HAL
open science

Epidemiological Forecasting with Model Reduction of Compartmental Models. Application to the COVID-19 Pandemic

Athmane Bakhta, Thomas Boiveau, Yvon Maday, Olga Mula

To cite this version:

Athmane Bakhta, Thomas Boiveau, Yvon Maday, Olga Mula. Epidemiological Forecasting with Model Reduction of Compartmental Models. Application to the COVID-19 Pandemic. Biology, 2020, 10 (1), pp.22. 10.3390/biology10010022 . hal-03117258

HAL Id: hal-03117258

https://hal.sorbonne-universite.fr/hal-03117258v1

Submitted on 21 Jan 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

biology

MDPI

Article

Epidemiological Forecasting with Model Reduction of Compartmental Models. Application to the COVID-19 Pandemic

Athmane Bakhta [1], Thomas Boiveau [2], Yvon Maday [3,4] and Olga Mula [5,6,*]

1 Service de Thermo-Hydraulique et de Mécanique des Fluides, CEA, Université Paris-Saclay, 91191 Gif-sur-Yvette, France; athmane.bakhta@cea.fr
2 Institut Carnot Smiles, Sorbonne Université, 75005 Paris, France; thomas.boiveau@upmc.fr
3 Sorbonne Université and Université de Paris, CNRS, Laboratoire Jacques-Louis Lions (LJLL), F-75005 Paris, France; maday@ann.jussieu.fr
4 Institut Universitaire de France, 75005 Paris, France
5 CEREMADE, CNRS, UMR 7534, Université Paris-Dauphine, PSL University, 75016 Paris, France
6 Inria, Commedia Team, 75012 Paris, France
* Correspondence: mula@ceremade.dauphine.fr

Simple Summary: Using tools from the reduced order modeling of parametric ODEs and PDEs, including a new positivity-preserving greedy reduced basis method, we present a novel forecasting method for predicting the propagation of an epidemic. The method takes a collection of highly detailed compartmental models (with different initial conditions, initial times, epidemiological parameters and numerous compartments) and learns a model with few compartments which best fits the available health data and which is used to provide the forecasts. We illustrate the promising potential of the approach to the spread of the current COVID-19 pandemic in the case of the Paris region during the period from March to November 2020, in which two epidemic waves took place.

Abstract: We propose a forecasting method for predicting epidemiological health series on a two-week horizon at regional and interregional resolution. The approach is based on the model order reduction of parametric compartmental models and is designed to accommodate small amounts of sanitary data. The efficiency of the method is shown in the case of the prediction of the number of infected people and people removed from the collected data, either due to death or recovery, during the two pandemic waves of COVID-19 in France, which took place approximately between February and November 2020. Numerical results illustrate the promising potential of the approach.

Keywords: COVID-19; epidemiology; forecasting; model reduction; reduced basis

check for updates

Citation: Bakhta, A.; Boiveau, T.; Maday,Y.; Mula, O. Epidemiological Forecasting with Model Reduction of Compartmental Models. Application to the COVID-19 Pandemic. *Biology* **2021**, *10*, 22. https://dx.doi.org/10.3390/biology10010022

Received: 3 November 2020
Accepted: 23 December 2020
Published: 31 December 2020

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Providing reliable epidemiological forecasts during an ongoing pandemic is crucial to mitigate the potentially disastrous consequences for global public health and the economy. As the ongoing pandemic of COVID-19 sadly illustrates, this is a daunting task in the case of new diseases due to the incomplete knowledge of the behavior of the disease and the heterogeneities and uncertainties in the health data count. Despite these difficulties, many forecasting strategies exist, and we can cast them into two main categories: the first type is purely data-based and involves statistical and learning methods such as time series analysis, multivariate linear regression, grey forecasting or neural networks [1–5]; the second approach uses epidemiological models, which are appealing since they provide an interpretable insight of the mechanisms of the outbreak. They also provide high flexibility in the level of detail to describe the evolution of a pandemic, ranging from simple compartmental models that divide the population into a few exclusive categories to highly detailed descriptions involving numerous compartments or even agent-based models (see, e.g., [6–8] for general

references on mathematical epidemiological models and [9–11] for some models focused on COVID-19). One salient drawback of using epidemiological models for forecasting purposes lies in the very high uncertainty in the estimation of the relevant parameters. This is due to the fact that the parameters cannot often be inferred from real observations, and the available data are insufficient or too noisy to provide any reliable estimation. The situation is aggravated by the fact that the number of parameters can quickly become large even in moderately simple compartmental models [10]. As a result, forecasting with these models involves making numerous a priori hypotheses which can sometimes be difficult to justify by data observations.

In this paper, our goal is to forecast the time-series of infected, removed and dead patients with compartmental models that involve as few parameters as possible in order to infer these series solely from the data. The available data are only given for hospitalized people; one can nevertheless estimate the total number of infected people through an adjustment factor taken from the literature. Such a factor takes into account the proportion of asymptomatic people and infected people who do not go to hospital. The model that includes the least number of parameters is probably the susceptible–infected–removed (SIR) model [12], which is based on a partition of the population into the following groups:

- Uninfected people, called susceptible (S);
- Infected and contagious people (I), with more or less marked symptoms;
- People removed (R) from the infectious process, either because they were cured or unfortunately died after being infected.

If N denotes the total population size that we assume to be constant over a certain time interval $[0, T]$, we have

$$N = S(t) + I(t) + R(t), \forall t \in [0, T],$$

and the evolution from S to I and from I to R is given for all $t \in [0, T]$ by

$$\begin{aligned} \frac{dS}{dt}(t) &= -\frac{\beta I(t)S(t)}{N} \\ \frac{dI}{dt}(t) &= \frac{\beta I(t)S(t)}{N} - \gamma I(t) \\ \frac{dR}{dt}(t) &= \gamma I(t). \end{aligned}$$

The SIR model has only two parameters:

- $\gamma > 0$ represents the recovery rate. In other words, its inverse γ^{-1} can be interpreted as the length (in days) of the contagious period;
- $\beta > 0$ is the transmission rate of the disease. It essentially depends on two factors: the contagiousness of the disease and the contact rate within the population. The larger this second parameter is, the faster the transition from susceptible to infectious will be. As a consequence, the number of hospitalized patients may increase very quickly and may lead to a collapse of the health system [13]. Strong distancing measures such as confinement can effectively act on this parameter [14,15], helping to keep it low.

Our forecasting strategy is motivated by the following observation: by allowing the parameters β and γ to be time-dependent, we can find optimal coefficients $\beta^*(t)$ and $\gamma^*(t)$ that exactly fit any series of infected and removed patients. In other words, we can perfectly fit any observed health series with an SIR model with time-dependent coefficients.

As we explain below, the high fitting power stems from the fact that the parameters β and γ are searched in $L^\infty([0, T], \mathbb{R}_+)$—the space of essentially bounded measurable functions. For our forecasting purposes, however, this space is too large to give any predictive power, and we need to find a smaller manifold that simultaneously has good fitting and forecasting properties. To this end, we developed a method based on model order reduction. The idea of the method was to find a space with a reduced dimension

that can host the dynamics of the current epidemic. This reduced space is learnt from a series of detailed compartmental models based on precise underlying mechanisms of the disease. One major difficulty in these models is the fitting of the correct parameters. In our approach, we do not seek to estimate these parameters; instead, we consider a large range of possible parameter configurations with a uniform sampling that allows us to simulate virtual epidemic scenarios in a longer range than the fitting window $[0, T]$. We next cast each virtual epidemic from the family of detailed compartmental models into the family of SIR models with time-dependent coefficients, as explained below. This procedure yields time-dependent parameters β and γ for each detailed virtual epidemic. The set of all such β (or γ) is then condensed into a reduced basis with a small dimension. We finally use the available health data on the time window $[0, T]$ to find the functions β and γ from the reduced space that best fit the current epidemic over $[0, T]$. Since the reduced basis functions are defined over a longer time range $[0, T + \tau]$ with $\tau > 0$ (e.g., two weeks), the strategy automatically provides forecasts from T to $T + \tau$. Its accuracy will be related to the pertinence of the mechanistic mathematical models that have been used in the learning process.

We note that an important feature of our approach is that all virtual simulations are considered equally important in the first stage, and the procedure automatically learns what the best scenarios (or linear combinations of scenarios) to describe the available data are. Moreover, the approach even mixes different compartmental models to accommodate these available data. This is in contrast to other existing approaches which introduce a strong a priori belief regarding the quality of a certain particular model. Note also that we can add models related to other illness and use the large manifold to fit a possible new epidemic. It is also possible to mix the current approach with other purely statistical or learning strategies by means of expert aggregation. One salient difference with these approaches which is important to emphasize is that our method hinges on highly detailed compartmental models which have clear epidemiological interpretations. Our collapsing methodology into the time-dependent SIR is a way of "summarizing" the dynamics into a few terms. One may expect that reducing the number of parameters comes at the cost of losing the interpretability of parameters, and this is true in general. Nevertheless, the numerical results of the present work show that a reasonable tradeoff between the "reduction of the number of parameters" and "interpretability of these few parameters" can be achieved.

The paper is organized as follows. In Section 2, we present the forecasting method in the case of a single region with a constant population. For this, in Section 2.1, we briefly introduce the epidemiological models involved in the procedure, namely the SIR model with time-dependent coefficients and more detailed compartmental models used for the training step. In Section 2.2, after proving that the SIR model with time-dependent coefficients in $L^\infty([0, T])$ is able to fit any admissible epidemiological evolution (as explained below), we present the main steps of the forecasting method. The method involves a collapsing step from detailed models to SIR models with time-dependent coefficients and model reduction techniques. We detail these points in Sections 2.3 and 2.4. In Section 3, we explain how the method can easily be extended to a multi-regional context involving population mobility and regional health data observations (provided, of course, that mobility data are available). In Section 3.1, we begin by clarifying that the nature of the mobility data will dictate the kind of multi-regional SIR model to use in this context. In Section 3.2, we outline how to adapt the main steps of the method to the multi-regional case. Finally, in Section 4, we present numerical results for the the two pandemic waves of COVID-19 in France in 2020, which took place approximately between February and November 2020. Concluding comments are given in Section 5, followed by two Appendices A and B that present details about the processing of the data noise and the forecasting error.

2. Methodology for a Single Region

For the sake of clarity, we first consider the case of a single region with a constant population and no population fluxes with other regions. Here, the term region is generic and may be applied to very different geographical scales, ranging from a full country to a department within a country or even smaller partitions of a territory.

2.1. Compartmental Models

The final output of our method is a mono-regional SIR model with time-dependent coefficients as explained below. This SIR model with time-dependent coefficients is evaluated with reduced modeling techniques involving a large family of models with finer compartments proposed in the literature. Before presenting the method in the next section, we here introduce all the models that we use in this paper along with useful notations for the rest of the paper.

2.1.1. SIR Models with Time-Dependent Parameters

We fit and forecast the series of infected and removed patients (dead and recovered) with SIR models where the coefficients β and γ are time-dependent:

$$\begin{aligned}\frac{dS}{dt}(t) &= -\frac{\beta(t)I(t)S(t)}{N}\\ \frac{dI}{dt}(t) &= \frac{\beta(t)I(t)S(t)}{N} - \gamma(t)I(t)\\ \frac{dR}{dt}(t) &= \gamma(t)I(t).\end{aligned}$$

In the following, we use bold-faced letters for past-time quantities. For example, $\mathbf{f} := \{f(t) : 0 \leq t \leq T\}$ for any function $\mathbf{f} \in L^\infty([0,T])$. Using this notation, for any given $\boldsymbol{\beta}$ and $\boldsymbol{\gamma} \in L^\infty([0,T])$ we denote by

$$(\mathbf{S}, \mathbf{I}, \mathbf{R}) = \texttt{SIR}(\boldsymbol{\beta}, \boldsymbol{\gamma}, [0,T])$$

the solution of the associated SIR dynamics in $[0,T]$.

2.1.2. Detailed Compartmental Models

Models involving many compartments offer a detailed description of epidemiological mechanisms at the expense of involving many parameters. In our approach, we use them to generate virtual scenarios. One of the initial motivations behind the present work is to provide forecasts for the COVID-19 pandemic; thus, we have selected the two following models which are specific for this disease, but note that any other compartmental model [9,10,16] or agent-based simulation could also be used.

- First model, `SEI5CHRD`: This model is inspired by the one proposed in [10]. It involves 11 different compartments and a set of 19 parameters (see Table 1). The dynamics of the model are illustrated in Figure 1, and the system of equations reads as follows:

$$
\begin{aligned}
\frac{dS}{dt}(t) &= -\frac{1}{N}S(t)\big(\beta_p I_p(t) + \beta_a I_a(t) + \beta_{ps} I_{ps}(t) + \beta_{ms} I_{ms}(t) + \beta_{ss} I_{ss}(t) + \beta_H H(t) + \beta_C C(t)\big) \\
\frac{dE}{dt}(t) &= \frac{1}{N}S(t)\big(\beta_p I_p(t) + \beta_a I_a(t) + \beta_{ps} I_{ps}(t) + \beta_{ms} I_{ms}(t) + \beta_{ss} I_{ss}(t) + \beta_H H(t) + \beta_C C(t)\big) - \varepsilon E(t) \\
\frac{dI_p}{dt}(t) &= \varepsilon E(t) - \mu_p I_p(t) \\
\frac{dI_a}{dt}(t) &= p_a \mu_p I_p(t) - \mu I_a(t) \\
\frac{dI_{ps}}{dt}(t) &= p_{ps}(1-p_a)\mu_p I_p(t) - \mu I_{ps}(t) \\
\frac{dI_{ms}}{dt}(t) &= p_{ms}(1-p_a)\mu_p I_p(t) - \mu I_{ms}(t) \\
\frac{dI_{ss}}{dt}(t) &= p_{ss}(1-p_a)\mu_p I_p(t) - \mu I_{ss}(t) \\
\frac{dC}{dt}(t) &= p_c \mu I_{ss}(t) - (\lambda_{C,R} + \lambda_{C,D})C(t) \\
\frac{dH}{dt}(t) &= (1-p_c)\mu I_{ss}(t) - (\lambda_{H,R} + \lambda_{H,D})H(t) \\
\frac{dR}{dt}(t) &= \lambda_{C,R} C(t) + \lambda_{H,R} H(t) \\
\frac{dD}{dt}(t) &= \lambda_{C,D} C(t) + \lambda_{H,D} H(t)
\end{aligned}
$$

The different parameters involved in the model are described in Table 2 and detailed in the appendix of [10].

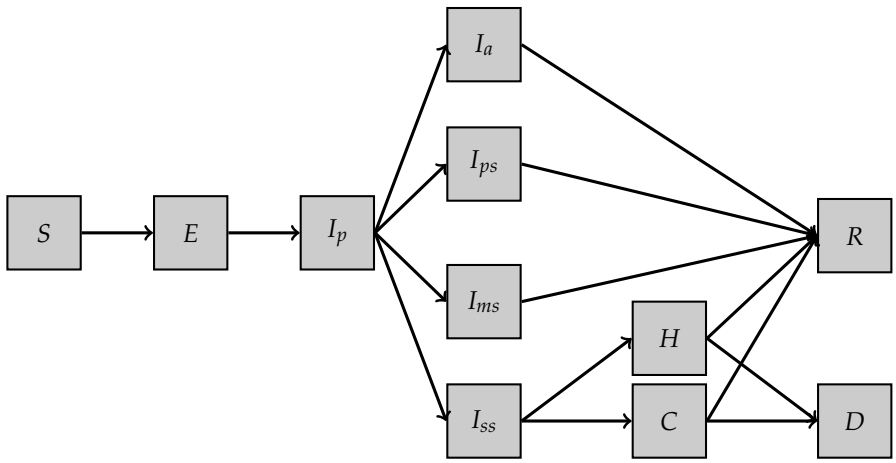


Figure 1. `SEI5CHRD` model.

Table 1. Description of the compartments in Model SEI5CHRD.

Compartment	Description
S	Susceptible
E	Exposed (non infectious)
I_p	Infected and pre-symptomatic (already infectious)
I_a	Infected and a-symptomatic (but infectious)
I_{ps}	Infected and paucisymptomatic
I_{ms}	Infected with mild symptoms
I_{ss}	Infected with severe symptoms
H	Hospitalized
C	Intensive Care Unit
R	Removed
D	Dead

Table 2. Description of the parameters involved in Model SEI5CHRD.

Parameter	Description
β_p	Relative infectiousness of I_p
β_a	Relative infectiousness of I_a
β_{ps}	Relative infectiousness of I_{ps}
β_{ms}	Relative infectiousness of I_{ms}
β_{ss}	Relative infectiousness of I_{ss}
β_H	Relative infectiousness of I_H
β_C	Relative infectiousness of I_C
ε^{-1}	Latency period
μ_p^{-1}	Duration of prodromal phase
p_a	Probability of being asymptomatic
μ^{-1}	Infectious period of I_a, I_{ps}, I_{ms}, I_{ss}
p_{ps}	If symptomatic, probability of being paucisymptomatic
p_{ms}	If symptomatic, probability of developing mild symptoms
p_{ss}	If symptomatic, probability of developing severe symptoms (note that $p_{ps} + p_{ms} + p_{ss} = 1$)
p_C	If severe symptoms, probability of going in C
λ_{CR}	If in C, daily rate entering in R
λ_{CD}	If in C, daily rate entering in D
λ_{HR}	If hospitalized, daily rate entering in R
λ_{HD}	If hospitalized, daily rate entering in D

We denote by

$$\begin{aligned}\mathbf{u} = \texttt{SEI5CHRD}(&\mathbf{u}_0, \beta_p, \beta_a, \beta_{ps}, \beta_{ms}, \beta_{ss}, \beta_H, \beta_C,\\ &\varepsilon, \mu_p, p_a, \mu, p_{ps}, p_{ms}, p_{ss}, p_C,\\ &\lambda_{CR}, \lambda_{CD}, \lambda_{HR}, \lambda_{HD}, [0, T])\end{aligned}$$

the parameter-to-solution map where $\mathbf{u} = (\mathbf{S}, \mathbf{E}, \mathbf{I}_p, \mathbf{I}_a, \mathbf{I}_{ps}, \mathbf{I}_{ms}, \mathbf{I}_{ss}, \mathbf{C}, \mathbf{H}, \mathbf{R}, \mathbf{D})$.

- Second model, SE2IUR: This model is a variant of the one proposed in [9]. It involves five different compartments (see Table 3) and a set of six parameters. The dynamics of the model are illustrated in Figure 2 and the set of equations is as follows:

$$
\begin{aligned}
\frac{dS}{dt}(t) &= -\frac{1}{N}\beta S(t)(E_2(t)+U(t)+I(t)) \\
\frac{dE_1}{dt}(t) &= \frac{1}{N}\beta S(t)(E_2(t)+U(t)+I(t)) - \delta E_1(t) \\
\frac{dE_2}{dt}(t) &= \delta E_1(t) - \sigma E_2(t) \\
\frac{dI}{dt}(t) &= \nu\sigma E_2(t) - \gamma_1 I(t) \\
\frac{dU}{dt}(t) &= (1-\nu)\sigma E_2(t) - \gamma_2 U(t) \\
\frac{dR}{dt}(t) &= \gamma_1 I(t) + \gamma_2 U(t)
\end{aligned}
$$

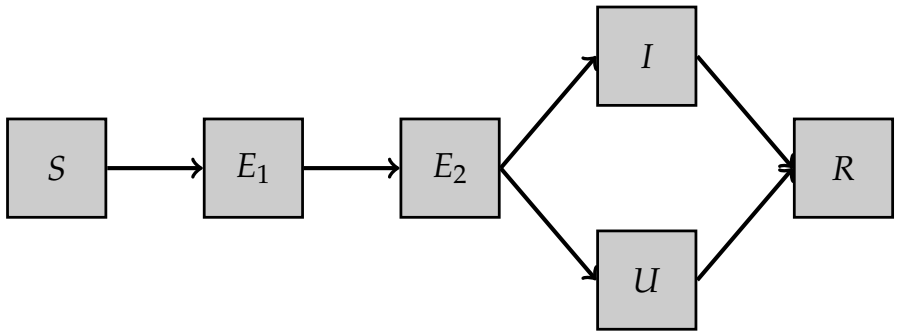


Figure 2. SE2IUR model.

We denote by

$$\mathbf{u} = \texttt{SE2IUR}(\mathbf{u}_0, \beta, \delta, \sigma, \nu, \gamma_1, \gamma_2, [0, T])$$

the parameter-to-solution map where $\mathbf{u} = (\mathbf{S}, \mathbf{E1}, \mathbf{E2}, \mathbf{I}, \mathbf{U}, \mathbf{R})$. The different parameters involved in the model are described in Table 4.

Table 3. Description of the compartments in model SE2IUR.

Compartment	Description
S	Susceptible
E_1	Exposed (non infectious)
E_2	Infected and pre-symptomatic (already infectious)
I	Infected and symptomatic
U	Un-noticed
R	dead and removed

Table 4. Description of the parameters involved in model SE2IUR.

Parameter	Description
β	Relative infectiousness of I, U, E_2
δ^{-1}	Latency period
σ^{-1}	Duration of prodromal phase
ν	Proportion of I among $I+U$
γ_1	If I, daily rate entering in R
γ_2	If U, daily rate entering in R

- Generalization: In the following, we abstract the procedure as follows. For any `Detailed_Model` with d compartments involving a vector $\mu \in \mathbb{R}^p$ of p parameters, we denote by

$$\mathbf{u}(\mu) = \texttt{Detailed_Model}(\mu, [0, \widetilde{T}]), \quad \mathbf{u}(\mu) \in L^\infty([0, \widetilde{T}], \mathbb{R}^d)$$

  the parameter-to-solution map, where $\widetilde{T}$ is some given time simulation that can be as large as desired because this is a virtual scenario. Note that, because the initial condition of the illness at time 0 is not known, we include the initial condition $\mathbf{u}_0$ in the parameter set.

2.2. Forecasting Based on Model Reduction of Detailed Models

We assume that we are given health data in a time window $[0, T]$, where $T > 0$ is assumed to be the present time. The observed data are the series of infected people, denoted I_{obs}, and removed people, denoted R_{obs}. They are usually given at a national or a regional scale and on a daily basis. For our discussion, it is useful to work with time-continuous functions, and $t \to I_{\text{obs}}(t)$ denotes the piecewise constant approximation in $[0, T]$ from the given data (and similarly for $R_{\text{obs}}(t)$). Our goal is to give short-term forecasts of the series in a time window $\tau > 0$ whose size is about two weeks. We denote by $I(t)$ and $R(t)$ the approximations to the series $I_{\text{obs}}(t)$ and $R_{\text{obs}}(t)$ at any time $t \in [0, T]$.

As mentioned above, we propose to fit the data and provide forecasts with SIR models with time-dependent parameters $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$. The main motivation for using such a simple family is that it possesses optimal fitting and forecasting properties for our purposes, as explained above. We define the cost function

$$\mathcal{J}(\boldsymbol{\beta}, \boldsymbol{\gamma} \mid I_{\text{obs}}(t), R_{\text{obs}}(t), [0, T]) := \int_0^T \Big(|I(t) - I_{\text{obs}}(t)|^2 + |R(t) - R_{\text{obs}}(t)|^2 \Big) \mathrm{d}t \tag{1}$$

such that

$$(\mathbf{S}, \mathbf{I}, \mathbf{R}) = \texttt{SIR}(\boldsymbol{\beta}, \boldsymbol{\gamma}, [0, T]),$$

and the fitting problem can be expressed at the continuous level as the optimal control problem of finding

$$J^* = \inf_{(\boldsymbol{\beta}, \boldsymbol{\gamma}) \in L^\infty([0,T]) \times L^\infty([0,T])} \mathcal{J}(\boldsymbol{\beta}, \boldsymbol{\gamma} \mid \mathbf{I}_{\text{obs}}, \mathbf{R}_{\text{obs}}, [0, T]). \tag{2}$$

The following result ensures the existence of a unique minimizer under very mild constraints.

Proposition 1. *Let $N \in \mathbb{N}^*$ and $T > 0$. For any real-valued functions $S_{obs}, I_{obs}, R_{obs}$ of class $\mathcal{C}^1$, defined on $[0, T]$ satisfying*

(i) $S_{obs}(t) + I_{obs}(t) + R_{obs}(t) = N$ *for every* $t \in [0, T]$,
(ii) S_{obs} *in nonincreasing on* $[0, T]$,
(iii) R_{obs} *is nondeacreasing on* $[0, T]$,

there exists a unique minimizer $(\boldsymbol{\beta}^*_{obs}, \boldsymbol{\gamma}^*_{obs})$ *to Equation (2).*

Proof. One can set

$$\begin{cases} \beta^*_{\text{obs}}(t) & := -\dfrac{N}{I_{\text{obs}}(t) S_{\text{obs}}(t)} \dfrac{dS_{\text{obs}}}{dt}(t) \\ \gamma^*_{\text{obs}}(t) & := \dfrac{1}{I_{\text{obs}}(t)} \dfrac{dR_{\text{obs}}}{dt}(t) \end{cases} \tag{3}$$

so that

$$(\mathbf{S}_{\text{obs}}, \mathbf{I}_{\text{obs}}, \mathbf{R}_{\text{obs}}) = \texttt{SIR}(\boldsymbol{\beta}^*, \boldsymbol{\gamma}^*, [0, T])$$

and

$$\mathcal{J}(\boldsymbol{\beta}^*_{\text{obs}}, \gamma^*_{\text{obs}}, [0,T]) = 0$$

which obviously implies that $J^* = 0$. $\square$

Note that one can equivalently set

$$\begin{cases} \beta^*_{\text{obs}}(t) & := -\dfrac{N}{I_{\text{obs}}(t)S_{\text{obs}}(t)} \dfrac{dS_{\text{obs}}}{dt}(t) \\ \gamma^*_{\text{obs}}(t) & := \dfrac{1}{I_{\text{obs}}(t)} \left[\dfrac{dI_{\text{obs}}}{dt}(t) - \dfrac{\beta^*_{\text{obs}}(t) I_{\text{obs}}(t) S_{\text{obs}}(t)}{N} \right] \end{cases}$$

or again

$$\begin{cases} \gamma^*_{\text{obs}}(t) & := \dfrac{1}{I_{\text{obs}}(t)} \dfrac{dR_{\text{obs}}}{dt}(t) \\ \beta^*_{\text{obs}}(t) & := \dfrac{N}{I_{\text{obs}}(t)S_{\text{obs}}(t)} \left[\dfrac{dI_{\text{obs}}}{dt}(t) - \gamma^*_{\text{obs}}(t) I_{\text{obs}}(t) \right] \end{cases}$$

This simple observation means that there exists a time-dependent SIR model which can perfectly fit the data of any (epidemiological) evolution that satisfies properties (i), (ii), and (iii). In particular, we can perfectly fit the series of sick people with a time-dependent SIR model (with a smoothing of the local oscillations due to noise). Since the health data are usually given on a daily basis, we can approximate $\boldsymbol{\beta}^*_{\text{obs}}, \gamma^*_{\text{obs}}$ by approximating the derivatives by classical finite differences in Equation (3).

The fact that we can build $\boldsymbol{\beta}^*_{\text{obs}}$ and γ^*_{obs} such that $\mathcal{J}(\boldsymbol{\beta}^*_{\text{obs}}, \gamma^*_{\text{obs}}) = J^* = 0$ implies that the family of time-dependent SIR models is rich enough not only to fit the evolution of any epidemiological series but also to deliver perfect predictions of the health data. It is however important to note that since the $\boldsymbol{\beta}^*_{\text{obs}}, \gamma^*_{\text{obs}}$ are derived exclusively from the data and depend on time, we lose the direct interpretations of these coefficients in terms of the length of the contagious period or the transmission rate that these coefficients present when they are considered constant. The great approximation power comes also at the cost of defining the parameters $\boldsymbol{\beta}$ and γ in $L^\infty([0,T])$ which is a space that is too large to be able to define any feasible prediction strategy.

In order to pin down a smaller manifold where these parameters may vary without sacrificing much in terms of the fitting and forecasting power, our strategy is the following:

1. Learning phase: The fundamental hypothesis of our approach is that the specialists of epidemiology have understood the mechanisms of infection transmission sufficiently well. The second hypothesis is that these accurate models suffer from the proper determination of the parameters they contain. We thus propose to generate a large number of virtual epidemics, following these mechanistic models, with parameters that can be chosen in the vicinity of the suggested parameter values, including the different initial conditions.
   (a) Generate virtual scenarios using detailed models with constant coefficients:
      - Define the notion of a `Detailed_Model` which is most appropriate for the epidemiological study. Several models could be considered. For our numerical application, the detailed models are defined in Section 2.1.
      - Define an interval range $\mathcal{P} \subset \mathbb{R}^p$ where the parameters μ of `Detailed_Model` vary. We call the solution manifold $\mathcal{U}$ the set of virtual dynamics over $[0, T+\tau]$, namely

$$\mathcal{U} := \{\mathbf{u}(\mu) = \texttt{Detailed_Model}(\mu, [0, T+\tau]) \ : \ \mu \in \mathcal{P}\}.$$

      - Draw a finite training set

$$\mathcal{P}_{\text{tr}} = \{\mu_1, \ldots, \mu_K\} \subseteq \mathcal{P}$$

of $K \gg 1$ parameter instances and compute $\mathbf{u}(\mu) = \texttt{Detailed_Model}(\mu, [0, T+\tau])$ for $\mu \in \mathcal{P}_{\text{tr}}$. Each $\mathbf{u}(\mu)$ is a virtual epidemiological scenario. An important detail for our prediction purposes is that the simulations are done in $[0, T+\tau]$; that is, we simulate not only in the fitting time interval but also in the prediction time interval. We call

$$\mathcal{U}_{\text{tr}} = \{\mathbf{u}(\mu) \ : \ \mu \in \mathcal{P}_{\text{tr}}\}$$

the training set of all virtual scenarios.

(b) Collapse every detailed model $\mathbf{u}(\mu) \in \mathcal{U}_{\text{tr}}$ into an SIR model following the ideas explained in Section 2.3. For every $\mathbf{u}(\mu)$, the procedure gives time-dependent parameters $\boldsymbol{\beta}(\mu)$ and $\boldsymbol{\gamma}(\mu)$ and associated SIR solutions $(\mathbf{S}, \mathbf{I}, \mathbf{R})(\mu)$ in $[0, T+\tau]$. This yields the sets

$$\mathcal{B}_{\text{tr}} := \{\boldsymbol{\beta}(\mu) \ : \ \mu \in \mathcal{P}_{\text{tr}}\} \quad \text{and} \quad \mathcal{G}_{\text{tr}} := \{\boldsymbol{\gamma}(\mu) \ : \ \mu \in \mathcal{P}_{\text{tr}}\}. \tag{4}$$

(c) **Compute reduced models:**
We apply model reduction techniques using $\mathcal{B}_{\text{tr}}$ and $\mathcal{G}_{\text{tr}}$ as training sets in order to build two bases

$$\mathrm{B}_n = \text{span}\{b_1, \ldots, b_n\}, \quad \mathrm{G}_n = \text{span}\{g_1, \ldots, g_n\} \subset L^\infty([0, T+\tau], \mathbb{R}),$$

which are defined over $[0, T+\tau]$. The space B_n is such that it approximates all functions $\boldsymbol{\beta}(\mu) \in \mathcal{B}_{\text{tr}}$ well (or all $\boldsymbol{\gamma}(\mu) \in \mathcal{G}_{\text{tr}}$ can be well approximated by elements of G_n). In Section 2.4, we outline the methods we have explored in our numerical tests.

2. Fitting on the reduced spaces: We next solve the fitting problem (2) in the interval $[0, T]$ by searching $\boldsymbol{\beta}$ (or $\boldsymbol{\gamma}$) in B_n (or in G_n) instead of in $L^\infty([0, T])$; that is,

$$J^*_{(\mathrm{B}_n, \mathrm{G}_n)} = \min_{(\boldsymbol{\beta}, \boldsymbol{\gamma}) \in \mathrm{B}_n \times \mathrm{G}_n} \mathcal{J}(\boldsymbol{\beta}, \boldsymbol{\gamma} \mid \mathbf{I}_{\text{obs}}, \mathbf{R}_{\text{obs}}, [0, T]). \tag{5}$$

Note that the respective dimensions of B_n and G_n can be different; for simplicity, we consider them to be equal in the following. Obviously, since B_n and $\mathrm{G}_n \subset L^\infty([0, T])$, we obtain

$$J^* \leq J^*_{(\mathrm{B}_n, \mathrm{G}_n)},$$

but we numerically observe that the function $n \mapsto J^*_{(\mathrm{B}_n, \mathrm{G}_n)}$ decreases very rapidly as n increases, which indirectly confirms the fact that the manifold generated by the two above models accommodates the COVID-19 epidemic well.
The solution of problem (5) gives us the coefficients $(c_i^*)_{i=1}^n$ and $(\tilde{c}_i^*)_{i=1}^n \in \mathbb{R}^n$ such that the time-dependent parameters

$$\beta_n^*(t) = \sum_{i=1}^n c_i^* b_i(t), \quad \forall t \in [0, T+\tau],$$
$$\gamma_n^*(t) = \sum_{i=1}^n \tilde{c}_i^* g_i(t).$$

achieve the minimum (5).

3. Forecast: For a given dimension n of the reduced spaces, we propagate in $[0, T+\tau]$ the associated SIR model, as follows:

$$(\mathbf{S}_n^*, \mathbf{I}_n^*, \mathbf{R}_n^*) = \texttt{SIR}(\boldsymbol{\beta}_n^*, \boldsymbol{\gamma}_n^*, [0, T+\tau])$$

The values $I_n^*(t)$ and $R_n^*(t)$ for $t \in [0, T[$ are by construction close to the observed data $\mathbf{I}_{\text{obs}}, \mathbf{R}_{\text{obs}}$ (up to some numerical optimization error). The values $I_n^*(t)$ and $R_n^*(t)$ for $t \in [T, T+\tau]$ are then used for prediction.

4. Forecast combination/aggregation of experts (optional step): By varying the dimension n and using different model reduction approaches, we can easily produce a collection of different forecasts, and thus the question of how to select the best predictive model arises. Alternatively, we can also resort to forecast combination techniques [17]: denoting $(I_1, R_1), \ldots, (I_P, R_P)$ as the different forecasts, the idea is to search for an appropriate linear combination

$$I^{\text{FC}}(t) = \sum_{p=1}^{P} w_p I_p(t)$$

and perform a similar operation for R. Note that these combinations do not need to involve forecasts from our methodology only; other approaches such as time series forecasts could also be included. One simple forecast combination is the average in which all alternative forecasts are given the same weight $w_p = 1/P$, $p = 1, \ldots P$. More elaborate approaches consist in estimating the weights that minimize a loss function involving the forecast error.

Before going into detail regarding some of the steps, three points should be highlighted:

1. To bring out the essential mechanisms, we have idealized some elements in the above discussion by omitting certain unavoidable discretization aspects. To start with, the ODE solutions cannot be computed exactly but only up to a certain level of accuracy given by a numerical integration scheme. In addition, the optimal control problems (2) and (5) are non-convex. As a result, in practice, we can only find a local minimum. Note, however, that modern solvers find solutions which are very satisfactory for all practical purposes. In addition, note that solving the control problem in a reduced space as in (5) could be interpreted as introducing a regularizing effect with respect to the control problem (2) in the full $L^\infty([0, T])$ space. It is to be expected that the search of global minimizers is facilitated in the reduced landscape.
2. `routine-IR` and `routine-`$\boldsymbol{\beta\gamma}$: A variant for the fitting problem (5) as studied in our numerical experiments is to replace the cost function $\mathcal{J}(\boldsymbol{\beta}, \boldsymbol{\gamma} \mid \mathbf{I}_{\text{obs}}, \mathbf{R}_{\text{obs}}, [0, T])$ by the cost function

$$\widetilde{\mathcal{J}}(\boldsymbol{\beta}, \boldsymbol{\gamma} \mid \boldsymbol{\beta}^*_{\text{obs}}, \boldsymbol{\gamma}^*_{\text{obs}}, [0, T]) := \int_0^T \left(|\boldsymbol{\beta} - \boldsymbol{\beta}^*_{\text{obs}}|^2 + |\boldsymbol{\gamma} - \boldsymbol{\gamma}^*_{\text{obs}})|^2 \right) \mathrm{d}t. \tag{6}$$

In other words, we use the variables $\boldsymbol{\beta}^*_{\text{obs}}$ and $\boldsymbol{\gamma}^*_{\text{obs}}$ from (3) as observed data instead of working with the observed health series $\mathbf{I}_{\text{obs}}$, $\mathbf{R}_{\text{obs}}$. In Section 4, we refer to the standard fitting method as `routine-IR` and to this variant as `routine-`$\boldsymbol{\beta\gamma}$.
3. The fitting procedure works both on the components of the reduced basis and the initial time of the epidemics to minimize the loss function; however, for simplicity, this last optimization is not reported here.

2.3. Details on Step 1-(b): Collapsing the Detailed Models into SIR Dynamics

Let

$$\mathbf{u}(\mu) = \texttt{Detailed_Model}(\mu, [0, T+\tau]) \ \in L^\infty([0, T+\tau], \mathbb{R}^d)$$

be the solution in $[0, T+\tau]$ to a detailed model for a given vector of parameters $\mu \in \mathbb{R}^d$. Here, d is possibly large ($d = 11$ in the case of the `SEI5CHRD` model and $d = 5$ in the case of `SE2IUR`'s model). The goal of this section is to explain how to collapse the detailed dynamics $\mathbf{u}(\mu)$ into `SIR`dynamics with time-dependent parameters. The procedure can be understood as a projection of a detailed dynamics into the family of dynamics given by `SIR` models with time-dependent parameters.

For the SEI5CHRD model, we collapse the variables by setting

$$\begin{aligned}\mathbf{S}^{\text{col}} &= \mathbf{S} + \mathbf{E}\\ \mathbf{I}^{\text{col}} &= \mathbf{I}_p + \mathbf{I}_a + \mathbf{I}_{ps} + \mathbf{I}_{ms} + \mathbf{I}_{ss} + \mathbf{C} + \mathbf{H}\\ \mathbf{R}^{\text{col}} &= \mathbf{R} + \mathbf{D}\end{aligned}$$

Similarly, for the SE2IUR model, we set

$$\begin{aligned}\mathbf{S}^{\text{col}} &= \mathbf{S} + \mathbf{E}_{1i}\\ \mathbf{I}^{\text{col}} &= \mathbf{E}_{2i} + \mathbf{I}_i + \mathbf{U}_i\\ \mathbf{R}^{\text{col}} &= \mathbf{R}\end{aligned}$$

Note that $\mathbf{S}^{\text{col}}$, $\mathbf{I}^{\text{col}}$ and $\mathbf{R}^{\text{col}}$ depend on μ, but we have omitted this dependence to simplify the notation.

Once the collapsed variables are obtained, we identify the time-dependent parameters $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ of the SIR model by solving the fitting problem

$$(\boldsymbol{\beta}, \boldsymbol{\gamma}) \in \underset{(\boldsymbol{\beta},\boldsymbol{\gamma}) \in L^\infty([0,T+\tau],\mathbb{R}) \times L^\infty([0,T+\tau],\mathbb{R})}{\arg\inf} \mathcal{J}(\boldsymbol{\beta}, \boldsymbol{\gamma} \,|\, \mathbf{I}^{\text{col}}, \mathbf{R}^{\text{col}}, [0, T+\tau]) \tag{7}$$

where

$$(\mathbf{S}, \mathbf{I}, \mathbf{R}) = \texttt{SIR}(\boldsymbol{\beta}, \boldsymbol{\gamma}, [0, T+\tau]).$$

Note that problem (7) has the same structure as problem (2), with the difference arising from the fact that the collapsed variables $\mathbf{I}^{\text{col}}$, $\mathbf{R}^{\text{col}}$ in (7) play the role of the health data $\mathbf{I}_{\text{obs}}$, $\mathbf{R}_{\text{obs}}$ in (2). Therefore, it follows from Proposition 1 that problem (7) has a very simple solution as it suffices to apply formula (3) to solve it. Note here that the exact derivatives of $\mathbf{S}^{\text{col}}$, $\mathbf{I}^{\text{col}}$, and $\mathbf{R}^{\text{col}}$ can be obtained from the `Detailed_Model`.

Since the solution $(\boldsymbol{\beta}, \boldsymbol{\gamma})$ to (7) depends on the parameter μ of the detailed model, repeating the above procedure for every detailed scenario $\mathbf{u}(\mu)$ for any $\mu \in \mathcal{P}_{\text{tr}}$ yields the two families of time-dependent functions $\mathcal{B}_{\text{tr}} = \{\boldsymbol{\beta}(\mu) \,:\, \mu \in \mathcal{P}_{\text{tr}}\}$ and $\mathcal{G}_{\text{tr}} = \{\boldsymbol{\gamma}(\mu) \,:\, \mu \in \mathcal{P}_{\text{tr}}\}$ defined in the interval $[0, T+\tau]$ as introduced in Section (4).

2.4. Details of Model Order Reduction

Model order reduction is a family of methods aiming at the approximation of a set of solutions of parametrized PDEs or ODEs (or related quantities) with linear spaces, which are called reduced models or reduced spaces. In our case, the sets to approximate are

$$\mathcal{B} = \{\boldsymbol{\beta}(\mu) \,:\, \mu \in \mathcal{P}\} \quad \text{and} \quad \mathcal{G} = \{\boldsymbol{\gamma}(\mu) \,:\, \mu \in \mathcal{P}\},$$

where each μ is the vector of parameters of the detailed model which take values over $\mathcal{P}$, and $\boldsymbol{\beta}(\mu)$ and $\boldsymbol{\gamma}(\mu)$ are the associated time-dependent coefficients of the collapsed SIR evolution. In the following, we view $\mathcal{B}$ and $\mathcal{G}$ as subsets of $L^2([0,T])$, and we denote by $\|\cdot\|$ and $\langle\cdot,\cdot\rangle$ its norm and inner product. Indeed, in view of Proposition 1, $\mathcal{B}$ and $\mathcal{G} \subset L^\infty([0,T]) \subset L^2([0,T])$.

Continuing the discussion if $\mathcal{B}$ (the same will hold for $\mathcal{G}$), of we measure performance in terms of the worst error in the set $\mathcal{B}$, the best possible performance that reduced models of dimension n can achieve is given by the Kolmogorov n-width:

$$d_n(\mathcal{B})_{L^2([0,T])} := \inf_{\substack{Y \in L^2([0,T]) \\ \dim(Y) \le n}} \max_{u \in \mathcal{B}} ||u - P_Y u||$$

where P_Y is the orthogonal projection onto the subspace Y. In the case of measuring errors in an average sense, the benchmark is given by

$$\delta_n^2(\mathcal{B},\nu)_{L^2([0,T])} := \inf_{\substack{Y\in L^2([0,T])\\ \dim(Y)=n}} \int_{\mathcal{P}} ||u(y) - P_Y u(y)||^2 \mathrm{d}\nu(y)$$

where ν is some given measure on $\mathcal{P}$.

In practice, building spaces that meet these benchmarks is generally not possible. However, it is possible to build sequences of spaces for which the error decay is comparable to that given by $\left(d_n(\mathcal{B})_{L^2([0,T])}\right)_n$ or $\left(\delta_n(\mathcal{B})_{L^2([0,T])}\right)_n$. As a result, when the Kolmogorov width decays quickly, the constructed reduced spaces will deliver a very good approximation of the set $\mathcal{B}$ with few modes (see [18–21]).

We next present the reduced models used in our numerical experiments. Other methods could, of course, be considered, and we refer readers to [22–25] for general references on model reduction. We continue the discussion in a fully discrete setting in order to simplify the presentation and keep it as close to the practical implementation as possible. All the claims below could be written in a fully continuous sense at the expense of introducing additional mathematical objects such as certain Hilbert–Schmidt operators to define the continuous version of Singular Value Decomposition (SVD).

We build the reduced models using the two discrete training sets of functions $\mathcal{B}_{\text{tr}} = \{\boldsymbol{\beta}_i\}_{i=1}^K$ and $\mathcal{G}_{\text{tr}} = \{\gamma_i\}_{i=1}^K$ from $\mathcal{B}$ and $\mathcal{G}$, where K denotes the number of virtual scenarios considered. The sets have been generated in step 1-(b) of our general pipeline (see Section 2.2).

We consider a discretization of the time interval $[0, T+\tau]$ into a set of $Q \in \mathbb{N}^*$ points as follows: $\{t_1 = 0, \cdots, t_P = T, \cdots, t_Q = T+\tau\}$ where $P < Q$. Thus, we can represent each function $\boldsymbol{\beta}_i$ as a vector of Q values

$$\boldsymbol{\beta}_i = (\beta_i(t_1), \cdots, \beta_i(t_Q))^T \in \mathbb{R}_+^Q.$$

and thus assemble all the functions of the family $\{\boldsymbol{\beta}_i\}_{i=1}^K$ into a matrix $M_{\mathcal{B}} \in \mathbb{R}_+^{Q\times K}$. The same remark applies for the family $\{\gamma_i\}_{i=1}^K$ which gives a matrix $M_{\mathcal{G}} \in \mathbb{R}_+^{Q\times K}$.

1. SVD: The eigenvalue decomposition of the correlation matrix $M_{\mathcal{B}}^T M_{\mathcal{B}} \in \mathbb{R}^{K\times K}$ gives

$$M_{\mathcal{B}}^T M_{\mathcal{B}} = V\Lambda V^T,$$

where $V = (v_{i,j}) \in \mathbb{R}^{K\times K}$ is an orthogonal matrix and $\Lambda \in \mathbb{R}^{K\times K}$ is a diagonal matrix with non-negative entries, which we denote as λ_i and present in decreasing order. The $\ell^2(\mathbb{R}^Q)$-orthogonal basis functions $\{\mathbf{b}_1, \ldots, \mathbf{b}_K\}$ are then given by the linear combinations

$$\mathbf{b}_i = \sum_{j=1}^K v_{j,i}\boldsymbol{\beta}_j, \quad 1 \le i \le K.$$

For $n \le K$, the space

$$\mathrm{B}_n = \mathrm{span}\{\mathbf{b}_1, \ldots \mathbf{b}_n\}$$

is the best n-dimensional space to approximate the set $\{\boldsymbol{\beta}_i\}_{i=1}^K$ in the average sense. We have

$$\delta_n(\{\boldsymbol{\beta}_i\}_{i=1}^K)_{\ell^2(\mathbb{R}^{Q+1})} = \left(\frac{1}{K}\sum_{i=1}^K ||\boldsymbol{\beta}_i - P_{\mathrm{B}_n}\boldsymbol{\beta}_i||^2_{\ell^2(\mathbb{R}^{Q+1})}\right)^{1/2} = \left(\sum_{i>n}^K \lambda_i\right)^{1/2}$$

and the average approximation error is given by the sum of the tail of the eigenvalues. Therefore, the SVD method is particularly efficient if there is a fast decay of the eigenvalues, meaning that the set $\mathcal{B}_{\text{tr}} = \{\boldsymbol{\beta}_i\}_{i=1}^K$ can be approximated by only few

modes. However, note that, by construction, this method does not ensure positivity in the sense that $P_{\mathrm{B}_n}\beta_i(t)$ may become negative for some $t \in [0, T]$ although the original function $\boldsymbol{\beta}_i(t) \geq 0$ for all $t \in [0, T]$. This is due to the fact that the vectors $\mathbf{b}_i$ are not necessarily nonnegative. As we will see later, in our study, ensuring positivity especially for extrapolation (i.e., forecasting) is particularly important and motivates the next methods.

2. Nonnegative Matrix Factorization (NMF, see [26,27]): NMF is a variant of SVD involving nonnegative modes and expansion coefficients. In this approach, we build a family of non-negative functions $\{\mathbf{b}_i\}_{i=1}^n$ and we approximate each $\boldsymbol{\beta}_i$ with a linear combination
$$\boldsymbol{\beta}_i^{\mathtt{NMF}} = \sum_{j=1}^{n} a_{i,j}\mathbf{b}_j, \quad 1 \leq i \leq K, \tag{8}$$
where for every $1 \leq i \leq K$ and $1 \leq j \leq n$, the coefficients $a_{i,j} \geq 0$ and the basis function $\mathbf{b}_j \geq 0$. In other words, we solve the following constrained optimization problem:
$$(W^*, B^*) \in \underset{(W,B)\in\mathbb{R}_+^{K\times n}\times\mathbb{R}_+^{n\times Q}}{\arg\min} \|M_{\mathcal{B}} - WB\|_F^2.$$
We refer readers to [27] for further details on the NMF and its numerical aspects.

3. The Enlarged Nonnegative Greedy (ENG) algorithm with projection on an extended cone of positive functions: We now present our novel model reduction method, which is of interest in itself as it allows reduced models that preserve positivity and even other types of bounds to be built. The method stems from the observation that NMF approximates functions in the cone of positive functions of span$\{\mathbf{b}_i \geq 0\}_{i=1}^n$ since it imposes $a_{i,j} \geq 0$ in the linear combination (8). However, note that the positivity of the linear combination is not equivalent to the positivity of the coefficients $a_{i,j}$ since there are obviously linear combinations involving very small $a_{i,j} < 0$ for some j which may still deliver a nonnegative linear combination $\sum_{j=1}^n a_{i,j}\mathbf{b}_j$. We can thus widen the cone of linear combinations yielding positive values by carefully including these negative coefficients $a_{i,j}$. One possible strategy for this is proposed in Algorithm 1, which describes a routine that we call `Enlarge_Cone`. The routine
$$\{\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_n\} = \mathtt{Enlarge_Cone}[\{\mathbf{b}_1, \ldots, \mathbf{b}_n\}, \varepsilon]$$
takes a set of nonnegative functions $\{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ as an input and modifies each function $\mathbf{b}_i$ by iteratively adding negative linear combinations of the other basis functions $\mathbf{b}_j$ for $j \neq i$ (see line 11 of the routine). The coefficients are chosen in an optimal way so as to maintain the positivity of the final linear combination while minimizing the L^∞-norm. The algorithm returns a set of functions
$$\boldsymbol{\psi}_i = \mathbf{b}_i - \sum_{j\neq i} \sigma_j^i \mathbf{b}_j, \quad \forall i \in \{1, \ldots, n\}$$
with $\sigma_j^i \geq 0$. Note that the algorithm requires the setting of a tolerance parameter $\varepsilon > 0$ for the computation of the σ_j^i.
Once we have run `Enlarge_Cone`, the approximation of any function $\boldsymbol{\beta}$ is then sought as
$$\boldsymbol{\beta}^{(EC)} = \underset{c_1,\ldots,c_n\geq 0}{\arg\min} \|\boldsymbol{\beta} - \sum_{j=1}^{n} c_j\boldsymbol{\psi}_j\|_{L^2([0,T+\tau])}^2$$
We emphasize that the routine is valid for any set of nonnegative input functions. We can thus apply `Enlarge_Cone` to the functions $\{\mathbf{b}_i \geq 0\}_{i=1}^n$ from NMF but also to the functions selected by a greedy algorithm such as the following:

- For $n = 1$, find
$$\mathbf{b}_1 = \underset{\boldsymbol{\beta}\in\{\boldsymbol{\beta}_i\}_{i=1}^K}{\arg\max} \|\boldsymbol{\beta}\|^2_{L^2([0,T+\tau])}$$
- At step $n > 1$, we have selected the set of functions $\{\mathbf{b}_1, \ldots, \mathbf{b}_{n-1}\}$. We next find
$$\mathbf{b}_n = \underset{\boldsymbol{\beta}\in\{\boldsymbol{\beta}_i\}_{i=1}^K}{\arg\max} \min_{c_1,\ldots,c_n\geq 0} \|\boldsymbol{\beta} - \sum_{j=1}^{n-1} c_j \mathbf{b}_j\|^2_{L^2([0,T+\tau])}$$

In our numerical tests, we call the Enlarged Nonnegative Greedy (ENG) method the routine involving the above greedy algorithm combined with our routine `Enlarge_Cone`.

Algorithm 1 `Enlarge_Cone`$[\{\mathbf{b}_1, \ldots, \mathbf{b}_n\}, \varepsilon] \to \{\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_n\}$.

Input: Set of nonnegative functions $\{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$. Tolerance $\varepsilon > 0$.
  for $i \in \{1, \ldots, n\}$ **do**
    Set $\sigma^i_j = 0, \quad \forall j \neq i$.
    for $\ell \in \{1, \ldots, n\}$ **do**
      $\alpha^{i,*}_\ell = \arg\max\{\alpha \geq 0 \ : \ \left[\mathbf{b}_i - \sum_{j\neq i} \sigma^i_j \mathbf{b}_j - \alpha \mathbf{b}_\ell\right](t) > 0, \ \forall t \in [0, T+\tau]\}$
      $\sigma^i_\ell = \sigma^i_\ell + \frac{\alpha^{i,*}_\ell}{2}$
      while $\alpha^{i,*}_\ell \geq \varepsilon$ **do**
        $\alpha^{i,*}_\ell = \arg\max\{\alpha \geq 0 \ : \ \left[\mathbf{b}_i - \sum_{j\neq i} \sigma^i_j \mathbf{b}_j - \alpha \mathbf{b}_\ell\right](t) > 0, \ \forall t \in [0, T+\tau]\}$
        $\sigma^i_\ell = \sigma^i_\ell + \frac{\alpha^{i,*}_\ell}{2}$
      end while
    end for
    $\boldsymbol{\psi}_i = \mathbf{b}_i - \sum_{j\neq i} \sigma^i_j \mathbf{b}_j$
  end for
Output: $\{\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_n\}$

We remark that, if we work with positive functions that are upper bounded by a constant $L > 0$, we can ensure that the approximations, denoted as Ψ, and written as a linear combination of basis functions, will also be between these bounds 0 and L by defining on the one hand, and as we have just done, a cone of positive functions generated by the above family $\{\psi_i\}_i$, and on the other hand, considering the base of the functions $L - \varphi$, φ to be above the set all greedy elements of the reduced basis, to which we also apply the enlargement of these positive functions. We then require the approximation to be written as a positive combination of the first (positive) functions and for $L - \Psi$ to also be written as a combination with positive components in the second basis.
In this frame, the approximation appears under the form of a least-square approximation with $2n$ linear constraints on the n coefficients, expressing the fact that the coefficients are nonnegative in the two above transformed bases.

In addition to the previous basis functions, it is possible to include more general/generic basis functions such as polynomial, radial, or wavelet functions in order to guarantee simple forecasting trends. For instance, one can add affine functions in order to include the possibility of predicting with a simple linear extrapolation to the range of possible forecasts offered by the reduced model. Given the overall exponential behavior of the health data, we have added an exponential function of the form $b_0(t) = \xi \exp(-\xi' t)$ (or $g_0(t) = \psi \exp(-\psi' t)$) to the reduced basis functions $\{b_1, \ldots, b_n\}$ (or $\{g_1, \ldots, g_n\}$) where the real-valued nonnegative parameters ξ, ξ', ψ, ψ' are obtained through a standard exponential regression from $\boldsymbol{\beta}^*_{\text{obs}}$ (or γ^*_{obs}) associated with the targeted profiles of infectious people;

that is, the profiles defined in the time interval $[0, T]$ that should be extrapolated to $]T, \tau]$. In other words, the final reduced models that we use for forecasting are

$$\mathrm{B}_{n+1} = \mathrm{span}\{b_0, b_1, \ldots, b_n\}, \quad \mathrm{G}_{n+1} = \mathrm{span}\{g_0, g_1, \ldots, g_n\} \subset L^\infty([0, T+\tau], \mathbb{R}),$$

Indeed, including the exponential functions in the reduced models gives easy access to the overall behavior of the parameters β and γ; the rest of the basis functions generated from the training sets catch the higher-order approximations and allow then to improve the extrapolation.

Remark 1. ***Reduced models on*** $\mathcal{I} = \{\boldsymbol{I}(\mu) : \mu \in \mathcal{P}\}$ ***and*** $\mathcal{R} = \{\boldsymbol{R}(\mu) : \mu \in \mathcal{P}\}$ *Instead of applying model reduction to the sets $\mathcal{B}$ and $\mathcal{G}$, as we do in our approach, we could apply the above techniques directly to the sets of solutions $\mathcal{I}$ and $\mathcal{R}$ of the SIR models with time-dependent coefficients in $\mathcal{B}$ and $\mathcal{G}$. In this case, the resulting approximation would however not follow SIR dynamics.*

3. Methodology for Multiple Regions Including Population Mobility Data

The forecasting method of Section 2.2 for a single region can be extended to the treatment of multiple regions involving population mobility. The prediction scheme is based on a multi-regional SIR with time-dependent coefficients. Compared to other more detailed models, the main advantage of our approach is that it drastically reduces the number of parameters to be estimated. Indeed, detailed multi-regional models such as multi-regional extensions of the above `SEI5CHRD` and `SE2IUR` models from Section 2.3 require a number of parameters that quickly grows with the number P of regions involved. Their calibration thus requires large amounts of data which, in addition, may be unknown, very uncertain, or not available. In a forthcoming paper, we will apply the fully multi-regional setting for the post-lockdown period.

The structure of this section is the same as above for the case of a single region. In Section 3.1, we begin by introducing the multi-regional SIR model with time-dependent coefficients and associated detailed models. As with any multi-regional model, mobility data are required as input data, and the nature and level of detail of the available data motivates certain choices regarding the modeling of the multi-regional SIR (as well as the other detailed models). We then present in Section 3.2 the general pipeline, in which we emphasize the high modularity of the approach.

3.1. Multi-Regional Compartmental Models

In the spirit of fluid flow modeling, there are essentially two ways of describing mobility between regions:

- In a Eulerian description, we take the regions as fixed references for which we record incoming and outgoing travels;
- In a Lagrangian description, we follow the motion of people living in a certain region and record their travels in the territory. We can expect this modeling to be more informative regarding the geographical spread of the disease, but it comes at the cost of additional details regarding the home region of the population.

Note that both descriptions hold at any coarse or fine geographical level, in the sense that what we call the regions could be taken to be full countries, departments within a country, or very small geographical partitions of a territory. We next describe the multi-regional SIR models with the Eulerian and Lagrangian descriptions of population fluxes, which form- the output of our methodology.

3.1.1. Multi-Regional SIR Models with Time-Dependent Parameters

Eulerian description of population flux: Assume that we have P regions and the number of people in region i is N_i for $i = 1, \ldots, P$. Due to mobility, the population in each

region varies, so N_i depends on t. However, the total population is assumed to be constant and equal to N; that is,

$$N = \sum_{i=1}^{P} N_i(t), \quad \forall t \geq 0.$$

For any $t \geq 0$, let $\lambda_{i\to j}(t) \in [0,1]$ be the probability that people from i travel to j at time t. In other words, $\lambda_{i\to j}(t)N_i(t)\delta t$ is the number of people from region i that have travelled to region j between time t and $t+\delta t$. Note that we have

$$\sum_{j=1}^{P} \lambda_{i\to j}(t) = 1, \quad \forall t \geq 0.$$

Since, for any $\delta t \geq 0$,

$$N_i(t+\delta t) = N_i(t) - \sum_{j\neq i} \lambda_{i\to j}(t)N_i(t)\delta t + \sum_{j\neq i} \lambda_{j\to i}(t)N_j(t)\delta t$$

dividing by δt and taking the limit $\delta t \to 0$ yields

$$\frac{\mathrm{d}N_i}{\mathrm{d}t}(t) = -\sum_{j\neq i} \lambda_{i\to j}(t)N_i(t) + \sum_{j\neq i} \lambda_{j\to i}(t)N_j(t).$$

Note that we have

$$\sum_{i=1}^{P} \frac{\mathrm{d}N_i}{\mathrm{d}t}(t) = 0, \quad \forall t \geq 0.$$

Thus, $\sum_i N_i(t) = \sum_i N_i(0) = N$, which is consistent with our assumption that the total population is constant.

The time evolution of the N_i is known in this case if we are given the $\lambda_{i\to j}(t)$ from Eulerian mobility data. In addition to these mobility data, we also have the data of the evolution of infected and removed people, and our goal is to fit a multi-regional SIR model that is in accordance with this data. Thus, we propose the following model.

Denoting S_i, I_i and R_i as the number of susceptible, infectious and removed people in region i at time t, we first have the relation

$$N_i(t) = S_i(t) + I_i(t) + R_i(t) \quad \Leftrightarrow \quad 1 = \frac{S_i(t)}{N_i(t)} + \frac{I_i(t)}{N_i(t)} + \frac{R_i(t)}{N_i(t)}.$$

Note that from the second relation, it follows that

$$0 = \frac{\mathrm{d}}{\mathrm{d}t}\frac{S_i}{N_i} + \frac{\mathrm{d}}{\mathrm{d}t}\frac{I_i}{N_i} + \frac{\mathrm{d}}{\mathrm{d}t}\frac{R_i}{N_i}. \tag{9}$$

To model the evolution between compartments, one possibility is the following SIR model:

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\frac{S_i}{N_i} &= -\left(\beta_i\lambda_{i\to i}\frac{I_i}{N_i} + \sum_{j\neq i}\beta_j\lambda_{j\to i}\frac{I_j}{N_j}\right)\frac{S_i}{N_i} \\
\frac{\mathrm{d}}{\mathrm{d}t}\frac{I_i}{N_i} &= -\frac{\mathrm{d}}{\mathrm{d}t}\frac{S_i}{N_i} - \gamma_i\frac{I_i}{N_i} \\
\frac{\mathrm{d}}{\mathrm{d}t}\frac{R_i}{N_i} &= \gamma_i\frac{I_i}{N_i},
\end{aligned} \tag{10}$$

The parameters β_i, γ_i, N_i depend on t, but we have omitted this dependence for ease of reading. Introducing the compartmental densities

$$s_i = \frac{S_i}{N_i}, \quad i_i = \frac{I_i}{N_i}, \quad r_i = \frac{R_i}{N_i},$$

the system equivalently reads

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} s_i &= -\left(\beta_i \lambda_{i\to i} i_i + \sum_{j\neq i} \beta_j \lambda_{j\to i} i_j \right) s_i \\ \frac{\mathrm{d}}{\mathrm{d}t} i_i &= -\frac{\mathrm{d}}{\mathrm{d}t} s_i - \gamma_i i_i \\ \frac{\mathrm{d}}{\mathrm{d}t} r_i &= \gamma_i i_i, \end{aligned} \tag{11}$$

Before going further, some points should be highlighted:

- The model is consistent in the sense that it satisfies (9), and when $P = 1$, we recover the traditional SIR model;
- Under lockdown measures, $\lambda_{i\to j} \approx \delta_{i,j}$ and the population $N_i(t)$ remains practically constant. As a result, the evolution of each region is decoupled from the others, and each region can be addressed with a mono-regional approach;
- The use of β_j in Equation (11) is debatable. When people from region j arrive in region i, it may be reasonable to assume that the contact rate is β_i;
- The use of $\lambda_{j\to i}$ in Equation (11) is also very debatable. The probability $\lambda_{j\to i}$ was originally defined to account for the mobility of people from region j to region i without specifying the compartment. However, in Equation (11), we need the probability of mobility of infectious people from region j to region i, which we denote by $\mu_{j\to i}$ in the following. It seems reasonable to think that $\mu_{j\to i}$ may be smaller than $\lambda_{j\to i}$, because as soon as people become symptomatic and suspect an illness, they will probably not move. Two possible options would be as follows:
  - We could try to estimate $\mu_{j\to i}$. If symptoms arise, for example, 2 days after infection and if people recover in 15 days on average, then we could say that $\mu_{j\to i} = 2/15\lambda_{j\to i}$.
  - As the above seems to be quite empirical, another option would be to use $\lambda_{j\to i}$ and absorb the uncertainty in the values of the β_j that can be fitted.
- We choose not to add mobility in the R compartment as this does not modify the dynamics of the epidemic spread; only adjustments in the population sizes are needed.

Lagrangian description of population flux: We call the above description Eulerian because we have fixed the regions as a fixed reference. Another possibility is to follow the trajectories of inhabitants of each region, in the same spirit as when we follow the trajectories of fluid particles.

Let S_i, I_i, and R_i now be the number of susceptible, infectious and removed people who are resident in region $i, i = 1, \ldots P$. It is reasonable to assume that $S_i(t) + I_i(t) + R_i(t)$ is constant in time. However, not all the residents of region i may be in that region at time t. Let $\lambda^{(i)}_{j\to k}(t)$ be the probability that susceptible people resident in i travel from region j to region k at time t. With this notation, $\lambda^{(i)}_{i\to i}(t)$ is the probability that susceptible people resident at i remain in region i at time t. Similarly, let $\mu^{(i)}_{j\to k}(t)$ be the probability that infectious people resident in i travel from region j to k at time t. Thus, the total number of susceptible and infectious people that are in region i at time t is

$$\mathcal{S}_i(t) = \sum_{k=1}^{P}\sum_{j=1}^{P}\left(\lambda_{j\to i}^{(k)}(t) - \lambda_{i\to j}^{(k)}(t)\right)S_k(t)$$
$$\mathcal{I}_i(t) = \sum_{k=1}^{P}\sum_{j=1}^{P}\left(\mu_{j\to i}^{(k)}(t) - \mu_{i\to j}^{(k)}(t)\right)S_k(t)$$

We can thus write the evolution over S_i, I_i, R_i as

$$\begin{aligned}\frac{\mathrm{d}S_i}{\mathrm{d}t} &= -\sum_{j=1}^{P}\sum_{k=1}^{P}\beta_k(t)\lambda_{j\to k}^{(i)}(t)S_i(t)\mathcal{I}_k(t)\\ \frac{\mathrm{d}I_i}{\mathrm{d}t} &= -\frac{\mathrm{d}S_i}{\mathrm{d}t} - \gamma_i(t)I_i(t)\\ \frac{\mathrm{d}R_i}{\mathrm{d}t} &= \gamma_i(t)I_i(t)\end{aligned} \tag{12}$$

Note that $S_i(t) + I_i(t) + R_i(t)$ is constant, which is consistent with the fact that, in our model,

$$\frac{\mathrm{d}}{\mathrm{d}t}(S_i + I_i + R_i) = 0.$$

We emphasize that, to implement this model, the Lagrangian mobility data $\lambda_{j\to k}^{(i)}$ are required for all $(i, j, k) \in \{1, \ldots, P\}^3$.

Notation: In the following, we gather the compartmental variables in vectors

$$\vec{\mathbf{S}} := (\mathbf{S})_{i=1}^{P}, \quad \vec{\mathbf{I}} := (\mathbf{I})_{i=1}^{P}, \quad \vec{\mathbf{R}} := (\mathbf{R})_{i=1}^{P}$$

as well as the time-dependent coefficients

$$\vec{\boldsymbol{\beta}} = (\boldsymbol{\beta})_{i=1}^{P}, \quad \vec{\boldsymbol{\gamma}} = (\boldsymbol{\gamma})_{i=1}^{P}.$$

For any $\vec{\boldsymbol{\beta}}$ and $\vec{\boldsymbol{\gamma}} \in (L^\infty([0, T]))^P$, we denote by

$$(\vec{\mathbf{S}}, \vec{\mathbf{I}}, \vec{\mathbf{R}}) = \texttt{Multiregional_SIR}(\vec{\mathbf{S}}(0), \vec{\mathbf{I}}(0), \vec{\mathbf{R}}(0), \vec{\boldsymbol{\beta}}, \vec{\boldsymbol{\gamma}}, [0, T])$$

the output of any of the above multi-regional SIR models. For simplicity, in what follows, we omit the initial condition in the notation.

3.1.2. Detailed Multi-Regional Models with Constant Coefficients

In the spirit of the multi-regional SIR, one can formulate detailed multi-regional versions of more detailed models such as those introduced in Section 2.1. We omit the details for the sake of brevity.

3.2. Forecasting for Multiple Regions with Population Mobility

Similar to the mono-regional case, we assume that we are given health data in $[0, T]$ in all regions. The observed data in region i are the series of infected people, denoted as I_i^{obs}, and recovered people, denoted as R_i^{obs}. They are usually given at a national or a regional scale and on a daily basis.

We propose to fit the data and provide forecasts with SIR models with time-dependent parameters $\boldsymbol{\beta}_i$ and γ_i for each region i. As in the mono-regional case, we can prove that such a simple family possesses optimal fitting properties for our purposes. In the current case, the cost function reads

$$\mathcal{J}(\vec{\boldsymbol{\beta}}, \vec{\boldsymbol{\gamma}} \,|\, \vec{\mathbf{I}}_{\text{obs}}, \vec{\mathbf{R}}_{\text{obs}}, [0,T]) := \sum_{i=1}^{P} \int_0^T \Big(|I_i(t) - I_i^{\text{obs}}(t)|^2 + |R_i(t) - R_i^{\text{obs}}(t)|^2 \Big) \mathrm{d}t$$
$$\text{such that } \Big(\vec{\mathbf{S}}, \vec{\mathbf{I}}, \vec{\mathbf{R}}\Big) = \texttt{Multiregional_SIR}\Big(\vec{\boldsymbol{\beta}}, \vec{\boldsymbol{\gamma}}, [0,T]\Big),$$

and the fitting problem is the optimal control problem of finding

$$J^* = \inf_{\vec{\boldsymbol{\beta}}, \vec{\boldsymbol{\gamma}} \in (L^\infty([0,T]))^P \times (L^\infty([0,T]))^P} \mathcal{J}(\vec{\boldsymbol{\beta}}, \vec{\boldsymbol{\gamma}} \,|\, \vec{\mathbf{I}}_{\text{obs}}, \vec{\mathbf{R}}_{\text{obs}}, [0,T]). \tag{13}$$

The following proposition ensures the existence of a unique minimizer under certain conditions. To prove this, it is useful to remark that any of the above multi-regional SIR models (see (11) and (12)) can be written in the general form

$$\begin{aligned} \frac{\mathrm{d}\vec{\mathbf{S}}}{\mathrm{d}t} &= \mathrm{M}(\Lambda(t), \vec{\mathbf{S}}(t), \vec{\mathbf{I}}(t))\vec{\boldsymbol{\beta}} \\ \frac{\mathrm{d}\vec{\mathbf{I}}}{\mathrm{d}t} &= -\frac{\mathrm{d}\vec{\mathbf{S}}}{\mathrm{d}t} - \mathrm{diag}(\mathbf{I}(t))\,\vec{\boldsymbol{\gamma}} \\ \frac{\mathrm{d}\vec{\mathbf{R}}}{\mathrm{d}t} &= \mathrm{diag}(\mathbf{I}(t))\vec{\boldsymbol{\gamma}}, \end{aligned}$$

where, by a slight change of notation, the vectors $\vec{\mathbf{S}}$, $\vec{\mathbf{I}}$ and $\vec{\mathbf{R}}$ are the densities of population in the case of the Eulerian approach (see Equation (11)). They are classical population numbers in the case of the Lagrangian approach (see Equation (12)). $\mathrm{diag}(\mathbf{I}(t))$ is the $P \times P$ diagonal matrix with diagonal entries given by the vector $\mathbf{I}(t)$. $\mathrm{M}(\Lambda(t), \vec{\mathbf{S}}(t), \vec{\mathbf{I}}(t))$ is a matrix of size $P \times P$ that depends on the vectors of susceptible and infectious people $\vec{\mathbf{S}}(t), \vec{\mathbf{I}}(t)$ and on the mobility matrix Λ. In the case of the Eulerian description, $\Lambda(t) = (\lambda_{i,j}(t))_{1 \le i,j \le P}$ and in the case of the Lagrangian approach $\Lambda(t)$ is the $P \times P \times P$ tensor $\Lambda(t) = (\lambda_{j,k}^{(i)}(t))_{1 \le i,j,k \le P}$. For example, in the case of the Eulerian model (12), the matrix M reads

$$\mathrm{M}(\Lambda(t), \vec{\mathbf{S}}(t), \vec{\mathbf{I}}(t)) = -\,\mathrm{diag}(\vec{\mathbf{S}}(t))\Lambda^T\,\mathrm{diag}(\vec{\mathbf{I}}(t)) = -(S_i \lambda_{i \to j} I_j)_{1 \le i,j \le P} \tag{14}$$

Proposition 2. *If $\mathrm{M}(\Lambda(t), \vec{\boldsymbol{S}}(t), \vec{\boldsymbol{I}}(t))$ is invertible for all $t \in [0,T]$, then there exists a unique minimizer $(\vec{\boldsymbol{\beta}}^*, \vec{\boldsymbol{\gamma}}^*)$ to problem (13).*

Proof. Since we assume that $\mathrm{M}(\Lambda(t), \vec{\mathbf{S}}(t), \vec{\mathbf{I}}(t))$ is invertible for every $t \in [0,T]$, we can set

$$\begin{cases} \vec{\boldsymbol{\beta}}^*(t) & := M^{-1}(t)\frac{\mathrm{d}\vec{\mathbf{S}}}{\mathrm{d}t} \\ \vec{\boldsymbol{\gamma}}^*(t) & := \mathrm{diag}^{-1}(I(t))\frac{\mathrm{d}\vec{\mathbf{R}}}{\mathrm{d}t} \end{cases}$$

or equivalently

$$\begin{cases} \vec{\boldsymbol{\beta}}^*(t) & := M^{-1}(t)\frac{\mathrm{d}\vec{\mathbf{S}}}{\mathrm{d}t} \\ \vec{\boldsymbol{\gamma}}^*(t) & := -\,\mathrm{diag}^{-1}(I(t))\Big(\frac{\mathrm{d}\vec{\mathbf{I}}}{\mathrm{d}t} + \mathrm{M}(\Lambda(t), \vec{\mathbf{S}}(t), \vec{\mathbf{I}}(t))\vec{\boldsymbol{\beta}}^*\Big) \end{cases}$$

so that

$$(\vec{\mathbf{S}}_{\text{obs}}, \vec{\mathbf{I}}_{\text{obs}}, \vec{\mathbf{R}}_{\text{obs}}) = \texttt{Multiregional_SIR}\Big(\vec{\boldsymbol{\beta}}^*, \vec{\boldsymbol{\gamma}}^*, [0,T]\Big)$$

and

$$\mathcal{J}(\vec{\boldsymbol{\beta}}^*, \vec{\boldsymbol{\gamma}}^* \,|\, \vec{\mathbf{I}}_{\text{obs}}, \vec{\mathbf{R}}_{\text{obs}}, [0,T]) = 0$$

which implies that $J^* = 0$. □

Before continuing, let us comment on the invertibility of $\mathrm{M}(\Lambda(t), \vec{\mathbf{S}}(t), \vec{\mathbf{I}}(t))$ which is necessary in Proposition 2. A sufficient condition to ensure this is if the matrix is diagonally dominant row-wise or column-wise. This yields certain conditions on the mobility matrix

$\Lambda(t)$ with respect to the values of $\vec{\mathbf{S}}(t), \vec{\mathbf{I}}(t)$. For example, if M is defined as in Equation (14), the matrix is diagonally dominant in each row if for every $1 \leq i \leq P$,

$$\lambda_{i\to i} > \sum_{j\neq i} \lambda_{i\to j} \frac{I_j}{I_i}.$$

Similarly, if for every $1 \leq j \leq P$,

$$\lambda_{j\to j} > \sum_{i\neq j} \lambda_{i\to j} \frac{S_i}{S_j},$$

then the matrix is diagonally dominant for each column and guarantees invertibility. Note that any of the above conditions is satisfied in situations with little or no mobility where $\lambda_{i\to i} \approx \delta_{i,j}$.

Now that we have exactly defined the set-up for the multi-regional case, we can follow the same steps in Section 2.2 to derive forecasts involving model reduction for the time-dependent variables $\vec{\boldsymbol{\beta}}$ and $\vec{\boldsymbol{\gamma}}$.

4. Numerical Results

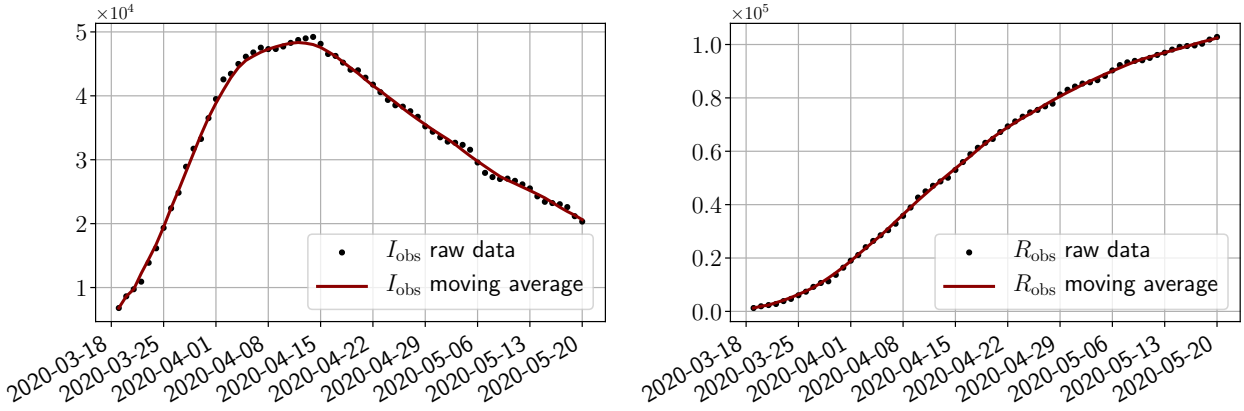
In this section, we apply our forecasting method to the ongoing COVID-19 pandemic, which spread in the year 2020 in France and started approximately in February. Particular emphasis is placed on the first pandemic wave, for which we consider the period from 19 March to 20 May 2020. Due to the lockdown imposed between 17 March and 11 May, inter-regional population mobility was drastically reduced in that period. Studies using anonymized Facebook data have estimated the reduction to be 80% (see [28]). As a result, it is reasonable to treat each region independently from the rest, and we apply the mono-regional setting in Section 2. Here, we focus on the case of the Paris region, and we report different forecasting errors obtained using the methods described in Section 2. Some forecasts are also shown for the second wave for the Paris region between 24 September and 25 November.

The numerical results are presented as follows. Section 4.1 explains the sources of health data. Sections 4.2.1 and 4.2.2 explore the results in detail and present a study of the forecasting power of the methods in a two-week horizon. Section 4.2.3 displays forecasts for the second wave. Section 4.2.4 aims to illustrate the robustness of the forecasting over longer periods of time. A discussion of the fitting errors of the methods is given in Appendix A. Additional results highlighting the accuracy of the forecasts are shown in Appendix B.

4.1. Data

We use public data from Santé Publique France (https://www.data.gouv.fr/en/datasets/donnees-hospitalieres-relatives-a-lepidemie-de-covid-19/) to get the numbers $I_{\text{obs}}(t)$ of infected and $R_{\text{obs}}(t)$ of removed people. As shown in Figure 3, the raw data present some oscillations at the scale of a week, which are due to administrative delays for the cases to be officially reported by hospitals. For our methodology, we have smoothed the data by applying a 7 day moving average filter. In order to account for the total number of infected people, we also multiply the data by an adjustment factor $f_{\text{adj}} = 15$ as stated in the literature (indeed, it is said in [29] that "of the 634 confirmed cases, a total of 306 and 328 were reported to be symptomatic and asymptomatic", and in [10], it is claimed that the probability of developing severe symptoms for a symptomatic patient is 0 for children, 0.1 for adults and 0.2 for seniors; thus, if one takes $p = 0.13$ as an approximate value of these probabilities, one may estimate the adjustment factor as $f_{\text{adj}} = \frac{634}{328} \times \frac{1}{p} \approx 15$). Obviously, this factor is uncertain and could be improved in the light of further retrospective studies of the outbreak. However, note that when $S \simeq N$, which is the case at the start of the epidemic, the impact of this factor is negligible in the dynamics as can be understood

from (3). In addition, since we use the same factor to provide a forecast of hospitalized people, the influence on the choice is minor.

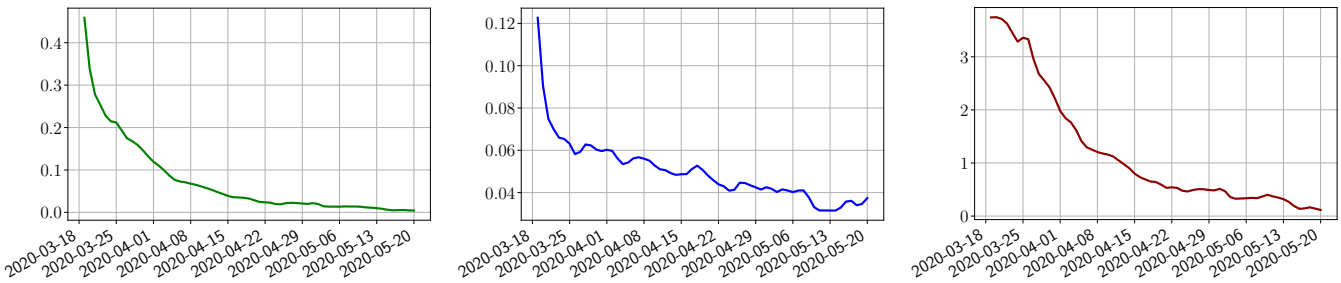


(**a**) Infected. (**b**) Removed.

Figure 3. Data from $t_0 = 19/03/2020$ **to** $T = 20/05/2020$.

4.2. Results

Using the observations $I_{obs}(t)$ and $R_{obs}(t)$, we apply a finite difference scheme in Formula (3) to derive $\beta^*_{obs}(t)$ and $\gamma^*_{obs}(t)$ for $t \in [0, T]$. Figure 4 shows the values of these parameters as well as the basic reproduction number $R^*_{0,obs} = \beta^*_{obs}/\gamma^*_{obs}$ for the first pandemic wave in Paris.



(**a**) β^*_{obs}. (**b**) γ^*_{obs}. (**c**) $R^*_{0,obs}$.

Figure 4. β^*_{obs}, γ^*_{obs}, $R^*_{0,obs} = \beta^*_{obs}/\gamma^*_{obs}$ deduced from the data from $t_0 = 19/03/2020$ to $T = 20/05/2020$.

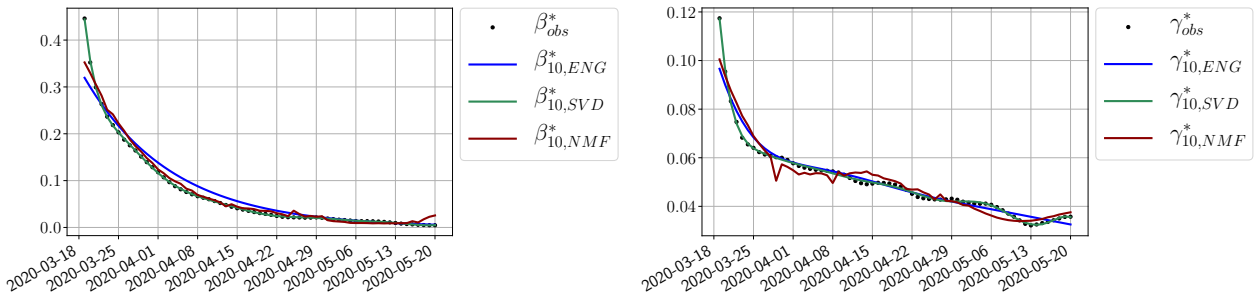
We next follow the steps presented in Section 2.2 to obtain the forecasts. In the learning phase (step 1), we use two parametric detailed models of `SE2IUR` and `SEI5CHRD` types to generate training sets $\mathcal{B}_{tr}$ and $\mathcal{G}_{tr}$ composed of $K = 2618$ training functions $\boldsymbol{\beta}(\mu)$ and $\boldsymbol{\gamma}(\mu)$ where μ are uniformly sampled in the set of parameters $\mathcal{P}_{tr}$ in the vicinity of the parameter values suggested in the literature [9,10]. Based on these training sets, we finish step 1 by building three types of reduced models: SVD, NMF and ENG (see Section 2.4).

Given the reduced bases B_n and G_n, we next search for the optimal $\boldsymbol{\beta}^*_n \in \mathrm{B}_n$ and $\boldsymbol{\gamma}^*_n \in \mathrm{G}_n$ that best fit the observations (step 2 of our procedure). For this fitting step, we consider two loss functions:

1. `routine-IR`: loss function $\mathcal{J}(\boldsymbol{\beta}, \boldsymbol{\gamma} \mid \mathbf{I}_{obs}, \mathbf{R}_{obs}, [0, T])$ from (1),
2. `routine-`$\boldsymbol{\beta\gamma}$: loss function $\widetilde{\mathcal{J}}(\boldsymbol{\beta}, \boldsymbol{\gamma} \mid \boldsymbol{\beta}^*_{obs}, \boldsymbol{\gamma}^*_{obs}, [0, T])$ from (6)

We study the performance of each of the three reduced models and the impact of the dimension n of the reduced model in terms of the fitting error. The presentation of these results is presented in Appendix A in order not to overload the main discussion. The main conclusion is that the fitting strategy using SVD-reduced bases provides smaller errors than NMF and ENG, especially when we increase the number of modes n. This is illustrated

in Figure 5 where we show the fittings obtained with `routine-`$\boldsymbol{\beta\gamma}$ and $n = 10$ for the first wave (from $t_0 = 19/03/2020$ to $T = 20/05/2020$). We observe that SVD is the best at fitting β^*_{obs} and γ^*_{obs}, while ENG produces a smoother fitting of the data. Although the smoother fitting with ENG yields larger fitting errors than SVD, we see in the next section that it yields better forecasts.



(**a**) β. (**b**) γ.

Figure 5. Fitting from $t_0 = 19/03/2020$ **to** $T = 20/05/2020$.

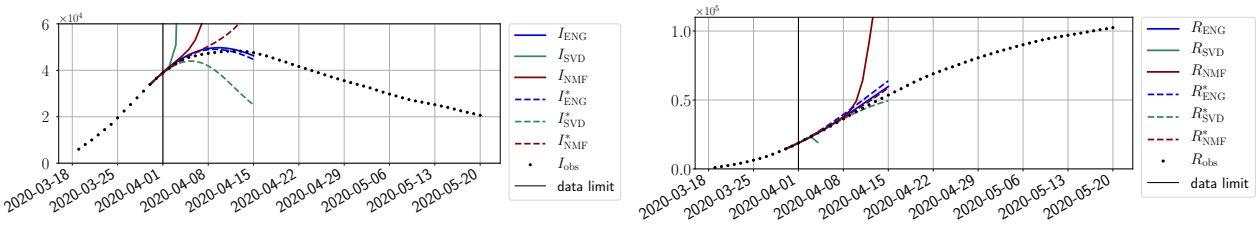
4.2.1. Forecasting for the First Pandemic Wave with a 14 Day Horizon

In this section, we illustrate the short-term forecasting behavior of our method. We consider a forecasting window of $\tau = 14$ days and we examine several different starting days in the course of the first pandemic wave. The results are shown in Figures 6–14. Recall that that the forecasting uses the coefficients of the reduced bases obtained by the fitting procedure but also the optimal initial condition of the forecast that minimizes the error on the three days prior to the start of the forecast. For each given fitting strategy (`routine-IR`, `routine-`$\boldsymbol{\beta\gamma}$) and each given type of reduced model (SVD, NMF, ENG), we have chosen to plot an average forecast computed with predictions using reduced dimensions $n \in \{5, 6, 7, 8, 9, 10\}$. This choice is a simple type of forecast combination, but of course other more elaborate aggregation options could be considered. The labels of the plots correspond to the following:

- I_{SVD}, I_{NMF}, I_{ENG}, R_{SVD}, R_{NMF}, R_{ENG} are the average forecasts obtained using `routine-`$\boldsymbol{\beta\gamma}$.
- I^*_{SVD}, I^*_{NMF}, I^*_{ENG}, R^*_{SVD}, R^*_{NMF}, R^*_{ENG} are the average forecasts obtained using `routine-IR`.

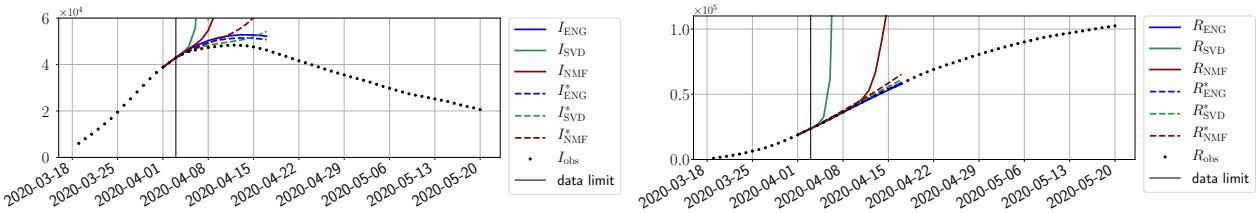
The main observation from Figures 6–14 is that the ENG-reduced model is the most robust and accurate forecasting method. Fitting ENG with `routine-IR` or `routine-`$\boldsymbol{\beta\gamma}$ does not seem to lead to large differences in the quality of the forecasts, but `routine-`$\boldsymbol{\beta\gamma}$ seems to provide slightly better results. This claim is further confirmed by the study of the numerical forecasting errors of the different methods shown in Appendix B.

Figures 6–14 also show that the SVD-reduced model is very unstable and provides forecasts that blow up. This behavior illustrates the dangers of overfitting, namely that a method with high fitting power may present poor forecasting power due to instabilities. In the case of SVD, the instabilities stem from the fact that approximations are allowed to take negative values. This is the reason why NMF, which incorporates the nonnegative constraint, performs better than SVD. One of the reasons why ENG outperforms NMF is the enlargement of the cone of nonnegative functions (see Section 2.4). It is important to note that, with ENG, the reduced bases are directly related to well-chosen virtual scenarios, whereas SVD and NMF rely on matrix factorization techniques that provide purely artificial bases. This makes forecasts from ENG more realistic and therefore more reliable.



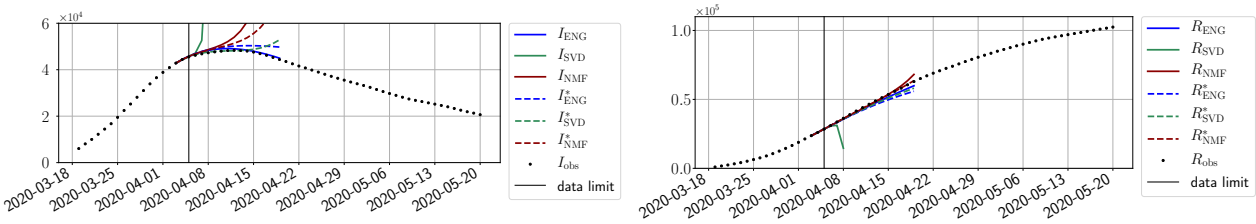
(**a**) Infected. (**b**) Removed.

Figure 6. Fourteen-day forecasts starting from $T = 01/04$.



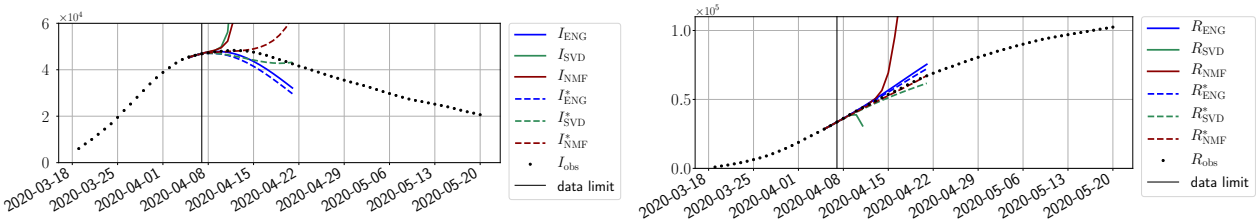
(**a**) Infected. (**b**) Removed.

Figure 7. Fourteen-day forecasts starting from $T = 03/04$.



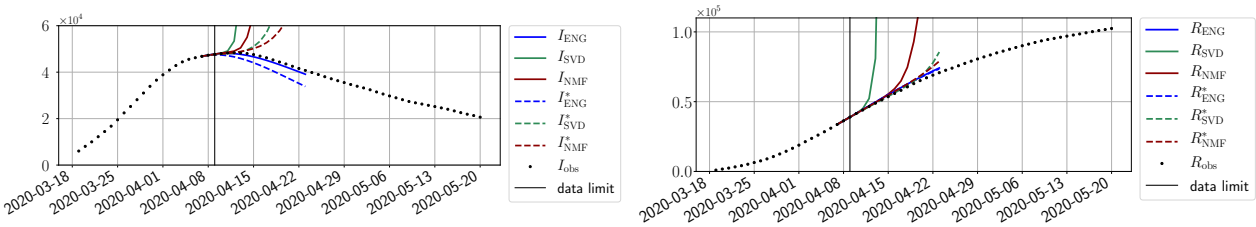
(**a**) Infected. (**b**) Removed.

Figure 8. Fourteen-day forecasts starting from $T = 05/04$.



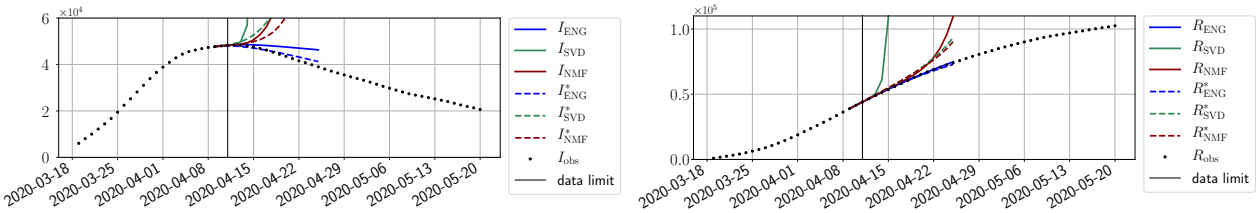
(**a**) Infected. (**b**) Removed.

Figure 9. Fourteen-day forecasts starting from $T = 07/04$.



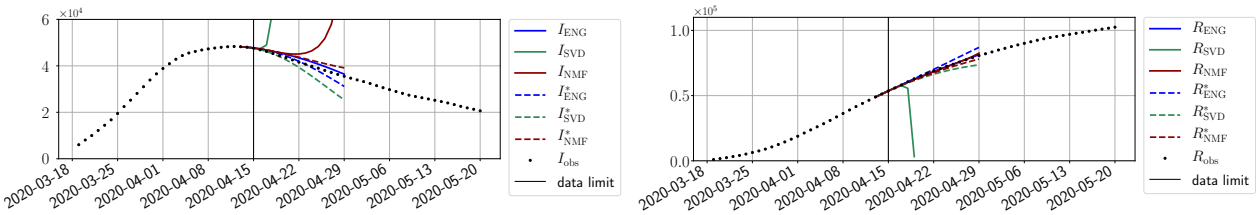
(**a**) Infected. (**b**) Removed.

Figure 10. Fourteen-day forecasts starting from $T = 09/04$.



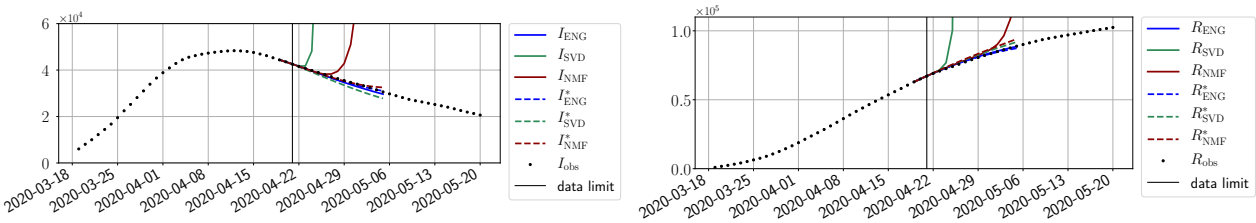
(**a**) Infected. (**b**) Removed.

Figure 11. Fourteen-day forecasts starting from $T = 11/04$.



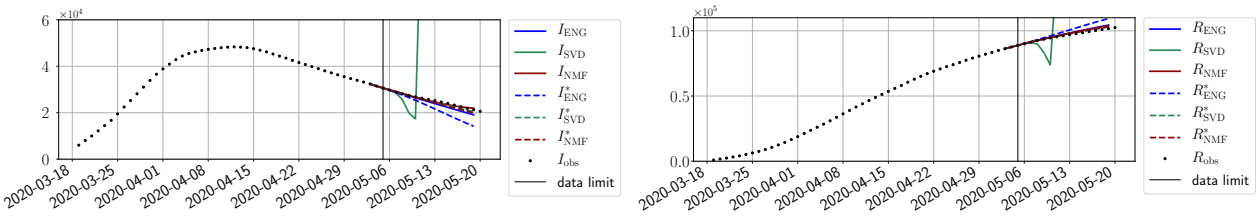
(**a**) Infected. (**b**) Removed.

Figure 12. Fourteen-day forecasts starting from $T = 15/04$.



(**a**) Infected. (**b**) Removed.

Figure 13. Fourteen-day forecasts starting from $T = 21/04$.



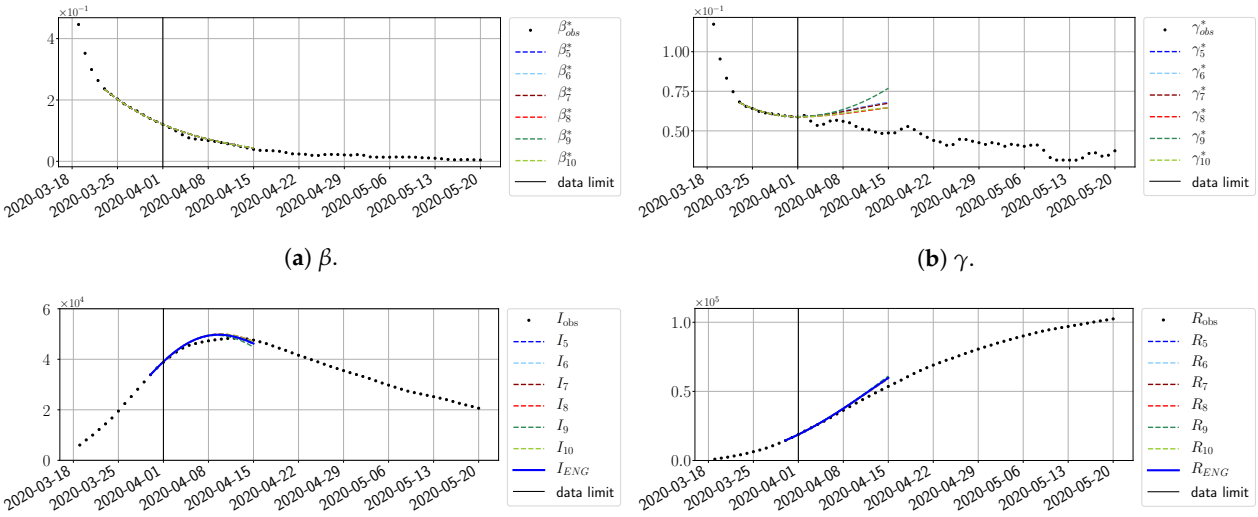
(**a**) Infected. (**b**) Removed.

Figure 14. Fourteen-day forecasts starting from $T = 05/05$.

4.2.2. Focus on the Forecasting with ENG

For our best forecasting method (`routine-`$\boldsymbol{\beta\gamma}$ using ENG), we plot in Figures 15–23 the forecasts for each dimension $n = 5$ to 10. The plots give the forecasts on a 14 day-ahead window for β, γ, and the resulting evolution of the infected I and removed R. We see that the method performs reasonably well for all values of n, proving that the results of the previous section with the averaged forecast are not compensating for spurious effects which could occur for certain values of n. We have chosen to display the inaccurate forecasts from 3 April, 7 April, and 11 April as they are among the worst predictions obtained using this method; however, it is important to mention that, despite the lack of accuracy in these cases, plausible epidemic behaviors remain, with different but realistic evolutions for β and γ compared to the actual evolution. Note that the method was able to predict the peak of the epidemic several days in advance (see Figure 15). We also observe that the prediction on γ

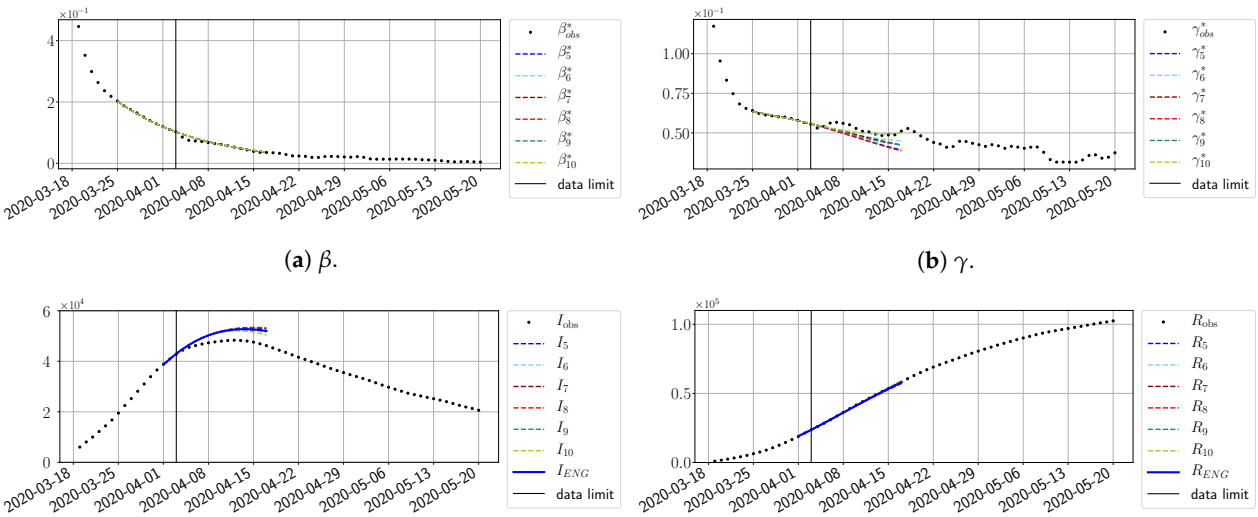
is difficult at all times due to the fact that γ^*_{obs} presents an oscillatory behavior. Despite this difficulty, the resulting forecasts for I and R are very satisfactory in general.



(**a**) β. (**b**) γ.

(**c**) Infected. (**d**) Removed.

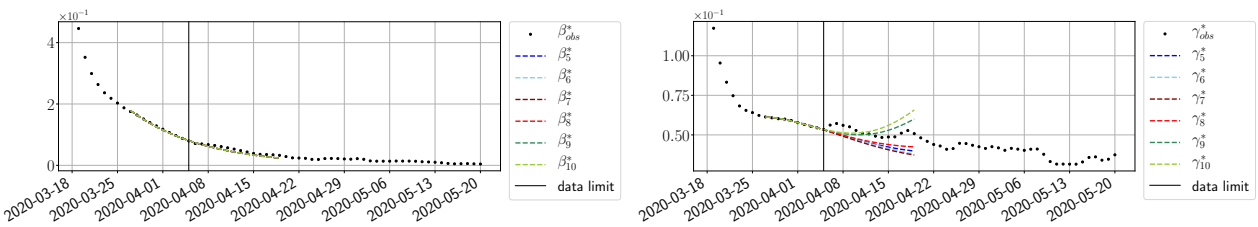
Figure 15. Enlarged Nonnegative Greedy (ENG) forecast from $T = 01/04$.



(**a**) β. (**b**) γ.

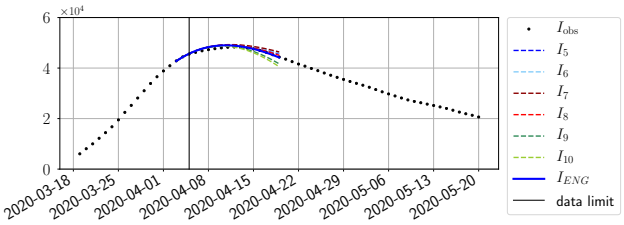
(**c**) Infected. (**d**) Removed.

Figure 16. ENG forecast from $T = 03/04$.

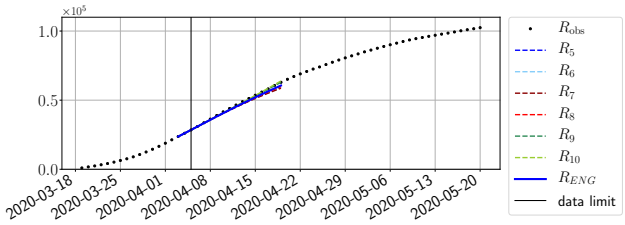


(**a**) β. (**b**) γ.

Figure 17. *Cont.*

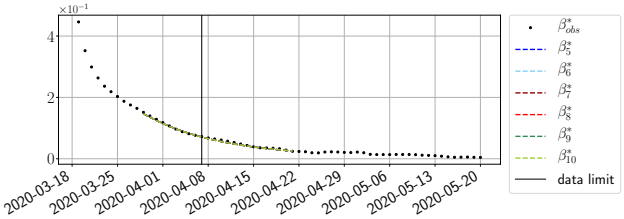


(**c**) Infected.

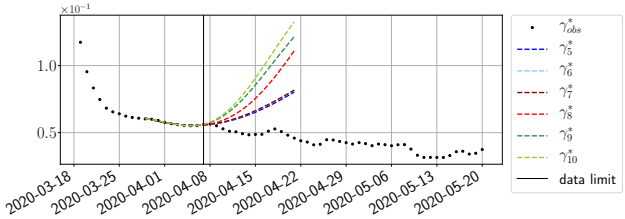


(**d**) Removed.

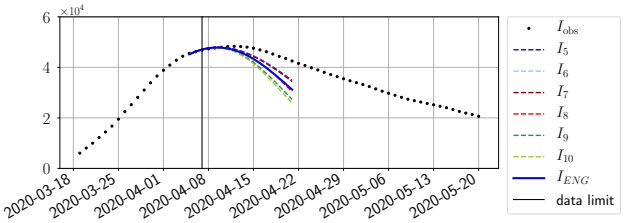
Figure 17. ENG forecast from $T = 05/04$.



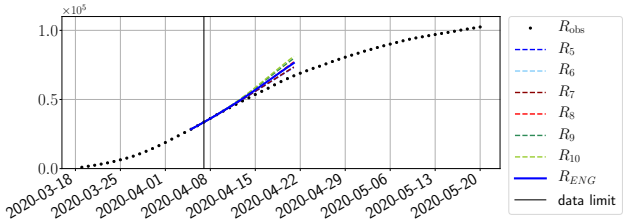
(**a**) β.



(**b**) γ.

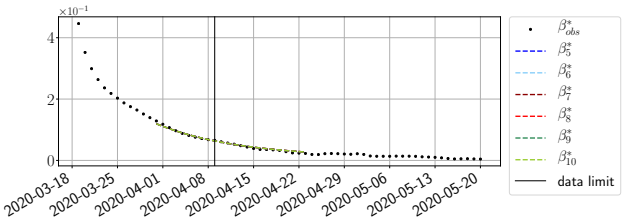


(**c**) Infected.

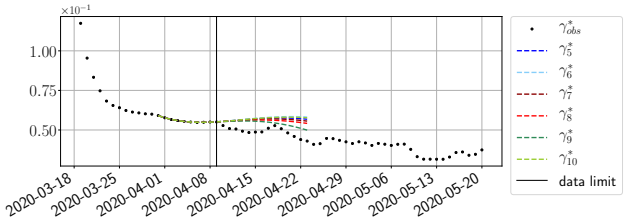


(**d**) Removed.

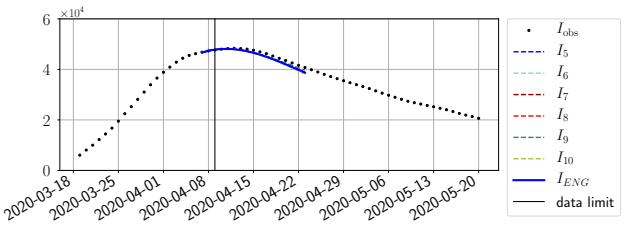
Figure 18. ENG forecast from $T = 07/04$.



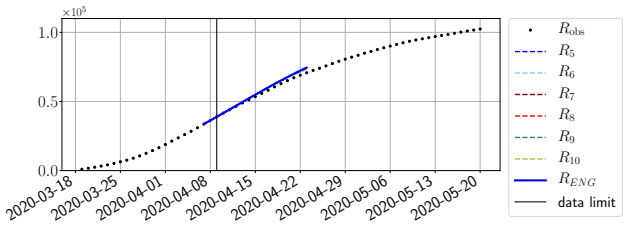
(**a**) β.



(**b**) γ.

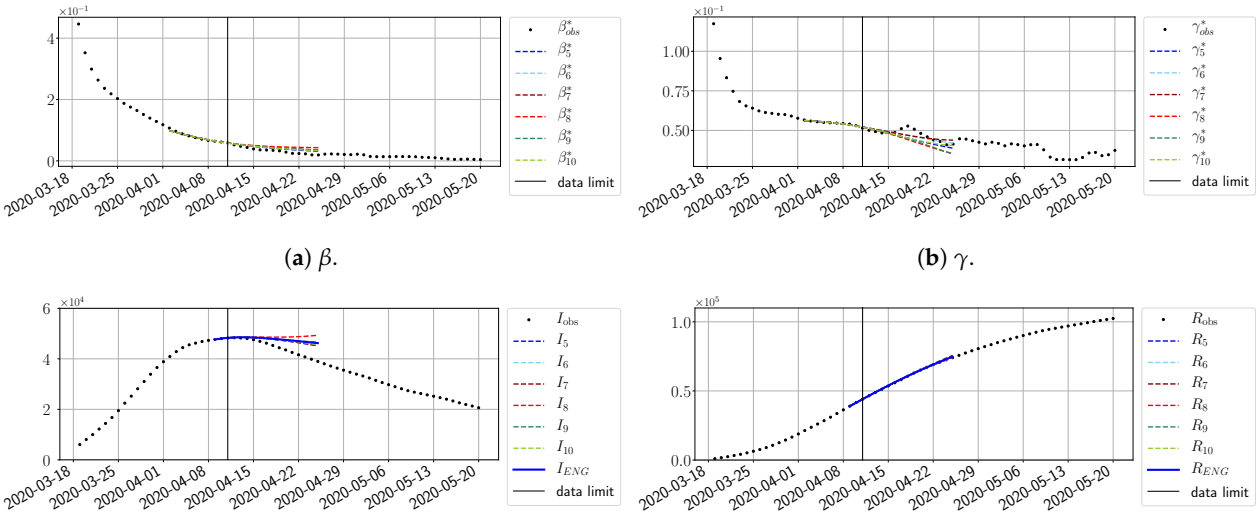


(**c**) Infected.



(**d**) Removed.

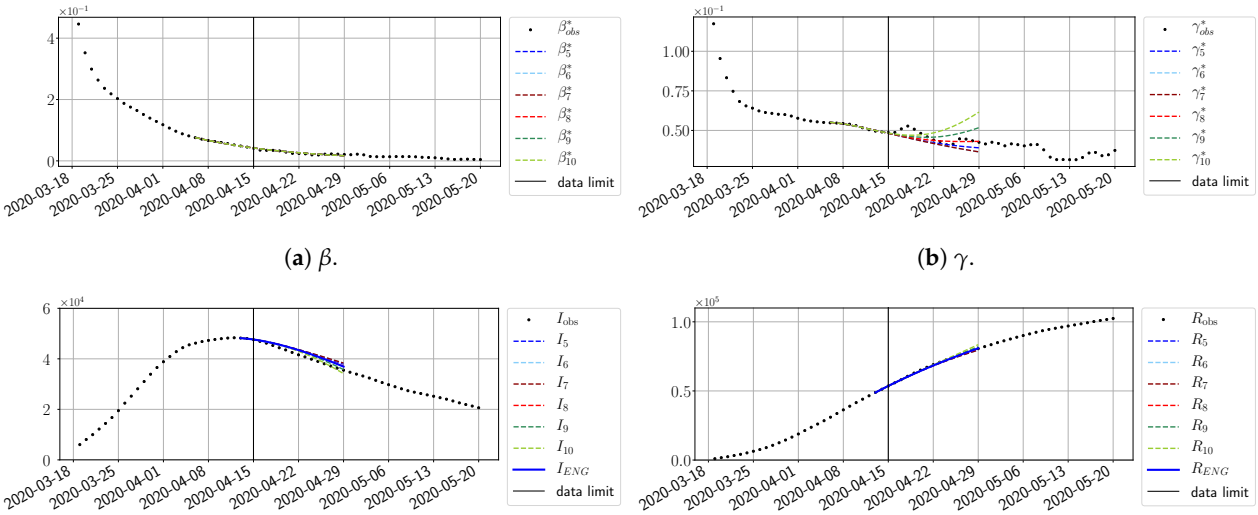
Figure 19. ENG forecast from $T = 09/04$.



(**a**) β. (**b**) γ.

(**c**) Infected. (**d**) Removed.

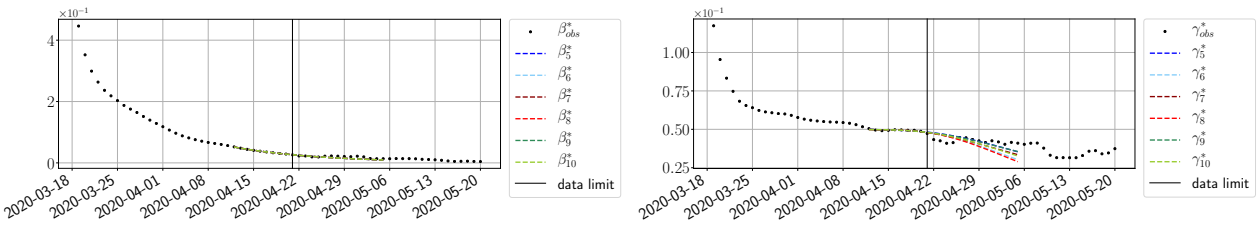
Figure 20. ENG forecast from $T = 11/04$.



(**a**) β. (**b**) γ.

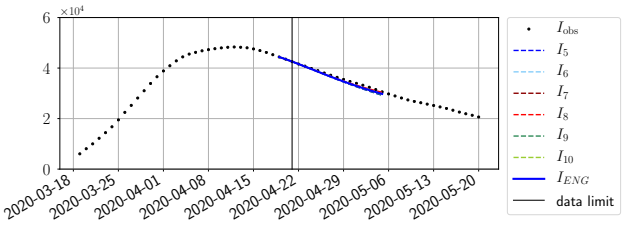
(**c**) Infected. (**d**) Removed.

Figure 21. ENG forecast from $T = 15/04$.

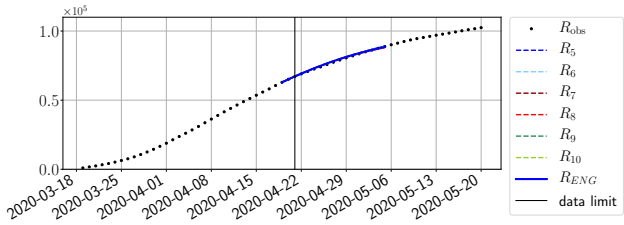


(**a**) β. (**b**) γ.

Figure 22. *Cont.*

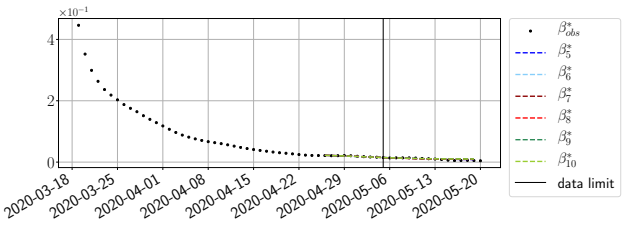


(**c**) Infected.

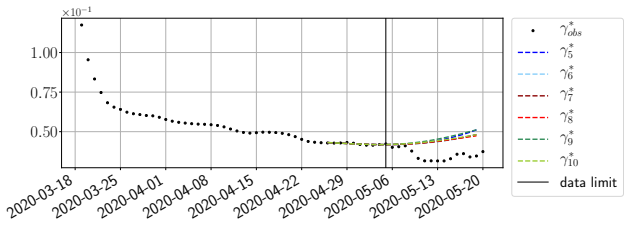


(**d**) Removed.

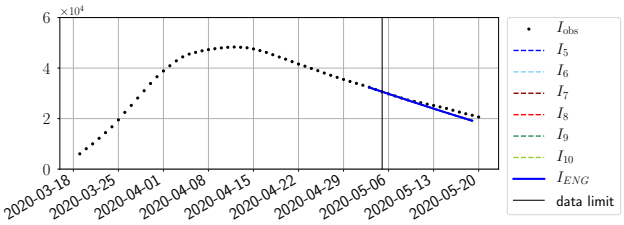
Figure 22. ENG forecast from $T = 21/04$.



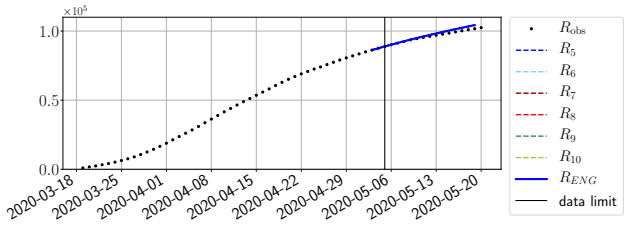
(**a**) β.



(**b**) γ.



(**c**) Infected.

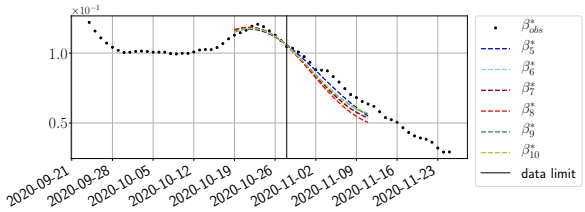


(**d**) Removed.

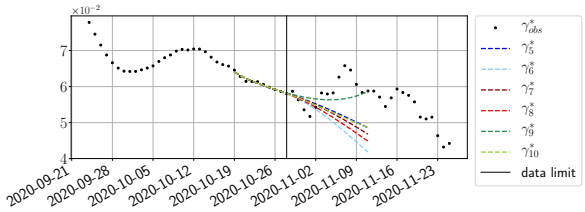
Figure 23. ENG forecast from $T = 05/05$.

4.2.3. Forecasting of the Second Wave with ENG

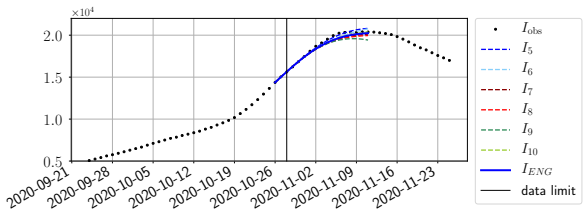
The review took place during the month of November 2020 as the second COVID-19 pandemic wave hit France. We took advantage of this to enlarge the body of numerical results, and we provide some example forecasts with ENG for this wave in Figures 24–26. As the figures illustrate, the method provides very satisfactory forecasts in a 14 day-ahead window. We again observe a satisfactory prediction of the second peak (Figures 24–26) and the same difficulty in forecasting γ due to the oscillations in γ_{obs}, but this has not greatly impacted the quality of the forecasts for I and R.



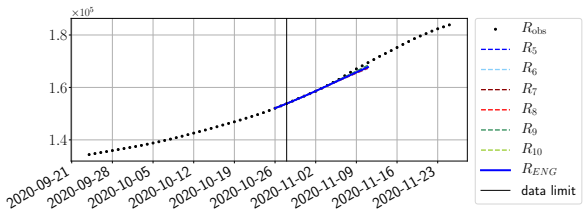
(**a**) β.



(**b**) γ.

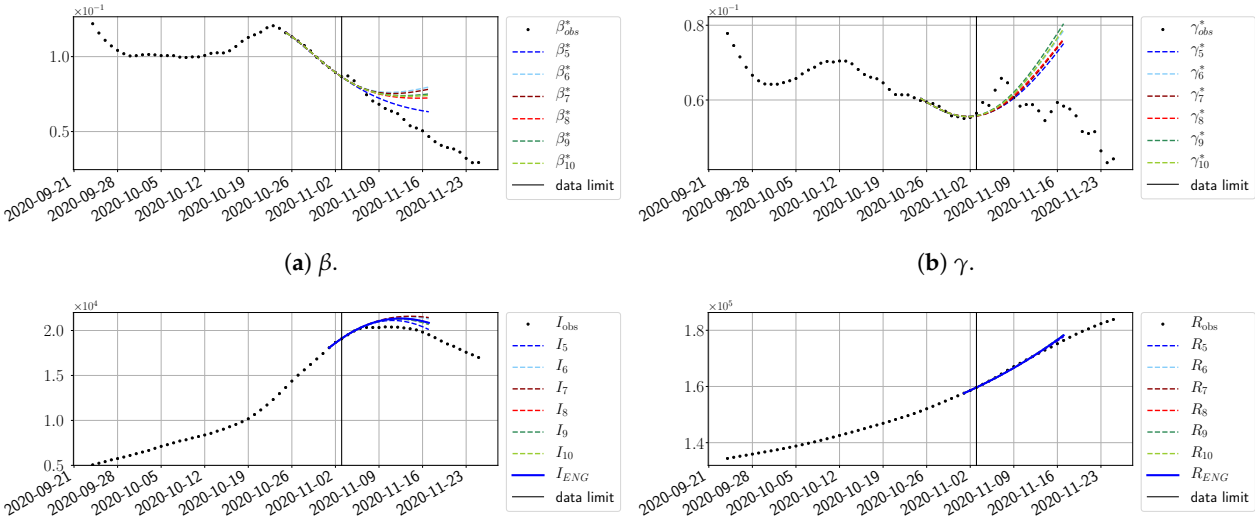


(**c**) Infected.



(**d**) Removed.

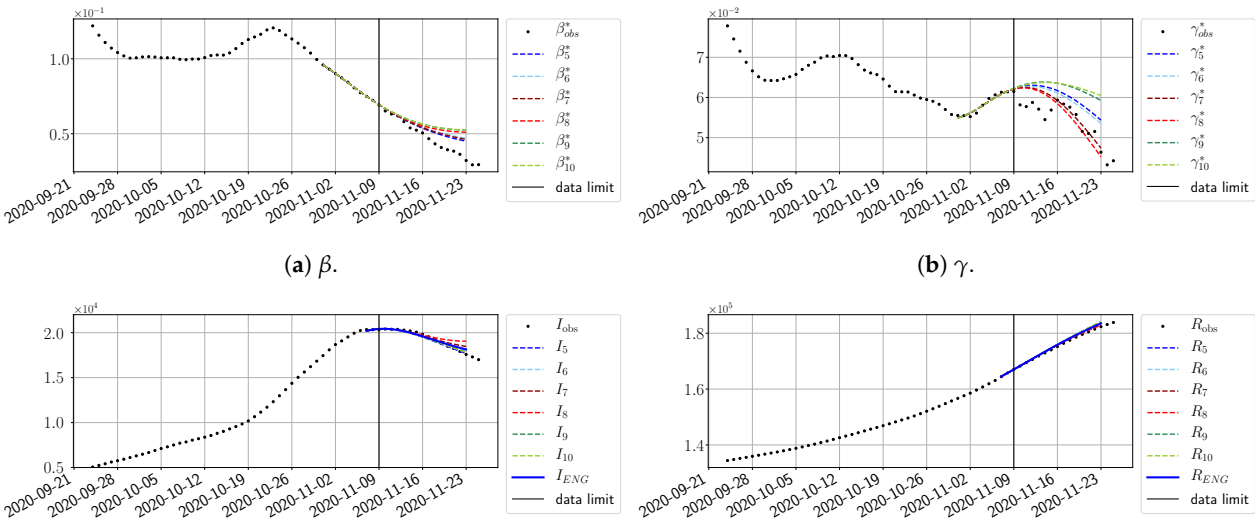
Figure 24. ENG forecast from $T = 28/10$.



(**a**) β. (**b**) γ.

(**c**) Infected. (**d**) Removed.

Figure 25. ENG forecast from $T = 03/11$.



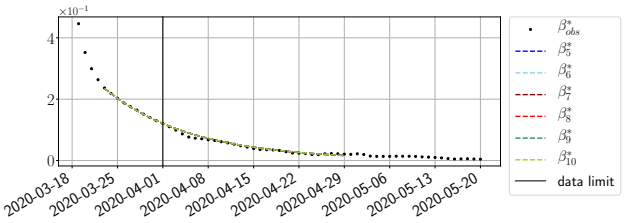
(**a**) β. (**b**) γ.

(**c**) Infected. (**d**) Removed.

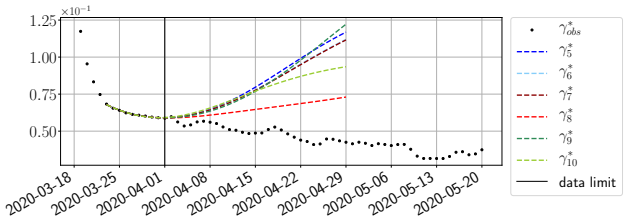
Figure 26. ENG forecast from $T = 09/11$.

4.2.4. Forecasts with ENG with a 28 Day-Ahead Window

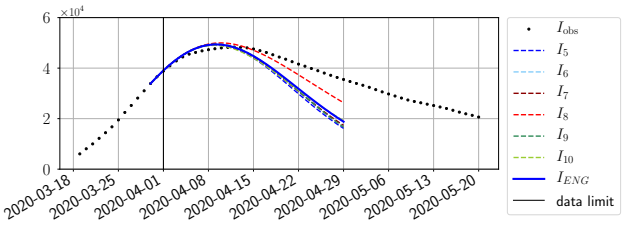
To conclude this section, we extend the forecasting window to 28 days instead of 14 and study whether the introduced ENG method still provides satisfactory forecasts. As shown in Figures 27–32, the results of the methods are quite stable for large windows. This shows that, in contrast to standard extrapolation methods using classical linear or affine regressions, the reduced basis catches the dynamics of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ not only locally but also at extended time intervals.



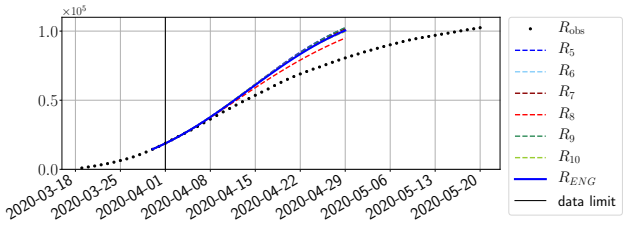
(**a**) β.



(**b**) γ.

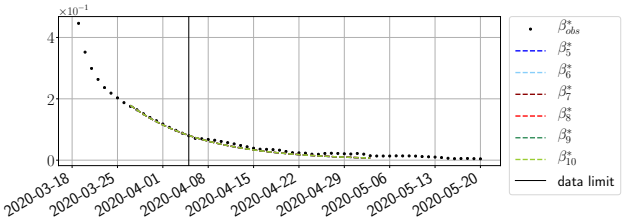


(**c**) Infected.

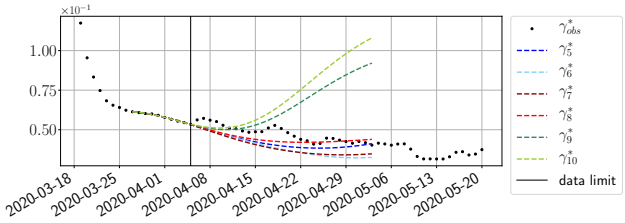


(**d**) Removed.

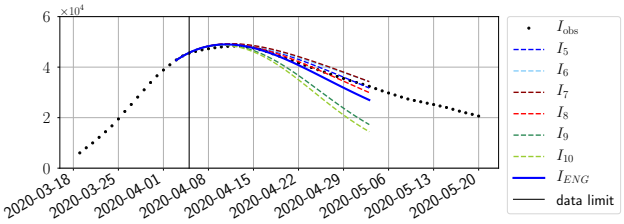
Figure 27. ENG forecast from $T = 01/04$.



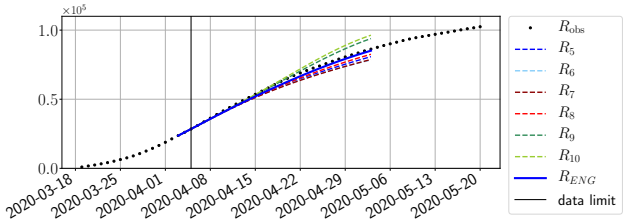
(**a**) β.



(**b**) γ.

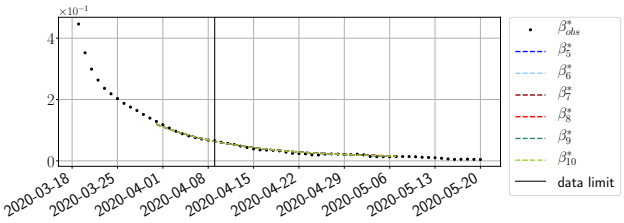


(**c**) Infected.

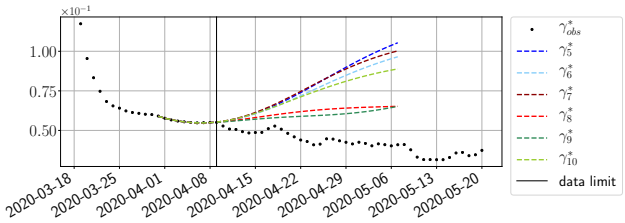


(**d**) Removed.

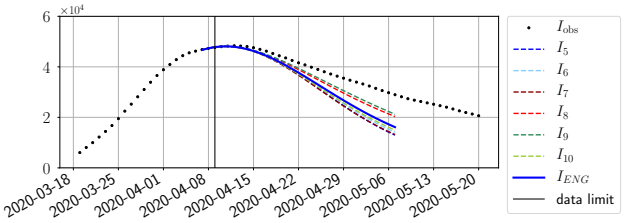
Figure 28. ENG forecast from $T = 05/04$.



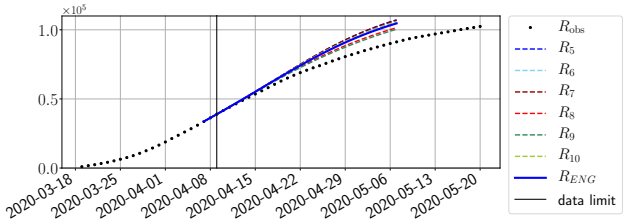
(**a**) β.



(**b**) γ.

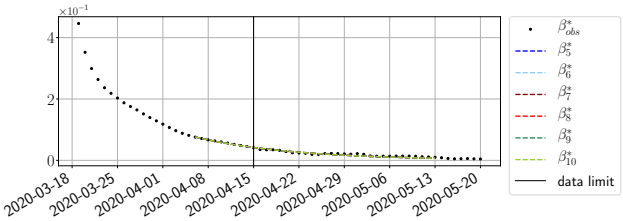


(**c**) Infected.

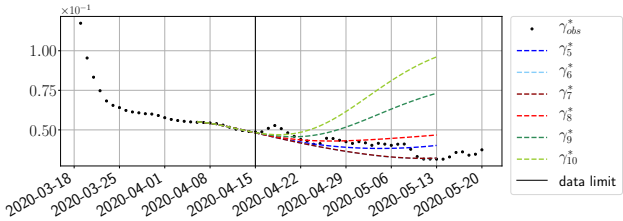


(**d**) Removed.

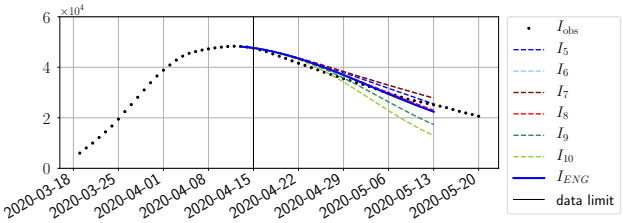
Figure 29. ENG forecast from $T = 05/04$.



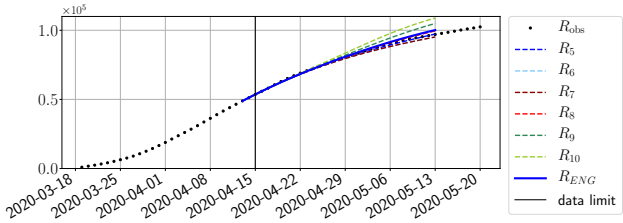
(**a**) β.



(**b**) γ.

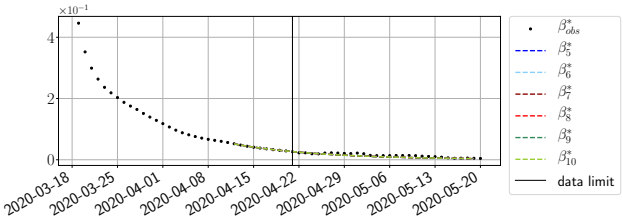


(**c**) Infected.

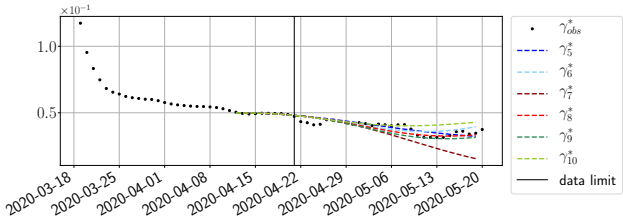


(**d**) Removed.

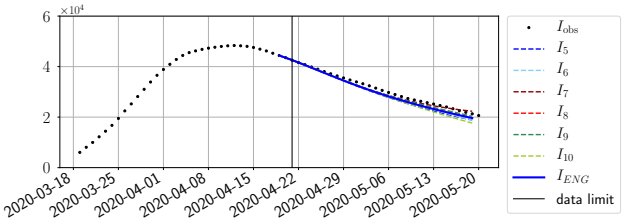
Figure 30. ENG forecast from $T = 15/04$.



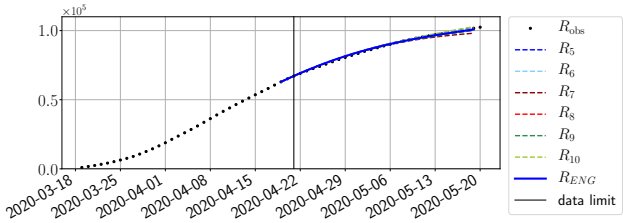
(**a**) β.



(**b**) γ.

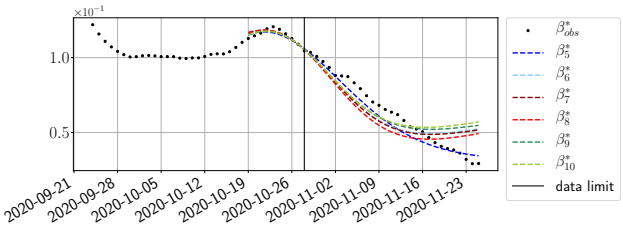


(**c**) Infected.

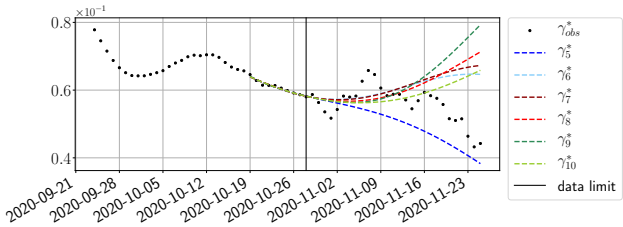


(**d**) Removed.

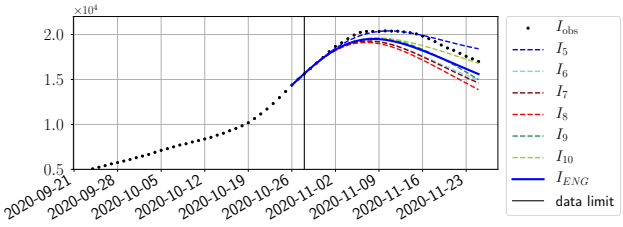
Figure 31. ENG forecast from $T = 21/04$.



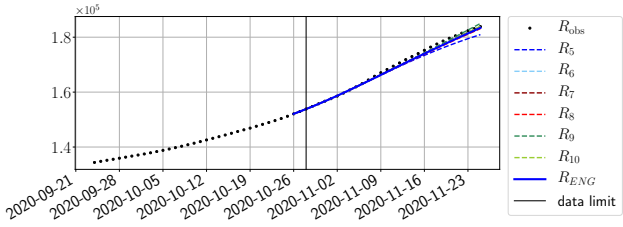
(**a**) β.



(**b**) γ.



(**c**) Infected.



(**d**) Removed.

Figure 32. ENG forecast from $T = 28/10$.

5. Conclusions

We have developed an epidemiological forecasting method based on reduced modeling techniques. Of the different strategies that have been explored, the one that outperforms the rest in terms of robustness and forecasting power involves reduced models that are built with an Enlarged Nonnegative Greedy (ENG) strategy. This method is novel and of interest in itself as it allows reduced models that preserve positivity and even other types of bounds to be built. Despite the fact that ENG does not have optimal fitting properties (i.e., interpolation properties), it is well suited for forecasting since, due to the preservation of the natural constraints of the coefficients, it generates realistic dynamics with few modes. The results have been presented for a mono-regional test case, and we plan to extend the present methodology to a multi-regional setting using mobility data.

Last but not least, we would like to emphasize that the developed strategy is very general and could be applied to the forecasting of other types of dynamics. The specificity of each problem may, however, require adjustments in the reduced models. This is exemplified in the present problem through the construction of reduced models that preserve positivity.

Author Contributions: All authors contributed equally. All authors have read and agreed to the published version of the manuscript.

Funding: Part of this research has been supported by the Emergences project grant "Models and Measures" of the Paris city council and by the generous donation made available by Alkan together with complementary funding given by the Sorbonne University Foundation.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data available on request.

Acknowledgments: This research has been undertaken in the context of the project Pandemia/Covidia, the GIS Obepine, and the "Facing the Virus" initiative within PSL University. They all aim at a better understanding of the COVID-19 pandemic, and we kindly thank their members for fruitful discussions. We also thank Gabriel Turinici for his feedback on the multi-regional models.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Analysis of Model Error and Observation Noise

In this section, we study the impact of observation noise and model error on the quality and behavior of the fitting step. The elements that impact the accuracy of our procedure are the following:

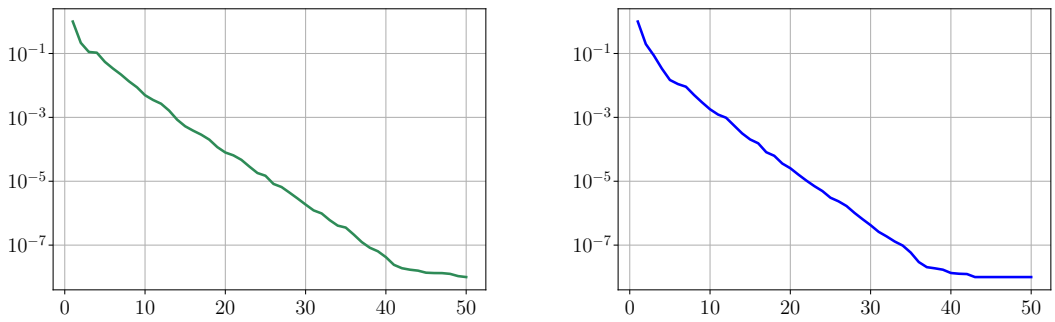
1. Observation noise: The observed health data $\mathbf{I}_{\text{obs}}$ and $\mathbf{R}_{\text{obs}}$ present several sources of noise. There may be some inaccuracies in the reporting of cases, and the data are then post-processed. This eventually yields the noisy time series $\mathbf{I}_{\text{obs}}$ and $\mathbf{R}_{\text{obs}}$ for which it is difficult to estimate the uncertainty. These noisy data are then used to produce (through finite differences) β^*_{obs} and γ^*_{obs}, which are in turn also noisy.
2. Model errors: Two types of model errors are identified:
   (a) Intrinsic model error on $\mathcal{B}$ and $\mathcal{G}$: The families of detailed models that we use are rich, but they may not cover all possible evolutions of $\mathbf{I}_{\text{obs}}$ and $\mathbf{R}_{\text{obs}}$. In other words, our manifolds $\mathcal{B}$ and $\mathcal{G}$ may not perfectly cover the real epidemiological evolution. Thsi error motivated the introduction of the exponential functions b_0 and g_0 described in Section 2.4.
   (b) Sampling error of $\mathcal{B}$ and $\mathcal{G}$: The size of the training sets $\mathcal{B}_{\text{tr}}$ and $\mathcal{G}_{\text{tr}}$ and the dimension n of the reduced models B_n and G_n also limit the approximation capabilities of the continuous target sets $\mathcal{B}$ and $\mathcal{G}$.

In this section, we aim to disentangle these effects in order to give a better insight of the fitting error behavior for the real data.

Appendix A.1. Study of the Impact of the Sampling Error: Fitting a Virtual Scenario

Our set of virtual epidemiological dynamics is $\mathcal{U}$. After the collapsing step, the manifolds for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are $\mathcal{B}$ and $\mathcal{G}$. These sets are potentially infinite, and we have used finite training subsets $\mathcal{B}_{\text{tr}} \subseteq \mathcal{B}$ and $\mathcal{G}_{\text{tr}} \subseteq \mathcal{G}$ to build the reduced models B_n and G_n.

First, we consider the eigenvalues for β and γ when performing an SVD decomposition for the virtual scenarios. Figure A1 shows a rapid decay of the eigenvalues obtained by SVD decomposition and shows that we can obtain a very good approximation of elements of $\mathcal{B}_{\text{tr}} \subseteq \mathcal{B}$ and $\mathcal{G}_{\text{tr}} \subseteq \mathcal{G}$ with only a few modes.

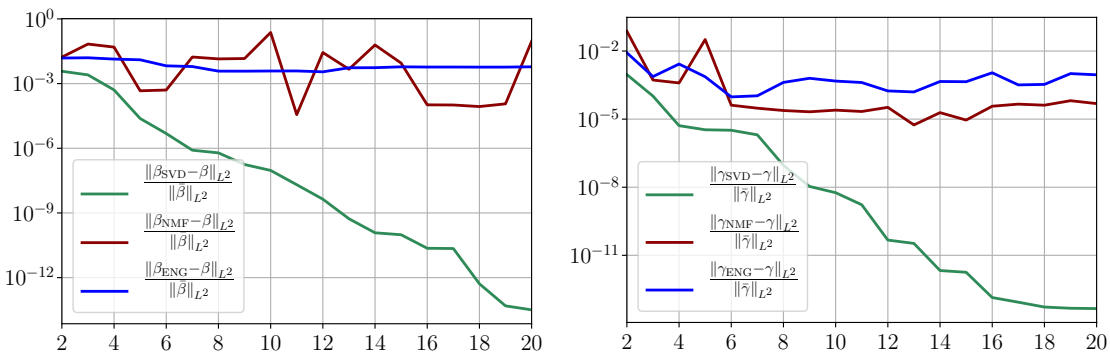


(**a**) Normalized eigenvalues for β vs. n. (**b**) Normalized eigenvalues for γ vs. n.

Figure A1. Training step: Decay of SVD eigenvalues.

Among the sources of noise and model error given at the beginning of Appendix A, we can study the impact of the sampling error (point 2-b) as follows. For this, we start by examining the fitting error on an interval of 45 days for two functions $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ which belong to the manifold $\mathcal{B}$ and $\mathcal{G}$ and are in the training sets $\mathcal{B}_{\text{tr}}$ and $\mathcal{G}_{\text{tr}}$. This error will serve as a benchmark, which we use to compare the fitting errors of functions that are not in the training sets.

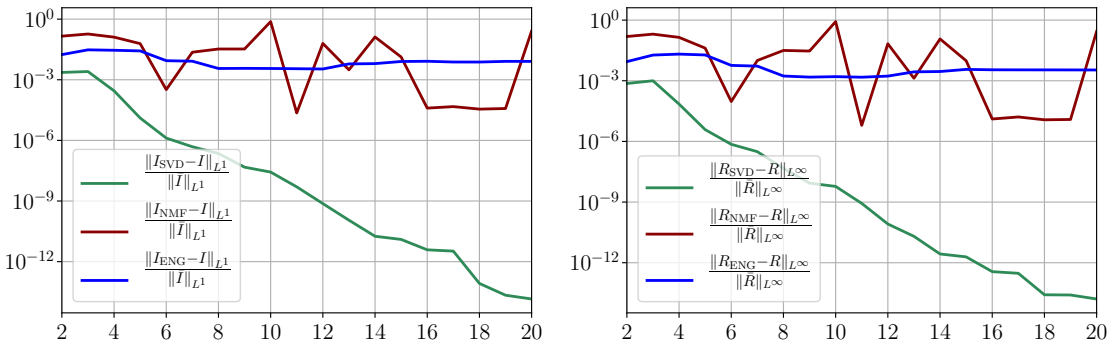
Figure A2 shows relative fitting errors $\|\beta_n^* - \beta_{\text{obs}}^*\|_{L^2([t_0,T])} / \|\bar{\beta}_{\text{obs}}^*\|_{L^2([t_0,T])}$ and $\|\gamma_n^* - \gamma_{\text{obs}}^*\|_{L^2([t_0,T])} / \|\bar{\gamma}_{\text{obs}}^*\|_{L^2([t_0,T])}$ using SVD, NMF and ENG-reduced bases with $n = 2, \ldots, 20$, where the notation $\bar{x}$ denotes the mean of the quantity x over $[0, T]$. We observe that, for SVD, the fitting errors behave similarly to the error decay of the training step (Figure A1).



(**a**) L^2 relative error of β vs. n. (**b**) L^2 relative error of γ vs. n.

Figure A2. Study of sampling errors: Fitting errors of functions $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ from $\mathcal{B}$ and $\mathcal{G}$ that belong to the training sets $\mathcal{B}_{\text{tr}}$ and $\mathcal{G}_{\text{tr}}$.

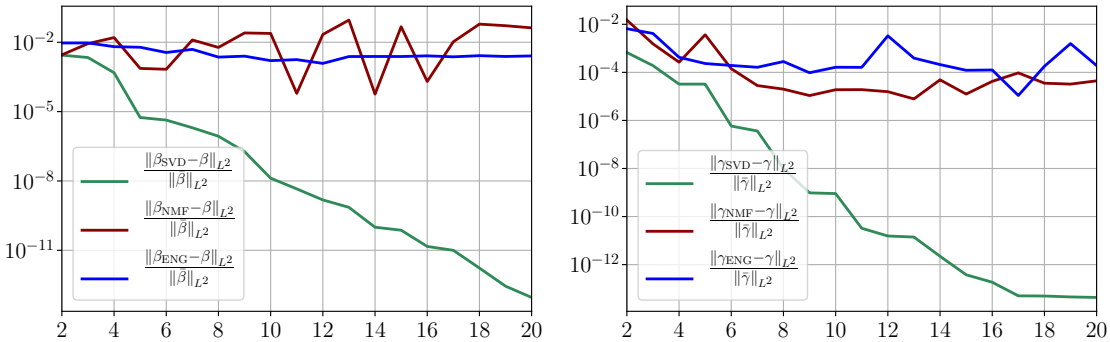
Figure A3 shows the L^1 and L^∞ errors obtained after the propagation of the fittings of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$. In both metrics, the error for **I** and **R** obtained using SVD decreases below 10^{-12}. When NMF and ENG are used, the error decreases and stagnates at 10^{-2} for both **I** and **R**.



(**a**) L^1 relative error of I vs. n. (**b**) L^∞ relative error of R vs. n.

Figure A3. Study of sampling errors: Errors on **I** and **R** obtained by the susceptible–infected–removed (SIR) model using the fitted $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$.

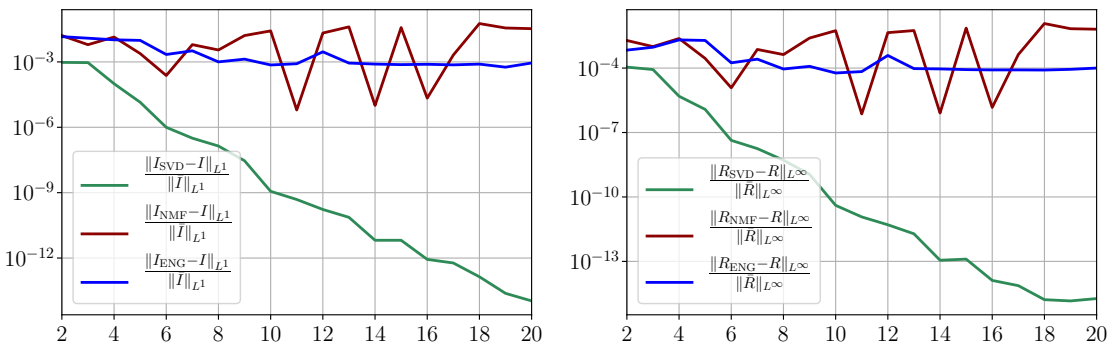
Now, we consider the fitting error for two functions $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ which belong to the manifold $\mathcal{B}$ and $\mathcal{G}$ but which are not in the training sets $\mathcal{B}_{\mathrm{tr}}$ and $\mathcal{G}_{\mathrm{tr}}$. Figure A4 shows the fitting error on the virtual scenario considered using SVD, NMF and ENG-reduced bases for $n = 2, \ldots, 20$. We note that the quality of the fittings in each method is very similar to that of Figure A3 where the functions were taken in the training sets. This illustrates that the sampling error does not play a major role in our experiments.



(**a**) L^2 relative error of β vs. n. (**b**) L^2 relative error of γ vs. n.

Figure A4. Study of sampling errors: Fitting errors of functions $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ from $\mathcal{B}$ and $\mathcal{G}$ but which do not belong to the training sets $\mathcal{B}_{\mathrm{tr}}$ and $\mathcal{G}_{\mathrm{tr}}$.

Figure A5 shows the L^1 and L^∞ errors obtained after the propagation of the fittings of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$. In both metrics, the error for **I** and **R** obtained using SVD decreases to 10^{-14}. When NMF and ENG are used, the error decreases and stagnates at 10^{-3} and 10^{-4} for both **I** and **R**, respectively.

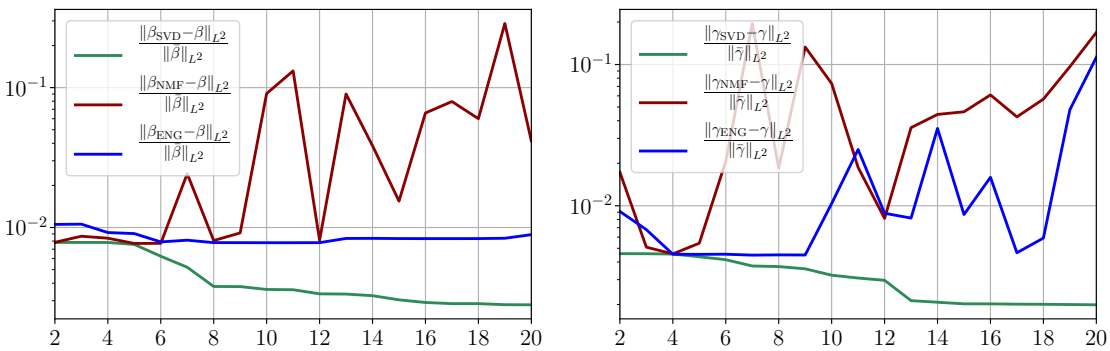


(**a**) L^1 relative error of I vs. n. (**b**) L^∞ relative error of R vs. n.

Figure A5. Study of sampling errors: Errors on **I** and **R** obtained by the SIR model using the fitted $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$.

Appendix A.2. Study of the Impact of Noisy Data and Intrinsic Model Error

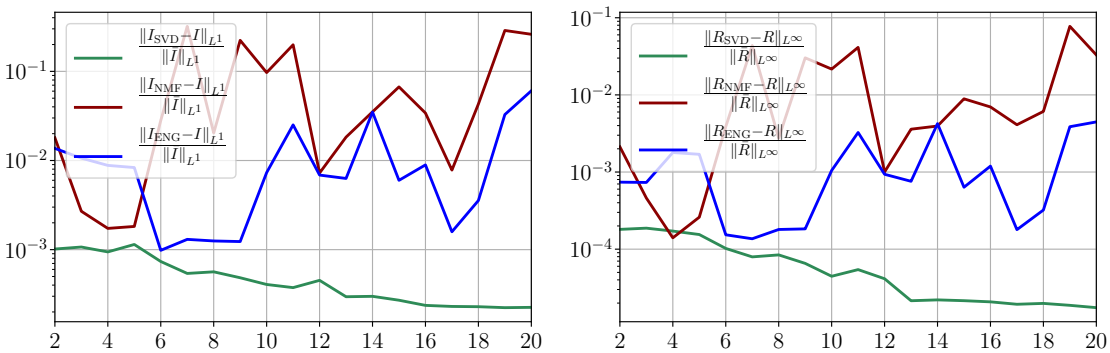
To investigate the impact of noise in the observed data, we now add noise to the two previously chosen functions $\beta \in \mathcal{B}$ and $\gamma \in \mathcal{G}$, and we study the fitting error for this noisy data. The level of the noise has been chosen to be the same level as that estimated for the real dynamics. In order to estimate the noise, we performed a fitting of the real data using SVD-reduced bases; the level of noise is defined as the difference between this fitting and the real data. This level of noise is then added to the virtual scenario considered here. Figure A6 shows the fitting error on β and γ using SVD, NMF and ENG-reduced bases for $n = 2, \ldots, 20$.



(**a**) L^2 relative error of β vs. n. (**b**) L^2 relative error of γ vs. n.

Figure A6. Fitting errors of β and γ in a noisy virtual scenario.

We observe that the noise strongly deteriorates the fitting error obtained using NMF, and the error becomes oscillatory and very unstable. For ENG, the error remains low and consistently around 10^{-2} for β. We observe the same behavior for γ with instabilities arising for $n > 10$. For SVD, the error is lower than in the ENG case and slowly decreases as n increases. Figure A7 shows the L^1 and L^∞ errors obtained after the propagation of the fittings of β and γ. In line with the observations from Figure A6, the error obtained using NMF is very unstable. Using ENG, we observe a decay from $n = 2$ to $n = 7$ and oscillations that remain around 10^{-2} for I and 10^{-3} for R. The decay observed for SVD is slow and steady; the error nearly reaches 10^{-4} for I and 10^{-5} for R when $n = 20$.

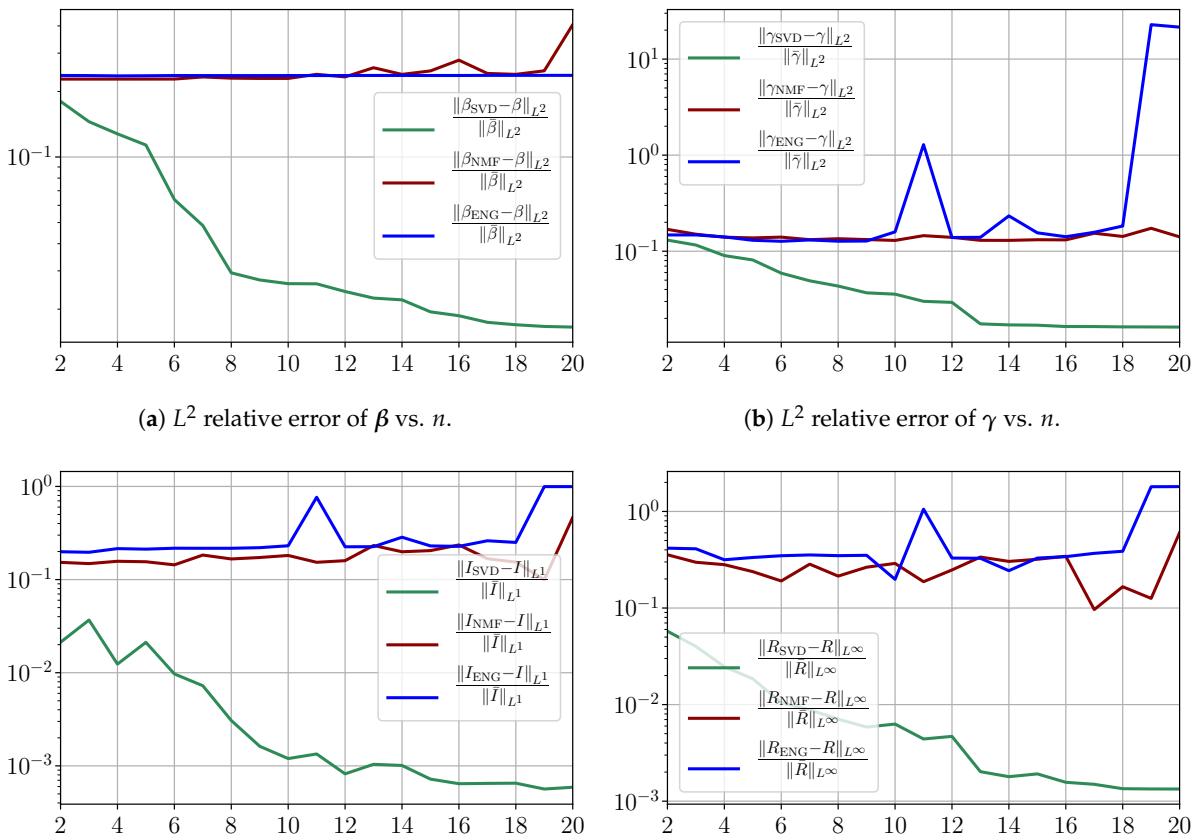


(**a**) L^1 relative error of I vs. n. (**b**) L^∞ relative error of R vs. n.

Figure A7. Fitting errors of I and R on noisy data.

Finally, it is necessary to add the intrinsic model error on top of the previous sampling error and observation noise. In so doing, the main change is that the previous fitting error plots from Figure A6 have essentially the same behavior but the error values are increased depending on the degree of model inaccuracy. We have therefore disentangled all the effects of model error and noisy data, and all the observations from this section thus give a better insight regarding the fitting on the real data.

Figure A8 summarizes the fitting results for an example fitting period taken from $t_0 = 19/03$ to $T = 03/05$. Figure A8a,b shows the fitting error on $\boldsymbol{\beta}$ and γ using SVD, NMF and ENG-reduced bases for $n = 2, \ldots, 20$. Figure A8c,d shows the L^1 and L^∞ relative errors on $\mathbf{I}_n$ and $\mathbf{R}_n$ after the propagation of the fittings of β_n^* and γ_n^*.



(**a**) L^2 relative error of $\boldsymbol{\beta}$ vs. n.

(**b**) L^2 relative error of γ vs. n.

(**c**) L^1 relative error of I vs. n.

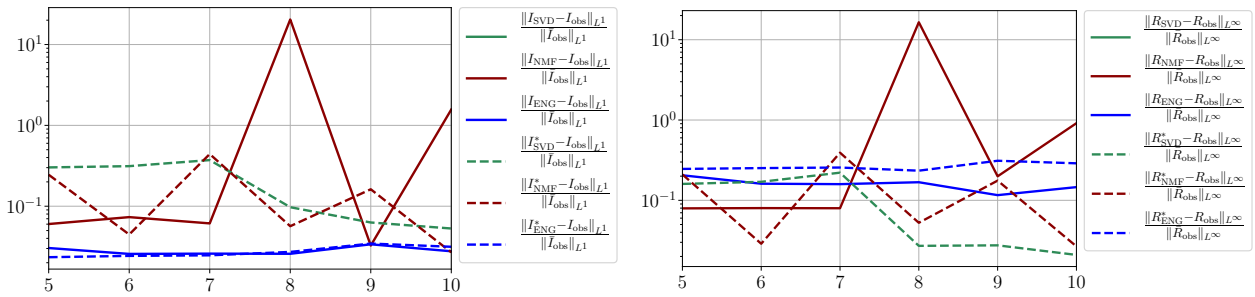
(**d**) L^∞ relative error of R vs. n.

Figure A8. Fitting errors from $t_0 = 19/03/2020$ to $T = 03/05/2020$.

From Figure A8a,b, we observe that the fitting error with SVD decreases at a moderate rate as the dimension n of the reduced basis is increased. The error with NMF and ENG does not decrease and oscillates around a certain constant error value of the order 10^{-1}. Note that this value is small and yields small errors in the approximation of $\mathbf{I}$ and $\mathbf{R}$, as Figure A8c,d illustrates.

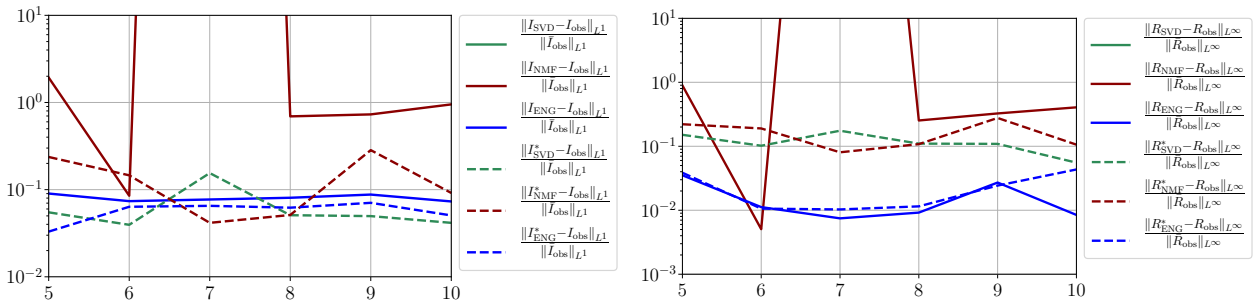
Appendix B. Study of Forecasting Errors

In this section, we present a detailed study of the forecasting errors for each different fitting strategy (`routine-IR`, `routine-`$\boldsymbol{\beta\gamma}$), reduced model (SVD, NMF, ENG) and starting date T. We anticipate the main conclusion announced in Section 4.2.1: ENG fitted with `routine-`$\boldsymbol{\beta\gamma}$ outperforms the other reduced models and is the most robust and accurate reduced model to use in a forecasting strategy. Figure A9 shows the relative errors of a 14 day forecast from $T = 01/04$ for each forecasting method and each reduced basis. Similarly Figures A10–A17 show the forecasts' relative errors from different times, respectively.



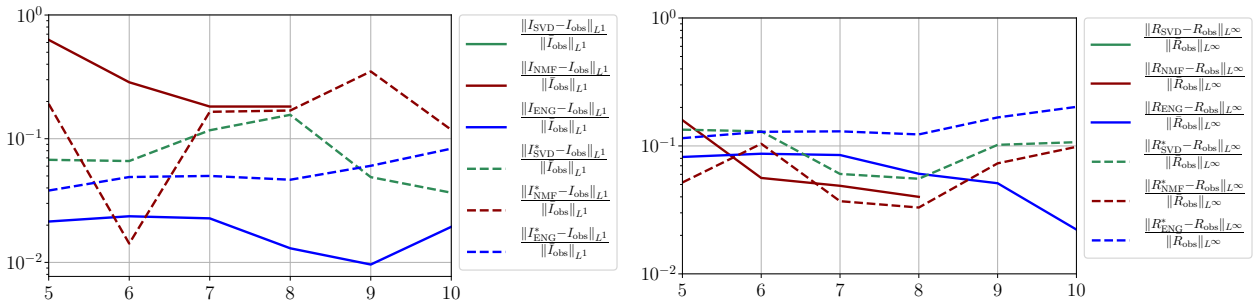
(**a**) L^1 relative error of I vs. n. (**b**) L^∞ relative error of R vs. n.

Figure A9. Forecasting errors of I and R (from $T = 01/04$).



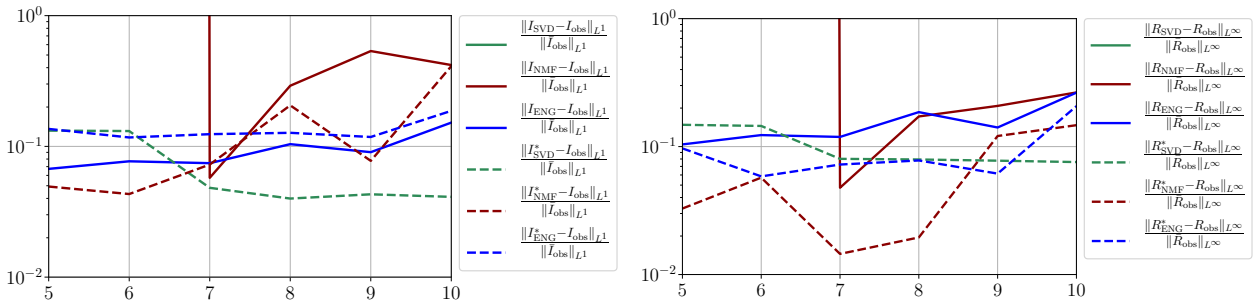
(**a**) L^1 relative error of I vs. n. (**b**) L^∞ relative error of R vs. n.

Figure A10. Forecasting errors of I and R (from $T = 03/04$).



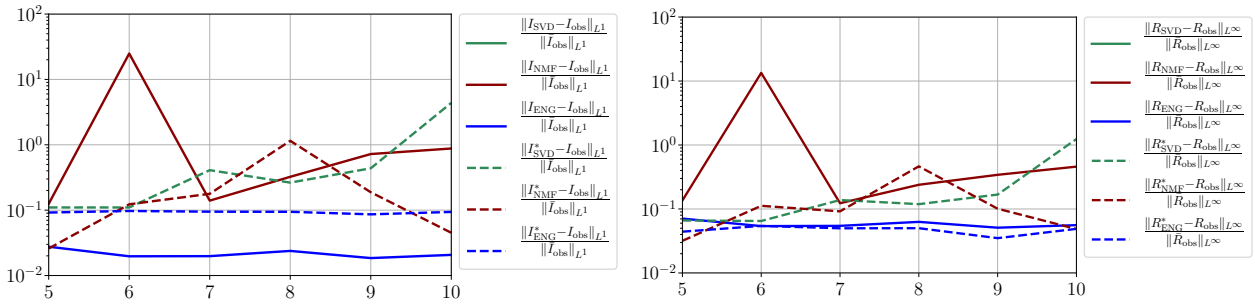
(**a**) L^1 relative error of I vs. n. (**b**) L^∞ relative error of R vs. n.

Figure A11. Forecasting errors of I and R (from $T = 05/04$).



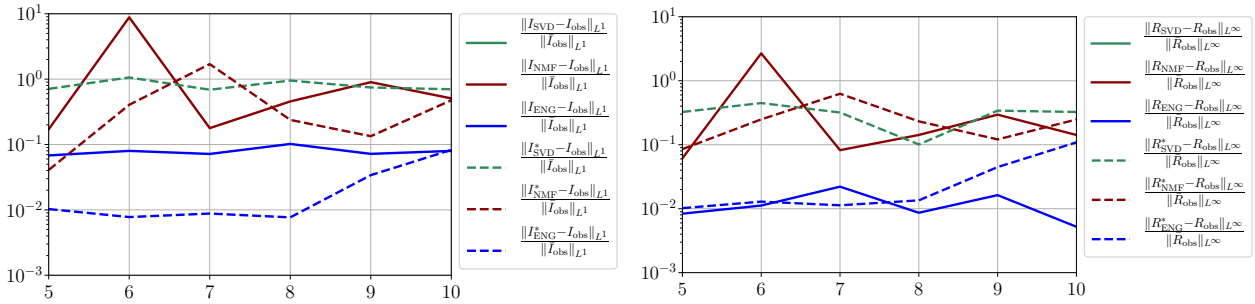
(**a**) L^1 relative error of I vs. n. (**b**) L^∞ relative error of R vs. n.

Figure A12. Forecasting errors of I and R (from $T = 07/04$).



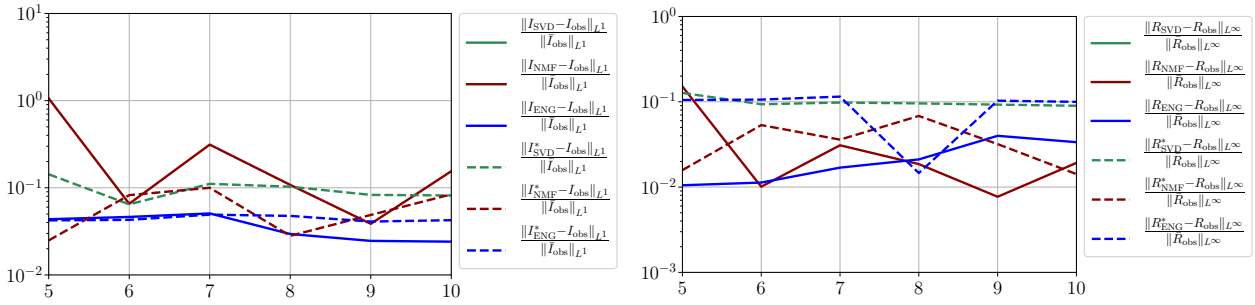
(**a**) L^1 relative error of I vs. n. (**b**) L^∞ relative error of R vs. n.

Figure A13. Forecasting errors of I and R (from $T = 09/04$).



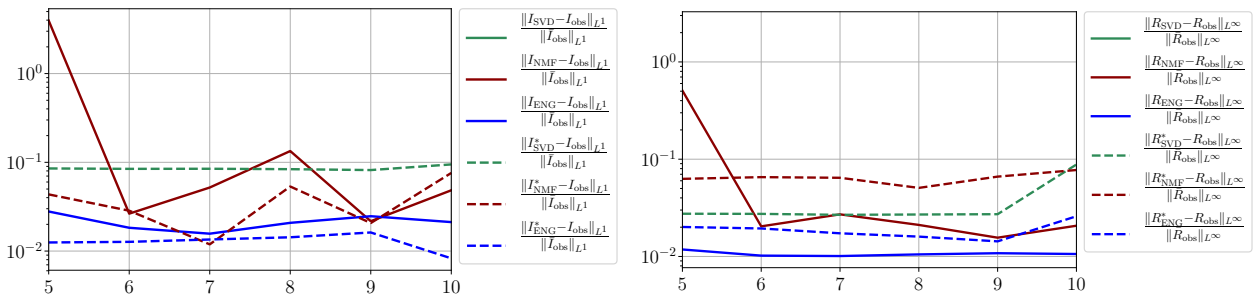
(**a**) L^1 relative error of I vs. n. (**b**) L^∞ relative error of R vs. n.

Figure A14. Forecasting errors of I and R (from $T = 11/04$).



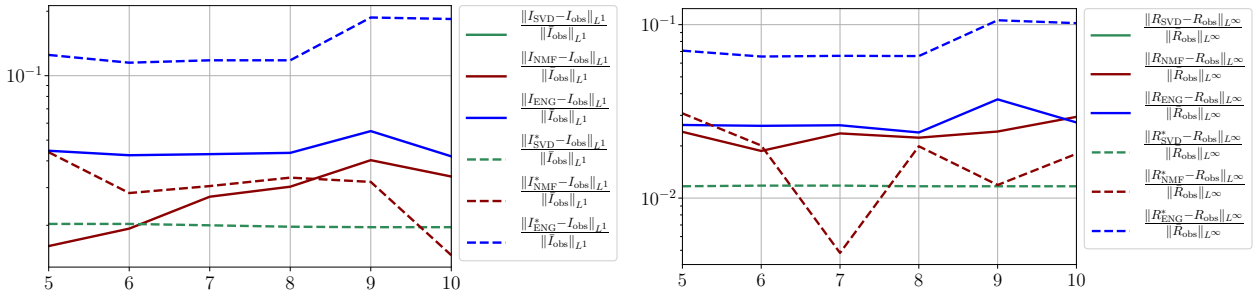
(**a**) L^1 relative error of I vs. n. (**b**) L^∞ relative error of R vs. n.

Figure A15. Forecasting errors of I and R (from $T = 15/04$).



(**a**) L^1 relative error of I vs. n. (**b**) L^∞ relative error of R vs. n.

Figure A16. Forecasting errors of I and R (from $T = 21/04$).



(**a**) L^1 relative error of I vs. n. (**b**) L^∞ relative error of R vs. n.

Figure A17. Forecasting errors of I and R (from $T = 05/05$).

We observe that the quality of the forecast depends on the reduced basis but also strongly on the starting day T from which the forecast is produced. The forecasts using `routine-`$\boldsymbol{\beta\gamma}$ with SVD and NMF are not accurate and most often blow up. When `routine-IR` is used with SVD and NMF, the forecasts are more robust as there is no observed explosion of error. Reduced bases from ENG consistently lead to the the best forecast being obtained using either `routine-`$\boldsymbol{\beta\gamma}$ and `routine-IR`; by inspecting the error on Figures A9–A17 and the averaged forecasts obtained in Section 4.2.1, we conclude that `routine-`$\boldsymbol{\beta\gamma}$ with ENG-reduced bases provides slightly better forecasts.

References

1. Ceylan, Z. Estimation of COVID-19 prevalence in Italy, Spain, and France. *Sci. Total. Environ.* **2020**, *729*, 138817. [CrossRef]
2. Anastassopoulou, C.; Russo, L.; Tsakris, A.; Siettos, C. Data-based analysis, modelling and forecasting of the COVID-19 outbreak. *PLoS ONE* **2020**, *15*, e0230405. [CrossRef]
3. Fang, X.; Liu, W.; Ai, J.; He, M.; Wu, Y.; Shi, Y.; Shen, W.; Bao, C. Forecasting incidence of infectious diarrhea using random forest in Jiangsu Province, China. *BMC Infect. Dis.* **2020**, *20*, 1–8. [CrossRef] [PubMed]
4. Roda, W.C.; Varughese, M.B.; Han, D.; Li, M.Y. Why is it difficult to accurately predict the COVID-19 epidemic? *Infect. Dis. Model.* **2020**, *5*, 271–281. [CrossRef] [PubMed]
5. Liu, Z.; Magal, P.; Seydi, O.; Webb, G. Predicting the cumulative number of cases for the COVID-19 epidemic in China from early data. *arXiv* **2020**, arXiv:2002.12298.
6. Keeling, M.J.; Rohani, P. *Modeling Infectious Diseases in Humans and Animals*; Princeton University Press: Princeton, NJ, USA, 2011.
7. Martcheva, M. *An Introduction to Mathematical Epidemiology*; Springer: Berlin/Heidelberg, Germany, 2015; Volume 61.
8. Brauer, F. Mathematical epidemiology: Past, present, and future. *Infect. Dis. Model.* **2017**, *2*, 113–127. [CrossRef] [PubMed]
9. Magal, P.; Webb, G. Predicting the number of reported and unreported cases for the COVID-19 epidemic in South Korea, Italy, France and Germany. *Medrxiv* **2020**, *509*, 110501. [CrossRef]
10. Di Domenico, L.; Pullano, G.; Sabbatini, C.E.; Boëlle, P.Y.; Colizza, V. Impact of lockdown in Île-de-France and possible exit strategies. *BMC Med.* **2020**, *18*, 240. [CrossRef]
11. Flaxman, S.; Mishra, S.; Gandy, A.; Unwin, H.J.T.; Mellan, T.A.; Coupland, H.; Whittaker, C.; Zhu, H.; Berah, T.; Eaton, J.W.; et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature* **2020**, *584*, 257–261. [CrossRef]
12. Kermack, W.O.; McKendrick, A.G.; Walker, G.T. A contribution to the mathematical theory of epidemics. *Proc. R. Soc. Lond. Ser. Contain. Pap. Math. Phys. Character* **1927**, *115*, 700–721. [CrossRef]
13. Armocida, B.; Formenti, B.; Ussai, S.; Palestra, F.; Missoni, E. The Italian health system and the COVID-19 challenge. *Lancet Public Health* **2020**, *5*, e253. [CrossRef]
14. Roques, L.; Klein, E.K.; Papaïx, J.; Sar, A.; Soubeyrand, S. Impact of Lockdown on the Epidemic Dynamics of COVID-19 in France. *Front. Med.* **2020**, *7*, 274. [CrossRef] [PubMed]
15. Ferguson, N.; Laydon, D.; Nedjati Gilani, G.; Imai, N.; Ainslie, K.; Baguelin, M.; Bhatia, S.; Boonyasiri, A.; Cucunuba Perez, Z.; Cuomo-Dannenburg, G.; et al. Report 9: Impact of Non-Pharmaceutical Interventions (NPIs) to Reduce COVID19 Mortality and Healthcare Demand. *Imp. Coll. Lond.* **2020**, *10*, 77482. [CrossRef]
16. Liu, Z.; Magal, P.; Seydi, O.; Webb, G. A COVID-19 epidemic model with latency period. *Infect. Dis. Model.* **2020**, *5*, 323–337.
17. Poncela, P.; Rodríguez, J.; Sánchez-Mangas, R.; Senra, E. Forecast combination through dimension reduction techniques. *Int. J. Forecast.* **2011**, *27*, 224–237. [CrossRef]
18. Binev, P.; Cohen, A.; Dahmen, W.; DeVore, R.; Petrova, G.; Wojtaszczyk, P. Convergence Rates for Greedy Algorithms in Reduced Basis Methods. *SIAM J. Math. Anal.* **2011**, *43*, 1457–1472. [CrossRef]
19. DeVore, R.; Petrova, G.; Wojtaszczyk, P. Greedy algorithms for reduced bases in Banach spaces. *Constr. Approx.* **2013**, *37*, 455–466. [CrossRef]
20. Cohen, A.; DeVore, R. Kolmogorov widths under holomorphic mappings. *IMA J. Numer. Anal.* **2016**, *36*, 1–12. [CrossRef]

21. Maday, Y.; Mula, O.; Turinici, G. Convergence analysis of the Generalized Empirical Interpolation Method. *SIAM J. Numer. Anal.* **2016**, *54*, 1713–1731. [CrossRef]
22. Quarteroni, A.; Manzoni, A.; Negri, F. *Reduced Basis Methods for Partial Differential Equations: An Introduction*; Springer: Berlin/Heidelberg, Germany, 2015; Volume 92.
23. Benner, P.; Cohen, A.; Ohlberger, M.; Willcox, K. *Model Reduction and Approximation: Theory and Algorithms*; SIAM: Philadelphia, PA, USA, 2017; Volume 15.
24. Hesthaven, J.S.; Rozza, G.; Stamm, B. *Certified Reduced Basis Methods for Parametrized Partial Differential Equations*; Springer: Berlin/Heidelberg, Germany, 2016; Volume 590.
25. Maday, Y.; Patera, A. Reduced basis methods. In *Model Order Reduction*; Benner, P., Grivet-Talocia, S., Quarteroni, A., Rozza, G., Schilders, W., Silveira, L., Eds.; De Gruyter: Oxford, UK, 16 December 2020; Chapter 4, pp. 139–179.
26. Paatero, P.; Tapper, U. Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics* **1994**, *5*, 111–126. [CrossRef]
27. Gillis, N. The why and how of nonnegative matrix factorization. *Regul. Optim. Kernels, Support Vector Mach.* **2014**, *12*, 257–291.
28. Atif, J.; Cappé, O.; Kazakçi, A.; Léo, Y.; Massoulié, L.; Mula, O. *Initiative Face au Virus Observations sur la Mobilité Pendant L'épidémie de COVID-19*; Technical Report; PSL University: Paris, France, 2020.
29. Mizumoto, K.; Kagaya, K.; Zarebski, A.; Chowell, G. Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020. *Eurosurveillance* **2020**, *25*, 2000180. [CrossRef]