A data first approach to modelling Covid-19

Jayanti Prasad*

Khagol-20, 38/1, Panchavti, Pashan, Pune (India) - 411008

May 22, 2020

Abstract

The primary data for Covid-19 pandemic is in the form of time series for the number of confirmed, recovered and dead cases. This data is updated every day and is available for most countries from multiple sources such as [Gar20b, iD20]. In this work we present a two step procedure for model fitting to Covid-19 data. In the first step, time dependent transmission coefficients are constructed directly from the data and, in the second step, measures of those (minimum, maximum, mean, median etc.,) are used to set priors for fitting models to data. We call this approach a "data driven approach" or "data first approach". This scheme is complementary to Bayesian approach and can be used with or without that for parameter estimation. We use the procedure to fit a set of SIR and SIRD models, with time dependent contact rate, to Covid-19 data for a set of 45 most affected countries. We find that SIR and SIRD models with constant transmission coefficients cannot fit Covid-19 data for most countries (mainly because social distancing, lockdown etc., make those time dependent). We find that any time dependent contact rate, which falls gradually with time, can help to fit SIR and SIRD models for most of the countries. We also present constraints on transmission coefficients and basic reproduction number $\mathfrak{R}_0$ as well as effective reproduction number $\mathfrak{R}(t)$. The main contributions of our work are as follows. (1) presenting a two step procedure for model fitting to Covid-19 data (2) constraining transmission coefficients as well as $\mathfrak{R}_0$ and $\mathfrak{R}(t)$, for a set of most affected countries and (3) releasing a python package **PyCov19** [Pra20] that can used to fit a set of compartmental models with time varying coefficients to Covid-19 data.

1 Introduction

At present the world is going through an unprecedented crisis of pandemic Covid-19 caused by a novel form of coronavirus, named Sars-CoV-2 which was passed to the human from bats in the Wuhan city of China, some time in December 2019 [Org20a, Org20b, ea20h, ea20t, ea20v, ea20d, ea20r, ea20l]. Till the middle of May 2020 the virus has reached almost all the parts of the world resulting in more than four million people infected and more than a quarter million deaths [Wor20]. The measures to contain the virus medically by developing a vaccine are going on war footing. However, the success is

*prasad.jayanti@gmail.com

still expected to be a few years away [ea20f]. Till a fraction of the population develop (herd) immunity or the vaccine is ready, the only means to contain the pandemic are social measures (social distancing, contact tracing etc.,) and enhanced hygiene practices [ea06, ea20s, ea20p].

Some of the most important problems related to Covid-19 research are (1) estimating the controlling parameters of the pandemic, (2) making short term predictions using mathematical-statistical modeling which can help in mitigating policies (3) simulating the growth of the epidemic by taking into account as many contributing effects as possible and (4) quantifying the impact of mitigation measures, such as lockdown etc [ea20j].

Modeling Covid-19 pandemic with compartmental models of Kermack and McKendrick (for an introduction see [KR08, Li18, BC18]) has been one of the most active problems in the recent times [ea20p, ea20a, ea20c, ea20e, FP20, ea20m, Oli20]. There have been alternative approaches also such as [RMI20] where statistical considerations are being taken into account for predictions. In one of the studies [FP20] it is argued that the data for the confirmed, recovered and dead, all three can easily fit a power law model with similar coefficients. The main attractive feature of these data driven approaches is that the complexity of the model being considered is determined by the data and not by theoretical expectations.

In the present work we follow a middle approach and fit two compartmental models, named SIR and SIRD with some modification, to the Covid-19 data. One of the main reasons to consider these models has been that the Covid-19 data is available only for the **S**usceptible, **I**nfected, **R**ecovered and **D**ead compartments (for the notations used here and other places in the present work see table (1)). It may be true that a large fraction of the population which may be **E**xposed (defined later) play an important role in the dynamics of the pandemic however, it is hard to get reliable numbers for that. Apart from that, a large number of undocumented cases [ea20l] may have significant influence on the spread of the pandemic.

A brief summary of the work presented here is as follows.

In §2 we give a brief introduction to the compartmental models and introduce the notations and variables used in the work. In particular, we discuss the SIR model in §2.1 and the SEIR and the SIRD models in §2.2 and §2.3 respectively. One of the major parts of the work presented here is to study the time dependence of the contact rate β, we introduce a set of parametric models of $\beta(t)$ in §2.4. We discuss the time series data used in the study in the §3 by giving an example of Italy which is one of the most affected countries. The main results of our work are given in §4 and in §5. In §4 we discuss the reconstruction (regression) procedure for the set of transmission coefficients as well as for the effective reproduction (defined later), number $\mathfrak{R}(t)$. Parameter estimation is discussed in §5. The main conclusions of our work with a summary and some important points are discussed in §6.

2 Mathematical Modeling

Mathematical modeling of epidemics that started a century back with the seminal work of Kermack and McKendrick (see [NC08, KR08, Li18] for an introduction) is still the main framework most commonly used. The main idea of the Kermack and McKendrick's compartmental models is that every individual in a society belongs to one of the M compartments and the total number of individuals belonging to different compartments keep changing with time. The minimum value M can have is two, for the Susceptible-Infected-Susceptible (SIS) model, in which the recovery does not guarantee that one will not get the infection again [KR08].

During an epidemic phase an individual can go through many stages from being perfectly healthy to the recovered one after an infection, with or without any immunity (short or long term) or may die. If we represent every stage with a compartment and keep the track of the number of individuals in each compartment then we can easily model the dynamics of the epidemic. This approach is very similar to the approach taken in astronomy where we count the number of stars in different stages of their life to understand the stellar evolution.

In principle we can have any number of logical compartments but in practice we should consider only those compartments for which we have the counts data, in particular for model fitting. Taking into account the fact that we have data only for the number of confirmed, recovered and dead population, the only compartmental model that meets the requirement is the SIRD model. If we consider the recovered and dead together we get the SIR model as is discussed in the next section.

One of the important compartments that also is commonly considered is the 'exposed' one and represents the population which have received the infection but cannot pass to others, before a certain period called the incubation period. If we consider exposed population also then we get the SEIR model that also is discussed below. Three compartmental models SIR, SIRD, and SEIR are shown in the (a), (b) and (c) panels of Figure (1) respectively (for more detail one can refer to [KR08, Het00, ea97, Oli20]).

If we identify the compartments with the nodes of a graph then the transmission between different compartments, as is represented by a set of coefficients, can be considered the edges of the graph. Some of the nodes may have multiple edges and some of the edges could be bi-directional also. The main challenge of the modeling a pandemic like Covid-19 is not the scarcity of mathematical models but it is of the reliable data for the compartments being considered.

2.1 SIR Model

The most basic compartmental model is the SIR model which is shown in (a) of the Figure(1) and is described by the following set of equations:

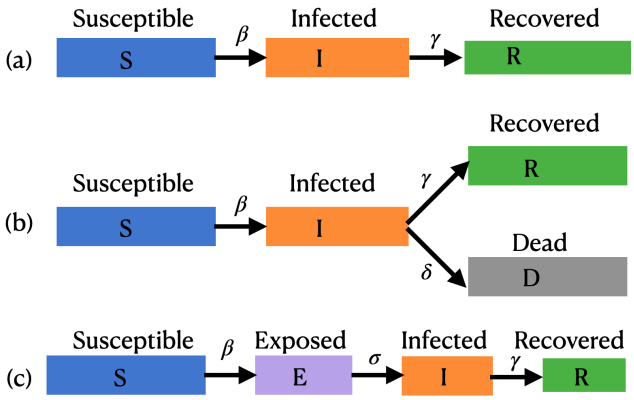

Figure 1: In the compartmental models the total population N is divided into a set of compartments as shown in the figure. The number of compartments and their connectivity depends on the detail of the model being considered. If we consider these compartments as nodes of a graph then there are transmission coefficients for every connecting edge that determine how effective that edge is in changing the population of the connected compartments. In (a) and (c) representing SIR and SEIR models, the compartments are connected in a linear way, however, for the case (b), representing the SIRD model, there is a branching also. Since the total population must remain a constant so the rates of change along all the connecting edges must add to zero.

S. No.	**Notation**	**Description**
1	**S**, **I**, **R**, **D**, **E**	Name of the compartments
2	S, I, R, D, E	Population in the compartments
3	$\mathcal{C}, \mathcal{R}, \mathcal{D}$	Cumulative numbers
4	$\mathfrak{R}_0$	Basic reproduction number
5	$\mathfrak{R}(t)$	Effective reproduction number

Table 1: Notations used in the work

$$\begin{aligned} \frac{dS}{dt} &= -\beta \frac{S}{N} I \\ \frac{dI}{dt} &= \beta \frac{S}{N} I - \gamma I \\ \frac{dR}{dt} &= \gamma I \end{aligned} \tag{1}$$

Here β and γ are the transmission coefficients, also called the contact rate and the recovery removal rate respectively, and $1/\beta$ and $1/\gamma$ represent the mean duration of infectiousness and the average period of infectivity (see [Het00, ea03a]), respectively.

In general there is some time lag between acquiring an infection and becoming infectious. However, in the SIR model it is ignored and an assumption is made that individuals become infectious immediately upon getting an infection. This is a very strong assumption and the main reasons for

making this is that we do not know reliably how many people are actually 'exposed', or have the virus but are still not infectious (cannot pass it to others). One of the ways to address this problem could be contact tracing and assuming that anyone who has come into contact with an infected person is an exposed one. However, this assumption is as strong as the assumption made in the SIR model.

If we do not consider the birth, death and movement of people then the following condition must be satisfied.

$$S + I + R = N \text{ (constant) } \text{ or } \frac{dS}{dt} + \frac{dI}{dt} + \frac{dR}{dt} = 0, \tag{2}$$

here S, I and R is the population of the **S**, **I** and **R** compartments respectively.

In equation (1) the transmission coefficient β is one of the most important parameters of the epidemic dynamics and can be written as the product of the contact rates (the average number of contacts per person per time) and the transmission probability (the probability of disease transmission on contact between a susceptible and an infectious person). As has been mentioned that the transmission coefficient γ can be identified with the recovery rate which is nothing but the inverse of the infectious period (during which an infected person can pass the virus to other healthy people).

In general, the equations (1) is solved with the following initial conditions:

$$S(t = 0) = S_0 > 0, \quad I(t = 0) = I_0 > 0, \quad \text{and} \quad R(t = 0) = R_0 = 0. \tag{3}$$

The second equation from 1 can be written as:

$$\frac{dI}{dt} = \beta \left(\frac{S}{N} - \frac{\gamma}{\beta} \right) I, \tag{4}$$

and for $S/N > \gamma/\beta$ we get a positive infection rate.

Here we define one of the most important parameters of an epidemic in terms of the ratio β/γ, called the **basic reproduction number** $\mathfrak{R}_0$, when considered a constant, and called the **effective reproduction number** $\mathfrak{R}(t)$, when considered a function of time. The most common definition [KR08] of it is that it is *the average number of secondary cases arising from an average primary case in an entirely susceptible population.* Note that in the text we may also use just "reproduction number" and the meaning of it will depend on the context. Some of the studies such as [ea20a] call $\mathfrak{R}_0$and $\mathfrak{R}(t)$both as basic reproduction number, however, we follow the convention used in [ea03b, ea03a, Cob20].

The basis reproduction number $\mathfrak{R}_0$ is the main measure which quantifies the transmissibility of the virus and $\mathfrak{R}_0 > 1$, sets a chain of transmissions leading an exponential growth of the pandemic. We can keep $\mathfrak{R}_0 < 1$, by minimizing the contact rates (social distancing etc.,), lowering the infectiousness of the infected people (by treating them or putting them in a quarantine etc.) and reducing the susceptibility of the healthy people by vaccination etc., (for detail see [ea05]).

The SIR model is one of the most basic models and can be easily generalized by one or more of the following ways:

1. **Adding more compartments:** Depending on the type of pandemic and other details we are interested in we can add more compartments to the SIR model. These compartments can fit in between the existing ones (for example as shown in Figure(1) (c) for SEIR case) or can branch out from the existing once (as shown in Figure(1) (b) for the SIRD case). With every new compartment added we must include the transmission coefficients for the connecting edges and also need an initial population for the new compartment being added [BC18, KB20, ea20b, ea20u].

2. **Heterogeneous population:** In the basic SIR model we consider a homogeneous population which share transmission coefficients such as β and γ. This may not be true in practice, for example, people from different groups (based on age, medical conditions, gender etc, etc) may have different contact rate β and/or recovery rate γ [ea20p, RR20].

3. **Variable transmission coefficients:** Theoretically transmission coefficients such as β and γ are considered constant, however, in practice they can vary with time due to multiple factors as the pandemic spreads [Cob20, P.20, GG20]. For example, social distancing and other precautions such as hand wash etc., may help to lower β. In a similar way a better understanding of the disease and the urgency with which medical and testing facilities are brought online may improve the recovery rate γ.

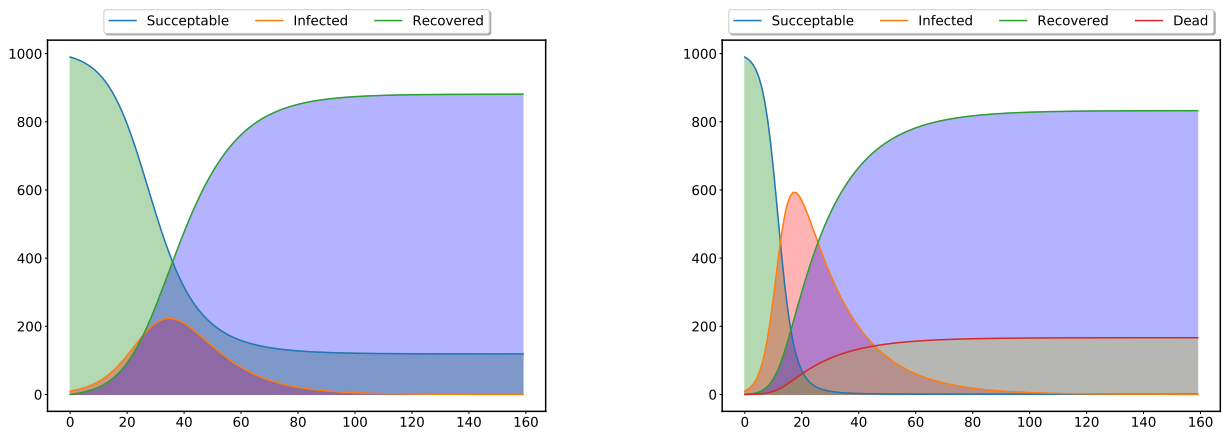

Figure 2: The left panel in the figure shows the standard SIR model with three type of population (compartments) - Susceptible, Infected and Recovered and the right panel shows the SIRD model with four type of population - on extra compartment for the dead population. The dynamics of the epidemic depends on the transmission coefficients β, γ and δ and the starting values of the different types of population. The values of the parameters and the initial population in different compartments used in the figure are only for the purpose on an illustration.

2.2 SEIR model

If we relax the assumption that the people who get the infection become infectious instantly and consider a latent period to the the onset of infectiousness there is a fraction of population (compartment) which has been exposed to the virus but will become infectious only after some latent period $1/\sigma$, then the model is called **S**usceptible-**E**xposed-**I**nfected-**R**ecovered (SEIR) model represented by the following set of equations [KR08]

$$\begin{aligned}
\frac{dS}{dt} &= -\beta\frac{S}{N}I \\
\frac{dE}{dt} &= \beta\frac{S}{N}I - \sigma E \\
\frac{dI}{dt} &= \sigma E - \gamma I \\
\frac{dR}{dt} &= \gamma I
\end{aligned} \tag{5}$$

Note that if we combine the second and third equation above we get:

$$\frac{d(E+I)}{dt} = \beta\frac{S}{N}I - \gamma I = \beta\left(\frac{S}{N} - \frac{\gamma}{\beta}\right)I \tag{6}$$

From the above equation we can see that population in the **E** and **I** compartments together can grow with time only when the fraction of the susceptible population is greater than the inverse of the reproduction number :

$$\frac{S}{N} > \frac{1}{\mathfrak{R}} \quad \text{where} \quad \mathfrak{R} = \frac{\beta}{\gamma} \tag{7}$$

There are many forms of SEIR equations which are in common use (see [BC18, ea20p, ea20u, ea20b, P.20, ea20n]) however, equation (5) is the simplest one and does not include natural deaths. One of the common practices with the SEIR model has been to consider the incubation period $1/\sigma$ a constant, and estimates it from some other observations. The SEIR modal is quite complex as compared to the SIR model and we cannot find the number of exposed people exactly at time $t = 0$ for evolving the equations and so the approach used to define $\mathfrak{R}$ no longer works. Thanks to the new generation matrix models [BC18] it is still possible to write $\mathfrak{R}$ in a close form for this case also.

2.3 SIRD model

One of the serious drawbacks of the SIR model is that people who recover and who die are treated in the same way - there are no separate compartments for the dead and recovered people. This drawback can be addressed by separating the compartments for the dead and recovered population as is done in the SIRD model described with the following set of equations (for a detail discussion see [ea20a, Vil20]).

$$
\begin{aligned}
\frac{dS}{dt} &= -\beta \frac{S}{N} I \\
\frac{dI}{dt} &= \beta \frac{S}{N} I - (\delta + \gamma) I \\
\frac{dR}{dt} &= \gamma I \\
\frac{dD}{dt} &= \delta I
\end{aligned} \tag{8}
$$

Here a new transmission coefficient δ has been introduced which we can identify with the death rate. One of the advantages of the SIRD model is that it has three transmission coefficients β, γ and δ and we have the data for three time series $I(t), R(t)$ and $D(t)$ available so it is possible to compute the time dependency of all the three coefficients as well as the reproduction number $\mathfrak{R}$.

The aim of any mitigation measures may be one or more of the followings:

1. Lower the contact or infection rate β.
2. Lower the mortality rate δ.
3. Increase the recovery rate γ.

The SIRD model provides us a framework to estimate or fit all these parameters. In one of the coming sections we will discuss how we can reconstruct the transmission coefficients β, δ and γ as well as $\mathfrak{R}$ from the data by a direct reconstruction approach.

The basic reproduction rate for the SIRD model can be written in the following way [ea20a]:

$$
\mathfrak{R} = \frac{\beta}{\gamma + \delta}, \tag{9}
$$

or,

$$
\frac{1}{\mathfrak{R}} = \frac{1}{\mathfrak{R}_\gamma} + \frac{1}{\mathfrak{R}_\delta}, \tag{10}
$$

where $\mathfrak{R}_\gamma = \beta/\gamma$ and $\mathfrak{R}_\delta = \beta/\delta$. If apart from death and recovery there is some other channel that can lower the population in the **I** compartment, for example if infected people move out from that region with transmission coefficient η then we can write:

$$
\frac{1}{\mathfrak{R}} = \frac{1}{\mathfrak{R}_\gamma} + \frac{1}{\mathfrak{R}_\delta} + \frac{1}{\mathfrak{R}_\eta}, \tag{11}
$$

with $\mathfrak{R}_\eta = \beta/\eta$. A more realistic model will have multiple compartments (nodes), either connected in series or some branching out from others, with data to constrain the transmission coefficients (edges). Apart from this, realistic models may also require to consider different transmission coefficients of different subgroups (based on age etc.,). Incorporating, all these considerations will lead to very complex models having very less connect with the actual data we have.

2.4 Time dependent β models

As a pandemic triggers various containment measures [ea20s, DG20, ea20k, TS20, RR20, ea20c] such as lockdown, social distancing, improved hygiene practices etc., are taken and that lead to transmission coefficients such as β becoming time dependent [GG20, ea20m, FP20, ea20e, ea20i]. Apart from this, the drop in the susceptible population also decreases β (see [ea03a, Cob20]).

Lockdown has been one of the most common mitigation measures followed all over the world and, in its extreme form, we can assume that once it starts the contact rate between susceptible and infectious people drops to zero. In general, the lockdown starts on a fixed day t_l and has a duration (time scale) we call τ (we will be using both τ and corresponding decay rate $\mu = 1/\tau$ in the discussion). We can incorporate these two parameters into the modeling of $\beta(t)$ in many different ways and a set of three common choices is given below:

1. **Polynomial Suppression** [ea20m]:

$$\beta(t) = \begin{cases} \beta_0, & \text{if} \quad t \leq t_l \\ \beta_0 \left[1 - \mu \left(\frac{t-t_l}{t}\right)\right], & \text{otherwise} \end{cases} \tag{12}$$

   This model is discussed here just for an example and we do not expect the variation of $\beta(t)$ as slow as linear one. This expression shows that $\beta(t)$ starts with an initial value β_0 and after time t_l it starts decreasing linearly with a constant rate of $\mu = 1/\tau$ and finally becomes $\beta_0(1-\mu)$ at $t = \infty$.

2. tanh **Suppression**:

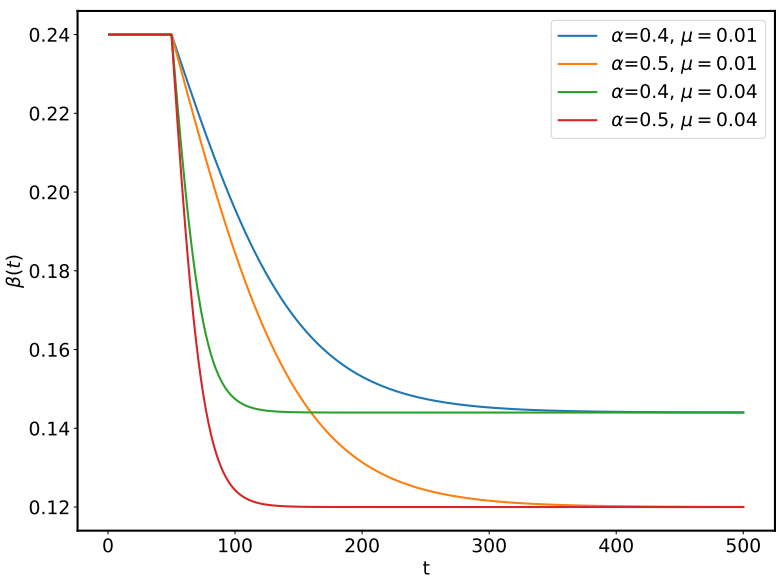

Figure 3: This figure shows the effect of α and μ on $\beta(t)$ for the **tanh** model.

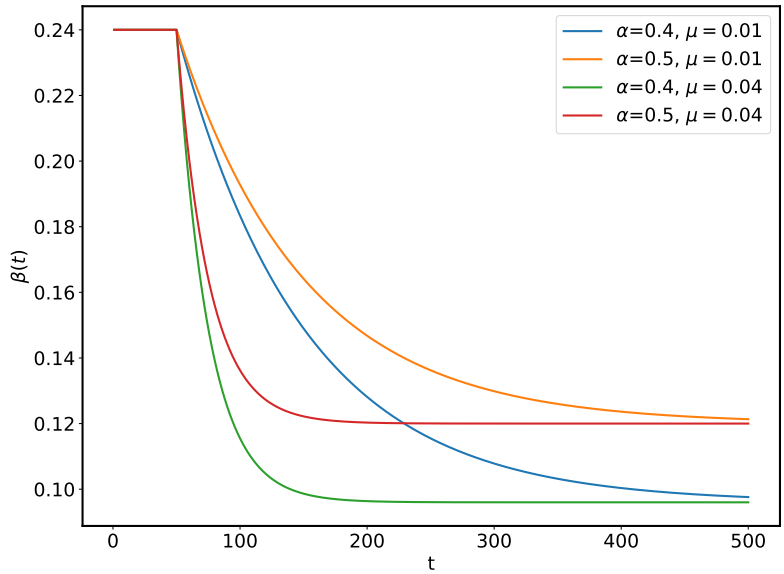

Figure 4: This figure shows the effect of α and μ on $\beta(t)$ for the **exp** model.

$$\beta(t) = \begin{cases} \beta_0, & \text{if} \quad t \leq t_l \\ \beta_0 \left[1 - \alpha * \tanh\left(\frac{t-t_l}{\tau}\right)\right], & \text{otherwise} \end{cases} \tag{13}$$

This form of suppression of $\beta(t)$ starts with a constant value β_0 at some $t = t_l$ and keeps decaying for period represented by τ and finally settles to a final value $\beta_0(1-\alpha)$ as is shown in Figure (3). This can be written in the following way also:

$$\alpha = 1 - \frac{\beta_{\text{min}}}{\beta_{\text{max}}} \tag{14}$$

From equation (13) we can also write :

$$\mu = \frac{1}{\tau} = \frac{1}{t - t_l} \operatorname{arctanh}\left[\frac{1}{\alpha}\left(1 - \frac{\beta(t)}{\beta_0}\right)\right] \tag{15}$$

Equations (14) and equation (15) are important to find the priors for α and μ once we know the priors for β and this will be discussed again in §4 and will be used in parameter estimation in §5.

3. **Exponential Suppression** [FP20, ea20o, ea20n]:

$$\beta(t) = \begin{cases} \beta_0, & \text{if} \quad t \leq t_l \\ \beta_0 \left[(1-\alpha)\exp\left(\frac{-(t-t_l)}{\tau}\right) + \alpha\right], & \text{otherwise} \end{cases} \tag{16}$$

This model is similar to the tanh model and in this case also $\beta(t)$ starts from some initial value β_0 and after decreasing for a period and finally approaches to a constant value $\beta_0\alpha$ at $t = \infty$ as

is shown in Figure (4). Note that the transmission coefficient β may decay with time without any intervention also as is discussed in [ea97] for plants.

In this case also we can write :

$$\alpha = \frac{\beta_{\min}}{\beta_{\max}}, \tag{17}$$

and

$$\mu = \frac{1}{\tau} = -\frac{1}{t - t_l} \log \left[\frac{1}{1-\alpha} \left(\frac{\beta(t)}{\beta_0} - \alpha \right) \right] \tag{18}$$

These equations also will be used to find the priors for parameter estimation.

In one of the studies [ea20o] it has been argued that even the time of recovery $1/\gamma$ may also vary with time due to the improvement in medical understanding of the epidemic and facilities and that also can be modeled as an exponential function. There have been other physically well motivated exponentially decaying forms also such as given in [Vil20] in which β starts from starting value β_0 and decay with rate $1/\tau$ finally becomes β_1.

$$\beta(t) = \beta_0(t) \exp(-t/\tau) + \beta_1(1 - \exp(-t/\tau)). \tag{19}$$

The author argues that β_1 depends on the policy decisions leading to behavioral changes. This model is different from the model we are considering only in the respect that it considers the "lockdown" from the beginning i.e., $t = 0$.

The time dependent β models as are discussed above and shown in Figure (5) share a common property that before a certain time t_l, that we can identify with the day on which lockdown starts, β has a constant value β_0 and after that it starts decreasing with a rate that depend on the parameter $\mu = 1/\tau$. The effect of the suppression in β is controlled by the parameter μ and for its zeros values all the models become constant β models. From Figure (5) we can conclude that different models can lead to the same amounts of "flattening" of the curve with a different choice of parameters so there is no preferred model for the suppression.

The SIR model with constant transmission coefficients is applicable only in the situation when the pandemic is let to grow without any intervention. In the real world once a pandemic starts interventions of different kinds (social, medical etc.,) are considered to reduce the rate at which the the epidemic spreads. These interventions can be easily taken into account by considering a time dependent (decaying) growth rate (β). As we can see from the above figure that a decaying (exponentially) β helps to contain the disease by lowering the height of the peak as shown in Figure (5).

3 Data

The primary data for Covid-19 is in terms of three times series for the count of confirmed $\mathcal{C}(t)$, recovered $\mathcal{R}(t)$ and dead $\mathcal{D}(t)$, persons for every country. By definition all the three times series

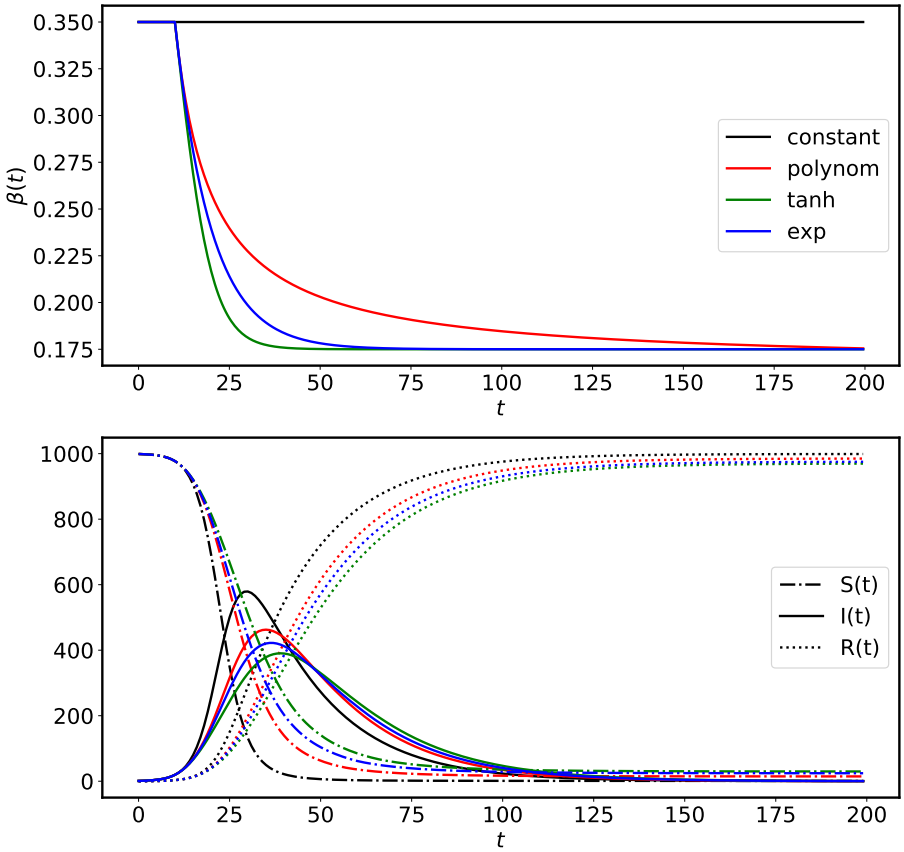

Figure 5: The top panel in the figure shows the β suppression models as are discussed in the text. The solid and dashed line in the bottom panel show the number of infected people $I(t)$ and the number of recovered people $R(t)$ for the SIR model corresponding to different β models. From the bottom panel we can see that all the models have similar effects, however, for the tanh case the suppression in the peak of $I(t)$ is maximum. This figure is for an illustration the values of parameters have been chosen carefully to highlight the effects.

are non-decreasing functions of time, as are shown in Figure (6) for Italy. The data for Covid-19 is provided by the John Hopkins University [ea20g, Gar20a, Gar20b] and is updated on daily basis and interactive tools are also provided for data exploration. The worldometer website [Wor20] and **our world in data** [iD20] also provide an up to-date data for most countries with some extra information, such as the numbers for the active cases, critical cases and the test conducted for a million population of the country.

The time series which are shown in Figure (6) grow exponentially in the beginning and then settle to a slower power law growth. If we look at the daily new cases, as shown in Figure (7), they look like broad peaked functions with increasing fluctuations around the the current value. Time series for a set of forty five countries used in the analysis are shown in the Figure (8).

There are many factors, known and unknown, which determine the behavior of these time series,

S. No	country	starting_date	num_days	population
1	**US**	2020-02-21	91	331,002,651
2	**Russia**	2020-03-13	70	145,934,462
3	**Brazil**	2020-03-10	73	212,559,417
4	**UK**	2020-03-01	82	67,886,011
5	**Spain**	2020-02-28	84	46,754,778
6	**Italy**	2020-02-22	90	60,461,826
7	**France**	2020-02-28	84	65,273,511
8	**Germany**	2020-02-27	85	83,783,942
9	**Turkey**	2020-03-17	66	84,339,067
10	**Iran**	2020-02-23	88	83,992,949
11	**India**	2020-03-05	78	1,380,004,385
12	**Peru**	2020-03-13	70	32,971,854
13	**China**	2020-03-11	72	1,439,323,776
14	**Canada**	2020-03-04	79	37,742,154
15	**SaudiArabia**	2020-03-12	71	34,813,871
16	**Mexico**	2020-03-15	68	128,932,753
17	**Chile**	2020-03-13	70	19,116,201
18	**Belgium**	2020-03-05	78	11,589,623
19	**Pakistan**	2020-03-13	70	220,892,340
20	**Netherlands**	2020-03-04	79	17,134,872
21	**Qatar**	2020-03-11	72	2,881,053
22	**Ecuador**	2020-03-14	69	17,643,054
23	**Belarus**	2020-03-16	67	9,449,323
24	**Sweden**	2020-03-04	79	10,099,265
25	**Switzerland**	2020-03-01	82	8,654,622
26	**Portugal**	2020-03-08	75	10,196,709
27	**Singapore**	2020-02-05	107	5,850,342
28	**Bangladesh**	2020-03-24	55	164,689,383
29	**UAE**	2020-03-07	76	9,890,402
30	**Ireland**	2020-03-10	73	4,937,786
31	**Indonesia**	2020-03-11	72	273,523,615
32	**Poland**	2020-03-11	72	37,846,611
33	**Ukraine**	2020-03-20	63	43,733,762
34	**SA**	2020-03-14	69	59,308,690
35	**Kuwait**	2020-02-26	86	4,270,571
36	**Colombia**	2020-03-15	68	50,882,891
37	**Romania**	2020-03-11	72	19,237,691
38	**Israel**	2020-03-08	75	8,655,535
39	**Japan**	2020-02-06	98	126,476,461
40	**Austria**	2020-03-04	79	9,006,398
41	**Egypt**	2020-03-08	75	102,334,404
42	**DR**	2020-03-19	64	10,847,910
43	**Philippines**	2020-03-10	73	109,581,078
44	**Denmark**	2020-03-08	75	5,792,202
45	**Serbia**	2020-03-13	70	8,737,371
46	**Panama**	2020-03-13	70	4,314,767

Table 2: A summary for the countries used for modelling

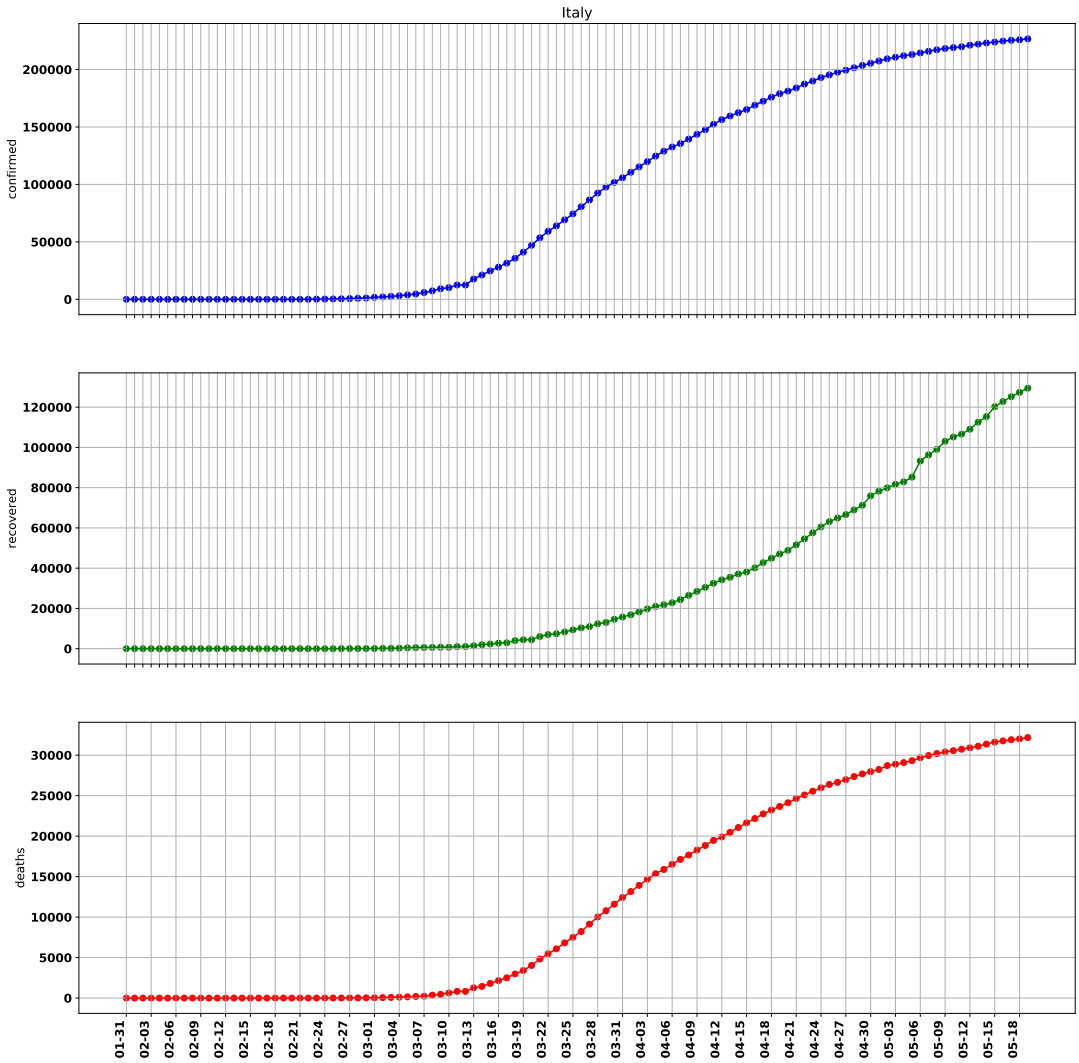

Figure 6: The total number of confirmed, recovered and dead case for Italy.

such as the nature of the diseases/virus, the health profile of the population, availability of medical infrastructure, age-distribution, social mixing, personal hygiene and may be geographical location and genetic makeup of the population. Compartmental models, such as SIR, SIRD, SEIR etc., help to simplify the problem by replacing a large number of controlling parameters with a very small set of well motivated parameters - the transmission coefficients.

The compartmental models predict that how the population in different compartments change with time. In the SIR model the population in **S** and **R** compartments can only decrease and increase respectively, however, in the **I** compartment it can increases as well as decreases. On the onset of the epidemic almost everyone is in the **S** compartment with a very small fraction in **I** and no one in the **R** compartment. At the end of the pandemic everyone is in the **R** compartment with no one in **S**

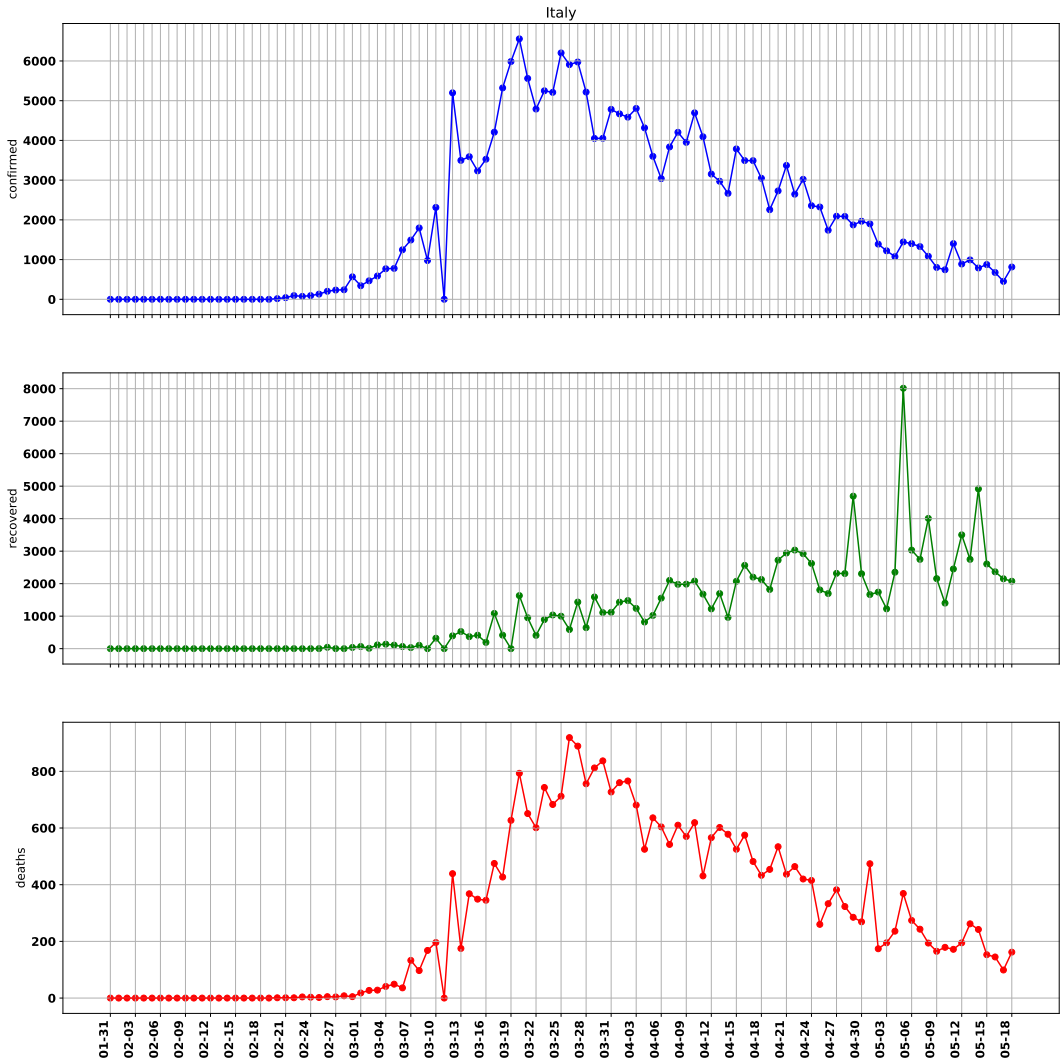

Figure 7: The number of new confirmed, recovered and dead case for Italy.

or **I** compartment (see Figure (1). The same happens for SIRD model also where the decrease in **I** compartment happens due to recovery (R) and deaths (D).

The time series $I(t)$ for the population in compartment **I** can be obtained by subtracting $\mathcal{R}$ and $\mathcal{D}$ from $\mathcal{C}$.

$$I(t) = \mathcal{C}(t) - \mathcal{R}(t) - \mathcal{D}(t) \tag{20}$$

For a set of 45 countries the time series of $I(t)$ are shown in Figure (8). The time series $\mathcal{S}, \mathcal{I}, \mathcal{R}$ and $\mathcal{D}$ have very weak dependency on each other - the only constrains they have to satisfy is that the sum total of the population in different compartment must add to the total population. The number of people $I(t)$ in the **I** compartment at time t does depend on all the three transmission rates β, γ, and

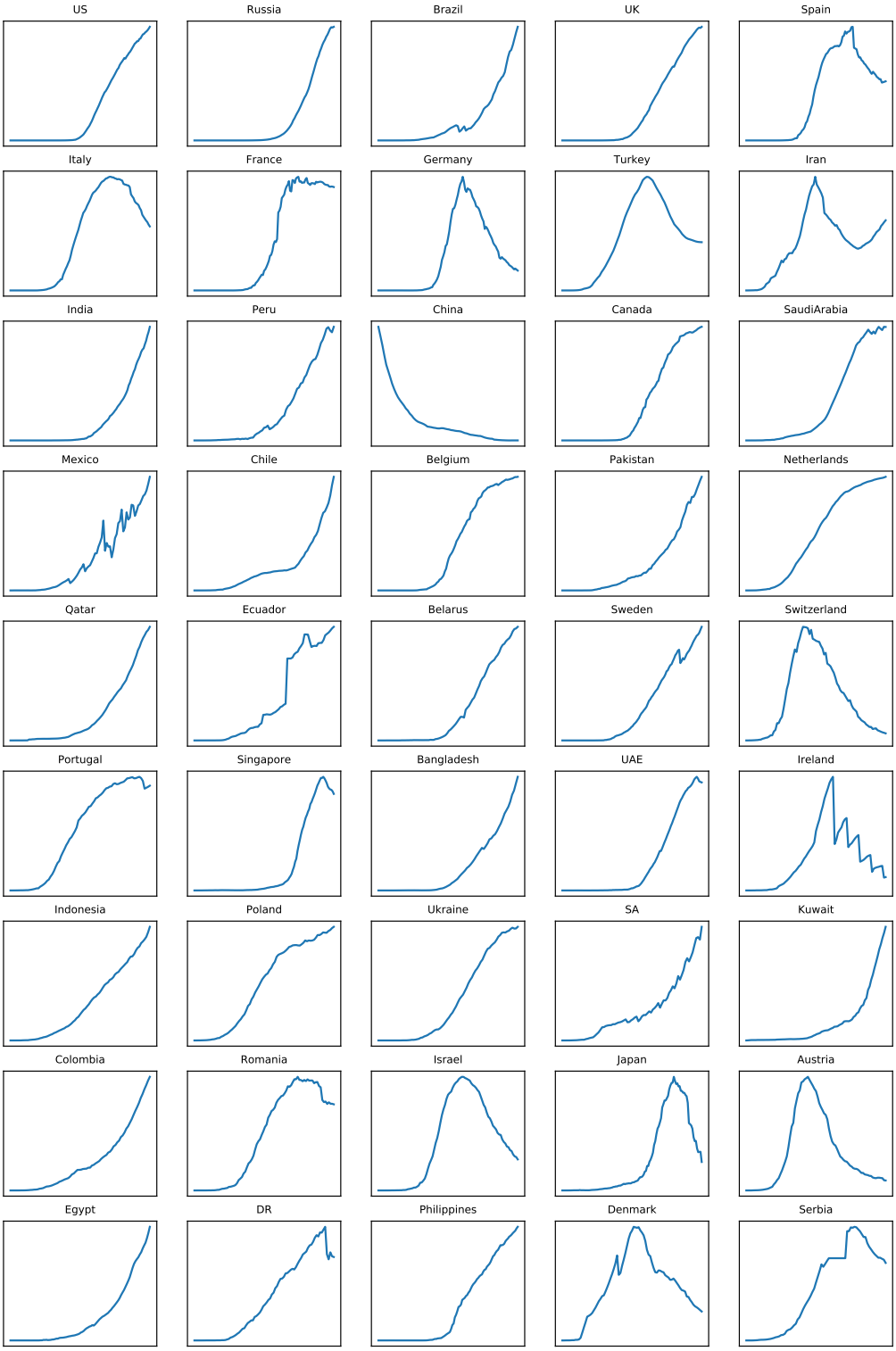

Figure 8: Time series for a set 45 countries we used for modeling.

δ, therefore it is a good measure which we can fit to a compartmental model, such as SIR or SIRD and can get constraints.

4 Reconstruction

In this and next section we present the main results of the study in the form of demonstrating a reconstruction procedure for the time dependent transmission coefficients $\beta(t)$, $\gamma(t)$, $\delta(t)$ and the effective reproduction number $\mathfrak{R}(t)$. We consider an example of Italy and India for this procedure. Note that this approach is common and can be used to understand the variation of the transmission coefficients with time as a result of interventions. The main advantages of this approach is that there are no parameters to adjust and so the results are easy to reproduce.

The approach we use here is similar to as used in [ea20e, GG20]. In this approach the evolution equations are written in a discretized form as shown in equation (21).

$$\begin{aligned} S_{t+1} &= S_t - \beta(t)\frac{S_t}{N}I_t \\ I_{t+1} &= I_t + \beta(t)\frac{S_t}{N}I_t - \gamma(t)I_t \\ R_{t+1} &= R_t + \gamma(t)I_t \end{aligned} \tag{21}$$

From the third equation we can write:

$$\gamma(t) = \frac{R_{t+1} - R_t}{I_t}, \tag{22}$$

and using this and second equation from (21) we get,

$$\beta(t) = \frac{(I_{t+1} - I_t) + (R_{t+1} - R_t)}{I_t}. \tag{23}$$

Note that by definition $R_{t+1} \geq R_t$, so $\gamma(t) \geq 0$, however, we may have $I_{t+1} \leq I_t$ also, $\beta(t)$ may become negative also once the population in the compartment **I** starts decreasing.

Here an important assumption is being made and that is the fraction of susceptible population S/N is close to unity which may be true at the beginning of the epidemic. Once we have expressions for the time dependent β and γ we can also written an expression for the time dependent reproduction number in the following way:

$$\mathfrak{R}_{SIR}(t) = \frac{\beta(t)}{\gamma(t)} = \frac{(I_{t+1} - I_t) + (R_{t+1} - R_t)}{R_{t+1} - R_t} = 1 + \frac{I_{t+1} - I_t}{R_{t+1} - R_t} \tag{24}$$

Following the similar procedure we can write the SIRD equations in the following discretized form:

$$
\begin{aligned}
S_{t+1} &= S_t - \beta(t)\frac{S_t}{N}I_t \\
I_{t+1} &= I_t + \beta(t)\frac{S_t}{N}I_t - (\gamma(t)+\delta(t))I_t \\
R_{t+1} &= R_t + \gamma(t)I_t \\
D_{t+1} &= D_t + \delta(t)I_t
\end{aligned} \tag{25}
$$

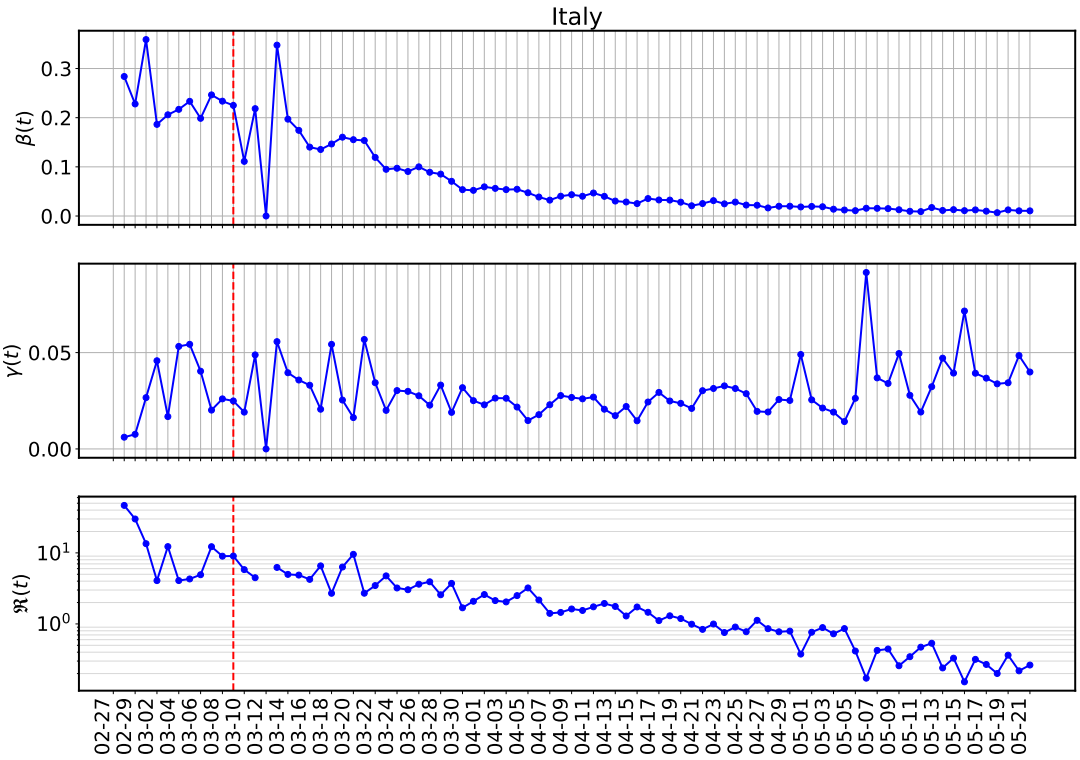

Figure 9: Reconstruction of the transmission coefficients $\beta(t)$, $\gamma(t)$ and the effective reproduction number $\mathfrak{R}(t)$ with the SIR model for Italy. In this and next figures the red vertical dashed line shows the day of public lockdown enforcement.

From these questions we can write :

$$
\begin{aligned}
\beta(t) &= \frac{(I_{t+1}-I_t)+[(R_{t+1}-R_t)+(D_{t+1}-D_t)]}{I_t} \\
\gamma(t) &= \frac{R_{t+1}-R_t}{I_t} \\
\delta(t) &= \frac{D_{t+1}-D_t}{I_t},
\end{aligned} \tag{26}
$$

and can write the expression for the reproduction number:

$$
\mathfrak{R}_{SIRD}(t) = \frac{\beta}{\gamma+\delta} = \frac{\Delta I_t + \Delta R_t + \Delta D_t}{\Delta R_t + \Delta D_t} = 1 + \frac{\Delta I_t}{\Delta R_t + \Delta D_t}, \tag{27}
$$

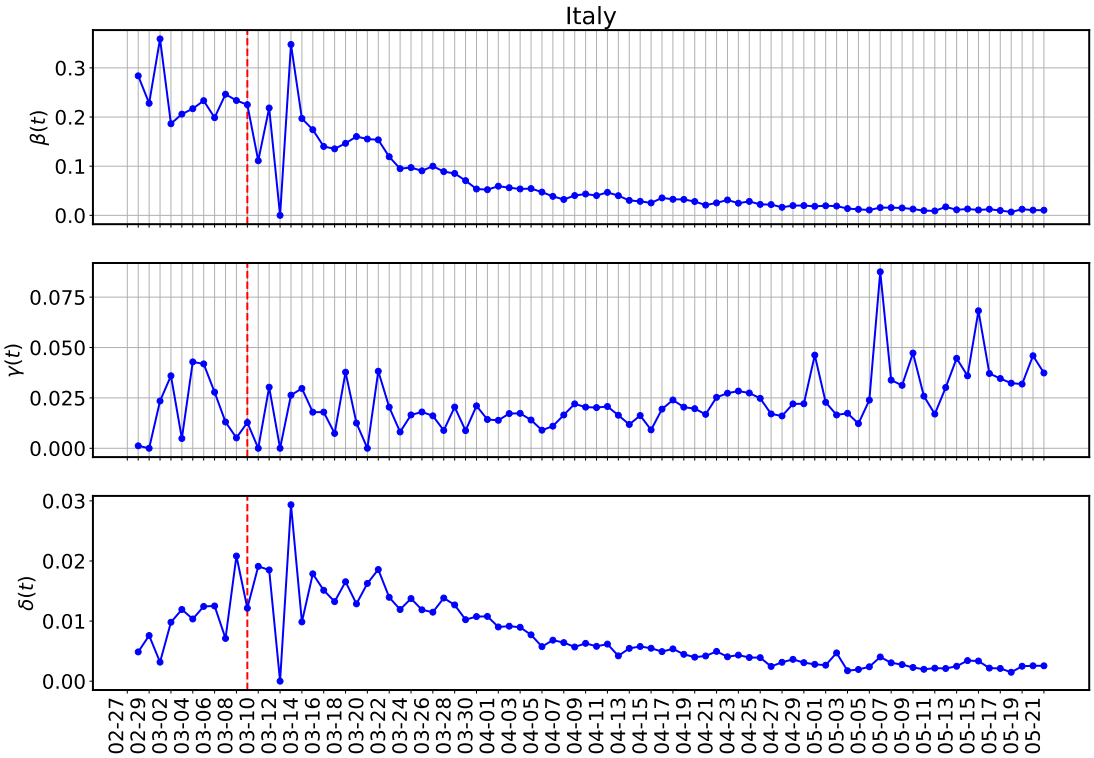

Figure 10: Reconstruction of the transmission coefficients $\beta(t)$, $\gamma(t)$ and $\delta(t)$ with the SIRD model for Italy.

where $\Delta X_t = X_{t+1} - X_t$ with $X = I, R$ and D. This equation is identical to equation (24) if we do not count dead and recovered separately i.e., replace $\Delta R_t + \Delta D_t$ with ΔR_t. One of the interpretations of R is that it is a ratio of two rates and so in case we are interested finding out two separates measures for γ and δ, we can also write :

$$\mathfrak{R}_\gamma = \frac{\beta}{\gamma} \quad \text{and} \quad \mathfrak{R}_\delta = \frac{\beta}{\delta}, \tag{28}$$

and so,

$$\frac{1}{\mathfrak{R}} = \frac{1}{\mathfrak{R}_\gamma} + \frac{1}{\mathfrak{R}_\delta} \tag{29}$$

The procedure as discussed above can be used to know the variation of the transmission coefficients β, γ, δ and effective reproduction $\mathfrak{R}(t)$with time. In order to follow this procedure we need to abandon first few data points which have very high noise. As explained above occasionally we may also have negative values of $\mathfrak{R}(t)$.

In figure (9) we show the reconstruction for $\beta(t), \gamma(t)$ and $\mathfrak{R}(t)$for Italy with SIR model and in figure (10) the same is shown for $\beta(t), \gamma(t)$ and $\delta(t)$ in case of SIRD model. We show $\mathfrak{R}(t)$as well as $\mathfrak{R}_\gamma(t)$ and $\mathfrak{R}_\delta(t)$ for Italy in the case of SIRD model in figure (11). Similar figure are also shown for India in figure (12), (13) and (14). In all the figures the vertical red dashed lines mark the date of lockdown.

One of the main attractive feature of the split of $\mathfrak{R}$ for δ and γ is that it is sensitive to individual values of γ and δ and not just their sum $\gamma + \delta$. For the case when $\gamma = \delta$ we get the usual reproduction

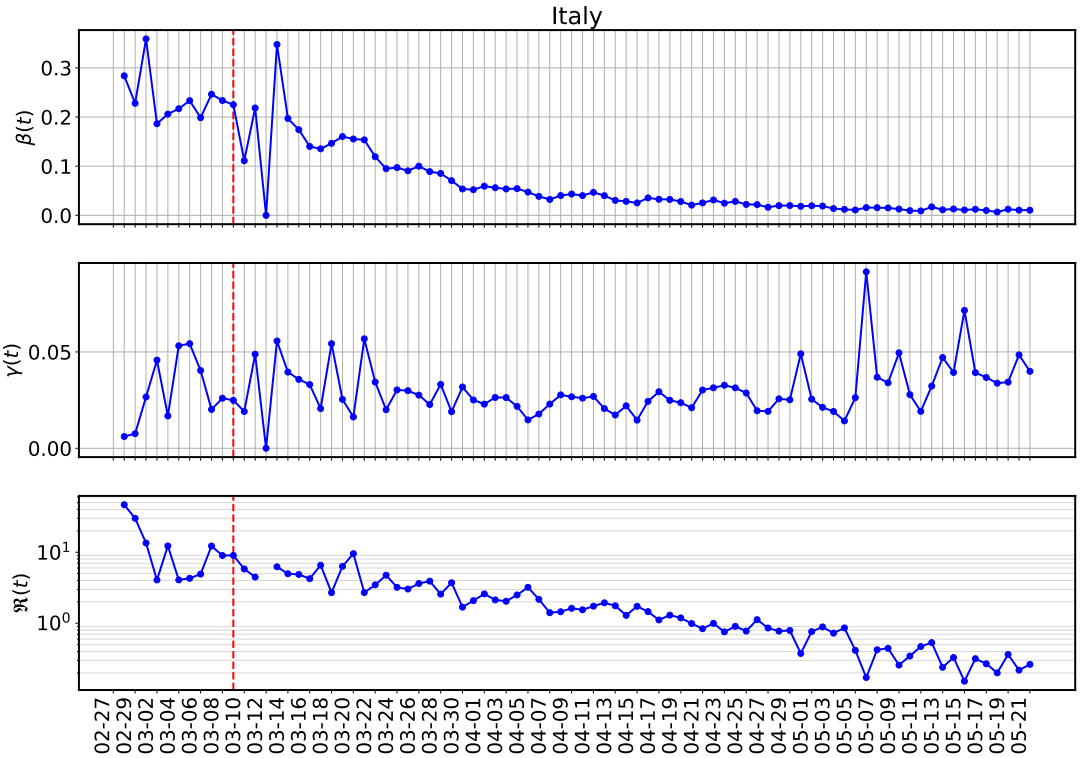

Figure 11: Reconstruction of the and the effective reproduction number $\mathfrak{R}_\gamma(t)$, $\mathfrak{R}_{\delta(t)}$ and $\mathfrak{R}(t)$ with the SIRD model for Italy.

number.

4.1 Priors

One of the import uses of the reconstruction procedure we have discussed here is to find the priors (minimum, maximum and best fit) values for the parameters to be fitted. Once we have the estimates for $\beta(t), \gamma(t)$ and $\delta(t)$ from the above procedure we can easily find $X_{\min}, X_{\max}, X_0$, values (with $X = \beta, \gamma, \delta$. Here, X_0 is the approximate point for the parameter that is needed in many optimization procedure which iteratively find the solution. Since in the present work we use parametric form of $\beta(t)$, so we need priors for the parameters of $\beta(t)$ i.e., β_1, α, μ and τ which can find from the reconstructed $\beta(t)$ (see §2.4 for detail).

5 Model fitting and parameter estimation

We consider a set of six compartmental models, three belonging to the SIR and three to the SIRD class. The models are different from each other in terms of the choice for the epidemiological class (SIR or SIRD) or the model for the contact rate $\beta(t)$ (see §2.4 for detail). A summary of the models is given in Table (3).

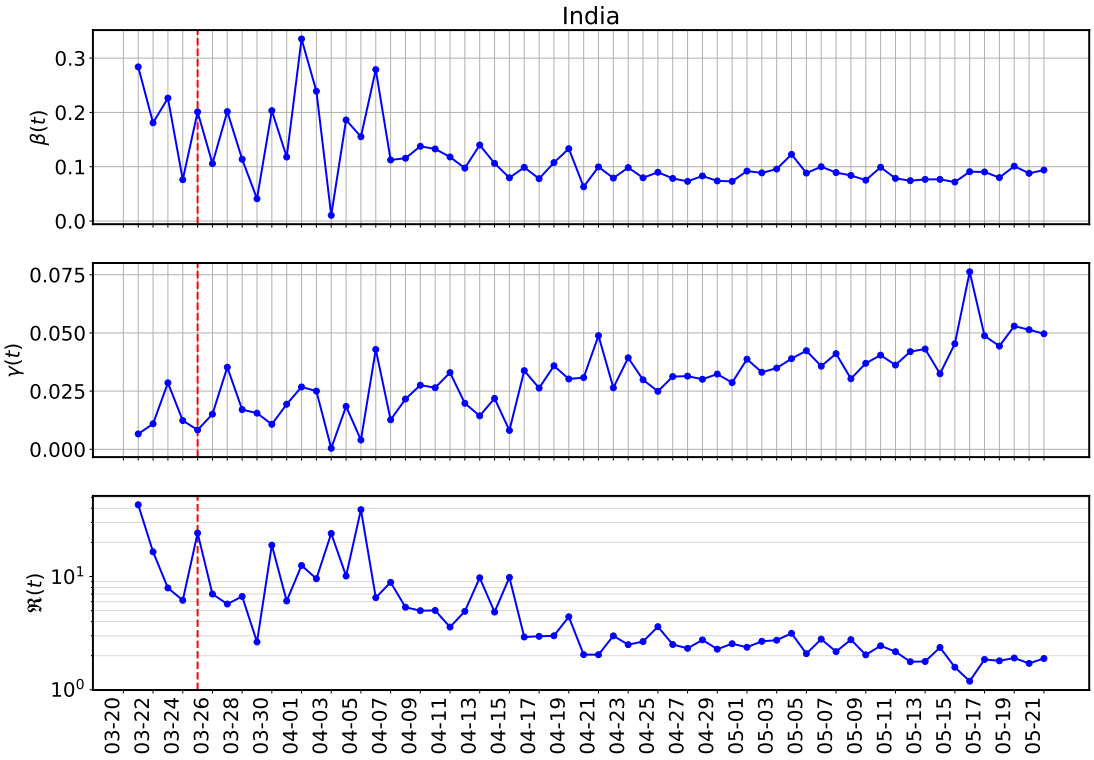

Figure 12: Reconstruction of the transmission coefficients $\beta(t)$, $\gamma(t)$ and the effective reproduction number $\mathfrak{R}(t)$ with the SIR model for India.

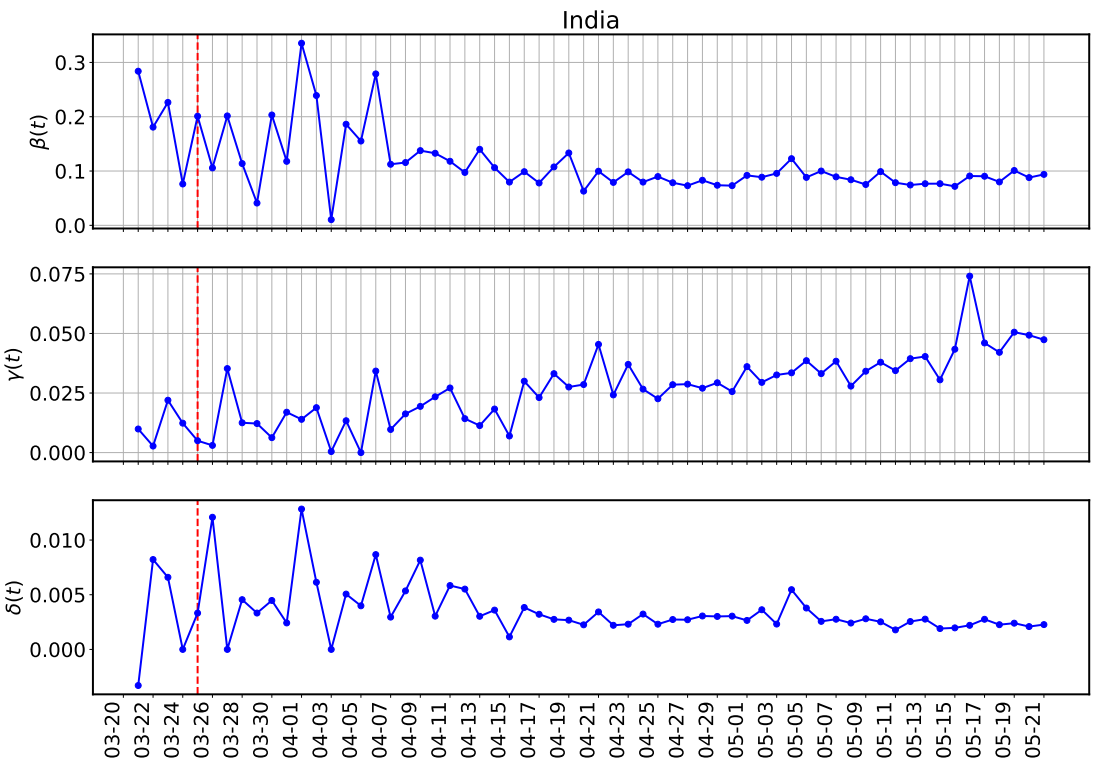

Figure 13: Reconstruction of the transmission coefficients $\beta(t)$, $\gamma(t)$ and $\delta(t)$ with the SIRD model for India.

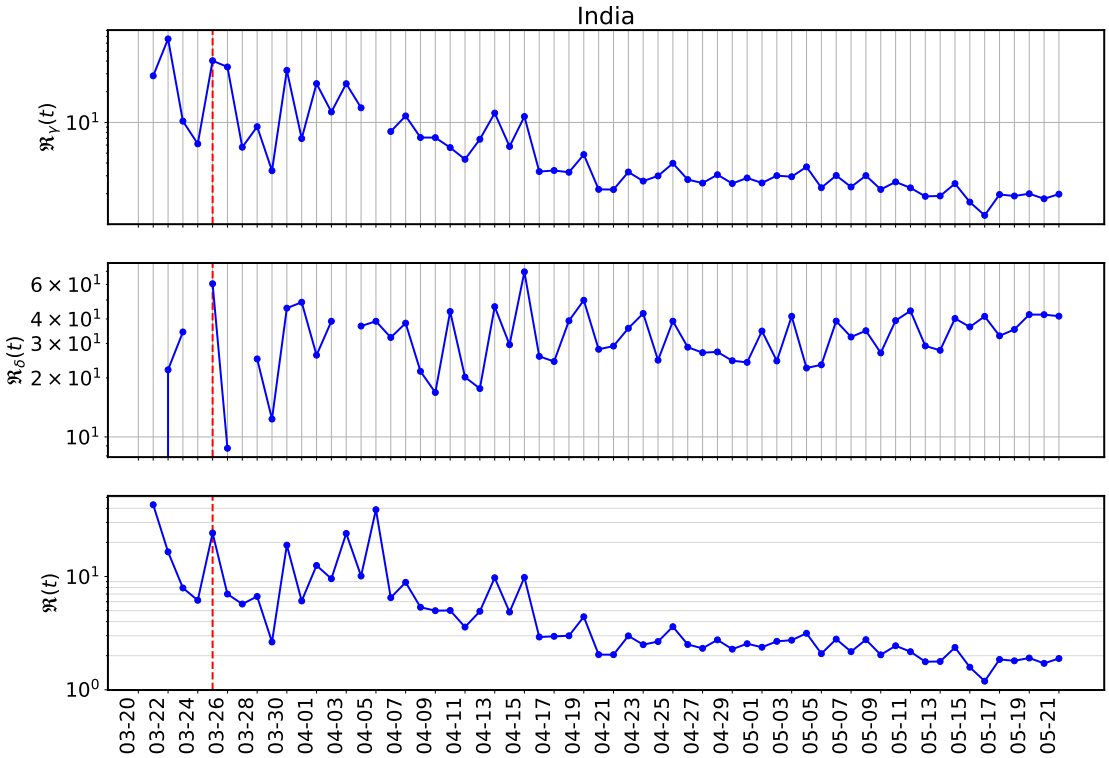

Figure 14: Reconstruction of the and the effective reproduction number $\mathfrak{R}_\gamma(t)$, $\mathfrak{R}_{\delta(t)}$ and $\mathfrak{R}(t)$ with the SIRD model for India.

Model	**Epidemiological model**	$\beta(t)$ **Model**	**Parameters**
1	SIR	exp	$(\beta, \gamma), (\beta_0, \alpha, \mu, t_l)$
2	SIR	tanh	$(\beta, \gamma), (\beta_0, \alpha, \mu, t_l)$
3	SIRD	exp	$(\beta, \gamma, \delta), (\beta_0, \alpha, \mu, t_l)$
4	SIRD	tanh	$(\beta, \gamma, \delta), (\beta_0, \alpha, \mu, t_l)$
5	SIR	exp	$(\beta, \gamma), (\beta_0, \mu)$
6	SIRD	exp	$(\beta, \gamma, \delta), (\beta_0, \mu)$

Table 3: A summary of the models being considered for the analysis

Note that in the model (5) and (6), $\beta(t)$ starts decaying from the very beginning (in place of starting from a particular day representing the date of the lockdown) with a constant rate μ.

In any fitting procedure the choice of the loss function depends on what we wish to fit. In the common least square fitting we use the sum of the squares of the offsets as the loss function. However, there is a problem here with the data we have for that choice. The time series we wish to fit have small values at the beginning and very large values at the later stage, so the the fitting is biased towards the points which have large values. One of the solutions for this could be to fit the log of the time series but then the fitting becomes biased toward small values, in the beginning (or later stage when

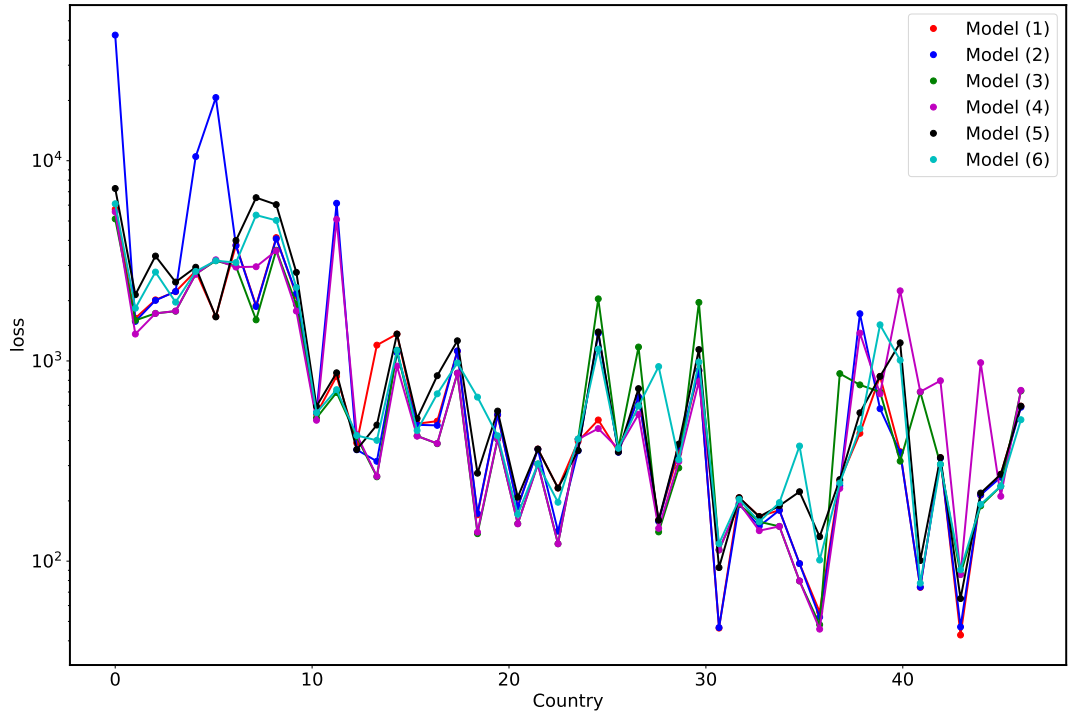

Figure 15: This figure shows the value of the loss function at the best fit point. From the figure we can see that different models fit equally well.

the peaks falls).

We decide to use the loss function of the ordinary least square which fits the data points close to the peak (having higher values) more accurately than other data points. We found this useful for the following two reasons:

1. The peak in the time series is an important feature, in particular its location and height, therefore any loss function biased towards it is justified.

2. For the short term predictions only the data points close to the dates of prediction is important, so using a loss function that fits later points (having higher values) more accurately than the noisy data points in the beginning is favorable.

The loss function which we used for fitting the data to SIR and SIR models is given below.

$$\mathcal{L}^c(\theta|\mathbb{D}) = \frac{1}{n_d}\sum_{l=1}^{l=n_d}\frac{w_l}{n_t}\sum_{i=1}^{i=n_t}(X^o_{l,i} - X^p_{l,i})^2 \tag{30}$$

The variables used in the above equations are defined in table (4):

For SIR and SIRD models we fit multiple time series together so we must weight the sum of the squares of the offsets for different time series since they have very different values - the value of $I(t)$ is generally few orders of magnitude higher than $R(t)$ and $D(t)$. We use the following weights for this

Variable	Name	Description
$\mathbb{D}$	Primary Data	$\mathcal{C}, \mathcal{R}, \mathcal{D}$
θ	Fitting parameters	As defined in the last column of Table (3)
X_l^o	Fitting data	$I, R+D$ for SIR and $I,\ R,\ D$ for SIRD
X_l^p	Predicted data	prediction from the model
n_t	Days	Number of days for which we have data
n_d	Number of time series used	2 for SIR and 3 for SIRD
c	country	label for the country
w_l	weight	Used for fitting multiple time series together

Table 4: A summary of the variables

purpose:

$$w_l = 1 - \frac{\bar{X}_l}{\sum_l \bar{X}_l}, \tag{31}$$

where $\bar{X}_l$ is the average of the time series $X(t)$.

We use the **solve_ivp** and **minimize** modules from the `Scipy` [Sci20] for integrating the differential equations and minimize the cost function respectively. The loss function given by equation (30) represents the Root Mean Square Deviation (RMSD) and we use its final value as a measure of the goodness of the fit, and is shown by the points of different colors (for different models) in the Figure (15). From this figure we can see that it is hard to conclude that which of the models fit the data best (has always lowest RMSD). We can also notice from the figure that the model (6) is less sensitive towards the choice of a country for fitting the model (has the smallest fluctuations).

A list of fitting parameter for the different models is given in Table (3). For the SIR class of models, model (1) and (2) we have five fitting parameters named, $\gamma, \beta_0, \alpha, \mu, t_l$ and for the SIRD models, model (3) and (4) we have six fitting parameters named $\gamma, \beta_0, \alpha, \mu, t_l$ and δ. As we can notice that for model (1) to model (4) four of the parameters are associated with $\beta(t)$ and for model (5) and (6) the variation of $\beta(t)$ is controlled by just two parameter - β_0, the initial value of β and its decay rate $\mu = 1/\tau$.

The best fit values of the fitting parameters with their 90 % CI (standard deviation) as well as the median values are given Table (5). The tables also give the estimate for the effective reproduction number $\mathfrak{R}(t)$which is a derived quantity here. Note that for computing $\mathfrak{R}_0$we have extrapolated the value of $\mathfrak{R}(t)$to the last date for which the data is being used here.

A histogram of the effective reproduction number for the different models being considered is shown in Figure (16) and detail values of that for different countries, which include the average values as well as 90% CI (StdDev), are given in Table (6).

Models	Parametes						
	S.No.	**Parameter**	**Aver**	**Std Dev**	**Median**	**Min**	**Max**
model (1)	1	γ	0.0350	0.0268	0.0275	0.0050	0.1590
	2	β_0	0.3028	0.1292	0.2975	0.0090	0.6560
	3	α	0.1292	0.2052	0.0330	0.0010	0.9130
	4	μ	0.0958	0.0845	0.0705	0.0100	0.4000
	5	t_l	9.2555	11.3031	2.4900	0.0000	40.8920
	6	$\mathfrak{R}(t)$	1.6133	1.3235	1.4485	0.0000	6.7690
model (2)	1	γ	0.0470	0.0617	0.0280	0.0050	0.4100
	2	β_0	0.2867	0.1229	0.2680	0.0010	0.5230
	3	α	0.8168	0.2482	0.9320	0.0040	1.0000
	4	μ	0.0677	0.0675	0.0455	0.0100	0.3650
	5	t_l	10.6100	15.0991	2.1895	0.0000	71.1930
	6	$\mathfrak{R}(t)$	1.6110	1.3100	1.2860	0.0000	6.6590
model (3)	1	δ	0.0048	0.0053	0.0030	0.0000	0.0270
	2	γ	0.0584	0.0818	0.0280	0.0000	0.3490
	3	β_0	0.3222	0.1342	0.3105	0.0120	0.6550
	4	α	0.2245	0.2977	0.0990	0.0010	1.0000
	5	μ	0.0867	0.0876	0.0645	0.0100	0.4570
	6	t_l	9.4041	12.6319	3.0380	0.0000	45.7770
	7	$\mathfrak{R}(t)$	1.5971	1.2192	1.2215	0.0060	6.7840
model (4)	1	δ	0.0042	0.0039	0.0030	0.0000	0.0220
	2	γ	0.0577	0.0853	0.0275	0.0000	0.4110
	3	β_0	0.2924	0.1304	0.2770	0.0010	0.6300
	4	α	0.7540	0.2975	0.8780	0.0010	1.0000
	5	μ	0.0766	0.1063	0.0455	0.0100	0.5440
	6	t_l	10.2211	13.1231	1.7215	0.0000	49.8160
	7	$\mathfrak{R}(t)$	1.6105	1.2643	1.4000	0.0010	6.7190
model (5)	1	γ	0.0360	0.0283	0.0265	0.0050	0.1600
	2	β_0	0.3264	0.1299	0.3025	0.0010	0.6020
	3	μ	0.0356	0.0187	0.0365	0.0100	0.1010
	4	$\mathfrak{R}(t)$	1.2654	1.0931	1.0965	0.0000	5.8410
model (6)	1	δ	0.0046	0.0049	0.0030	0.0000	0.0240
	2	γ	0.0417	0.0489	0.0255	0.0000	0.2740
	3	β_0	0.3342	0.1251	0.3260	0.0080	0.6020
	4	μ	0.0332	0.0162	0.0360	0.0100	0.0700
	5	$\mathfrak{R}(t)$	1.1739	1.0386	1.0335	0.0420	5.8410

Table 5: A summary of the fitting parameters for all the models.

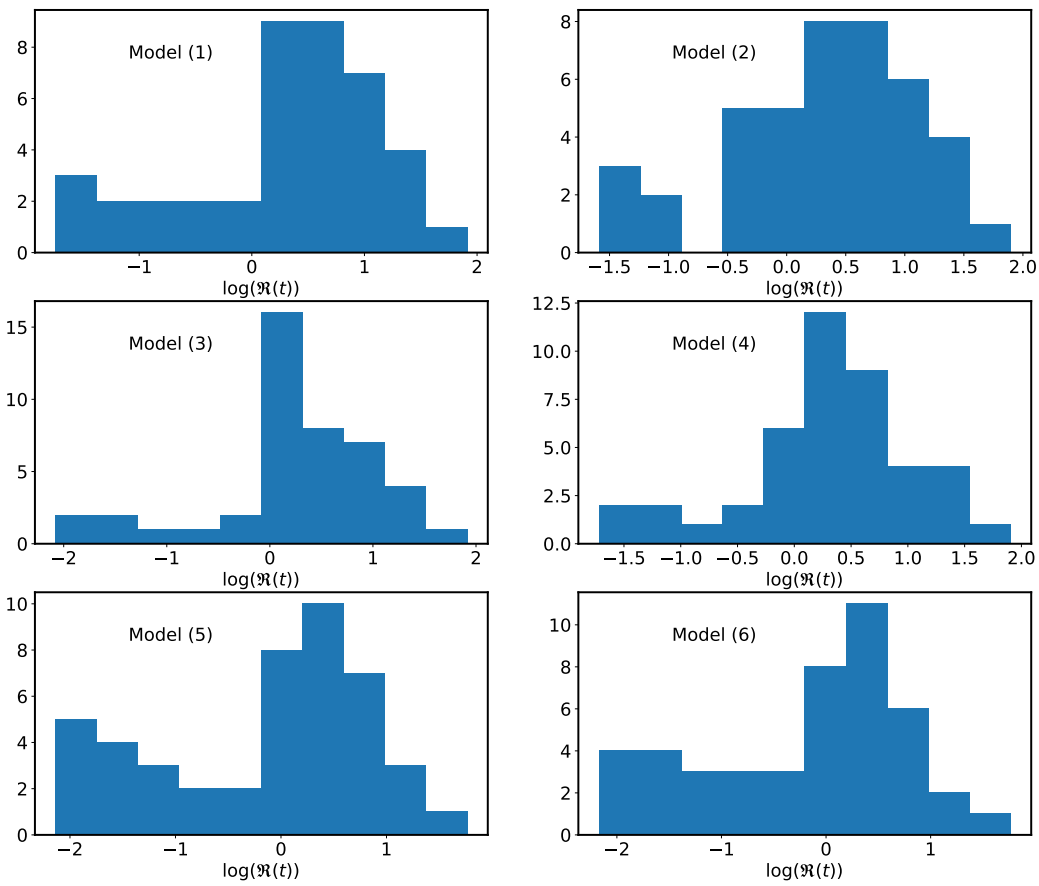

Figure 16: Histogram of the effective reproduction number $\mathfrak{R}(t)$with different models.

S. No	country	Model (1)	Model (2)	Model (3)	Model (4)	Model (5)	Model (6)	Aver	StdDev
1	**US**	1.47	3.206	2.057	1.447	1.383	1.383	1.824	0.661
2	**Russia**	1.939	1.568	2.898	1.475	2.824	2.814	2.253	0.61
3	**Brazil**	2.06	2.063	2.105	2.104	1.545	1.519	1.899	0.26
4	**UK**	3.456	3.592	3.988	4.126	1.453	1.453	3.011	1.125
5	**Spain**	0.275	0.212	0.276	0.381	0.141	0.141	0.238	0.084
6	**Italy**	0.196	1.66	0.21	0.27	0.193	0.208	0.456	0.539
7	**France**	0.482	0.582	0.491	0.573	0.253	0.265	0.441	0.134
8	**Germany**	0.203	0.204	0.16	0.543	0.296	0.296	0.284	0.126
9	**Turkey**	0.072	0.032	0.053	0.032	0.403	0.395	0.164	0.166
10	**Iran**	0.741	0.717	0.689	0.764	0.172	0.186	0.545	0.26
11	**India**	1.784	1.907	2.012	1.769	1.878	1.87	1.87	0.081
12	**Peru**	2.062	1.214	1.986	1.207	1.621	1.733	1.637	0.336
13	**China**	0.0	0.002	0.125	0.001	0.0	0.042	0.028	0.046
14	**Canada**	1.611	1.054	1.073	1.027	0.66	0.66	1.014	0.319
15	**SaudiArabia**	1.765	0.627	1.782	1.479	1.763	1.842	1.543	0.426
16	**Mexico**	1.16	1.191	1.181	1.193	1.026	1.014	1.128	0.077
17	**Chile**	2.104	2.148	2.144	2.159	1.607	1.607	1.962	0.251
18	**Belgium**	1.178	1.229	1.216	1.26	0.323	0.343	0.925	0.419
19	**Pakistan**	2.863	2.879	2.862	2.887	1.835	1.015	2.39	0.722
20	**Netherlands**	2.616	2.729	2.91	3.287	0.349	0.499	2.065	1.18
21	**Qatar**	6.769	6.659	6.784	6.719	5.841	5.841	6.436	0.422
22	**Ecuador**	1.427	1.115	1.351	0.956	1.379	1.31	1.256	0.166
23	**Belarus**	1.921	1.633	1.603	1.667	1.911	1.982	1.786	0.155
24	**Sweden**	2.59	1.343	1.564	1.557	1.054	1.131	1.54	0.508
25	**Switzerland**	0.011	0.655	0.971	0.002	0.163	0.163	0.328	0.361
26	**Portugal**	1.111	1.658	0.937	1.18	0.962	0.876	1.121	0.262
27	**Singapore**	4.198	4.22	1.13	4.237	3.544	3.544	3.479	1.093
28	**Bangladesh**	3.025	2.729	3.226	2.343	2.998	0.864	2.531	0.796
29	**UAE**	1.209	1.774	1.159	1.786	1.815	1.815	1.593	0.29
30	**Ireland**	0.175	0.218	0.809	0.179	0.208	0.217	0.301	0.228
31	**Indonesia**	2.385	2.446	2.227	2.308	1.438	1.504	2.051	0.416
32	**Poland**	0.699	1.208	1.152	1.385	0.683	0.695	0.97	0.287
33	**Ukraine**	1.545	0.987	1.605	1.258	1.504	1.587	1.414	0.223
34	**SA**	2.031	2.056	2.032	2.056	1.694	1.533	1.9	0.208
35	**Kuwait**	3.314	3.317	3.33	3.329	2.654	1.37	2.886	0.72
36	**Colombia**	3.309	3.334	3.174	3.204	2.153	2.017	2.865	0.556
37	**Romania**	0.482	0.336	1.365	0.317	0.49	0.52	0.585	0.357
38	**Israel**	0.019	1.483	0.952	1.415	0.118	0.114	0.684	0.623
39	**Japan**	1.181	0.0	1.18	1.068	1.176	0.594	0.866	0.439
40	**Austria**	0.006	0.0	0.006	1.842	0.245	0.245	0.391	0.658
41	**Egypt**	3.032	3.042	1.088	1.084	2.245	2.373	2.144	0.806
42	**DR**	1.149	1.05	1.227	1.063	1.139	1.214	1.14	0.067
43	**Philippines**	2.319	2.087	1.889	2.0	1.014	1.061	1.728	0.505
44	**Denmark**	0.317	0.384	0.318	1.709	0.127	0.134	0.498	0.55
45	**Serbia**	0.915	0.714	0.976	0.25	0.903	0.958	0.786	0.254
46	**Panama**	1.036	0.84	1.195	1.183	1.027	1.052	1.056	0.118

Table 6: Effective reproduction number $\mathfrak{R}(t)$for countries with different models

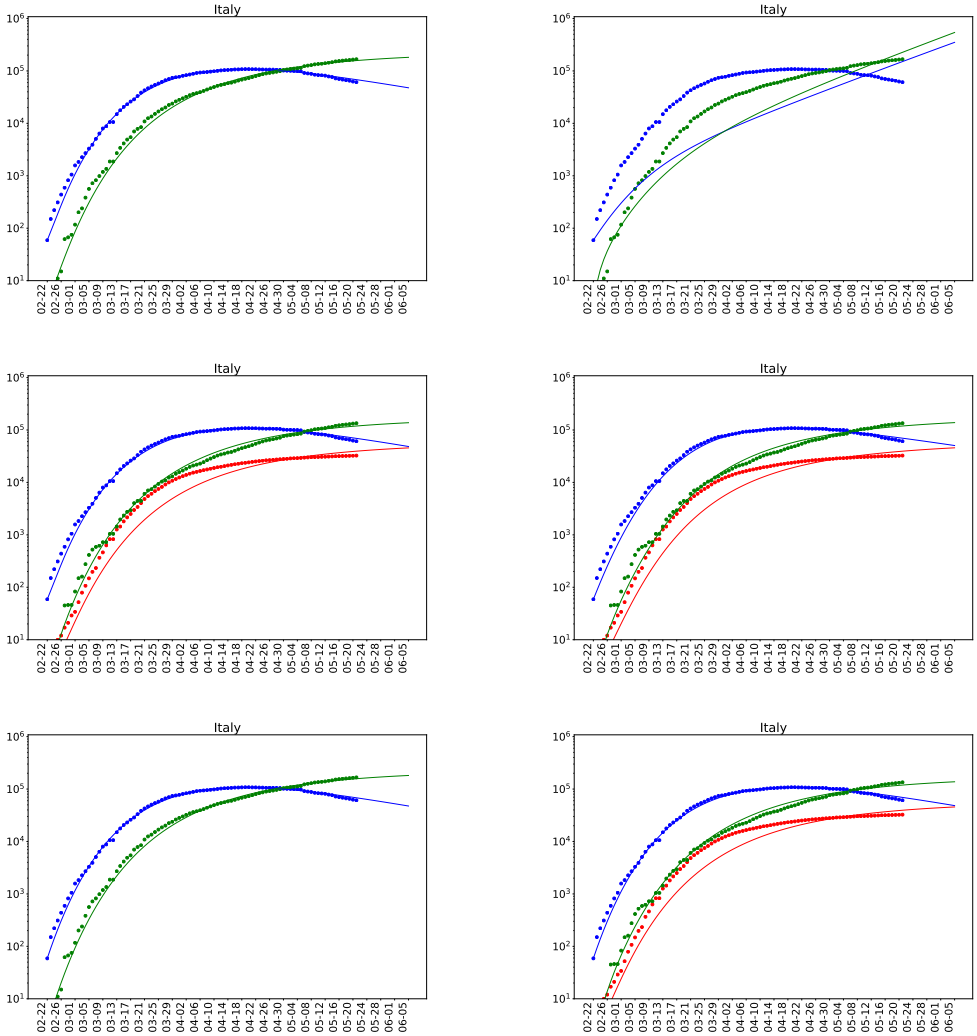

Figure 17: Fitting models with data for Italy. The panel from the top left to the bottom right are for the Models (1) to Model (6) respectively.

6 Summary and conclusions

Covid-19 is a global crisis and understanding its impact on different systems of the modern human life (medical, social, economical etc.,) and the responses presented is an important exercise to carry out. We understand that despite being a global phenomenon, the impact of Covid-19 in terms of the loss to life and the resourced being exhausted depend on the local conditions as well as on the mitigation measures taken locally. However, we believe that the global picture of the crisis does help to plan and take policy decisions at the local scale also.

Full understanding of any pandemic, in particular like Covid-19 which does not have any other examples in the history (in terms of the scale and impact), may become available only when it is over and the facts and figure presented here may have very short life. However, we still believe that any quick timely insight may help a lot in terms of the planning for the worse. Knowing very well that all the mathematical models are wrong but some are useful, we believe that mathematical models which are presented in this work may help to develop some insight about the crisis. A brief summary of the work presented here is as follows.

In §1 we have given a very brief introduction of the problem being addressed and reviewed some of the key works about Covid-19 which motivated the present work. A brief introduction of the mathematical framework used in the work in §2, in particular we have review a set of compartmental models SIR, SEIR and SIRD in §2. We have also discussed a set of of parametric models for one of the transmission coefficients $\beta(t)$, in §2.4. We have discussed the data being used in the work in §3.

The main results of the present work are discussed in §4 and §5. In §4 we have reviewed a reconstruction procedure for the transmission coefficients and basic reproduction number $\mathfrak{R}(t)$. This procedure does not depend on the choice of any parameter and can be easily generalized for other similar models also. We have presented the best-fit values of the parameters with their 90% CI, §5 in the form a set of tables. We have presented the values of the parameters in the following two forms:

1. Model based

2. Country based

All the fitting parameters for the models being considered are summarized in Table (**??**). The country based parameters are given in terms of a set of two tables (6) where every row represents a country. From the country based table we can see that the estimate are quite consistent - do not change much from one model to another. In place of presenting all the parameters for countries we present only the basic reproduction number corresponding to different models.

We present Table (2) which in the appendix which have some basic information about the countries being considered for the modeling. In order to see how different models fits the data we have given the plots for all the six models for a set of countries in the figures from (**??**).

The work we presented here assumes that spreading of a pandemic like Covid-19 happens homogeneously in space and time, however, we know that it is far from true. As the experience [ea03b] shows that "super-spread" events (SSEs) or rare events where, one particular infectious person interacts with a very large number of susceptible people over a short period of time have the maximum impact. In these situation the average measures like $\mathfrak{R}_0$ are not very informative. In the present work we data for a set of countries to constrain the parameters of the SIR and SIRD model one similar exercise with SIRD model for India is done in [ea20q].

Acknowledgment

The author would like to thank Dr. Gaurav Goswami for comments and feedback. At present the author works as an independent researcher and data scientist and the work presented here is not supported by any public or private agency. The author will be thankful to any agency, individual or individuals who come forward to sponsor/support this and other similar works on Covid-19.

References

[BC18] J. Blackwood and L. M. Childs. An introduction to compartmental modeling for the budding infectious disease modeler. *Letters in Biomathematics*, 5:1:195–221, 2018. doi:10.1080/23737867.2018.1509026.

[Cob20] S. Cobey. Modeling infectious disease dynamics. *Science*, 2020. doi:10.1126/science.abb5659 (2020).

[DG20] E. Daddi and M. Giavalisco. Early forecasts of the evolution of the covid-19 outbreaks and quantitative assessment of the effectiveness of countering measures. *ArXiv e-prints*, 2020, 2004.08365.

[ea97] C. A. G. et al. Analysis and fitting of an sir model with host response to infection load for a plant disease. *Philos Trans R Soc Lond B Biol Sci.*, 352(1351):353–364, 1997. doi:10.1098/rstb.1997.0026.

[ea03a] L. et al. Transmission dynamics and control of severe acute respiratory syndrome. *Science*, 300(5627):1966–1970, 2003. doi:10.1126/science.1086616.

[ea03b] R. et al. Transmission dynamics of the etiological agent of sars in hong kong: Impact of public health interventions. *Science*, 300:1961–1966, 2003. doi:10.1126/science.1086478.

[ea05] N. M. F. et al. Strategies for containing an emerging influenza pandemic in southeast asia. *Nature*, 437:209–214, 2005. doi:10.1038/nature04017.

[ea06] N. M. F. et al. Strategies for mitigating an influenza pandemic. *Nature*, 442:448–452, 2006. doi:10.1038/nature04795.

[ea20a] A. et al. Data-based analysis, modelling and forecasting of the covid-19 outbreak. *PLOS ONE*, 2020. doi:10.1371/journal.pone.0230405.

[ea20b] A. et al. Accounting for symptomatic and asymptomatic in a seir-type model of covid-19. *ArXiv e-prints*, 2020, 2004.01805.

[ea20c] C. et al. Assessing the efficiency of different control strategies for the coronavirus (covid-19) epidemic. *ArXiv e-prints*, 2020, 2004.03539.

[ea20d] C. et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in wuhan, china: a descriptive study. *The Lancet*, 395:507–513, 2020. doi:10.1016/S0140-6736(20)30211-7.

[ea20e] C. et al. A time-dependent sir model for covid-19 with undetectable infected persons. *ArXiv e-prints*, 2020, 2003.00122.

[ea20f] C. et al. A strategic approach to covid-19 vaccine r&d. *Science*, 2020. doi:10.1126/science.abc5312.

[ea20g] D. et al. An interactive web-based dashboard to track covid-19 in real time. *The Lancet*, 20:533–534, 2020. doi:10.1016/S1473-3099(20)30120-1.

[ea20h] G. et al. The species severe acute respiratory syndrome-related coronavirus: classifying 2019-ncov and naming it sars-cov-2. *Nat Microbiology*, 5:536–544, 2020. doi:10.1.1/jpb001.

[ea20i] H. et al. Monitoring the spread of covid-19 by estimating reproduction numbers over time. *ArXiv e-prints*, 2020, 2004.08557.

[ea20j] J. D. et al. Inferring change points in the spread of covid-19 reveals the effectiveness of interventions. *Science*, 2020. doi:10.1126/science.abb9789.

[ea20k] K. et al. Early dynamics of transmission and control of covid-19: a mathematical modelling study. *The Lancet*, 20(5):553–558, 2020. doi:10.1016/S1473-3099(20)30144-4.

[ea20l] L. et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (sars-cov-2). *Science*, 368(6490):489–493, 2020. doi:10.1126/science.abb3221.

[ea20m] L. et al. Preliminary analysis of covid-19 spread in italy with an adaptive seird model. *ArXiv e-prints*, 2020, 2003.09909.

[ea20n] L. et al. A modified seir model to predict the covid-19 outbreak in spain and italy: simulating control scenarios and multi-scale epidemics. *MedRxiv e-prints*, 2020, 2003.09909. doi:10.1101/2020.03.27.20045005.

[ea20o] P. et al. A data driven analysis and forecast of an seiard epidemic model for covid-19 in mexico. *ArXiv e-prints*, 2020, 2004.08288.

[ea20p] P. et al. The effect of control strategies to reduce social mixing on outcomes of the covid-19 epidemic in wuhan, china: a modelling study. *The Lancet*, 5:261–270, 2020. doi:10.1016/S2468-2667(20)30073-6.

[ea20q] S. C. et al. Studying the progress of covid-19 outbreak in india using sird model. *medRxiv*, 2020. doi:10.1101/2020.05.

[ea20r] S. et al. Covid-19: Epidemiology, evolution, and cross-disciplinary perspectives. *Trends in Molecular Medicine*, 26(5):483–495, 2020. doi:10.1016/j.molmed.2020.02.008.

[ea20s] W. et al. The global impact of covid-19 and strategies for mitigation and suppression. *Imperial College COVID-19 Response Team*, 2020. doi:10.25561/77735.

[ea20t] W. et al. A new coronavirus associated with human respiratory disease in china. *Nature*, 579:270–273, 2020. doi:10.1038/s41586-020-2012-7.

[ea20u] Y. et al. Modified seir and ai prediction of the epidemics trend of covid-19 in china under public health interventions. *Journal of thoracic disease*, 12(3):165–174, 2020. doi:10.21037/jtd.2020.02.64.

[ea20v] Z. et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, 579:270–273, 2020. doi:10.1038/s41586-020-2012-7.

[FP20] D. Fanelli and F. Piazza. Analysis and forecast of covid-19 spreading in china, italy and france. *Chaos, Solitons and Fractals*, 134:109761, 2020, 2003.06031. doi:10.1016/j.chaos.2020.109761.

[Gar20a] L. Gardner. Mapping 2019-ncov. `https://systems.jhu.edu/research/public-health/ncov/`, 2020.

[Gar20b] L. Gardner. Novel coronavirus (covid-19) cases. `https://github.com/CSSEGISandData/COVID-19`, 2020.

[GG20] M. D. Gaurav Goswami, Jayanti Prasad. Extracting the effective contact rate of covid-19 pandemic. *ArXiv e-prints*, 2020, 2004.07750.

[Het00] H. W. Hethcote. The mathematics of infectious diseases. *SIAM Review*, 42(4):599–653, 2000.

[iD20] O. W. in Data. Data on covid-19 (coronavirus) confirmed cases, deaths, and tests. `https://github.com/owid/covid-19-data`, 2020.

[KB20] P. S. Kathakali Biswas. Space-time dependence of corona virus (covid-19) outbreak. *ArXiv e-prints*, 2020, 2003.03149.

[KR08] M. J. Keeling and P. Rohani. *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press, 2008.

[Li18] M. Y. Li. *An Introduction to Mathematical Modeling of Infectious Diseases*. Springer International Publishing, 2018.

[NC08] G. NC and F. C. Mathematical models of infectious disease transmission. *Nat Rev Microbiol.*, 6:477–87, 2008. doi:10.1038/nrmicro1845.

[Oli20] G. Oliveira. Refined compartmental models, asymptomatic carriers and covid-19. *ArXiv e-prints*, 2020, 2004.14780. doi:10.1101/2020.04.14.20065128.

[Org20a] W. H. Organization. Novel coronavirus (2019-ncov) situation report - 1. `https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports`, 2020.

[Org20b] W. H. Organization. Report of the who-china joint mission on coronavirus disease 2019 (covid-19). `https://www.who.int/publications-detail/report-of-the-who-china-joint-mission-on-coronavirus-disease-2019-(covid-19)`, 2020.

[P.20] T. P. A time-dependent seir model to analyse the evolution of the sars-covid-2 epidemic outbreak in portugal. *Bull World Health Organ*, 2020. doi:10.2471/BLT.20.258392.

[Pra20] J. Prasad. A python package for fitting covid-19 data. `https://github.com/jayanti-prasad/PyCov19`, 2020.

[RMI20] P. M. Robert Marsland III. Data-driven modeling reveals a universal dynamic underlying the covid-19 pandemic under social distancing. *ArXiv e-prints*, 2020, 2004.10666.

[RR20] S. R. and A. R. Age-structured impact of social distancing on the covid-19 epidemic in india. *ArXiv e-prints*, 2020, 2003.12055.

[Sci20] SciPy. Scientific computing tools for python. `https://www.scipy.org/`, 2020.

[TS20] J. C. Tridip Sardar, Sk Shahid Nadim. Assessment of 21 days lockdown effect in some states and overall india: A predictive mathematical study on covid-19 outbreak. *ArXiv e-prints*, 2020, 2004.03487.

[Vil20] Villaverde. Estimating and simulating a sird model of covid-19 for many countries, states, and cities. `https://cepr.org/active/publications/discussion_papers/dp.php?dpno=14711`, 2020.

[Wor20] Worldometers. Covid-19 coronavirus pandemic. `https://www.worldometers.info/coronavirus/`, 2020.