

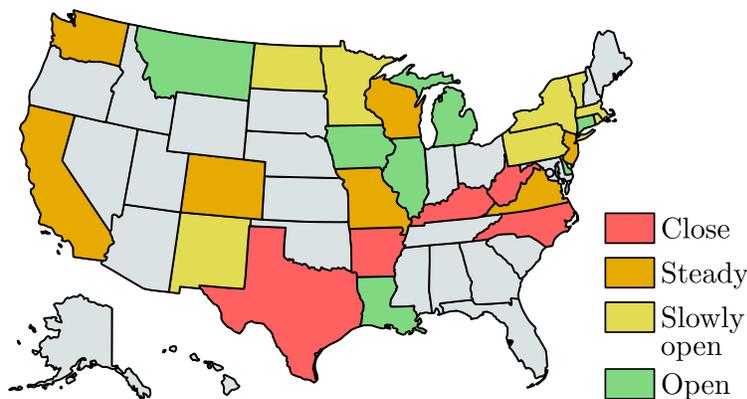
# Safety-Critical Control of Active Interventions for COVID-19 Mitigation

Aaron D. Ames<sup>1</sup>, Tamás G. Molnár<sup>2</sup>, Andrew W. Singletary<sup>1</sup> & Gábor Orosz<sup>2</sup>

<sup>1</sup>California Institute of Technology

<sup>2</sup>University of Michigan, Ann Arbor

The world has recently undergone the most ambitious mitigation effort in a century<sup>1</sup>, consisting of wide-spread quarantines aimed at preventing the spread of COVID-19<sup>2</sup>. The use of influential epidemiological models<sup>3-6</sup> of COVID-19 helped to encourage decision makers to take drastic non-pharmaceutical interventions. Yet, inherent in these models are often assumptions that the active interventions are static, e.g., that social distancing is enforced until infections are minimized, which can lead to inaccurate predictions that are ever evolving as new data is assimilated. We present a methodology to dynamically guide the active intervention by shifting the focus from viewing epidemiological models as systems that evolve in autonomous fashion to control systems with an “input” that can be varied in time in order to change the evolution of the system. We show that a safety-critical control approach<sup>7</sup> to COVID-19 mitigation gives active intervention policies that formally guarantee the safe evolution of compartmental epidemiological models. This perspective is applied to current US data on cases while taking into account reduction of mobility, and we find that it accurately describes the current trends when time delays<sup>8</sup> associated with incubation and testing are incorporated. Optimal active intervention policies are synthesized to determine future mitigations necessary to bound infections, hospitalizations, and death, both at national and state levels. We therefore provide means in which to model and modulate active interventions with a view toward the phased reopenings that are currently beginning across the US and the world in a decentralized fashion. This framework can be converted into public policies, accounting for the fractured landscape of COVID-19 mitigation in a safety-critical fashion.



**Figure 0.** Illustration of the safety-critical active intervention policies developed in this paper applied at the state level (for states with sufficient data). The states are colored according to whether it is safe to open further (green), slowly open (yellow) hold the current mitigation efforts steady (orange), or increase mitigation (red). This is determined based upon an active intervention policy that formally guarantees bounded hospitalizations and deaths.

**Introduction.** As COVID-19 spreads throughout the world<sup>9–11</sup>, due to the novelty of the virus and the resulting lack of pharmaceutical options necessary to suppress infection<sup>12</sup>, unprecedented mitigation steps to slow its progression were taken in the form of non-pharmaceutical interventions<sup>3,13</sup>, e.g., social distancing, mask-wearing, quarantining, and stay-at-home orders. It is largely agreed upon that these slowed the spread of the virus<sup>2,14</sup>, thereby saving lives. Yet studies have shown that if these active interventions had been enforced even a week earlier<sup>15</sup>, the result would have been a substantial reduction in deaths. As a means of mitigating the spread of COVID-19, the question therefore becomes: when, where, and how does one decide to take non-pharmaceutical interventions? This question is especially relevant<sup>16</sup> as restrictions are being relaxed in a decentralized fashion across the US and throughout the world.

Due to the pressing need to understand past and future mitigation efforts, and the corresponding role of active interventions, there has been a surge of recent papers on the modeling of COVID-19<sup>5,17–20</sup>. Epidemiological models for predicting the spread of COVID-19 often utilize dynamical systems obtained from so-called “compartmental” models wherein the compartments are chosen to reflect different populations of interest<sup>21–23</sup>, e.g., susceptible (*S*), infected (*I*), recovered (*R*), etc. More compartments can be added allowing for higher fidelity models, although one must be careful of overfitting the largely increased number of parameters in more complex models. The most fundamental (and elementary) of these compartmental models is the SIR model, which has recently been used in modeling of COVID-19<sup>24</sup>. Examples of more complex compartmental models applied for COVID-19 include the SEIR<sup>25,26</sup> and SIRT<sup>27</sup> models, which involve exposed (*E*) and threatened (*T*) populations, and the SIDARTHE model<sup>5</sup> which adds even more compartments. While these models have been found to be useful when modeling the spread of COVID-19 and the corresponding mitigation procedures, e.g., stay-at-home orders, the approaches are fundamentally based on autonomous dynamics<sup>28,29</sup> as they do not have a time-varying *control input* that can dynamically change the evolution of the system. We propose a different approach: applying *safety-critical control* methods to guide active non-pharmaceutical interventions wherein we can actively predict the interventions needed to maintain safety by viewing compartmental models as *control systems*.

| <b>Background</b>  | <b>Main Findings</b>  | <b>Policy Implications</b>   |
|--|---|--|
| Epidemiological models provide a powerful tool to guide the mitigation of COVID-19. Yet these models are dynamical systems that must be constantly updated as new data is assimilated. Policy decisions, therefore, fail to account for future active interventions, e.g., social distancing and stay-at-home orders, that can change the evolution of these models. | Viewing epidemiological models as control systems allows for the design of active intervention policies that can mitigate COVID-19 while modulating these mitigation efforts as function of time. With this viewpoint, safety-critical policies are synthesized that guarantee safety of the system by bounding the infected, hospitalized, and deceased populations. | The safety-critical active interventions synthesized can be used to formulate and inform public and governmental policy on lifting or increasing mitigation efforts in a centralized and decentralized fashion. These policies may, therefore, guide mitigation efforts at local or national levels to ensure hospitals do not reach capacity, and overall deaths are limited. |

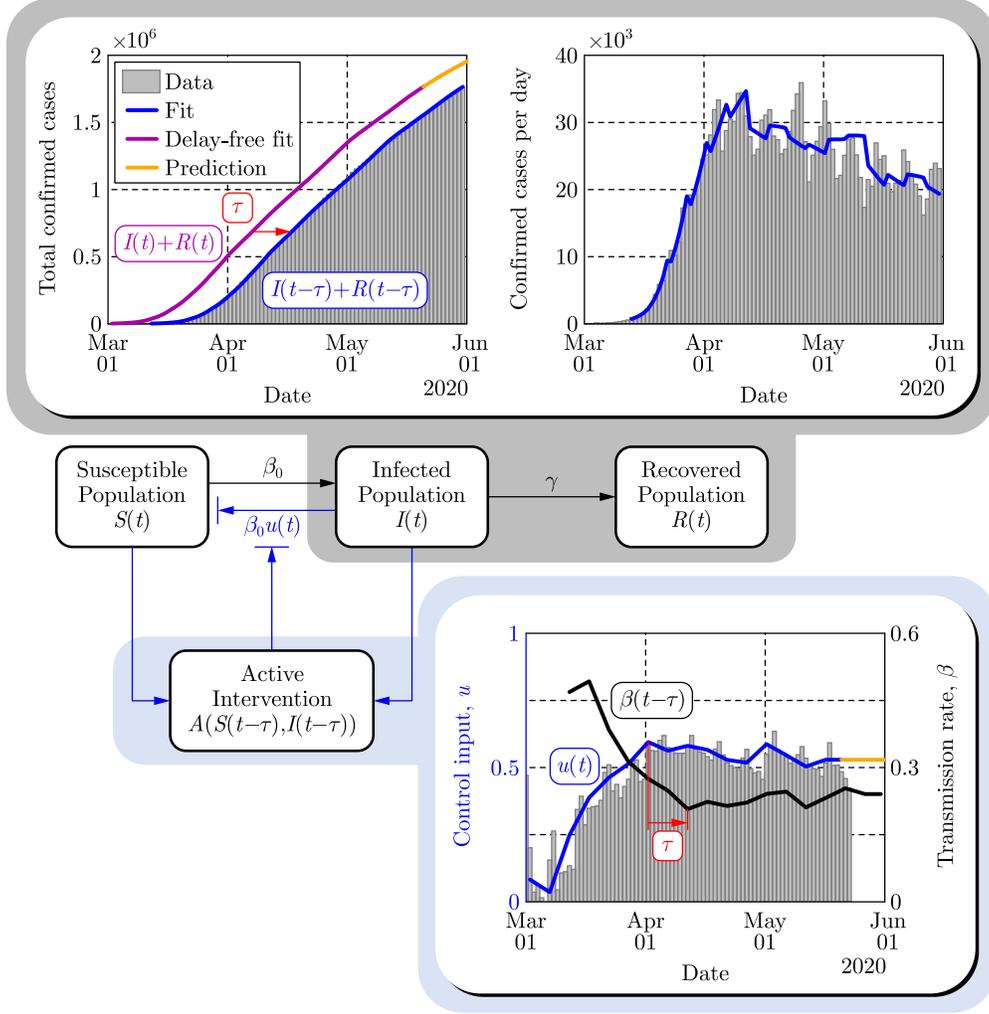
**Table 1.** Policy summary.

**SIR model as a Control System.** At the core of our approach is a fundamental shift in perspective on epidemiological models: from viewing them as dynamical systems that evolve in an autonomous fashion, to that of control systems for which the evolution can be dynamically modified. In many ways, this is the *de facto* manner in which these models are implemented, if only in an implicit fashion, as they are constantly updated as new data is assimilated, e.g., as changes in social distancing are observed, predictive models are updated<sup>30</sup>. We, therefore, will formalize this perspective by making the control aspect of epidemiological models explicit. Note that viewing compartmental epidemiological models as control systems is not unique<sup>31,32</sup>, but has found only limited application to COVID-19<sup>33</sup> and has yet to enjoy formal guarantees on safety. Additionally, there are examples of control-theoretic concepts being applied, namely in the the context of time-varying<sup>27,34</sup> and state-varying<sup>6,35</sup> choices of the transmission rate; these can be viewed as time- and state-varying inputs to a control system. Our approach differs in that we wish to synthesize *active intervention policies* (i.e., feedback control laws) that will determine future actions to take based upon past observations of the states of the systems.

To motivate the methodology utilized throughout this paper, we will begin by considering the fundamental epidemiological compartmental model: the SIR model<sup>21,23</sup>. Importantly, the approach introduced herein can be applied to *any* compartmental model, and will subsequently be applied to a more descriptive model. The SIR model consists of a *susceptible* population  $S$ , *infected* population  $I$ , and *recovered* population  $R$ . We can view the evolution of these populations as a control system where active interventions, expressed by the control input  $u(t)$ , modulate the rate of change of the infected population:

$$\begin{aligned}
 \dot{S}(t) &= -\frac{\beta_0}{N} \left( 1 - \underbrace{u(t)}_{\text{Control Input}} \right) S(t)I(t), \\
 \dot{I}(t) &= \frac{\beta_0}{N} \left( 1 - \underbrace{u(t)}_{\text{Control Input}} \right) S(t)I(t) - \gamma I(t), \\
 \dot{R}(t) &= \gamma I(t).
 \end{aligned} \tag{1}$$

Here the total population  $N = S(t) + I(t) + R(t)$  is constant,  $\beta_0 > 0$  is the transmission rate (when no intervention is present) and  $\gamma > 0$  is the recovery rate, yielding the reproduction number:  $R_0 = \beta_0/\gamma$ . This model relates to the traditional SIR model via the time-varying transmission rate  $\beta(t) = \beta_0(1 - u(t))$ . Time-varying  $\beta(t)$  has been considered<sup>20</sup>; for example, we can utilize the policy  $u(t) = -A \cos(\omega t)$  in the SIR model<sup>36</sup> to recover models of seasonal variations in infection<sup>37</sup>. In the setting considered here, taking  $u(t) \equiv 0$  corresponds to *no* intervention, yielding the traditional SIR model with  $\beta(t) \equiv \beta_0$ , whereas  $u(t) \equiv 1$  can be viewed as *maximum* intervention, full and complete quarantine of the population. In the latter case the infected population decays to zero exponentially,  $I(t) = e^{-\gamma t} I(0)$ , since the susceptible population is isolated. These effects can be seen, for example, in the Chinese response to COVID-19 and the corresponding drop in  $R_0$  documented<sup>19</sup>.



**Figure 1.** Paradigm shift wherein compartmental models are viewed as control systems rather than dynamical systems. This is illustrated on the populations  $I(t)$  and  $R(t)$  of the SIR model (top panels) wherein the control input  $u(t)$  is modulated based upon the intervention policies estimated from mobility data (bottom panel). The time delay  $\tau = 10$  days is highlighted to emphasize that the observed data corresponds to the delayed counterparts of the populations, and this delay also appears in the active intervention policy:  $u(t) = A(S(t - \tau), I(t - \tau))$ , given in Eq. (2).

An illustration of the SIR model as a control system is shown in Fig. 1 where the interactions between the compartments are denoted by arrows with appropriate rate constants indicated. The blue arrow represents the time dependent modulation of the transmission rate  $\beta(t)$ . The control input  $u(t)$  is estimated from mobility data<sup>38</sup> in the US between March 2 and May 20, 2020 by assuming  $u = 0$  at the beginning this period when no non-pharmaceutical interventions were present. Furthermore, the parameters  $\beta_0$ ,  $\gamma$  and  $N$  of the SIR model were fitted to the recorded number of confirmed cases:  $I(t - \tau) + R(t - \tau)$ . Fitting for the time delay  $\tau$ , in the corresponding transmission rate  $\beta(t - \tau)$ , reveals that the COVID-19 data<sup>39</sup> depicted publicly<sup>40</sup> are delayed by  $\tau \approx 10$  days. This time delay originates from the incubation time of the virus (i.e., people are being infectious before being symptomatic) and the time needed for testing<sup>33,41,42</sup>. That is, the

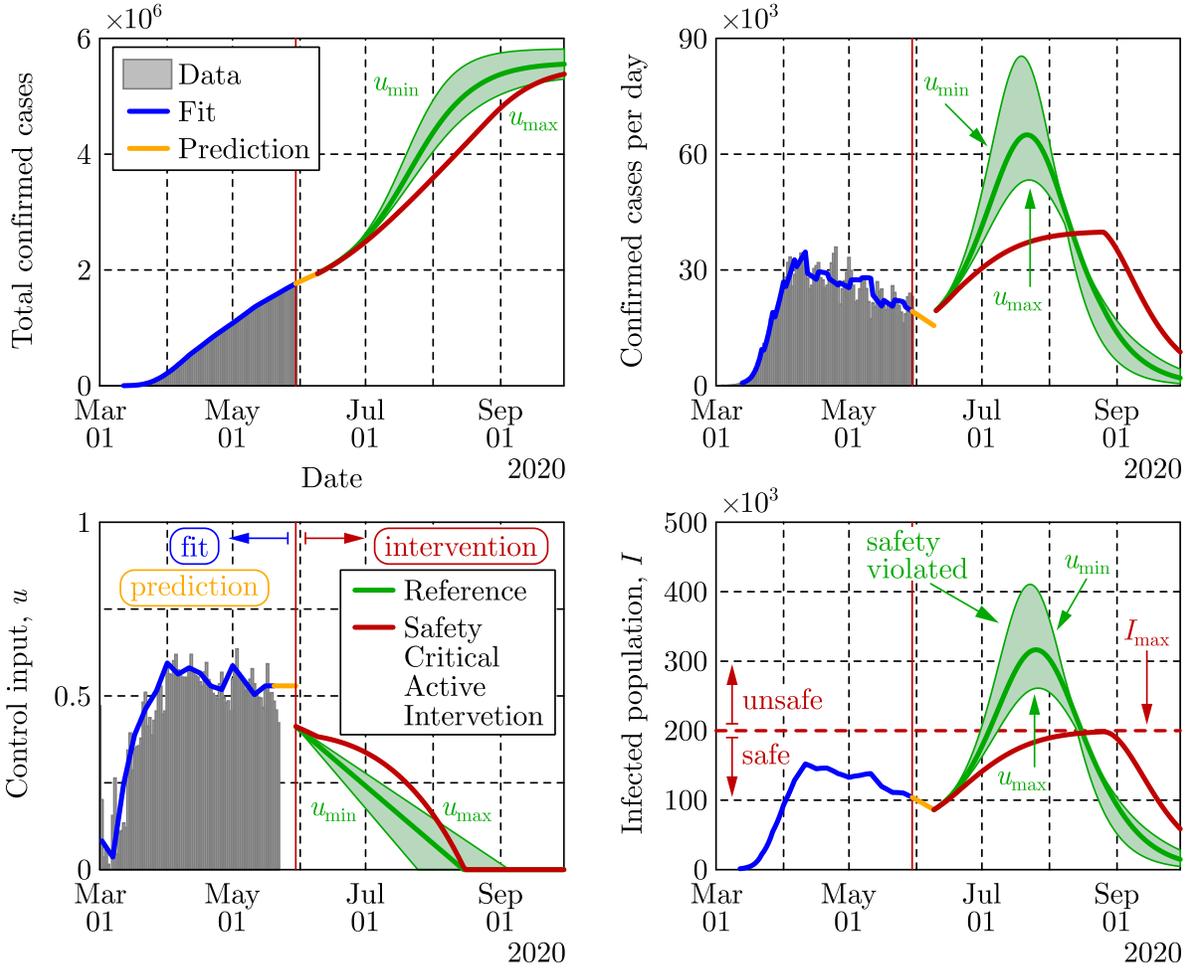
data corresponds the number of confirmed cases  $\tau$  days ago while the real current number could be much higher. For example, in mid-March, when interventions were introduced in the US,  $I(t - \tau) + R(t - \tau)$  was reported to be in the range of a few thousand, while the real number  $I(t) + R(t)$  is estimated to be more than a hundred thousand. This delay also appears in the active intervention policies which depend on the state of the system, i.e.,  $u(t) = A(S(t - \tau), I(t - \tau))$ , and it therefore must be compensated for in order to ensure the safety of these policies. Finally, we remark that when fitting the model (1) to the aforementioned data one may obtain good fits while setting  $N$  in the range from 7.5 million up to 330 million (see the Methods section for additional details). Smaller values encode the fact that not everyone susceptible is necessarily exposed when the total number of infected is small relative to the total population, as well as the fact that the total number of infections is underreported<sup>43</sup>. In Fig. 1 we used the lowest value  $N = 7.5$  million; the consequences of this choice will be discussed in the context of active interventions.

**Safety-Critical Control for Active Intervention.** Utilizing the paradigm of epidemiological models as control systems, we can synthesize active intervention policies, i.e., inputs to Eq. (1) expressed as functions of the populations of the compartmental model. A special case of this is referred to as shield immunity<sup>6</sup>, wherein the policy  $u(t) = \frac{\alpha R(t)}{N + \alpha R(t)}$  with  $\alpha \geq 0$  was chosen. Our goal is to synthesize active intervention policies so as to achieve desired safety-critical behaviors, that is, to guarantee that the system, with the policy applied, evolves in a safe fashion. Concretely, we may quantify safety in the context of the SIR model as limiting the total number of infected persons:  $I(t) \leq I_{\max}$ . To achieve such goal, we leverage the framework of control barrier functions<sup>7</sup> which gives necessary and sufficient conditions on the safety, along with tools to generate active intervention policies that ensure safety.

While there may exist multiple safe policies, it is beneficial to chose one which minimizes the active intervention  $u(t)$ , since more aggressive interventions potentially result in the lose of jobs and other economic and physiological effects<sup>44,45</sup>. The active intervention policy, i.e., feedback control law, that gives the minimal possible (pointwise optimal) interventions so as to ensure the safety of the system can be explicitly calculated (as described in Methods):

$$u(t) = A(S(t), I(t)) := \max \left\{ 0, 1 - \frac{\gamma}{\beta_0} \frac{N}{S(t)} \frac{I_{\max}}{I(t)} \right\} \quad \Rightarrow \quad I(t) \leq I_{\max}. \quad (2)$$

Notice the *activation function* or *rectified linear unit*<sup>46</sup>  $\text{ReLU}(x) = \max\{0, x\}$  can be used to express the policy; notably, this also used in neural networks in the context of machine learning<sup>47</sup>. This highlights that interventions only become “active” when safety is in danger of being violated. However, if one simply uses the obtained feedback control law in the SIR model with time delay  $\tau$ , i.e., substitutes  $u(t) = A(S(t - \tau), I(t - \tau))$  into Eq. (1), safety cannot be ensured due to the delay. In order to compensate for this delay we construct predictors<sup>48</sup> (as described in Methods) and use the predicted states  $S_p(t)$  and  $I_p(t)$  in the active intervention policy:  $u(t) = A(S_p(t), I_p(t))$ . If the predictions are accurate, i.e.,  $S_p(t) = S(t)$  and  $I_p(t) = I(t)$ , then the delay-free control design can ensure safety. Such predictors play an essential role in making the active intervention policies, synthesized from control barrier functions, implementable in the presence of time delay<sup>8</sup>.



**Figure 2.** Application of the safety-critical active intervention policy in Eq. (2) that keeps the number of infected people under a given limit  $I_{max}$ , to the SIR model in Eq. (1) with the parameters that yielded Fig. 1. The safety-critical policy is compared against a reference policy where the control input is reduced linearly. The epidemic ends relatively early due to the reduced population  $N$ , used in the model.

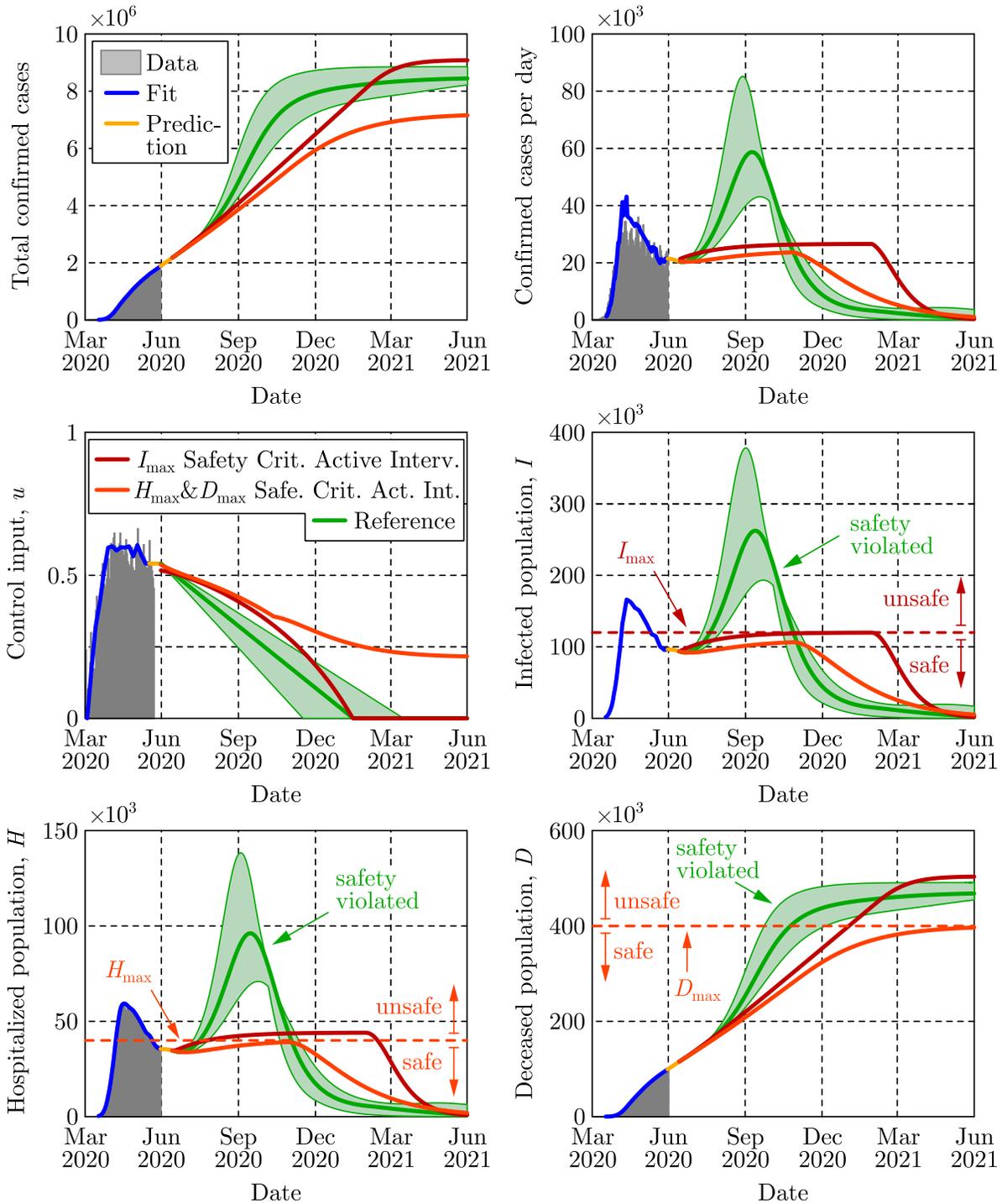
Figure 2 depicts the results of applying the safety-critical active intervention policy in Eq. (2) to the SIR model in Eq. (1) while compensating for the 10 days delay using predictors. The control barrier function is able to keep the infected population under  $I_{max} = 200,000$  while gradually driving the control input (active intervention) to zero, i.e., mitigation methods can eventually be removed. Notice that this opening strategy decreases the control input very slowly at the beginning followed by a faster opening toward the end. As a reference we also show the results of another opening strategy where the control input is reduced to zero linearly in time. In this case the number of infections peaks at a much higher value putting a large burden on the health system. While this figure vividly illustrates the use of safety-critical active intervention, and the benefits thereof, it also predicts that all restrictions can be lifted by mid-July. This is due to the use of the simplified SIR model that was considered to illustrate the concepts presented and, more specifically, due to the fact that the model heavily depends on the  $N$  (chosen to be 7.5 million when fitting the data). Selecting

a larger  $N$  would yield a longer mitigation period: the time period where active intervention is necessary, i.e., where Eq. (2) is non-zero, can be calculated as  $T \approx \frac{N}{\beta_0 I_{\max}} \left( \frac{\beta_0 S_0}{\gamma N} - 1 \right)$ , where  $S_0$  is the size of the susceptible population when the controller in Eq. (2) is initiated. Increasing  $N$  increases the period for which active intervention is necessary i.e., when the safety critical intervention policy is applied to the overly simplistic SIR model. In order to make predictions more reliable it is necessary to use a higher fidelity compartmental model. Moreover, doing so allows for additional safety-critical constraints to be considered, including hospitalization and death.

**Safety-Critical Active Interventions for the SIHRD Model.** The safety-critical approach to active intervention can be applied to more complex compartmental models, viewed as control systems. To better capture other salient populations for which safety is critical, we consider the SIHRD model (detailed in Methods) which includes the  $S$ ,  $I$  and  $R$  populations of the SIR model together with hospitalized and deceased populations,  $H$  and  $D$ , respectively<sup>20,23</sup>. The equations governing this model are, therefore, similar to those in Eq. (1) with the addition of dynamics governing the evolution of populations associated with hospitalization and deaths. Correspondingly, the control input again appears via the time varying transmission rate  $\beta(t) = \beta_0(1 - u(t))$ , while  $\gamma$  still denotes the recovery rate of the infected population. The additional parameters  $\lambda > 0$ ,  $\nu > 0$  and  $\mu > 0$  represent the hospitalization rate, recovery rate in hospitals and death rate, respectively. These rates are obtained by fitting the model to the data together with the effective population  $N$  that becomes 13.2 million for this model (as discussed in Methods).

The evolution of the SIHRD model is shown in Fig. 3 relative to US data, including mobility data, where the fits accurately capture the data for the infected, hospitalized and deceased populations to present day (the predictive power of this model is illustrated in Methods). Also illustrated in Fig. 3 are policies that allow for future mitigation designed using the safety-critical paradigm. Safety-critical active intervention policies can be synthesized for the SIHRD model, wherein the additional compartments allow for the consideration of safety constraints aimed at limiting hospitalization and death. In particular, we will consider two active interventions policies: one policy analogous to Eq. (2) aimed at limiting the infected population, and another policy aimed at simultaneously limiting both the number of hospitalized and dead. The results of applying these two policies are shown in Fig. 3, with the specific controllers detailed in Methods. Note, additional policies could be considered, i.e., ones bounding the populations in any compartment, or any combination thereof.

The first safety critical policy considered aims to limit the number of infected, i.e.,  $I(t) \leq I_{\max}$ , with results qualitatively similar to those of the SIR model in Fig. 2. Again mitigation measures are enforced over the same duration as a linear “opening up” policy while the optimality of the safety-critical policy results in substantially fewer infections at the peak. The second safety critical policy aims to limit hospitalizations ( $H(t) \leq H_{\max}$ ) based upon hospital capacity, while simultaneously limiting deaths ( $D(t) \leq D_{\max}$ ). Achieving these objectives, as indicated in Fig. 3, requires maintaining a non-zero input for a longer duration, i.e., some form of mitigation must be practiced for an extended period to limit overall death. This reflects the practices of countries

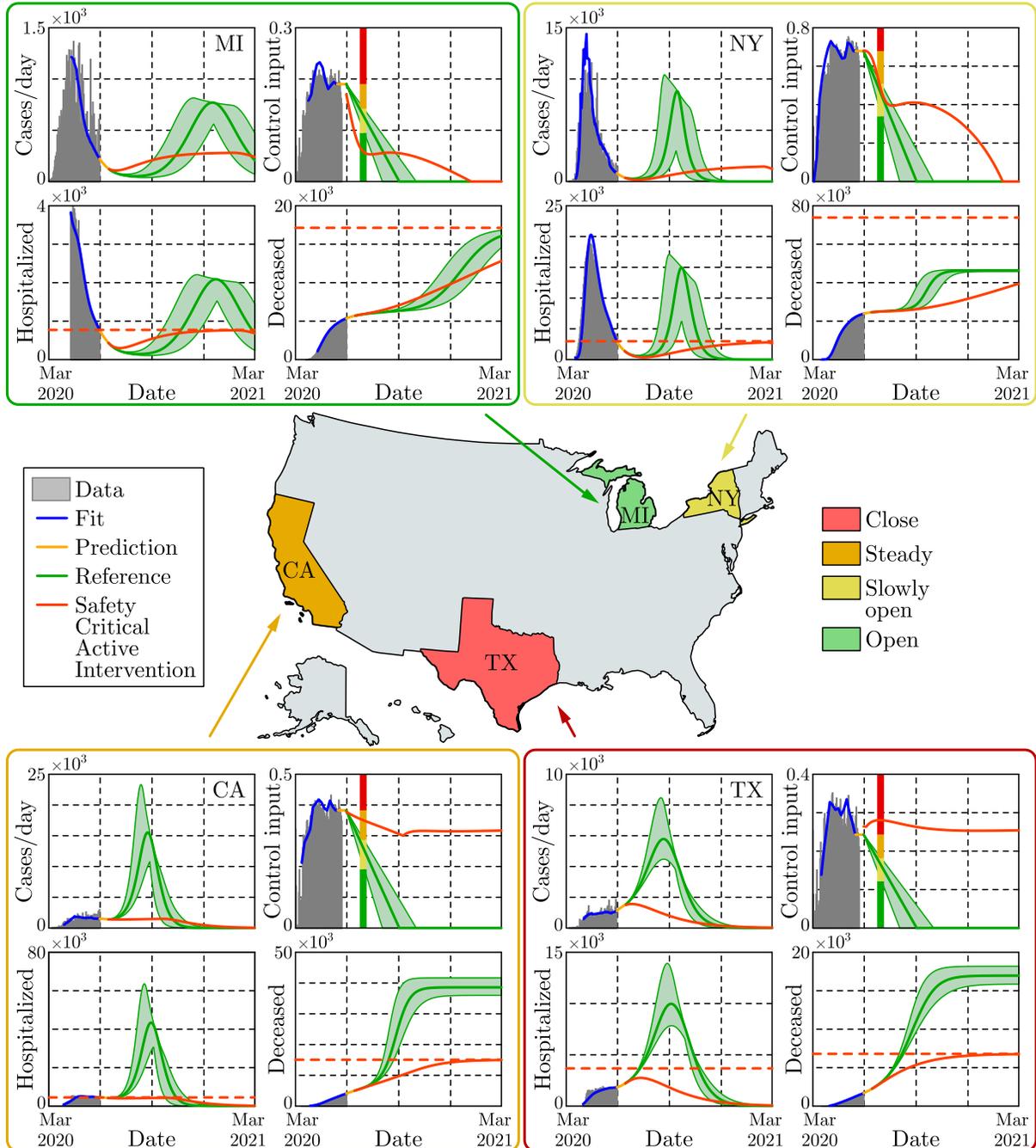


**Figure 3.** Two safety-critical active intervention policies applied to the SIHRD model. The red policy keeps the number of infected under  $I_{\max}$  as in Fig. 2 while the dark orange policy keeps the number of hospitalized under  $H_{\max}$  and also keeps the number of deaths under  $D_{\max}$ . The reference policy, that is linear in time, fails to maintain safety and results in a spike in infections and hospitalizations.

that successfully mitigated the first wave of the epidemic<sup>49</sup>. Importantly, both of the synthesized safety-critical active intervention policies guarantee the safety constraints while simultaneously minimizing mitigation—compared against the naive linear reference policy which would drive the number of hospitalized above the limit  $H_{\max}$ , and result in large number of deceased persons. This indicates the important role that active intervention policies can play in guaranteeing safety, encoded by limiting hospitalizations and deaths.

The safety-critical policies synthesized above can also be applied to smaller geographical areas. This is especially relevant from a practical perspective, as specific mitigation efforts are determined at a state level in the US. In Fig. 4, the results are shown for four different states with safety-critical active intervention policies simultaneously bounding hospitalization and death; the safety bounds  $H_{\max}$  and  $D_{\max}$  were chosen as outlined in Methods, and different bounds can be used based upon state-level public policy. Different states require different levels of mitigation as highlighted by the color of each state. The gating criterion for state level mitigation was, as a proof of concept, determined by the value of the safety-critical control input 30 days after the start of active intervention; other criterion could be used based upon public policy. In this case, Michigan may open up, i.e., relax its mitigation efforts relatively quickly, reducing the control input to less than 50% of its current value in 30 days, yet mitigation efforts must be kept in place throughout the year. Qualitatively similar behavior can be seen in the case of New York, though active interventions cannot be reduced as quickly—if relaxed too quickly the result is a second spike in infections equal to the first already experienced. By comparison, California needs to very slowly relax its mitigation efforts and settle into a steady state mitigation at 80% of its current value, or the result is an outbreak with very high number of hospitalized and substantially more death. Texas should increase its current mitigation efforts to avoid a sudden and significant rise of infections, hospitalizations and death. In the case of both California and Texas, the way in which they open has a profound effect on the total hospitalizations and deaths, with deaths more than doubling if a naive opening up policy is implemented. Therefore, the safety-critical approach can determine the optimal way in which states should open—assuming good data at the state level—thereby informing policy that has the potential to dramatically reduce hospitalizations and deaths.

**Summary.** The approach taken in this paper revolves around a new paradigm: viewing compartmental epidemiological models as control systems, viz. Eq. (1). Importantly, this perspective allows one to view these models not as systems that evolve independent of human behavior, but rather as systems where human behavior is an input that can *actively* modify their evolution (cf. Fig. 1). In this setting, we are able to synthesize active intervention policies that can serve to guide future mitigation efforts. We specifically synthesized safety-critical policies that formally guarantee that the evolution of compartmental models—the SIR and SIHRD—stay within “safe sets.” These safe sets encode bounds on the number of infected, hospitalized, and deceased populations. Closed form expressions for optimal active intervention policies were synthesized, as in Eq. (2), that ensure safety. To demonstrate this approach, US COVID-19 data on cases, hospitalizations and deaths were utilized to fit the static parameters of the SIR and SIHRD models. The active component of the control system, i.e., the control input, was synthesized utilizing mobility data;



**Figure 4.** Safety-critical active interventions at the state level for four states: California, Michigan, New York, and Texas. The SIHRD model, viewed as a control system, was fit to the data for each state through May 30, 2020. From this, safety-critical active intervention policies that simultaneously bound hospitalizations and deaths are synthesized. The color of each state is determined by the control input 30 days after the start of the safety-critical active interventions, as indicated by the vertical line in the control input plots. The safety-critical policy is compared against the naive linear opening up reference policy which violates the safety bounds—resulting in over twice the deaths in the at risks states: California and Texas. This illustrates that the way in which states open up has important ramifications.

the result was models with predictive power. Projecting into the future while compensating for the incubation and testing delays, the active intervention policies were applied and compared against “naive opening up” policies. It was shown that the safety-critical policies that limit hospitalizations and deaths greatly outperformed these reference policies (Fig. 2), and this was demonstrated at both the national (Fig. 3) and the state level (Fig. 4).

**Policy Implications.** The safety-critical approach to active intervention can directly inform public policy. To wit, the results presented demonstrate that epidemiological models (viewed as control systems) can capture the role of human action in mitigating COVID-19; both to describe observed data, and to actively modulate future behavior. Active intervention policies (feedback control laws) can, therefore, be used to guide non-pharmaceutical actions that should be taken to achieve a desired outcome with regard to the COVID-19 pandemic—or unforeseen future pandemics. Of particular concern are mitigation efforts devoted to ensuring safety; this encodes the desire to limit the infected, hospitalized and deceased population. The safety-critical active intervention policy presented herein results in concrete guidance on future mitigation efforts needed to achieve these guarantees. These actions can be at a local, state, national or international level depending on the ability to guide active interventions among these populations. The end result can be codified in tangible and specific public policies on “opening up”, i.e., on lifting or increasing mitigation efforts. As demonstrated throughout this paper on COVID-19 data and the corresponding epidemiological models, safety-critical active interventions—if properly encoded as public policy—have the ability to ensure available hospital capacity and save lives.

## Methods

**Safety-Critical Control for Guaranteed Safety.** Safety can be framed as set invariance<sup>50–52</sup> in the context of control systems and controller synthesis. Let  $\mathcal{R}^n$  be the state space of the compartmental model of interest, consisting of  $n$ -dimensional Euclidean space, with  $n$  the number of compartments, i.e., for the SIR model  $n = 3$  and for the SIHRD model  $n = 5$ . A state  $\mathbf{x} \in \mathcal{R}^n$  consists of values of the populations, e.g.,  $\mathbf{x} = [S, I, R]^\top$  for the SIR model. A safety constraint is a function  $h : \mathcal{R}^n \rightarrow \mathcal{R}$  that encodes the safe behavior of the system through:

$$\text{Safe set: } \quad \mathcal{C} := \{\mathbf{x} \in \mathcal{R}^n : h(\mathbf{x}) \geq 0\}, \quad (3)$$

wherein the goal is for the system to evolve in this safe set. For example, for the SIR model  $h(S, I, R) = I_{\max} - I$ , with the set  $\mathcal{C}$  containing the states for which  $I \leq I_{\max}$ . The goal is to give (necessary and sufficient) conditions for control systems, and synthesize corresponding policies, that render this set forward invariant, i.e., that keep the system safe.

A *control system* (in control affine form) is a first order nonlinear differential equation with a control input:

$$\dot{\mathbf{x}}(t) = f(\mathbf{x}(t)) + g(\mathbf{x}(t)) \underbrace{u(t)}_{\text{Control Input}}, \quad (4)$$

where  $\mathbf{x} \in \mathcal{R}^n$  and  $u \in \mathcal{R}$  is the scalar valued control input (note that all of the methods presented also hold for vector valued control inputs). All compartmental models can be expressed in the general form of Eq. (4); which becomes an autonomous dynamical system (as they are typically modeled) for  $u(t) \equiv 0$ , i.e., the system evolves according to  $\dot{\mathbf{x}}(t) = f(\mathbf{x}(t))$ . The addition of the control input  $u(t)$ , as was done in Eq. (1), allows one to modify the evolution of the system to achieve desired behaviors. This modification is done via *control laws* or *policies*:  $u(t) = K(\mathbf{x}(t))$ . The result is a *closed loop* dynamical system:  $\dot{\mathbf{x}}(t) = f(\mathbf{x}(t)) + g(\mathbf{x}(t))K(\mathbf{x}(t))$ , wherein  $\mathbf{x}(t)$  is a solution to this system with initial condition  $\mathbf{x}(0) = \mathbf{x}_0$ .

We are interested in guarantees of safety framed as set invariance per Eq. (3). Thus, we say that the control system in Eq. (4) is *safe* with the policy  $u(t) = K(\mathbf{x}(t))$  if  $\mathbf{x}_0 \in \mathcal{C}$  implies that  $\mathbf{x}(t) \in \mathcal{C}$  for all  $t \geq 0$ , where  $\mathbf{x}(t)$  is a solution to the closed loop system with the policy applied. By the definition of the safe set in Eq. (3), safety is thus equivalent to satisfying the safety constraint for all time:  $h(\mathbf{x}(t)) \geq 0$ . Safety-critical control addresses the fundamental question: *how does one synthesize control policies that render the set  $\mathcal{C}$  safe*, i.e., control policies such that safety constraint  $h(\mathbf{x}(t)) \geq 0$  is satisfied for the closed loop system? To achieve safe behavior for the control system in Eq. (4) representing an abstract compartmental model, we leverage the framework of control barrier functions<sup>7</sup>. This is a new methodology for controller synthesis that has its bases in a long and rich history of set invariance for dynamical systems<sup>50–52</sup>. In particular, by considering the function  $h(\mathbf{x})$  that defines the safe set  $\mathcal{C}$ , we wish to find conditions on the rate of change of this function that guarantee forward set invariance; conditions that can be checked over the entire set  $\mathcal{C}$  and thereby used to synthesize control policies.

It was discovered<sup>7</sup> that necessary<sup>1</sup> and sufficient conditions for forward set invariance are given by lower bounding the rate of change of  $h$  when differentiated along  $\mathbf{x}(t)$  with respect to time:

$$\frac{d}{dt}h(\mathbf{x}(t)) \geq -\alpha h(\mathbf{x}(t)) \quad \iff \quad \mathcal{C} \text{ is safe} \quad \iff \quad h(\mathbf{x}(t)) \geq 0, \quad (5)$$

for  $\alpha > 0$  and all  $t \geq 0$ . The importance of the derivative condition is that it can be checked at every point of time with respect to the input  $u$ . Thus,  $h$  is a *control barrier function*<sup>7</sup> (CBF) if there exists a  $u(t)$  such that:

$$\frac{d}{dt}h(\mathbf{x}(t)) = \dot{h}(\mathbf{x}(t), u(t)) = \underbrace{\frac{\partial h}{\partial \mathbf{x}} f(\mathbf{x}(t))}_{:=L_f h(\mathbf{x}(t))} + \underbrace{\frac{\partial h}{\partial \mathbf{x}} g(\mathbf{x}(t)) u(t)}_{:=L_g h(\mathbf{x}(t))} \geq -\alpha h(\mathbf{x}(t)). \quad (6)$$

As a result, for a control barrier function, one can synthesize a policy that ensures safety by choosing a controller  $u(t)$  that satisfies Eq. (6). For example, if  $L_g h(\mathbf{x}) \neq 0$  then  $h$  is a control barrier function as  $u(t)$  satisfying Eq. (6) can be explicitly solved for through the pseudoinverse. We seek to do this in an optimal way so as to minimize the amount of active intervention.

With the goal of achieving safety while minimizing the input—as is the case with compartmental epidemiological models where we wish to minimize the active intervention—the control law synthesis problem can be framed as an optimization problem; specifically, a quadratic program:

$$\begin{aligned} u(t) = K(\mathbf{x}(t)) = \underset{u \in [0,1]}{\operatorname{argmin}} \quad & u^2 \\ \text{s.t.} \quad & L_f h(\mathbf{x}(t)) + L_g h(\mathbf{x}(t))u \geq -\alpha h(\mathbf{x}(t)). \end{aligned} \quad (7)$$

Note that here we limit  $u \in [0, 1]$  since this corresponds to the interval of active interventions with  $u = 0$  denoting no intervention and  $u = 1$  denoting complete intervention, e.g., fully isolating the infected population. Importantly, one can explicitly solve the optimization problem in Eq. (7) to get a closed form expression:

$$u(t) = K(\mathbf{x}(t)) = \begin{cases} -\frac{L_f h(\mathbf{x}(t)) + \alpha h(\mathbf{x}(t))}{L_g h(\mathbf{x}(t))} & \text{if } L_f h(\mathbf{x}(t)) < -\alpha h(\mathbf{x}(t)) \\ 0 & \text{if } L_f h(\mathbf{x}(t)) \geq -\alpha h(\mathbf{x}(t)). \end{cases} \quad (8)$$

For this choice of control law, the closed loop system is safe and, additionally, the minimal input is optimally chosen. This is represented by the conditional statement, wherein  $u = 0$  if the natural dynamics of the system satisfy the control barrier function condition in Eq. (6).

---

<sup>1</sup>Technically, for necessity,  $\alpha$  must be chosen to be an extended class  $\mathcal{K}$  function<sup>7</sup> not a constant. We utilize a constant for simplicity of exposition and without loss of generality.

If  $L_g h(\mathbf{x}(t)) > 0$ , Eq. (8) simplifies to:

$$u(t) = K(\mathbf{x}(t)) = \max \left\{ 0, -\frac{L_f h(\mathbf{x}(t)) + \alpha h(\mathbf{x}(t))}{L_g h(\mathbf{x}(t))} \right\} = \text{ReLU} \left( -\frac{L_f h(\mathbf{x}(t)) + \alpha h(\mathbf{x}(t))}{L_g h(\mathbf{x}(t))} \right), \quad (9)$$

wherein the min becomes the max when  $L_g h(\mathbf{x}(t)) < 0$ . It is this formulation that leads to the active intervention policies that we will synthesize for both the SIR and SIHRD compartmental models, viewed as control systems.

**Application of Safety-Critical Methods to the SIR Model.** Consider the SIR model, viewed as a control system, as given in Eq. (1). This is clearly of the form of the general control system given in Eq. (4) wherein  $\mathbf{x} = [S, I, R]^\top \in \mathcal{R}^3$  and:

$$f(\mathbf{x}(t)) = \begin{bmatrix} -\frac{\beta_0}{N} S(t)I(t) \\ \frac{\beta_0}{N} S(t)I(t) - \gamma I(t) \\ \gamma I(t) \end{bmatrix}, \quad g(\mathbf{x}(t)) = \begin{bmatrix} \frac{\beta_0}{N} S(t)I(t) \\ -\frac{\beta_0}{N} S(t)I(t) \\ 0 \end{bmatrix}. \quad (10)$$

As a result,  $\dot{\mathbf{x}}(t) = f(\mathbf{x}(t))$  is just the standard SIR model—viewed as an autonomous dynamical system. As indicated above, the safety constraint  $I(t) \leq I_{\max}$  leads to the function  $h(I) = I_{\max} - I$  defining the safe set  $\mathcal{C} = \{[S, I, R]^\top \in \mathcal{R}^3 : I \leq I_{\max}\}$  as in Eq. (3). For the safety function  $h(I) = I_{\max} - I$  calculating Eq. (6) yields:

$$\dot{h}(S(t), I(t)) = -\dot{I}(t) = \underbrace{-\frac{\beta_0}{N} S(t)I(t) + \gamma I(t)}_{L_f h(S(t), I(t))} + \underbrace{\frac{\beta_0}{N} S(t)I(t)}_{L_g h(S(t), I(t))} u(t) \geq -\alpha(I_{\max} - I(t)). \quad (11)$$

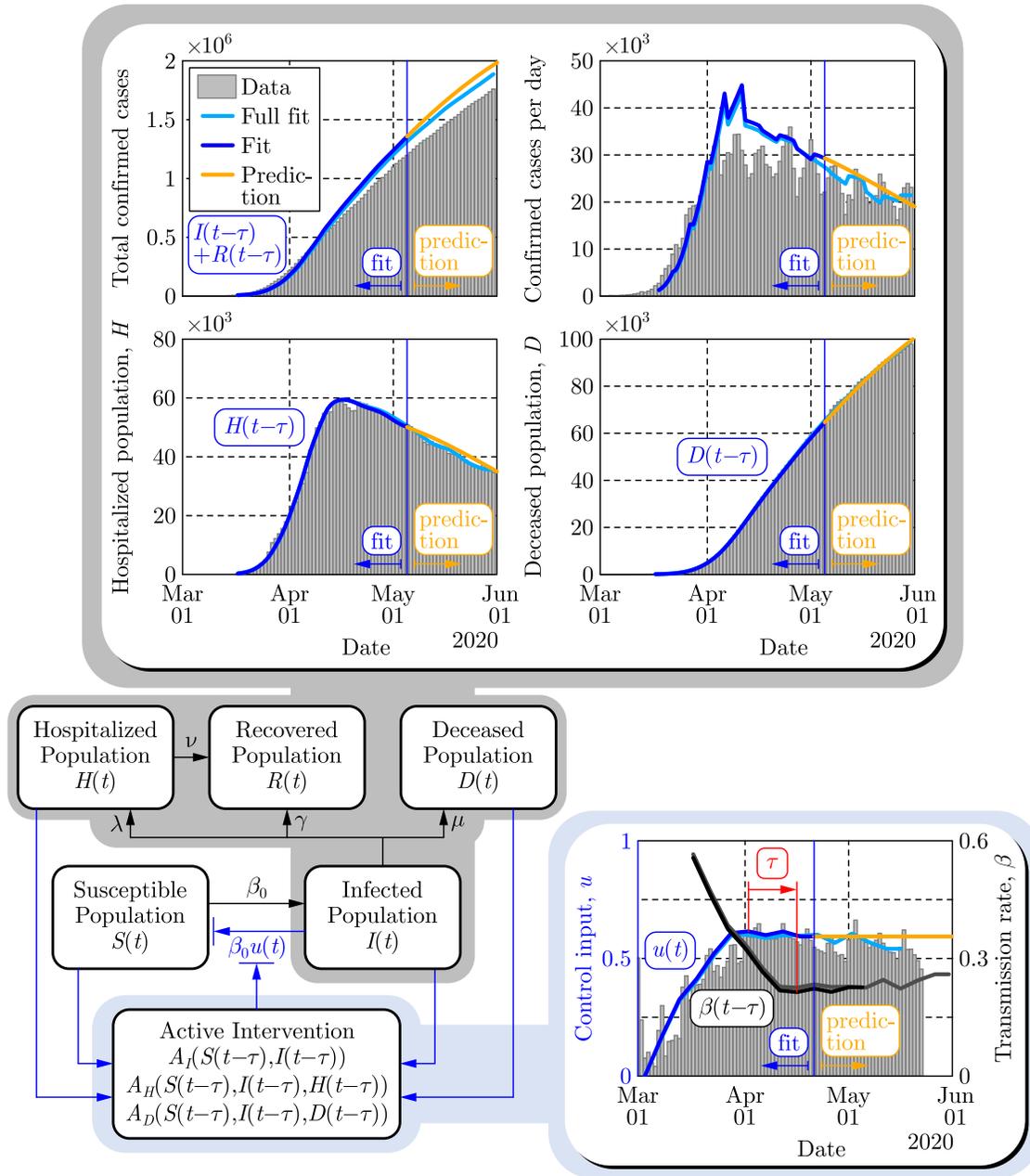
It follows that  $h$  is a control barrier function since  $I \neq 0$  and  $S \neq 0$  corresponds to having nonzero infected or susceptible populations, and therefore,  $L_g h(S(t), I(t)) \neq 0$ . The explicit solution in Eq. (9) to the optimization-based controller in Eq. (7) becomes:

$$u(t) = A(S(t), I(t)) = \text{ReLU} \left( 1 - \frac{\alpha(I_{\max} - I(t)) + \gamma I(t)}{\frac{\beta_0}{N} S(t)I(t)} \right) \Rightarrow I(t) \leq I_{\max}, \quad (12)$$

if  $I(0) \leq I_{\max}$ , since in the domain of interest  $S > 0, I > 0 \Rightarrow L_g h(S(t), I(t)) > 0$ . By picking  $\alpha = \gamma$ , the control law in Eq. (12) yields Eq. (2) which was used in Fig. 2.

**Safety-Critical Control Applied to the SIHRD Model.** The SIHRD model is a compartmental epidemiological model that extends the SIR model to include two additional compartments related to hospitalized and deceased populations. These additional compartments will be important in the synthesis of safety-critical controllers that bound these populations. The SIHRD model—viewed as a control system—is illustrated in Fig. 5, where  $S, I$  and  $R$  are the same populations as in the SIR model,  $H$  denotes the population that is currently hospitalized due to the virus (and assumed not to transmit to the susceptible population as a result), and  $D$  is the deceased population. The rate constants are indicated along the arrows linking the compartments:  $\beta_0$  is the transmission rate

while  $\gamma$  and  $\nu$  are the recovery rates of the infected and hospitalized populations, respectively. Additionally,  $\lambda$  represents the hospitalization rate and  $\mu$  is the mortality rate. These parameters are coupled via  $1/(\gamma + \lambda + \mu)$  which is the characteristic infectious period of the virus, accounting for hospitalizations and deaths, after which there is assumed to be no transmission.



**Figure 5.** Illustration of the predictive power of the SIHRD model. The model parameters are estimated using the US data up to May 5, 2020 (dark blue), shown as a vertical blue line, and then used to predict forward for 25 days until May 30, 2020 (yellow). These are compared to a fit where the data was used until May 30 (light blue).

When casting the model in the form of Eq. (4), we use  $\mathbf{x} = [S, I, H, R, D]^\top \in \mathcal{R}^5$  and obtain the following control system:

$$\underbrace{\begin{bmatrix} \dot{S}(t) \\ \dot{I}(t) \\ \dot{H}(t) \\ \dot{R}(t) \\ \dot{D}(t) \end{bmatrix}}_{\dot{\mathbf{x}}(t)} = \underbrace{\begin{bmatrix} -\frac{\beta_0}{N}S(t)I(t) \\ \frac{\beta_0}{N}S(t)I(t) - (\gamma + \lambda + \mu)I(t) \\ \lambda I(t) - \nu H(t) \\ \gamma I(t) + \nu H(t) \\ \mu I \end{bmatrix}}_{f(\mathbf{x}(t))} + \underbrace{\begin{bmatrix} \frac{\beta_0}{N}S(t)I(t) \\ -\frac{\beta_0}{N}S(t)I(t) \\ 0 \\ 0 \\ 0 \end{bmatrix}}_{g(\mathbf{x}(t))} \underbrace{u(t)}_{\text{Control Input}}. \quad (13)$$

As indicated in Fig. 5, one may design active intervention policies<sup>2</sup> for the control input  $u(t)$  that modulates the transmission rate:  $\beta(t) = \beta_0(1 - u(t))$ . In particular, we are interested in synthesizing safety-critical active intervention policies that bound infections, hospitalization and death for the SIHRD model. The corresponding safety functions are given by:

$$\begin{aligned} h_I(I) &:= I_{\max} - I, \\ h_H(H) &:= H_{\max} - H, \\ h_D(D) &:= D_{\max} - D, \end{aligned} \quad (14)$$

with corresponding safe sets  $\mathcal{C}_I$ ,  $\mathcal{C}_H$ , and  $\mathcal{C}_D$  defined as in Eq. (3). In the case of  $h_I$ , a similar calculation to that in Eq. (11) yields the active intervention policy (analogous to Eq. (12)):

$$u(t) = A_I(S(t), I(t)) = \text{ReLU} \left( 1 - \frac{\alpha_I(I_{\max} - I(t)) + (\gamma + \lambda + \mu)I(t)}{\frac{\beta_0}{N}S(t)I(t)} \right) \Rightarrow I(t) \leq I_{\max}, \quad (15)$$

assuming  $I(0) \leq I_{\max}$ , wherein we selected  $\alpha_I = (\gamma + \lambda + \mu)/10$  in Fig. 3.

For the safety functions,  $h_i$  for  $i \in \{H, D\}$ , associated with hospitalization and death, additional steps are needed to synthesize the active intervention policy. In particular, the input  $u(t)$  does not appear when differentiating these functions as was the case in Eq. (6). Yet, we know by Eq. (5) that sufficient conditions for the sets  $\mathcal{C}_i$  to be safe are given by  $\dot{h}_i + \alpha_i h_i \geq 0$  for  $i \in \{H, D\}$ , where now  $\dot{h}_i$  does not depend on the input  $u(t)$  as  $L_g h_i(\mathbf{x}) = 0$ . As a result, define the following *extended safety functions*<sup>50,53</sup>:

$$h_i^e(\mathbf{x}(t)) := \dot{h}_i(\mathbf{x}(t)) + \alpha_i h_i(\mathbf{x}(t)) = \underbrace{\frac{\partial h_i}{\partial \mathbf{x}} f(\mathbf{x}(t))}_{\dot{h}_i(\mathbf{x}(t))=L_f h_i(\mathbf{x}(t))} + \alpha_i h_i(\mathbf{x}(t)), \quad (16)$$

with associated safe sets:  $\mathcal{C}_i^e = \{\mathbf{x} \in \mathcal{R}^5 : h_i^e(\mathbf{x}) \geq 0\}$ . Importantly,  $h_i^e$  are now themselves control

<sup>2</sup>Note that the delayed values of the different populations appear in the corresponding feedback laws which will be compensated using predictors based on the fitted model. This will be described in detail later in Methods.

barrier functions, wherein the condition in Eq. (6) becomes:

$$\dot{h}_i^e(\mathbf{x}(t), u(t)) = \underbrace{\frac{\partial \dot{h}_i}{\partial \mathbf{x}} f(\mathbf{x}(t))}_{:=L_f^2 h_i(\mathbf{x}(t))} + \underbrace{\frac{\partial \dot{h}_i}{\partial \mathbf{x}} g(\mathbf{x}(t))}_{:=L_g L_f h_i(\mathbf{x}(t))} u(t) + \alpha_i \dot{h}_i(\mathbf{x}(t)) \geq -\alpha_i^e h_i^e(\mathbf{x}(t)), \quad (17)$$

with  $i \in \{H, D\}$ .

This allows us to synthesize optimal active intervention policies as in Eq. (7) via:

$$u(t) = A_i(\mathbf{x}(t)) = \underset{u \in [0,1]}{\operatorname{argmin}} \quad u^2 \quad (18)$$

s.t.  $L_f^2 h_i(\mathbf{x}(t)) + L_g L_f h_i(\mathbf{x}(t))u \geq -\alpha_i^e h_i^e(\mathbf{x}(t)) - \alpha_i \dot{h}_i(\mathbf{x}(t)),$

which can be converted into a conditional statement in the form of Eq. (8). Applying these constructions to the SIHRD model result in the optimal active intervention policies:

$$u(t) = A_H(S(t), I(t), H(t))$$

$$= \operatorname{ReLU} \left( 1 - \frac{\alpha_H \alpha_H^e (H_{\max} - H(t)) + (\nu - \alpha_H - \alpha_H^e)(\lambda I(t) - \nu H(t)) + (\gamma + \lambda + \mu)\lambda I(t)}{\lambda \frac{\beta_0}{N} S(t) I(t)} \right)$$

$$\Rightarrow H(t) \leq H_{\max} \quad (19)$$

assuming the initial condition satisfies<sup>3</sup>  $H(0) \leq H_{\max}$  and  $h_H^e(\mathbf{x}(0)) \geq 0$ , and

$$u(t) = A_D(S(t), I(t), D(t))$$

$$= \operatorname{ReLU} \left( 1 - \frac{\alpha_D \alpha_D^e (D_{\max} - D(t)) + (\gamma + \lambda + \mu - \alpha_D - \alpha_D^e)\mu I(t)}{\mu \frac{\beta_0}{N} S(t) I(t)} \right)$$

$$\Rightarrow D(t) \leq D_{\max}, \quad (20)$$

assuming  $D(0) \leq D_{\max}$  and  $h_D^e(\mathbf{x}(0)) \geq 0$ . We selected  $\alpha_H = (\gamma + \lambda + \mu)/10$ ,  $\alpha_H^e = \nu/10$  and  $\alpha_D = \alpha_D^e = (\gamma + \lambda + \mu)/10$  to generate Figs. 3 and 4 as described in the next section.

**Enforcing Multiple Safety Constraints Encoded as Control Barrier Functions.** In addition to enforcing safety constraints via individual barrier functions, we can simultaneously enforce multiple safety constraints. We will demonstrate this in the context of enforcing both the safety constraints associated with hospitalization and death,  $h_H \geq 0$  and  $h_D \geq 0$  as given in Eq. (14), for the SIHRD model. Note the same concepts apply if we wanted to simultaneously enforce  $h_I \geq 0$  or any combination of the constraints  $h_I \geq 0$ ,  $h_H \geq 0$  and  $h_D \geq 0$ . Similarly, these ideas can be applied to multiple safety constraints for more complex compartmental models, e.g., the SIDARTHE model<sup>5</sup>.

<sup>3</sup>For the extended control barrier functions<sup>50,53</sup>,  $\mathbf{x}(0) \in \mathcal{C} \cap \mathcal{C}^e$  implies that  $\mathbf{x}(t) \in \mathcal{C} \cap \mathcal{C}^e$  for all  $t \geq 0$ .

In order to limit the number of hospitalized and deceased populations in the SIHRD model, encoded by  $h_H(H) = H_{\max} - H \geq 0$  and  $h_D(D) = D_{\max} - D \geq 0$ , while minimizing the active intervention  $u$ , we consider the quadratic program (QP):

$$\begin{aligned} u(t) = A_{HD}(\mathbf{x}(t)) = \operatorname{argmin}_{u \in [0,1]} u^2 & \quad (21) \\ \text{s.t. } L_f^2 h_H(\mathbf{x}(t)) + L_g L_f h_H(\mathbf{x}(t))u & \geq -\alpha_H^e h_H^e(\mathbf{x}(t)) - \alpha_H \dot{h}_H(\mathbf{x}(t)) \\ L_f^2 h_D(\mathbf{x}(t)) + L_g L_f h_D(\mathbf{x}(t))u & \geq -\alpha_D^e h_D^e(\mathbf{x}(t)) - \alpha_D \dot{h}_D(\mathbf{x}(t)). \end{aligned}$$

Here, we simultaneously enforce the (extended) barrier function condition in Eq. (17) for the functions  $h_H^e$  and  $h_D^e$  that imply satisfaction of  $\dot{h}_H \geq -\alpha_H h_H$  and  $\dot{h}_D \geq -\alpha_D h_D$  as desired. In general, it is not guaranteed that a QP with multiple constraints is feasible without a relaxation term<sup>7</sup> but in this case due to the special structure of the control barrier functions considered, a solution can be guaranteed and stated in closed form.

To see this, we begin by noting that for  $h_i, i \in \{I, H, D\}$ , defined as in Eq. (14):

$$\frac{1}{\lambda} L_g L_f h_H(\mathbf{x}(t)) = \frac{1}{\mu} L_g L_f h_D(\mathbf{x}(t)) = L_g h_I(\mathbf{x}(t)). \quad (22)$$

Therefore, defining

$$\Phi_i(\mathbf{x}(t)) := L_f^2 h_i(\mathbf{x}(t)) + \alpha_i^e h_i^e(\mathbf{x}(t)) + \alpha_i \dot{h}_i(\mathbf{x}(t)), \quad (23)$$

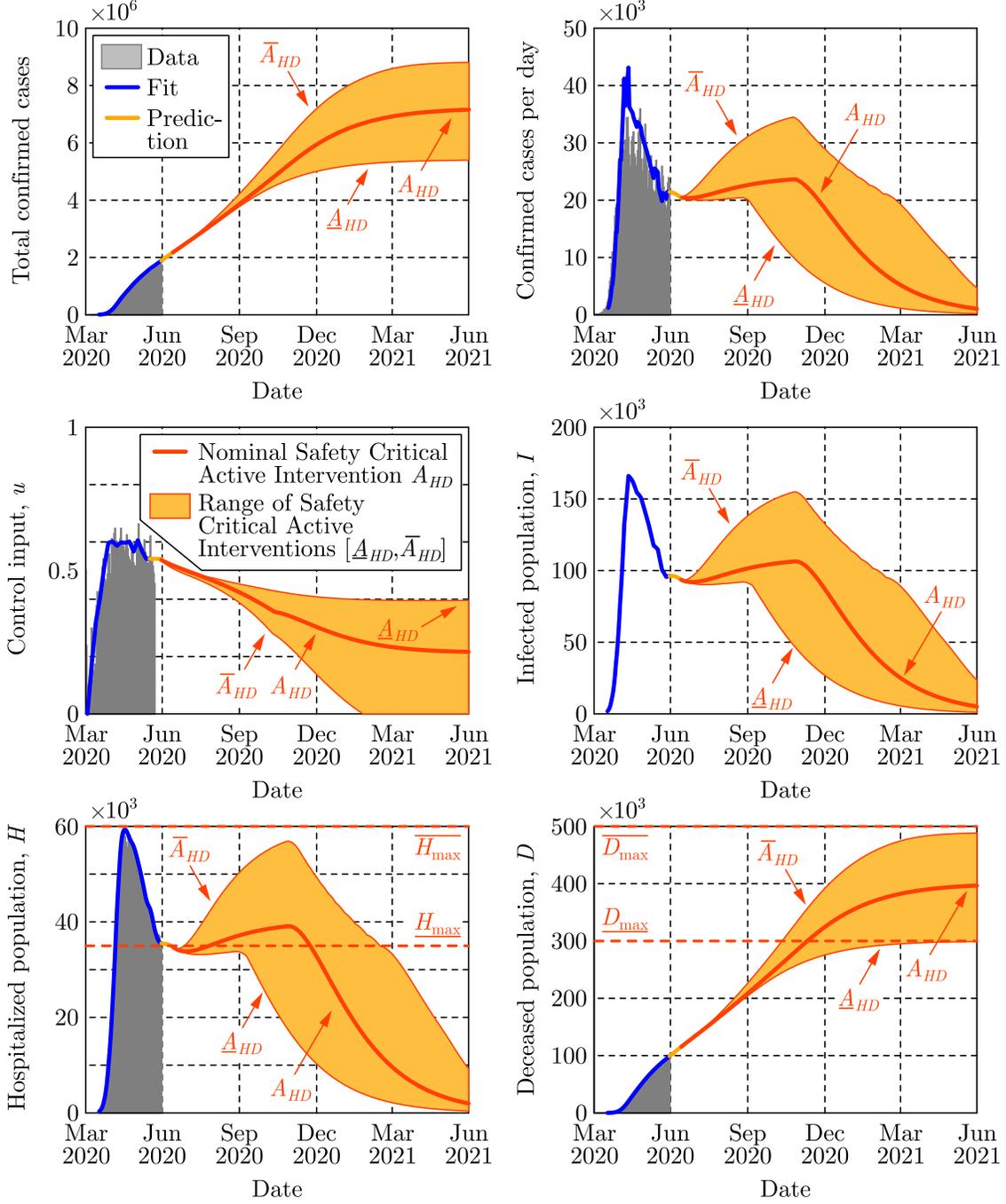
the conditions in Eq. (21) can be restated as a single inequality constraint:

$$\min \left\{ \frac{1}{\lambda} \Phi_H(\mathbf{x}(t)), \frac{1}{\mu} \Phi_D(\mathbf{x}(t)) \right\} + L_g h_I(\mathbf{x}(t)) u(t) \geq 0. \quad (24)$$

Thus, we can explicitly solve for the QP with this single constraint, yielding the same general form as Eq. (8) which leads to Eq. (9). In particular, when  $L_g h_I(\mathbf{x}(t)) > 0$ , the result is the explicit form for a controller that simultaneously enforces  $h_H(\mathbf{x}(t)) \geq 0$  and  $h_D(\mathbf{x}(t)) \geq 0$  for  $t > 0$ :

$$u(t) = A_{HD}(\mathbf{x}(t)) = \operatorname{ReLU} \left( -\frac{\min \left\{ \frac{1}{\lambda} \Phi_H(\mathbf{x}(t)), \frac{1}{\mu} \Phi_D(\mathbf{x}(t)) \right\}}{L_g h_I(\mathbf{x}(t))} \right) = \max \{A_H(\mathbf{x}(t)), A_D(\mathbf{x}(t))\}, \quad (25)$$

assuming  $h_i(\mathbf{x}(0)) \geq 0$  and  $h_i^e(\mathbf{x}(0)) \geq 0$  for  $i \in \{H, D\}$ . Therefore, the combined conditions on bounding hospitalizations and death can be achieved via taking the maximum of the controllers  $A_H$  and  $A_D$ , as given in (19) and (20), respectively. This controller is applied in Fig. 3 using US data and in Fig. 4 using state-level data (as described later in Methods). Figure 6 further demonstrates the range of safety-critical behaviors one can achieve with the active intervention policy in Eq. (25), wherein a range of maximum hospitalizations and deaths are considered leading to a range of active interventions.



**Figure 6.** Application via the active intervention policy in Eq. (25) utilizing multiple barrier functions, where a range of bounds are considered for hospitalization from  $\bar{H}_{max}$  to  $H_{max}$ , and death from  $\bar{D}_{max}$  to  $D_{max}$ . These bounds lead to safety-critical policies  $\bar{A}_{HD}$  and  $A_{HD}$ , respectively. These policies, and their corresponding range of interventions, are compared against the nominal safety-critical policy  $A_{HD}$  used in Fig. 3. Observe that the higher bounds on hospitalization and death,  $\bar{H}_{max}$  and  $\bar{D}_{max}$ , lead to the control input going to zero, while lower (and hence more stringent) bounds require sustained active intervention.

**Time Delays and Controller Synthesis with Predictors.** While delays have been considered in the past in epidemic models<sup>54–56</sup>, they were only used to modify the autonomous dynamics of the forecasting models and thus were not considered in the context of control systems and active interventions. Here we discuss how delays affect the active intervention policies designed for the delay-free system and how to utilize predictors to compensate for the delay.

If there exists a *measurement delay*  $\tau$  in a control system, cf. Eq. (4), then at time  $t$  only the *delayed state*  $\mathbf{x}(t - \tau)$  is available via measurements, while the *instantaneous state*  $\mathbf{x}(t)$  of the system is unknown. Therefore, the controller must rely on the delayed state, which modifies the control law from  $u(t) = K(\mathbf{x}(t))$  to  $u(t) = K(\mathbf{x}(t - \tau))$ , yielding the time-delayed closed loop system:

$$\dot{\mathbf{x}}(t) = f(\mathbf{x}(t)) + g(\mathbf{x}(t))K(\underbrace{\mathbf{x}(t - \tau)}_{\text{Delay}}). \quad (26)$$

Therefore, since measurement delay affects the dynamics of the closed loop system via time delays, they are typically undesirable—especially since they are often source of instability or reduced performance<sup>57</sup>. In the context of safety-critical control, as considered herein, time delays may lead to violation of the safety condition if one designs a controller by assuming no delays. If the delay is large, the measured delayed state may be significantly different from the instantaneous state due to the evolution of the system over the delay interval. This prevents the delay-free control design from guaranteeing safety.

A possible solution to overcome the poor performance caused by delays is the application of *predictor feedback control*<sup>48,58–61</sup>. Predictor feedback utilizes an *internal model* in order to predict the current instantaneous state of the system from delayed measurements. Prediction is made over the delay interval based on the delayed state  $\mathbf{x}(t - \tau)$  resulting in a *predicted state*  $\mathbf{x}_p(t)$  that is an estimation of the instantaneous state  $\mathbf{x}(t)$ , i.e.,  $\mathbf{x}_p(t) \approx \mathbf{x}(t)$ . Then, one can utilize the predicted state in the controller designed for the delay-free system. This leads to  $u(t) = K(\mathbf{x}_p(t))$  yielding:

$$\dot{\mathbf{x}}(t) = f(\mathbf{x}(t)) + g(\mathbf{x}(t))K(\mathbf{x}_p(t)). \quad (27)$$

If the prediction is accurate, i.e.,  $\mathbf{x}_p(t) = \mathbf{x}(t)$ , it eliminates the delays from the system.

To provide the predicted state  $\mathbf{x}_p(t)$  to the controller, one needs to anticipate how the closed loop system evolves under the predictor feedback controller over the delay interval  $[t - \tau, t]$ . This can be achieved by introducing  $\theta \in [t - \tau, t]$  and substituting  $t = \theta$  and  $\mathbf{x}(\theta) = \mathbf{x}_p(\theta)$  into Eq. (27):

$$\dot{\mathbf{x}}_p(\theta) = f(\mathbf{x}_p(\theta)) + g(\mathbf{x}_p(\theta))K(\mathbf{x}_p(\theta)). \quad (28)$$

This equation is the *internal model* used for prediction. To obtain the predicted state  $\mathbf{x}_p(t)$  required at time  $t$  by the controller  $u(t) = K(\mathbf{x}_p(t))$ , Eq. (28) can be numerically integrated over  $\theta \in [t - \tau, t]$  with initial condition  $\mathbf{x}_p(t - \tau) = \mathbf{x}(t - \tau)$  consisting of the most recent available measurement. Assuming a reasonable model,  $\mathbf{x}(\theta) \approx \mathbf{x}_p(\theta)$  for all  $\theta \in [t - \tau, t]$  wherein  $\mathbf{x}(t) \approx \mathbf{x}_p(t)$ , and so the predictor eliminates the time delay from the closed loop system in Eq. (27).

Figure 2 was generated by directly utilizing Eq. (28) as internal model. Prediction can be further improved by noticing that during the initial time  $\theta \in [t - \tau, t_0]$  (if  $t \in [t_0, t_0 + \tau]$ ), the system is not yet affected by an active intervention starting at  $t_0$ . Thus, one can apply the nominal (fitted) model with the corresponding fitted control input (see the yellow curve in Fig. 2) to calculate the predicted state during this initial interval, and Eq. (28) can be used afterwards. In Figs. 3 and 4, this more accurate prediction algorithm was utilized, although using Eq. (28) only could already compensate the effects of the time delay and managed to maintain safety in Fig. 2.

**Predictors for Compartmental Models.** In the control of COVID-19, the measurement delay originates from the incubation period and testing, and it can go up to about two weeks ( $\tau \approx 14$ ). Over these two weeks, the number of infected population may have increased significantly. As a result, the infected, hospitalized and deceased populations may be much closer to the safety limit than what the data shows. Thus, active intervention policies should be applied earlier than suggested by the delay-free controller, otherwise the populations of interest may overshoot and increase above the safe limit. The predictor feedback control technique accounts for the two-weeks measurement delay by predicting what could be the true number of infected population currently. Then, the delay-free control law can be applied utilizing the predictor, and it will therefore maintain safety if the prediction is accurate enough.

Accurate prediction requires knowing the delay. As shown in Fig. 1, the delay can be identified by model fitting to compartmental data and by utilizing mobility data. Furthermore, prediction requires an accurate internal model. For the SIR model, one can utilize the closed loop system:

$$\underbrace{\begin{bmatrix} \dot{S}_p(\theta) \\ \dot{I}_p(\theta) \\ \dot{R}_p(\theta) \end{bmatrix}}_{\dot{\mathbf{x}}_p(\theta)} = \underbrace{\begin{bmatrix} -\frac{\beta_0}{N} S_p(\theta) I_p(\theta) \\ \frac{\beta_0}{N} S_p(\theta) I_p(\theta) - \gamma I_p(\theta) \\ \gamma I_p(\theta) \end{bmatrix}}_{f(\mathbf{x}_p(\theta))} + \underbrace{\begin{bmatrix} \frac{\beta_0}{N} S_p(\theta) I_p(\theta) \\ -\frac{\beta_0}{N} S_p(\theta) I_p(\theta) \\ 0 \end{bmatrix}}_{g(\mathbf{x}_p(\theta))} A(S_p(\theta), I_p(\theta)), \quad (29)$$

which can be integrated over the interval  $\theta \in [t - \tau, t]$  with initial conditions  $S_p(t - \tau) = S(t - \tau)$ ,  $I_p(t - \tau) = I(t - \tau)$  and  $R_p(t - \tau) = R(t - \tau)$  to get the predicted states  $S_p(t)$  and  $I_p(t)$  required by the controller at time  $t$ . Similarly, for the SIHRD model prediction can be made using:

$$\underbrace{\begin{bmatrix} \dot{S}_p(\theta) \\ \dot{I}_p(\theta) \\ \dot{H}_p(\theta) \\ \dot{R}_p(\theta) \\ \dot{D}_p(\theta) \end{bmatrix}}_{\dot{\mathbf{x}}_p(\theta)} = \underbrace{\begin{bmatrix} -\frac{\beta_0}{N} S_p(\theta) I_p(\theta) \\ \frac{\beta_0}{N} S_p(\theta) I_p(\theta) - (\gamma + \lambda + \mu) I_p(\theta) \\ \lambda I_p(\theta) - \nu H_p(\theta) \\ \gamma I_p(\theta) + \nu H_p(\theta) \\ \mu I_p(\theta) \end{bmatrix}}_{f(\mathbf{x}_p(\theta))} + \underbrace{\begin{bmatrix} \frac{\beta_0}{N} S_p(\theta) I_p(\theta) \\ -\frac{\beta_0}{N} S_p(\theta) I_p(\theta) \\ 0 \\ 0 \\ 0 \end{bmatrix}}_{g(\mathbf{x}_p(\theta))} A_i(\mathbf{x}_p(\theta)), \quad (30)$$

for  $i \in \{I, H, D, HD\}$  depending on which active intervention policy is being utilized, i.e., whether we wish to bound infections, hospitalizations, deaths, or a combination thereof. According to Fig. 5, a short term (i.e., two-week) prediction can be made accurately with such model. Therefore, the result is safety of the system even in the presence of the time delay.

**Parameter Identification for Compartmental Models.** As mentioned previously, a wide variety of model parameters can be used to fit COVID-19 data, yielding drastically different predictions for the evolution of the different populations. In order to provide a model with high predictive ability, the parameters must be determined in a way that reflects their physical meaning while the model reproduces the available data. Since the model is being treated as a control system, the sequence of inputs that reflect the intervention applied thus far can be estimated. To ensure that the estimated inputs are realistic, they are initialized using mobility data from SafeGraph<sup>38</sup>, a data company that aggregates anonymized location data from numerous applications in order to provide insights about physical places. Such data allow us to quantify the increased time people stay at home to increase social distancing. To enhance privacy, SafeGraph excludes census block group information if fewer than five devices visited an establishment in a month from a given census block group.

The data fitting problem identifying the parameters of the SIR model in Eq. (1) can be formulated as an unconstrained, nonlinear optimization problem:

$$\min_{\beta, i_0, \alpha_0, \alpha_1, \tau} \frac{\|(I + R) - P\|}{\|P\|} + \frac{\|\beta - Q\|}{\|Q\|}. \quad (31)$$

The first term in the objective function seeks to minimize the difference between the total cases  $(I + R)$  given by the model and the number of positive test cases from the data  $P$ . The second term seeks to minimize the difference between the transmission rate  $\beta$  and some function  $Q$  that models the expected transmission rate based on mobility data. To obtain  $(I + R)$  at different time moments, we integrate Eq. (1) and this integration results in the nonlinearity in the objective function.

The optimization problem in Eq. (31) is used to find the parameters of the SIR model in Eq. (1) as follows. The first decision variable is the vector  $\beta \in \mathbb{R}^k$ ,  $k = \text{floor}(T/K)$ , representing the time-varying transmission rate  $\beta_j = \beta_0(1 - u(t_j))$ ,  $j \in \{1, 2, \dots, k\}$ . This corresponds to a control input  $u(t_j)$  that is updated every  $K$  days over the time span  $T$ . The value  $K = 5$  days was used in order to prevent overfitting and account for the inability to define new policies on a daily basis. The decision variable  $i_0 = \frac{I(0)}{I(0)+R(0)} \in [0, 1]$  signifies the initial values of different populations, while decision variables  $\alpha_0, \alpha_1, \tau$  are used to scale and delay the mobility data according to:

$$Q(t) = \alpha_0 + \alpha_1(1 - T_s(t - \tau)), \quad T_s(t) = \frac{T_h(t) - \min(T_h(t))}{1 - \min(T_h(t))}, \quad (32)$$

where  $T_h$  is the median percentage time spent at home. Recall that the delay  $\tau$  exists due to the testing delay and the incubation period of the virus<sup>42</sup>, and it is bounded to be between 5 to 16 days. Finally, the parameter  $\beta_0$  is extracted from  $\beta$  and  $Q$  as  $\beta_0 = \max_{j \in \{1, \dots, k\}, t \in [0, T]} \{\beta_j, Q(t)\}$ . When fitting the SIR model to the US national data we obtain  $\beta_0 = 0.51$  1/day and  $\tau = 10$  days, while we set the parameter  $\gamma = 0.2$  1/day to correspond to the average characteristic recovery time observed in the data. The fit for the SIR model is shown in Fig. 1 where the prediction along the 10 day period of the delay  $\tau$  is highlighted. These predictions are used in Fig. 2 when applying the safety-critical active intervention policy in Eq. (2).

It is important to note that for the SIR model, and the corresponding optimization problem in Eq. (31),  $N$  is fixed at a value below the total population. This is necessary due to the underreporting of infections<sup>43</sup>, along with the fact that many people will never have any contact with infected individuals. Additionally, setting  $N$  to be the entire population of the US in the SIR model would require many years for the virus to die out even with strict social distancing. Since our goal with the SIR model was to demonstrate the concepts of safety-critical active interventions, rather than to predict the size of the populations accurately, we decided to fix  $N$  to be 7.5 million. Alternatively, one can make  $N$  a decision variable—this did not improve results for the SIR model, but will prove useful for the higher-fidelity SIHRD model as explained below.

Analogous to the optimization problem for the SIR model, the optimization problem to estimate the parameters of the SIHRD model in Eq. (13) can be framed as:

$$\min_{N, \beta, i_0, \gamma, \lambda, \nu, \mu, \alpha_0, \alpha_1, \tau} \frac{\|(I + H + R + D) - P\|}{\|P\|} + \frac{\|\beta - Q\|}{\|Q\|} + h \frac{\|H - H_d\|}{\|H_d\|} + d \frac{\|D - D_d\|}{\|D_d\|}. \quad (33)$$

This formula accounts for the number of hospitalizations  $H_d$  and the number of deaths  $D_d$  in the data. The weights  $h > 1$  and  $d > 1$  are set to reflect that the data for hospitalizations and deaths are inherently less uncertain than the data on total positive test cases. In order to obtain the populations  $I + H + R + D$ ,  $H$ , and  $D$  as function of time we integrate the model in Eq. (13) which contains the new decision variables  $N, \gamma, \lambda, \nu, \mu$ . Due to the increased complexity of the model, the parameter  $\gamma$  is no longer fixed, as it was in the SIR model, instead, the value of  $1/(\gamma + \lambda + \mu)$  is implicitly constrained to be roughly within 2.5-7.5 days through individual bounds on the decision variables. Lastly, the characteristic time at which someone recovers from the hospital  $1/\nu$  is bounded between 2.5 and 4 days. Finally, differing from the SIR model optimization, the population  $N$  is now a decision variable with a lower bound of 4% and an upper bound of 10% of the total population of the nation or state of interest. We note that the value of  $N$  does not change the qualitative conclusions regarding safety-critical active interventions.

These optimization problems were solved using the pagmo<sup>62</sup> C++ library, and the solutions were verified using a variety of its solvers including CMA-ES, differential evolution, NSGA-II, and several solvers in the NLOPT suite. Table 2 provides the obtained parameters as computed by CMA-ES with a population size of 400 and a generation number of 400. The predictive power of the SIHRD model is illustrated by the fit in Fig. 5 over a 25 day horizon which provides accurate predictions for both the hospitalized  $H$  and the deceased  $D$  populations. This tight fit on  $H$  and  $D$  comes at the cost of a worse fit on the infected populace  $I$ , as reflected in the graph on total confirmed cases. However, as noted previously, the uncertainty in the total case number data is much higher than that of the  $H$  and  $D$  compartments, so this mismatch is not surprising. In fact, it could be argued that the actual number of infected persons is higher than reported and this model, as a result of its utilization of hospitalizations and deaths, actually captures this higher number. For example, in Fig. 5 the SIHRD model predicts a higher peak in the number of cases per day in the first half of April—this better describes the hospitalization and death data, and could provide a more accurate picture of cases per day due to the lack of testing at that time.

|        |     | Parameters |           |          |           |       |       |            |            |        |
|--------|-----|------------|-----------|----------|-----------|-------|-------|------------|------------|--------|
|        |     | $N(\%)$    | $\beta_0$ | $\gamma$ | $\lambda$ | $\mu$ | $\nu$ | $\alpha_0$ | $\alpha_1$ | $\tau$ |
| Region | USA | 4.01       | 0.57      | 0.07     | 0.14      | 0.01  | 0.39  | 0.16       | 0.39       | 14     |
|        | CA  | 4.00       | 0.57      | 0.07     | 0.25      | 0.01  | 0.26  | 0.31       | 0.26       | 15     |
|        | MI  | 5.71       | 0.44      | 0.07     | 0.25      | 0.04  | 0.25  | 0.34       | 0.10       | 16     |
|        | NY  | 4.28       | 0.60      | 0.07     | 0.08      | 0.01  | 0.27  | 0.12       | 0.46       | 14     |
|        | TX  | 4.00       | 0.37      | 0.20     | 0.04      | 0.01  | 0.09  | 0.23       | 0.15       | 15     |
|        | AR  | 9.18       | 0.44      | 0.20     | 0.15      | 0.01  | 0.40  | 0.34       | 0.10       | 16     |
|        | CO  | 4.00       | 0.61      | 0.07     | 0.21      | 0.02  | 0.37  | 0.23       | 0.38       | 13     |
|        | CT  | 4.00       | 0.63      | 0.07     | 0.25      | 0.03  | 0.40  | 0.38       | 0.25       | 15     |
|        | DE  | 4.00       | 0.50      | 0.12     | 0.23      | 0.01  | 0.40  | 0.40       | 0.10       | 14     |
|        | IA  | 4.00       | 0.54      | 0.20     | 0.17      | 0.01  | 0.40  | 0.39       | 0.14       | 16     |
|        | IL  | 4.68       | 0.45      | 0.07     | 0.25      | 0.02  | 0.31  | 0.35       | 0.10       | 16     |
|        | KY  | 10.00      | 0.50      | 0.14     | 0.25      | 0.02  | 0.26  | 0.40       | 0.10       | 8      |
|        | LA  | 4.00       | 0.48      | 0.07     | 0.25      | 0.02  | 0.33  | 0.34       | 0.14       | 15     |
|        | MA  | 8.32       | 0.60      | 0.07     | 0.25      | 0.02  | 0.38  | 0.33       | 0.27       | 13     |
|        | MN  | 4.00       | 0.56      | 0.15     | 0.22      | 0.03  | 0.40  | 0.40       | 0.16       | 16     |
|        | MO  | 4.00       | 0.42      | 0.07     | 0.25      | 0.02  | 0.20  | 0.32       | 0.10       | 12     |
|        | MT  | 4.03       | 0.33      | 0.07     | 0.23      | 0.02  | 0.34  | 0.23       | 0.10       | 8      |
|        | NC  | 4.00       | 0.63      | 0.17     | 0.18      | 0.01  | 0.40  | 0.32       | 0.31       | 16     |
|        | ND  | 4.00       | 0.38      | 0.20     | 0.07      | 0.01  | 0.40  | 0.28       | 0.10       | 5      |
|        | NJ  | 7.43       | 0.71      | 0.07     | 0.19      | 0.02  | 0.37  | 0.28       | 0.43       | 16     |
|        | NM  | 4.00       | 0.58      | 0.20     | 0.19      | 0.02  | 0.40  | 0.39       | 0.19       | 16     |
|        | PA  | 4.00       | 0.57      | 0.07     | 0.25      | 0.02  | 0.38  | 0.33       | 0.23       | 16     |
|        | RI  | 8.79       | 0.60      | 0.20     | 0.14      | 0.01  | 0.40  | 0.36       | 0.24       | 16     |
|        | VA  | 4.00       | 0.77      | 0.17     | 0.25      | 0.02  | 0.26  | 0.40       | 0.37       | 16     |
|        | VT  | 10.00      | 0.54      | 0.20     | 0.23      | 0.03  | 0.26  | 0.40       | 0.14       | 14     |
|        | WA  | 10.00      | 0.57      | 0.16     | 0.25      | 0.03  | 0.34  | 0.40       | 0.17       | 8      |
|        | WI  | 10.00      | 0.50      | 0.14     | 0.25      | 0.02  | 0.31  | 0.40       | 0.10       | 5      |
|        | WV  | 4.00       | 0.58      | 0.14     | 0.25      | 0.02  | 0.23  | 0.35       | 0.22       | 16     |

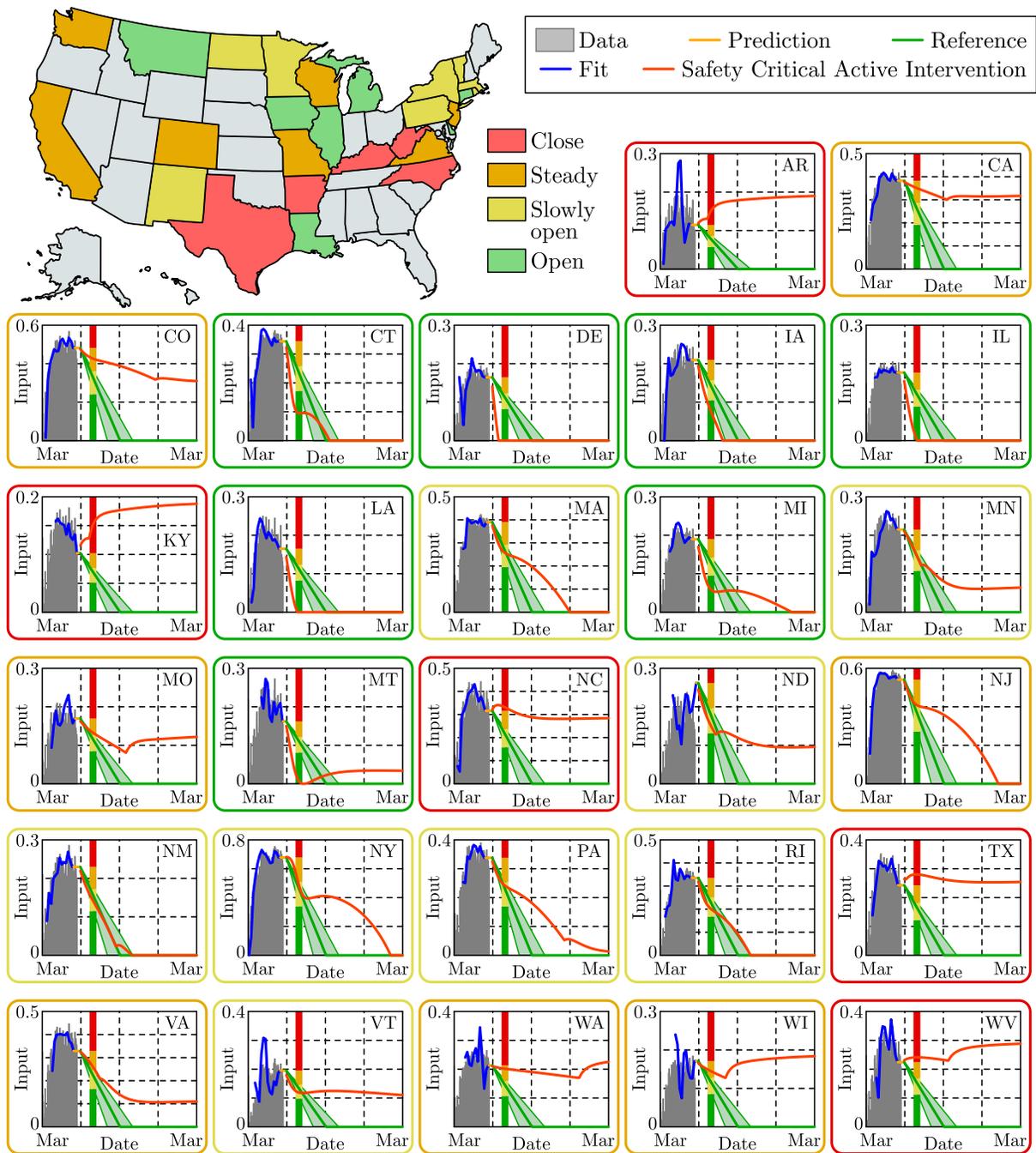
**Table 2.** Parameter estimates for the SIHRD model in different states, including the national level, obtained by solving the optimization problem in Eq. (33). Rough agreement between the states can be seen, but parameters vary due to differing exposure levels, social distancing protocols, population dynamics, and testing capacity. Note that the five states with a significantly smaller  $\tau$  have values shifted by roughly one week from the average, due to the cyclical nature of the data.

**State-Level Safety-Critical Active Interventions.** To provide a case study in using the safety-critical control framework to determine public policy, we apply the approach described in the main body of the paper, and detailed in Methods, to all states of the US that have sufficient data. A state is considered to have sufficient data if it has recorded values of positive cases, hospitalizations, and deaths for 50 consecutive days. While the model parameters can be estimated over shorter time periods, a sufficient window of data to verify the prediction accuracy was also required. To obtain the SIHRD model parameters at the state level, the optimization problem in Eq. (33) was solved using the CMA-ES algorithm with  $h = 2$  and  $d = 4$  to reflect the relative certainty of the data on hospitalizations and deaths, and thus, the desire to fit the former two populations more accurately. The parameters are identified per the methods described above, with the values shown in Table 2. Utilizing the fitted models, we develop a proof-of-concept reopening policy.

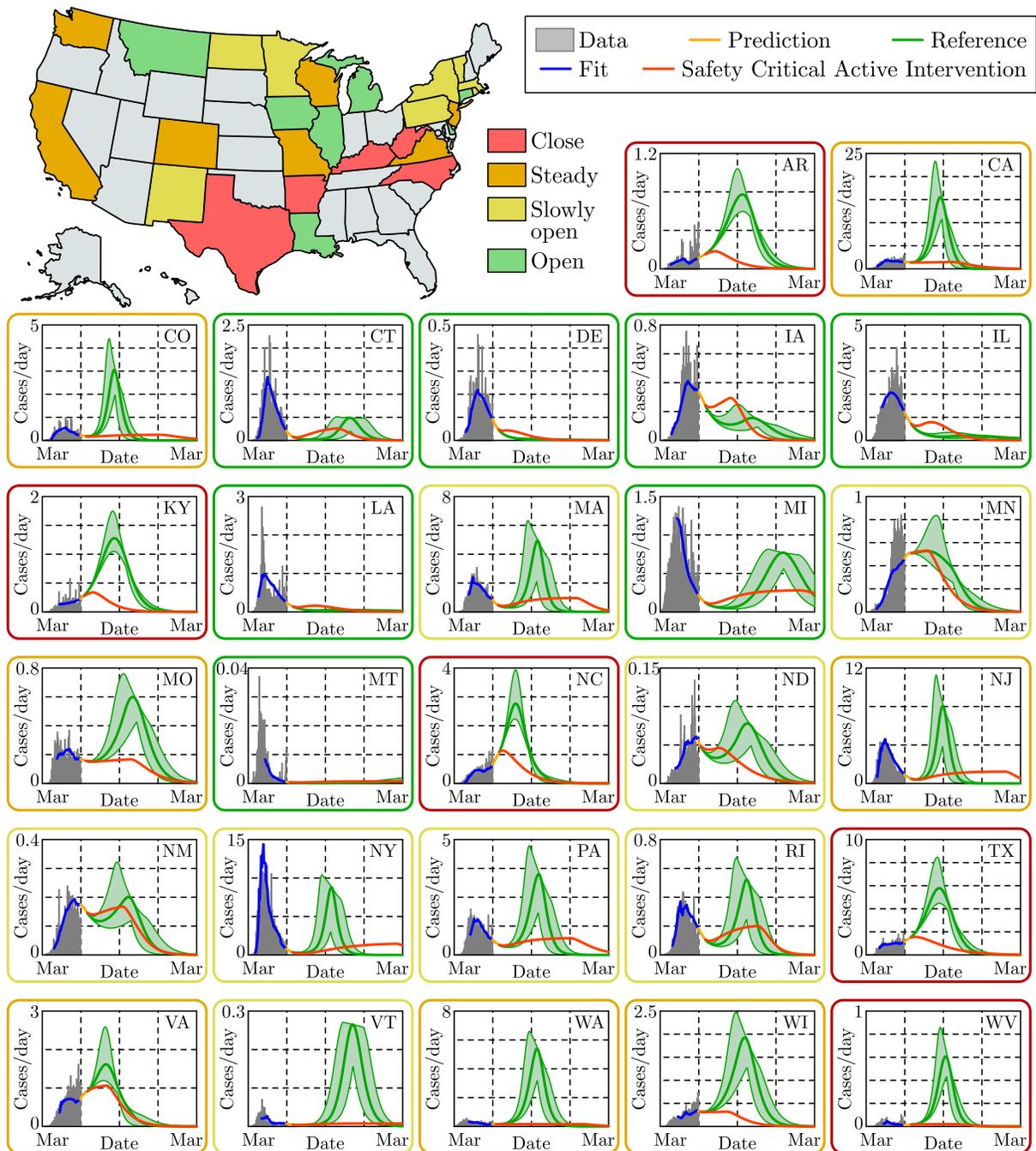
To develop a template policy on whether a state should reopen, and if so how aggressively, we apply the safety-critical framework to the SIHRD model for each state. Specifically, we utilize the safety-critical active intervention policy  $u(t) = A_{HD}(\mathbf{x}(t))$  given in Eq. (25). In this policy, there are two constants to be chosen: the maximum number of hospitalizations  $H_{\max}$ , and the maximum number of deceased  $D_{\max}$ . The value of  $H_{\max}$  for each state was chosen to be the last fitted hospitalization value (i.e., the last point of the blue curves in Figs. 4 and 9) to prevent hospitalizations from getting worse than the most recent recorded data. To guarantee safety, per Eq. (19), the initial condition must satisfy:  $h_H^e(\mathbf{x}(0)) = -\dot{H}(0) + \alpha_H(H_{\max} - H(0)) \geq 0$ . If  $H_{\max}$  did not satisfy this (which is the case for states where the most recent value of hospitalizations is the largest),  $H_{\max}$  is increased to the smallest feasible value:  $H_{\max} = \dot{H}(0)/\alpha_H + H(0)$  to ensure the feasibility of safety-critical active intervention. The value of  $D_{\max}$  was chosen to be three times the value at the start time of the active intervention (i.e., the start point of dark orange curves in Figs. 4 and 10). Although this choice is arbitrary, it satisfies the condition  $h_D^e(\mathbf{x}(0)) = -\dot{D}(0) + \alpha_D(D_{\max} - D(0)) \geq 0$  for all states and allows us to keep consistency throughout the states.

The proof-of-concept reopening framework is based upon the behavior of the control input  $u(t)$ . Since this safety-critical policy is (pointwise) optimal, i.e., solves the optimization problem in Eq. (21) at each  $t$ , there is no better instantaneous mitigation approach. Therefore, if this policy says to increase mitigations then a state should close down and if this policy rapidly reduces interventions the state can continue to open. To translate the control input into an applicable policy we utilize a “gating criterion” that is based on the value of the control input 30 days after the beginning of the active intervention period; see the colored vertical lines in Figs. 4 and 7 that determine the colors of states on the map. If the value of the control input after 30 days is greater than its initial value, the state should close down (states in red). If the input has reduced by less than 25%, mitigation efforts should be kept steady (states in orange). If the control input has decreased by more than 25% but less than 50%, the state can slowly reopen (states in yellow). Finally, a state where the input has decreased more than 50% in this 30 day period can reopen at a more rapid rate (states in green). While this is a very simplistic policy, created to provide uniformity across states, it demonstrates one way to utilize safety-critical methods to inform public policy. Indeed, using the entirety of the control input can result in more informed state level policy decisions.

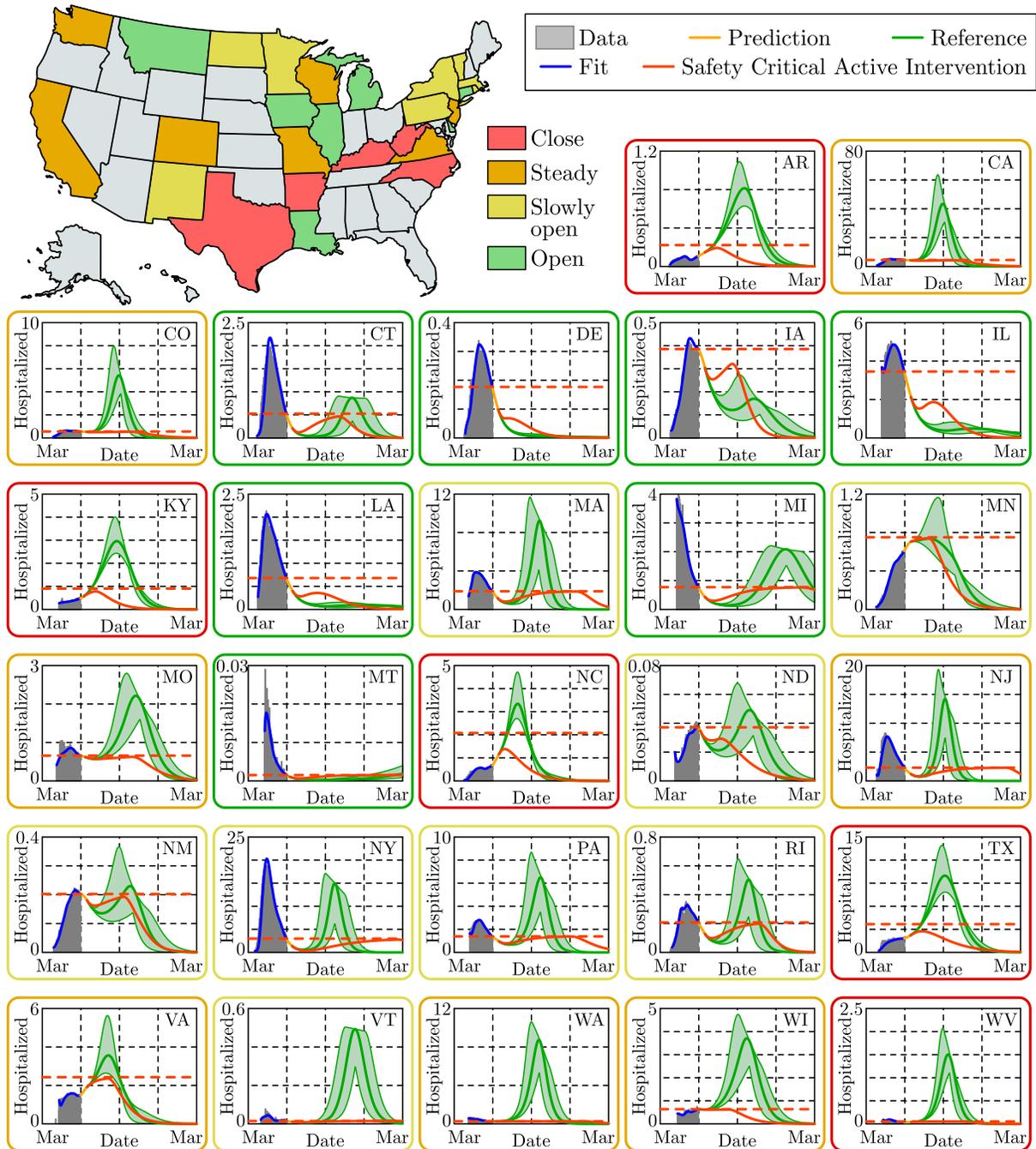
The criterion outlined above was applied to state level SIHRD models, with the results shown in Figs. 7-10. The safety-critical control inputs, that determine the state by state recommendations, are shown in Fig. 7. Also shown in that figure is the reference naive reopening policy that decays linearly in time, reaching zero on September 1 (the start of the school year). Fig. 8 shows the cases per day for the safety-critical and reference intervention policies; a second spike in cases can be seen for states that are yellow, orange and red for the reference policy. This spike is also seen in Fig. 9 in the context of hospitalizations, where  $H_{\max}$  is exceeded for the reference policy for all states while the safety-critical policy does not exceed  $H_{\max}$  as the theory implies. Finally, Fig. 10 shows the total deaths, wherein large death rates are seen—especially in states in red—for the reference policy while the safety-critical policy does not exceed  $D_{\max}$ . These figures, therefore, show that safety-critical control of active intervention can be used to synthesize state level reopening policies and, if followed, can limit hospitalizations and deaths.



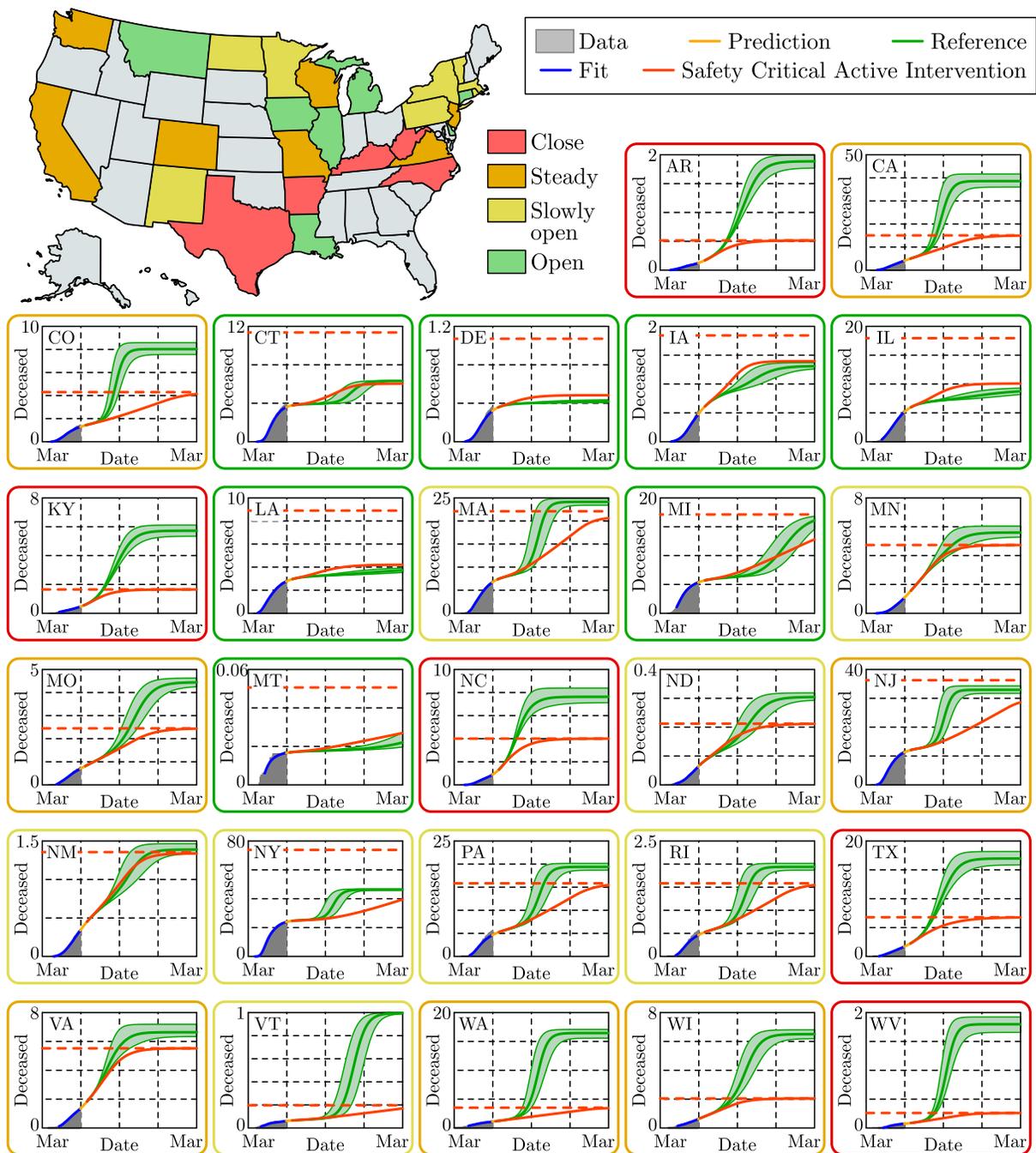
**Figure 7.** State level active intervention policies, along with the control input fit to US mobility data (grey) through May 30, 2020. Based upon the optimal safety-critical control policy (dark orange), state level mitigation efforts are recommended, i.e., a 30 day gating criterion is utilized where if at this time the safety-critical policy is increasing the state should close (red), if it has decreased by less than 25% it should hold mitigation efforts steady (orange), by less than 50% it should slowly reopen (yellow), and more than 50% it can open at a faster pace (green). This is compared against the naive linear reopening policy serving as a reference (green tube); this reference will be used in subsequent figures to demonstrate the consequences of a naive opening.



**Figure 8.** State level prediction for the cases per day over a year time span. The vertical axis measured in 1000's of people, and state level data (grey) is shown through May 30, 2020 along with the fit from the SIHRD model (blue). The effects of different active intervention policies are shown: the optimal safety-critical active intervention policy (dark orange) and the linear naive reopening policy (green). Both of these policies, at the state level, are shown in Fig. 7. In almost all states, the result is a second spike in cases—often larger than the original spike—if a naive reopening policy is used.



**Figure 9.** State level predictions for the number of persons hospitalized, with the vertical axis measured in 1000's of people. State level data (grey), along with the corresponding fit of the SIHRD model (blue) for hospitalizations; note the accurate fit of this data. The optimal safety-critical active intervention (dark orange) is compared against the naive reopening reference policy (green), with both policies shown in Fig. 7. Of particular note, for the safety-critical policy the number of hospitalizations is guaranteed to be bounded above by  $H_{\max}$  (horizontal dashed dark orange line), and it can be seen that this safety constraint is satisfied for all states. This can be compared against the reference policy where  $H_{\max}$  is exceeded for all states colored yellow, orange and red and even some states colored green. This indicates that without careful mitigation policies, hospital capacity constraints can be easily violated.



**Figure 10.** State level predictions for the total number of deaths, with the vertical axis measured in 1000's of persons. The data for the number of deceased persons as recorded through May 30, 2020 (grey) is shown along with the fit of the SIHRD model (blue); as with the hospitalization data in Fig. 9, the fit of the data is remarkable. The results of applying the safety-critical active intervention policy (dark orange) and the naive reference policy (green), where these policies are shown for each state in Fig. 7. The safety-critical policy guarantees that the total number of deaths stays under the upper bound  $D_{\max}$ , indicated by a horizontal dashed dark orange line. Note that for states that should close down as determined by the the algorithm illustrated in Fig. 7, i.e., the states indicated in red, there is a dramatic difference in the total number of deaths between the safety-critical and reference policies indicating the essential role of proper mitigation.

## References

1. Gates, B. Responding to Covid-19—a once-in-a-century pandemic? *New England Journal of Medicine* **382**, 1677–1679 (2020).
2. Flaxman, S. *et al.* Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature* (2020).
3. Ferguson, N. *et al.* Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand (2020).
4. IHME. COVID-19 Projections. <https://covid19.healthdata.org/united-states-of-america> (2020). [Online].
5. Giordano, G. *et al.* Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nature Medicine* (2020).
6. Weitz, J. S. *et al.* Modeling shield immunity to reduce COVID-19 epidemic spread. *Nature Medicine* (2020).
7. Ames, A. D., Xu, X., Grizzle, J. W. & Tabuada, P. Control barrier function based quadratic programs for safety critical systems. *IEEE Transactions on Automatic Control* **62**, 3861–3876 (2016).
8. Orosz, G. & Ames, A. D. Safety functionals for time delay systems. In *2019 American Control Conference (ACC)*, 4374–4379 (IEEE, 2019).
9. Holshue, M. L. *et al.* First case of 2019 novel coronavirus in the United States. *New England Journal of Medicine* **382**, 929–936 (2020).
10. Phelan, A. L., Katz, R. & Gostin, L. O. The novel coronavirus originating in Wuhan, China: challenges for global health governance. *Jama* **323**, 709–710 (2020).
11. Lescure, F.-X. *et al.* Clinical and virological data of the first cases of COVID-19 in Europe: a case series. *The Lancet Infectious Diseases* **20**, 697–706 (2020).
12. Le, T. T. *et al.* The COVID-19 vaccine development landscape. *Nature Reviews Drug Discovery* **19**, 305–306 (2020).
13. Kraemer, M. U. *et al.* The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science* **368**, 493–497 (2020).

14. Matrajt, L. & Leung, T. Evaluating the effectiveness of social distancing interventions to delay or flatten the epidemic curve of coronavirus disease. *Emerging Infectious Disease Journal* **26** (2020).
15. Pei, S., Kandula, S. & Shaman, J. Differential effects of intervention timing on COVID-19 spread in the United States. *medRxiv* 2020.05.15.20103655 (2020).
16. Hennekens, C. H., George, S., Adirim, T. A., Johnson, H. & Maki, D. G. The emerging pandemic of coronavirus and the urgent need for public health leadership. *The American Journal of Medicine* **133**, 648–650 (2020).
17. Arenas, A. *et al.* A mathematical model for the spatiotemporal epidemic spreading of COVID19. *medRxiv* 2020.03.21.20040022 (2020).
18. Bertozzi, A. L., Franco, E., Mohler, G., Short, M. B. & Sledge, D. The challenges of modeling and forecasting the spread of COVID-19. *arXiv preprint* 2004.04741 (2020).
19. Kucharski, A. J. *et al.* Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *The Lancet Infectious Diseases* **20**, 553–558 (2020).
20. Fernández-Villaverde, J. & Jones, C. I. Estimating and simulating a SIRD model of COVID-19 for many countries, states, and cities. Working Paper 27128, National Bureau of Economic Research (2020).
21. Kermack, W. O. & McKendrick, A. G. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society A* **115**, 700–721 (1927).
22. Hethcote, H. W. Three basic epidemiological models. In *Applied Mathematical Ecology*, 119–144 (Springer, 1989).
23. Hethcote, H. W. The mathematics of infectious diseases. *SIAM review* **42**, 599–653 (2000).
24. Batista, M. Estimation of the final size of the coronavirus epidemic by the SIR model (2020). [Online at ResearchGate].
25. Blyuss, K. B. & Kyrychko, Y. N. Effects of latency and age structure on the dynamics and containment of COVID-19. *medRxiv* 2020.04.25.20079848 (2020).
26. Lopez, L. R. & Rodo, X. A modified SEIR model to predict the COVID-19 outbreak in Spain and Italy: simulating control scenarios and multi-scale epidemics. *medRxiv* 2020.03.27.20045005 (2020).

27. Dandekar, R. & Barbastathis, G. Quantifying the effect of quarantine control in Covid-19 infectious spread using machine learning. *medRxiv* 2020.04.03.20052084 (2020).
28. Anderson, R. M. & May, R. *Infectious disease of humans: Dynamics and control* (Oxford University Press, 1992).
29. Stépán, G. Great delay in a predator-prey model. *Nonlinear Analysis* **10**, 913–929 (1986).
30. Zhao, S. & Chen, H. Modeling the epidemic dynamics and control of COVID-19 outbreak in China. *Quantitative Biology* **8**, 11–19 (2020).
31. Elhia, M., Rachik, M. & Benlahmar, E. Optimal control of an SIR model with delay in state and control variables. *ISRN Biomathematics* **2013**, 403549 (2013).
32. Bolzoni, L., Bonacini, E., Soresina, C. & Groppi, M. Time-optimal control strategies in SIR epidemic models. *Mathematical Biosciences* **292**, 86–96 (2017).
33. Casella, F. Can the COVID-19 epidemic be controlled on the basis of daily test reports? *arXiv preprint* 2003.06967 (2020).
34. Anderson, S. C. *et al.* Estimating the impact of COVID-19 control measures using a Bayesian model of physical distancing. *medRxiv* 2020.04.17.20070086 (2020).
35. Franco, E. A feedback SIR (fSIR) model highlights advantages and limitations of infection-based social distancing. *arXiv preprint* 2004.13216 (2020).
36. Dietz, K. The incidence of infectious diseases under the influence of seasonal fluctuations. In *Mathematical Models in Medicine*, 1–15 (Springer, 1976).
37. Buonomo, B., Chitnis, N. & d’Onofrio, A. Seasonality in epidemic models: a literature review. *Ricerche di Matematica* **67**, 7–25 (2018).
38. SafeGraph. <https://www.safegraph.com> (2020). [Online].
39. The COVID Tracking Project. <https://covidtracking.com/> (2020). [Online].
40. Global COVID-19 Tracker & Interactive Charts. <https://coronavirus.1point3acres.com/> (2020). [Online].
41. Chen, Y., Cheng, J., Jiang, Y. & Liu, K. A time delay dynamical model for outbreak of 2019-nCoV and the parameter identification. *arXiv preprint* 2002.00418 (2020).

42. Pei, S. & Shaman, J. Initial simulation of SARS-CoV2 spread and intervention effects in the continental US. *medRxiv* 2020.03.21.20040303 (2020).
43. Li, R. *et al.* Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). *Science* **368**, 489–493 (2020).
44. Kawohl, W. & Nordt, C. Covid-19, unemployment, and suicide. *The Lancet Psychiatry* **7**, 389–390 (2020).
45. Douglas, M., Katikireddi, S. V., Taulbut, M., McKee, M. & McCartney, G. Mitigating the wider health effects of covid-19 pandemic response. *BMJ* **369** (2020).
46. Xu, B., Wang, N., Chen, T. & Li, M. Empirical evaluation of rectified activations in convolutional network. *arXiv preprint* 1505.00853 (2015).
47. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
48. Krstic, M. *Delay Compensation for Nonlinear, Adaptive, and PDE Systems* (Birkhäuser, 2009).
49. Lai, S. *et al.* Effect of non-pharmaceutical interventions to contain COVID-19 in China. *Nature* (2020).
50. Ames, A. D. *et al.* Control barrier functions: Theory and applications. In *2019 18th European Control Conference (ECC)*, 3420–3431 (IEEE, 2019).
51. Aubin, J.-P. *Viability theory* (Springer Science & Business Media, 2009).
52. Blanchini, F. & Miani, S. *Set-theoretic methods in control* (Springer, 2008).
53. Nguyen, Q. & Sreenath, K. Exponential control barrier functions for enforcing high relative-degree safety-critical constraints. In *2016 American Control Conference (ACC)*, 322–328 (IEEE, 2016).
54. Kyrychko, Y. N. & Blyuss, K. B. Global properties of a delayed SIR model with temporary immunity and nonlinear incidence rate. *Nonlinear Analysis: Real World Applications* **6**, 495–507 (2005).
55. Rost, G. & Wu, J. SEIR epidemiological model with varying infectivity and infinite delay. *Nonlinear Analysis: Real World Applications* **5**, 389–402 (2008).

56. McCluskey, C. C. Complete global stability for an SIR epidemic model with delay – Distributed or discrete. *Nonlinear Analysis: Real World Applications* **11**, 55–59 (2010).
57. Insperger, T. & Stépán, G. *Semi-Discretization for Time-Delay Systems: Stability and Engineering Applications* (Springer, 2011).
58. Michiels, W. & Roose, D. Time delay compensation in unstable plants using delayed state feedback. In *Proceedings of the IEEE Conference on Decision and Control* (Orlando, FL, USA, 2001).
59. Krstic, M. & Bekiaris-Liberis, N. Compensation of infinite-dimensional input dynamics. *Annual Reviews in Control* **34**, 233–244 (2010).
60. Molnár, T. G. & Insperger, T. On the robust stabilizability of unstable systems with feedback delay by finite spectrum assignment. *Journal of Vibration and Control* **22**, 649–661 (2016).
61. Karafyllis, I. & Krstic, M. *Predictor feedback for delay systems: Implementations and approximations* (Birkhäuser, Basel, 2017).
62. Biscani, F. & Izzo, D. esa/pagmo2: pagmo 2.15.0 (2020). URL <https://doi.org/10.5281/zenodo.3738182>.

**Acknowledgements** The authors would like to thank Franca Hoffmann for her insights into compartmental epidemiological models and Gábor Stépán for discussions regarding non-pharmaceutical interventions in Europe. This research is supported in part by the National Science Foundation, CPS Award #1932091.

**Competing Interests** The authors declare that they have no competing financial interests.

**Correspondence** Correspondence and requests for materials should be addressed to Aaron Ames (email: [ames@caltech.edu](mailto:ames@caltech.edu)) and Gábor Orosz (email: [orosz@umich.edu](mailto:orosz@umich.edu)).