This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2019.2930363, IEEE Transactions on Vehicular Technology

Nonlinear Dynamic Chaos Theory Framework for Passenger Demand Forecasting in Smart City

Benedetta Picano, *Student Member, IEEE* Romano Fantacci, *Fellow, IEEE,* and Zhu Han, *Fellow, IEEE.*

***Abstract*—Recently chaos theory has emerged as a powerful tool to address forecasting problems of nonlinear time series, since it is able to meet the dynamical and geometrical structures of very complex systems, reaching higher accuracy on the prediction values than the classical approaches. This paper aims at applying the chaos theory principles to different problems, in order to pursue high levels of accuracy on the predicted results. After the verification of the chaotic behavior of the datasets taken into analysis through the largest Lyapunov exponent research, the detection of the suitable embedding dimension and time delay has been carried out, in order to reconstruct the phase space of the underlying dynamical systems. Three different predictive methods have been proposed for different datasets. Finally, the performance comparison with the moving average model, a deep neural network based strategy, and a chaos theory based algorithm recently proposed in literature has been provided.**

***Index Terms*—Chaos Theory, Forecasting, Nonlinear Time Series Analysis.**

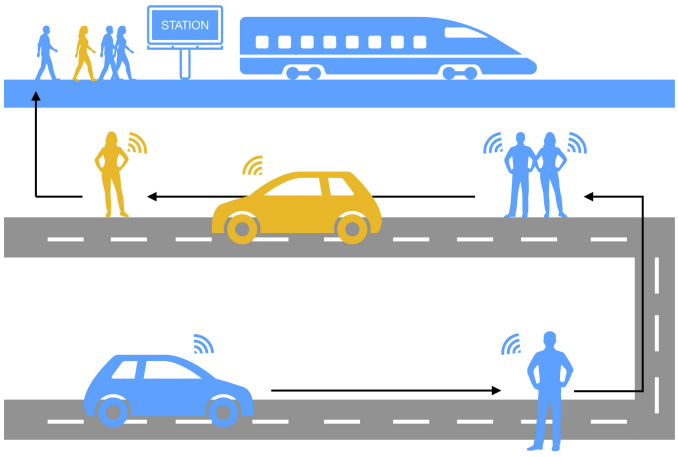

Figure 1. Ride-sharing service in smart city

I. Introduction

Recently, the ever increasing diffusion of vehicles has resulted in cities with remarkable levels of urban traffic, fatalities, injuries, and congestion [1]. Within this context, the smart city concept has emerged, opening the doors towards several promising applications. In particular, an unprecedented chance to create a wide variety of new services has been provided [1]–[3], involving the efficient data-collection and data-processing, and providing support to smart infrastructure, smart healthcare, smart governance, smart mobility, smart technology, etc. [3]. Furthermore, many research efforts have been made to design modern solutions to control mobility and traffic, especially in order to lower road congestion and improve the transportation efficiency [4], [5]. Therefore, the Intelligent Transportation Systems (ITS) has gained much attention, favoring the development of proper strategic solutions to reduce the drawbacks due to the growing spread of vehicles. In particular, in order to limit the traffic congestion, one of the main objective is the reduction of the utilization of cars, by promoting the usage of public transportation during the rush hours. Transportation network companies (TNC) such as Uber or Lyft offer peer-to-peer ride-sharing services to move people from and to homes and offices, or to the public transportation stations. These types of services constitute a promising way to ensure to people a fast and comfortable solution to reduce the utilization of their own cars. Within this context, by considering the TNC perspective, the accurate prediction of traffic demands plays a crucial role to properly allocate resources and, consequently, to avoid resource waste or delays on services provision. For all these reasons, traffic forecasting and mobility forecasting have gained significant momentum in the area of ITS. The application of forecasting procedures to the ITS [6]–[8] area focuses on the prediction of traffic conditions, a given number of hours ahead in the future.

In general terms, we refer to short-, medium-, and long-term forecasting. The first one involves predictions with horizons from few minutes up to few days ahead, and the second one deals with time horizons from few days to few months ahead. Instead, the long-term forecasting is referred to predictions from quarters to years ahead [9]. Although medium and long-term forecasting generally involve the prediction about risk management and profitability planning, the short-term forecasting is frequently applied to traffic demands and mobility prediction because of its satisfactory accuracy [9], [10]. The short-term forecasting has been extensively studied in literature, and many different methods have been proposed [10], [11].

Roughly speaking, the whole family of the predictive methods can be divided into two different main branches: the approaches based on the classic time series analysis (TSA), and the techniques based on the deep learning. The former branch includes methods such as the Auto-Regressive (AR), the Auto-Regressive and Moving Average (ARMA), or the Auto-Regressive Integrated Moving Average (ARIMA) [12]–[14], while the methods based on the latter approach are the strategies based on Artificial Neural Network (ANN), Support Vector Machine (SVM), Support Vector Regression (SVR) and

B. Picano and R. Fantacci are with the Department of Information Engineering, University of Florence, Florence, Italy e-mail: name.surname@unifi.it.

Z. Han is with the University of Houston, Houston, TX 77004 USA (e-mail:zhan2@uh.edu), and also with the Department of Computer Science and Engineering, Kyung Hee University, Seoul, South Korea, 446-701.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2019.2930363, IEEE Transactions on Vehicular Technology

so on [15]–[18]. TSA has been widely used for a very long time to solve a vast number of forecasting problems, by guaranteeing limited complexity and good performance on prediction accuracy. However, recently, the ever increasing complexity on the time series, due to the intrinsic complexity of the current problems, has lead to a performance reduction on accuracy applying the TSA. Therefore, despite TSA has provided good solution to many problems during past years, recently the high level complexity of the current problems scenarios have limited the applicability of the TSA due to its inability in performing valuable predictions on time series with complex behaviors. From the other side, the deep learning approaches reach very accurate results, but require a large amount of data to be trained, procedure that is extremely computationally expensive. Therefore, the overall prediction performance results compromised by these drawbacks that constitute a strong bottleneck, since many application scenarios need to rapidly predict new values and do not have a large samples set to work on.

In this situation, Chaos Theory (CT) has emerged as a powerful tool to perform nonlinear TSA [10], [19]. More in depth, CT studies the behavior of the nonlinear dynamic systems that exhibit strong sensitivity to the initial system conditions, in which the irregular behavior hides determinism. The apparent irregular behavior in the observed time series space is a typical characteristic of the nonlinear time series. Furthermore, the degree of complexity of a time series is generally strictly related to the nonlinearity of the underlying system. CT provides, under certain conditions on the time series taken into account (i.e., nonlinearity and chaotic behavior), useful strategies to address and manage such complexity, by resulting in high levels of accuracy prediction, avoiding the expensive training procedure, typical for example of the deep learning approaches. However, it is important to note that one of the major drawback of the CT based strategies is that it requires the verification of the chaotic trend of the analyzed time series, since it performs good results in presence of chaotic features.

The chaotic nonlinear time series class, in the corresponding *phase space*, exhibit a strange attractor with a regular structure that allow us to observe its geometrical properties and dynamics in order to predict the future behavior of the original time series.

This paper proposes different applications of the CT principles to solve the forecasting problem. The prediction has been conducted on real data in Chengdu from Didi, a Chinese TNC, and two sets from Google dataset search, concerning the Uber pickup requests in New York City and in Bangalore, respectively. More in depth, the main contributions of this work are:

- Validation of the chaotic behavior of the considered datasets, provided by the largest Lyapunov exponent analysis. The analysis consist of estimating the divergence rate of close trajectories associated to the scalar time series, during its evolution;
- Applications of the CT principles to design three different predictive algorithms for different datasets. In particular, the reconstruction of the phase space for each dataset has been pursuit, and proper forecasting algorithms proposed. Furthermore, the forecasting algorithms are based on local predictive mechanisms, and for the third dataset a hybrid approach is presented, combining both local and global approximations;
- Comparison of the proposed approaches with the well-known moving average (MA) model [20], that presented in [21], based on the CT as well, and the deep learning approach proposed in [22]. System performance has been provided in terms of mean squared error, mean absolute deviation and mean absolute percentage error.

The rest of paper is organized as follows. Section II proposes an in-depth review of the related literature. In Section III, we propose the problem formulation, and in Section IV the nonlinear time series analysis based on CT is addressed. Section V presents the forecasting algorithms proposed, and in Section VI the experimental results are shown. Finally, the conclusions are drawn in Section VII.

II. Related Literature

The classical TSA methodologies are applied in paper [6],[12],[23],[7],[8]. Paper[6] proposes a short-term spatio-temporal forecasting approach to estimate the future taxi-passengers demand. The method aims at predicting the number of service requests that emerge at taxi ranks, by exploiting the real-time information exchanged among taxis. The paper combines both the predictive ARIMA and the time-varying Poisson models to realize the passengers demand prediction [12], [23]. Authors in [7] model the univariate vehicular traffic flow with the seasonal ARIMA, providing theoretical evidences about the suitability of this model in solving the short-term traffic conditions forecasting problems. Similarly, in [8], the study of traffic forecasting problem on large IEEE802.11 infrastructures is addressed. More in depth, authors in [8] evaluate the performance of many modified versions of the moving average and ARIMA algorithms, at different time scale, to forecast the access points load in wireless networks. Paper [8] highlights the importance of fine-grained prediction horizons and recent past data, to obtain high levels of accuracy on the forecast values. Furthermore, many works based on machine learning approaches have been proposed. Examples are represented by paper [10], [24], [25]. Within the short-term traffic prediction, [10] combines the CT principles with the SVM, to improve the accuracy on the forecast values. In particular, authors in [10] adopt as measure of similarity the dynamic time warping to mitigate the negative effects of possible bursty points outside the neighborhood area of the processed point.

In [24] is provided a short-term passenger demand forecasting of light rail services. In particular, authors in [24] propose a novel neural networks model to fit non-stationary time series, aiming at minimizing the prediction error. Furthermore, the model formulated in [24] is based on the multi-layer perceptron one and the back-propagation algorithm is applied during the training process. Work [25] compares the forecasting performance applying both ANNs and Box Jenkins methods to airline passenger demand, calculated over the past five year daily data. An ANN strategy is also adopted in paper [26],

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2019.2930363, IEEE Transactions on Vehicular Technology

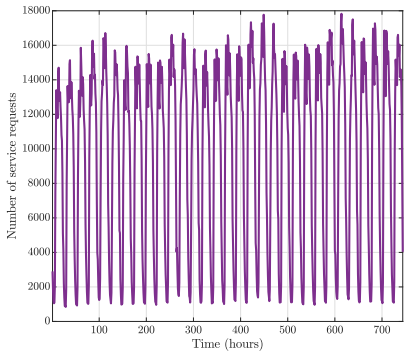

Figure 2. The trend behavior of Dataset 1.

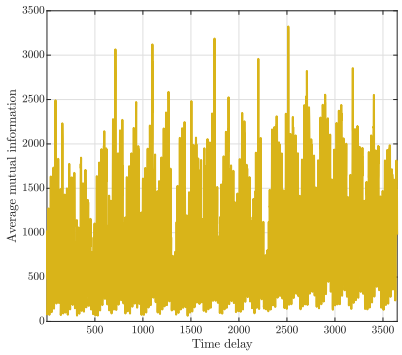

Figure 3. The trend behavior of Dataset 2.

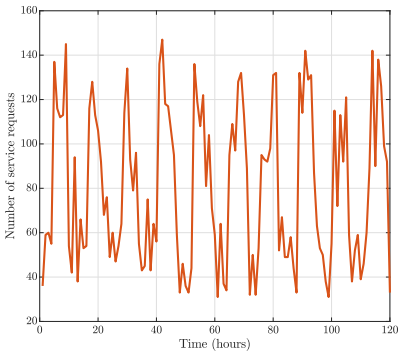

Figure 4. The trend behavior of Dataset 3.

in which a multilayer perceptron neural network is adopted to forecast the lightning occurrences. The recurrent neural networks are sequentially applied in paper [27], where the main objective is the real-time prediction of the taxi demand in the city of New York. Paper [22] proposes a forecasting framework based on the combination of the feed forward neural network and the long short term memory approach, to evaluate the electricity consumption. Differently, paper [28] predicts the urban traffic passengers flows by proposing a predictive structure based on convolutional neural networks and a graph representation of the traffic data, in order to extract the spatio-temporal information of the analyzed samples series. A spatio-temporal analysis has been provided also in [29], in which the demand for shared bicycles in three typical subway stations in the city of Beijing is predicted. The forecasting framework proposed in [29] is based on an improved version of the Xgboost method and the idea of sliding window. Within the CT approaches, in papers [21], [30], CT is applied to forecast the electricity price. Specifically, in [30] the electricity price is modeled as a multivariate time series, since it depends on many different factors. Then, authors in [30] provide the phase space reconstruction of the corresponding chaotic time series and find the forecasting function by fitting all points in the phase space, by applying the Elman model, which is one of the most common recurrent neural network. Paper [21] addresses the same problem presented in [30] but, after phase space reconstruction, in order to improve the accuracy on the predicted results, authors use the add-weighted one-rank multi-steps prediction method [31]. Authors in [32] apply CT principles as well, to forecast the grids load data. Subsequently to the phase space reconstruction, paper [32] proposes the introduction of a weight on the largest Lyapunov exponent with the aim at mitigating the effect of significantly divergent trajectories, in order to reduce the forecasting error. This implies that during the prediction process, points close to the processed one with divergent trajectory will give a lower contribution on the prediction. Differently, paper [33] exploits CT to predict the urban daily water demand. In [33], the prediction of each point is made by considering the behavior, in the phase space, of only its nearest point. Another situation in which chaotic and nonlinear behaviors are widely present, is that of wind power generation. Within this context, the wind power exhibits fluctuations very difficult to predict.

Paper [34] deals with the system power forecasting, through the use of CT combined with the wavelet packet one. In [34] the wavelet packet theory is used to decompose the history of wind power data between high and low frequency components. Then, the frequency components are reconstructed with the single branch and the phase space is built for each single branch. Whether, during the process, the time series shows a non-chaotic behavior, prediction is performed by using back propagation neural network, otherwise through CT. The usage of both wavelet and CT has been also adopted in [35] to predict traffic in wireless sensor networks. This paper proposes the application of CT principles to both the high and low frequency parts of the original signal, and builds the predictive function considering the near points most influential than the far ones. Paper [36] addresses the forecasting of the load of power. In particular, it provided a short term prediction by combining CT with the fuzzy approach. In the method proposed by [36], closer the points in the phase space are to the value to be forecast, the greater is their impact on the predicted result. Finally, paper [37] aims at predicting a nonlinear time series for human actions and dynamic textures synthesis through a CT approach. Specifically, the phase space is reconstructed considering the corresponding multivariate time series and future predictions are made using a nonparametric data driven model, based on a kernel which is a decreasing function of the distance from the point that has to be predicted [38]. Then, the future multivariate time series values are built by extracting the univariate time series from the reconstructed phase space.

The main contributions of this paper are the analysis of the behavior of three different and real service request dataset, deriving from Didi, the biggest TNC in China, and from the Google data search engine. In fact, for each dataset, the largest Lyapunov exponent has been analyzed and its chaotic behavior verified. Furthermore, in order to pursuit a CT based approach, the reconstruction of the phase space associated to each scalar time series taken into account has been addressed. Finally, three different CT predictive algorithms, one for each dataset considered, have been proposed and comparison with the MA approach, those presented in paper [21] and paper [22] has

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2019.2930363, IEEE Transactions on Vehicular Technology

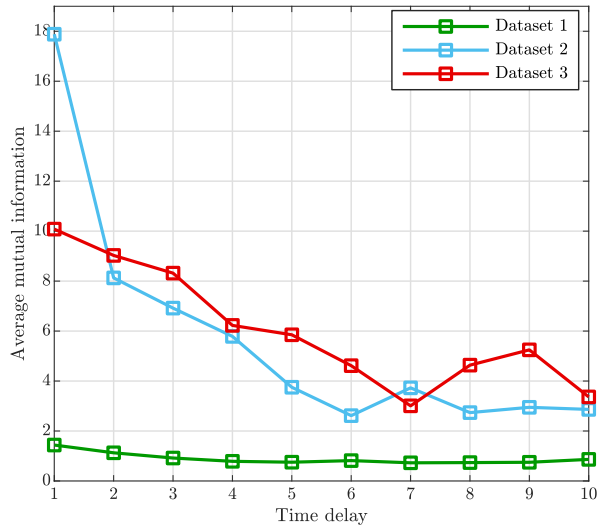

Figure 5. The optimal value of τ for Dataset 1, Dataset 2 and Dataset 3.

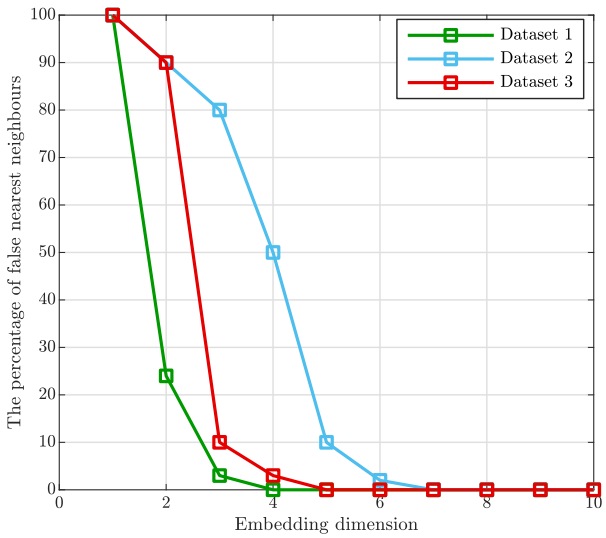

Figure 6. The optimal value of m for Dataset 1, Dataset 2 and Dataset 3.

been provided, in terms of mean squared error, mean absolute error and mean percentage error.

III. Problem Statement

Given a scalar time series $\mathcal{X} = \{x_i\}_{i=1}^{N}$, x_i represents the value observed at time i and N is the whole number of samples of the time series $\mathcal{X}$. More in depth, $\mathcal{X}$ comes out from a previous sampling procedure in the city of interest. As regards the Didi dataset, the city of Chengdu has been modeled as a rectangle $\mathcal{R}$ of dimension $P \times Q$, where $\rho_{p,q}$, $p \in [0, P)$ and $q \in [0, Q)$, represents the region with coordinates p and q. A service request demand r_u is represented as a quintuple in the form $(r_{u,id}, r_{u,pc}, r_{u,pt}, r_{u,dc}, r_{h,dt})$, in which $r_{u,id}$ is the request ID, $r_{u,pc}$ the pick-up coordinates, $r_{u,pt}$ is the pick-up time, $r_{h,dc}$ identifies the coordinates of the destination, and $r_{u,dt}$ the time of arrival at destination. Our analysis has been conducted by grouping all the requests based on the pick-up time. Therefore, time has been partitioned into equal slots starting from 0. Hence, the k-th time slot is identified by the interval $[k \times S, (k+1) \times S)$, where S is the time span of the interval. Consequently, the passenger demand at the interval k, i.e., x_k, is given by

$$x_k = |\{u \in [0, \mathcal{N}) : r_{u,pc} \in \mathcal{R} \wedge r_{h,pt} \in [k \times S, (k+1) \times S)\}|, \tag{1}$$

where $\mathcal{N}$ is the total number of received requests and $|\cdot|$ means the number of elements belonging to the set. The other two sets have been taken on Google dataset search, and each request of the first of them is identified by the triplets (π_d, π_t, π_a), where π_d is the pick-up date, π_t the pick-up time and π_a the pick-up address. Differently, each request of the last dataset is composed of six fields, i.e, $(\mu_r, \mu_p, \mu_d, \mu_s, \mu_u, \mu_o)$, where μ_r is the request identifier number, μ_p represents the pick-up point, μ_d the driver number, μ_s the status of the ride (completed or ongoing), and μ_u and μ_o are the pick-up time and the drop off time, respectively. Given the time series $\mathcal{X}$, forecasting problem involves the prediction about the future behavior, δ steps ahead in the future. This paper deals with the short-term forecasting, aiming at minimizing the forecasting error that, generally speaking, is a measure of the gap between the predicted and the real value of the time series analyzed. Despite there exist many different metrics to evaluate the forecasting error [11], this paper considers the minimization of the mean squared error (MSE) defined as

$$MSE = \frac{1}{M} \sum_{i=1}^{M} (\hat{x}_{i+\delta} - x_{i+\delta})^2, \tag{2}$$

where M represents the number of the samples in test data, and $\hat{x}_{i+\delta}$ and $x_{i+\delta}$ are the actual and the predicted values at time $i+\delta$. In addition, in order to provide an exhaustive analysis, we have also considered the mean percent error (MAPE) defined as

$$MAPE = \frac{1}{M} \sum_{i=1}^{M} \left| \frac{\hat{x}_{i+\delta} - x_{i+\delta}}{x_{i+\delta}} \right| \cdot 100, \tag{3}$$

and the mean absolute deviation (MAD) given by

$$MAD = \frac{1}{M} \sum_{i=1}^{M} |\hat{x}_{i+\delta} - x_{i+\delta}|. \tag{4}$$

It is important to note that metrics (2) and (4) highlight the variability of the forecasting error, while (3) expresses the error in terms of percentage on the actual data.

In the following section, in order to clarify the motivation behind the insight of the proposed forecasting algorithms, the technical background about CT principles is provided and the phase space reconstruction procedure explained.

IV. Chaos Theory Approach

The class of the chaotic nonlinear dynamical systems includes the nonlinear dynamical systems whose behavior is unpredictable on the long term, and exhibit strong sensitivity to

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2019.2930363, IEEE Transactions on Vehicular Technology

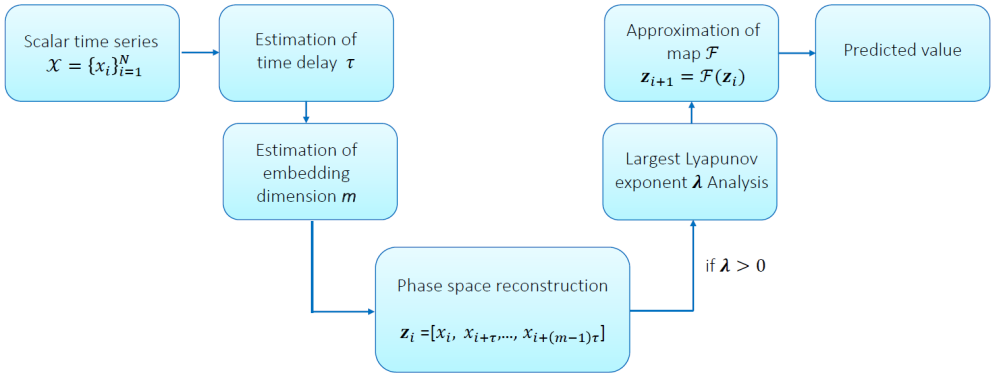

Figure 7. Scalar time series analysis and prediction procedure based on CT

$$F_{fnn}(T) = \frac{\sum_{i=1}^{N-m-1} \Theta\left(\frac{|\mathbf{z}_i^{(m+1)} - \mathbf{z}_j^{(m+1)}|}{|\mathbf{z}_i^{(m)} - \mathbf{z}_j^{(m)}|} - T\right)\Theta(\frac{\sigma}{T} - |\mathbf{z}_i^{(m)} - \mathbf{z}_j^{(m)}|)}{\sum_{i=1}^{N-m-1} \Theta(\frac{\sigma}{T} - |\mathbf{z}_i^{(m)} - \mathbf{z}_j^{(m)}|)}; \tag{5}$$

the initial conditions, that implies small changes in the initial state result in significant differences on the final states [39], [40]. Generally speaking, the nonlinear dynamical systems with chaotic behavior present apparent irregular trend which hides the deterministic features of these systems. A system state is specified by a m dimensional vector $\mathbf{z}$, while the system dynamics can be expressed by [19]

$$\mathbf{z}_{n+1} = \mathcal{F}(\mathbf{z}_n), \tag{6}$$

in which $\mathbf{F}$ is a m dimensional map. It is important to highlight the relation between equation (6) and the set $\mathcal{X}$. Indeed, the elements of $\mathcal{X}$ are a sequence of scalar measurements, therefore, the conversion of the observations into state vectors has to be performed. In order to execute such conversion, the *phase space* reconstruction procedure [37], [41] has been conducted. After the phase space reconstruction procedure, the important features of chaotic time series can be caught by analyzing the underlying dynamics and the geometrical structure of its corresponding *attractor*, i.e., the set of values to which the system tends to evolve.

For the sake of simplicity, hereafter we refer to the dataset derived from Didi as Dataset 1, Dataset 2 for the dataset of the Uber pickups in New York City, and finally Dataset 3 for the Uber service requests in Bangalore.

In the following, the paper provides a brief theoretical background about the phase space reconstruction in Section IV-A, then in Section IV-B the time delay is presented. In Section IV-C the embedding dimension is discussed. Finally, Section IV-D presents the largest Lyapunov exponent analysis.

A. Phase Space Reconstruction

Given the chaotic scalar time series $\mathcal{X}$, the first step towards the comprehension of its behavior is the reconstruction of the phase space (PSR) associated with it, in order to pursuit the analysis of its strange attractor, as illustrated in Figure 7. Due to the Takens' delay embedding theorem, which affirms the existence of a map between the real phase space and its reconstructed version [37], the PSR is provided by associating, to each element x_i in $\mathcal{X}$, a vector in the form

$$\mathbf{z}_i = [x_i, x_{i+\tau}, \ldots, x_{i+(m-1)\tau}], \tag{7}$$

where τ is the *time delay* and m is the *embedding dimension*. As detailed in Section IV-B and Section IV-C, a proper choice of the values of τ and m is crucial to pursuit a suitable PSR.

B. Time Delay Estimation

As it is evident from (7), m and τ strongly impact the transformation of the scalar time series to the phase space vectors represented by $\mathbf{z}_i$. According to this, time delay τ rules the temporal distance between two successive points in the phase space vector, i.e., $\mathbf{z}_i$ and $\mathbf{z}_{i+1}$, and its optimal value minimizes the redundancy between points $x_{i+\tau}$ and x_i, maximizing the knowledge about $x_{i+\tau}$ from x_i. In this paper we compute the exact value of delay τ with the approach based on the evaluation of the amount of mutual information between pairs of points $x_{i+\tau}$ and x_i, with τ that minimizes the mutual information between observations [42].

In practical terms, for each measured data, the histogram of the probability distribution of the data is created. Then, by varying time the delay τ, the mutual information x_i and $x_{i+\tau}$ results expressed by [43]

$$I(\tau) = \sum_{f,j} p_{f,j}(\tau) \log p_{f,j} - 2\sum_{f} \log p_f, \tag{8}$$

where p_f is the probability that x_i is in the f-th bin of the histogram, while $p_{f,j}$ is the probability that x_i and $x_{i+\tau}$ fall in the f-th and j-th bin of the histogram, respectively.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2019.2930363, IEEE Transactions on Vehicular Technology

As depicted in Figure 5, the values of τ for Dataset 1, Dataset 2 and Dataset 3 are $\tau = 5, 6$ and 7, respectively. Indeed, Figure 5, shows the plot of the mutual information I when τ changes. In order to select the proper value for time delay τ, i.e., the minimum time delay, the optimal value of τ is in correspondence of the first local minimum of the I function.

C. Embedding Dimension Estimation

As in the case of τ, to choose a proper value of m is very important to observe the determinism underlying the system associated to time series $\mathcal{X}$, with the least computational effort. In this paper, we adopt the false nearest neighbors method [44]. The idea behind this method is that, since the presence of chaos can induce an exponential divergence on the trajectories of two nearby points, there exist some points in the data set that are neighbors in the embedding space but for which their temporal evolution exponentially diverges. Hence, this method consists, for each point in $\mathcal{X}$, in finding its nearest neighbor in m dimension. Then, the ratio between their distance in $m+1$ dimensions and m dimensions is calculated. Finally, whether the resulted ratio is greater than a fixed threshold r, the neighbor is considered false.

Hence, given a fixed threshold T, the false nearest neighbors function can be defined as in (5) on the top of the previous page. in which σ is the standard deviation of the data, j is index of the nearest point, and Θ is the Heaviside step function given by

$$\Theta(\nu) = \begin{cases} 0 & \nu < 0; \\ 1 & \nu \geq 0. \end{cases} \tag{9}$$

Figure 6 shows the F_{fnn} function by varying the embedding dimension m, and the proper value of m is $m = 3, 6$ and 4, respectively. Indeed, for each set of data, the right value of m is the value for which is minimum the number of false nearest neighbors, hence, graphically, it is in correspondence of the last m value before the plot of F_{fnn} drops to zero.

D. Largest Lyapunov Exponent

Once the PSR has been pursued, in order to verify the chaotic behavior of $\mathcal{X}$, we analyze the largest Lyapunov exponent. There exist many approaches to check the presence of chaos in a time series [21], and the study of the largest Lyapunov exponent is one of the most used. The main idea behind such technique is the study of the distance of two close vectors in the phase space over the time transition [45]. Hence, considering two trajectories $\mathbf{y}$ and $\mathbf{x}$, i.e., solutions of (6), close in the state space, the evolution of their mutual distance is given by

$$\mathbf{y}_{n+1} - \mathbf{x}_{n+1} = \mathbf{J}_n(\mathbf{y}_n - \mathbf{x}_n) + O(||\mathbf{y}_n - \mathbf{x}_n||^2), \tag{10}$$

where $\mathbf{J}_n$ is the $m \times m$ Jacobian matrix of $\mathbf{F}$. Then, supposing Λ_i the eigenvalue of $\mathbf{J}$, the Lyapunov exponents are given by [19]

$$\lambda_i = \lim_{N \to \infty} \frac{1}{2N} ln|\Lambda_i^{(N)}|. \tag{11}$$

Hence, the study of the largest Lyapunov exponent aims at evaluating the sign of the value of the largest Lyapunov exponent, that represents the rate of separation of close trajectories in the phase space. Since the positive value of such divergence rate is a strong signature of the presence of chaos, it represents a suitable criterion for establishing the chaotic nature of a time series. In this respect, this paper computes the largest Lyapunov exponent by applying the Rosenstein method [46], which is based on the estimation of the local divergence rates of trajectories over the whole data set in the phase space. Specifically, the local divergence is estimated on the neighborhood of each point of $\mathcal{X}$ in the phase space. The general idea of the method is the measurement, for each x_i in $\mathcal{X}$, of the expansion rate in a particular time span δ of the trajectories. In particular, the expansion rate for time span δ is given by

Table I
LARGEST LYAPUNOV EXPONENT

Dataset	Lyapunov Exponent
Dataset 1	0.7
Dataset 2	1.6
Dataset 3	0.3

$$\mathcal{E}(\delta) = \frac{1}{t} \sum_{i=1}^{N} \ln\left(\frac{1}{|\mathcal{N}(\mathbf{z}_i)|} \sum_{\mathbf{z}_z \in \mathcal{N}_\epsilon(\mathbf{z}_i)} |\mathbf{z}_{i+\delta} - \mathbf{z}_{z+\delta}| \right); \tag{12}$$

where $\mathcal{N}(\mathbf{z}_i)$ is the neighborhood of point $\mathbf{z}_i$ in the phase space. More in depth, $\mathcal{N}_\epsilon(\mathbf{z}_i)$ derived from the selection of vectors $\mathbf{z}_z$ in m dimension, closer than a given value ϵ in the max norm. Hence, in order to define the neighborhood of $\mathbf{z}_i$, it is necessary to determine the indices z for which

$$||\mathbf{z}_z - \mathbf{z}_i|| \leq \epsilon. \tag{13}$$

The details of algorithm are reported in [46] and in our case, the resulted largest Lyapunov for all the three dataset analyzed has been reported in Table I. Since the largest Lyapunov exponents are a real number greater than zero, the considered time series indeed exhibit a chaotic behavior [19], [30], which justify the validation of our proposed chaotic framework.

V. TIME SERIES FORECASTING

The general idea behind the prediction of the behavior of $\mathcal{X}$, is the approximation of the map $\mathcal{F}$ in (6).

In order to predict the future behaviors of Dataset 1, Dataset 2, Dataset 3, we apply the PSR and the suitable values for time delay τ and embedding dimension m have been calculated in accordance with Sections IV and IV-C. In general term, there exist two main approaches to address the approximation problem of $\mathcal{F}$ function in (6): the local and the global approximation approach. One of the most used local prediction method is the neighbors based prediction, that evaluates the future behavior of the points belonging to a neighborhood around the point which has to be predicted. Then, the resulted forecast value is given by the average of the values of the neighbors points. Both the algorithms proposed for Dataset 1 and Dataset 2 constitute two general improved versions of the classical neighbors based prediction approach, in which each term is properly weighted. Despite different

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2019.2930363, IEEE Transactions on Vehicular Technology

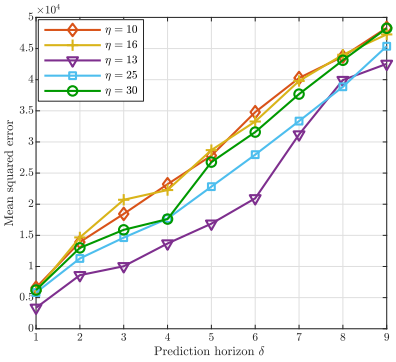

Figure 8. The predicted values by varying the number of considered neighbors for Dataset 1.

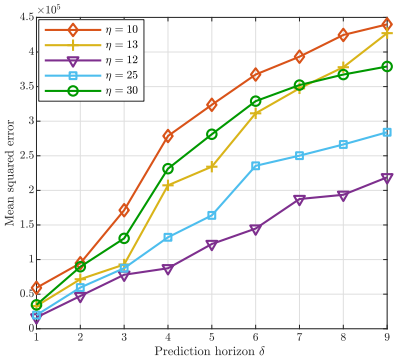

Figure 9. The predicted values by varying the number of considered neighbors for Dataset 2.

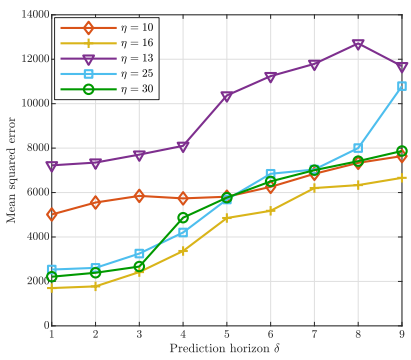

Figure 10. The predicted values by varying the number of considered neighbors for Dataset 3.

dataset may require different weight definitions, typically, the attribution of high weights to points close to that needing prediction, or to those that exhibit a high degree of similarity with the point to be forecast, could result in accuracy of prediction, compared with the standard version of the modified method. One strength point of local methods is that it does not involve any model for $\mathcal{F}$, while, as better explained later, the determination of the suitable number of neighbors to consider for the prediction is not a trivial issue. Furthermore, when determinism is weak on the dataset or for long term predictions, local methods result ineffective, since they tend to reproduce past trends. Differently, the algorithm proposed for Dataset 3 also considers the global approach. Roughly speaking, the global approaches constitute a more advanced technique compared to the local methods, but they often lead to more difficult problem management since modeling $\mathcal{F}$ implies the determination of many parameters.

Summarizing, as concerns the first two datasets, we propose two prediction algorithms based on a local approximation of $\mathcal{F}$. Differently, the prediction procedure for Dataset 3 constitutes a hybrid approach between the local and global approximation. In order to predict the value δ steps ahead of x_i, i.e. $x_{i+\delta}$, with a local approximation of $\mathcal{F}$, we consider the neighborhood around $\mathbf{z}_i$ in the phase space, of radius ϵ, i.e., $\mathcal{N}_\epsilon(\mathbf{z}_i)$.

A. Forcasting Algorithm for Dataset 1

The algorithm that we propose to predict the future behavior of Dataset 1, i.e., the number of Didi requests collected in a given hour in the city of Chengdu, is based on the evaluation, for each $\mathbf{z}_j \in \mathcal{N}_\epsilon(\mathbf{z}_i)$, of its trajectory and its value at time $j+\delta$, hence $x_{j+\delta}$. Then, the weighted mean value over all the future behaviors of the points belonging to $\mathcal{N}_\epsilon(\mathbf{z}_i)$ is computed, and the weight β_j is defined as

$$\beta_j = \frac{1}{|x_{j+\delta} - x_i|}. \tag{14}$$

As it is straightforward to note, the higher is the similarity between x_j and $x_{j+\delta}$, the higher is the value of β_j. Consequently, the predicted value of $x_{i+\delta}$, i.e., $\hat{x}_{i+\delta}$, being η the number of points in $\mathcal{N}_\epsilon(\mathbf{z}_i)$, is given by

$$\hat{x}_{i+\delta} = \frac{1}{\eta} \sum_{\mathbf{z}_j \in \mathcal{N}_\epsilon(\mathbf{z}_i)} \beta_i x_{i+\delta}. \tag{15}$$

Hence, the whole prediction procedure for Dataset 1 can be summarized as follows

- compute the optimal value for time delay τ according to Section IV-B;
- compute the optimal value for embedding dimension m according to Section IV-C;
- build $\mathcal{N}_\epsilon(\mathbf{z}_i)$;
- for each $\mathbf{z}_j \in \mathcal{N}_\epsilon(\mathbf{z}_i)$, measure $x_{j+\delta}$;
- for each $\mathbf{z}_j \in \mathcal{N}_\epsilon(\mathbf{z}_i)$, calculate β_j;
- determine $\hat{x}_{i+\delta}$ in accordance with (15).

The details of the forecast procedure are illustrated in Algorithm 1.

Algorithm 1 Forecasting Algorithm for Dataset 1

1: **Input:** chaotic time series $\mathcal{X}$, prediction horizon δ, radius ϵ, t index of the value to forecast;
2: Select the delay time τ suitable for $\mathcal{X}$, by evaluating the average mutual information in accordance with subsection IV-B;
3: Select the embedding dimension m suitable for $\mathcal{X}$, by applying the false nearest neighbors method in accordance with subsection IV-C;
4: Build neighborhood $\mathcal{N}_\epsilon(\mathbf{z}_i)$ around $\mathbf{z}_i$ and radius ϵ;
5: **for each** $\mathbf{z}_j \in \mathcal{N}_\epsilon(\mathbf{z}_j)$ **do**
6: evaluate its value $x_{j+\delta}$ along the trajectory of x_i, δ steps ahead;
7: compute its weight α_j as detailed in (14) ;
8: **end for**
9: Find the forecast value of $\hat{x}_{j+\delta}$ as expressed in (15);
10: **Output:** forecast value $\hat{x}_{j+\delta}$.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2019.2930363, IEEE Transactions on Vehicular Technology

Algorithm 2 Forecasting Algorithm for Dataset 2

1: **Input:** chaotic time series $\mathcal{X}$, prediction horizon δ, radius ϵ, t index of the value to forecast;
2: Select the delay time τ suitable for $\mathcal{X}$, by evaluating the average mutual information in accordance with subsection IV-B;
3: Select the embedding dimension m suitable for $\mathcal{X}$, by applying the false nearest neighbors method in accordance with subsection IV-C;
4: Build neighborhood $\mathcal{N}_\epsilon(\mathbf{z}_i)$ around $\mathbf{z}_i$ and radius ϵ;
5: **for each** $\mathbf{z}_j \in \mathcal{N}_\epsilon(\mathbf{z}_i)$ **do**
6: evaluate its value $x_{j+\delta}$ along the trajectory of x_j, δ steps ahead;
7: compute its weight γ_j as detailed in (16) ;
8: **end for**
9: Find the forecast value of $\hat{x}_{i+\delta}$ as expressed in (17);
10: **Output:** forecast value $\hat{x}_{i+\delta}$.

B. Forcasting Algorithm for Dataset 2

In order to predict the trend of Dataset 2, hence the number of total Uber pickups requests for a given hour in the city of New York, we define the following weight based on the similarity between the point object of prediction x_i and the neighbor point x_j. Hence, γ_j is defined as

$$\gamma_j = \frac{1}{|x_i - x_j|}. \tag{16}$$

Consequently, the prediction formula is expressed by

$$\hat{x}_{i+\delta} = \frac{1}{\eta} \sum_{\mathbf{z}_j \in \mathcal{N}_\epsilon(\mathbf{z}_i)} \gamma_i x_{i+\delta}. \tag{17}$$

The whole prediction procedure has been reported in the follows and reported in Algorithm 2, and its behavior is summarized in the following

- compute the optimal values for time delay τ and embedding dimension m according to Section IV-B and Section IV-C, respectively;
- build $\mathcal{N}_\epsilon(\mathbf{z}_i)$;
- for each $\mathbf{z}_j \in \mathcal{N}_\epsilon(\mathbf{z}_i)$, measure $x_{j+\delta}$;
- for each $\mathbf{z}_j \in \mathcal{N}_\epsilon(\mathbf{z}_i)$, calculate γ_j as reported in (16) ;
- determine $\hat{x}_{i+\delta}$ in accordance with (17).

C. Forcasting Algorithm for Dataset 3

This algorithm aims at predicting the behavior of Dataset 3, hence the number of total uber service requests in Banglore in a given hour. The algorithm prediction for Dataset 3 is a hybrid approach between the local approximation and the global one. Specifically, in order to improve the accuracy about the very short term forecasting, we provide a global nonlinear approximation of $\mathcal{F}$ given by the radial basis function [19] defined as follows

$$\mathbf{z}_{i+1} = \mathbf{F}(\mathbf{z}_i) = \theta_0 + \sum_{w=1}^{g} \theta_w \phi(|\mathbf{z}_i - \zeta_w|), \tag{18}$$

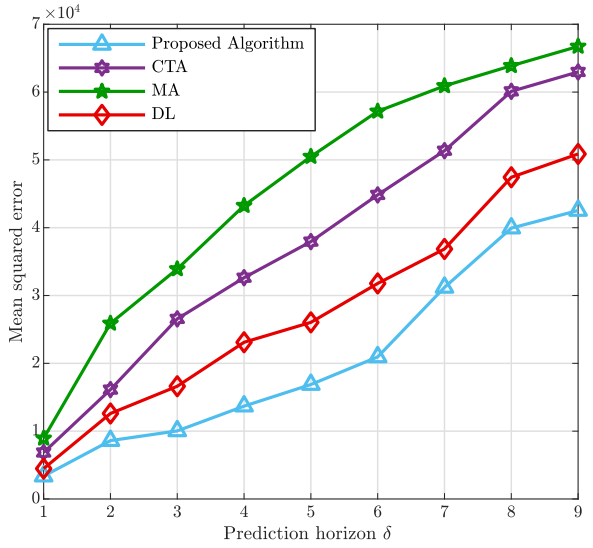

Figure 11. Forecasting error expressed in terms of mean squared error for Dataset 1.

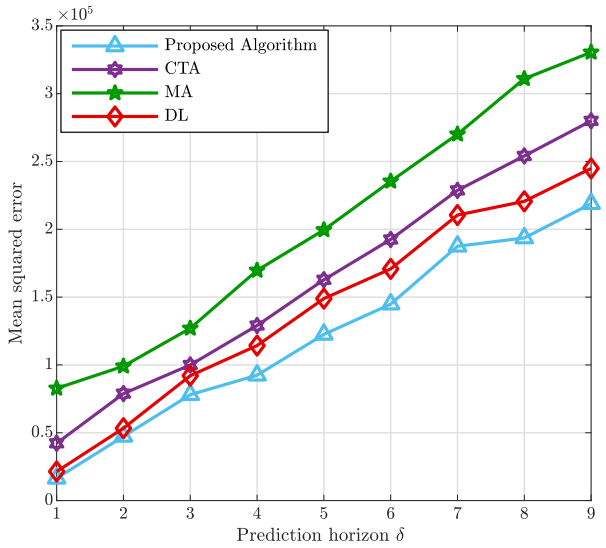

Figure 12. Forecasting error expressed in terms of mean squared error for Dataset 2.

where ζ_w are the g centers of the attractor, ζ_0 and ζ_w are coefficients, and ϕ is the Lorentzian function expressed by

$$\phi(\psi) = \frac{1}{[1 + (\frac{r}{a})^2]}, \tag{19}$$

with a constant. The joint utilization of both the approximation models presented in (17) and (18) has been pursued and, while the first approach ensures a lower long-term forecasting error, the second one improves the performance of the short-term forecasting. The prediction procedure acts as follows

- compute the most suitable value for time delay τ according to Section IV-B;

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2019.2930363, IEEE Transactions on Vehicular Technology

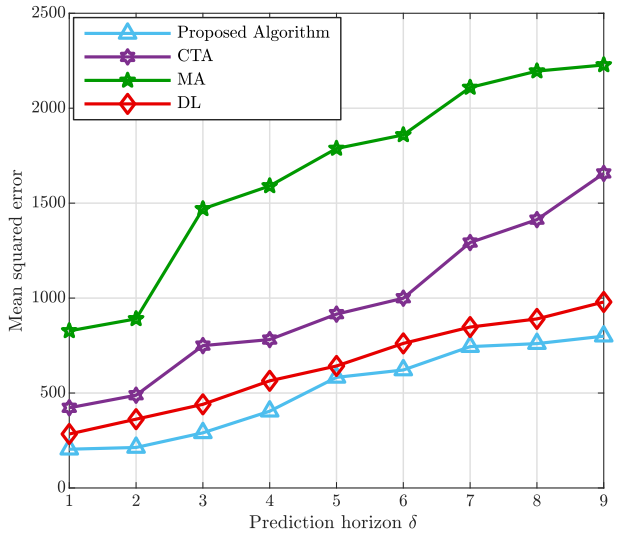

Figure 13. Forecasting error expressed in terms of mean squared error for Dataset 3.

Table II
ORDER OF MAGNITUDE OF COMPLEXITY

Time delay estimation	$\mathcal{O}(N \cdot \log N)$
Embedding Dimension	$\mathcal{O}(N^2+4\cdot\xi\cdot t+\xi^2\cdot N)$
Phase Space Reconstruction	$\mathcal{O}(\lambda \cdot m)$
Nearest neighbors procedure	$\mathcal{O}(N)$

- compute the most suitable value for embedding dimension m according to Section IV-C;
- build the approximation of $\mathcal{F}$ according to (18);
- build $\mathcal{N}_\epsilon(\mathbf{z}_i)$;
- for each $\mathbf{z}_j \in \mathcal{N}_\epsilon(\mathbf{z}_i)$, measure x_{j+1};
- for each $\mathbf{z}_j \in \mathcal{N}_\epsilon(\mathbf{z}_i)$, calculate γ_j;
- assign to $\hat{x}_{i+1}$ the most accurate value between the values obtained from the application of the local and the global approximation.

The details about the forecasting method for Dataset 3 is illustrated in Algorithm 3.

D. Practical Considerations

Since the value of ϵ directly impacts on the number of considered neighbors, the choice of such value crucially affects the prediction performance. Due to the fact that there not exists an exact method to determine the optimal value for ϵ [21], in Figure 8, Figure 9 and Figure 10 is reported the predictive performance of the algorithms designed for the three datasets. The better behavior is evident considering a number of neighbors $\eta = 13$ and $\eta = 12$ for Dataset 1 and Dataset 2, respectively, while for Dataset 3 the suitable value is $\eta = 16$. The time complexity is approximately the same for all the three algorithms previously presented. Let $\mathcal{X}$ be the time series composed of N scalar values, the estimation of the optimal value of τ requires a computational complexity in the order of $\mathcal{O}(N \cdot \log N)$, while by applying the procedure to find the suitable embedding dimension m the maximum amount of time taken is $\mathcal{O}(N^2 + 4 \cdot \xi \cdot t + \xi^2 \cdot N)$, where ξ is the number of considered m values. Furthermore, the phase space reconstruction exhibits a computational complexity of $\mathcal{O}(\lambda \cdot m)$, where $\lambda = N - (m-1)\tau$. Finally, the nearest neighbors procedure has a complexity in the order of $\mathcal{O}(N)$. Hence, we can conclude that the overall time complexity is

$$\mathcal{O}(N\cdot\log N)+\mathcal{O}(N^2+4\cdot\xi\cdot N+\xi^2\cdot N)+\mathcal{O}(\lambda\cdot m)+\mathcal{O}(N). \quad (20)$$

The order of magnitude of the forecasting based on CT strategy steps are reported in Table II.

Algorithm 3 Forecasting Algorithm for Dataset 3

1: **Input:** chaotic time series $\mathcal{X}$, prediction horizon δ, radius ϵ, i index of the value to forecast;
2: Select the delay time τ suitable for $\mathcal{X}$, by evaluating the average mutual information in accordance with subsection IV-B;
3: Select the embedding dimension m suitable for $\mathcal{X}$, by applying the false nearest neighbors method in accordance with subsection IV-C;
4: build the global approximation of $\mathcal{F}$ as in (18);
5: build neighborhood $\mathcal{N}_\epsilon(\mathbf{z}_i)$ around $\mathbf{z}_i$ and radius ϵ;
6: $y = 1$;
7: **for each** $\mathbf{z}_j \in \mathcal{N}_\epsilon(\mathbf{z}_j)$ **do**
8: evaluate $h = x_{j+y}$ along the trajectory of x_i, y steps ahead;
9: compute its weight γ_j as detailed in (16) ;
10: **end for**
11: Find the forecast value of $\hat{x}_{j+1}$ as expressed in (17);
12: select the most accurate forecast value between the ones originated from the local and the global approach;
13: repeat until the prediction is δ steps ahead;
14: **Output:** forecast value $\hat{x}_{j+\delta}$.

VI. EXPERIMENTAL RESULTS

In this paper we analyzed three sets of data related to the nonlinear time series represented by Datatset 1 derives from sampled data collected in the city of Chengdu, from Didi Chuxing, the biggest TNC in China. The dataset contains the passengers requests of one month, from 11/01/2016 to 11/30/2016, and the whole dataset contains more than 6.11 million of passengers requests. The whole area has been divided into 20 $\times$ 20 same-size grids. Every grid is a square with sides equal to 700 meters, and the longitude of the focus area is from 30.60E to 30.73E, the latitude is from 104.00N to 104.15N, while the considered surface is about 207.35 $10\,\mathrm{km}^2$. Furthermore, Dataset 2 and Dataset 3 have been retrieved from *Google Dataset Search* where they are named as *Uber pickups in New York City* and *Uber request data* respectively. Dataset 2 derives from a sampling period from April to September

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2019.2930363, IEEE Transactions on Vehicular Technology

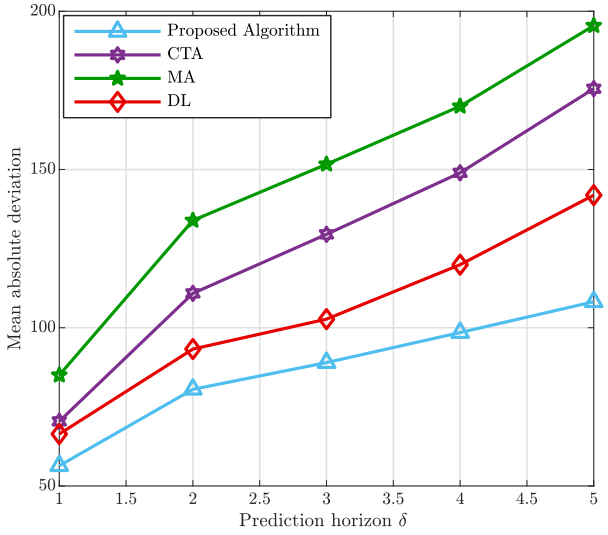

Figure 14. Mean absolute deviation for Dataset 1.

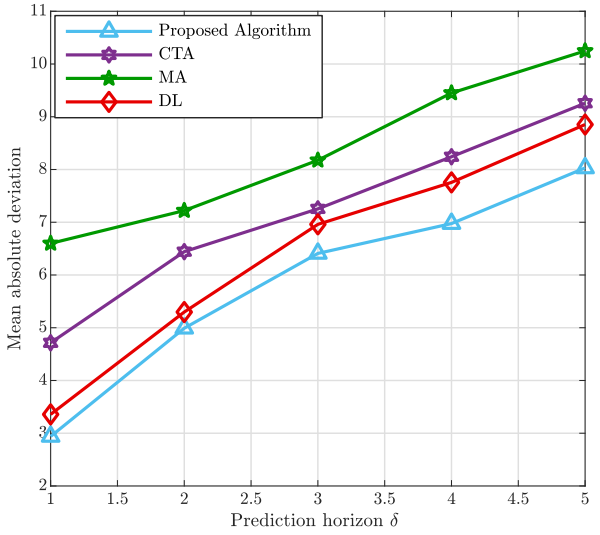

Figure 16. Mean absolute deviation for Dataset 3.

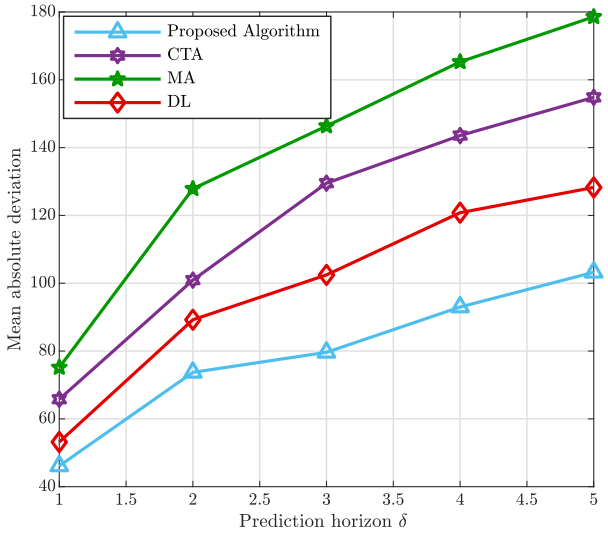

Figure 15. Mean absolute deviation for Dataset 2.

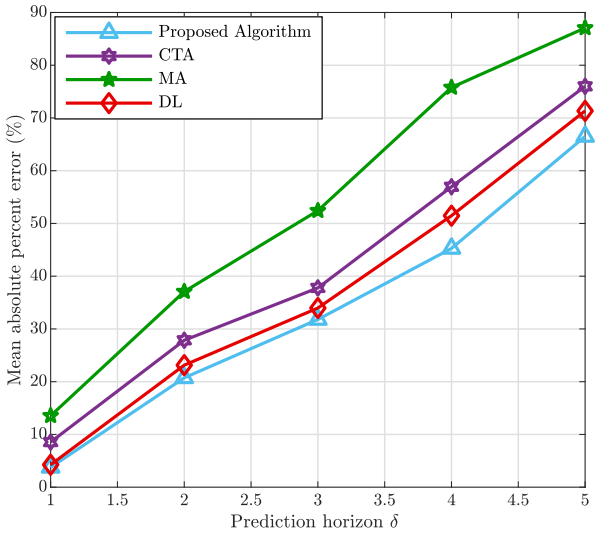

Figure 17. Mean absolute percentage error for Dataset 1.

2014, while Dataset 3 is referred to Uber service requests data in Bangalore from 11/07/2016 to 15/07/2016.

In order to provide an exhaustive analysis, we compare our methods with the well known moving average (MA) model [20], with a complexity linear in the number of the samples considered to provide forecasting, and with the CT based approach proposed in [21] (CTA) whose complexity is in the order of magnitude of $\mathcal{O}(N^2)$. Furthermore, we also propose performance comparison with the deep neural network method (DL) designed in [22]. Furthermore, all the approaches have been applied by using 3 days of samples to forecast 5 and 9 hours ahead.

Figures 17, 18 and 19 show the performance comparison among the proposed algorithm, the CTA, the MA, and the DL models in terms of MSE. Despite all the four predictive approaches get worse as the prediction horizon increases, it is clearly evident as the proposed forecasting methods reach better results in all the three application dataset, by considering the same number of training days for all the methodologies applied. In fact, each algorithm guarantees a higher accuracy respect to the CTA, the MA, and the deep neural network strategies. In order to analyze the strategies performance in terms of measure of the variability of the forecast errors, MSE and MAD have been represented in Figures 11 - 16. As is evident to note, in Figures 11 - 16 the forecasting accuracy is better when the proposed algorithms is adopted. Then, it is clearly evident that both the MSE and MAD increase for high values of δ. It is due to the general difficulty in predicting

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2019.2930363, IEEE Transactions on Vehicular Technology

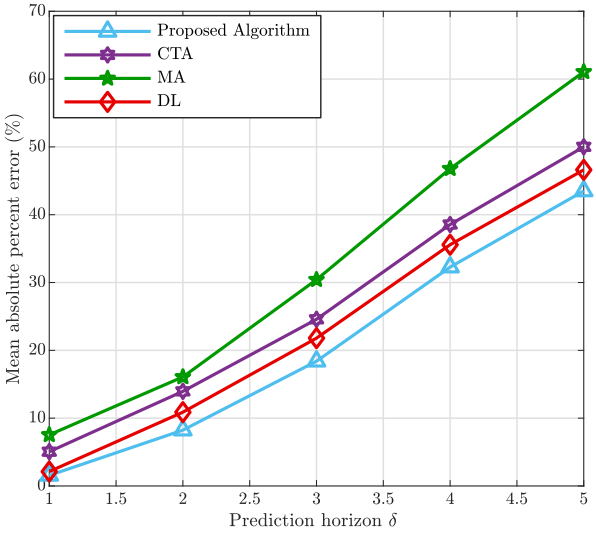

Figure 18. Mean absolute percentage error for Dataset 2.

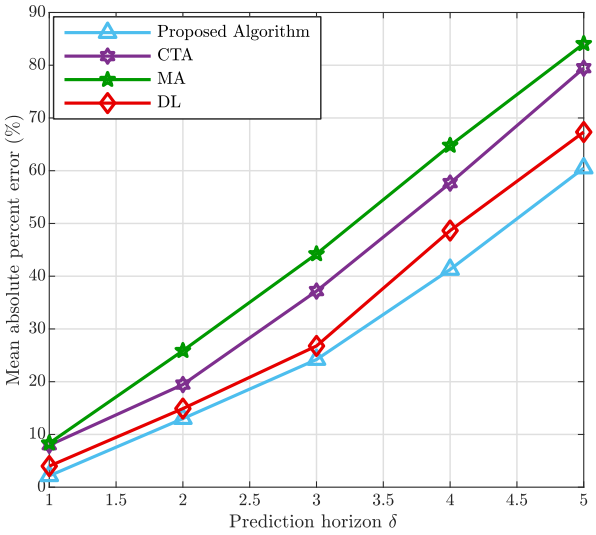

Figure 19. Mean absolute percentage error for Dataset 3.

behavior for long interval times. In order to better quantify the meaning of the MSE and MAD, Figures 17 and 19 show the performance of the proposed algorithms in comparison with the considered alternatives, expressed in terms of MAPE metric. The results confirm the good performance of the proposed approach in comparison with the alternatives taken into account, for all the three different strategies proposed for each dataset. In conclusion, the proposed CT approaches provide a suitable solution to forecast values in complex and nonlinear dynamical systems, by investigating and capturing their underlying dynamics and geometrical structure, in the corresponding reconstructed phase space, and chasing the time series behavior, guaranteeing more accuracy than the MA, CTA, and the DL models.

VII. Conclusions

This paper addresses the problem of the prediction of the service requests for the TNCs. In particular, different algorithms for different real datasets have been presented. The predictive methods designed for the three analyzed dataset are based on the CT principles and the corresponding phase space has been reconstructed, the chaotic behavior studied, through the analysis of the largest Lyapunov exponent. Furthermore, a different CT based algorithm has been proposed for the different datasets studied. The validity of the proposed strategies have been confirmed by simulations and comparison with the MA, the CT based approach presented in [21], and the one discussed in [22]. Finally, system performance has been expressed in terms of mean squared, mean absolute error and mean percent forecasting error.

Acknowledgement

This work was partially supported by the Project "GAUChO—A Green Adaptive Fog Computing and Networking Architecture" funded by the MIUR, Progetti di Ricerca di Rilevante Interesse Nazionale, Bando 2015 under Grant $2015YPXH4W\_004$. (Corresponding author: Romano Fantacci), and by US MURI AFOSR MURI 18RT0073 , NSF CNS-1717454, CNS-1731424, CNS-1702850, CNS-1646607.

References

[1] S. Yi, H. Wei, L. Xiao-bo, L. Bin, and L. Tao, "Study of two-dimensional motion estimation technique in vehicle tracking," in *the 6th International Conference on ITS Telecommunications*, Chengdu, China, June 2006.

[2] S. Rhee, "Catalyzing the internet of things and smart cities: Global city teams challenge," in *1st International Workshop on Science of Smart City Operations and Platforms Engineering (SCOPE) in partnership with Global City Teams Challenge (GCTC) (SCOPE - GCTC)*, Vienna, Austria, April 2016.

[3] N. Mohamed, J. Al-Jaroodi, S. Lazarova-Molnar, I. Jawhar, and S. Mahmoud, "A service-oriented middleware for cloud of things and fog computing supporting smart city applications," in *IEEE SmartWorld Congress*, San Francisco, CA, August 2017.

[4] M. Pečar and G. Papa, "Transportation problems and their potential solutions in smart cities," in *International Conference on Smart Systems and Technologies (SST)*, Osijek, Croatia, October 2017, pp. 195–199.

[5] T. Hu, G. Zheng, and T. Liao, "Smart mobility: Evaluation of demand-responsive transit systems in chiayi city," in *International Smart Cities Conference (ISC2)*, Wuxi, China, September 2017.

[6] L. Moreira-Matias, J. Gama, M. Ferreira, J. Mendes-Moreira, and L. Damas, "Predicting taxi–passenger demand using streaming data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 3, pp. 1393–1402, September 2013.

[7] B. M. Williams and L. A. Hoel, "Modeling and forecasting vehicular traffic flow as a seasonal arima process : Theoretical basis and empirical results," *J. Transp. Eng.*, vol. 129, no. 6, pp. 664–672, November / December 2003.

[8] M. Papadopouli, E. Raftopoulos, and H. Shen, "Evaluation of short-term traffic forecasting algorithms in wireless networks," in *2nd Conference on Next Generation Internet Design and Engineering*, Valencia, Spain, April 2006.

[9] R. Weron, "Electricity price forecasting: A review of the state-of-the-art with a look into the future," *International Journal of Forecasting*, vol. 30, no. 4, pp. 1030 – 1081, October 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0169207014001083

[10] X. Liu, X. Fang, Z. Qin, C. Ye, and M. Xie, "A short-term forecasting algorithm for network traffic based on chaos theory and svm," *Journal of Network and Systems Management*, vol. 19, no. 4, pp. 427–447, December 2011. [Online]. Available: https://doi.org/10.1007/s10922-010-9188-3

[11] D. C. Montgomery, C. L. Jennings, and M. Kulahci, *Time Series Analysis and Forecasting*, 2nd ed. Cambridge, UK: Wiley, 2015.

[12] G. Box, *Box and Jenkins: Time Series Analysis, Forecasting and Control*. London: Palgrave Macmillan UK, 2013, pp. 161–215. [Online]. Available: https://doi.org/10.1057/9781137291264_6

[13] S. Lee and D. Fambro, "Application of subset autoregressive integrated moving average model for short-term freeway traffic volume forecasting," vol. 1678, pp. 179–188, November 1999.

[14] C. K. Moorthy and B. G. Ratcliffe, "Short term traffic forecasting using time series methods," *Transportation Planning and Technology*, vol. 12, no. 1, pp. 45–56, July 1988.

[15] A. G. Salman, B. Kanigoro, and Y. Heryadi, "Weather forecasting using deep learning techniques," in *International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, Depok, Indonesia, October 2015, pp. 281–285.

[16] X. Qiu, L. Zhang, Y. Ren, P. N. Suganthan, and G. Amaratunga, "Ensemble deep learning for regression and time series forecasting," in *IEEE Symposium on Computational Intelligence in Ensemble Learning (CIEL)*, Singapore, December 2014.

[17] Y. He, J. Deng, and H. Li, "Short-term power load forecasting with deep belief network and copula models," in *9th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, Hangzhou, China, August 2017, pp. 191–194.

[18] F. Liu, F. Xu, and S. Yang, "A flood forecasting model based on deep learning algorithm via integrating stacked autoencoders with bp neural network," in *IEEE Third International Conference on Multimedia Big Data (BigMM)*, Laguna Hills, CA, April 2017, pp. 58–61.

[19] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*, 2nd ed. Cambridge: Cambridge University Press, 2003.

[20] P. M. T. Broersen, "The best order of long autoregressive models for moving average estimation," in *the 8th European Signal Processing Conference (EUSIPCO 1996)*, Trieste, Italy, September 1996.

[21] H. Cui and X. Song, "Research on electricity price forecasting based on chaos theory," in *International Seminar on Future Information Technology and Management Engineering*, Leicestershire, UK, November 2008, pp. 398–401.

[22] W. Chandramitasari, B. Kurniawan, and S. Fujimura, "Building deep neural network model for short term electricity consumption forecasting," in *2018 International Symposium on Advanced Intelligent Informatics (SAIN)*, Aug 2018, pp. 43–48.

[23] A. Ihler, J. Hutchins, and P. Smyth, "Adaptive event detection with time-varying poisson processes," in *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '06, New York, NY, January 2006, pp. 207–216. [Online]. Available: http://doi.acm.org/10.1145/1150402.1150428

[24] D. Celebi, B. Bolat, and D. Bayraktar, "Light rail passenger demand forecasting by artificial neural networks," in *International Conference on Computers Industrial Engineering*, Troyes, France, July 2009, pp. 239–243.

[25] S. M. T. F. Ghomi and K. Forghani, "Airline passenger forecasting using neural networks and box-jenkins," in *12th International Conference on Industrial Engineering (ICIE)*, Tehran, Iran, January 2016, pp. 10–13.

[26] N. H. Abdullah, R. Adnan, A. M. Samad, and F. Ahmat Ruslan, "Lightning forecasting modelling using artificial neural network (ann): Case study sultan abdul aziz shah airport or skypark subang," in *2018 IEEE Conference on Systems, Process and Control (ICSPC)*, Dec 2018, pp. 1–4.

[27] J. Xu, R. Rahmatizadeh, L. Bölöni, and D. Turgut, "Real-time prediction of taxi demand using recurrent neural networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 8, pp. 2572–2581, Aug 2018.

[28] J. Li, H. Peng, L. Liu, G. Xiong, B. Du, H. Ma, L. Wang, and M. Zakirul Alam Bhuiyan, "Graph cnns for urban traffic passenger flows prediction," in *2018 IEEE SmartWorld, Ubiquitous Intelligence Computing, Advanced Trusted Computing, Scalable Computing Communications, Cloud Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, Oct 2018, pp. 29–36.

[29] Y. Cui, W. Lv, Q. Wang, and B. Du, "Usage demand forecast and quantity recommendation for urban shared bicycles," in *2018 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC)*, Oct 2018, pp. 238–2388.

[30] Z. Liu, H. Yang, and M. Lai, "Electricity price forecasting model based on chaos theory," in *International Power Engineering Conference*, Leicestershire, UK, November 2005.

[31] M. Cai, F. Cai, A. Shi, B. Zhou, and Y. Zhang, "Chaotic time series prediction based on local-region multi-steps forecasting model," in *Advances in Neural Networks - ISNN*, F.-L. Yin, J. Wang, and C. Guo, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 418–423.

[32] H. Wang and D. Chi, "Short-term load prediction based on chaos time series theory," in *Second International Conference on Intelligent Computation Technology and Automation*, vol. 2, Changsha, China, October 2009, pp. 189–192.

[33] K. Chang, J. Gao, M. Chen, and Y. Yuan, "Urban daily water demand short-term forecasts based on the chaos theory," in *Fourth International Conference on Natural Computation*, vol. 4, Jinan, China, October 2008, pp. 696–699.

[34] J. Mao, X. Zhang, and J. Li, "Wind power forecasting based on chaos and wavelet packet theory," in *13th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, Wuhan, China, May 2018, pp. 1604–1608.

[35] G. Li, C. Zhu, and X. Li, "Application of chaos theory and wavelet to modeling the traffic of wireless sensor networks," in *International Conference on Biomedical Engineering and Computer Science*, Wuhan, China, April 2010.

[36] S. Kawauchi, H. Sugihara, and H. Sasaki, "Development of very-short-term load forecasting based on chaos theory," *Electrical Engineering in Japan*, vol. 148, no. 2, pp. 55–63. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/eej.10322

[37] A. Basharat and M. Shah, "Time series prediction by chaotic modeling of nonlinear dynamical systems," in *IEEE 12th International Conference on Computer Vision*, Kyoto, Japan, September 2009, pp. 1941–1948.

[38] E. A. Nadaraya, "On estimating regression," *Theory Pb. Appl.*, 1964.

[39] E. Ott, *Chaos in Dynamical Systems*. Cambridge: Cambridge University Press, 2002.

[40] K. T. Alligood, T. D. Sauer, and J. A. Yorke, *Chaos: an Introduction to Dynamical Systems*. New York, USA: Spinger, 1997.

[41] F. Takens, "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence, Warwick 1980*, D. Rand and L.-S. Young, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 1981, pp. 366–381.

[42] A. M. Fraser and H. L. Swinney, "Independent coordinates for strange attractors from mutual information," *Phys. Rev. A*, vol. 33, no. 2, pp. 1134–1140, February 1986. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.33.1134

[43] Z. Liu, "Chaotic time series analysis," vol. 2010, no. 1-2, pp. 1–31, 2010.

[44] M. B. Kennel, R. Brown, and H. D. I. Abarbanel, "Determining embedding dimension for phase-space reconstruction using a geometrical construction," *Phys. Rev. A*, vol. 45, no. 6, pp. 3403–3411, March 1992. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.45.3403

[45] H. Mori and S. Urano, "Short-term load forecasting with chaos time series analysis," in *Proceedings of International Conference on Intelligent System Application to Power Systems*, Jan 1996, pp. 133–137.

[46] M. T. Rosenstein, J. J. Collins, and C. De Luca, "A practical method for calculating largest lyapunov exponents from small data set," vol. 65, no. 1-2, pp. 117–134, May 1993.