

# Title: Multi-cluster models for epidemic spread and evaluation based on data-driven parameterization

Duration: 1 year

Submitted to SERB under Short-term MATRICS special call on Mathematical Modeling and Computations for COVID-19 Infections

Principal Investigator: Arzad Alam Kherani, IIT Bhilai.

Co-PI: Prof. D. Manjunath, Department of Electrical Engineering, IIT Bombay.

Co-PI: Dr. Rishi Ranjan Singh, IIT Bhilai.

## 1. Introduction

Establishing models that reflect the epidemiological transmission mechanism is required to provide insights into the spatio-temporal process of the infectious disease outbreaks. A good model greatly facilitates the design mechanisms to optimally control the spread through feasible interventions. The key difficulty in developing useful models is the complexity of the interplay between microscopic and macrodynamic evolution in networks. Abstract models incorporate varying degrees of details and are very helpful in providing a high-level understanding of the process. Many such models are well known and widely understood. These models typically, do not account for the differences between individuals, social classes, community behaviours, etc and rely on a key assumption of a uniform whole. Thus, these models need to be tailored significantly to be used 'in the field.'

Newer social mechanisms of disease control are expected to evolve now that the world population is trained for self-quarantine. Also, the behaviour of different *clusters* of society towards such social prevention mechanism requires an in-depth understanding of the definition of clusters themselves. Use of the term *cluster* in this proposal is conceptual at best at the time of writing the proposal and relates to the recent experiences of different behaviour observed across a diverse country like India. Terms like hot-spots may be related to our notion of *clusters*, but we feel that there is much more to the phenomenon than just geographically defined hot-spots.

In the Indian context, the approach will be to quantify the parameters of the model for different classifications of clusters. We will define and identify the clusters (social-economical, geographical, institutional, commercial, etc.) and use publicly available data to quantify the parameters for the model that we have proposed/analysed. We will be using an approach similar to that used in IndiaSIM and will provide the projections for infrastructure requirements (like, number of hospital beds to be prepared in advance, number of expected relevant medical equipment, etc.). We will also provide detailed analysis and scientific guidelines on the best way of utilization of resources (available medical personnel, hospital facilities, etc.) in the case of dynamic evolution of an epidemic.

## 2. Background and Motivation

A huge body of literature exists that provide study of dynamics of evolution of epidemics, and usually assume a homogeneous population. Most of these models study the effect of several preventive measures on a large population. Such large population averages are easier to model than estimating fine grained behavior, however, the latter are more important for study of effectiveness of intervention.

Standard SIR and variants (e.g., SEIR) can model a *homogeneous* population, but suffer from the following limitations:

- parametrizing ( $\beta$ ,  $\mu$ ,  $\gamma$ ) these models is a significantly non trivial exercise,
- statistics derived based on data collected from other societies don't easily translate and they are at best just a guide.

Demographics (population densities, income, age, living conditions), social behavior, public infrastructure, workplace, nutrition levels and infection-vulnerability, evolving virulence of the virus, climatic conditions, are just some of the factors that need to be considered in the parametrization of the models, exposure to TB/Malaria and vaccines,

In the wake of the ongoing COVID-19 outbreak, some India-specific and some general studies have become available, providing several predictions on the way the COVID-19 spread is expected to evolve. We summarise some of those in the following:

[1] is the widely cited and is popularised through regular appearances in the mainstream media. However, the modelling details of this system does not appear to be publicly available. It is too early (at the time of writing this proposal) to validate their

prediction of infection peak in late May, with significant hospitalizations in mid-May 2020, however the prediction of infection take-off around 10-15 April seems to be holding true.

A recent study reported in [2] has attracted significant attention. This reference assumes that contact rates depend on age demographics, predicts around 1 lakh infections around 15 April 2020, peaking to 180 Million cumulative infections round late June. We are yet to validate the predictions. However, what can be said for sure is that the real data available does not match with the dramatic change in numbers as forecast in the paper (the paper predicted a dramatic negative growth in number of infected individuals). We do not wish to hazard a guess on the reasons behind this failure in the proposal phase, and would be studying the possible reasons for this as more data becomes available (including, the efficiency and usage of testing, random vs targeted testing, variety of strains, etc.) We do acknowledge that the authors of [2] did not have the luxury of hindsight as we may have in a couple of months from now during the execution of this project.

[3] has reported genetic study of strains from Wuhan, US, India, Italy, Nepal and pointed out that Indian sequence has maximum number of mutations. They speculate that mutations will lead to variations in spread and mortality. Thus, a key takeaway for us is to be able to model concurrent presence of such mutations. Also see [4] and [5] for recent efforts towards understanding COVID-19 spread. There are of course many more available via Google Scholar that essentially reinterpret the standard SIR models with minor variations with the preliminary data and make strong assumptions and hence their strong predictions should be read with care.

### 3. Challenges and Observations

Several challenges obviously present themselves.

- We believe that there is a need to account for the differences in the transmission rates in the various stages of disease evolution and transmission models---symptomatic, asymptomatic, presymptomatic, and Environmental.
- The key factor in modeling the spread is to model the rate of spread. This requires us to understand colocation processes, population mobility, population mixing.

One of the key challenges is to find mappings from demographic information to model parameters. Several key factor (non-exhaustive) that may influence the model parameters are expected to include the population composition, income and job profile, transport resources, prevalence of other diseases that can affect immunity and/or mask identifying COVID infection.

Incorporating the model parameters with the information/insight so gained, and with an eye on the number of distinct clusters exhibited in the parameter sets, would be essential in a diverse and vast country like India.

Some of the relevant questions are:

- The spatio-temporal data on testing is sparse and possibly has missing components. For example, the interpretation of the testing data requires us to know the protocol used to test and the nature of the test. In the absence of this information, interpolation methods need to be developed.
- Many diseases that are prevalent in India have symptoms similar to those from Covid-19 infections. How do we separate them?
- What information can we obtain if the testing method is known, (e.g., serum or swab)
- How do we use the temporal information of measurement data to update model parameters

### 4. Deliverables of the proposed project

1. Clear characterization/definition of clusters
2. Multi-cluster model for spread of epidemic and impact of preventive measures/interventions
3. Data-driven estimation of the parameters of the studied models
4. Identification of parameters for different strains and a model for predicting the extent of mutation
5. Packaging the solution in a usable software form and make it available for ready use by Government of India on need basis (with emphasis on ease of use)

References:

[1] <https://cddep.org/covid-19/>

[2] Singh & Adihikari: <https://arxiv.org/pdf/2003.12055.pdf>

[3] Mandal et al. <http://www.ijmr.org.in/preprintarticle.asp?id=281325;type=0>

[4] Dorigatti et al, <https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-College-COVID19-severity-10-02-2020.pdf>

[5] Fivethirtyeight: <https://fivethirtyeight.com/features/why-its-so-freaking-hard-to-make-a-good-covid-19-model/>