Uncovering Quasi-periodic Nature of Physical Systems: A Case Study of Signalized Intersections

Suddhasattwa Das¹, Shakib Mustavee², Shaurya Agarwal²

Abstract—This paper presents a novel approach to analyze quasiperiodically driven dynamical systems. It aims to develop a complete data-driven framework for modeling such unknown dynamics. To achieve this, we characterize Koopman eigenfrequencies as generating frequencies of the quasiperiodic driver of the system. We compute true eigenfrequencies of Koopman operators by applying the theory of Reproducing Kernel Hibert Space (RKHS) and results from ergodic theory. We also demonstrate the decomposition of quasiperiodically driven dynamics into two components, i) the quasiperiodic driving source with generating frequencies and ii) the driven nonlinear dynamics. A unique aspect of the proposed framework is that it applies to the analysis of systems where the periodic component is either non-dominant or even absent. As a case study, we analyze a system of nine traffic signalized intersections. The proposed framework accurately reconstructs the measured queue lengths of the signalized intersections and makes stable long-term predictions.

Index Terms—Koopman operator, Reproducing Kernel Hibert Space, Quasiperiodic systems, Signalized intersections

I. INTRODUCTION

Many physical phenomena show periodicities, such as oscillations and limit cycles, characterized by a single frequency or period. There are also many phenomena, such as the arrival of full foliage, the arctic ice cycle, and planetary dynamics, in which there are multiple driving frequencies. As a result of which the states of the system do not show exact and regular periodicity. Such systems are called quasiperiodic systems. More complicated systems exhibit chaos and a high degree of nonlinearity but having a quasiperiodic component as a driving source. Examples are astronomical systems [1], climate systems [2], [3], geophysical flows on periodic domains [4], epidemics [5], and computational neuroscience [6].

The goal of this paper is to develop a robust method for reconstructing quasiperiodically driven dynamical systems using the Koopman operator framework. Koopman operator is originally a linear operator that describes the spatiotemporal evolution of a nonlinear dynamical system. Projection of observables on the eigenfunction of the Koopman operator is described quasiperiodicity in terms of eigenfunctions of the Koopman operator. However, an accurate estimation of the eigenfunctions and eigenfrequencies has remained an elusive task. Techniques such as DMD [7], [8], deep neural networks [9], [10], or Fourier averaging [1] are inadequate when the dynamics has a substantial chaotic component.

Our novel approach uses a technique from RKHS interpolation theory developed in [11] to extract the true eigenfrequencies. After that, we proceed to identify different components of the dynamics in a manner similar in structure to [6]. This approach leads to a robust theoretical and numerical method to reconstruct quasiperiodically driven systems having strong chaotic components by identifying their true eigenfrequencies. As a case study, we study the queue length dynamics on a corridor of traffic signalized intersections. Such a system has quasiperiodic driving sources, but the measurement data also reflects unpredictable phenomena such as accidents, road constructions, and human factors. It makes the study more challenging.

Contributions: The contributions of this paper is as follows:

• We provide an alternative mathematical formulation of the quasiperiodic coordinates in terms of the Koopman operator that correctly recovers both the periodic and chaotic part of the dynamics.

• The proposed framework is even applicable to the systems where the periodic component is either non-dominant or even absent.

• We describe the physical significance of the quasiperiodic sources obtained from the proposed technique

Outline: Rest of the paper is arranged as follows: The dynamical systems theory related to quasiperiodically driven systems is described in Section II-A. We discuss the relevant concepts and techniques in Section II-B and Section II-C. The proposed data-driven implementation is described in Section III. Section IV performs a case study using real measurements from traffic intersections and discusses the results.

II. PROPOSED FRAMEWORK

A. Formulation

A quasiperiodically driven system has the form

\[ \begin{align*}
\theta_{n+1} &= \theta_n + \rho \mod T^d, \\
x_{n+1} &= f(x_n, \theta_n)
\end{align*} \] (1)

where \( f \) is some nonlinear function. The angular coordinate \( \theta \) is a point on a \( d \)-dimensional torus \( T^d \). \( \theta \) represents the phase of a driving quasiperiodic system. The vector \( \rho \) is called the rotation vector [12], [13]. It represents the angular
increments at each step for each of the coordinates of $\theta$. If the underlying system arises from a continuous time system by taking samples at intervals $\Delta t$, then $\tilde{\rho} = \Delta \tilde{\omega}$ for some angular frequency vector $\tilde{\omega}$. The variable $x$ lies in some abstract or unknown manifold $\mathcal{X}$. Let $\Omega := \mathbb{T}^d \times \mathcal{X}$. Thus (1) is a one-way coupled or skew-product dynamical system on the space $\Omega$:

$$F: \Omega \rightarrow \Omega, \quad \begin{pmatrix} \theta_{n+1} \\ x_{n+1} \end{pmatrix} = F(\begin{pmatrix} \theta_n \\ x_n \end{pmatrix}) := \begin{pmatrix} \theta_n + \tilde{\omega} \\ g(\theta_n, x_n) \end{pmatrix}. \quad (2)$$

The space $\Omega$ is unknown, and its points are pairs of points from $\mathbb{T}^d$ and $\mathcal{X}$ respectively. This abstract formulation applies to all quasiperiodically driven systems. We further assume the dynamics in (1) to be of the form

$$\begin{align*}
\theta_{n+1} &= \theta_n + \Delta t \tilde{\omega} \mod 2\pi \\
x_{n+1} &= g_{\text{per}}(\theta_n) + g_{\text{chaos}}(\theta_n, x_n)
\end{align*} \quad (3)$$

Thus the task is to find (i) the quasiperiodicity dimension $d$ and the rotation vector $\tilde{\omega}$; and the functions (ii) $g_{\text{per}}$ and (iii) $g_{\text{chaos}}$. The functions and spaces above will be assumed unknown, and possibly a low-dimensional or partial observation space $\Omega$. It generates a sequence of $k$-dimensional data points $\{y_n := (\theta_n, x_n) : n = 0, 1, 2, \ldots\}$, where $(\theta_n, x_n)$ is a trajectory of the dynamics in (2) under $F$.

**B. Koopman operator and its spectrum**

The Koopman operator $U$ is essentially a time-shift operator. It operates on functions instead of points on the phase space. Given a function $\phi: \Omega \rightarrow \mathbb{R}$, $U\phi$ is the function defined as

$$(U\phi)(z) := \phi(Fz), \quad \forall z \in \Omega, \quad (4)$$

where $F$ is the underlying dynamical system (2). $\phi$ can be interpreted as a measurement or observation on the phase space $\Omega$, and $U\phi$ is the evolution/transformation of this measurement with the dynamics.

**a) Koopman eigenfrequencies**: Since $U$ is unitary, its spectrum must lie on the unit circle of the complex plane. The eigenvalues of $U$ correspond to the point spectrum, and any eigenfunction $\zeta$ has a corresponding eigenvalue of the form $e^{i\omega}$ for some $\omega \in \mathbb{R}$. Thus

$$\begin{align*}
(U^n\zeta)(z) &= \zeta(F^n z) = e^{i\omega n}\zeta(z), \quad \forall n \in \mathbb{N}.
\end{align*} \quad (5)$$

$\omega$ is called the Koopman eigenfrequency corresponding to $\zeta$. Thus the time-evolution of Koopman eigenfunctions is equivalent to multiplication by $e^{i\omega n}$ as a function of time $n$.

$U$ always has the constant functions as eigenfunctions with eigenfrequency 0. $U$ may or may not have other eigenfrequencies. The collection of eigenfunctions and (eigen-)frequencies have an algebraic structure to them. For any two frequencies $\omega_1, \omega_2$, and integers $a, b$, $a\omega_1 + b\omega_2$ is also a frequency [11]. If the system has at least one nonzero frequency, then it has all harmonics of that frequency and thus infinitely many frequencies. A collection of eigenfrequencies is said to be independent if no integer linear combination is an integer. If the system has two independent frequencies, then all its frequencies are together dense on the real line. A collection of frequencies will be called a basis or generating set of eigenfrequencies if they are independent, and all frequencies of the system can be generated by taking integer linear combinations of frequencies from this set. There is no unique choice of a basis, but all bases will have the same cardinality $d$, called the quasiperiodicity dimension. $d$ is a fixed finite number if $\Omega$ is a finite-dimensional manifold.

**b) Koopman eigenfunctions and torus dynamics**: Koopman eigenfunctions reveal quasiperiodic dynamics embedded in the system. Any $k$ Koopman eigenfunctions leads to a rotation on a $k$-dimensional torus:

$$\begin{align*}
\zeta_1, \ldots, \zeta_k &\mapsto R_{\tilde{\omega}}(\zeta_1, \ldots, \zeta_k) \\
\mathbb{T}^k &\rightarrow \mathbb{T}^k
\end{align*} \quad (6)$$

Here, $R_{\tilde{\omega}}$ is a rotation by the vector $\tilde{\omega}$ of frequencies. If these eigenfrequencies are independent, then this map will be surjective. Taking $d = k$ implies that the dynamics has an embedded factor torus rotation of the same dimension as the quasiperiodicity dimension.

This completes our examination of the quasiperiodic structure of the dynamics (2). We next discuss some techniques from Functional Analysis for reconstructing the quasiperiodic component and its complement.

**C. Kernels and integral operators**

A kernel is a function $k: M \times M \rightarrow \mathbb{R}$ on some space $M$. The quantity $k(x, y)$ is measure of similarity, closeness or distance between two points $x, y \in M$. Kernel based methods have been used very effectively to obtain information such as statistical manifolds [14], geometric information [15], and dynamical information such as tracer flows [4], stable/unstable foliations [16], Koopman spectrum [11], [17]. The techniques in this paper are based on [11]. We shall use the Gaussian kernel

$$k_\epsilon(x, y) := \exp \left( -\frac{1}{\epsilon} d(x, y)^2 \right),$$

where $\epsilon$ is called the bandwidth parameter, and $d(\cdot, \cdot)$ is some notion of metric or distance on the space.

**a) Delay-coordinates**: The data sequence $y_n$ is obtained through an observation $Y$. However $Y$ may not be a one-to-one map and its values may not correspond to unique states in $\Omega$. We convert $Y$ into an embedding using the method of delay coordinates [18], by incorporating $Q$ delays to get the map $Y^{(Q)}: \Omega \rightarrow \mathbb{R}^{Q+1}$:

$$Y^{(Q)}(\omega) := (Y(\omega), Y(F^1\omega), \ldots, Y(F^Q\omega)).$$

Thus the delay coordinate version of each point $y_n$ is

$$y_n \leftrightarrow y_n^{(Q)} := (y_n, y_{n+1}, \ldots, y_{n+Q}).$$

We next use the Gaussian shape function to implicitly obtain a kernel $k_\epsilon: \Omega \times \Omega \rightarrow \mathbb{R}$ as follows

$$k_\epsilon(z, z') := \exp \left( -\frac{1}{\epsilon} \left\| Y^{(Q)}(z) - Y^{(Q)}(z') \right\|^2 \right). \quad (6)$$
Even if the two states $z, z'$ are unknown, the left-hand side in (6) can be computed since the right-hand side only uses the observation map $Y$. We next modify $k_{e}$ by a process called 

**bistochastic normalization** [19] to get a kernel $p_{e}$ which is symmetric, Markovian, and more adapted to the non-uniform distribution of the data.

$$
\tilde{k}_{e}(z, z') := \frac{k_{e}(z, z')}{\deg_{R}(z) \deg_{L}(z')^{1/2}}.
$$

$p_{e}(z, z') := \tilde{k}_{e}(z, z')\tilde{k}_{e}(z', z')d\mu(z')$. 

**b) Kernel integral operator:** Associated to the kernel $p_{e} : \Omega \times \Omega \to \mathbb{R}$ is the integral operator $P_{e}$, which operates on $L^{2}(\mu)$ functions $\phi : \Omega \to \mathbb{R}$ as

$$
(P_{e}\phi)(z) := \int_{\Omega} p_{e}(z, z')\phi(z')d\mu(z').
$$

$P_{e}$ is a compact, symmetric operator on $L^{2}(\mu)$ [e.g. 19]. Moreover $P_{e}$ has a complete basis of eigenfunctions

$$
P_{e}\phi_{j} = \lambda_{j}\phi_{j}, \quad j = 0, 1, 2, \ldots ,
$$

where the indexing is done so that the $\lambda_{j}$s are in decreasing order. Due to the normalizations carried out, we have $\phi_{1} \equiv 1_{\Omega}$, the constant function equal to 1 everywhere. Moreover, the eigenvalues satisfy $1 = \lambda_{1} \geq \lambda_{2} \geq \lambda_{3} \geq \ldots > 0$. Also importantly, the $\phi_{j}$s are an orthonormal basis, i.e.,

$$
\langle \phi_{i}, \phi_{j} \rangle_{L^{2}(\mu)} = \int_{\Omega} \phi_{i}^{*}(z)\phi_{j}(z)d\mu(z) = \delta_{i,j}.
$$

All these properties of the $\lambda_{j}$ and $\phi_{j}$ are useful for kernel-based learning, in which we recreate or extrapolate unknown functions from some samples using these $\phi_{j}$s as a basis.

**c) Kernel based learning:** One of the main advantages of kernel-based approaches is that while the $\phi_{j}$ can be approximated to any degree of accuracy by solving an eigenvalue equation of a data-driven matrix; they can be easily extended from vectors to a continuous function over the entire data space $\mathbb{R}^{k((Q+1)}$ and thus on $\Omega$. The $\phi_{j}$s also happen to be left singular vectors of the asymmetric operator $\tilde{K}_{e}$, with $\sqrt{\lambda_{j}}$ and $\gamma_{j}$ being the associated singular values and right singular vectors. Thus for an arbitrary point $z \in \mathbb{R}^{k((Q+1)}$, we have

$$
\phi_{j}(z) = \lambda_{j}^{-1} \int_{\Omega} p_{e}(z, z')\phi_{j}(z')d\mu(z') = \lambda_{j}^{-1} \int_{\Omega} \tilde{k}_{e}(z, z')\gamma_{j}(z')d\mu(z').
$$

These $\phi_{j}$ for the basis for learning any function $f : \mathbb{R}^{k(Q+1)} \to \mathbb{R}^{d}$. It is done by first computing the components of $f$ along the first $L$ $\phi_{j}$

$$
f_{i} := \langle \phi_{i}, f \rangle_{L^{2}(\mu)} = \int_{\mathbb{R}^{k(Q+1)}} \phi_{i}^{*}(z)f(z)d\mu(z),
$$

and then taking the sum/integral

$$
f(y) \approx \sum_{i=1}^{L} f_{i}(y) = \int_{\mathbb{R}^{k(Q+1)}} p_{e}(y, z) \sum_{i=1}^{L} \frac{f_{i}}{\lambda_{i}} \phi_{i}(z)d\mu(z).
$$

The parameter $L$ is called the spectral truncation parameter; it is the size of the hypothesis space.

We next describe the use of the fast-Fourier transform to derive the Koopman eigenfrequencies but on the $L$ eigenfunctions instead of the raw data.

**III. THE DATA-DRIVEN PROCEDURE**

In the data-driven approach, all of the entities described in Section II-C have a data-driven analog. We begin with the invariant measure $\mu$ itself. It will be replaced by $\mu_{N} = \frac{1}{N} \sum_{n=1}^{N} \delta_{y_{n}(Q)}$, the average of the Dirac-delta measure on the data points $y_{n}(Q)$. These are called sampling/empirical measures. Their integrals with respect to these sampling measures are given by

$$
\int_{\mathbb{R}^{k(Q+1)}} \phi d\mu_{N} = \frac{1}{N} \sum_{n=1}^{N} \phi \left( \delta_{y_{n}(Q)} \right),
$$

for every continuous test function $\phi : \mathbb{R}^{k(Q+1)} \to \mathbb{R}$. The kernel integral operators $K_{i}, \tilde{K}$ and $P$ will be approximated as $N \times N$ matrices $[K]$, $[\tilde{K}]$ and $[P]$.

$$
[K]_{i,j} := k_{e}(y_{i}(Q), y_{j}(Q)), \quad \tilde{d} := \frac{1}{N}\mathbb{E}_{y,z} [K]_{i,z}.
$$

Next compute

$$
\tilde{q} := \frac{1}{N} [K] D^{-1}, \quad D := \text{diag}(\tilde{d}).
$$

$\tilde{d}$ and $\tilde{q}$ are called left and right degree vectors respectively. Finally compute

$$
[\tilde{K}] := D^{-1} [K] Q^{-1/2}, \quad Q := \text{diag}(\tilde{q}).
$$

The matrix $[P] := [\tilde{K}] [\tilde{K}]^{*}$ is not computed explicitly. See [19, Algorithm 1] for more details. Compute the top $L$ singular values $1 = \sigma_{1} \geq \ldots \geq \sigma_{L}$ of $[\tilde{K}]$ and the corresponding left eigenvectors $\bar{q}_{1}, \ldots, \bar{q}_{L}$ and right singular vectors $\bar{q}_{1}, \ldots, \bar{q}_{L}$. Set $\lambda_{i} := \sigma_{i}^{2}$, for $i = 1, \ldots, L$. The set of vectors $\left\{ \phi_{i}^{l} : l = 1, \ldots, L \right\}$ and $\left\{ \bar{q}_{i}^{l} : l = 1, \ldots, L \right\}$ are both orthonormal systems for $\mathbb{C}^{N}$. The eigenvectors have continuous extensions to any $y \in \mathbb{R}^{k(Q+1)}$ as

$$
\phi^{l}(y) := \frac{1}{N \lambda_{i}} \bar{q}_{i}^{l}(y)\bar{q}_{i}^{l} \bar{q}_{i}^{l} Q^{-1/2} \bar{q}_{i}^{l},
$$

$$
\tilde{k}_{e}(y) := \left( k_{e}(y, y_{1}(Q)), \ldots, k_{e}(y, y_{N}(Q)) \right).
$$

The function $\bar{q}_{i}$ is continuous as the vector $\tilde{k}_{e}(y)$ is a continuous function of $y$. If $y$ in the above equation is substituted by one of the data-points $y_{n}(Q)$, then by design $\phi^{l}(y_{n}(Q)) = \phi^{l}_{n}$. Thus $\bar{q}_{i}$ is indeed a continuous extension of the vector $\bar{q}_{i}$. This feature of extendability and easy evaluation at arbitrary points is one of the most powerful tools of kernel-based methods.
a) RKHS based spectral filtering: The following procedure was described in [11, Algorithm 1], and accepts as parameters \( \epsilon_1, \epsilon_2 > 0 \) and integer \( L_0 > 1 \). Let \( F_N \) denote the discrete Fourier transform on vectors of length \( N \). Let \( [\Phi] \) be the \( N \times L \) matrix whose \( l \)-th column is \( \phi_l^\top \). Set \( \Lambda := \text{diag}(\lambda_1, \ldots, \lambda_L) \) and compute
\[
[\hat{\Phi}] := F[\Phi], \quad [H] := [\hat{\Phi}] \Lambda^{-0.5}.
\]
Next, compute the RKHS-norms as
\[
[W]_{n,1} := |[H]_{n,1}|, \quad [W]_{n,l+1} := [W]_{n,l} + |[H]_{n,l+1}|, \quad l = 1, \ldots, L - 1.
\]
Now let \( J = 1, \ldots, N \). Discard all the \( j \in J \) for which \( |W|_{j,L_0} < \epsilon_1 \). Of the remaining \( j \in J \), discard those \( j \) for which \( \ln |W|_j,L - \ln |W|_j,L_0 > \epsilon_2 \). Compute \( \omega_j = \frac{2\pi j}{N\Delta t} \) for all of the remaining \( j \in J \). These frequencies \( \omega_1 < \omega_2 < \ldots < \omega_m \) can be interpreted to be true Koopman eigenfrequencies with substantial presence in the original data.

b) Periodic and chaotic components: The identified frequencies \( 0 = \omega_1 < \omega_2 < \ldots < \omega_m \) are by no means exhaustive, they are only a finite subset of a usually infinite set of Koopman eigenfrequencies. However, they represent those (true) frequencies that have a significant presence in the data. The threshold \( \epsilon_1 \) is meant to be a numerical implementation of frequencies being significant. We next construct the periodic component \( g_{\text{per}} : \mathbb{R} \rightarrow \mathbb{R}^k \) as
\[
g_{\text{per}}(t) := \Re \sum_{j=1}^m [A]_{j,j} e^{i\omega_j t}, \quad (12)
\]
The \( m \times k \) matrix \( [A] \) is the least-squares solution to \( \Re [F] [A] = [Y] \), where \( Y \) is the data-matrix and \( [F] \) is \( N \times m \) matrix
\[
[F]_{n,j} := (2 - \delta_{j,1}) e^{i\omega_j}, \quad 1 \leq n \leq N, \quad 1 \leq j \leq m.
\]
Next, we construct the chaotic component
\[
g_{\text{chaos}} : \mathbb{R}^{k(Q+1)} \rightarrow \mathbb{R}^k, \quad g_{\text{chaos}} := \sum_{l=1}^L [E]_{l,:} \tilde{\phi}_l, \quad (13)
\]
where \( [Y_{\text{non}}] := [Y] - \Re [F] [A] \) and \( [E] := [\Phi]^\top [Y_{\text{non}}] \).

c) The reconstruction: We avoid the task of identifying a set of generating frequencies by directly using the selected frequencies in the approximation
\[
g_{\text{per}}(\tilde{\theta} + n\tilde{\omega}) \approx \sum_{j=1}^m a_j \exp(in\omega_j), \quad n = 0, 1, 2, \ldots \quad (14)
\]
Using this simplification in (14), and the formulas in (13) and (12), we create the following data-driven model of the dynamics:
\[
y_{n+1}^0 := g_{\text{per}}(n\Delta t) + g_{\text{chaos}}(y_n^0, \ldots, y_n^Q) \quad y_{n+1}^0 := y_n^0 \quad \cdots \quad y_{n+1}^Q := y_n^{Q-1} \quad (15)
\]
Here each \( y_n^i \in \mathbb{R}^k \), thus making the state vector \( y_n = (y_n^0, \ldots, y_n^Q) \) a vector in \( \mathbb{R}^{k(Q+1)} \). We have thus created a standalone dynamical system \( \mathbb{R}^{k(Q+1)} \) which is conjugate to the latent dynamics.

IV. CASE STUDY: SIGNALIZED INTERSECTION CORRIDOR

A. Data Description

The case study analyzes queue length measurements from nine adaptive traffic signals located on the Alafaya Trail (SR-434) in East Orlando, FL. The obtained data includes the details of each movement with the time, duration, queue length, and waiting time. It provides information on eight movements: north left (NL), north through (NT), south left (SL), south through (ST), east left (EL), east through (ET), west left (WL), and west through (WT). In this study, we focus on the queue length formation of northbound through movements. The raw data was processed and calibrated by Rahman, and et al. [20] and was resampled at regular intervals of \( \Delta t = 2 \) minutes. This work uses the processed data from [20].

![Northbound queue length formation at the corridor at \( t \)th time](image)

Fig. 1. Northbound queue length formation at the corridor at \( t \)th time

B. Problem Formulation

We formulate the signalized intersection corridor as a quasiperiodically driven dynamical system. The underlying dynamics of the system are high dimensional, and its governing equations are unknown. We hypothesize that the signalized intersection corridor system obeys a dynamics of the form \( \dot{Y} = F(Y) \), and the observed queue lengths are generated through some measurement function \( \hat{Y} \), as described in Section II-A. Note that the measurement \( \hat{Y} \) is not necessarily one-to-one. In particular, it may not be possible to connect the \( y_i \) with a dynamical rule of the form \( y_{i+1} = F(y_i) \). Rather \( y_i \) should be interpreted as a partial observation of the true state in \( \Omega \), in the \( i \)-th time frame. In Figure 1 we show the visual representation of \( y_i \). We use only this data to obtain a parameter-free reconstruction of the dynamical system.

C. Computation

We arrange the data in the form of a matrix \( [Y] \) with \( k = 9 \) columns. Each column corresponds to the traffic queue length.
as a function of time, at one among 9 intersections along Alfaya Trail. We used a bandwidth parameter of $\epsilon = 0.1$ for the Gaussian kernel. We compute a total of $L = 1001$ eigenfunctions for the bistochastic kernel. We set $\epsilon_1 = 0.1$ and then choose the parameters $L_0 = 5$ and $\epsilon_2 = 2.5$ using the heuristic approach shown in Figure 2. Lower the value of $\epsilon_2$, more frequencies get filtered out and higher the probability of the identification being correct. The two thresholds $\epsilon_1$ and $\epsilon_2$ are based on the asymptotic behavior in two different directions [11, Theorem 1, 4]. Combined, they provide a surer guarantee of identification of true eigenfrequencies and the discarding of spurious eigenvalues or pseudo-spectrum.

![Fig. 2. Choice of the thresholds $L_0, \epsilon_2$ for the traffic intersection data. $L_0$ was chosen to be 100, as it corresponds to the approximate location of an abrupt jump in the curve in the left panel. Given this choice of $L_0$, we plotted the quantity $\ln(\langle W(n, 0) \rangle/W(n, L_0))$ as a function of $n \in 0, \ldots, N/2$. We choose $\epsilon_2$ to be 2.5, which is approximately the first two of the points of inflection of the curve in the red panel, as indicated by the horizontal red and blue lines.]

\[ \frac{1}{N} \sum_{n=0}^{N-1} \ln W(n, 0) > \epsilon_1 \]

\[ \ln(W(n, l)/W(n, L_0)) \]

Fig. 2. Choice of the thresholds $L_0, \epsilon_2$ for the traffic intersection data. $L_0$ was chosen to be 100, as it corresponds to the approximate location of an abrupt jump in the curve in the left panel. Given this choice of $L_0$, we plotted the quantity $\ln(W(n, l)/W(n, L_0))$ as a function of $n \in 0, \ldots, N/2$. We choose $\epsilon_2$ to be 2.5, which is approximately the first two of the points of inflection of the curve in the red panel, as indicated by the horizontal red and blue lines.

D. Results

1) Generating Frequencies: This work analyzed 20,000 snapshots of queue lengths for the nine intersections to extract Koopman frequencies, i.e., generating frequencies of the quasiperiodic driving source of the dynamics. Figure 3 exhibits the spectrum of Koopman eigenfrequencies. The $x$ axis denotes period indices corresponding to the frequencies. We separate longer periods from the shorter ones and present them in two different panels for convenience of presentation. For both the panels, the $y$-axis shows $\{\ln(W(n, l)/W(n, L_0))\}$ which we have denoted as amplitude for the selected period indices $l$. Figure 3 shows that dominant periods are clustered around 1 hour, 2 hours, 3 hours, 6 hours, 12 hours, 14 hours, 3.5 days, 7 days, and 14 days. The identified periods correspond to the natural periods of the system. These periods are consistent with the results in [21]. The authors decomposed speed measurements from an intersection corridor via multiscale multifractal analysis (MMA). They reported that the dominant periodicities on weekdays are 7 days, 24, 12, 8, 6, and 3 hours while on weekends are 12 and 24 hours. In the present work, we did not differentiate weekend data from weekday data. However, we took a different set of observables (i.e., northbound through queue length data) and a completely different intersection corridor system, our identified quasiperiodic frequencies matched with that of [21]. This finding corroborates that the proposed technique can successfully identify generating frequencies of the quasiperiodic driving force of the intersection dynamical systems.

2) Reconstruction and Prediction: We reconstruct the original data from the decomposed parts, which is shown in Figure 4. The red curve is the output obtained from the reconstruction at each intersection. The curves closely following each other, including the moments when fluctuations occur. Although reconstructed dynamical models differ from the true system, this difference is inevitable in a learning problem. However, the reconstruction of $h_{\text{per}}$ is bounded, which guarantees that the dynamics under (3) would remain bounded, and the deviation of the trajectories also remain bounded. In Figure 5 we illustrate the error in reconstruction and prediction by computing the normalized relative error

\[ \text{error}_{rel}(i) := \frac{|y_n - \hat{y}_n|}{\max_{i} |y_n|}, i = 1, \ldots 9. \]  

Here $\hat{y}_n(i)$ denotes the output of the reconstructed system (15).

V. Conclusion

This work developed a data-driven framework for modeling quasi-periodic dynamical systems using Koopman theoretic approach. The proposed approach can handle dynamics with strong nonlinear and chaotic components. Thus, it is applicable across domains. We performed a case study using queue length data on a corridor of nine signalized intersections. The proposed approach accurately identified the generating
Fig. 5. Dynamical reconstruction and relative prediction error of NB queue length data. Normalized error for intersections in $n^{th}$ snapshot are indicated as $e^1(n), \ldots, e^9(n)$. The top two plots shows a sudden and qualitative difference in the error beyond the training period of $n = 20000$. The bottom two plots focus on the prediction performance beyond the training period. The system with the state initialized with the state $y_{t-20000}$ and iterated over the time window $n = [20000, 26000]$. An increased size of the moving average window reduces the effect of random outliers or events and diminishes the prediction error. Number of snapshots used to calculate moving averages in each cases is denoted as M.A and shown in each plot.

frequencies and the results for reconstruction and prediction are encouraging. The long-term prediction error remained bounded without exogenous inputs, unlike recurrent neural network-based methods such as long short-term memory (LSTM). Moreover, in comparison to deep NNs, the proposed technique is not a black-box approach. All these advantages make it a promising candidate for future research.

ACKNOWLEDGEMENTS

Authors would like to thank Dr. Samiul Hasan and his research group for providing access to the intersection dataset.

REFERENCES

[1] S. Das and J. Yorke, “Super convergence of ergodic averages for quasiperiodic orbits,” Nonlinearity, vol. 31, p. 391, 2018. DOI: 10.1088/1361-6544/aa99a6

[2] R. Vautard and M. Ghil, “Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series,” Phys. D, vol. 35, pp. 395–424, 1989. DOI: 10.1016/0167-2789(89)90077-8

[3] J. Slawinska and D. Giannakis, “Spatiotemporal pattern extraction with data-driven Koopman operators for convectively coupled equatorial waves,” in Proceedings of the 6th International Workshop on Climate Informatics, A. Banerjee, W. Ding, J. Dy, V. Lyubchich, and A. Rhines, Eds., Boulder, Colorado, 2016, pp. 49–52. DOI: 10.5065/D6K072N6

[4] D. Giannakis and S. Das, “Extraction and prediction of coherent patterns in incompressible flows through space-time Koopman analysis,” Phys. D, vol. 402, p. 132, 2021. DOI: 10.1016/j.physd.2019.132211

[5] S. Mustavee, S. Agarwal, C. Enyioha, and S. Das, “A linear dynamical perspective on epidemiology: Interplay between early covid-19 outbreak and human mobility,” arXiv preprint arXiv:2107.07380, 2021.

[6] N. Marrouch, J. Slawinska, D. Giannakis, and H. L. Read, “Data-driven koopman operator approach for computational neuroscience,” Annals of Mathematics and Artificial Intelligence, vol. 88, no. 11, pp. 1155–1173, 2020.

[7] H. Arbabi and I. Mezić, “Study of dynamics in post-transient flows using koopman mode decomposition,” Physical Review Fluids, vol. 2, no. 12, p. 124, 2017.

[8] K. R. Shabab, S. Mustavee, S. Agarwal, M. H. Zaki, and S. Das, “Exploring dmd-type algorithms for modeling signalised intersections,” arXiv preprint arXiv:2107.06369, 2021.

[9] E. Yeung, S. Kundu, and N. Hodas, “Learning deep neural network representations for koopman operators of nonlinear dynamical systems,” in 2019 American Control Conference (ACC). IEEE, 2019, pp. 4832–4839. DOI: 10.23919/ACC.2019.8815339

[10] L. Gonon and J. Ortega, “Reservoir computing universality with stochastic inputs,” IEEE transactions on neural networks and learning systems, vol. 31, no. 1, pp. 100–112, 2019. DOI: 10.1109/TNNLS.2019.2899649

[11] S. Das and D. Giannakis, “Koopman spectra in reproducing kernel Hilbert spaces,” Appl. Comput. Harmon. Anal., vol. 49, pp. 573–607, 2020. DOI: 10.1016/j.acha.2020.05.008

[12] M. Herman, Mesure de Lebesgue et nombre de rotation. Springer, 1979, vol. 597, pp. 271–293.

[13] V. Arnold, “Small denominators. i. mapping of the circumference onto itself,” Amer. Math. Soc. Transl. (2), vol. 46, pp. 213–284, 1965.

[14] S. Das, D. Giannakis, and E. Szekely. An information-geometric approach for feature extraction in ergodic dynamical systems, 2020. arXiv:stat/2004.02172 [Online]. Available: https://arxiv.org/pdf/2004.02172.pdf

[15] T. Berry and T. Sauer, “Density estimation on manifolds with boundary,” Comput. Statist. Data Anal., vol. 107, pp. 1–17, 2017. DOI: 10.1016/j.csda.2016.09.011

[16] T. Berry, R. Cressman, Z. Gregurić-Ferenček, and T. Sauer, “Time-scale separation from diffusion-mapped delay coordinates,” SIAM J. Appl. Dyn. Sys., vol. 12, pp. 618–649, 2013. DOI: 10.1137/12088183X

[17] S. Das and D. Giannakis, “Delay-coordinate maps and the spectra of Koopman operators,” J. Stat. Phys., vol. 175, no. 1–2, pp. 1107–1145, 2019. DOI: 10.1007/s10955-019-02272-w

[18] T. Sauer, J. A. Yorke, and M. Casdagli, “Embedology,” J. Stat. Phys., vol. 65, no. 3–4, pp. 579–616, 1991. DOI: 10.1007/bf01053745

[19] D. Giannakis, S. Das, and J. Slawinska, “Reproducing kernel Hilbert space compactification of unitary evolution groups,” Appl. Comput. Harmon. Anal., vol. 54, pp. 75–136, 2021. DOI: 10.1016/j.acha.2021.02.004

[20] R. Rahman and S. Hasan, “Real-time signal queue length prediction using long short-term memory neural network,” Neural Computing and Applications, vol. 33, no. 8, pp. 3311–3324, 2021.

[21] J. Wang, P. Shang, and X. Cui, “Multiscale multifractal analysis of traffic signals to uncover richer structures,” Physical Review E, vol. 89, no. 3, p. 032, 2014.