the next visited location, the model starts by identifying the most likely cluster, then infer the future location using the corresponding HMM. The authors, measure a prediction accuracy of about 13.85% with GPS trajectories. A further example, Feng et al. [17] employ the Kalman filter to predict the future location of a vehicle.

Nonetheless, these conventional *personal* models fail to predict visits to new locations [3, 4]. Hence, several *joint* methods that train on aggregated data are more and more employed to enhance the predictive performance of individuals' mobility. Asahara et al. [6] propose a Mixed Markov chain Model (MMM) that first classifies the users into groups of similar users. Next, to predict the future location of a user it trains an MC predictor for all users of the group to which she belongs. The authors compared the performance of the MMM, with the MC and HMM models. They report an accuracy of prediction of 16.9% (45.6%) for MC, 4.2% (2.41%) for the HMM, and 74.1% (64%) for the MMM when predicting transitions between locations with simulation (real-world GPS) data. Calabrese et al. [18] introduce a predictor that combines an individual's past mobility choices with collective behavior to help (i) predict the likelihood the user changes location and (ii) infer the type of geographical areas that should be similar to the type that interest the collectivity at the given time. Using a CDR dataset, the authors report a good level of accuracy in terms of prediction error (60% of the errors are zero). Alhasoun et al. [9] propose a Dynamic Bayesian Network approach that couples friends "similar strangers" records to increase the accuracy in predicting the next location. They report an accuracy of prediction of 60.03% by relying on timestamped geographical data in addition to social contact contextual information.

While most of the previous works base their forecasts on timestamped geographical data, the recent availability of enriched geo-tagged datasets with various contextual information brings new opportunities to enhance the prediction task [4]. Nonetheless, such data sources are not publicly available and require to follow a non-trivial process for their acquisition. Moreover, with the emerging requirements of privacy protection, handling such data raises serious privacy concerns. Therefore, scholars strive to consider privacy issues straightforwardly in the modeling and system or bypass them by employing the least possible data [13, 14].

Position of our work: Different from the existing works on individual mobility prediction. We focus on the exploration problem in the next-place prediction task by solely manipulating timestamped geographical data. In this regard, we propose a two-step exploration-aware mobility prediction framework. Here, prediction is firstly performed to forecast the next type (or purpose) of movement. We then leverage this knowledge to adapt predictors decisions at the inference of the next place of users.

3 Problem Formulation

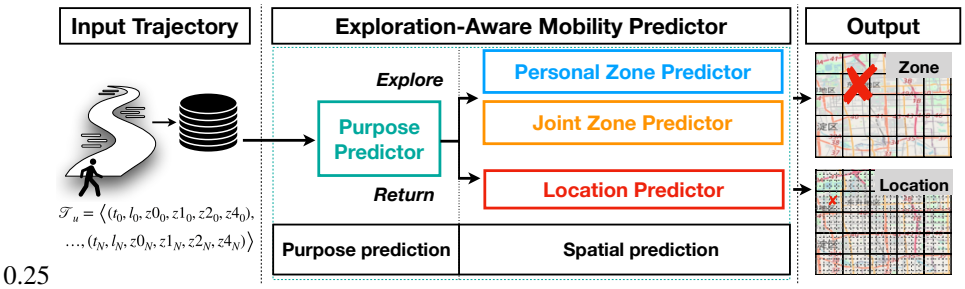


Figure 1: System overview.

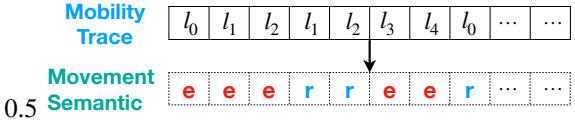


Figure 2: Adding movement semantic to a mobility trace.

Figure 3: Exploration-Aware Mobility Prediction Framework.

We leverage GPS traces and due to the popularity and high precision of such data, we suppose that users' locations are calculated based on the GPS system. Our approach can be adapted to other positioning systems. A GPS point is defined by its coordinates latitude, lat and longitude lon. The GPS mobility trace $\mathcal{T}_u$ of a user u can be defined as follows,

Definition 1 (***GPS Mobility trace*** $\mathcal{T}_u$**).** *it is an ordered sequence of GPS records reporting the locations visited by the user u during the data collection period. A GPS record $q =< u, t, lat, lon >$ contains the following information: the identifier of the user "u", the date of collection "t", the latitude "lat", and longitude "lon". Thus, the GPS mobility trajectory $\mathcal{T}_u$ of the user u can be written as, $\mathcal{T}_u = \langle (t_0, lat_0, lon_0), (t_1, lat_1, lon_1), \ldots (t_N, lat_N, lon_N) \rangle$.*

Due to GPS range errors, locations are usually defined by spatial grid IDs or Points of Interest. Following customary practice [4], we superpose uniform grids of size c meters $\times$ c meters on the geographical maps. Next, we project the GPS coordinates to convert them into spatial grid IDs. We consider different cell size: cells of size $(200m)^2$ that we refer to as *locations*, cells of size $(800m)^2$ that we refer to as *zones_0*, cells of size $(1km)^2$ that we refer to as *zones_1*, cells of size $(2km)^2$ that we refer to as *zones_2*, and cells of size $(4km)^2$ that we refer to as *zones_4*. Thus, the mobility traces are extended to $\mathcal{T}_u = \langle (t_0, l_0, z0_0, z1_0, z2_0, z4_0), \ldots, (t_N, l_N, z0_N, z1_N, z2_N, z4_N) \rangle$.

As mentioned earlier, in this work, we focus on the next-place prediction problem. Hereafter, we provide a general definition,

Definition 2 (***Next-place Mobility Prediction*** **).** *given the current location l_N of a user, the next-place mobility prediction is about predicting the next location l_{N+1} to which the user will make a transition.*

The next-place prediction encompasses two main tasks: (i) predicting *when* an individual will make a transition (ii) predicting *where* the individual will go next. In our approach, we relax the next-place prediction problem to *where will the individual go next?* by assuming that the transition time is already known.

The mobility prediction problem can be tackled in several ways depending on the characteristics of the data and the objectives of the forecast. It can be addressed either *directly*, i.e., by straightforwardly inferring future locations [7, 19], or *indirectly*, i.e., by forecasting other events such as social context, type of direction, and so forth, and based on that infer the next location [4, 3]. Besides, the prediction can also be *personal*, or *joint* [7, 3, 6]. When only timestamped geographical data are available, conventional models are usually *direct-personal* or *direct-joint*, i.e., they attempt to predict the future locations directly using the individual's data or the aggregated collective data [3, 4].

In contrast, in this work, we propose a simple *indirect* method for predicting transitions between places by utterly relying on timestamped geographical data. An individual's mobility trajectory can be viewed as a sequence of instants of returns interrupted by instants of explorations [1, 2]. Unlike conventional predictors leveraging the same type of data, we do not naively infer the future location based on the past history of the user, if next, she is more likely to discover a new one. As Figure 1 shows, the proposed mobility predictor consists of two parts: (i) type of movement prediction and (ii) spatial prediction. Details of these parts are presented in the following Sections.

4 Exploration-aware Mobility Predictors

Our proposed method is divided into two sequentially dependent modules: (i) purpose prediction, i.e., predicting the next type of movement (Section 4.1) and (ii) spacial prediction, i.e., inferring the next location or zone where the individual will be (Section 4.2). Both modules are detailed in the following.

4.1 Purpose prediction

Following recent practice [4, 2], the proposed movement prediction strategies adopt the subsequent movement dichotomy: (i) *explorations* or discoveries of new places e and (ii) visits of previously known locations termed *returns* r. Hence, the set of movements comprises two elements $\mathcal{M} = \{e, r\}$. The movement prediction task aims to answer the following question: *what will the individual do next? Explore or return?*

Considering the definition in Section 3, we convert the original GPS mobility trajectories of each user u into a sequence $\mathcal{T}_u = \langle (t_0, l_0, z0_0, z1_0, z2_0, z4_0), \ldots, (t_N, l_N, z0_N, z1_N, z2_N, z4_N) \rangle$. Next, as in [4], we assume that the first occurrence of a location l_x in $\mathcal{T}_u$ is an *exploration (cf. e)*, else it is a *return (cf. r)*. Thus, before browsing the mobility traces, each user u has an empty set of known locations $\mathcal{L}_u$. We then add movement semantic to each record $q_x \in \mathcal{T}_u$ in the mobility trace, by associating the label r in case $l_x \in \mathcal{L}_u$, otherwise we associate the label e as depicted in Figure 2. When a location is first met, it is added to the set of known locations $\mathcal{L}_u$.

Subsequently, we propose two approaches to forecast the next type of movement an individual will perform.

Successive Types of Movements Predictor (STMP): In the first approach, we ignore the temporal dimension. In other words, only the order of occurrence of the types of visits is considered but not the elapsed time. To forecast the type of the $N + 1^{th}$ movement of the user u, we construct the table exp_u that contains the number of successive explorations within the mobility trace $\mathcal{T}_u$ that comprises N records. When a user starts exploring a counter $nb\_exp \leftarrow 1$ is started. After each consecutive explorations, the counter is incremented $nb\_exp \leftarrow nb\_exp + 1$ until meeting a return or the end of the trace $\mathcal{T}_u$, the value of the counter is saved in the table exp_u and reset to 0. Each time an exploration event occurs after a return the process restarts again until reaching the end of the sequence (cf. Algorithm 1, lines 4–6). Likewise, we construct a table ret_u that contains the number of successive returns within $\mathcal{T}_u$ (cf. Algorithm 1, lines 12–13). Following, we compute two values to characterize exploration visits, $\mu_{exp} = mean(exp_u)$ and $\sigma_{exp} = std(exp_u)$ that are the average and standard deviation (respectively) of successive explorations (cf. Algorithm 1, line 20). Similarly, we compute the average and standard deviation of successive returns (cf. Algorithm 1, line 21). ***Final decision:*** According to the last type of movement, if the number of the successive same type of movement is included in the interval $[\mu_{type} \pm \sigma_{type}]$ with $type \in \{e, r\}$, then predict the same movement as next, else predict the opposite movement (cf. Algorithm 1, lines 22–26). For instance, if the last movement was an exploration e and the current number of successive exploration nb_exp_u is included in the interval $[\mu_{exp} \pm \sigma_{exp}]$, then an exploration e is predicted, else a return r is predicted.

Inter Exploration Interval Predictor (IEIP): Temporal occurrence of exploration visits is the main and unique parameter considered in prediction decisions. To predict the $N + 1^{th}$ type of movement, we consider the trace $\mathcal{T}_u$ of size N. We focus on exploration events, as previously shown in [2] the temporal exploration activities appear to be regular. Therefore, we compute the *Inter-Exploration Interval (IEI)*, i.e., the elapsed time between two consecutive explorations [2] (cf. Algorithm 2, lines 2–5). ***Final decision:*** If the elapsed time since the last exploration event is included in the interval $[\mu_{IEI} \pm \sigma_{IEI}]$, we forecast that the next movement is an exploration else we forecast a return (cf. Algorithm 2, lines 9–13).

4.2 Spatial Prediction

Recall that return visits are highly foreseeable due to their high temporal periodicity and partial regularity [4]. Likewise, exploration visits are not completely random if we consider a coarse-grained spatial scope [2]. In what follows, we propose two spatial predictors leveraging the results of purpose predictors to improve exploration-like forecasts: (i) Personal Spatial Predictor (PSP), (ii) Joint Spatial Predictor (JSP). The description of such predictors is preceded by the

Algorithm 1 Successive Types of Movements Predictor – STMP –

function *STMP* $(\mathcal{T}_u, \mathcal{L}_u)$
for l **in** $\mathcal{T}_u$ **do** ▷ Successive same types of movement calculation
  if $l \notin \mathcal{L}_u$ **then** ▷ Explorations
    $nb\_exp_u \leftarrow nb\_exp_u + 1$
    $\mathcal{L}_u.ADD(l)$
    $last \leftarrow$ **e**
    if $nb\_ret_u > 0$ **then** ▷ Successive returns interruption
      $ret_u.ADD(nb\_ret_u)$
      $nb\_ret_u \leftarrow 0$
    end if
  else ▷ Returns
    $nb\_ret_u \leftarrow nb\_ret_u + 1$
    $last \leftarrow$ **r**
    if $nb\_exp_u > 0$ **then** ▷ Successive explorations interruption
      $exp_u.ADD(nb\_exp_u)$
      $nb\_exp_u \leftarrow 0$
    end if
  end if
end for
$\mu_{exp}, \sigma_{exp} \leftarrow Stats(exp_u)$
$\mu_{ret}, \sigma_{ret} \leftarrow Stats(ret_u)$
if $last =$ **e** and $nb\_exp_u \in [\mu_{exp} \pm \sigma_{exp}]$ or $last =$ **r** and $nb\_ret_u \notin [\mu_{ret} \pm \sigma_{ret}]$ **then**
  return e ▷ Predict an exploration
else
  return r ▷ Predict a return
end if
end function

Algorithm 2 Inter Exploration Interval Predictor – IEIP –

function *IEIP* $(T_u, \mathcal{L}_u)$
for (t, l) **in** T_u **do** ▷ IEI sequence computation
  if $l \notin \mathcal{L}_u$ **then**
    $tab\_exp\_interval.ADD(t - last)$
    $last \leftarrow t$
  end if
end for
$\mu_{IEI}, \sigma_{IEI} \leftarrow Stats(tab\_exp\_interval)$
if $t_{N+1} - last \in [\mu_{IEI} \pm \sigma_{IEI}]$ **then**
  return e ▷ Predict an exploration
else
  return r ▷ Predict a return
end if
end function

introduction of three prediction methods used by PSP or JSP according to the type of movement issued by the purpose predictors.

4.2.1 Prediction methods

Considering the mobility trace $\mathcal{T}_u = \langle (t_0, l_0, z0_0, z1_0, z2_0, z4_0), \ldots, (t_N, l_N, z0_N, z1_N, z2_N, z4_N) \rangle$, of the user u, with N records, we firstly leverage three distinct methods (cf. ME) for next visits' prediction, accordingly adjusted to operate (i) on two spatial resolutions (i.e., location or zones) and (ii) with two mobility views (i.e., personal or joint):

(ME) MC Location-Predictor: This predictor gets as input the mobility trace of an individual only. To predict the next location where the user will go, the *MC Location-Predictor* considers the stochastic sequence of the visited locations $x_0^N = l_0, l_1, \ldots, l_N$. Next, it trains a first-order MC predictor on the $N_s = \frac{2}{3} \times N$ first elements. Following, the *MC Location-Predictor* forecasts the next location l_{N+1} that will be visited by the user.

(ME) Personal Zone-Predictor: We slightly modify the *MC Location-predictor* to predict the future coarse-grained zone, instead of locations, where the user will be. It considers the stochastic sequence of visited zones $x_0^N =$

$zx_0, zx_1, \ldots, zx_N$. Then, similar to the *MC Location-Predictor*, it trains a first-order MC predictor using the first $N_s = \frac{2}{3} \times N$ elements of the coarse-grained sequence. Afterward, it forecasts the next visited zone zx_{N+1}.

(ME) Joint Zone-Predictor: We go further here and design a benchmark predictor that leverages the collective exploratory mobility behavior of the population, as input. Besides, prediction is done in terms of zones where a user will perform an exploration, in case she is more prone to discover a new place. First, it constructs an Exploration Origin-Destination (EOD) matrix $\mathbf{E}(t)$ at time t. The matrix gives an estimation of the probability to make a discovery in a zone j after visiting the location i. More precisely, the EOD matrix is of size $n \times m$, where n is the number of different "Origins" and m is the number of the different "Destinations". The "Origins" set contains *only locations* after which explorations happened. Hence, it is at most equal to the total number of locations visited by the users. The "Destinations" contains *all the distinct zones* where users explore. Each $\mathbf{e}_{i,j}$ gives the probability of exploring in the zone j after visiting the location i. For instance, in Figure 4, if the user in the location $L3$ is more prone to explore at time t, the *Joint Zone-Predictor* first constructs the EOD(t) matrix. Next, it identifies the most likely zone X where users usually explore after being in $L3$, i.e., the zone with the highest e_{L3X}. Finally, it suggests the zone $X = Z1$ as the next spatial unit where the user will explore, i.e., discover a new location.

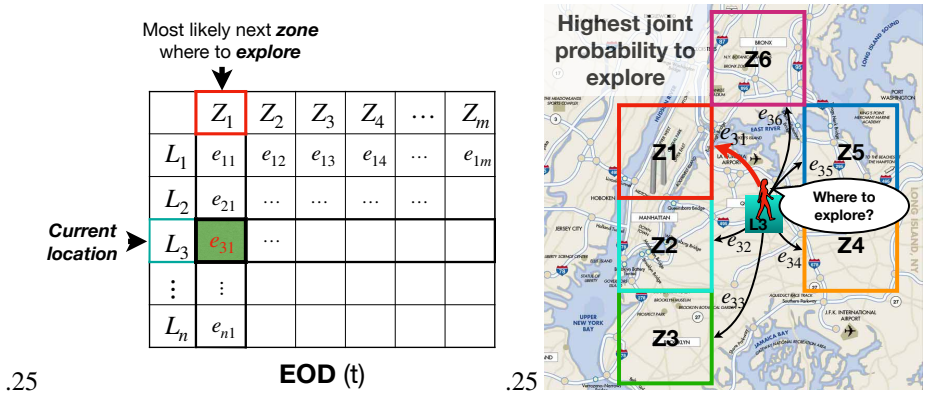


Figure 4: Joint Zone-Predictor.

4.2.2 Designed Spatial Predictors.

Using the aforementioned prediction methods, we design two spatial predictors:

Personal Spatial Predictor (PSP): The PSP takes as input the predicted type of movement. In case, the forecasted movement is a return, it uses the *MC Location-Predictor* to forecast the next visited location. Hence it provides a fine-grained intuition on where the user will be next. On the other side, if the predicted movement is an exploration, the PSP employs the *Personal Zone-Predictor*. Accordingly, a coarse-grained spatial unit is returned.

Joint Spatial Predictor (JSP): Similar to the PSP, the JSP takes the outcomes of a movement predictor as an entry. If the forecasted movement is a return, the *MC Location-Predictor* is used to infer the next location. Else, the *Joint Zone-Predictor* is used to infer the zone where an exploration might occur.

Note that the geographical accuracy of the proposed spatial predictors decays when the considered user is assumed to be exploring. Although this decay in performance, the inferred zones are of reasonable size that might lure and benefit many applications, such as recommendation systems or Multi-access Edge Computing infrastructures improvement.

5 Experimental settings

We present next the used data sources and describe the procedure ensued to complete the missing records and to filter out bad users. Finally, we present a simple mobility profiler that we apply to discuss the optimal use of the proposed prediction framework, whether it benefits the whole population or is more suited to users exhibiting a high exploration activity.

5.1 Datasets

Three real-world GPS mobility traces are used to evaluate the performance of our proposed exploration-aware predictor. The characteristics of the datasets are detailed in Table 1.

Macaco [20]: it collects 900k GPS records of users from more than 6 different countries lasting for about 34 months (from May 2015 to April 2018) with a frequency of one sample each 5 minutes. Each record includes a user ID, a timestamp, and location information, i.e., GPS coordinates (latitude and longitude). The dataset was collected by the MACACO project, but for project-related privacy policies, this dataset is not publicly available.

Privamov: [21]: it contains around 156 million GPS records of 100 volunteers from the city of Lyon in France. The data collection spans 15 months, from October 2014 to January 2016 with a frequency of sampling roughly equal to a few seconds. Every tuple consists of four parts, an anonymized user ID, the date of collection, the latitude, and longitude. The dataset was collected by the Privamov sensing campaign and is available on request. **Geolife:** [22]: it is a public dataset collected by Microsoft Research Asia, it collects more than 900 k GPS records of 182 individuals distributed in over 30 cities mainly in China, the USA, and Europe. The data collection lasted more than 64 months, from April 2007 to August 2012 with a frequency of 1 sample every 1≈5 seconds. Each GPS tuple contains an anonymized user ID, a timestamp, and location information (latitude, longitude, and altitude).

Dataset	Users	Duration	Frequency	Records
Macaco [20]	132	34 months	5 min	900 k
Privamov [21]	100	15 months	few seconds	156 M
Geolife [22]	182	64 months	1 to 5 seconds	900 k

Table 1: Datasets description.

5.2 Data handling

We extract the GPS mobility trajectory of each individual u, $\mathcal{T}_u = \langle (t_0, l_0, z0_0, z1_0, z2_0, z4_0), \ldots, (t_N, l_N, z0_N, z1_N, z2_N, z4_N) \rangle$. To eliminate the harmful sparsity effects, we complete the mobility traces according to steps in [23]. We identify three locations per user:

- **Workplace A** (l_{wp_A}): the most frequent daily location between 10 am and 11 am.
- **Workplace B** (l_{wp_B}): the most visited location between 2 pm and 5 pm.
- **Home** (l_H): the most prevalent place between 2 am and 6 am (night).

If a record is missing at time t_x between 10am – 11am / 2pm – 5pm / 2am – 6am, we complete the mobility trajectory $\mathcal{T}_u$ with a new tuple, with the timestamp t_x, a location $l_{wp_A}/l_{wp_B}/l_H$, and the associated zones.

Afterwards, we define a *complete day* for the GPS datasets as a day in which an individual has *on average* one record each 15 min. We filter out "bad" users and select only participants who have at least 1 month of complete days of data (i.e. have between 2688 and 8064 records). We are left with 266 users: 84 in Macaco, 77 in Privamov, and 105 in Geolife.

After filtering and completing the datasets, we uniformize their frequency of sampling and take one month of data for all the users. Finally, given the small number of users in each dataset, we aggregate them to have on trace with 266 users that we label *Agg_gps*.

To better understand the characteristics of the mobility traces and the types of movements (transition) habits of the population, we analyze the detailed information of data and draw several statistical features in Appendix A.1.

5.3 Identifying hard- and easy-to-predict users

Exploration activities are a key reason for the low accuracy of mobility prediction tasks [4, 2]. Individuals exhibiting a high tendency to explore are hence less likely to be foreseeable with predictors relying on past history. Therefore, to examine the efficiency of our proposed mobility prediction framework and investigate if it is more fitted to users who exhibit high exploration activities or is beneficial to all users. We propose a simple mobility profiler that seeks to isolate *hard-to-predict* from *easy-to-predict* users, according to their level of exploratory activities. For each individual u, with an N-long mobility trajectory, $\mathcal{T}_u = \langle (t_0, l_0, z0_0, z1_0, z2_0, z4_0), \ldots, (t_N, l_N, z0_N, z1_N, z2_N, z4_N) \rangle$, we train a simple MC predictor on the first $N_s = \frac{2}{3}N$ records and predict the future location $l_{N_{s+1}}$. Then, we increment N_s and repeat until $N_s = N$. Afterward, we compute the accuracy of prediction that is a commonly used prediction evaluation metric, and is given by, $accuracy = \frac{\text{number of correct predictions}}{\text{total number of predictions}}$. We also compute the exploration ratio for each user, $\alpha = \frac{\text{number of transitions of type exploration}}{\text{total number of transitions}}$.

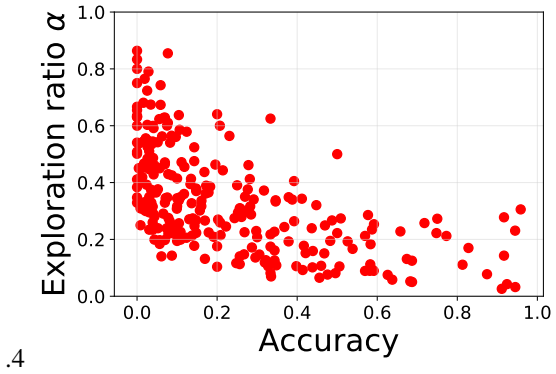


Figure 5: All users.

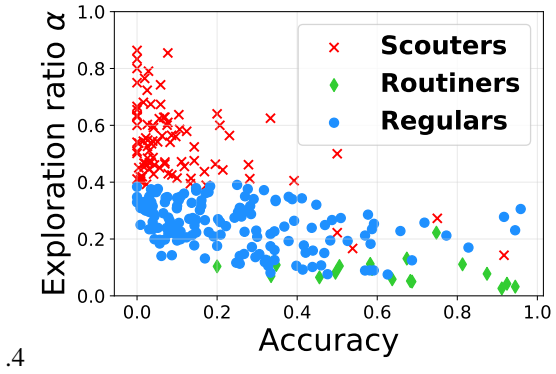


Figure 6: Per profile as proposed in [2].

Figure 7: Accuracy of prediction vs exploration ratio.

Figure 5 depicts the accuracy of prediction achieved by a first-order MC predictor against the exploration ratio. In general, we can observe that the MC predictor performs poorly with users having a high exploration ratio, particularly those holding a ratio above 0.4.

In Figure 6, we apply to the previous Figure 5 the profiling proposed in [2]. *Scouters* are defined as users with a high tendency to explore, *Routiners* are individuals who rarely interrupt their returning routine to explore, and *Regulars* exhibit an intermediate behavior. We can observe that *Scouters* typically have an exploration ratio above 0.4 and hold the lowest predictive scores.

Accordingly, hereafter, we will first evaluate the performance of the proposed predictor on the whole population. Next, we will apply the proposed framework only on the *hard-to-predict* users and employ a simple MC with the rest of the population as depicted in Figure 8.

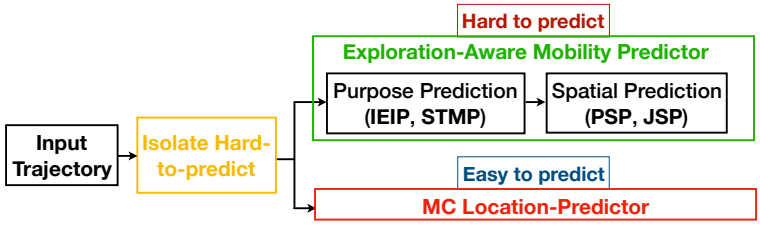


Figure 8: Global Prediction Framework.

We use the exploration ratio metric to determine if a user is *hard-to-predict* or not and variate the selection threshold in the set $Th = \{0.2, 0.4, 0.6\}$. So, a user holding an exploration score above a given threshold is classified as *hard-to-predict*, else as an easy-to-predict individual. Hence, low Th results in a higher number of hard-to-predict users

to whom we apply a double prediction (movement and spatial). Whereas, high Th allows selecting users exhibiting high exploratory activities to be a candidate for a double prediction, while others are subject to a direct spatial prediction.

According to Figure 8, a ratio below 0.2 isolates *Routiners*, who get high accurate predictions with traditional MC due to their *easy-to-predict* mobility. The ratio range [0.2,0.4] isolates Regulars users. Finally, the values higher than 0.6 identifies *Scouters* or *hard-to-predict* users (i.e., high probable users to get low traditional MC prediction accuracy) requiring the improvement of prediction methods. *Scouter* users trigger the exploration-enhanced MC predictor.

6 Movement Prediction Evaluation

Recall that our first goal is to infer an individual's next type of movement given the movement history. In what follows, we evaluate the performance of each of the STMP Algorithm 1 and the IEIP Algorithm 2. Given the imbalanced ratio between exploration and return transitions (see AppendixA.1 Figure 34), to measure the performance of the proposed movement predictors in forecasting each type of movement, we employ three widely used information retrieval measures:

- **Precision** P: it is a measure of relevance, it shows the ability of a classifier in not labeling as positive a sample that is negative. It is defined as the number of true positives T_p over the number of true positives plus the number of false positives F_p [24], $P = \frac{T_p}{T_p+F_p}$.
- **Recall** R: it measures the ability of a classifier in finding all positive samples. It is defined as the ratio between true positive T_p samples over the number of true positive T_p plus the number of false negatives F_n [24], $R = \frac{T_p}{T_p+F_n}$.
- f_1**-score** $F1$: it is the harmonic mean of precision and recall [24], it is given by, $F1 = 2 \times \frac{P \times R}{P+R}$.

By considering returns as positive events and explorations as negative events, a true positive result T_p refers to the correct prediction of a return, a false positive result F_p indicates that an exploration is predicted to be a return, a false negative F_n refers to a return predicted to be an exploration, and a true negative result T_n related to the correct prediction of an exploration (see Table 2).

	Actual Return	Actual Exploration
Predicted Return	T_p	F_p
Predicted Exploration	F_n	T_n

Table 2: Matrix of correct and misleading results.

We compare our proposed methods with the performance of the widely used state-of-the-art MC predictor. As in conventional personal predictors relying on location data, the MC predictor assumes that the next location a user will visit can be found in the set of known places [4, 3]. This means that the MC model is constantly forecasting returns. Thus, it holds the best scores in terms of predicting returns as a type of movement. Moreover, in view of the large proportion of returns compared to explorations (see Appendix A.1 Figure 34), the MC allows evaluating how often the proposed algorithms are accurate in predicting the next type of movement. Alternatively stated, it helps to tune the movement predictors in favor of explorations/returns or to have global satisfying results.

Return visits: Figures 12 and 16 represent the performance of both of the STMP (Algorithm 1) and IEIP (Algorithm 2) in accurately predicting the occurrence of return and exploration events respectively. The proposed algorithms are first applied for the whole population, then only for hard-to-predict users while for the easy-to-predict users an MC predictor is applied. For the hard-to-predict users selection, as previously stated, we vary the exploration ratio's selection threshold α in the set $Th = \{0.2, 0.4, 0.6\}$, i.e., a lower threshold induces a higher labeling of hard-to-predict users.

Figure 12 shows the precision, recall, and $f1$-score achieved by the STMP, the IEIP, and the conventional MC predictor. In Figure 9, we can see that with more than 60% of the population STMP and IETP predictors achieve a precision above 70% in predicting returns. This indicates that more than 70% of the time the forecasted returns are real revisits to known places. We also notice that the IEIP reaches the highest scores, the precision value is above 90% for more than 40% of the users. Furthermore, applying the proposed algorithms only on users classified as hard-to-predict engenders slight changes in the precision scores when predicting returns.

Figure 10 depicts that the STMP succeeds at least 80% of the time in predicting returns for 80% of the population. On the contrary, the IEIP that focuses on exploration visits comes off badly in forecasting returns when applied to all users. Furthermore, we can observe that applying the proposed algorithms only on users exhibiting high exploratory activities allows improving the recall score, notably, the IEIP. Actually, reducing the proportion of users to whom we apply the

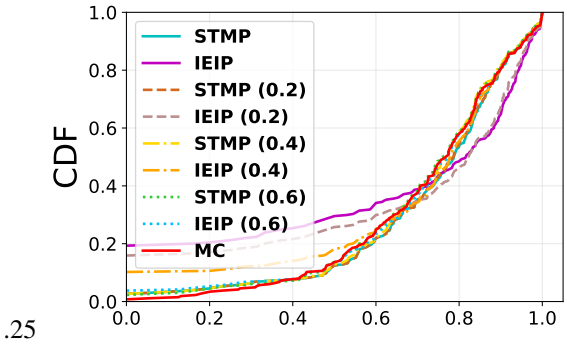


Figure 9: Precision

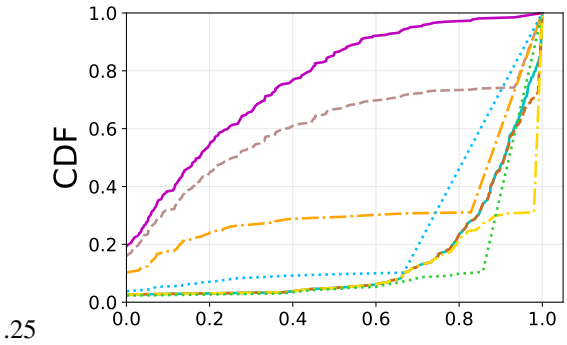


Figure 10: Recall

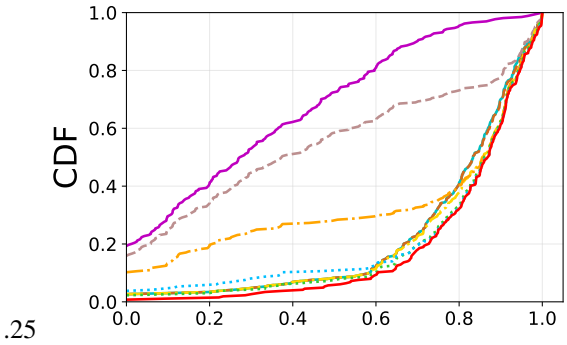


Figure 11: F1-Score

Figure 12: Performance comparison for returns forecasts (better seen in color).

proposed algorithms leads to an increase in the probability of predicting returns. Hence, the obtained recall scores are expected to improve with the decrease in the size of the selected hard-to-predict users. Curiously, the best average scores that we obtain with both of the STMP and IEIP movement predictors are when applying them to users having an exploration ratio above 40%, which approximately corresponds to the *Scouter* profile proposed in [2] [1].

As reported by the weighted average of precision and recall in Figure 11, the STMP performs better than the IEIP in predicting returns. Namely, the average $f1$-score held by STMP exceeds 80%, and in general, its performance is very close to the MC's. On the contrary, the average $f1$-score reached by IEIP is less than 40% when applied to all users. Furthermore, Figure 11 reveals that increasing the exploration ratio for hard-to-predict user selection increases the achieved performance by both algorithms.

Exploration visits: Figure 16 presents the performance evaluation of the STMP and IEIP only, provided that a conventional predictor such as MC will always predict returns and fail at each discovery of a new location.

[1] Note that in Figure 10 the curve corresponding to the MC is not represented as the recall score is equal to 1 for all users.

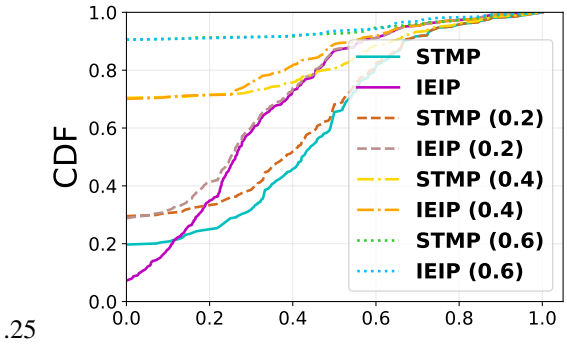


Figure 13: Precision

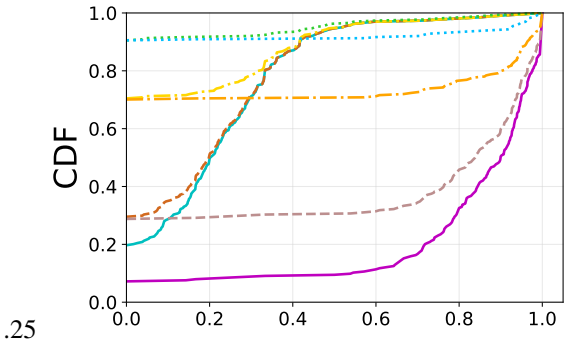


Figure 14: Recall

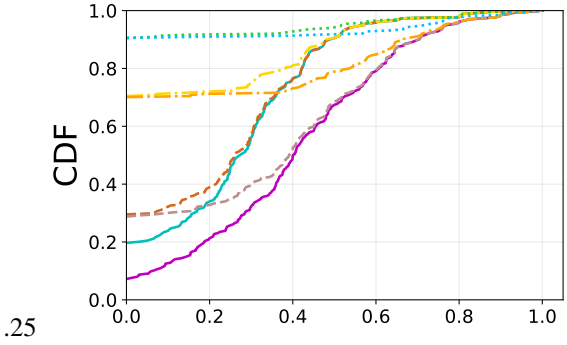


Figure 15: F1-Score

Figure 16: Performance comparison for exploration type of movement forecasts (better seen in color).

Figure 13 shows that the precision achieved by STMP in predicting explorations for 60% of the population surpasses 35%. Whereas for the same proportion of the population only 20% of the explorations inferred by IEIP are in fact discoveries of new places. The partial application of the proposed STMP and IEIP to the population leads to a decrease in the attained precisions. Indeed, for users classified as easy-to-predict, the application of the MC to them induces null scores when predicting explorations.

Figure 14 depicts that for 60% of the population at least 18% of the time explorations are correctly predicted by the STMP when applied for all users. Conversely, for the same percentage of users, more than 80% of moments of discovery are accurately foreseen by the IEIP. Similarly, to the measure of precision, the recall decreases with the decrease in the size of the selected population.

Figure 15 shows that the IEIP outperforms the STMP in predicting explorations, whether it is applied to all users or for the categories of users with a high proclivity to explore.

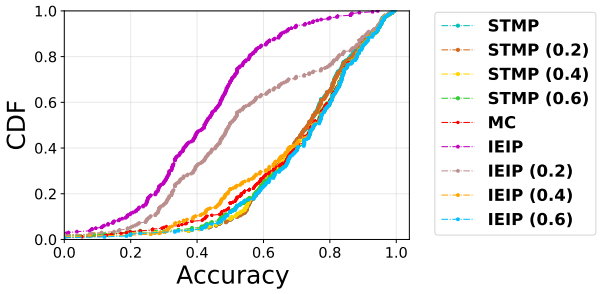


Figure 17: Accuracy of prediction of the type of movement.

General visits: Next, we want to investigate how often the proposed movement predictors are accurate in predicting the next type of movement. We compute the accuracy of prediction achieved by the MC predictor and each of the proposed movement predictors. Here, the accuracy of prediction is the ratio between the number of correctly predicted types of movement and the total number of predictions.

Figure 17 shows the accuracy of prediction in terms of types of movement achieved by STMP, IEIP, and MC. It is noted that apart from IEIP and IEIP (0.2), the other predictors perform almost equally well, but with a slight dominance of the IEIP (0.6). The accuracy of prediction is on average above 75% for the predictors except for the IEIP and IEIP (0.2) that have a score around 50%. Based on the aforementioned results, we can draw two main conclusions. First, the strategy adopted by conventional predictors, as MC, assumes constant returns what is a key lowering factor for the predictive performance. Figure 17 shows that on average more than 20% of users' transitions are explorations. Second, the proposed movement predictors are faced with a trade-off, gaining accuracy in predicting explorations at the cost of losing efficiency in forecasting returns. We point out that the proposed movement predictors are preliminary versions, that we aim to propose advanced versions in our future works. We claim here that the proposed movement predictors can be tuned to fit the requirements and needs of the using applications, unlike conventional models that focus on returns only.

7 Spatial Prediction Evaluation

We evaluate here the predictive power of the two proposed spatial predictors. First, the outcomes of the STMP and the IEIP predictors with their variations feed the spatial predictors. For the computation of the prediction accuracy, we consider a prediction to be correct if the inferred *location* or *zone* is correct. This is done through a comparison with three baseline predictors operating on different spatial scales: *MC Location-Predictor*, *MC Location-Oracle-Predictor*, and *MC Zone-Oracle-Predictor*.

- **MC Location-Predictor:** It is described in Section 4.2.
- **MC Location-Oracle-Predictor:** It performs as well as the *MC Location-Predictor* in predicting returns, but reaches perfect scores in forecasting explorations. Given the mobility trace of the user u, $\mathcal{T}_u = \langle (t_0, l_0, z0_0, z1_0, z2_0, z4_0), \ldots, t_x, l_x, z0_x, z1_x, z2_x, z4_x) \rangle$, if next the user makes a transition to a location l_{x+1} that is not present in $\mathcal{T}_u$ the *MC Location-Oracle-Predictor* will accurately predict l_{x+1}. This predictor holds the best feasible scores that an MC predictor endowed with a perfect movement predictor and spatial exploration forecaster can achieve.
- **MC Zone-Oracle-Predictor:** It follows the same strategy as the *MC Location-Oracle-Predictor*. Yet, it operates on a coarse-grained spatial resolution. Instead of predicting the next visited location, it predicts large zones. Given the mobility trace $\mathcal{T}_u = \langle (t_0, l_0, z0_0, z1_0, z2_0, z4_0), \ldots, (t_x, l_x, z0_x, z1_x, z2_x, z4_x) \rangle$, this predictor forecasts the next zone zc_{x+1} of size c km $\times$ c km where the user is going to move. Besides, it is always accurate in predicting the right zone in case the user is exploring a new place. It holds the best achievable performance and we take it as a reference to evaluate the efficiency of the proposed framework.

In what follows, we compare the performance of the PSP and JSP spatial predictors with the three baselines. As input, the spatial predictors receive the results from the STMP and IEIP movement schemes with their different settings. As previously indicated, locations are squared cells of size $(200m)^2$. We consider 4 distinct sizes for the zones: $(800m)^2$, $(1km)^2$, $(2km)^2$, and $(4km)^2$.

PSP: Figure 22 reports the prediction accuracy of the PSP predictor with the different inputs and zone sizes. First, we can see that the prediction accuracy of the PSP with the distinct inputs outperforms the *MC Location-Predictor*'s scores.

Second, applying the proposed framework to larger proportions of users allows increasing the overall accuracy of prediction. We have two hypotheses with regard to the last observation: (a) applying the proposed framework is relevant to the whole population (b) by increasing the number of considered users, the number of false-negative forecasts (i.e., returns predicted to be explorations) increase and given the coarse-grained spatial prediction, in this case, the predictive performances are enhanced [4]. Third, with the expansion of the size of the zones, the accuracy of prediction of the PSP combined with the different movement predictors grows to approach the *MC Location-Oracle-Predictor* performance. Notably, the accuracy of prediction is substantially improved when considering zones of size $(2km)^2$ or zones of size $(4km)^2$. We can also see that the PSP fed with the IEIP algorithm applied to the whole population slightly surpasses the score obtained by the *MC Location-Oracle-Predictor*. This is mainly due to false-negative forecasts.

JSP: Figure 27 depicts the prediction accuracy of the JSP and its different settings and the baseline spatial predictors. Unexpectedly, the accuracy of prediction is at most equal to the *MC Location-Predictor*. Besides, expanding the zone size helps in improving the minimal achieved scores only, but not the overall performance.

Spatial prediction for each type of movement: Next, to understand how the spatial predictors perform in predicting the locality of each type of movement. We report in Figure 30 and Figure 33 the CDF of the accuracy of the spatial prediction for each type of movement by the PSP and JSP respectively. We depict the results with zones of size $(4km)^2$. For other spatial resolutions of the zones, we provide descriptive tables in Appendix A.2. Figure 30 shows that the accuracy of predicting returns and explorations by the PSP predictor. First, we can see that the performance achieved by the combination of the PSP with the STMP in predicting returns is very close to the MC's, with slight improvements when applying partially the proposed framework to users exhibiting a high exploration activity. Specifically, when applying it to users having an exploration ratio α above 40% (see Figure 28). When the PSP takes as input the IEIP's outcomes, the improvement in the predicting returns is more noticeable. Namely, when using the proposed framework with hard-to-predict users (see Figure 29).

On the contrary, whereas the prediction accuracy of the *MC Location-Predictor* in forecasting explorations is equal to zero, the PSP achieves appealing performance. Notably, it is more beneficial when using it with all users. Recall that for users classified as easy-to-predict we apply the *MC Location-Predictor* for the forecasts.

In Figure 33, we depict the accuracy of predicting the locality of each type of movement by the JSP. First, for return forecasts, the JSP and PSP are alike, in view of the fact that they rely on the same location predictor and take the same inputs. Second, the difference in performance between the JSP and PSP emanates from exploration forecasts. Compared to the SPS, the JSP works poorly in predicting the locality of explorations. This suggests that the overall weak predictive power of the JSP presented in Figure 27 follows from the low potential of the Joint Zone-Predictor in forecasting the locality of explorations. *Unlike return patterns that can be common to many individuals that are strangers to each other, exploration patterns are more personal. Furthermore, we measured the Jaccard similarity for the top 5 most visited zones when exploring between users within the same city (Beijing). We report very lows scores, the similarity is at most equal to 0.22. This implies, that when it is about forecasting explorations it is better to rely on the individual's mobility behavior than looking at the global patterns.* Note, that similarities in exploration spatial patterns might be observed among users within the same social circle and who are not perfect strangers to one another.

8 Conclusions and Future Work

Individuals' novelty-seeking tendencies in mobility bring impacting effects to mobility prediction. We tackle this problem and propose a novel 2-step adaptive prediction framework composed of (i) a *purpose prediction* (i.e., exploration or return) that feeds (ii) a *spatial prediction*. Contrary to existing methods, the proposed framework does not naively rely on mobility return's regularity, but adjusts its forecasts when explorations are more probable to happen. We designed two purpose prediction algorithms that base their forecasts on coarse- or fine-grained regularities observed in exploratory and return visits. We then develop two spatial prediction tasks that take the outcomes of the purpose predictors as input and operate on different spatial scales. The two spatial predictors are similar in predicting returns, but employ different strategies in case the input is an exploration. The first relies its forecasts on the personal data of the considered user, whereas the second one exploits common exploratory tendencies among the population to infer the next coarse-grained zones. Not that, though opening privacy concerns, the leveraging of population mobility data is not an issue for some entities – such as telecom operators in resource allocation planning – due to their natural global view of the network population. Our proposed framework achieves interesting results both in inferring the next type of movement as well as in forecasting the spatial occurrence of the visits. Moreover, they confirm that exploration visits are not completely random if the spatial resolution is increased. Besides, we find that explorations are more personal contrary to returns where common patterns are shared between users. Unlike conventional methods that are always wrong when explorations occur, the proposed framework does predict new visits for spatial units of excellent

precision when networks planning are concerned; and reasonable precision for more personalized applications, as for recommendation services.

For future work, we aim at investigating temporal visiting patterns associated with the purposes of movements. Correlating time of the day with purposes of movement may bring an additional dimension on the exploration handing.

References

[1] M. C. Gonzalez, C. A. Hidalgo, A. L. Barabasi. Understanding individual human mobility patterns. *Nature*, 453:779–782, Jun. 2008.

[2] L. Amichi, A. Carneiro Viana, M. Crovella, and A. A.F. Loureiro. Understanding individuals' proclivity for novelty seeking. In *ACM SIGSPATIAL '20*, page 314–324, 2020.

[3] C. Yu, Y. Liu, D. Yao, L. T. Yang, H. Jin, H. Chen, and Q. Ding. Modeling user activity patterns for next-place prediction. *IEEE Systems Journal*, 11(2):1060–1071, 2017.

[4] A. Cuttone, S. Lehmann and M. C. Gonzalez. Understanding predictability and exploration in human mobility. *EPJ Data Science*, 7(1), Jan. 2018.

[5] H. Barbosa, M. Barthelemy, G. Ghoshal, C. R. James, M. Lenormand, T. Louail, R. Menezes, J. J. Ramasco, F. Simini, and M. Tomasini. Human mobility: Models and applications. *Physics Reports*, 734:1 – 74, 2018.

[6] A. Asahara, K. Maruyama, A. Sato, and K. Seto. Pedestrian-movement prediction based on mixed markov-chain model. In *ACM SIGSPATIAL '11*, page 25–33, 2011.

[7] M. Killijian S.n Gambs and M. N. del Prado Cortez. Next place prediction using mobility markov chains. In *MPM '12*, 2012.

[8] G. Gidófalvi and F. Dong. When and where next: Individual mobility prediction. In *Proceedings of the First ACM SIGSPATIAL MobiGIS '12*, page 57–64, 2012.

[9] F. Alhasoun. City scale next place prediction from sparse data through similar strangers. 2017.

[10] Xin Lu, Erik Wetter, Nita Bharti, Andrew J. Tatem and Linus Bengtsson . Approaching the Limit of Predictability in Human Mobility. *Scientific Reports*, 3(2923), 2013.

[11] W. Mathew, R. Raposo, and B. Martins. Predicting future locations with hidden markov models. In *UbiComp '12*, page 911–918, 2012.

[12] R. Chan. The cambridge analytica whistleblower explains how the firm used facebook data to sway elections, May 2020.

[13] K. Bonawitz, V. Ivanov, B. Kreuter, A. Marcedone,H. B. McMahan, S. Patel, D. Ramage, A. Segal, and K. Seth. Practical secure aggregation for privacy-preserving machine learning. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, page 1175–1191, Oct. 2017.

[14] A. Y. Xue, R. Zhang, Y. Zheng, X. Xie, J. Huang, and Z. Xu. Destination prediction by sub-trajectory synthesis and privacy protection against such prediction. In *Proceedings of The 29th IEEE International Conference on Data Engineering*, April 2013.

[15] Chao Z., Keyang Z., Q. Yuan, L. Zhang, Tim H., and Jiawei H. Gmove: Group-level mobility modeling using geo-tagged social media. In *Proceedings of the 22nd ACM SIGKDD*, KDD '16, page 1305–1314, 2016.

[16] L. Song, D. Kotz, Ravi J., and Xiaoning H. Evaluating next-cell predictors with extensive wi-fi mobility data. *IEEE T MOBILE COMPUT*, 5(12):1633–1649, 2006.

[17] H. Feng, C. Liu, Y. Shu, and O. WW Yang. Location prediction of vehicles in vanets using a kalman filter. *Wirel. Pers. Commun.*, 80(2):543–559, 2015.

[18] F. Calabrese, G. Di Lorenzo, and C. Ratti. Human mobility prediction based on individual and collective geographical preferences. In *13th IEEE T INTELL TRANSP*, pages 312–317, 2010.

[19] X. Lu, L. Bengtsson, and P. Holme. Predictability of population displacement after the 2010 haiti earthquake. *PNAS*, 109(29):11576–11581, 2012.

[20] K. Jaffres-Runser, G. Jakllari, T. Peng and V. Nitu. Crowdsensing Mobile Content and Context Data: Lessons Learned in the Wild. In *PerCom Workshops*, 2017.

[21] S. BenMokhtar, A. Boutet, L. Bouzouina, P. Bonnel, O. Brette, L. Brunie, M. Cunche, S. D'Alu, V. Primault, P. Raveneau, H. Rivano, R. Stanica. PRIVA'MOV: Analysing Human Mobility Through Multi-Sensor Datasets. In *NetMob*, Apr. 2017.

[22] W.Y Ma Y. Zheng, X. Xie. Geolife: A collaborative social networking service among user, location and trajectory. In *IEEE T KNOWL DATA EN*, volume 33, pages 32–40, 2010. https://www.microsoft.com/en-us/download/details.aspx?id=52367.

[23] G. Chen, A. Carneiro Viana, M. Fiore, and C. Sarraute. Complete Trajectory Reconstruction from Sparse Mobile Phone Data. *EPJ Data Science*, October 2019.

[24] C. Goutte and E. Gaussier. A probabilistic interpretation of precision, recall and f-score, with implication for evaluation. In *ECIR*, pages 345–359. Springer, 2005.

Acknowledgements

We would like to thank the research agencies CAPES, CNPq, FAPEMIG, and FAPESP (grant 18/23064-8) and the support from INRIA, Sorbonne UPMC, LINCS, ANR (French National Research Agency) MITIK project - call PRC AAPG2019.

A Appendix

A.1 Data

In Figure 34, we report the CDF of the proportions of transitions of types exploration (Exp) and returns (Ret). We can see that the majority of the population has a low exploration activity, more than 60% of the population has an exploration ratio lower than 40%. Inversely, all users depict a high returning activity, 80% of the users have a degree of return above 50%. Indeed, routine patterns are embedded in today's societies.

Figure 35 reports the average number of successive same type of movement $\mu(type)$ against the standard deviation $\sigma(type)$. First, we can see that the average number of successive returns is substantially higher than explorations. This means that individuals usually spend longer periods revisiting known places than discovering new ones. Moreover, we observe that the standard deviation takes small values for the average number of sequential exploration, implying that although the assigned randomness to this type of movement, explorations occurrences are less irregular than thought.

Figure 36 shows the hourly fluctuations of the IEI mean and standard deviation. We can see that the IEI mean varies throughout the week, it is lower during day time, particularly on weekends. This means that exploration activities are more numerous during the daytime and increase on weekends. Yet, we can also observe a regularity in exploratory visits during weekdays.

A.2 Spatial Prediction

In Table 3 and Table 4 we report the average accuracies of prediction achieved by the PSP and JSP in predicting the spatial occurrence of explorations while varying the spatial resolution.

PSP	**Explorations**			
Spatial Units	$(800m)^2$	$(1km)^2$	$(2km)^2$	$(4km)^2$
STMP	0.39	0.41	0.50	0.63
STMP (0.2)	0.26	0.28	0.35	0.44
STMP (0.4)	0.10	0.11	0.14	0.18
STMP (0.6)	0.03	0.03	0.05	0.06
IEIP	0.34	0.36	0.45	0.58
IEIP (0.2)	0.22	0.24	0.32	0.4
IEIP (0.4)	0.09	0.09	0.12	0.16
IEIP (0.6)	0.02	0.02	0.04	0.05

Table 3: Ratio of correctly predicted explorations and returns for the PSP.

From Table 3 we can see that the accuracy of predicting the spatial units where explorations might occur is relatively high for the PSP. Additionally, it increases with the increase of zones size. On the contrary, Table 4 shows that the JSP performs poorly in predicting the locality of exploration events, and the changes in performance are not noticeable with the change in the spatial resolution

JSP	**Explorations**			
Spatial Units	$(800m)^2$	$(1km)^2$	$(2km)^2$	$(4km)^2$
STMP	0.17	0.17	0.16	0.14
STMP (0.2)	0.11	0.11	0.11	0.10
STMP (0.4)	0.03	0.05	0.05	0.04
STMP (0.6)	0.01	0.01	0.02	0.04
IEIP	0.15	0.16	0.15	0.14
IEIP (0.2)	0.11	0.11	0.11	0.10
IEIP (0.4)	0.04	0.04	0.05	0.04
IEIP (0.6)	0.01	0.01	0.02	0.02

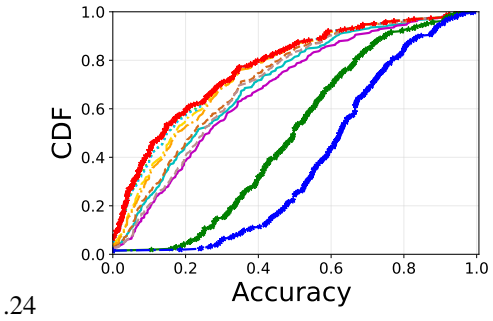
Table 4: Ratio of correctly predicted explorations and returns for the JSP.

Exploration-Aware Mobility Prediction Framework		
Modules	1- Purpose Prediction (Section 4.1)	2- Spatial Prediction (Section 4.2)
Description	Aims to predict the next type of movement: an exploration or a return	Aims to predict the spatial locality of the visits
Methods	• **STMP**: bases it forecasts on the average number of successive explorations and returns • **IEIP**: bases its forecasts on the time intervals between explorations	• **PSP**: bases its predictions by solely using. personal data • **JSP**: when the anticipated movement is an exploration, the spatial predictions is made by looking at the group-level exploratory behavior

Table 5: A general overview of the exploration-aware mobility prediction framework and the used methods.

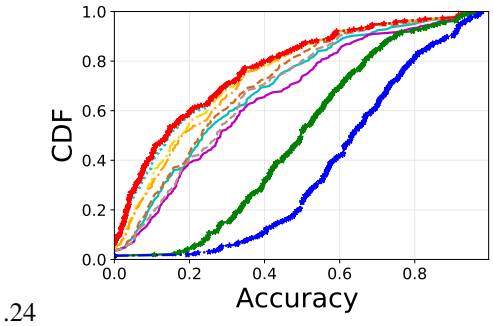
A.3 Exploration-aware mobility prediction framework

Table 5 gives a general overview of the algorithms employed within each prediction step of the exploration-aware mobility predictor.



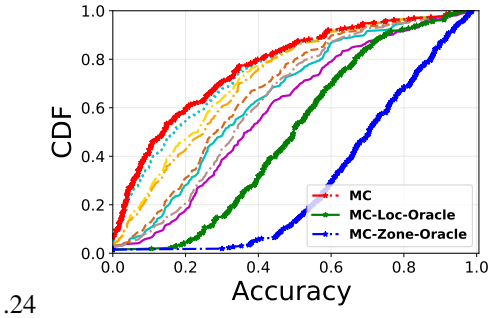
.24

Figure 18: Zones = 800 m × 800 m



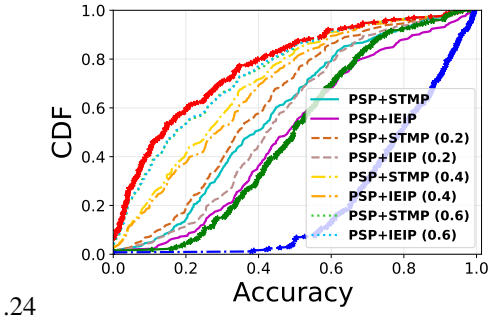
.24

Figure 19: Zones = 1 km × 1 km



.24

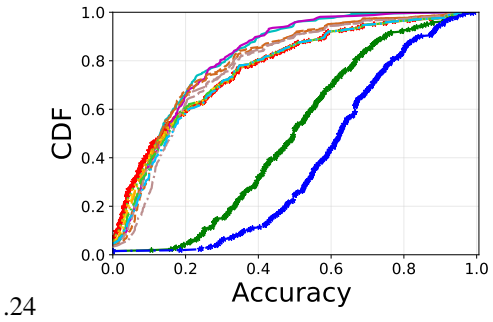
Figure 20: Zones = 2 km × 2 km



.24

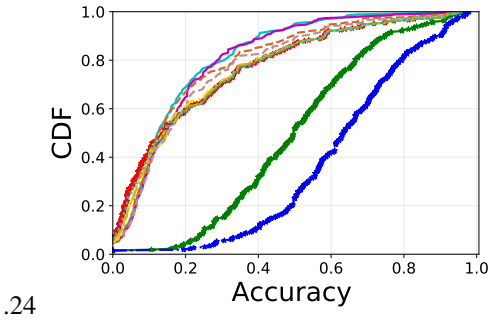
Figure 21: Zones = 4 km × 4 km

Figure 22: Accuracy of the Personal Spatial Prediction (common labels, better seen in color).



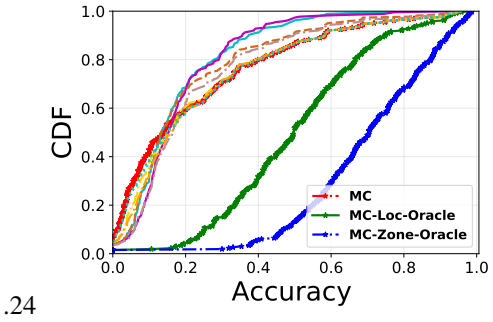
.24

Figure 23: Zones = 800 m × 800 m



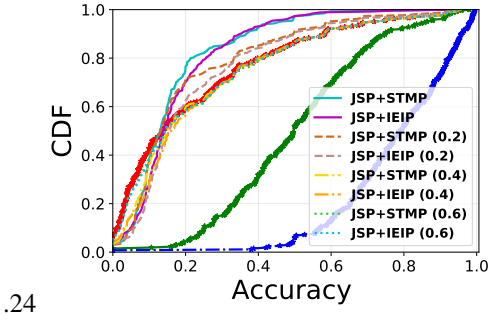
.24

Figure 24: Zones = 1 km × 1 km



.24

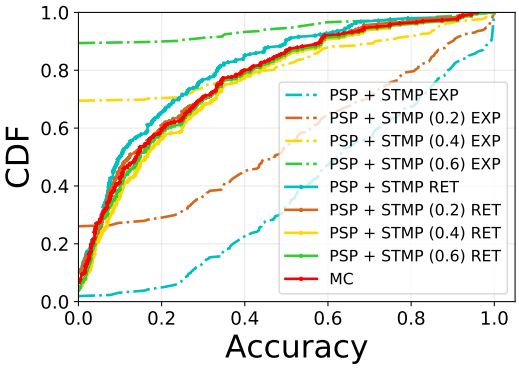
Figure 25: Zones = 2 km × 2 km



.24

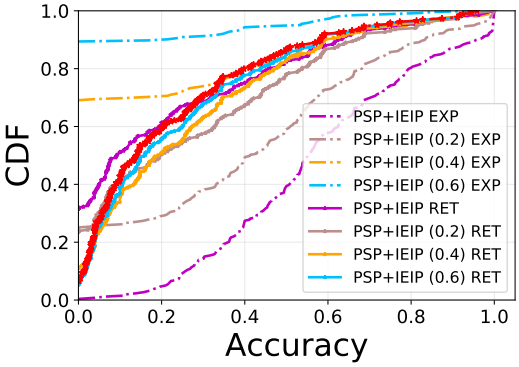
Figure 26: Zones = 4 km × 4 km

Figure 27: Accuracy of the Joint Spatial Prediction (common labels, better seen in color).



.24

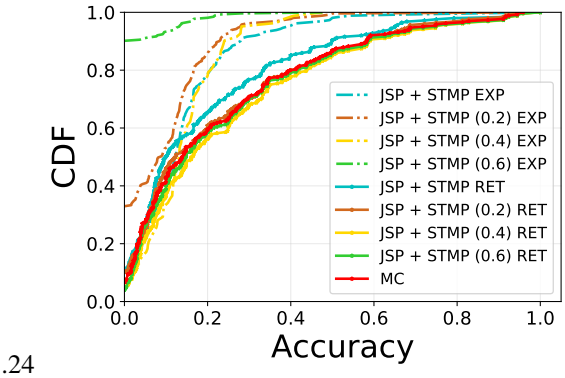
Figure 28: STMP



.24

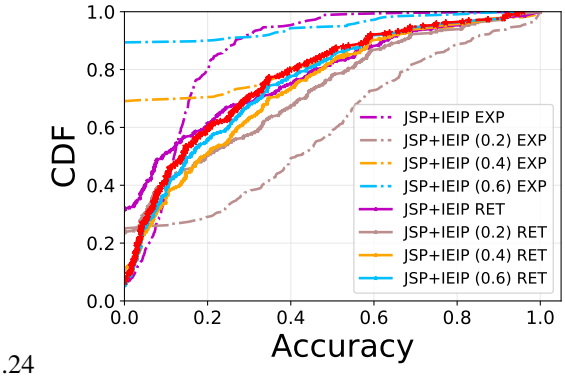
Figure 29: IEIP

Figure 30: Accuracy of the PSP for each type of movement.



.24

Figure 31: STMP



.24

Figure 32: IEIP

Figure 33: Accuracy of the JSP for each type of movement.

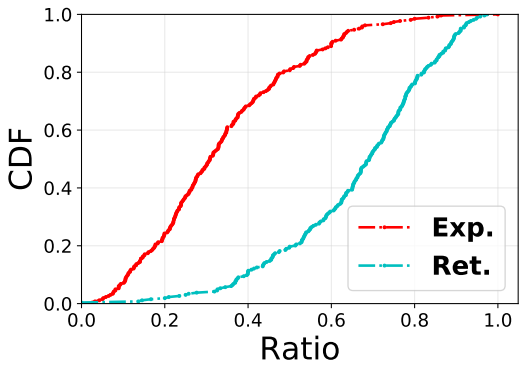


Figure 34: Ratio of types of transition.

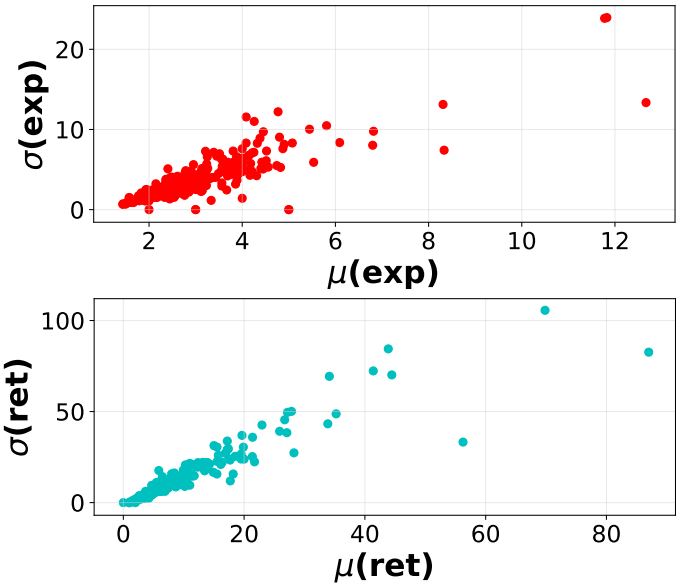


Figure 35: Number of successive types of movement mean and standard deviation.

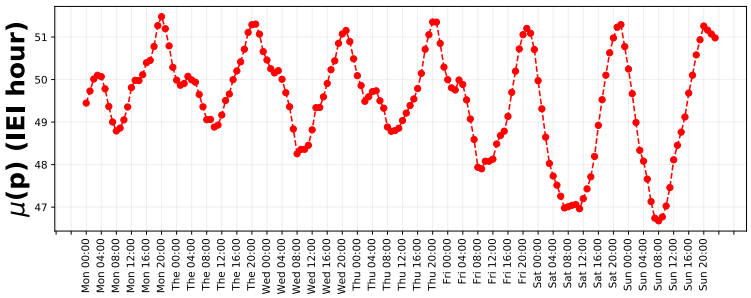


Figure 36: IEI mean.