A Mass-Conservation Model for Stability Analysis and Finite-Time Estimation of Spread of COVID-19

Hossein Rastgoftar, *Member, IEEE*, and Ella Atkins, *Senior Member, IEEE*

***Abstract*— The COVID-19 global pandemic has significantly impacted people throughout the United States and the World. While it was initially believed the virus was transmitted from animal to human, person-to-person transmission is now recognized as the main source of community spread. This article integrates data into physics-based models to analyze stability of the rapid COVID-19 growth and to obtain a data-driven model for spread dynamics among the human population. The proposed mass-conservation model is used to learn the parameters of pandemic growth and to predict the growth of total cases, deaths, and recoveries over a finite future time horizon. The proposed finite-time prediction model is validated by finite-time estimation of the total numbers of infected cases, deaths, and recoveries in the United States from March 12, 2020 to December 9, 2020.**

***Index Terms*— Finite-time estimation, finite-time modding, pandemic growth stability.**

I. Introduction

THE FIRST confirmed case of COVID-19 was identified in Wuhan, China, in December 2019 and the virus has rapidly grown across the world since then. Studies have shown that the virus is dominantly transmitted through close contact with infected people and contaminated surfaces as well as respiratory droplets [1], [2]. Recent studies support that the incubation period of the disease is within 14 days [3], [4]. However, there is uncertainty about the median incubation period of the virus as researchers have reported median incubation periods of 4 days [4], 5.1 days [5], and 6.4 days [6]. Metapopulation Dynamics [7]–[9], Mean-Field Theory [10]–[13], the SIR method [14], [15], and SEIR dynamics [16]–[18] are existing models used to estimate the dynamics of infectious disease. Atkeson applies the SIR model to estimate and evaluate the negative impacts of the growth of COVID 19 in the United States. Metapopulation Dynamics is used in [19] to evaluate the impact of domestic and international travel on the spread of COVID-19. The SEIR model was used to analyze the epidemic of COVID-19 in mainland China [20] and the Diamond Princess cruise ship [21]. Fanelli and Piazza [12] applies mean-field theory to analyze and predict the spread of COVID-19 in China, Italy, and France. Also, van Doremalen *et al.* [2] applies a Bayesian regression model to analyze the aerosol stability of COVID-19.

Several researchers have studied the negative impacts of COVID-19 on daily life. Negative psychological outcomes of COVID-19 were studied [22]–[24]. Zhang and Ma [28] and Liang *et al.* [29] investigated the influences of the COVID-19 outbreak on mental health and family support in China. Researchers have also investigated the negative impacts of COVID-19 on the economy [27]–[30], environment [31], gender equality, education [32], healthcare system [33], tourism [34], [35], agriculture, food security, and supply chain management [36].

The COVID 19 global pandemic has had a devastating impact throughout the United States and the world. To date, more than 46 900 000 confirmed cases have been identified and there have been 1 210 000 deaths worldwide. The numbers are still increasing and being updated daily by the World Health Organization [37], Centers for Disease Control and Prevention [1], and European Centre for Disease Prevention and Control [38]. The delay in responding effectively to the pandemic spread has exacerbated its negative impacts on public health, economy, education, and other sectors across the World, and revealed an urgent need for a plan. The ever-evolving consequences of COVID-19 and inevitable future pandemics will be better controlled if knowledge is gained from this outbreak to support construction and validation of a more informed pandemic model and intervention plan.

The primary objective of this article is to develop a data-driven model for finite-time estimation of the growth of a pandemic across the human population to effectively predict the spread of disease. This model can facilitate taking required preventive nonpharmaceutical interventions when there is no vaccine available. To this end, we first apply the mass conservation law to model evolution of the total number of cases, deaths, and recoveries across the human population. We then develop a constrained optimization problem to obtain parameters of the proposed dynamics based on statistics of the number of infected cases, deaths, and recoveries recorded in a time-sliding window. By learning the parameters of the proposed model, we develop a finite-time dynamics model to continuously predict the number of infection cases, deaths, and recoveries within a finite number of future days. The proposed prediction dynamics is validated by finite time estimation of COVID-19 infection cases, deaths, and recoveries in the

Manuscript received April 13, 2020; revised November 2, 2020 and December 15, 2020; accepted January 6, 2021. Date of publication February 2, 2021; date of current version August 2, 2021. This work was supported by the National Science Foundation under Award 1914581 and Award 1739525. *(Corresponding author: Hossein Rastgoftar.)*

Hossein Rastgoftar is with the Department of Mechanical Engineering, Villanova University, Villanova, PA 19085 USA, and also with the Department of Aerospace Engineering, University of Michigan, Ann Arbor, MI 48109 USA (e-mail: hossein.rastgoftar@villanova,edu).

Ella Atkins is with the Department of Aerospace Engineering, University of Michigan, Ann Arbor, MI 48109 USA (e-mail: ematkins@umich.edu).

Digital Object Identifier 10.1109/TCSS.2021.3050476

United States from March 12, 2020 to December 9, 2020. Compared to the existing Mean-Field Theoretic, SIR, and SEIR models, our proposed method can estimate the growth of a pandemic for a shorter future time horizon but with better accuracy. This is because the proposed method consistently learns and incorporates real-time data into growth modeling of the outbreak disease. As the result partial information about incubation period and exposure to the virus has minimal impact on predicting the spread of the pandemic disease.

The secondary objective of this article is provide a criterion for stability of a pandemic. In particular, the proposed stability criterion is applied to analyze the stability of the growth of COVID-19 in the United States where the publicly available data provided in [39] is used to incorporate day-to-day information about the number of confirmed cases, deaths, and recoveries into modeling of spread dynamics.

This article is organized as follows. A problem statement presented in Section II is followed by stability analysis and parameter estimation approaches developed in Section III. A novel finite-state prediction dynamics are obtained in Section III-B. Results are presented in Section V followed by concluding remarks in Section VI.

II. Problem Statement

Let t be the total number of cases with confirmed disease, d be the total number of deaths, r be the total number of recovered cases, u_t be the total number of new infected cases, u_d be the number of new deaths, and u_r be the number of new recovered cases. Using the mass conversation law, the spread dynamics of COVID-19 is given by

$$\mathbf{x}[k] = \mathbf{x}[k-1] + \mathbf{u}[k-1] \tag{1}$$

where $k = 1, 2 \cdots$ denotes a calendar day, $\mathbf{x}[k] = \begin{bmatrix} t[k] & d[k] & r[k] \end{bmatrix}^T$ is the state vector, and $\mathbf{u}[k] = \begin{bmatrix} u_t[k] & u_d[k] & u_r[k] \end{bmatrix}^T$ is the input vector. This article assumes that the existing data of COVID-19 is valid, thus, $\mathbf{x}$ is observable at every day k from the establishment of the pandemic. Given the total number of confirmed cases, deaths, and recoveries, the number of active cases becomes

$$k = 1, 2, \ldots, \qquad a[k-1] = t[k-1] - r[k-1] - d[k-1]. \tag{2}$$

The underlying mass-conservation model is also used to estimate the growth of COVID-19. We define $\hat{\mathbf{x}}[k] = \begin{bmatrix} \hat{t}[k] & \hat{d}[k] & \hat{r}[k] \end{bmatrix}^T$ and $\hat{\mathbf{u}}[k] = \begin{bmatrix} \hat{u}_t[k] & \hat{u}_d[k] & \hat{u}_r[k] \end{bmatrix}^T$ as the estimations of the actual state vector $\mathbf{x}[k]$ and the actual input vector $\hat{\mathbf{u}}[k]$ at day k, where the estimation of the growth of the pandemic is modeled by dynamics

$$\hat{\mathbf{x}}[k] = \hat{\mathbf{x}}[k-1] + \hat{\mathbf{u}}[k-1]. \tag{3}$$

Estimations of new total cases, denoted by $\hat{u}_t$, new deaths, denoted by $\hat{u}_d$, and new recoveries, denoted by $\hat{u}_r$, are defined as follows:

$$\hat{u}_t[k-1] = \sum_{h=1}^{N_\tau} \omega_h[k] a[k-h], \tag{4a}$$

$$\hat{u}_d[k-1] = \sum_{h=1}^{N_\tau} \lambda_h[k] a[k-h], \tag{4b}$$

$$\hat{u}_r[k-1] = \sum_{h=1}^{N_\tau} \theta_h[k] a[k-h] \tag{4c}$$

where $\omega_h : \mathbb{N} \to \mathbb{R}_{\geq 0}$, $\lambda_h : \mathbb{N} \to \mathbb{R}_{\geq 0}$, and $\theta_h : \mathbb{N} \to \mathbb{R}_{\geq 0}$.

Given the above problem specification, the main objective of this article is to study the following two problems.

1) *Problem 1—Estimation and Stability Analysis of Pandemic Growth Dynamics:* We use the underlying mass conservation law to obtain $N_\tau[k]$, $\omega_h[k]$, $\lambda_h[k]$, and $\theta_h[k]$ for $h \in \{1, \ldots, N_\tau\}$ at day k such that estimation vector $\hat{\mathbf{x}}[k]$ converges to actual state vector $\mathbf{x}[k]$. We also analyze the stability of the growth of COVID-19 by analyzing the stability of estimation dynamics (3).
2) *Problem 2—Growth Prediction:* Assuming the total number of infected cases, deaths, and recoveries are known in a time sliding window of length N_τ at days $k - N_\tau$, $\cdots$, and $k-1$, we propose a model based on mass conservation to predict the infected cases, deaths, and recoveries within the next M_τ days, at days $k+1$, $\cdots$, and $k + M_\tau$.

III. Estimation of COVID-19 Spread

We first use the mass conservation law to obtain a data-driven time-varying dynamics in Section III-A to model the spread of COVID-19. We then propose an optimization problem in Section III-B to determine parameters of the proposed model by relying on the existing data of COVID-19.

A. Estimation Dynamics

Let active cases given by (2) be expressed as follows:

$$k = 1, 2, \ldots, \qquad a[k-1] = \mathbf{a}\hat{\mathbf{x}}[k-1] \tag{5}$$

where

$$\mathbf{a} = \begin{bmatrix} 1 & -1 & -1 \end{bmatrix} \in \mathbb{R}^{1\times 3}. \tag{6}$$

By knowing $\hat{\mathbf{x}}[k - N_\tau], \cdots, \hat{\mathbf{x}}[k-1]$, we can use (4) and express $\hat{u}_t[k-1]$, $\hat{u}_d[k-1]$, and $\hat{u}_r[k-1]$ as follows:

$$\hat{\mathbf{u}}[k-1] = \sum_{h=1}^{N_t} \mathbf{G}_h[k-h]\hat{\mathbf{x}}[k-h] \tag{7}$$

where

$$\mathbf{G}_h = \mathbf{K}_h \mathbf{a}, \tag{8a}$$

$$i \in \mathcal{V},\ h = 1, \ldots, N_\tau,\ k = 1, 2, \ldots, \qquad \mathbf{K}_h[k] = \begin{bmatrix} \omega_h[k] \\ \lambda_h[k] \\ \theta_h[k] \end{bmatrix}. \tag{8b}$$

Substituting $\mathbf{a} = \begin{bmatrix} 1 & -1 & -1 \end{bmatrix}$ and (8b), matrix $\mathbf{G}_h$ is obtained as follows:

$$\mathbf{G}_h[k] = \begin{bmatrix} \omega_h[k] & -\omega_h[k] & -\omega_h[k] \\ \lambda_h[k] & -\lambda_h[k] & -\lambda_h[k] \\ \theta_h[k] & -\theta_h[k] & -\theta_h[k] \end{bmatrix} \tag{9}$$

for $h = 1, \ldots, N_\tau$ at discrete time k. Replacing $\hat{\mathbf{u}}$ by (7), the estimation dynamics (3) is converted to

$$\hat{\mathbf{x}}[k] = (\mathbf{I} + \mathbf{G}_1)\hat{\mathbf{x}}[k-1] + \sum_{h=2}^{N_\tau} \mathbf{G}_h \hat{\mathbf{x}}[k-h] \tag{10}$$

for $k = 1, 2, \cdots$ and $h \in \{1, \ldots, N_\tau\}$, where $\mathbf{I} \in \mathbb{R}^{3\times 3}$ is the identity matrix. Fig. 1 shows a block diagram of estimation dynamics (10).

1) Finite-Time Estimation Dynamics: Define vector

$$\hat{\mathbf{Y}}[k-1] = \left[\hat{\mathbf{x}}^T[k-N_\tau] \cdots \hat{\mathbf{x}}^T[k-1]\right]^T \in \mathbb{R}^{3N_\tau \times 1} \quad (11)$$

providing estimated number of infected cases, deaths, and recoveries in the past N_τ days for day k. If $\hat{\mathbf{x}}$ is updated by (10), $\hat{\mathbf{Y}}[k]$ is updated by finite-time estimation dynamics

$$\hat{\mathbf{Y}}[k] = \mathbf{\Lambda}[k]\hat{\mathbf{Y}}[k-1] \quad (12)$$

where

$$\mathbf{\Lambda}[k] = \begin{bmatrix} \mathbf{0} & \mathbf{I} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I} \\ \mathbf{G}_{N_\tau}[k] & \mathbf{G}_{N_\tau-1}[k] & \cdots & \mathbf{I}+\mathbf{G}_1[k] \end{bmatrix} \in \mathbb{R}^{3N_\tau \times 3N_\tau}. \quad (13)$$

It is noted that $\mathbf{I} \in \mathbb{R}^{3\times 3}$ is the identity matrix and $\mathbf{0} \in \mathbb{R}^{3\times 3}$ is a zero-entry matrix.

2) Dynamics of Spread of Active Cases: By considering (2), (6), and (8a), the estimation dynamics (10) can be expressed as follows:

$$\hat{\mathbf{x}}[k] = \hat{\mathbf{x}}[k-1] + \sum_{h=1}^{N_\tau} \mathbf{K}_h a[k-h]. \quad (14)$$

Premultiplying both sides of (14) by $\mathbf{a}$, (14) is converted to

$$\mathbf{a}\hat{\mathbf{x}}[k] = \mathbf{a}\hat{\mathbf{x}}[k-1] + \sum_{h=1}^{N_\tau} \mathbf{a}\mathbf{K}_h a[k-h]. \quad (15)$$

Now, define

$$\gamma_h[k] := \mathbf{a}\mathbf{K}_h[k] = \omega_h[k] - \lambda_h[k] - \theta_h[k] \quad (16)$$

at day k for $h \in \{1, \ldots, N_\tau\}$. We can replace $\mathbf{a}\hat{\mathbf{x}}[k]$, $\mathbf{a}\hat{\mathbf{x}}[k-1]$, and $\mathbf{a}\mathbf{K}_h[k]$ by $a[k]$, $a[k-1]$, and $\gamma_h[k]$, respectively, in (15). Then, dynamics of spread of active cases is given by

$$a[k] = (1+\gamma_1[k])a[k-1] + \cdots + \gamma_{N_\tau}[k]a[k-N_\tau]. \quad (17)$$

Theorem 1: The growth of active cases, governed by dynamics (17), reaches the stability if there exists a day k_s such that eigenvalues of matrix

$$\mathbf{\Gamma}[k] = \begin{bmatrix} 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ \gamma_{N_\tau}[k] & \gamma_{N_\tau-1}[k] & \cdots & 1+\gamma_1[k] \end{bmatrix} \in \mathbb{R}^{N_\tau \times N_\tau} \quad (18)$$

are all inside the unit disk centered at the origin at every day $k > k_s$.

Proof: Given the spread dynamics (3) with control input $\mathbf{u}$ defined by (4), the dynamics of the growth of active cases becomes

$$\mathbf{z}[k+1] = \mathbf{\Gamma}[k]\mathbf{z}[k] \quad (19)$$

where

$$\mathbf{z}[k] = \left[a[k-N_\tau] \cdots a[k-1]\right]^T. \quad (20)$$

Eigenvalues $\sigma_1[k], \cdots, \sigma_{N_\tau}[k]$ are the roots of the following characteristic polynomial:

$$\sigma^{N_\tau} - (1+\gamma_1)\sigma^{N_\tau-1} - \gamma_2\sigma^{N_\tau-2} - \cdots - \gamma_{N_\tau} = 0. \quad (21)$$

The growth of active cases achieves stability at day k_s if the eigenvalues of matrix $\mathbf{\Gamma}[k]$ all occur inside the unit disk centered at the origin at every $k \geq k_s$. If $a[k]$ achieves stability at time $k \geq k_s$, then $u_t[k]$, u_d, u_r, defined by (4), reach stability at time $k > k_s$.

■

B. Parameter Estimation

Let set

$$\mathcal{H} = \{1, \ldots, N_{\tau,\max}\} \quad (22)$$

define possible incubation periods of COVID-19 ranging from one day to $N_{\tau,\max}$ days. We define

$$h \in \mathcal{H}, \qquad \mathrm{C}_h\left(\tilde{\mathbf{K}}_1[k], \ldots, \tilde{\mathbf{K}}_h[k]\right) = \frac{1}{2}\sum_{j=1}^{h} \tilde{\mathbf{K}}_j^T \mathbf{P}_j \tilde{\mathbf{K}}_j \quad (23)$$

as the cost function for determining N_τ, $\mathbf{K}_1, \cdots, \mathbf{K}_{N_\tau}$, where $\mathbf{P}_h \in \mathbb{R}^{h\times h}$ is positive-definite and diagonal, and the gain vectors $\tilde{\mathbf{K}}_1[k], \cdots, \tilde{\mathbf{K}}_h[k]$ are candidates for $\mathbf{K}_1, \cdots, \mathbf{K}_h$ per (8b). This articleuses $N_{\tau,\max} = 14$ to obtain presented results.

1) Matrix $\mathbf{P}_h$: Weight matrix $\mathbf{P}_h = \left[p_{i,h}\right] \in \mathbb{R}^{h\times h}$ is a positive-definite and diagonal matrix defined based on the likelihood of the incubation period of COVID-19. We use the longnormal distribution to estimate the incubation period of COVID-19 by

$$i = 1, \ldots, h, \qquad \frac{e^{-\frac{\left(\ln\frac{i}{m}\right)^2}{2\beta^2}}}{\beta i\sqrt{2\pi}} \quad (24)$$

where m and β are the median and standard deviation of the distribution where $i \in \{1, \ldots, h\}$. Per the results reported in [40], $m = 6.8$ and $\beta = 3.4$ are chosen to determine matrix $\mathbf{P}_h$.

2) Assignment of N_τ and Gains $\mathbf{K}_1, \cdots, \mathbf{K}_{N_\tau}$: The cost function C_h depends on real-valued vectors $\tilde{\mathbf{K}}_1, \cdots, \tilde{\mathbf{K}}_h$ and discrete-valued variable $h \in \mathcal{H}$. We determine $\mathbf{K}_1, \cdots, \mathbf{K}_{N_\tau}$, and N_τ by solving the following optimization problem:

$$\left(\bar{\mathbf{K}}_1^{*[k]}, \ldots, \bar{\mathbf{K}}_h{}^*[k], h^{*[k]}\right) = \underset{h\in\mathcal{H}}{\operatorname{argmin}} \left(\underset{\tilde{\mathbf{K}}_1,\ldots,\tilde{\mathbf{K}}_h}{\operatorname{argmin}} \left(\sum_{j=1}^{h} \tilde{\mathbf{K}}_j^T[k]\mathbf{P}_j\tilde{\mathbf{K}}_j[k]\right)\right) \quad (25)$$

subject to inequality constraints

$$j = 1, \ldots, h, \qquad \tilde{\mathbf{K}}_j[k] \geq \mathbf{0} \quad (26)$$

and equality constraint

$$\mathbf{x}[k] - \mathbf{x}[k-1] = \sum_{j=1}^{h} \tilde{\mathbf{K}}_j \mathbf{a} a[k-j] \quad (27)$$

where $h = \mathcal{H}$. By solving the above optimization problem, $\mathbf{K}_1[k] = \bar{\mathbf{K}}_1^*[k]$, $\cdots$, $\mathbf{K}_{N_\tau}[k] = \bar{\mathbf{K}}_h^*[k]$, and $N_\tau = h^*$ are used to assign matrix $\mathbf{\Lambda}[k]$, defined in (13).

Remark 1: For $h \in \{1, \ldots, N_{\tau,\max}\}$, $\bar{\mathbf{K}}_1[k], \cdots, \bar{\mathbf{k}}_h[k]$ denote the optimal gain vectors obtained by solving the following quadratic programming optimization problem with linear equality and inequality constraints:

$$\begin{aligned}\big(\bar{\mathbf{K}}_1[k], \ldots, \bar{\mathbf{K}}_h[k]\big) &= \underset{\tilde{\mathbf{K}}_1,\ldots,\tilde{\mathbf{K}}_h}{\operatorname{argmin}} \mathrm{C}_h\big(\tilde{\mathbf{K}}_1[k], \ldots, \tilde{\mathbf{K}}_h[k]\big) \\ &= \underset{\tilde{\mathbf{K}}_1,\ldots,\tilde{\mathbf{K}}_h}{\operatorname{argmin}} \left(\frac{1}{2}\sum_{j=1}^{h} \tilde{\mathbf{K}}_j^T[k]\mathbf{P}_j\tilde{\mathbf{K}}_j[k]\right) \quad (28)\end{aligned}$$

subject to constraints (26) and (27). Then, we can write

$$\mathrm{C}^*_{h^*}[k] = \min\big\{\mathrm{C}^*_1[k], \ldots, \mathrm{C}^*_{N_\tau,\max}[k]\big\}$$

where $\mathrm{C}^*_h = \mathrm{C}_h\big(\bar{\mathbf{K}}_1, \ldots, \bar{\mathbf{K}}_h\big)$ and $\bar{\mathbf{K}}_1[k], \cdots, \bar{\mathbf{K}}_h[k]$ are assigned by solving (28) subject to constraints (26) and (27). Hence

$$\mathrm{C}^*_{h^*} = \underset{h\in\mathcal{H}}{\operatorname{argmin}} \left(\underset{\tilde{\mathbf{K}}_1,\ldots,\tilde{\mathbf{K}}_h}{\operatorname{argmin}} \left(\sum_{j=1}^{h} \tilde{\mathbf{K}}_j^T[k]\mathbf{P}_j\tilde{\mathbf{K}}_j[k]\right)\right) \quad (29)$$

and $N_\tau[k] = h^*[k]$. The flowchart on the left-hand side of Fig. 1 and the algorithm presented in Table I provides a method for solving the above optimization problem.

Define

$$\mathbf{Y}_h[k-1] = \big[\mathbf{x}^T[k-1] \cdots \mathbf{x}^T[k-h]\big]^T \in \mathbb{R}^{3h\times 1}$$

and

$$\mathbf{V}_h = \mathrm{vec}\Big(\big[\tilde{\mathbf{K}}_1 \cdots \tilde{\mathbf{K}}_h\big]^T\Big) \in \mathbb{R}^{3h\times 1}$$

for $h = 1, \ldots, N_{\tau,\max}$ where "vec" is the matrix vectorization operator. The inner-loop constrained quadratic optimization problem (28), presented in Remark (1), can be defined as follows:

$$\min_{\mathbf{V}_h} = \frac{1}{2}\mathbf{V}_h^T(\mathbf{I}_3 \otimes \mathbf{P}_h)\mathbf{V}_h \quad (30)$$

for $h = 1, \ldots, N_{\tau,\max}$ subject to

$$\big(\mathbf{I}_3 \otimes \mathbf{z}^T[k]\big)\mathbf{V}_h = \mathbf{X}_h[k] - \mathbf{X}_h[k-1], \quad (31a)$$

$$\mathbf{V}_h \geq \mathbf{0}_{3h\times 1} \quad (31b)$$

where $\mathbf{0}_{3h\times 1} \in \mathbb{R}^{3h\times 1}$ is a zero-entry matrix and $\otimes$ is the Kronecker product symbol. We use the MATLAB command "quadprog" to obtain gain vectors $\bar{\mathbf{K}}_1, \cdots, \bar{\mathbf{K}}_h$ for $h \in \big\{1, \cdots N_{\tau,\max}\big\}$. More specifically,

$$\bar{\mathbf{K}}_j = \big(\mathbf{I}_3 \otimes \big[\delta_{j1} \cdots \delta_{jj} \cdots \delta_{jh}\big]\big)\bar{\mathbf{V}}_h$$

for $j = 1, \cdots, h$ where $\bar{\mathbf{V}}_h$ is the solution of the above quadratic programming problem, minimizing cost function (30) and satisfying constraints (31a) and (31b), and δ_{ji} is the Kronecker delta defined as follows:

$$\delta_{ji} = \begin{cases} 1 & j = i,\ i \in \mathcal{H} \\ 0 & j \neq i,\ i \in \mathcal{H} \end{cases}. \quad (32)$$

By knowing matrix $\mathbf{\Lambda}[k]$ at day k, a prediction dynamics is developed in Section IV to estimate the number of infected cases, deaths, and recoveries within the next M_τ days.

IV. Finite-Step Prediction Dynamics

Let $\mathbf{x}[k+1-N_\tau], \cdots, \mathbf{x}[k]$ be known at day k. Prediction of the state x at day $k + M_\tau$ is denoted by $\mathbf{x}_{p,M_\tau}[k + M_\tau] = \big[t_{p,M_\tau}[k + M_\tau]\ d_{p,M_\tau}[k + M_\tau]\ r_{p,M_\tau}[k + M_\tau]\big]^T$ where $t_{p,M_\tau}[k + M_\tau]$, $d_{p,M_\tau}[k + M_\tau]$, and $r_{p,M_\tau}[k + M_\tau]$ are the predictions for the total numbers of infected cases, deaths, and recoveries at day $k + M_\tau$. Note that subscript "p" denote "prediction" and $\mathbf{x}_{p,M_\tau}[k + M_\tau]$ predicts total infected cases, deaths, and recoveries at day $k+M_\tau$ where $M_\tau \in \{1, \ldots, N_\tau\}$. We define the prediction state vector $\mathbf{S}_{M_\tau}[k] = \big[\mathbf{s}_{i,M_\tau}[k]\big] \in \mathbb{R}^{3N_\tau\times 1}$ where

$$\mathbf{s}_{i,M_\tau}[k] = \begin{cases} \mathbf{x}[k+i+M_\tau-N_\tau] & i \leq N_\tau - M_\tau \\ \mathbf{x}_{p,M_\tau}[k+i+M_\tau-N_\tau] & N_\tau - M_\tau < i \leq N_\tau. \end{cases} \quad (33)$$

It is noted that $\mathbf{S}_0[k] = \mathbf{Y}[k]$ if $M_\tau = 0$. $\mathbf{S}_{M_\tau}[k]$ is updated by the following prediction dynamics:

$$\begin{cases} \mathbf{S}_{M_\tau+1}[k] = \mathbf{\Lambda}[k]\mathbf{S}_{M_\tau}[k] \\ \mathbf{O}_{M_\tau}[k] = \Big[\mathbf{0}_{3M_\tau\times 3(N_\tau-M_\tau)}\ \mathbf{I}_{3M_\tau}\Big]\mathbf{S}_{M_\tau} \end{cases} \quad (34)$$

subject to initial condition

$$\mathbf{S}_0[k] = \mathbf{Y}[k] \quad (35)$$

where $\mathbf{0}_{3M_\tau\times 3(N_\tau-M_\tau)} \in \mathbb{R}^{3M_\tau\times 3(N_\tau-M_\tau)}$, $\mathbf{I}_{3M_\tau} \in \mathbb{R}^{3M_\tau\times 3M_\tau}$ is the identity matrix, and $\mathbf{O}_{M_\tau}$ is the output vector of prediction dynamics (34).

Algorithm 1 Assignment of $N_\tau[k]$ and Gain Vectors $\mathbf{K}_1[k], \cdots, \mathbf{K}_\tau[k]$ at Day k

Result: Obtain $N_\tau[k]$, $\mathbf{K}_1[k], \cdots, \mathbf{K}_{N_\tau}[k]$
Initialization: Set $h = 0$ and $N_{\tau,\max} = 14$; choose a very large $\mathrm{C}_{\min}$ ($\mathrm{C}_{\min} = 10^{12}$);
while $h \leq N_{\tau,\max}$ **do**
  Assign $\tilde{\mathbf{K}}_1, \cdots,$ and $\tilde{\mathbf{K}}_h$ such that:
    1) $\mathrm{C}_h\big(\tilde{\mathbf{K}}_1, \ldots, \tilde{\mathbf{K}}_h\big)$ is minimized;
    2) Constraints (26) and (27) are satisfied.
  if $\mathrm{C}_h\big(\tilde{\mathbf{K}}_1, \ldots, \tilde{\mathbf{K}}_1\big) \leq \mathrm{C}_{\min}$ **then**
    $\mathrm{C}_{\min} \leftarrow \mathrm{C}_h\big(\tilde{\mathbf{K}}_1, \ldots, \tilde{\mathbf{K}}_h\big)$;
    $N_\tau \leftarrow h$;
    $\mathbf{K}_1 \leftarrow \tilde{\mathbf{K}}_h$;
    $\vdots$
    $\mathbf{K}_{N_\tau} \leftarrow \tilde{\mathbf{K}}_h$.
  end
  $h \leftarrow$ h+1
end

V. Results

We consider the spread of COVID-19 in the United States over 272 days from March 12, 2020 to December 9, 2020 where the number of infected cases ($t[k]$), deaths ($d[k]$), and recoveries ($r[k]$) are incorporated from [39] for $k = 1, \ldots, 272$. We chose $N_{\tau,\max} = 14$ to obtain parameters of the proposed model. Therefore, the statistics of the first 14 days, from March 12 to March 25, are only used to learn

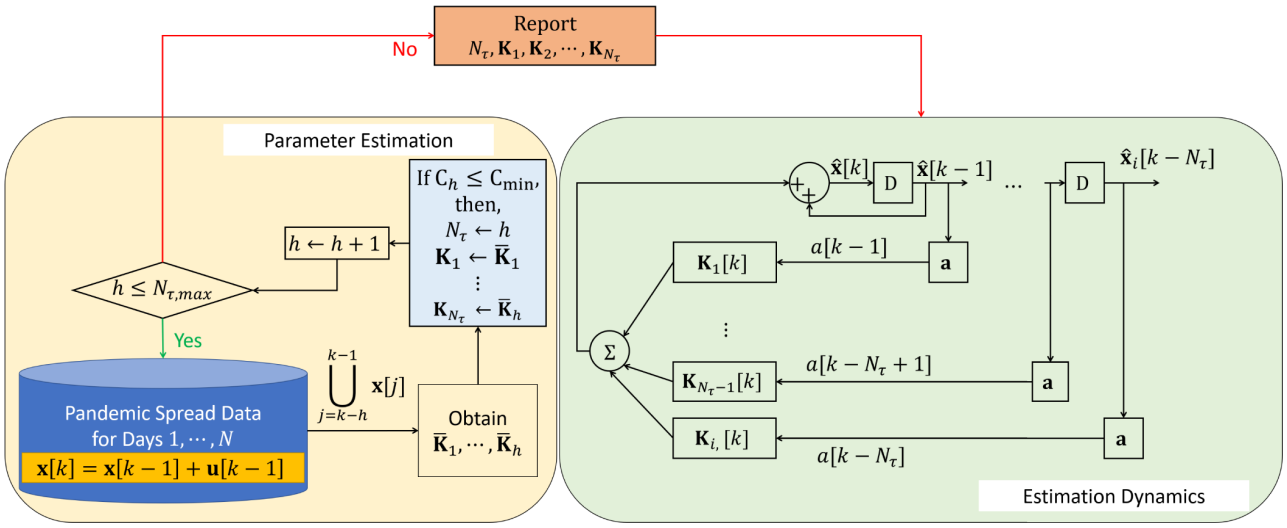

Fig. 1. Block diagram of the proposed mass-conservation model for the spread of a pandemic disease.

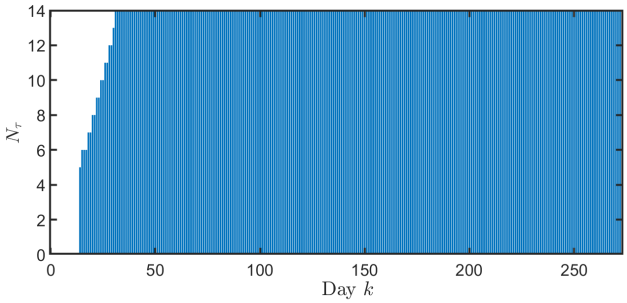

Fig. 2. $N_\tau[k]$ versus day k.

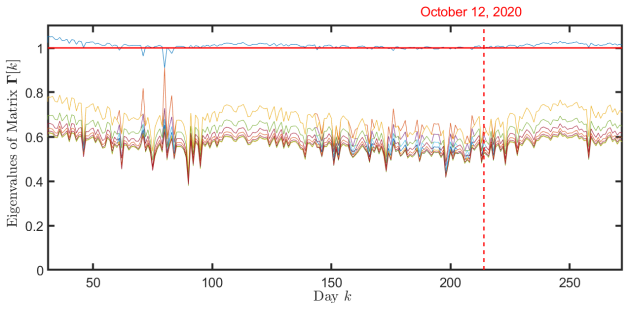

Fig. 3. Eigenvalues of matrix $\mathbf{\Gamma}[k]$.

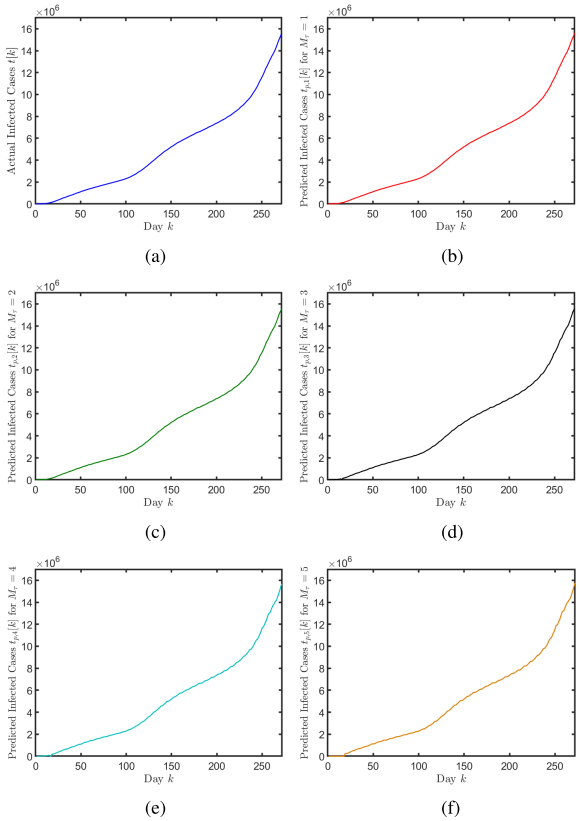

Fig. 4. Total number of infected cases for $k = 15, \ldots, 272$. (a) Total number of infected cases reported in [39]. The predicted numbers of infected cases are shown by red, green, black, cyan, and orange for (b) $M_\tau = 1$, (c) $M_\tau = 2$, (d) $M_\tau = 3$, (e) $M_\tau = 4$, and (f) $M_\tau = 5$, respectively.

parameters of the proposed model for $k \geq 15$, and we predict the growth of the total infected cases, deaths, and recoveries after March 26, 2020. As a result, the plots provided in this section do not provide information for $k = 1, \ldots, 14$.

We obtain gains $\mathbf{K}_1[k], \cdots, \mathbf{K}_{N_\tau}[k]$ and $N_\tau[k]$ by solving optimization problem (25) subject to constraints (26) and (27). In Fig. 2, $N_\tau[k]$ is plotted versus k for days 15 through 272. It is seen that $N_\tau[k] = 14$ for $k \geq 31$. Therefore, $\gamma_1[k] = \mathbf{K}_1[k]\mathbf{a}, \cdots, \gamma_{14}[k] = \mathbf{K}_{14}[k]\mathbf{a}$, and matrix $\mathbf{\Gamma}[k] \in \mathbf{R}^{14\times 14}$ are computed for every $k \geq 31$. Eigenvalues of matrix $\mathbf{\Gamma}[k]$ are plotted versus k in Fig. 3. One of the eigenvalues of matrix $\mathbf{\Gamma}[k]$ that has the largest magnitude at

$$\mathbf{O}_{M_\tau}[k] = \begin{cases} \emptyset & M_\tau = 0 \\ \left[\mathbf{x}^T_{p,M_\tau}[k+1] \cdots \mathbf{x}^T_{p,M_\tau}[k+M_\tau]\right]^T \in \mathbb{R}^{3M_\tau \times 1} & 1 \leq M_\tau \leq N_\tau \end{cases} \tag{36}$$

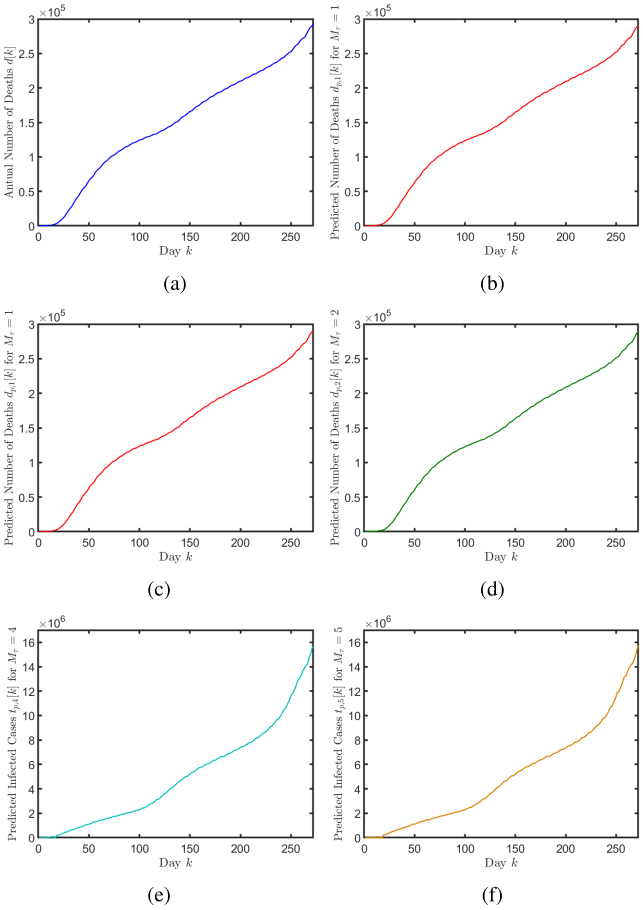

Fig. 5. Total number of deaths for $k = 15, \ldots, 272$: (a) Total number of deaths reported in [39]. The predicted numbers of deaths are shown by red, green, black, cyan, and orange for (b) $M_\tau = 1$, (c) $M_\tau = 2$, (d) $M_\tau = 3$, (e) $M_\tau = 4$, and (f) $M_\tau = 5$, respectively. (a) Actual number of deaths. (b) Predicted deaths for $M_\tau = 1$. (c) Predicted deaths for $M_\tau = 2$. (d) Predicted deaths for $M_\tau = 3$. (e) Predicted deaths for $M_\tau = 4$. (f) Predicted deaths for $M_\tau = 5$.

day k is considered as the first eigenvalue of matrix $\mathbf{\Gamma}[k]$. It is seen that the magnitude of the first eigenvalue of matrix $\mathbf{\Gamma}[k]$ fluctuates near 1 while the magnitudes of the remaining eigenvalues of matrix $\mathbf{\Gamma}[k]$ are all between 0 and 1 for $k = 31, \ldots, 272$. Additionally, we observe that the magnitude of the first eigenvalue of matrix $\mathbf{\Gamma}[k]$ is greater than 1 every day after October 12, 2020 ($k = 214$).

Therefore, the growth of the total infected cases, deaths, and recoveries have not reached stability although the magnitudes of eigenvalues of matrix $\mathbf{\Gamma}$ are all less than 1 some days after March 12, 2020.

In Figs. 4–6 actual and predicted numbers of infected cases, deaths, and recoveries are plotted for $k = 14$ through $k = 272$. To quantify the error of the proposed prediction model, we define

$$e_{t,M_\tau}[k] = \frac{t_{p,M_\tau}[k] - t[k]}{t[k]} \tag{37a}$$

$$e_{r,M_\tau}[k] = \frac{r_{p,M_\tau}[k] - r[k]}{r[k]} \tag{37b}$$

$$e_{r,M_\tau}[k] = \frac{t_{r,M_\tau}[k] - r[k]}{r[k]} \tag{37c}$$

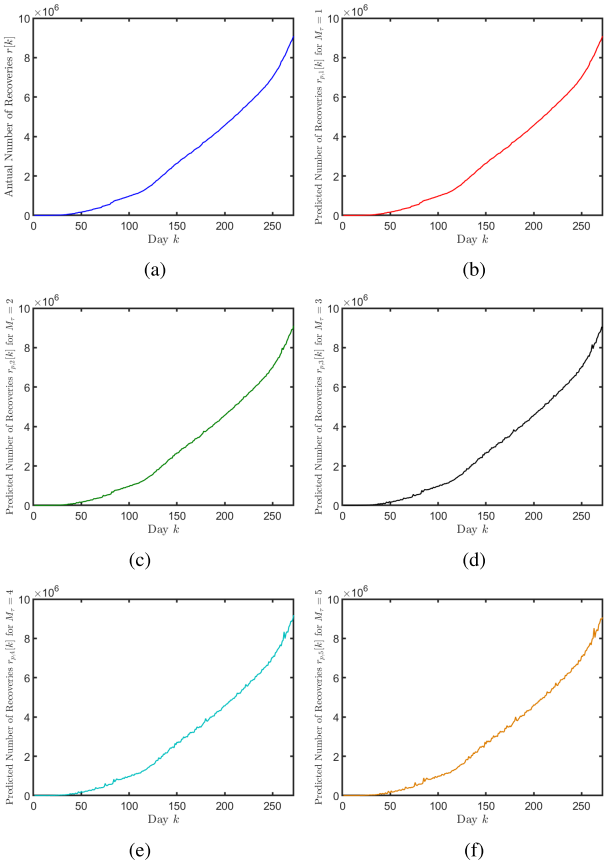

Fig. 6. Total number of recoveries for $k = 15, \ldots, 272$: (a) Total number of recoveries reported in [39]. The predicted numbers of recoveries are shown by red, green, black, cyan, and orange for (b) $M_\tau = 1$, (c) $M_\tau = 2$, (d) $M_\tau = 3$, (e) $M_\tau = 4$, and (f) $M_\tau = 5$, respectively.

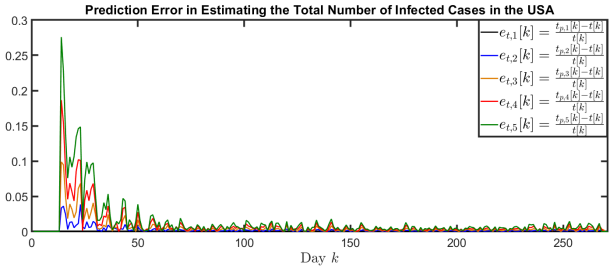

Fig. 7. Relative error of finite-time prediction of infected cases for $M_\tau = 1, \ldots, 5$ at day $k = 15, \ldots, 272$.

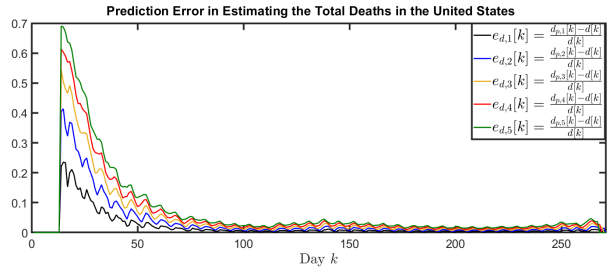

Fig. 8. Relative error of finite-time prediction of deaths for $M_\tau = 1, \ldots, 5$ at day $k = 15, \ldots, 272$.

as the relative error formulas in predicting the numbers of infected cases, deaths, and recoveries at day k. In Figs. 7–9, relative errors in predicting the total infected cases, deaths,

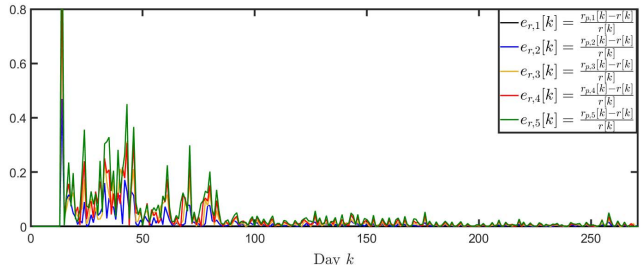

Fig. 9. Relative error of finite-time prediction of recoveries for $M_\tau = 1, \ldots, 5$ at day $k = 15, \ldots, 272$.

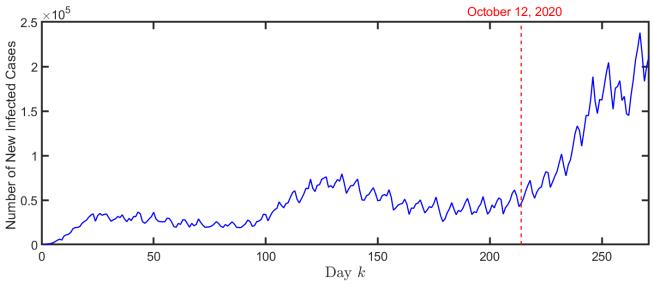

Fig. 10. New infected cases at day $k = 15, \ldots, 272$.

and recoveries are plotted for $M_\tau = 1, \ldots, 5$ at day $k = 15, \ldots, 272$. It is seen that the relative errors are significantly decreased after $k > 70$.

VI. Conclusion and Future Work

This article described a new data-driven approach, inspired by the law of mass conservation, to model and analyze the stability of COVID-19 spread dynamics. We developed an algorithm to learn the parameters of the proposed mass conservation-based model based on the history of infected cases, deaths, and recoveries recorded in time sliding windows. We used the history of total infected cases, deaths, and recoveries to develop a prediction model for estimating the numbers of infected cases, deaths, and recoveries within a receding finite time horizon. We evaluated the accuracy of the proposed prediction model by estimating the statistics of COVID-19 in the United States from March 12, 2020 to December 9, 2020. Results show that the relative estimation errors of the total number of infected cases, deaths, and recoveries are less than 3% after June 1.

The results of our model show that COVID-19 growth is unstable for all days after October 12, 2020 per Fig. 3. This result is consistent with the status of pandemic growth in the United States after mid-October. As shown in Fig. 10, the total number of new COVID-19 cases has significantly increased in the United States after October 12, 2020.

In future work, we plan to determine the underlying relation between control gains $\omega_1[k], \cdots, \omega_{N_\tau}[k]$, $\lambda_1[k], \cdots, \lambda_{N_\tau}[k]$, $\theta_1[k], \cdots, \theta_{N_\tau}[k]$ and state-wide and nationwide executive order release in the United States. To this end, we will use the existing SIR model to obtain the projection of disease spread and model the pandemic by an autonomous dynamics. By comparing the projected SIR dynamics and the nonautonomous conservation-based dynamics proposed in this article, control gains can be quantified and related to executive orders for every US state/district.

Furthermore, we plan to develop a novel decision-making model for optimal planning of infection control actions in the presence of uncertainty and ambiguity. More specifically, we will apply the finite-estimation model, proposed in this article, to model the spread of a pandemic disease as a Markov process with a Markov decision process (MDPs) applied to optimize nonpharmaceutical actions under a full-state observability assumption. The proposed decision-making model can also be applied to evaluate the effectiveness of statewide and nationwide orders and recommendations aimed at avoiding or mitigating rapid case growth.

Acknowledgment

The authors thank Mr. Xun Liu for proofreading of the final version of this manuscript.

References

[1] C for Disease Control and Prevention. (2020). *Coronavirus (COVID-19)*. [Online]. Available: https://www.cdc.gov/coronavirus/2019-ncov/index.html

[2] N. van Doremalen *et al.*, "Aerosol and surface stability of SARS-CoV-2 as compared with SARS-CoV-1," *New England J. Med.*, vol. 382, pp. 1564–1567, Apr. 2020.

[3] Q. Li *et al.*, "Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia," *New England J. Med.*, pp. 1200–1207, Jan. 2020.

[4] C. Eastin and T. Eastin, "Clinical characteristics of coronavirus disease 2019 in China," *J. Emergency Med.*, vol. 58, no. 4, pp. 711–712, Apr. 2020.

[5] S. A. Lauer *et al.*, "The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: Estimation and application," *Ann. Internal Med.*, vol. 172, no. 9, pp. 577–582, May 2020.

[6] C.-C. Lai, T.-P. Shih, W.-C. Ko, H.-J. Tang, and P.-R. Hsueh, "Severe acute respiratory syndrome coronavirus 2 (sars-cov-2) and corona virus disease-2019 (COVID-19): The epidemic and the challenges," *Int. J. Antimicrobial Agents*, vol. 17, Feb. 2020, Art. no. 105924.

[7] B. Grenfell, "(Meta)population dynamics of infectious diseases," *Trends Ecology Evol.*, vol. 12, no. 10, pp. 395–399, Oct. 1997.

[8] I. Hanski, "Metapopulation dynamics," *Nature*, vol. 396, no. 6706, pp. 41–49, 1998.

[9] M. J. Keeling, O. N. Bjørnstad, and B. T. Grenfell, "Metapopulation dynamics of infectious diseases," in *Ecology, Genetics and Evolution of Metapopulations*. Amsterdam, The Netherlands: Elsevier, 2004, pp. 415–445.

[10] Q. Wu and S. Chen, "Mean field theory of epidemic spreading with effective contacts on networks," *Chaos, Solitons Fractals*, vol. 81, pp. 359–364, Dec. 2015.

[11] V. Nagy, "Mean-field theory of a recurrent epidemiological model," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 79, no. 6, Jun. 2009, Art. no. 066105.

[12] D. Fanelli and F. Piazza, "Analysis and forecast of COVID-19 spreading in China, Italy and France," *Chaos, Solitons Fractals*, vol. 134, May 2020, Art. no. 109761.

[13] A. Kleczkowski and B. T. Grenfell, "Mean-field-type equations for spread of epidemics: The small world' model," *Phys. A, Stat. Mech. Appl.*, vol. 274, nos. 1–2, pp. 355–360, Dec. 1999.

[14] K. Biswas and P. Sen, "Space-time dependence of corona virus (COVID-19) outbreak," 2020, *arXiv:2003.03149*. [Online]. Available: http://arxiv.org/abs/2003.03149

[15] B. F. Maier and D. Brockmann, "Effective containment explains subexponential growth in confirmed cases of recent COVID-19 outbreak in mainland China," 2020, *arXiv:2002.07572*. [Online]. Available: http://arxiv.org/abs/2002.07572

[16] M. Y. Li, J. R. Graef, L. Wang, and J. Karsai, "Global dynamics of a SEIR model with varying total population size," *Math. Biosci.*, vol. 160, no. 2, pp. 191–213, Aug. 1999.

[17] V. Dukic, H. F. Lopes, and N. G. Polson, "Tracking epidemics with Google flu trends data and a state-space SEIR model," *J. Amer. Stat. Assoc.*, vol. 107, no. 500, pp. 1410–1426, Dec. 2012.

[18] Y. Liu, A. A. Gayle, A. Wilder-Smith, and J. Rocklöv, "The reproductive number of COVID-19 is higher compared to SARS coronavirus," *J. Travel Med.*, vol. 27, no. 2, Mar. 2020.

[19] M. Chinazzi *et al.*, "The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak," *Science*, vol. 368, no. 6489, pp. 395–400, Apr. 2020.
[20] L. Peng, W. Yang, D. Zhang, C. Zhuge, and L. Hong, "Epidemic analysis of COVID-19 in China by dynamical modeling," 2020, *arXiv:2002.06563*. [Online]. Available: http://arxiv.org/abs/2002.06563
[21] J. Rocklöv, H. Sjödin, and A. Wilder-Smith, "COVID-19 outbreak on the diamond princess cruise ship: Estimating the epidemic potential and effectiveness of public health countermeasures," *J. Travel Med.*, vol. 27, no. 3, May 2020.
[22] W. Cao *et al.*, "The psychological impact of the COVID-19 epidemic on college students in China," *Psychiatry Res.*, vol. 287, May 2020, Art. no. 112934.
[23] S. Li, Y. Wang, J. Xue, N. Zhao, and T. Zhu, "The impact of COVID-19 epidemic declaration on psychological consequences: A study on active Weibo users," *Int. J. Environ. Res. Public Health*, vol. 17, no. 6, p. 2032, Mar. 2020.
[24] M. T. Tull, K. A. Edmonds, K. M. Scamaldo, J. R. Richmond, J. P. Rose, and K. L. Gratz, "Psychological outcomes associated with Stay-at-Home orders and the perceived impact of COVID-19 on daily life," *Psychiatry Res.*, vol. 289, Jul. 2020, Art. no. 113098.
[25] Y. Zhang and Z. F. Ma, "Impact of the COVID-19 pandemic on mental health and quality of life among local residents in liaoning province, China: A cross-sectional study," *Int. J. Environ. Res. Public Health*, vol. 17, no. 7, p. 2381, Mar. 2020.
[26] L. Liang *et al.*, "The effect of COVID-19 on youth mental health," *Psychiatric Quart.*, vol. 4, pp. 1–12, Apr. 2020.
[27] M. Gupta, A. Abdelmaksoud, M. Jafferany, T. Lotti, R. Sadoughifar, and M. Goldust, "COVID-19 and economy," *Dermatologic Therapy*, pp. 1435–1443, Mar. 2020.
[28] W. McKibbin and R. Fernando, "The economic impact of COVID-19," *Econ. Time*, vol. 45, pp. 1–13, Oct. 2020.
[29] R. Baldwin and E. Tomiura, "Thinking ahead about the trade impact of COVID-19," *Econ. Time*, vol. 59, pp. 59–71, Oct. 2020.
[30] W. J. McKibbin and R. Fernando, "The global macroeconomic impacts of COVID-19: Seven scenarios," Austral. Nat. Univ., Canberra, ACT, Australia, Tech. Rep., 2020.
[31] M. A. Zambrano-Monserrate, M. A. Ruano, and L. Sanchez-Alcalde, "Indirect effects of COVID-19 on the environment," *Sci. Total Environ.*, vol. 728, Aug. 2020, Art. no. 138813.
[32] P. Sahu, "Closure of universities due to coronavirus disease 2019 (COVID-19): Impact on education and mental health of students and academic staff," *Cureus*, vol. 12, p. e7541, Apr. 2020.
[33] J. Sarasohn-Kahn. (2020). *The Median Hospital Charge In U.S. for COVID-19 Care Ranges From 34-45K*. [Online]. Available: https://www.healthpopuli.com/2020/07/17/the-median-hospital-charge-in-t%he-u-s-for-covid-19-care-ranges-from-34-45k/
[34] P. Niewiadomski, "COVID-19: From temporary de-globalisation to a re-discovery of tourism?" *Tourism Geographies*, vol. 23, pp. 1–26, Apr. 2020.
[35] R. T. R. Qiu, J. Park, S. Li, and H. Song, "Social costs of tourism during the COVID-19 pandemic," *Ann. Tourism Res.*, vol. 84, Sep. 2020, Art. no. 102994.
[36] R. Siche, "What is the impact of COVID-19 disease on agriculture?" *Scientia Agropecuaria*, vol. 11, no. 1, pp. 3–6, Mar. 2020.
[37] World Health Organization. (2020). *Novel Coronavirus-2019 Situation Reports*. [Online]. Available: https://gisanddata.maps.arcgis.com/apps/opsdashboard/index.html?fbclid=IwAR1smzyZx1mLzj3zOpv9cmWXUudS72comshBltYE2gw9QBk47BgRB3TNQKI#/bda7594740fd40299423467b48e9ecf6
[38] E. C. for Disease Prevention and Control. (2020). *Covid-19*. [Online]. Available: https://www.ecdc.europa.eu/en/covid-19-pandemic
[39] (2020). *Worldometer*. [Online]. Available: https://www.worldometers.info/coronavirus/country/us/
[40] J. A. Backer, D. Klinkenberg, and J. Wallinga, "Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20–28 January 2020," *Eurosurveillance*, vol. 25, no. 5, Feb. 2020, Art. no. 2000062.