



# On predictability of time series

Paiheng Xu <sup>a,b</sup>, Likang Yin <sup>a,b</sup>, Zhongtao Yue <sup>a,c,d</sup>, Tao Zhou <sup>a,c,\*</sup>

<sup>a</sup> Complex Lab, Web Science Center, University of Electronic Science and Technology of China, Chengdu 611731, People's Republic of China

<sup>b</sup> School of Hanhong, Southwest University, Chongqing 400715, People's Republic of China

<sup>c</sup> Big Data Research Center, University of Electronic Science and Technology of China, Chengdu 611731, People's Republic of China

<sup>d</sup> Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China, Chengdu 611731, People's Republic of China



## HIGHLIGHTS

- Time series prediction is a long-standing challenge with significant applications in disparate domains.
- Predictability indeed provides insights to the system itself.
- This work clarifies the method for predictability estimation.

## ARTICLE INFO

### Article history:

Received 6 September 2018

Received in revised form 1 February 2019

Available online 10 February 2019

### Keywords:

Predictability  
Human mobility  
Time series

## ABSTRACT

The method to estimate the predictability of human mobility was proposed in Song et al. (2010), which is extensively followed in exploring the predictability of disparate time series. However, the ambiguous description in the original paper leads to some misunderstandings, including the inconsistent logarithm bases in the entropy estimator and the entropy–predictability–conversion equation, as well as the details in the calculation of the Lempel–Ziv estimator, which further results in remarkably overestimated predictability. This paper demonstrates the degree of overestimation by four different types of theoretically generated time series and an empirical data set, and shows the intrinsic deviation of the Lempel–Ziv estimator for highly random time series. This work provides a clear picture on this issue and thus helps researchers in correctly estimating the predictability of time series.

© 2019 Elsevier B.V. All rights reserved.

## 1. Introduction

In recent years, the rapidly increasing usages of Global Positioning System (GPS), ranging from mobile phones, fitness bracelets and vehicle positioning systems, provide us with unprecedentedly rich information to capture and analyze human mobility patterns. As a result, the prediction of human mobility has received growing attention for its importance in traffic management [1–4], disaster response [5–7], epidemic prevention [8–10], and so on [11,12].

Though the prediction is getting more and more accurate due to advanced algorithms and the availability of vast data, it is not clear how well these algorithms perform versus the best possible prediction. Accordingly, the predictability of human mobility was proposed, which aims at measuring the theoretically maximum prediction accuracy  $\Pi_{max}$  for the given data. Song et al. proposed an entropic framework [13] to calculate the predictability  $\Pi_{max}$  by solving a limited case of the Fano

\* Corresponding author at: Complex Lab, Web Science Center, University of Electronic Science and Technology of China, Chengdu 611731, People's Republic of China.

E-mail address: [zhutou@ustc.edu](mailto:zhutou@ustc.edu) (T. Zhou).

inequality [14,15]. Empirical analysis on this basis [13] suggested that the predictability of human mobility could reach 93% on average.

Many scientists tried to design advanced predicting algorithms that can approach the predictability [16], such as the Markovian model [17], the neural network algorithm [18], the sequence-based model [4], and so on. The information from social networks [19,20], semantic labels [21], demographic characteristics [22], and activity patterns [23] are also utilized to improve the predicting algorithms.

Another interesting topic is to dig out significant factors that affect the predictability, such as the temporal and spatial resolution of data [24–28], the preference of exploration [25,29] and the data quality [30]. In particular, the predictability of next-timestep prediction is shown to have a very high upper bound ( $> 90\%$ ), which is mostly due to the stationarity in the human mobility [31,32]. This makes the next-place prediction a more challenging and attractive topic.

Some other researchers have tried different methods or modified entropic measures to quantify the predictability of human mobility, such as mutual information [33,34], instantaneous entropy [35,36], a contextual model which allows predictability to be assessed as the accuracy of the model in making predictions [37], and so on. Smith et al. [38] integrated real-world topological constraints into the calculation of the upper bound and presented a refined predictability of human mobility. To remove the stationarity in human mobility, Ikanovic and Mollgaard [32] aimed at the next-place prediction and proposed an alternative approach independent of the temporal scale. Yao et al. [39] proposed forecast entropy to measure the difficulty of predicting an observed time series based on the distributions of the time series in different spaces, and showed that the forecast entropy of a random system is clearly different from that of a deterministically chaotic system.

The analytical framework of predictability is also extended to other types of time series, such as human communication sequences [33,40,41], vehicular mobility [1–4,42], the IP address sequence of cyberattacks [43], stock price change [44], consumer visitation pattern [41,45], online user behaviors [20,34,46], electronic health records [47], and so on.

However, the ambiguous description in Ref. [13] leads to some misunderstandings, including the inconsistent logarithm bases in the entropy estimator and the entropy-predictability-conversion (EPC) equation, as well as the details in the calculation of the Lempel–Ziv (LZ) estimator. These misunderstandings will result in remarkably higher predictability than the true value. In this paper, we generate four types of artificial time series with controllable predictability, namely exploration sequence, random sequence, deterministic sequence and Markovian sequence, based on which we can quantify the deviation of theoretical estimation from the true value. We compare different possible understandings of the theoretical estimators in the literature, and clearly show the advantage of the right implementation, which is also consistent with previous theoretical entropy analyses on time series [48–50].

Many researchers have followed Ref. [13] to calculate the predictability of time series. Ciobanu et al. [51] studied mobile interactions collected at the Politehnica University of Bucharest and explored its predictability in the opportunistic networks. Li et al. [3] proposed an areas transition model to describe the vehicular mobility among the areas divided by the city intersections and examined the predictabilities of large-scale urban vehicular networks. Zhao et al. [4] obtained the theoretical predictability by entropy measure and used it to identify the effectiveness of different predicting algorithms. Xu et al. [42] defined travel time predictability as the probability to correctly predict the travel time by employing multiscale entropy. The predictability  $\Pi_{\max}$  obtained in the above-mentioned works may be higher than it supposed to be due to the unmatched choice of logarithm bases [11]. The present paper could refine our knowledge in the above-mentioned issues. This paper also raises a challenge about how to accurately estimate predictability for less-predictable time series since the LZ estimator fails in such case. The four types of artificially generated time series can also be treated as the touchstone for the validity of newly proposed methods in the future.

## 2. Theoretical analysis

Considering a historical sequence  $T = \{X_1, X_2, \dots, X_n\}$  (other time series with finite and discrete values of elements can be treated in the same way), Song et al. [13] adopted the actual entropy,

$$S = - \sum_{T' \subset T} P(T') \log_2[P(T')], \quad (1)$$

to measure the information capacity of such sequence, where  $P(T')$  represents the probability of finding a subsequence  $T'$  in the trajectory  $T$ . Based on the Fano inequality [14,15], they obtained the upper bound of the predictability  $\Pi_{\max}$  by solving the following EPC equation

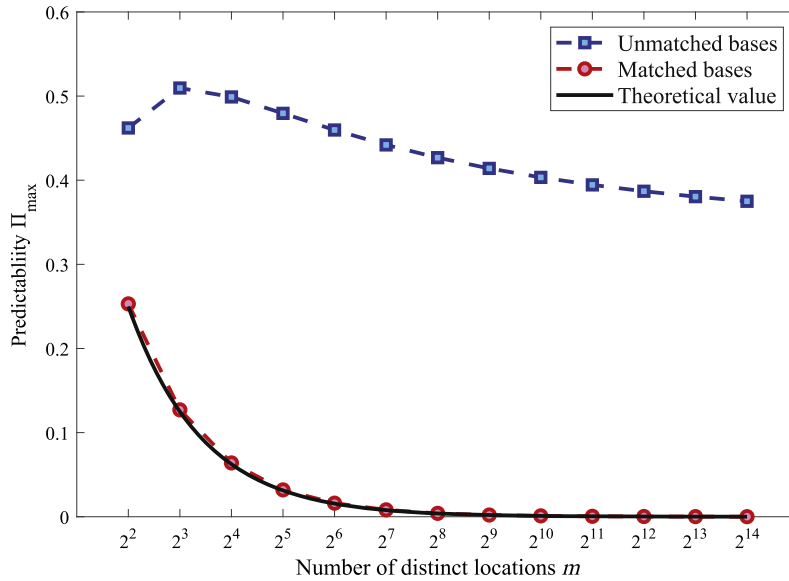
$$H = - [\Pi_{\max} \log_2 \Pi_{\max} + (1 - \Pi_{\max}) \log_2 (1 - \Pi_{\max})] + (1 - \Pi_{\max}) \log_2 (m - 1), \quad (2)$$

where  $m$  denotes the number of distinct locations appeared in  $T$ , and  $H$  is the entropy rate of  $T$ , mathematically defined as

$$H = \lim_{n \rightarrow \infty} S(X_1, X_2, \dots, X_n). \quad (3)$$

The direct computation of actual entropy is highly time-consuming and thus usually infeasible for real time series, therefore an estimator for actual entropy based on the LZ data compression method [48,49] is applied. For a time series with length  $n$ , the entropy is estimated by

$$S^{\text{est}} = \left( \frac{1}{n} \sum_i \Lambda_i \right)^{-1} \ln n, \quad (4)$$



**Fig. 1.** The predictability with unmatched bases (Eq. (4), blue squares) and matched bases (Eq. (5), red circles), compared with the theoretical value (black line) for exploration sequences with varying  $m$ . The value of  $\Lambda_i$  is set to be  $n - i + 2$  if every sub-sequence starting from  $X_i$  appears as a sub-sequence of  $\{X_1, X_2, \dots, X_{i-1}\}$ .

where  $\Lambda_i$  denotes the minimum length  $k$  such that the sub-sequence starting from position  $i$  with length  $k$  does not appear as a continuous sub-sequence of  $\{X_1, X_2, \dots, X_{i-1}\}$ .

The above description, extracted from Ref. [13], is ambiguous in two aspects. Firstly, the logarithm base in Eq. (4) is not clarified and thus was usually taken as the Euler's constant, namely  $e \approx 2.7183$  [11]. Secondly, when every sub-sequence starting from  $X_i$  appears as a sub-sequence of  $\{X_1, X_2, \dots, X_{i-1}\}$ , how to determine  $\Lambda_i$  is a puzzle.

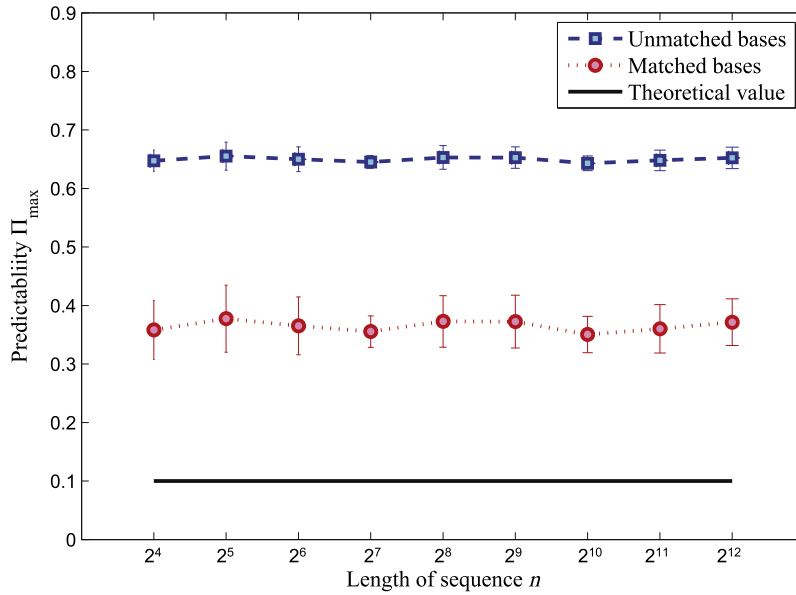
To quantify the accuracies of estimated predictability of different implementations, we consider four theoretical generators of time series with controllable predictability. Supposing there are  $m$  distinct locations in the constructed time series  $T$  with length  $n$ , the four types of sequences are as follows. (i) Exploration sequence. We set  $m = n$  and generate a random permutation with the  $m$  elements, so that every step in  $T$  can be considered as an exploration based on the previous historical trajectory. Since we do NOT know the information that the next location is a new location, the theoretical predictability should be  $1/m$ . (ii) Random sequence. Every elements in  $T$  are independently and randomly generated and thus the theoretical value should be  $1/m$  when  $n$  approaches infinity. (iii) Deterministic sequence. Without loss of generality, the constructed deterministic time series is  $T = \{X_1, X_2, \dots, X_m, X_1, X_2, \dots\}$ , whose predictability should converge to 1 when  $n$  approaches infinity. (iv) Markovian sequence. At each step, with probability  $p$ , the next location is determined by the same path as the deterministic sequence, say  $X_1 \rightarrow X_2 \rightarrow \dots \rightarrow X_m \rightarrow X_1 \rightarrow \dots$ , and with probability  $1 - p$ , the next location is randomly selected from the  $m$  candidates. Accordingly, the theoretical predictability should be  $p + (1 - p)/m$  when  $n$  approaches infinity.

We first consider the deviation caused by the unmatched logarithm bases in Eqs. (1) and (4), as described in Ref. [13]. According to the previous literature [48–50], the two bases should be the same. In particular, Grassberger [49] suggested to take the logarithm to base 2 in order to obtain  $S^{est}$  in bits. Therefore, we replace Eq. (4) by

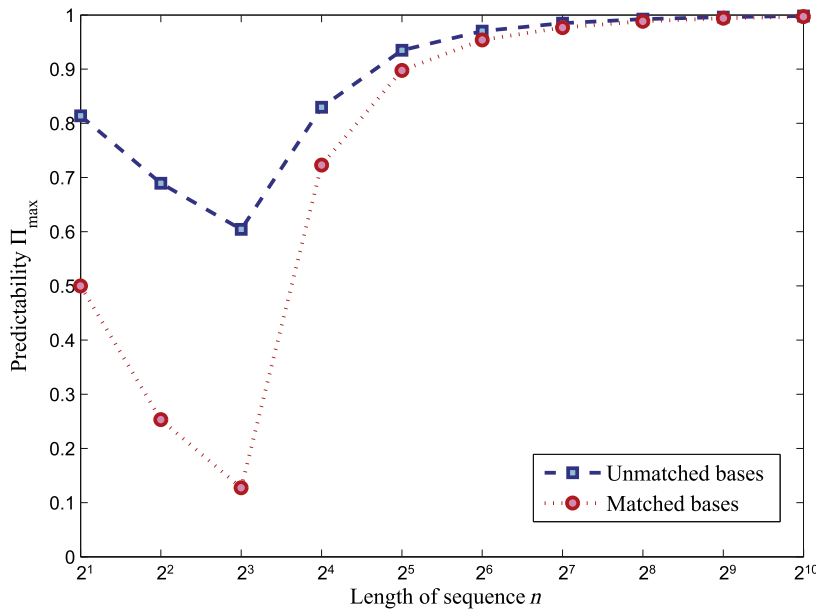
$$S^{est} = \left( \frac{1}{n} \sum_i \Lambda_i \right)^{-1} \log_2 n \quad (5)$$

to obtain the matched case (to replace  $\log_2$  in Eq. (4) by  $\ln$  will generate the same result). We first compare the two different cases based on the exploration sequences. As shown in Fig. 1, for the unmatched case, the predictability  $\Pi_{max}$  is larger than 0.35 even when  $m$  is  $2^{14}$ , while the theoretical value  $1/m$  should be already very close to zero. At the same time,  $\Pi_{max}$  obtained in the matched case (Eq. (5)) is in accordance with the theoretical value.

We further compare the matched and unmatched cases for random and deterministic sequences. Fig. 2 shows typical results with varying  $n$ , where one can observe that both matched and unmatched estimators are deviated from the theoretical value, while the matched case performs relatively better. In the other extreme relative to the random sequences, for deterministic sequences, when  $n$  becomes much larger than  $m$ , both estimators converges to the theoretical value 1 quickly, as shown in Fig. 3. Fig. 4 reports the results for Markovian sequences, whose predictability can be precisely controlled by adjusting the parameter  $p$ , with  $p = 0$  and  $p = 1$  corresponding to the two extremes, namely random sequences and deterministic sequences, respectively. Comparing the two curves with the same setting of  $\Lambda_i$  as Figs. 1–3 (marked

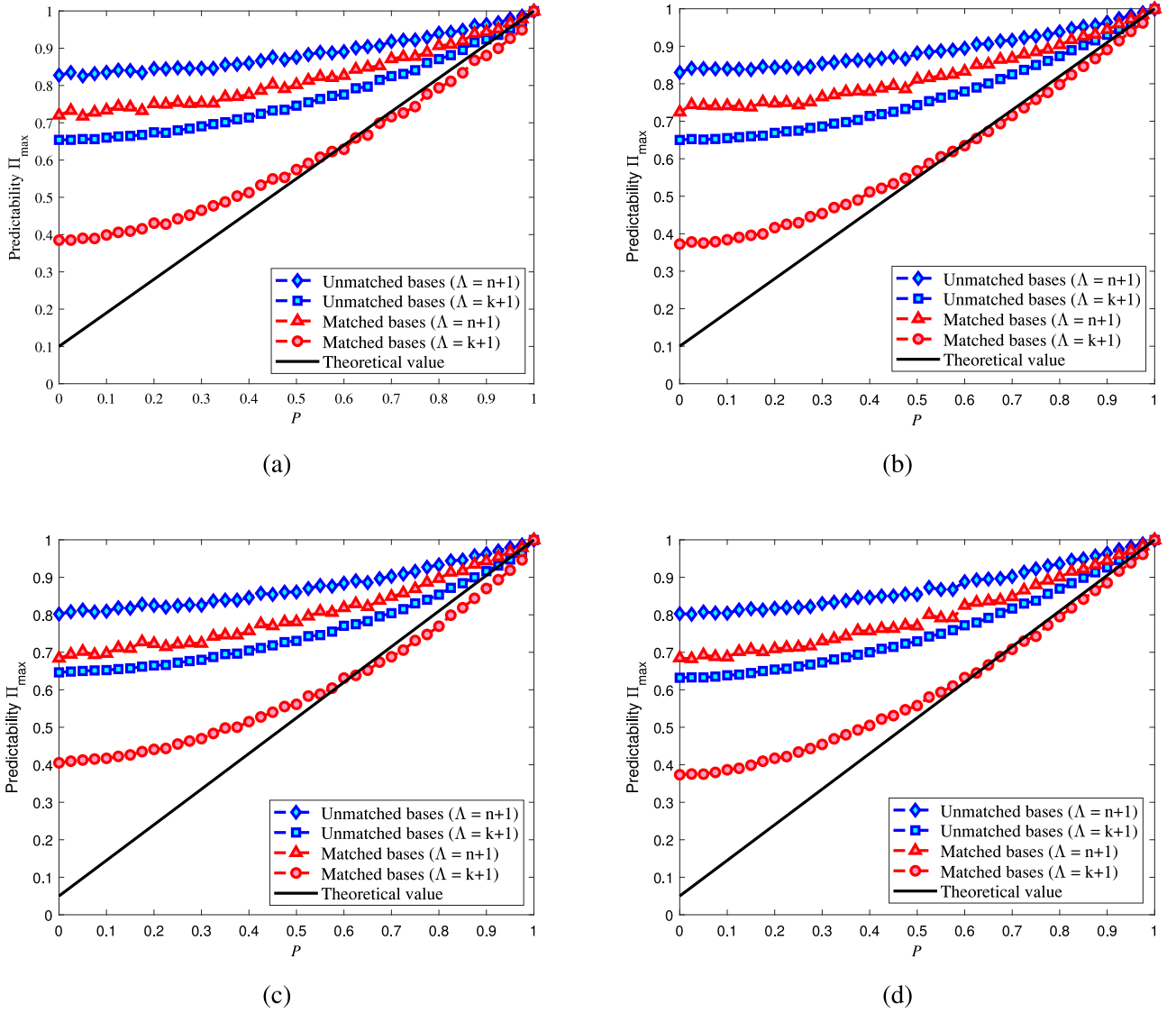


**Fig. 2.** The predictability with unmatched bases (Eq. (4), blue squares) and matched bases (Eq. (5), red circles), compared with the theoretical value (black line) for random sequences with varying  $n$ . The number of distinct locations  $m$  is fixed to be 10 (for other values of  $m$  ( $< n$ ), the results are the same). The simulation results are obtained by 100 independent implementations, with error bars denote the standard deviations. The value of  $\Lambda_i$  is set to be  $n - i + 2$  if every sub-sequence starting from  $X_i$  appears as a sub-sequence of  $\{X_1, X_2, \dots, X_{i-1}\}$ .

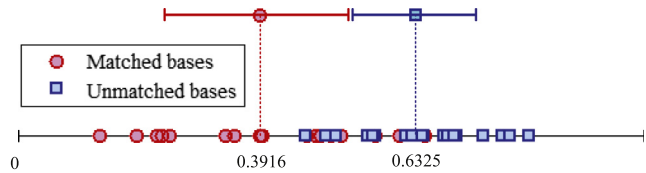


**Fig. 3.** The predictability with unmatched bases (Eq. (4), blue squares) and matched bases (Eq. (5), red circles) for deterministic sequences with varying  $n$ . The number of distinct locations  $m$  is fixed to be 10. The value of  $\Lambda_i$  is set to be  $n - i + 2$  if every sub-sequence starting from  $X_i$  appears as a sub-sequence of  $\{X_1, X_2, \dots, X_{i-1}\}$ .

as  $\Lambda = k + 1$  in Fig. 4), one can observe three phenomena: (i) the estimated predictability of unmatched case is always higher than that of matched case, (ii) the predictability of matched case is overall closer to the theoretical value than that of unmatched case, (iii) the deviation of LZ estimator for highly random sequences (i.e., small  $p$ ) is remarkably higher than that for more predictable sequences (i.e., large  $p$ ). For example, the deviation from theoretical value for matched bases (red circle) is 0.018 with  $p = 0.5$ , and 0.239 when  $p = 0$ .



**Fig. 4.** The predictability with unmatched bases (Eq. (4), in blue) and matched bases (Eq. (5), in red) for Markovian sequences with different  $p$ . The number of distinct locations  $m$  and the length of sequence  $n$  are set to be different values: (a)  $m = 10$ ,  $n = 100$ ; (b)  $m = 10$ ,  $n = 200$ ; (c)  $m = 20$ ,  $n = 100$ ; (d)  $m = 20$ ,  $n = 200$ . Other parameter settings ensuring  $1 \ll m \ll n$  will produce similar results. We also test different understandings of  $\Lambda_i$  when every sub-sequence starting from  $X_i$  appears as a sub-sequence of  $\{X_1, X_2, \dots, X_{i-1}\}$ , which are respectively marked as  $\Lambda = n + 1$  (Eq. (6)) and  $\Lambda = k + 1$  (Eq. (7)) in the plot. The simulation results are obtained by averaging over 100 independent implementations.



**Fig. 5.** Predictability distributions for the 22 valid samples, respectively obtained by applying unmatched bases (Eq. (4), blue squares) and matched bases (Eq. (5), red circles).  $\Lambda_i$  is determined by Eq. (7). The blue and red horizontal lines denote the corresponding error bars (i.e., standard deviations).

Another aspect we would like to clarify is how to determine the value of  $\Lambda_i$  when every sub-sequence starting from  $X_i$  appears as a sub-sequence of  $\{X_1, X_2, \dots, X_{i-1}\}$ . A straightforward treatment is to set  $\Lambda_i$  as

$$\Lambda_i = n + 1. \quad (6)$$

Let us look closely into the definition of  $\Lambda_i$ .  $\Lambda_i$  is the minimum length  $k$  such that the sub-sequence starting from position  $i$  with length  $k$  does not appear as a continuous sub-sequence of  $\{X_1, X_2, \dots, X_{i-1}\}$ , which can also be explained as one plus the length  $k_{\max}^{(i)}$  of the longest sub-sequence starting from position  $i$  that appears as a continuous sub-sequence of  $\{X_1, X_2, \dots, X_{i-1}\}$ , say

$$\Lambda_i = k_{\max}^{(i)} + 1. \quad (7)$$

Notice that, Eq. (7) is a unified explanation that can also be applied in the case when every sub-sequence starting from  $X_i$  appears as a sub-sequence of  $\{X_1, X_2, \dots, X_{i-1}\}$ , where  $k_{\max}^{(i)} = n - i + 1$  and thus  $\Lambda_i = n - i + 2$ . As shown in Fig. 4, estimator based on Eq. (7) performs much better than that based on Eq. (6). Some other possible alternatives of the understanding of  $\Lambda_i$  when every sub-sequence starting from  $X_i$  appears as a sub-sequence of  $\{X_1, X_2, \dots, X_{i-1}\}$ , such as  $\Lambda_i = 0$  and  $\Lambda_i = n$  performs no better or even worse than Eq. (6). So we can conclude that Eq. (7) is an proper and unified understanding of  $\Lambda_i$ , and indeed it has been applied in Figs. 1–3.

### 3. Empirical study

This section shows the difference between predictabilities estimated with unmatched and matched logarithm bases by a real data set recording interaction traces among 66 participants in the Politehnica University of Bucharest during March to May 2012 [51,52]. In the experiments, each participant carries an Android smartphone with tracing function that can identify other participants if they are close enough (by Bluetooth or AllJoyn). Therefore, we obtain a sequence of interacting persons for each participant. The original sequences describe next-timestep interactions and since the updates are frequency, they are many continuous sub-sequences consisting of the same interacting person. Therefore, we compress such a sub-sequence into only one element. For example, if a participant A's original interaction sequence is BBBBCCCDCCCCBB, it will be transformed into BCDCB.

After the above pretreatment, we remove all participants with sequence lengths no more than 5 and a few participants whose entropy rates do NOT converge according to Eq. (3) (also because of too small  $n$  compared with  $m$ ). Fig. 5 reports the estimated predictabilities of the remain 22 participants. In accordance with the theoretical analysis, for each participant, the estimated predictability by unmatched logarithm bases is always remarkably larger than that by matched bases, and the averaged values of  $\Pi_{\max}$  for the two cases are 0.63 and 0.39, respectively.

### 4. Conclusion

This paper briefly reviewed the framework proposed in Ref. [13] for quantifying the predictability of human mobility. The ambiguous description in Ref. [13] may lead to different understandings in some calculation details. We introduce some possible understandings, of which all the incorrect ones will result in overestimated predictability. When applying the considered method on human mobility or extending it to other discrete time series, we provide two clear suggestions. Firstly, the logarithm bases in the entropy estimator and the EPC equation should be the same. Secondly,  $\Lambda_i$  should be explained in a clear and unified way as one plus the length of the longest sub-sequence starting from position  $i$  that appears as a continuous sub-sequence of  $\{X_1, X_2, \dots, X_{i-1}\}$ .

Theoretical analysis on time series with controlled predictability showed that the estimator proposed in Ref. [13] failed when the time series is highly random. Therefore, at the end of this paper, we raise an open challenge for future study, that is, how to accurately estimate predictability for such less-predictable time series.

### Acknowledgments

The reported deviation from the true value was firstly found by Xiaoyong Yan and Zimo Yang. The authors would like to thank Zehui Qu for helpful discussion. This work was partially supported by the National Natural Science Foundation of China (Grants 61433014, 61673085, and 61603074).

### References

- [1] J. Wang, Y. Mao, J. Li, Z. Xiong, W.X. Wang, Predictability of road traffic and congestion in urban areas, *PLoS One* 10 (2015) e0121825.
- [2] W. Ren, Y. Li, S. Chen, D. Jin, L. Su, Potential predictability of vehicles visiting duration in different areas for large scale urban environment, in: *Proceedings of the Wireless Communications and Networking Conference*, IEEE Press, Shanghai, 2013, pp. 1674–1678.
- [3] Y. Li, D. Jin, P. Hui, Z. Wang, S. Chen, Limits of predictability for large-scale urban vehicular mobility, *IEEE Trans. Intell. Trans. Syst.* 15 (2014) 2671.
- [4] K. Zhao, D. Khryashchev, J. Freire, C. Silva, H. Vo, Predicting taxi demand at high spatial resolution: Approaching the limit of predictability, in: *Proceedings of the 2016 IEEE International Conference on Big Data*, IEEE Press, Washington, 2016, pp. 833–842.
- [5] X. Lü, L. Bengtsson, P. Holme, Predictability of population displacement after the 2010 Haiti earthquake, *Proc. Natl. Acad. Sci. USA* 109 (2012) 11576.
- [6] L. Bengtsson, X. Lü, A. Thorson, R. Garfield, J. Von Schreeb, Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: a post-earthquake geospatial study in Haiti, *PLoS Med.* 8 (2011) e1001083.
- [7] D.Y. Kenett, J. Portugali, Population movement under extreme events, *Proc. Natl. Acad. Sci. USA* 109 (2012) 11472.
- [8] S.T. Stoddard, A.C. Morrison, G.M. Vazquez-Prokopec, V.P. Soldan, T.J. Kochel, U. Kitron, J.P. Elder, T.W. Scott, The role of human movement in the transmission of vector-borne pathogens, *PLoS Neglect. Trop. Dis.* 3 (2009) e481.
- [9] V. Belik, T. Geisel, D. Brockmann, Natural human mobility patterns and spatial spread of infectious diseases, *Phys. Rev. X* 1 (2011) 011001.
- [10] D.H. Barmak, C.O. Dorso, M. Otero, H.G. Solari, Dengue epidemics and human mobility, *Phys. Rev. E* 84 (2011) 011901.



- [11] H. Barbosa, M. Barthelemy, G. Ghoshal, C.R. James, M. Lenormand, T. Louail, R. Menezes, J.J. Ramasco, F. Simini, M. Tomasini, Human mobility: Models and applications, *Phys. Rep.* 734 (2018) 1–74.
- [12] Y. Cao, J. Gao, D. Lian, Z. Rong, J. Shi, Q. Wang, Y. Wu, H. Yao, T. Zhou, Orderliness predicts academic performance: behavioural analysis on campus lifestyle, *J. R. Soc. Interface* 15 (2018) 20180210.
- [13] C. Song, Z. Qu, N. Blumm, A.-L. Barabási, Limits of predictability in human mobility, *Science* 327 (2010) 1018.
- [14] R.M. Fano, W. Wintringham, Transmission of information, *Phys. Today* 14 (1961) 56.
- [15] A. Brabazon, M. O'Neill, *Natural Computing in Computational Finance*, Springer Science & Business Media, 2008.
- [16] D. Lian, X. Xie, F. Zhang, N.J. Yuan, T. Zhou, Y. Rui, Mining location-based social networks: A predictive perspective, *IEEE Data Eng. Bull.* 38 (2015) 35.
- [17] X. Lü, E. Wetter, N. Bharti, A.J. Tatem, L. Bengtsson, Approaching the limit of predictability in human mobility, *Sci. Rep.* 3 (2013) 2923.
- [18] N. Mukai, N. Yoden, Taxi demand forecasting based on taxi probe data by neural network, in: *Intelligent Interactive Multimedia: Systems and Services*, Springer, Berlin, Heidelberg, 2012, pp. 589–597.
- [19] N.B. Ponieman, A. Salles, C. Sarraute, Human mobility and predictability enriched by social phenomena information, in: *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, ACM Press, Niagara Falls, 2013, pp. 1331–1336.
- [20] R. Jurdak, K. Zhao, J. Liu, M. Aboujaoude, M. Cameron, D. Newth, Understanding human mobility from twitter, *Plos One* 10 (2015) e0131469.
- [21] J.J.C. Ying, W.C. Lee, T.C. Weng, V.S. Tseng, Semantic trajectory mining for location prediction, in: *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, ACM Press, Chicago, 2011, pp. 34–43.
- [22] Z. Yang, D. Lian, N.J. Yuan, X. Xie, Y. Rui, T. Zhou, Indigenization of urban mobility, *Physica A* 469 (2017) 232.
- [23] C. Yu, Y. Liu, D. Yao, L.T. Yang, H. Jin, H. Chen, Q. Ding, Modeling user activity patterns for next-place prediction, *IEEE Syst. J.* 11 (2017) 1060.
- [24] M. Lin, W.J. Hsu, Z.Q. Lee, Predictability of individuals mobility with high-resolution positioning data, in: *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, ACM Press, Pittsburgh, 2012, pp. 381–390.
- [25] A. Cuttone, S. Lehmann, M.C. González, Understanding predictability and exploration in human mobility, *EPJ Data Sci.* 7 (2018) 2.
- [26] B.S. Jensen, J.E. Larsen, K. Jensen, J. Larsen, L.K. Hansen, Estimating human predictability from mobile sensor data, in: *Proceedings of the IEEE International Workshop on Machine Learning for Signal Processing*, IEEE Press, Kittilä, 2010, pp. 196–201.
- [27] X.Y. Yan, C. Zhao, Y. Fan, Z. Di, W.X. Wang, Universal predictability of mobility patterns in cities, *J. R. Soc. Interface* 11 (2014) 20140834.
- [28] X.Y. Yan, W.X. Wang, Z.Y. Gao, Y.-C. Lai, Universal model of individual and population mobility on diverse spatial scales, *Nature Commun.* 8 (2017) 1639.
- [29] L. Pappalardo, F. Simini, S. Rinzivillo, D. Pedreschi, F. Giannotti, A.-L. Barabási, Returners and explorers dichotomy in human mobility, *Nature Commun.* 6 (2015) 8166.
- [30] C. Iovan, A.M. Olteanu-Raimond, T. Couronn, Z. Smoreda, Moving and calling: Mobile phone data quality measurements and spatiotemporal uncertainty in human mobility studies, in: *Geographic Information Science at the Heart of Europe*, Springer, Switzerland, 2013, pp. 247–265.
- [31] X.Y. Yan, X.-P. Han, B.-H. Wang, T. Zhou, Diversity of individual mobility patterns and emergence of aggregated scaling laws, *Sci. Rep.* 3 (2013) 2678.
- [32] E.L. Ikanovic, A. Mollgaard, An alternative approach to the limits of predictability in human mobility, *EPJ Data Sci.* 6 (2017) 12.
- [33] T. Takaguchi, M. Nakamura, N. Sato, K. Yano, N. Masuda, Predictability of conversation partners, *Phys. Rev. X* 1 (2011) 011008.
- [34] W. Chen, Q. Gao, H. Xiong, Temporal predictability of online behavior in foursquare, *Entropy* 18 (2016) 296.
- [35] P. Baumann, S. Santini, On the use of instantaneous entropy to measure the momentary predictability of human mobility, in: *Proceedings of the IEEE 14th Workshop on Signal Processing Advances in Wireless Communications*, IEEE Press, Darmstadt, 2013, pp. 535–539.
- [36] J. McInerney, S. Stein, A. Rogers, N.R. Jennings, Exploring periods of low predictability in daily life mobility, *Pervasive Mobile Comput.* 9 (2013) 808.
- [37] D. Austin, R. M. Cross, T. Hayes, J. Kaye, Regularity and predictability of human mobility in personal space, *Plos One* 9 (2014) e90256.
- [38] G. Smith, R. Wieser, J. Goulding, D. Barrack, A refined limit on the predictability of human mobility, in: *Proceedings of the IEEE International Conference on Pervasive Computing and Communications*, IEEE Press, Budapest, 2014, pp. 88–94.
- [39] W. Yao, C. Essex, P. Yu, M. Davison, Measure of predictability, *Phys. Rev. E* 69 (2004) 066121.
- [40] L. Zhang, Y. Liu, Y. Wu, J. Xiao, Analysis of the origin of predictability in human communications, *Physica A* 393 (2014) 513.
- [41] Z.-D. Zhao, Z. Yang, Z.-K. Zhang, T. Zhou, Z.-G. Huang, Y.-C. Lai, Emergence of scaling in human-interest dynamics, *Sci. Rep.* 3 (2013) 3472.
- [42] T. Xu, X. Xu, Y. Hu, X. Li, An entropy-based approach for evaluating travel time predictability based on vehicle trajectory data, *Entropy* 19 (2017) 165.
- [43] Y.-Z. Chen, Z.-G. Huang, S. Xu, Y.-C. Lai, Spatiotemporal patterns and predictability of cyberattacks, *Plos One* 10 (2015) e0124472.
- [44] P. Fiedor, Frequency effects on predictability of stock returns, in: *Proceedings of the IEEE Conference on Computational Intelligence for Financial Engineering & Economics*, IEEE Press, London, 2014, pp. 247–254.
- [45] C. Krumme, A. Llorente, M. Cebrian, E. Moro, The predictability of consumer visitation patterns, *Sci. Rep.* 3 (2013) 1645.
- [46] R. Sinatra, M. Szell, Entropy and the predictability of online life, *Entropy* 16 (2014) 543.
- [47] D. Dahlem, D. Maniloff, C. Ratti, Predictability bounds of electronic health records, *Sci. Rep.* 5 (2015) 11865.
- [48] I. Kontoyiannis, P.H. Algoet, Y.M. Suhov, A.J. Wyner, Nonparametric entropy estimation for stationary processes and random fields with applications to english text, *IEEE Trans. Inform. Theory* 44 (1998) 1319.
- [49] P. Grassberger, Estimating the information content of symbol sequences and efficient codes, *IEEE Trans. Inform. Theory* 35 (1989) 669.
- [50] J. Ziv, A. Lempel, Compression of individual sequences via variable-rate coding, *IEEE Trans. Inform. Theory* 24 (1978) 530.
- [51] R.-I. Ciobanu, R.-C. Marin, C. Dobre, Interaction predictability of opportunistic networks in academic environments, *Trans. Emerg. Telecommun. Technol.* 25 (2014) 852.
- [52] R.-C. Marin, C. Dobre, F. Xhafa, Exploring predictability in mobile interaction, in: *Proceedings of the 3rd International Conference on Emerging Intelligent Data and Web Technologies*, IEEE Press, Bucharest, 2012, pp. 133–139.