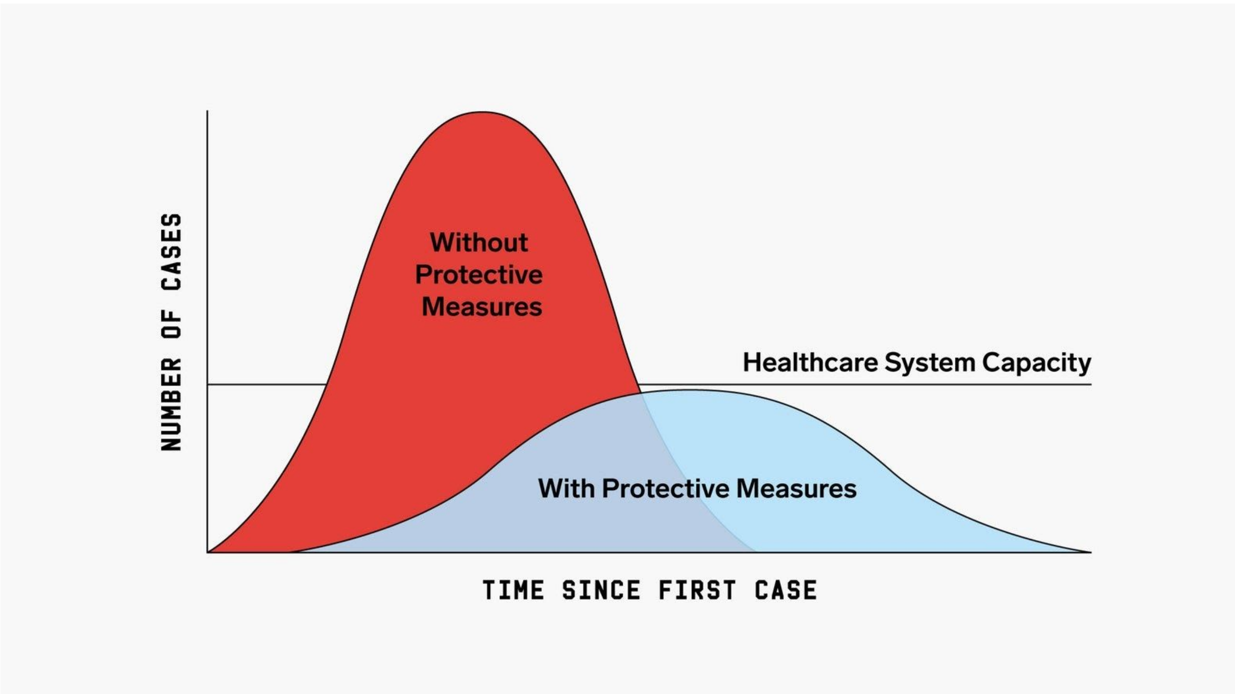
INTRO TO NEURAL COGNITIVE MODELING

COVID-19 ANALYSIS AND MODELING

Gauravdeep Singh BINDRA(2018201027) Rajat DUA(2018201066)



INTRODUCTION

Covid-19 disease (also called Coronavirus disease) is an infectious disease which was first identified in Wuhan in december 2019 had caused an ongoing pandemic around the world. It is caused by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2)

At the time of writing there have been 35,07,265 confirmed cases worldwide out of which 11,27,887 have recovered and 2,47,491 have died. (7.05% fatality rate)

India has done a good job of containing it till now with tough and timely measures like total nationwide lockdown. This lockdown has given the time to the center and state governments to make designated covid-19 hospitals and procure testing kits and PPE's. It has also avoided exponential growth of cases which leads to overflowing hospitals and increased fatality rate. India currently has 42,533 with 3.22% fatality rate.

Governments all over the world have been using mathematical models to predict the growth in the number of cases so that they can take decisions regarding lockdown and plan their testing strategy accordingly.

Compartmental Models are popular to effectively model infectious diseases. These models are represented through Differential Equations. SIR is a basic compartmental model. We have used the SIRD model.

We have also analysed the data, grouped countries on the basis of growth factor and performed S-R Trend analysis.

PREVIOUS WORK

We reference the research paper "Prediction of COVID-19 Disease Progression in India" by "Sourish Das" that was given by Sir.

Purpose

In this paper, the author develops an epidemiological SIR model and statistical machine learning model to predict disease progression in India. He implemented the SIR model to estimate the basic reproduction number R0 at the national and state level to identify which states require more attention. He also implemented the machine learning model to predict the number of cases ahead of time so Indian administration can be better prepared ahead of time.

Methodology Used

SIR Epidemiological Model

It is the Susceptible, Infected, Recovered (SIR) model.Let us define R0 as Basic Reproduction Number which tells the rate at which contagious people will infect other R0 number of people.

The SIR model can be described as,

$$\partial S/\partial t = -\beta\ SI/N$$

$$\partial I/\partial t = +\beta\ SI/N - \gamma I$$

$$\partial R/\partial t = +\gamma I$$

Where S,I,R denote the number of people in the population that are susceptible, infected and recovered and β is the transmission rate.

Each susceptible person contacts β people per day; a fraction I N of which are infectious. Therefore β SI/N moves out of the susceptible group and goes into the infected group. The parameter γ is the recovery rate, and γI is the flow out of the infected crowd and goes into the recovered group. The average duration a person spends in the infected group is 1/γ days. For Covid-19, 1/γ is around 14 days.

Statistical Machine Learning Model(SML)

Rate is per 100,000

Rate = Cases Population /Size × 100, 000.

Model the Rate as a function of time, country and time-country interaction in the following way:

$$ln\{Rate_{it} + 1\} = \beta_0 + \beta_1 t + \beta_2 t^2 + \cdots + \beta_p t^p\ + \alpha_i + \alpha_i t + \alpha_i t^2 + \cdots + \alpha_i t^p$$

where $Rate_{it}$ is the Rate of the i^{th} country at the t^{th} time point, α_i is the effect of i^{th} country, $\alpha_i t$ is the linear effect of time on the Rate of the i^{th} country, $\alpha_i t^2$ is the quadratic effect of time on the Rate of the i^{th} country.

Implementation

$R_0 = \beta/\gamma$

The R-package called, 'R0', was used to estimate R_0

Then for an estimated R_0 and γ (assumed to be 1/14), the disease progression, for the period, was simulated and new incidences were observed. Then the Mean Square Error was calculated.

$$MSE(\mu, \kappa) = \frac{1}{T}\sum_{t=1}^{T}\left(\hat{I}(t) - i_{obs}(t)\right)^2$$

Prediction

If lockdown works then actual confirmed cases for India should stay below 66,224 by May 01,2020.A Comparison of R0 between India and China The R0 for India for the first 22 days till the lockdown is around 2.5, like China. However, if we use the data, till 04-Apr-2020, then the R0 value is around 2.75. It indicates since the lockdown the situation has worsened.
State wise R0:
Madhya Pradesh (3.37) , Maharashtra (3.25), Tamil Nadu (3.09), Andhra Pradesh (2.96), Delhi (2.82) and West Bengal (2.77)

METHODS

SIRD MODEL

The SIRD model divides a fixed population of N individuals into 4 categories.

- S : Number of people who are Susceptible
- I : Number of infected individuals
- R : Number of people who have recovered
- D : Number of deaths

SIRD model assumes that a person once infected gains immunity on recovery and cant get infected again. But for novel coronavirus, the time period the immunity lasts hasn't been established yet.

We have chosen SIRD instead of SIR as SIR assumes that all infected people recover. That is not applicable to COVID-19 as it has a significant fatality rate.

Model:

$$S \xrightarrow{\beta I} I \xrightarrow{\gamma} R$$
$$I \xrightarrow{\alpha} D$$

α: Mortality rate [1/min]
β: Effective contact rate [1/min]
γ: Recovery rate [1/min]

Ordinary Differential Equation (ODE):

$$\frac{\mathrm{d}S}{\mathrm{d}T} = -N^{-1}\beta SI$$
$$\frac{\mathrm{d}I}{\mathrm{d}T} = N^{-1}\beta SI - (\gamma + \alpha)I$$
$$\frac{\mathrm{d}R}{\mathrm{d}T} = \gamma I$$
$$\frac{\mathrm{d}D}{\mathrm{d}T} = \alpha I$$

Where N = S + I + R + D

T : time elapsed from the start date

Non-dimensional Model

Set $(S, I, R, D) = N \times (x, y, z, w)$ and $(T, \alpha, \beta, \gamma) = (\tau t, \tau^{-1}\kappa, \tau^{-1}\rho, \tau^{-1}\sigma)$.
This results in the ODE

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -\rho xy$$
$$\frac{\mathrm{d}y}{\mathrm{d}t} = \rho xy - (\sigma + \kappa)y$$
$$\frac{\mathrm{d}z}{\mathrm{d}t} = \sigma y$$
$$\frac{\mathrm{d}w}{\mathrm{d}t} = \kappa y$$

N : Total Population

Ω : $coefficient(integer\ to\ simplify)$

The range of variables and parameters:

$$0 \leq (x, y, z, w, \kappa, \rho, \sigma) \leq 1$$
$$1 \leq \tau \leq 1440$$

Reproduction number can be defined as

$$R_0 = \rho(\sigma + \kappa)^{-1} = \beta(\gamma + \alpha)^{-1}$$

Example

For example, set $R_0 = 2.5, \kappa = 0.005, \rho = 0.2$ and initial values $(x_{(0)}, y_{(0)}, z_{(0)}, w_{(0)}) = (0.999, 0.001, 0, 0)$.

SIRF Model

As it is being reported that some cases died before the clininal diagonis, therefore we need to allow for movement directly from confirmed to fatal.

SIRF model divides the population into 5 categories:

- S : Susceptible
- S*: Confirmed but still uncategorized
- I : Confirmed and categorized as Infected
- R : Recovered
- F : Fatal

Measurement Variables

Confirmed = I + R + F

Model

$$
\begin{aligned}
\mathrm{S} \xrightarrow{\beta I} \mathrm{S}^* &\xrightarrow{\alpha_1} \mathrm{F} \\
\mathrm{S}^* &\xrightarrow{1-\alpha_1} \mathrm{I} \xrightarrow{\gamma} \mathrm{R} \\
& \qquad \mathrm{I} \xrightarrow{\alpha_2} \mathrm{F}
\end{aligned}
$$

α_1 : Mortality rate of S^* cases [-]
α_2 : Mortality rate of I cases [1/min]
β: Effective contact rate [1/min]
γ: Recovery rate [1/min]

Note: When $\alpha_1 = 0$, SIR-F model is the same as SIR-D model.

Ordinary Differential Equation (ODE):

$$
\begin{aligned}
\frac{\mathrm{d}S}{\mathrm{d}T} &= -N^{-1}\beta SI \\
\frac{\mathrm{d}I}{\mathrm{d}T} &= N^{-1}(1-\alpha_1)\beta SI - (\gamma + \alpha_2)I \\
\frac{\mathrm{d}R}{\mathrm{d}T} &= \gamma I \\
\frac{\mathrm{d}F}{\mathrm{d}T} &= N^{-1}\alpha_1 \beta SI + \alpha_2 I
\end{aligned}
$$

Where N = S + I + R + F, T = Time elapsed from the start date

Non-dimensional SIRF Model

Set $(S, I, R, F) = N \times (x, y, z, w)$ and $(T, \alpha_1, \alpha_2, \beta, \gamma) = (\tau t, \theta, \tau^{-1}\kappa, \tau^{-1}\rho, \tau^{-1}\sigma)$.
This results in the ODE

$$
\begin{aligned}
\frac{\mathrm{d}x}{\mathrm{d}t} &= -\rho xy \\
\frac{\mathrm{d}y}{\mathrm{d}t} &= \rho(1-\theta)xy - (\sigma + \kappa)y \\
\frac{\mathrm{d}z}{\mathrm{d}t} &= \sigma y \\
\frac{\mathrm{d}w}{\mathrm{d}t} &= \rho\theta xy + \kappa y
\end{aligned}
$$

Where N is the total population and τ is a coefficient ([min], is an integer to simplify).

The range of variables and parameters:

$$0 \leq (x, y, z, w, \theta, \kappa, \rho, \sigma) \leq 1$$
$$1 \leq \tau \leq 1440$$

Reproduction number can be defined as

$$R_0 = \rho(1-\theta)(\sigma+\kappa)^{-1} = \beta(1-\alpha_1)(\gamma+\alpha_2)^{-1}$$

Example

For example, set $R_0 = 2.5, \theta = 0.002, \kappa = 0.005, \rho = 0.2$ and initial values $(x_{(0)}, y_{(0)}, z_{(0)}, w_{(0)}) = (0.999, 0.001, 0, 0)$.

SR TREND ANALYSIS

Based on the relationship b/w S(Susceptible) and R(Recovered)

In SIRF Model

$$\frac{\mathrm{d}S}{\mathrm{d}T} = -\frac{\beta}{N} S I$$
$$\frac{\mathrm{d}R}{\mathrm{d}T} = \gamma I$$
$$I > 0$$
$$S \simeq N \text{ when } R = 0$$

Therefore,

$$\frac{\mathrm{d}S}{\mathrm{d}R} = -\frac{\beta}{N\gamma} S$$

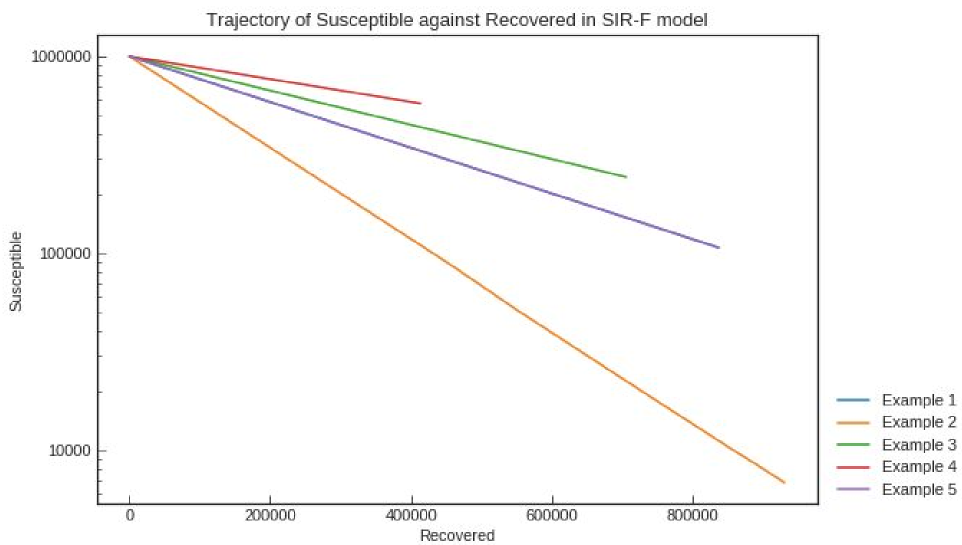
This leads to

$$S(R) = Ne^{-\frac{R\beta}{N\gamma}}$$

$$\log\left(S(R)\right) = \log\left(N\right) - \frac{R\beta}{N\gamma}$$

With constant $a = \frac{\beta}{N\gamma}$ and constant $b = \log N$,

$$\log S_{(R)} = -aR + b$$



GROWTH FACTOR

$$\text{Growth Factor} = \frac{\Delta C_n}{\Delta C_{n-1}}$$

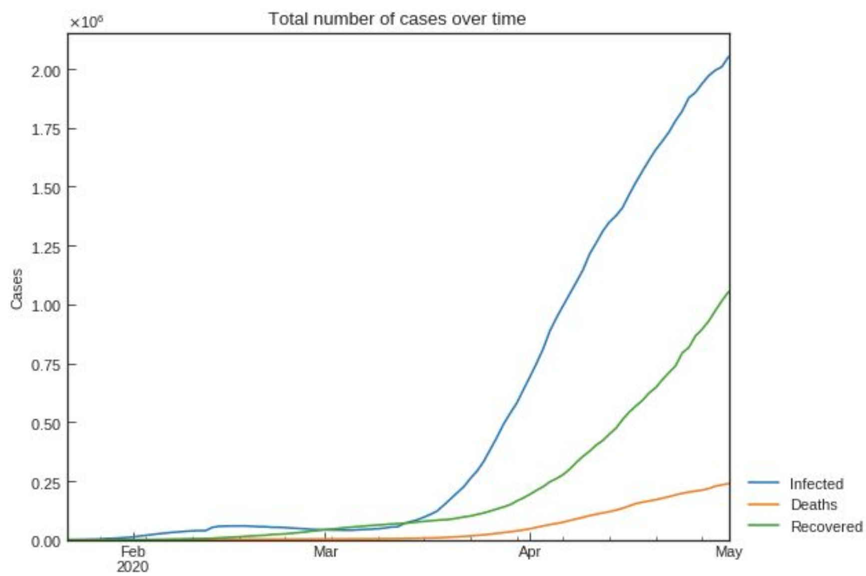
where C is the number of confirmed cases. We are grouping countries on the basis of growth factor into 3 categories.

1. Outbreaking : If the Growth Factor is >1 for the last 7 days.
2. Stopping: If the Growth Factor is <1 for the last 7 days.
3. At Crossroads: For other cases

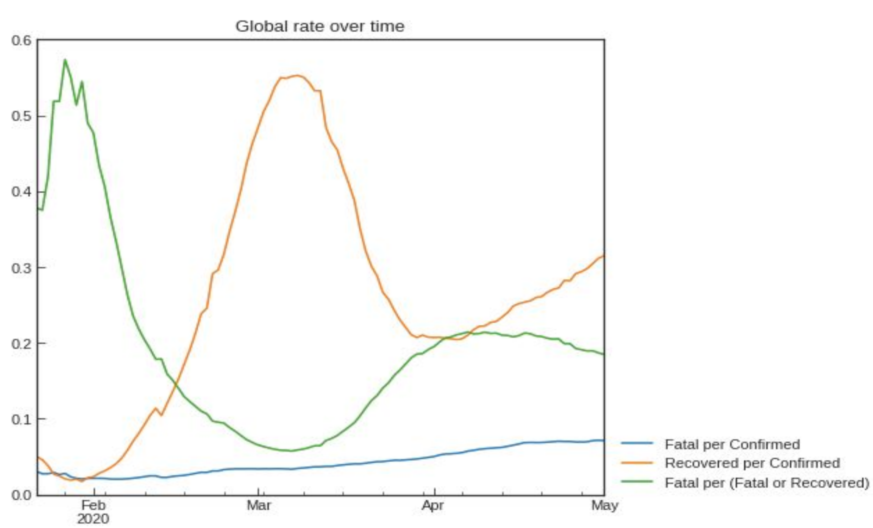
RESULTS

Visualizing Total Data

Worldwide Cases



Global Rate over Time



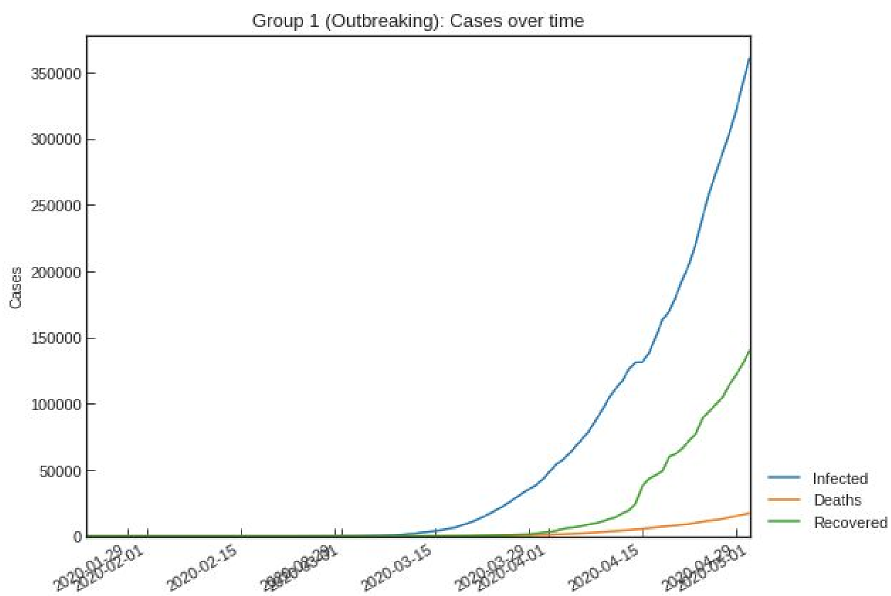
Growth Factor Analysis

List of outbreaking Countries

Out[76]:

```
'Mexico, Bahrain, Bulgaria, Estonia, Guatemala, Honduras, India, Latvia, Peru,
Qatar, Russia, Senegal, Afghanistan, Brazil, Kenya, Saudi Arabia, Bangladesh, A
rmenia, Mali, Albania, Burma, Congo, Ecuador, Dominican Republic, Egypt, El Sal
vador, South Africa, Montenegro, Pakistan, Rwanda, Cabo Verde, Luxembourg, Norw
ay, Kuwait, Jordan, Sierra Leone, Maldives, Azerbaijan, Mozambique, Gabon, Jama
ica, Malta, Costa Rica, Greece, Indonesia, Ukraine, Chile, Israel, Malawi, Mold
ova, North Macedonia, Sri Lanka, Uganda.'
```

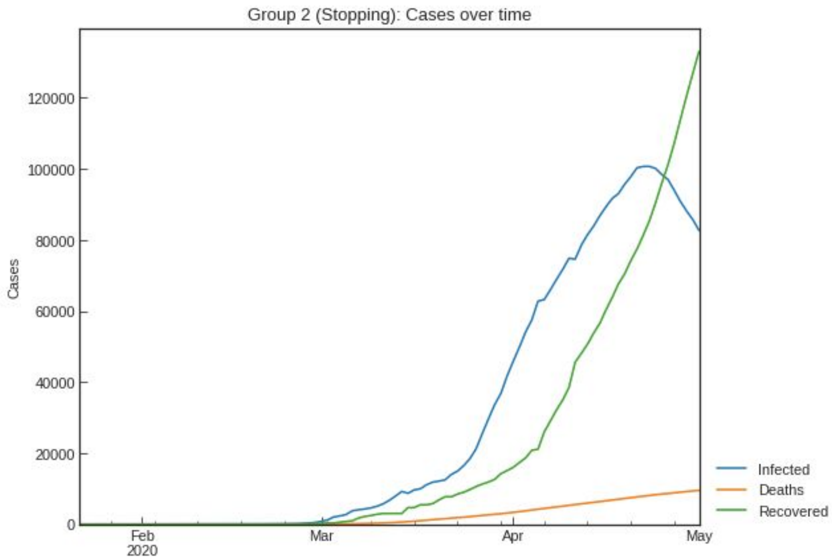
Cases in outreaking countries over time



List of Stopping Countries

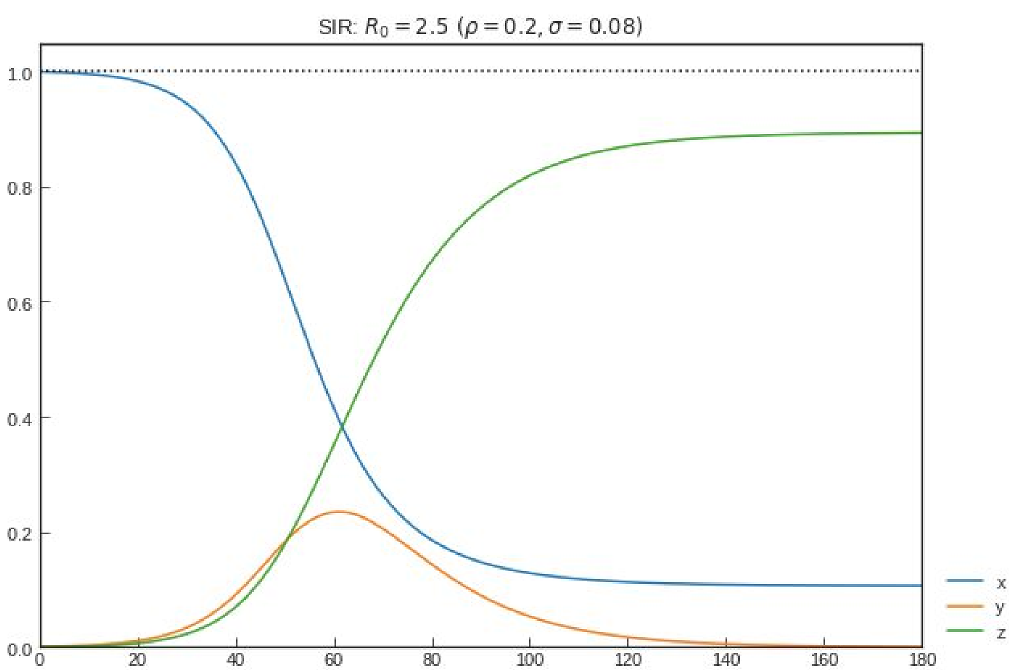
```
'Chad, Equatorial Guinea, Nicaragua, Syria, Zimbabwe, Botswana, Trinidad and To
bago, Cameroon, Andorra, Hong Kong, Nepal, Central African Republic, Iran, Papu
a New Guinea, Angola, Ghana, West Bank and Gaza, Liberia, Liechtenstein, Benin,
Libya, Bhutan, Turkey.'
```

Cases in Stopping Countries over time

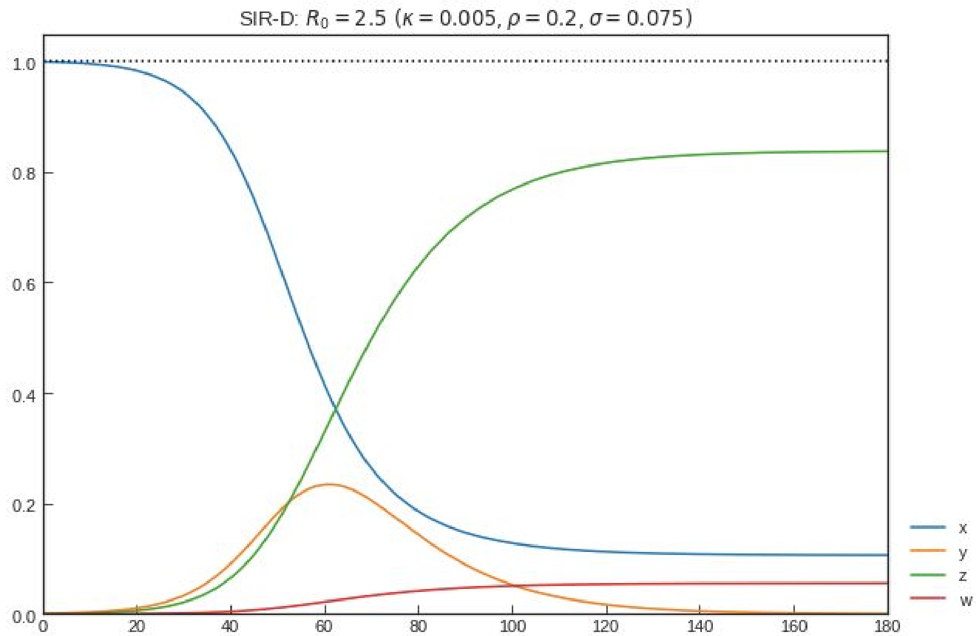


Example Runs

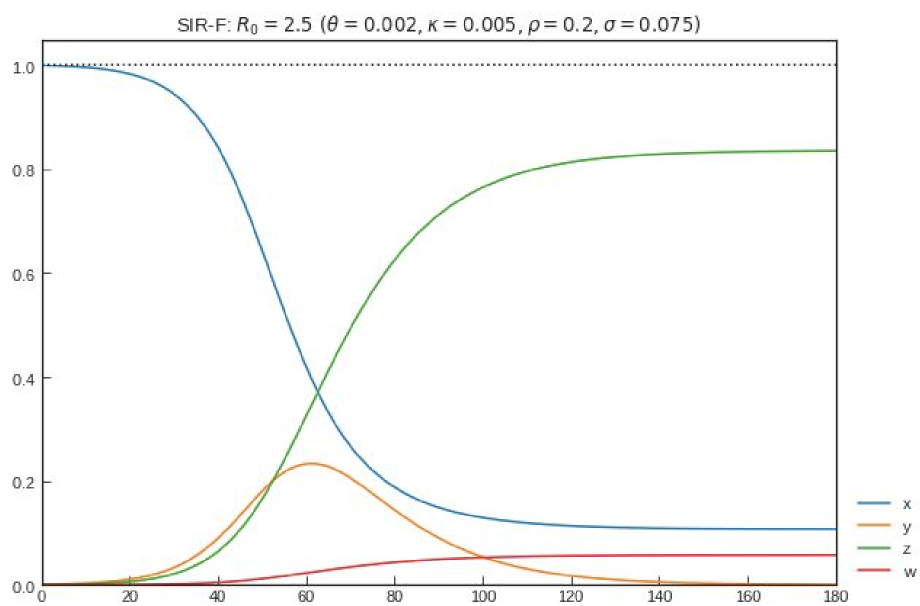
SIR Model



Example run of SIRD model



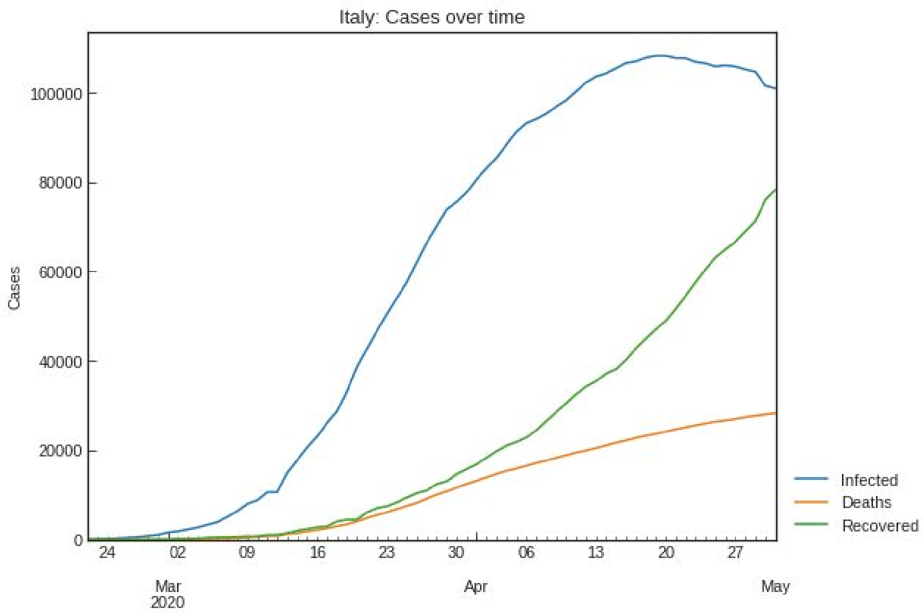
Example run of SIRF Model



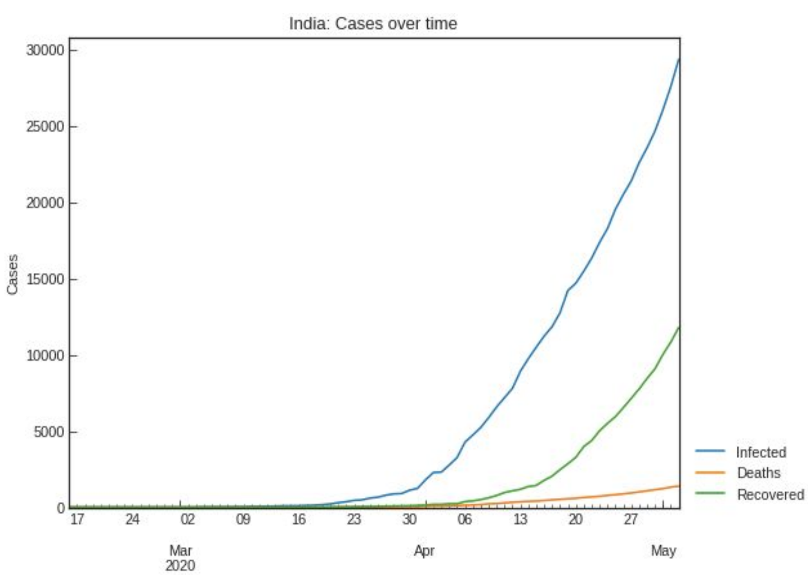
Comparison b/w Italy and India

1. Cases over time

Italy

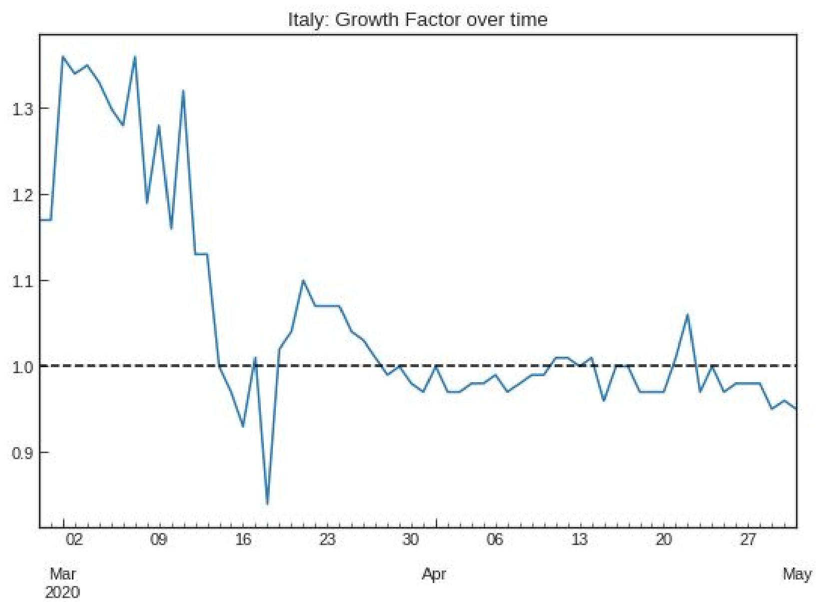


India

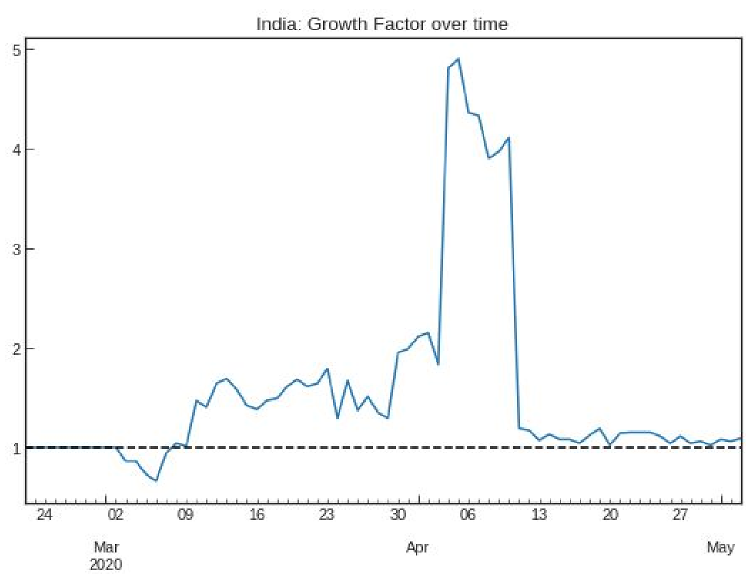


2. Growth Factor over time

Italy

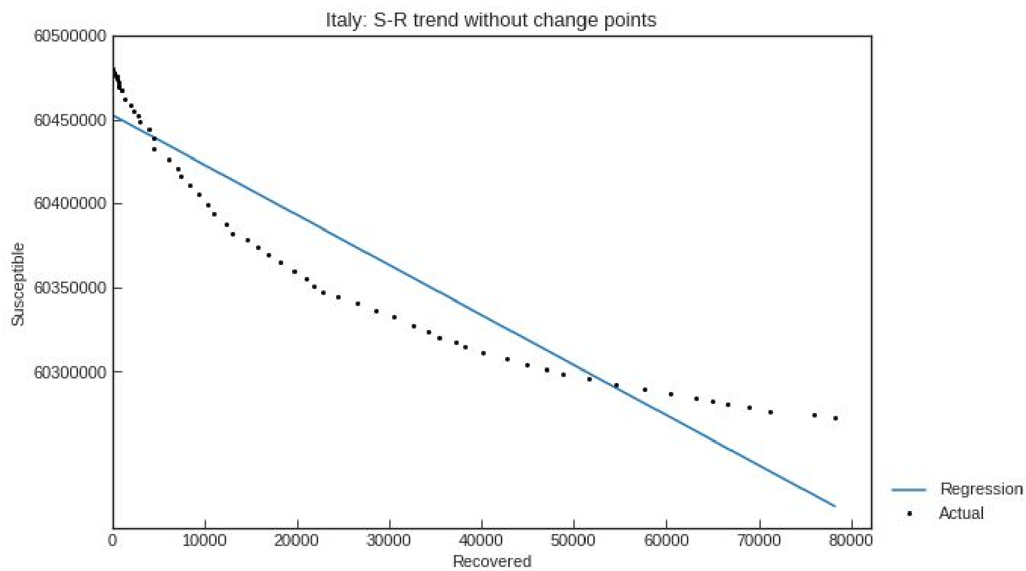


India

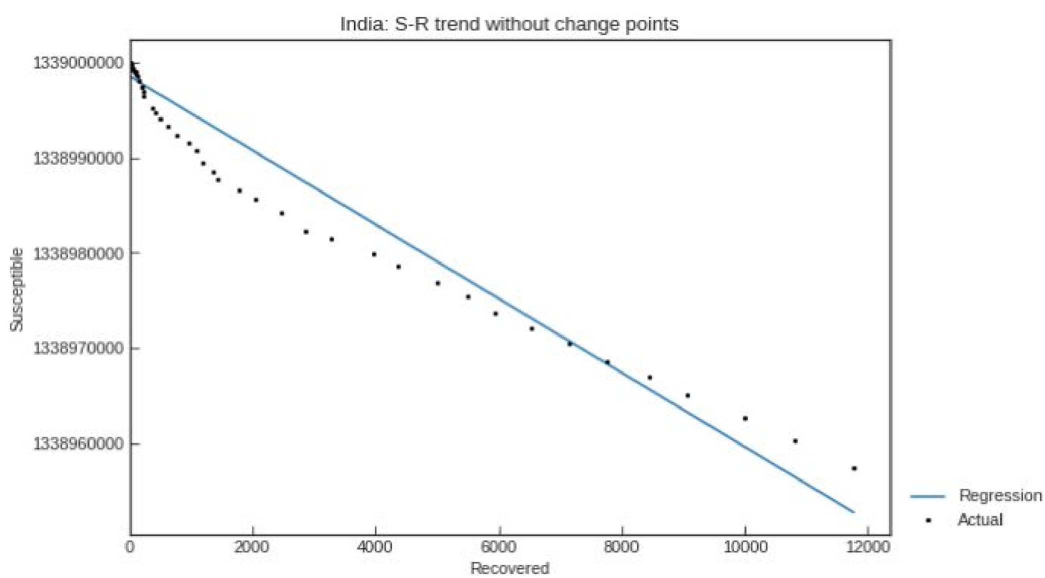


3. SR Trend Analyis

Italy

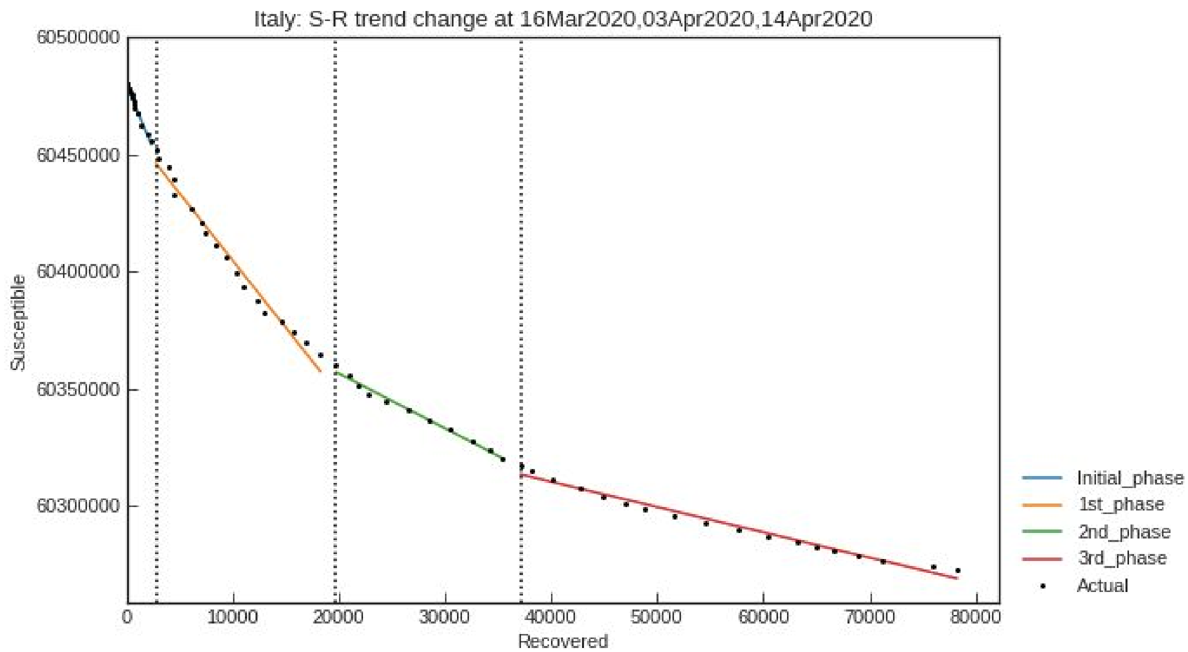


India

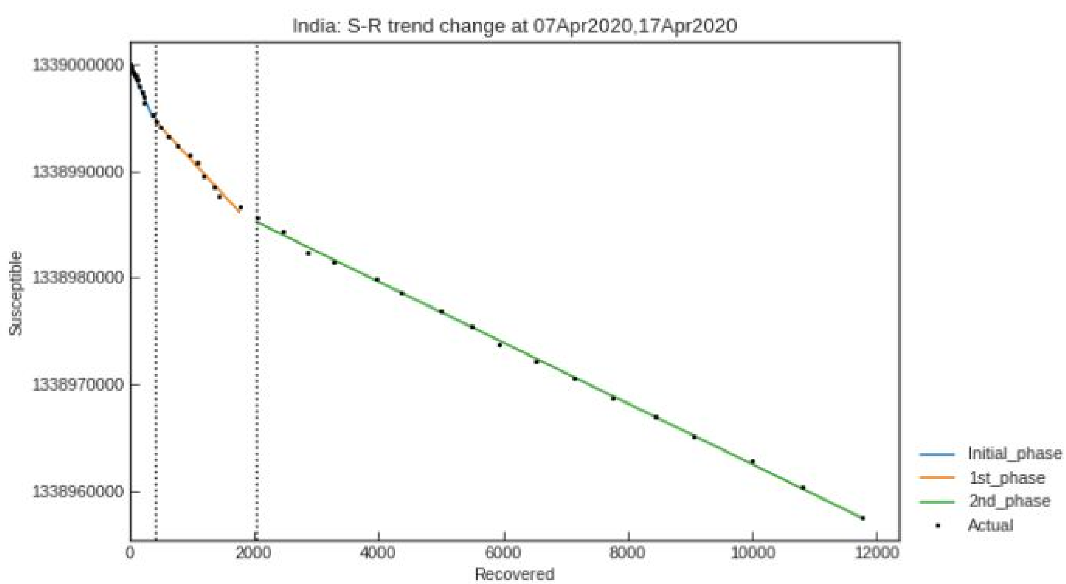


4. SR Trend Analysis (with change points)

Italy

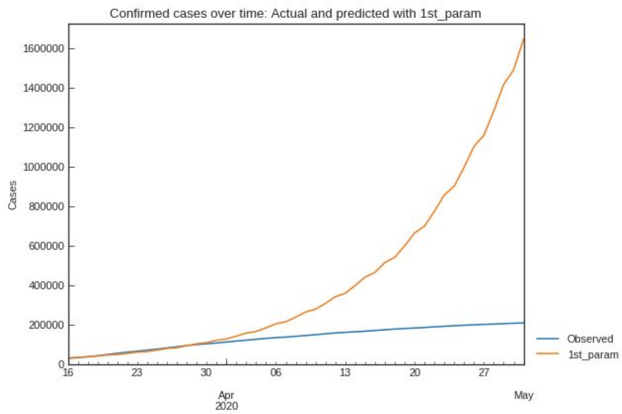


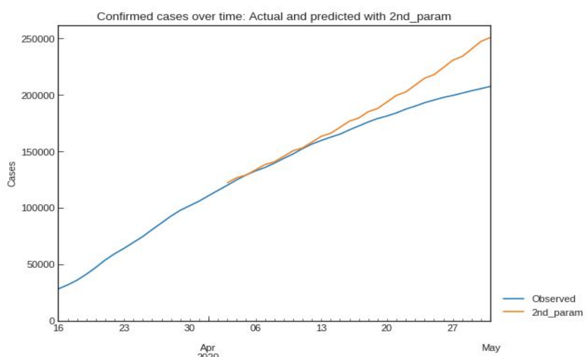
India



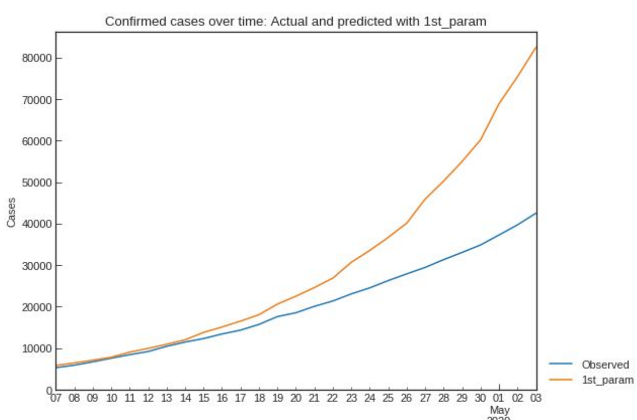
5. Phase Analysis

Italy





India



6. Paramters

Italy

	start_date	end_date	population	tau	theta	kappa	rho	sigma	Rt	score	alpha1 [-]	1/alpha2 [day]	1/beta [day]	1/gamma [day]
1st	17Mar2020	05Apr2020	60480000	1195	0.123718	0.000688	0.087324	0.016686	4.40	0.157248	0.124	1206	9	49
2nd	06Apr2020	11Apr2020	60480000	1195	0.142571	0.000243	0.036493	0.018510	1.67	0.011540	0.143	3408	22	44
3rd	12Apr2020	-	60480000	1195	0.144562	0.001075	0.025677	0.017515	1.18	0.027193	0.145	771	32	47

India

	start_date	end_date	population	tau	theta	kappa	rho	sigma	Rt	score	alpha1 [-]	1/alpha2 [day]	1/beta [day]	1/gamma [day]
1st	09Apr2020	15Apr2020	1339000000	1404	0.020438	0.001284	0.113989	0.018642	5.60	0.057279	0.020	759	8	52
2nd	16Apr2020	-	1339000000	1404	0.021594	0.000346	0.090090	0.028532	3.05	0.058634	0.022	2814	10	34

Comparison with past outbreaks

Table 1

Estimated Mean Values of $\mathscr{R}_0$ from Data.

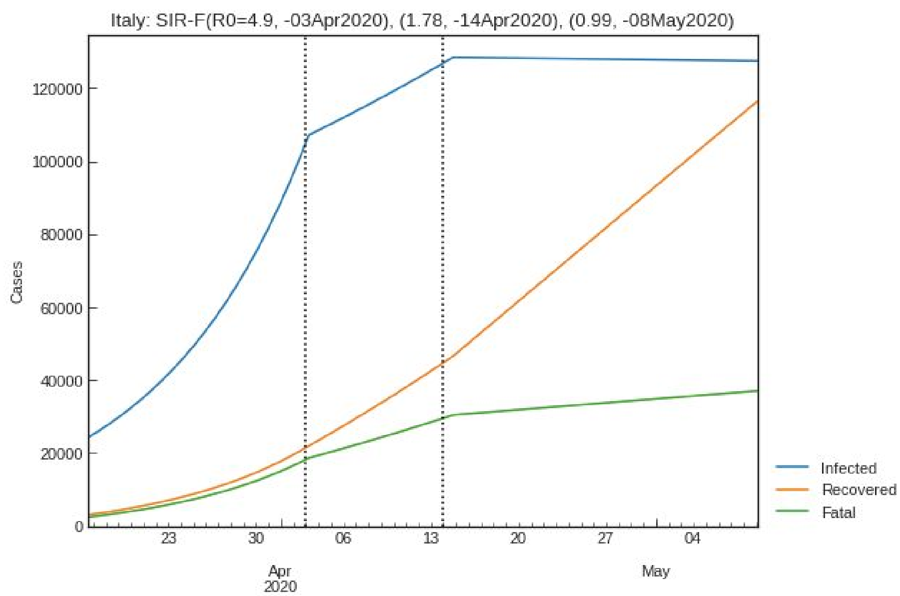
Disease outbreak and location	$\mathscr{R}_0$	Reference
Smallpox in Indian subcont. (1968–73)	4.5	Anderson and May (1991)
Poliomyelitis in Europe (1955–60)	6	Anderson and May (1991)
Measles in Ghana (1960–68)	14.5	Anderson and May (1991)
SARS epidemic in (2002–03)	3.5	Gumel et al. (2004)
1918 Spanish influenza in Geneva		
Spring wave	1.5	Chowell, Ammon, Hengartner, and Hyman (2006)
Fall wave	3.8	Chowell et al. (2006)
H2N2 influenza pandemic in US (1957)	1.68	Longini, Halloran, Nizam, and Yang (2004)
H1N1 influenza in South Africa (2009)	1.33	White, Archer, and Pagano (2013)
Ebola in Guinea (2014)	1.51	Althaus (2014)
Zika in South America (2015–16)	2.06	Gao et al. (2016)

(Source given in references)

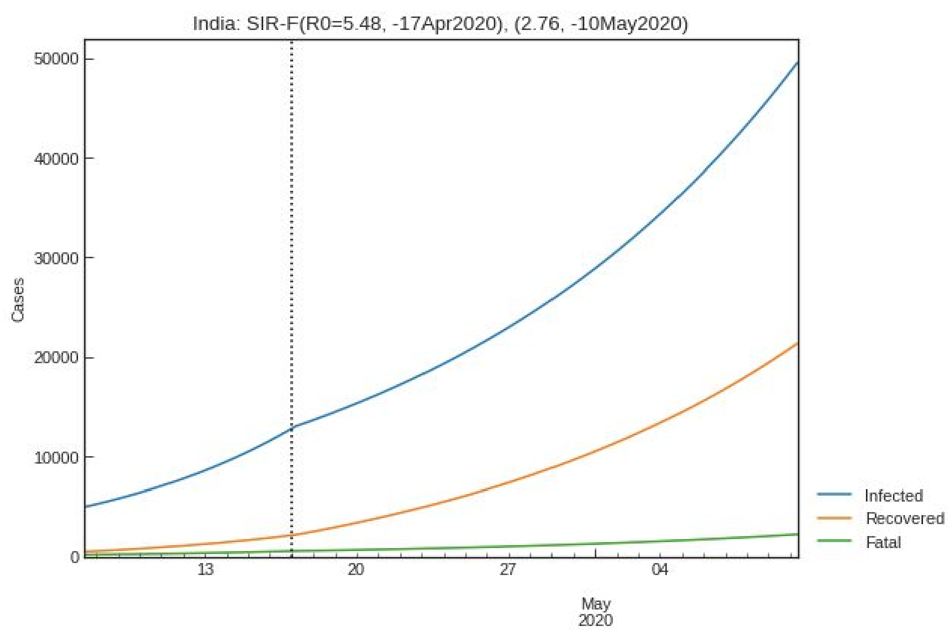
PREDICTIONS

For Next Week

Italy

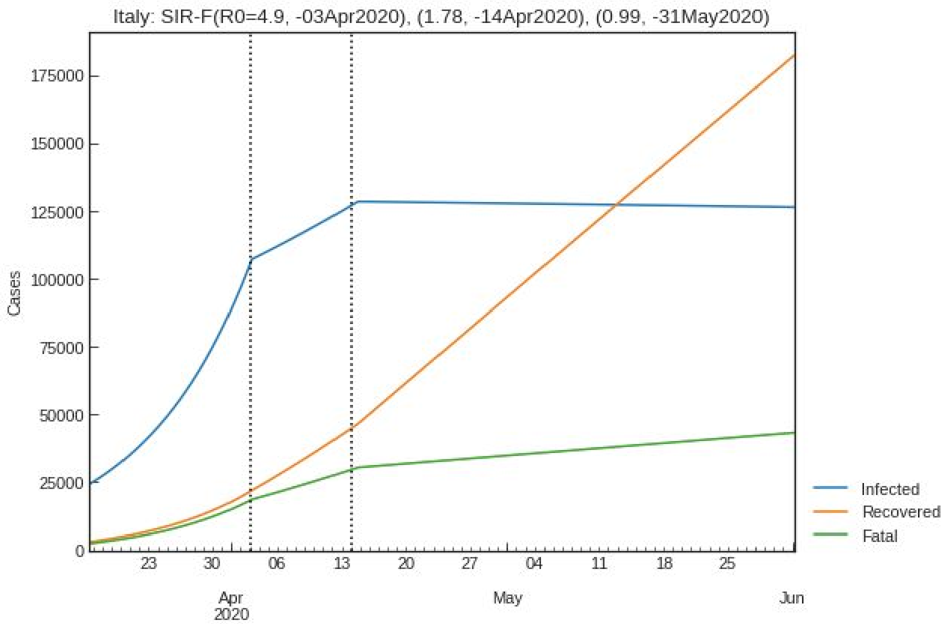


India

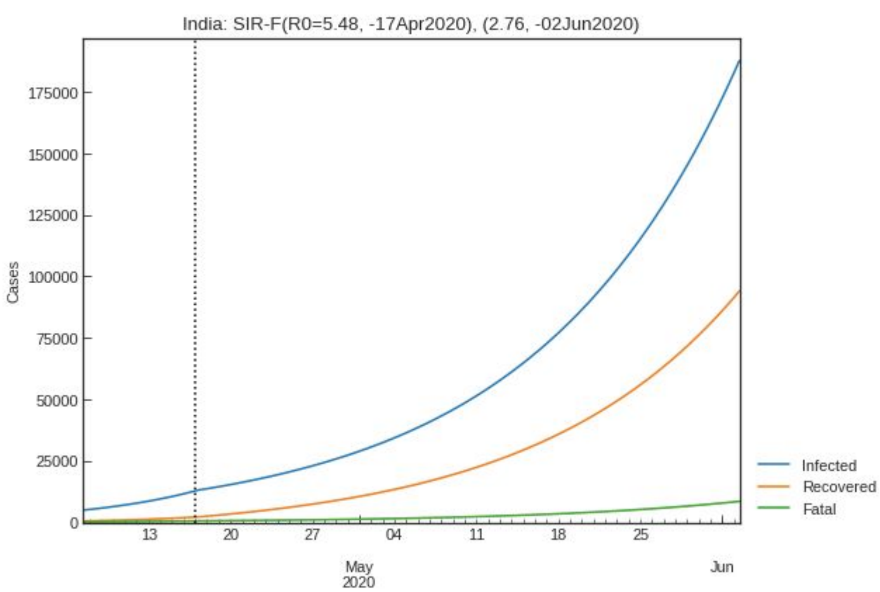


2. Next Month

Italy



India



REFERENCES

https://dc.etsu.edu/cgi/viewcontent.cgi?article=3102&context=etd

https://scipython.com/book/chapter-8-scipy/additional-examples/the-sir-epidemic-model

https://en.wikipedia.org/wiki/Compartmental_models_in_epidemiology

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6002118/

https://www.youtube.com/watch?v=Kas0tIxDvrg

https://towardsdatascience.com/infection-modeling-part-1-87e74645568a

Datasets Used

- Coronavirus Dataset : The number of cases :
  https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset
- Total Population :
  https://www.kaggle.com/dgrechka/covid19-global-forecasting-locations-population/metadata