Approximate Projection-Based Control of Networks

Max Z. Li, Karthik Gopalakrishnan, Hamsa Balakrishnan

***Abstract*— Modern infrastructures such as transportation and communication networks are large-scale systems with complex dependence structures between various sub-systems. Human-interpretable performance targets in such systems are often represented in terms of lower-dimensional projections of the high-dimensional state space. We consider the problem of designing control strategies for high-dimensional systems that lack a detailed model. To do so, we leverage the ability of *copulas* to represent dependant structures in high-dimensional data, and approximate the state space model through inverse sampling. We demonstrate the applicability of the control policies obtained from our methodology through a data-driven case study of controlling flight delays within the US air transportation network.**

I. Introduction

Many systems in the transportation, energy, robotics, and communication network domains have state spaces that are extremely high-dimensional and contain intricate dependencies. In other words, the system state can be represented by $\mathbf{x} \in \mathbb{R}^{N \times N}$, where N represents the number of nodes in the system, and contains possibly nonlinear dependencies among the elements of $\mathbf{x}$. Large-scale networked systems therefore yield high-dimensional state space representations, where each element of the state vector denotes the signal generated at a particular node.

Systems with high-dimensional state spaces have a few unique challenges associated with them. First, it is difficult to develop accurate, high-fidelity models for such large-scale complex systems, especially when there are no well-understood physical mechanisms governing the system interactions. Second, these systems typically operate under extremely noisy conditions requiring vast quantities of historical data to develop data-driven models. Additionally, these data-driven models are typically not interpretable, further reducing the confidence in the predictive performance of these models. Lastly, it may not be practically feasible to control every element in such a large-scale system. Thus, system operators may only be able to prescribe *high-level* requirements, or set system performance targets in some lower-dimensional projection of the system state. This requires the development of unique controllers for each system depending on the nature of the lower-dimensional performance target specification.

This work was partially supported by NSF CPS Award No. 1739505. Max Li was also supported by an NSF Graduate Research Fellowship.

MZL, KG, and HB are with the Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. {maxli, karthikg, hamsa}@mit.edu

A. Motivation and problem description

The focus of this paper is to present a strategy to analyze and control a high-dimensional (networked) system such that performance targets in a low-dimensional projected space are satisfied. For known system dynamics and performance requirements of certain parametric forms (e.g., quadratic costs), the problem may be analytically solvable (e.g., through LQR and LQG controllers). However, if the performance target is of the form $h(\mathbf{x}) \in \mathcal{F}$, where $\mathcal{F}$ is some feasible set, then the optimization problem is significantly more challenging, and often analytically intractable. The following example illustrates the need for lower-dimensional performance targets:

Example 1 (Flight delays in airport networks)**.** *Consider the network with airports as nodes, the delays at which we wish to model and control. Interdependencies between different airports due to flight-propagated delays, weather correlations, and passenger connectivity are complex to model. Consequently, it is difficult to develop a precise and accurate model for delays at each of the airports. Furthermore, system operators often only consider aggregate performance targets such as the total delay (i.e., the sum of all airport delays), the spatial distribution of delays across the network, or the variance of airport delays. When the system is disrupted, the system operator may want to drive the system to some desired configuration that satisfies aggregate performance targets.*

As a system expands in scale and complexity, its dynamics become more difficult to model. Consequently, an approach based on a lower-dimensional performance measure may be more interpretable and implementable in practice.

B. Our approach

Our proposed solution consists of two main components, namely, (1) generate a set of feasible high-dimensional states, and (2) map a desired trajectory satisfying performance targets in a lower-dimensional projection back to the high-dimensional state space. Figure 1 presents a flowchart depicting our framework. The first step uses a limited number of state observations and a multivariate Gaussian copula to represent the set of feasible high-dimensional system states. Copulas provide a way of relating the multivariate joint distributions to the univariate marginal distributions [1], and hence enable us to separately estimate the dependant structures in the multivariate distribution of the system state. The second step of our framework ensures trajectory conformance to lower-dimensional performance targets, and identifies the corresponding high-dimensional trajectory via an optimization subroutine. Our framework can accommodate a large class of lower-dimensional projections and is

flexible in terms of the objective function and cost structure for the optimization subroutine. This flexibility allows our framework to be generalized to a wide range of applications.

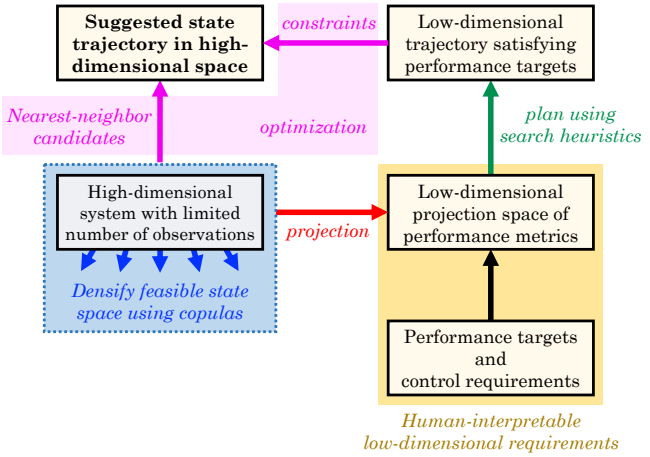

Fig. 1. Flowchart of the proposed approach.

C. Prior work

System identification tools can be used to develop models based on the observed input and output [2], [3]. However, many applications (including air traffic networks) are not amenable to these methods due to insufficient numbers of observations considering system complexity, noise, and dimensionality. Reduced-order models, which consider a lower-dimensional representation of the system, are a promising alternative to conventional system identification [4], [5]. The control of systems using reduced-order models has been widely studied [6], [7]. Although such reduced-order models could be used in the analysis of any high-dimensional system, our work focuses on using an optimization formulation to map a desired control trajectory from a low-dimensional space to a higher-dimensional state space using a multivariate Gaussian copula approximation.

Copulas provide a way to represent any multivariate distribution in terms of univariate marginal distributions of variables and the dependence structure (or copula) [8]. They have been widely-used in finance, e.g., risk management and option pricing [9], [10], regression estimation and inference [11], information retrieval [12], and generating latent space models [13]. The use of copulas for representing the states of a dynamical systems is a more recent line of work. Methods to construct state spaces from copulas [14] and to model parametric state space models (e.g., linear Gaussian) using inversion copulas [15] have been explored. In contrast to these prior works, our approach is non-parametric and uses a small set of historical observations to identify an appropriate copula that represents the high-dimensional state space.

The control of delays in airport networks has been a longstanding problem [16]. The use of low-dimensional, human-interpretable performance measures (e.g., total delay, spatial distribution of delays, see [17]) in a large-scale system makes this an ideal setting for our methodology.

D. Contributions

Our main contributions are two-fold:

(1) We develop a method to identify a sequence of future states and associated control actions for high-dimensional systems based on performance targets in a lower-dimensional space (Section III).
(2) We demonstrate how our approach can be used to drive the evolution of airport delays in accordance with aggregate performance targets, and discuss how it may be generalized to other networked settings (Section IV).

II. Mathematical Preliminaries

A. Notation

In this paper, boldface characters (e.g., $\mathbf{x}$, $\mathbf{U}$) indicate vectors, while a non-boldfaced equivalent with subscript i (e.g., x_i, U_i) indicates the i^{th} element of the corresponding vector. Superscript indices in parentheses indicate the observation index. Random variables are capitalized, and their lowercase counterparts indicate a realization. The functions f and F with random variables attached in subscripts indicate the probability density function (pdf) and cumulative distribution function (cdf), respectively. The space of symmetric positive semi-definite matrices of size N is denoted $\mathbb{S}^{N\times N}_{\succeq 0}$.

Suppose that we have a data set consisting of M observations, $\mathcal{O}_M = \{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(l)}, \ldots, \mathbf{x}^{(M)}\} \subset \mathcal{X}$, where $\mathcal{X} \subset \mathbb{R}^{N\times 1}$ denotes the space of feasible states. Then, $\mathbf{x}^{(l)} \in \mathcal{O}_M$ can be viewed as a realization of a multivariate random variable $\mathbf{X} = (X_1, \ldots, X_N) \in \mathbb{R}^{N\times 1}$ with cdf $F_{\mathbf{X}}(\mathbf{x}) = \mathbb{P}(X_1 \leq x_1, \ldots, X_N \leq x_N)$.

B. Representing multivariate cdfs using copulas

We provide the formal definition of a copula, and state Sklar's Theorem, which says that any multivariate cdf can be expressed by its marginals and an appropriate copula.

Definition 1 (N-dimensional copula)**.** *An N-dimensional copula, denoted by $C : [0,1]^N \to [0,1]$, is a cdf $C(\mathbf{u}) = C(u_1, \ldots, u_N)$ describing a random variable $\mathbf{U} \in [0,1]^N$. $C(\mathbf{u})$ is a copula if, and only if, it satisfies the following properties:*

(i) $C(\mathbf{u})$ is non-decreasing in each component u_i.
(ii) The i^{th} marginal distribution of $C(\mathbf{u})$, obtained by setting $u_j = 1, \forall j \neq i$, is equal to the cdf of a standard uniform random variable.
(iii) For scalars $a_i \leq b_i$, we have that

$$\mathbb{P}\left(\bigcap_{i=1}^{N} a_i \leq U_i \leq b_i\right) = C(b_1, \ldots, b_N) - C(a_1, \ldots, a_N).$$

Theorem 1 (Sklar's Theorem [8])**.** *Consider a N-dimensional cdf $F_{\mathbf{X}}$ with marginals $F_{X_1}, \ldots, F_{X_N}$. Then, there exists a copula C such that $F_{\mathbf{X}}(x_1, \ldots, x_N) = C(F_{X_1}(x_1), \ldots, F_{X_N}(x_N))$ for all $x_i \in \mathbb{R} \cup \{-\infty, \infty\}$ and $i = 1, \ldots, N$. Furthermore, if F_{X_i} is continuous for all $i = 1, \ldots, N$, then the copula C is unique.*

Sklar's theorem guarantees the unique existence of a copula corresponding to any continuous multivariate distribution.

It forms the basis of several algorithms for estimating the copula from data. This process typically uses parametric or non-parametric techniques to sequentially estimate the marginal distributions and the copula function.

C. State space description of a system

Given a networked system, we abstract the N signal-generating members of the system as vertices (the set V), and encode relationships between members i and j as weighted, undirected edges $(i,j) = (j,i) \in E \subseteq V \times V$. The vertices and edges constitute a graph $G = (V, E)$, and a signal supported on vertex i is $x_i \in \mathbb{R}$. We denote observation l of the vector of signals generated from this system, i.e., its state, by $\mathbf{x}^{(l)} \in \mathbb{R}^{N\times 1}$, with $\mathbf{x}^{(l)} = \left(x_1^{(l)}, \ldots, x_N^{(l)}\right)^\intercal$. If $\mathbf{x}^{(l)}$ was observed historically, then $\mathbf{x}^{(l)} \in \mathcal{X}$. We assume no other information about $\mathcal{X}$, apart from the historical observations.

III. Methodology

A. Estimating multivariate Gaussian copulas from data

We use a kernel density estimator to construct continuous marginal distributions, and find the copula density $c(\mathbf{u})$ by taking the appropriate partial derivatives of $C(\mathbf{u})$:

$$c(u_1, \ldots, u_N) = \frac{\partial^N}{\partial u_1 \cdots \partial u_N} C(u_1, \ldots, u_N). \tag{1}$$

From Sklar's Theorem, we can rewrite the unknown pdf $f_{\mathbf{X}}$ as the product of the copula density in (1) and the marginal density functions f_{X_i}, as shown in (2). We then estimate $f_{\mathbf{X}}$ through maximum likelihood estimation using $\mathcal{O}_M$.

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}) &= \frac{\partial^N}{\partial x_1 \cdots \partial x_N} F_{\mathbf{X}}(x_1, \ldots, x_N) \\ &= \frac{\partial^N}{\partial u_1 \cdots \partial u_N} C\left(\underbrace{F_{X_1}(x_1)}_{u_1}, \ldots, \underbrace{F_{X_N}(x_N)}_{u_N}\right) \\ &\quad \times \frac{\partial F_{X_1}(x_1)}{\partial x_1} \cdots \frac{\partial F_{X_N}(x_N)}{\partial x_N} \\ &= c(\mathbf{u}) \prod_{i=1}^{N} f_{X_i}(x_i). \end{aligned} \tag{2}$$

Suppose the copula is parameterized by $\boldsymbol{\Theta}$, denoted by $c(\mathbf{u}; \boldsymbol{\Theta})$. Let $\ell(\boldsymbol{\Theta}; \mathcal{O}_M)$ denote the log-likelihood of (2) with respect to the copula parameter, $\boldsymbol{\Theta}$, and historical state observations, $\mathcal{O}_M$. We have that

$$\begin{aligned} \ell(\boldsymbol{\Theta}; \mathcal{O}_M) &= \sum_{k=1}^{M} \ln c\left(F_{X_1}(x_1^{(k)}), \ldots, F_{X_N}(x_N^{(k)}); \boldsymbol{\Theta}\right) \\ &\quad + \sum_{k=1}^{M}\sum_{i=1}^{N} \ln f_{X_i}\left(x_i^{(k)}\right). \end{aligned} \tag{3}$$

Note that the log-likelihood is only over the copula parameter $\boldsymbol{\Theta}$. Since we have no information regarding parameterizations of marginal densities or cdf, we utilize the *canonical maximum likelihood* (CML), where empirical marginal distributions are first estimated based on $\mathcal{O}_M$, and these empirical marginal distributions are used to transform observations in $\mathcal{O}_M$ to uniform variates via the probability integral transform. Under CML estimation, (3) reduces to

$$\widehat{\boldsymbol{\Theta}} = \underset{\boldsymbol{\Theta}\in\mathcal{M}}{\operatorname{argmax}} \sum_{k=1}^{M} \ln c\left(\widehat{u_1}^{(k)}, \ldots, \widehat{u_N}^{(k)}; \boldsymbol{\Theta}\right), \tag{4}$$

where $\mathcal{M}$ is the space of copula parameters, and $\widehat{u_i}^{(k)} = \widehat{F}_{X_i}\left(x_i^{(k)}\right)$ is computed as the probability integral transform with an empirical estimate $\widehat{F}_{X_i}$ at each vertex i, given by the kernel density estimator in (5) with bandwidth h and the standard normal density function $\phi(t)$ as the smoothing kernel. Therefore,

$$\widehat{F}_{X_i}(x_i) = \frac{1}{M} \sum_{k=1}^{M} \left\{ \int_{-\infty}^{\frac{x_i - x_i^{(k)}}{h}} \phi(t)\, dt \right\}. \tag{5}$$

Now that we are able to consider a wide range of empirical marginal distributions, we impose a dependence structure by choosing a family of copulas over which we can carry out the CML estimation in (4). There are a variety of parametric copula families for bivariate distributions [1]; for multivariate distributions, the most flexible is the multivariate Gaussian copula, along with other possibilities such as t- and vine-based copulas [14]. We use CML to fit a multivariate Gaussian copula $C(\mathbf{u}; \boldsymbol{\rho})$ with the cdf and density $c(\mathbf{u}; \boldsymbol{\rho})$:

$$\begin{aligned} C(\mathbf{u}; \boldsymbol{\rho}) &= \Phi_{\boldsymbol{\rho}}\left(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_N)\right), \\ c(\mathbf{u}; \boldsymbol{\rho}) &= \det(\boldsymbol{\rho})^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\boldsymbol{\Xi}^\intercal\left(\boldsymbol{\rho}^{-1} - I_{N\times N}\right)\boldsymbol{\Xi}\right), \end{aligned}$$

where $\Phi_{\boldsymbol{\rho}}$ is a standardized multivariate normal distribution with correlation matrix $\boldsymbol{\rho} \in \mathbb{S}_{\succeq 0}^{N\times N}$ having unit diagonals, Φ^{-1} is the inverse cdf for a standard normal distribution, $\boldsymbol{\Xi} = \left(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_N)\right)^\intercal$, and $I_{N\times N}$ is the $N \times N$ identity matrix. Given our choice of the multivariate Gaussian copula, $\boldsymbol{\Theta} = \boldsymbol{\rho}$ and $\mathcal{M} = \mathbb{S}_{\succeq 0}^{N\times N}$, the CML estimation problem in (4) becomes

$$\widehat{\boldsymbol{\rho}} = \underset{\boldsymbol{\rho}\in\mathbb{S}_{\succeq 0}^{N\times N}}{\operatorname{argmax}} \left\{ -\frac{M}{2}\ln\det(\boldsymbol{\rho}) - \frac{1}{2}\sum_{k=1}^{M} \widehat{\boldsymbol{\Xi}}^{(k)\intercal}\widetilde{\boldsymbol{\rho}}\,\widehat{\boldsymbol{\Xi}}^{(k)} \right\}, \tag{6}$$

over valid correlation matrices, with $\widetilde{\boldsymbol{\rho}} = \boldsymbol{\rho}^{-1} - I_{N\times N}$ and $\widehat{\boldsymbol{\Xi}}^{(k)} = \left(\Phi^{-1}\left(\widehat{u_1}^{(k)}\right), \ldots, \Phi^{-1}\left(\widehat{u_N}^{(k)}\right)\right)^\intercal$.

After obtaining $\widehat{\boldsymbol{\rho}}$ from (6), we draw $\widetilde{M}$ samples residing in $[0,1]^N$ from $C(\mathbf{u}; \widehat{\boldsymbol{\rho}})$, where $\widetilde{M} \gg M$. We refer to these samples $\left\{\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(\widetilde{M})}\right\}$ as *simulated observations* drawn from the fitted multivariate Gaussian copula $C(\mathbf{u}; \widehat{\boldsymbol{\rho}})$. We transform these simulated observations back in conformance with $\mathcal{O}_M$ via the inverse probability integral transform, through the inverse of the empirical marginal distributions found via (5). With a slight overload on notation, we denote these transformed simulated observations as *copula-simulated state observations*, and define the set $\widehat{\mathcal{X}} \coloneqq \left\{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(\widetilde{M})}\right\}$ as the *approximate* space of feasible states. To avoid confusion, we denote the l^{th} historical state observations in $\mathcal{O}_M$ as $\mathbf{x}_{\mathcal{O}_M}^{(l)}$, in contrast to $\mathbf{x}^{(l)}$, the l^{th} copula-simulated state observation from $\widehat{\mathcal{X}}$.

B. Projection-based control

Recall that the objective of our projection-based control is to drive the system from some currently observed state $\mathbf{x}^{(0)}_{\mathcal{O}_M} \in \mathcal{X}$ through a sequence of desired intermediate states. Throughout the sequence, we only dictate performance targets in a lower-dimensional space. In our case, we chose to use a $\mathbb{R}^2$-projection of $\mathcal{X} \cup \widehat{\mathcal{X}}$ spanned by the 1-norm of the state vector $\|\mathbf{x}\|_1$ and its *total variation* (TV) with respect to the graphical system abstraction G.

Definition 2. *The total variation of the state vector $\mathbf{x}$ supported on the vertices of a graph with a weighted adjacency matrix $A = [a_{ij}]$, corresponding degree matrix $D = [d_{ij}]$ with $d_{ii} = \sum_j a_{ij}$ and 0 otherwise, and corresponding combinatorial graph Laplacian $\mathcal{L} = D - A$, is given by*

$$\mathrm{TV}(\mathbf{x}) = \frac{1}{2}\sum_{i \neq j} a_{ij}\left(x_i - x_j\right)^2 = \mathbf{x}^\intercal \mathcal{L}\mathbf{x}. \tag{7}$$

The choice of $\|\mathbf{x}\|_1$ is motivated by the fact that in positive signal-generating systems, this metric captures the total magnitude of signals across the entire system (e.g., the total delay in an airport network). The choice of $\mathrm{TV}(\mathbf{x}) = \mathbf{x}^\intercal\mathcal{L}\mathbf{x}$ reflects the fact that TV can be interpreted as a measure of signal smoothness, and has been used for outlier detection in graph signals [17]. Our projection-based control framework is agnostic to the specific choice of metrics; any reasonable set of low-dimensional metrics that captures key system performance characteristics may be used.

We define the projection $\mathrm{proj}_{\mathbb{R}^2} : \mathbb{R}^{N\times 1} \to \mathbb{R}^{2\times 1}$ that maps $\mathbf{x} \in \mathcal{X} \cup \widehat{\mathcal{X}}$ to $\left(\|\mathbf{x}\|_1, \sqrt{\mathrm{TV}(\mathbf{x})}\right)$. Note that the square root on TV ensures comparable units between the 1-norm and TV, which is a quadratic form in $\mathbf{x}$. The use of $\sqrt{\mathrm{TV}(\mathbf{x})}$ simplifies the task of defining geometric constraints in $\mathrm{im}\left(\mathrm{proj}_{\mathbb{R}^2}\mathcal{X} \cup \widehat{\mathcal{X}}\right)$. Let $\mathbf{x}^{(0)}_{\mathcal{O}_M}$ be the initial state of the system that has been observed, i.e., we have full knowledge of all signals $x_i^{(0)}$. We can compute its projection in our low-dimensional space as

$$\mathrm{proj}_{\mathbb{R}^2}\left(\mathbf{x}^{(0)}_{\mathcal{O}_M}\right) = \left(\left\|\mathbf{x}^{(0)}_{\mathcal{O}_M}\right\|_1, \sqrt{\mathrm{TV}\left(\mathbf{x}^{(0)}_{\mathcal{O}_M}\right)}\right). \tag{8}$$

Recall that our goal is to drive the system from $\mathbf{x}^{(0)}_{\mathcal{O}_M}$ to some unknown terminal state $\mathbf{x}^{(T)}$ via unknown intermediate states $\mathbf{x}^{(t)}$ in discrete time steps $t = 1, \ldots, T-1$, by only constraining the $\mathbb{R}^2$-projected system metrics $\mathrm{proj}_{\mathbb{R}^2}\left(\mathbf{x}^{(1)}\right), \ldots, \mathrm{proj}_{\mathbb{R}^2}\left(\mathbf{x}^{(T)}\right)$ to preset performance targets. In other words, we do not specify any entries $x_i^{(t)}$ in $\mathbf{x}^{(t)}, \forall t = 1, \ldots, T$, and instead consider all *candidate* copula-simulated state observations $\mathbf{x}^{(t)}_{\text{candidate}} \in \widehat{\mathcal{X}}$ that satisfy some performance target in the $\mathbb{R}^2$-projected space, i.e., some constraint on $\mathrm{proj}_{\mathbb{R}^2}\left(\mathbf{x}^{(t)}_{\text{candidate}}\right)$ and $\mathrm{proj}_{\mathbb{R}^2}\left(\mathbf{x}^{(t)}\right)$.

Let us consider an example of a geometric constraint on $\mathrm{im}\left(\mathrm{proj}_{\mathbb{R}^2}\mathcal{X} \cup \widehat{\mathcal{X}}\right)$ that could be used to ensure a specific type of conformance to performance targets in the $\mathbb{R}^2$-projected space. We first select performance target *anchors* $\mathbf{a}(t) \in \mathrm{im}\left(\mathrm{proj}_{\mathbb{R}^2}\widehat{\mathcal{X}}\right), \forall t = 1, \ldots, T$. We assume that, via the copula-based state space estimation described earlier in Section III-A, the number of copula-simulated state observations $\widetilde{M}$ was large enough such that $\mathrm{im}\left(\mathrm{proj}_{\mathbb{R}^2}\widehat{\mathcal{X}}\right)$ is dense around anchors $\mathbf{a}(t)$. Formally, let $\mathcal{B}\left((x_1, x_2); \varepsilon\right)$ be a ball in $\mathbb{R}^2$ with radius ε centered at (x_1, x_2). Then, for a small positive radius $\varepsilon > 0$, there exists a $\widetilde{M}_\varepsilon = \left|\widehat{\mathcal{X}}\right|$ such that

$$\mathcal{B}\left(\mathbf{a}(t); \varepsilon\right) \cap \mathrm{im}\left(\mathrm{proj}_{\mathbb{R}^2}\widehat{\mathcal{X}}\right) \neq \{\varnothing\}. \tag{9}$$

The anchors $\mathbf{a}(t)$ are the prescribed system performance targets $\mathrm{proj}_{\mathbb{R}^2}\left(\mathbf{x}^{(t)}\right)$ that candidate copula-simulated states $\mathbf{x}^{(t)}_{\text{candidate}} \in \widehat{\mathcal{X}}$ must adhere to at each corresponding time step t. We then solve the optimization problem in (10), where the geometric constraint encodes our requirement that the candidate copula-simulated states do not deviate more than $\delta_t > 0$ in Euclidean norm from anchor $\mathbf{a}(t)$ at time step t.

$$\begin{aligned} \mathbf{x}^{(t)}_* : \quad & \underset{\mathbf{x}^{(t)} \in \widehat{\mathcal{X}}}{\text{argmin}} \quad \left\|\mathbf{x}^{(t)} - \mathbf{x}^{(t-1)}_*\right\|_2 \\ & \text{s. t.} \quad \delta_t \geq \left\|\mathbf{a}(t) - \mathrm{proj}_{\mathbb{R}^2}\left(\mathbf{x}^{(t)}\right)\right\|_2 \\ & \qquad \mathbf{x}^{(0)}_* = \mathbf{x}^{(0)}_{\mathcal{O}_M} \\ & \qquad \forall t = 1, \ldots, T. \end{aligned} \tag{10}$$

The objective function reflects the fact that $\mathrm{proj}_{\mathbb{R}^2}$ is surjective, i.e., multiple candidate copula-simulated states could feasibly satisfy the geometric constraint. Although we do not know the underlying state dynamics from t to $t+1$, assuming we select a reasonable time-step, we prefer small changes at each vertex signal $x_i^{(t)}$. These assumptions are reasonable in real applications: For example, if we assume that the duration between time steps is 1-hour, we should not expect large variations in the delays at a particular airport, given the persistent nature of most disruptions (e.g., a slow-moving snowstorm). The objective function in (10) retrieves the current state that needs a *minimal-energy* evolution from the fully-known previous state, $\mathbf{x}^{(t-1)}_*$. Note that $\mathbf{x}^{(t-1)}_*$ is fully known because it is either the solution to the preceding optimization, or the known initial condition $\mathbf{x}^{(0)}_* = \mathbf{x}^{(0)}_{\mathcal{O}_M}$.

We define $\mathbb{T}_* \coloneqq \left\{\mathbf{x}^{(0)}_* = \mathbf{x}^{(0)}_{\mathcal{O}_M}, \mathbf{x}^{(1)}_*, \ldots, \mathbf{x}^{(T)}_*\right\}$ as the extrapolated system state trajectory starting at $\mathbf{x}^{(0)}_* = \mathbf{x}^{(0)}_{\mathcal{O}_M}$, obtained from solving (10) at each time step $t = 1, \ldots, T$. The control policy $\boldsymbol{\pi}^{(t)}_*$ at time $t = 0, \ldots, T-1$ is simply $\boldsymbol{\pi}^{(t)}_* = \mathbf{x}^{(t+1)}_* - \mathbf{x}^{(t)}_*$, with the initial condition $\mathbf{x}^{(0)}_* = \mathbf{x}^{(0)}_{\mathcal{O}_M}$. We note that for the setup in (10), it is possible to define a feasible set on the space of control actions by observing that

$$\boldsymbol{\pi}_{\min} \preceq_{\mathbb{R}} \boldsymbol{\pi}^{(t)}_* \preceq_{\mathbb{R}} \boldsymbol{\pi}_{\max}, \tag{11}$$

where $\preceq_{\mathbb{R}}$ is the element-wise inequality, $\boldsymbol{\pi}_{\min} = (\pi_{\min,1}, \cdots, \pi_{\min,N})^\intercal$, and $\boldsymbol{\pi}_{\max} = (\pi_{\max,1}, \ldots, \pi_{\max,N})^\intercal$, with

$$\pi_{\min,i} = \inf\left\{x_i^{(\tau)} - x_i^{(\sigma)} \,\middle|\, x_i^{(\tau)}, x_i^{(\sigma)} \in \mathcal{X} \cup \widehat{\mathcal{X}}\right\},$$
$$\pi_{\max,i} = \sup\left\{x_i^{(\tau)} - x_i^{(\sigma)} \,\middle|\, x_i^{(\tau)}, x_i^{(\sigma)} \in \mathcal{X} \cup \widehat{\mathcal{X}}\right\}.$$

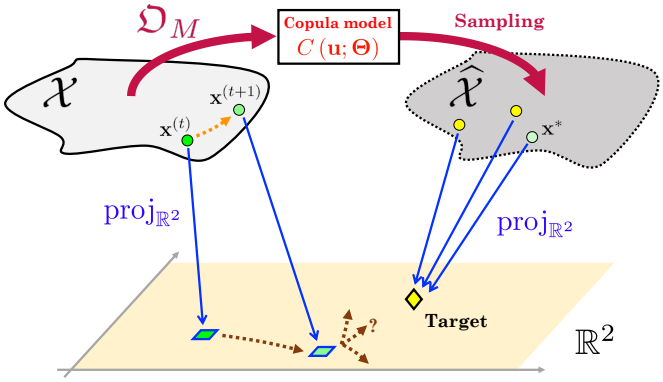

Fig. 2. Pictorial representation of our methodology.

C. Discussion

1) Interpretation and generalizability : Our framework provides a control policy for a dynamical system whose precise model is not known. Our two-step approach of first generating an approximate feasible state space, and then mapping low-dimensional performance targets and high-dimensional states through optimization is a proxy for some dynamics $\mathbf{x}^{(t)} = h\left(\mathbf{x}^{(t-1)}\right)$, where h is state-dependent. The optimization procedure in (10) can be thought of as a pseudo-inverse projection $\text{proj}_{\mathbb{R}^2}^{-1} : \mathbb{R}^{2\times 1} \to \widehat{\mathcal{X}}$ which selects a unique high-dimensional system state (since the objective function in (10) is convex) from the copula-approximated state space that satisfies performance target constraints. The proposed framework (Fig. 1) can be adapted to a range of general applications, as discussed in Section IV-C.

2) Technical remarks and caveats : When choosing the low-dimensional $\mathbb{R}^2$-projected trajectory that satisfies performance targets, we use a greedy approach at each time step in (10). We could have alternatively constructed a set of trajectory options, and then performed a breadth- or depth-first search to select one with the lowest cost. Furthermore, even though the projection operator $\text{proj}_{\mathbb{R}^2}$ we use is smooth and continuous, these are sufficient, but may not be necessary, conditions. The characterization of necessary and sufficient conditions on $\text{proj}_{\mathbb{R}^2}$ such that small perturbations in the low-dimensional projection space do not result in large perturbations in $\mathcal{X} \cup \widehat{\mathcal{X}}$ is an interesting question for future research.

It is important to bear in mind that copulas cannot provide more information than what is encoded in the historical observations. They should not be interpreted as stochastic processes. They also do not capture any time-varying information; they are merely a particular form of a multivariate probability distribution. We account for time-varying dependence structures in our case study by calibrating different copulas for different hours of the day.

IV. Airport Delay Network Example

A. Redistributing airport delays with conservation constraints

Let $\mathbb{T} = \left\{\mathbf{x}_{\mathcal{O}_M}^{(0)}, \ldots, \mathbf{x}_{\mathcal{O}_M}^{(T)}\right\} \subseteq \mathcal{O}_M$ be a sequence of observed system states that begin at time $t = 0$ and evolve until time $t = T$. In our air traffic network application, we could have a historical trajectory $\mathbb{T}$ that describes the state of airport delays from 6 a.m. until 10 a.m., in 1-hour intervals, in which case we would have $T = 4$. We can project each observed airport delay state $\mathbf{x}_{\mathcal{O}_M}^{(t)}$ onto the $\mathbb{R}^2$-space from before via $\mathbf{x}_{\mathcal{O}_M}^{(t)} \mapsto \left(\left\|\mathbf{x}_{\mathcal{O}_M}^{(t)}\right\|_1, \sqrt{\mathbf{x}_{\mathcal{O}_M}^{(t)\intercal} \mathcal{L} \mathbf{x}_{\mathcal{O}_M}^{(t)}}\right)$, and track the airport delays in terms of their total magnitude and spatial distribution across the airport network.

The proposed projection-based control framework can address the following operationally-important question: Could the airport delays be *redistributed* in a manner that preserves the total magnitude of delays (conservation of delays), but alleviates delay accumulation at specified airports? This question translates to controlling the system in the projected $\mathbb{R}^2$-space, and setting constraints on the 1-norm component. Our formulation for this problem is given by (12), where $\mathbf{x}_*^{(t)}$ is the solution to:

$$\begin{aligned} \underset{\mathbf{x}^{(t)} \in \widehat{\mathcal{X}}}{\arg\min} \quad & \left\{ \left\| \mathbf{x}^{(t)} - \mathbf{x}_*^{(t-1)} \right\|_2 + \lambda \left\| \mathbf{x}^{(t)} - \mathbf{x}_{\mathcal{O}_M}^{(t)} \right\|_2 \right. \\ & \left. + (1-\lambda) \mathbf{c}^\intercal \left(\mathbf{x}^{(t)} - \mathbf{x}_{\mathcal{O}_M}^{(t)} \right) \right\} \\ \text{s. t.} \quad & \left\| \mathbf{x}^{(t)} \right\|_1 \geq \left\| \mathbf{x}_{\mathcal{O}_M}^{(t)} \right\|_1 - \delta, \\ & \left\| \mathbf{x}^{(t)} \right\|_1 \leq \left\| \mathbf{x}_{\mathcal{O}_M}^{(t)} \right\|_1 + \delta, \\ & \mathbf{x}_*^{(0)} = \mathbf{x}_{\mathcal{O}_M}^{(0)}, \\ & \lambda \in [0,1], \\ & \forall t = 1, \ldots, T. \end{aligned} \tag{12}$$

Given the historical observation $\mathbf{x}_{\mathcal{O}_M}^{(t)}$, we search for an alternative state $\mathbf{x}_*^{(t)} \in \widehat{\mathcal{X}}$ from the copula-based sampling approximation of the feasible state space which minimizes the objective in (12). The first term of the objective function penalizes sudden transitions from the previous state, while the second term penalizes alternative states that deviate from their historical counterpart. The last term rewards delay reductions at specified airports via an appropriately-structured cost vector $\mathbf{c} \in \mathbb{R}^{N\times 1}$ given by (13), with S representing the subset of airports at which delay reductions are encouraged:

$$\mathbf{c} = [c_i] = \begin{cases} 1 & \text{if } i \in S \subseteq V \\ 0 & \text{otherwise.} \end{cases} \tag{13}$$

The λ parameter can be adjusted to weight historical conformity versus redistribution. The conservation constraint is on the 1-norm, and reflects the fact that system operators cannot simply remove delays without mechanisms such as flight cancellations.

B. Redistributing delay away from New York City airports

We apply (12) to an 11-hour long historical trajectory $\mathbb{T}_\diamond$, observed from 1 p.m. EDT on May 22 to 12 a.m. on May 23, 2014. Our networked system consists of $|V| = 30$ busiest US airports in terms of passenger enplanements. The vector of airport delays and the delay at airport i at hour t are $\mathbf{x}^{(t)}$ and $x_i^{(t)}$, respectively. The graph representation of the system is a complete, undirected correlation network, where edges

are weighted by the sample Pearson correlations between the hourly airport delays. Since airport delays are strongly dependent on the hour-of-day (e.g., because of heavy traffic during the morning and afternoon periods), we construct 24 different approximate state spaces $\widehat{\mathcal{X}}_0, \ldots, \widehat{\mathcal{X}}_\tau, \ldots, \widehat{\mathcal{X}}_{23}$, each from a copula model based on historical observations of $\mathbf{x}^{(t)}$ belonging to hour τ. Similarly, we have 24 different graph Laplacians $\mathcal{L}_0, \ldots, \mathcal{L}_\tau, \ldots, \mathcal{L}_{23}$, corresponding to the different hourly airport delay correlation networks. We use airport delay data for the 10-year period from 2008 through 2017 to construct $\widehat{\mathcal{X}}_\tau$. Therefore, $M = 3,653$ for each hour-category, τ. We sample $\widetilde{M} = 50,000 \gg M$ from each copula model in order to construct $\widehat{\mathcal{X}}_\tau$, for all $\tau = 0, \ldots, 23$.

Starting with the airport delay state vector $\mathbf{x}_*^{(0)} = \mathbf{x}_{\mathbb{T}_\diamond}^{(0)} \in \mathbb{T}_\diamond$ at 1 p.m. EDT, we solve (12) for $t = 1$ (2 p.m. EDT) through $t = 11$ (midnight EDT), with the slight modification that the minimization is performed over the appropriate $\widehat{\mathcal{X}}_\tau$ such that the hour category of τ matches that of the current time step t. We select a δ-tolerance of

$$\delta = 0.01 \times \max\left\{ \left\| \mathbf{x}_{\mathbb{T}_\diamond}^{(t)} \right\|_1 \;\middle|\; \mathbf{x}_{\mathbb{T}_\diamond}^{(t)} \in \mathbb{T}_\diamond \right\}.$$

For this example, $\delta \approx 19.6$ minutes, which is negligible in comparison to the total system delay. In essence, the delay-conservation constraints in (12) are enforced. We configure our cost vector $\mathbf{c}$ in accordance with (13) using $S = \{\text{EWR, JFK, LGA}\} \subset V$. In other words, we configure the third term in (12) to reward control policies that shift delay away from the three New York City (NYC) airports. Finally, we repeat the entire sequence of optimizations for differing $\lambda = \{0, 0.5, 1\}$. In this context, $\lambda = 0$ represents maximal efforts to shift delay away from the NYC airports with no penalties on not adhering to the historical trajectory $\mathbb{T}_\diamond$, whereas $\lambda = 1$ represents maximal adherence to the historical trajectory with no preference on redistributing delay away from the NYC airports. Decreasing λ is a proxy for increasing the workload of air traffic flow managers, since it requires greater efforts towards delay redistribution.

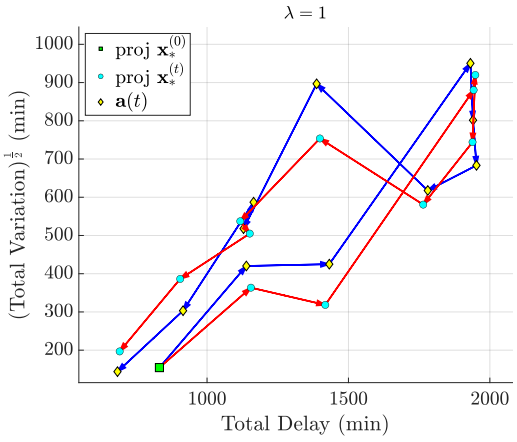

Fig. 3. $\mathbb{R}^2$-projected historically observed state trajectory (blue) and delay-conserved alternate state trajectory (red) with $\lambda = 1$ (maximal historical adherence).

We plot the $\mathbb{R}^2$-projected state trajectories in Figures 3 and 4 for $\lambda = 1$ and 0, respectively. Note that at each time step, the horizontal displacement between the $\mathbb{R}^2$-projected historical state (yellow diamonds) and $\mathbb{R}^2$-projected suggested state (cyan dots) is minimal, indicating that delay is conserved. We also observe, as expected, that for $\lambda = 1$, the suggested $\mathbb{R}^2$-projected trajectory $\mathbb{T}_*$ closely mirrors the historical $\mathbb{R}^2$-projected trajectory. On the other hand, for $\lambda = 0$, even though delay conservation is still enforced, the $\mathbb{R}^2$-projected state trajectories are quite different, indicating that delay redistribution is occurring over the set of airport vertices.

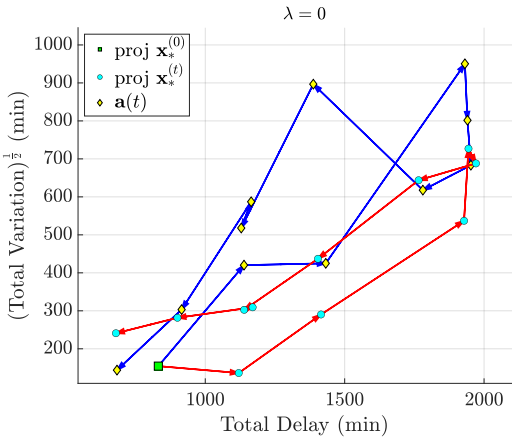

Fig. 4. $\mathbb{R}^2$-projected historically observed state trajectory (blue) and delay-conserved alternate state trajectory (red) with $\lambda = 0$ (maximal selective delay redistribution).

We plot the signals $x_{\text{EWR}}^{(t)}$, $x_{\text{JFK}}^{(t)}$, and $x_{\text{LGA}}^{(t)}$ for the entire 11-hour period, for all four state trajectories (baseline case, i.e., $\mathbf{x}_{\mathbb{T}_\diamond}^{(t)} \in \mathbb{T}_\diamond$, and $\mathbf{x}_*^{(t)} \in \mathbb{T}_*$ for $\lambda = \{1, 0.5, 0\}$) in Figure 5. For all three NYC airports (and all other airport vertices), $\lambda = 1$ strongly penalizes large values of $\left\| \mathbf{x}_*^{(t)} - \mathbf{x}_{\mathbb{T}_\diamond}^{(t)} \right\|_2$, so we observe adherence to the baseline delay values for $x_{\text{EWR}}^{(t)}$, $x_{\text{JFK}}^{(t)}$, and $x_{\text{LGA}}^{(t)}$. By setting $\lambda = 0$, the effects of selectively redistributing delay away from EWR, JFK, and LGA are very pronounced, as exhibited by much lower values for $x_{\text{EWR}}^{(t)}$, $x_{\text{JFK}}^{(t)}$, and $x_{\text{LGA}}^{(t)}$ compared to the baseline delay values. However, we know from the $\mathbb{R}^2$-projected state trajectories as well as the delay-conservation constraints in (12) that the total delay in the network must be conserved (within tolerance δ) at all time steps t. This indicates that, for $\lambda = 0$, the optimized control policy redistributed delay away from the NYC airports, and to the other airports in the network.

C. Generalizability of our framework

We conclude our case study example by highlighting five aspects of versatility provided by our framework: (1) the choice and selection of a copula family; (2) the form of the objective function; (3) the cost structure; (4) the constraints enforcing low-dimensional performance targets; and (5) the choice of projection metrics. In terms of (1), although we chose to use the family of multivariate Gaussian copulas, there have been interesting results on copula goodness-of-fit testing [18] and avoiding copula family misspecification [19]. While copula family selection is outside the scope of this

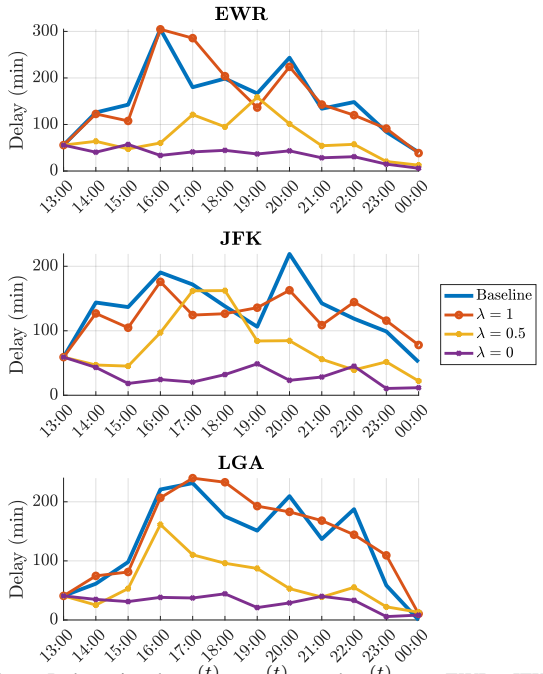

Fig. 5. Delay signals $x^{(t)}_{\text{EWR}}$, $x^{(t)}_{\text{JFK}}$, and $x^{(t)}_{\text{LGA}}$ at EWR, JFK, and LGA, respectively, across three separate solutions of (12) with differing λ parameter values.

work, our framework only requires that a copula is available to be sampled from, regardless of its specific family.

With regards to (2)-(5), we note that the optimization problem (12) for the airport delay network example was readily extendable from (10). Suppose, instead, that we want to prescribe control actions over a bike-share network, where certain vertices (bike stations) are rewarded for maintaining a steady inventory of bikes during the bike redistribution process. We could easily adapt a version of this problem from (12) by changing the sign on the third term of the objective function. Furthermore, we could assign priorities to different vertices by switching to a non-binary cost vector in (13). Now suppose that the bike-share system operator has access to bike depots located around the city, effectively allowing for a certain number of sink- and source-vertices in the network. This new degree of freedom can be captured by relaxing the conservation constraints in (12), i.e., by choosing a more lenient buffer, δ. Lastly, our choice of $\text{proj}_{\mathbb{R}^2}$ was made based on our preference of two aggregate system performance measures; the system operator could alternatively choose a modified $\widetilde{\text{proj}}_{\mathbb{R}^2}$ that captures other aggregate performance measures. Some canonical examples would include $\mathbb{E}\left[\mathbf{x}^{(t)}\right]$, $\text{Var}\left[\mathbf{x}^{(t)}\right]$, and $\left\|\mathbf{x}^{(t)}\right\|_\infty$.

V. Concluding Remarks

We demonstrated the ability of copulas to provide an observable approximation for the feasible state space of a complex system that has intricate dependence structures, as well as different distributions of sub-system behavior. In particular, copulas provide a way to separate estimating marginal distributions and estimating dependence structures. We fit historical state observations to the family of multivariate Gaussian copulas, and constructed an approximate state space through copula sampling and the inverse probability integral transform. We then proposed a control framework for large-scale networks that sets performance targets in a lower-dimensional projection of the full state space, and provides future system states by selecting from candidate states in the copula-approximated state space. We demonstrated the applicability of our proposed methodology through a case study of flight delays in the US airport network, as well as its effectiveness in selectively redistributing airport delays away from a particular subset of airports. A question of ongoing interest relates to the sample complexity of our copula-approximated state space.

References

[1] R. B. Nelson, *An Introduction to Copulas*. New York, NY: Springer, 2006, pp. 7–49. [Online]. Available: https://doi.org/10.1007/0-387-28678-0_2

[2] J. Sjöberg, Q. Zhang, L. Ljung, A. Benveniste, B. Deylon, P.-Y. Glorennec, H. Hjalmarsson, and A. Juditsky, *Nonlinear black-box modeling in system identification: a unified overview*. Linköping University, 1995.

[3] L. Ljung, "System identification," *Wiley Encyclopedia of Electrical and Electronics Engineering*, pp. 1–19, 1999.

[4] M. Khalil, S. Adhikari, and A. Sarkar, "Linear system identification using proper orthogonal decomposition," *Mechanical Systems and Signal Processing*, vol. 21, no. 8, pp. 3123–3145, 2007.

[5] B. Peherstorfer and K. Willcox, "Dynamic data-driven reduced-order models," *Computer Methods in Applied Mechanics and Engineering*, vol. 291, pp. 21–41, 2015.

[6] M. Bergmann, L. Cordier, and J.-P. Brancher, "Optimal rotary control of the cylinder wake using proper orthogonal decomposition reduced-order model," *Physics of fluids*, vol. 17, no. 9, p. 097101, 2005.

[7] M. Samimy, M. Debiasi, E. Caraballo, A. Serrani, X. Yuan, J. Little, and J. Myatt, "Feedback control of subsonic cavity flows using reduced-order models," *Journal of Fluid Mechanics*, vol. 579, pp. 315–346, 2007.

[8] A. Sklar, "Fonctions de répartition à n dimensions et leurs marges," *Publ. Inst. Statist. Univ. Paris*, vol. 8, pp. 229–231, 1959.

[9] P. Embrechts, A. J. McNeil, and D. Straumann, *Correlation and Dependence in Risk Management: Properties and Pitfalls*. Cambridge University Press, 2002, p. 176–223.

[10] E. Kole, K. Koedijk, and M. Verbeek, "Selecting copulas for risk management," *Journal of Banking & Finance*, vol. 31, no. 8, pp. 2405 – 2423, 2007.

[11] H. Noh, A. E. Ghouch, and T. Bouezmarni, "Copula-based regression estimation and inference," *Journal of the American Statistical Association*, vol. 108, no. 502, pp. 676–688, 2013.

[12] C. Eickhoff, A. P. de Vries, and K. Collins-Thompson, "Copulas for information retrieval," in *SIGIR '13*, 2013, p. 663–672.

[13] P. Wanigasekara, "Latent state space models for prediction," Ph.D. dissertation, Massachusetts Institute of Technology, 2016.

[14] A. Kreuzer, L. D. Valle, and C. Czado, "Bayesian multivariate non-linear state space copula models," *arXiv preprint arXiv:1911.00448*, 2019.

[15] M. S. Smith and W. Maneesoonthorn, "Inversion copulas from non-linear state space models with an application to inflation forecasting," *International Journal of Forecasting*, vol. 34, no. 3, pp. 389–407, 2018.

[16] A. Jacquillat and A. R. Odoni, "Endogenous control of service rates in stochastic and dynamic queuing models of airport congestion," *Transportation Research Part E: Logistics and Transportation Review*, vol. 73, pp. 133–151, 2015.

[17] K. Gopalakrishnan, M. Z. Li, and H. Balakrishnan, "Identification of outliers in graph signals," in *2019 IEEE 58th Conference on Decision and Control (CDC)*. IEEE, 2019, pp. 4769–4776.

[18] J.-D. Fermanian, "Goodness-of-fit tests for copulas," *Journal of Multivariate Analysis*, vol. 95, no. 1, pp. 119 – 152, 2005.

[19] B. Li and M. G. Genton, "Nonparametric identification of copula structures," *Journal of the American Statistical Association*, vol. 108, no. 502, pp. 666–675, 2013.