Sensitivity Analysis for COVID-19 Epidemiological Models within a Geographic Framework

Zhongying Wang
zhongyin@usc.edu
University of Southern California
Los Angeles, California

Orhun Aydin
oaydin@usc.edu
University of Southern California
Los Angeles, California

Abstract

Spatial sciences and geography have been integral to the modeling of and communicating information pertaining to the COVID-19 pandemic. Epidemiological models are being used within a geographic context to map the spread of the novel SARS-CoV-2 virus and to make decisions regarding state-wide interventions and allocating hospital resources. Data required for epidemiological models are often incomplete, biased, and available for a spatial unit more extensive than the one needed for decision-making. In this paper, we present results on a global sensitivity analysis of epidemiological model parameters on an important design variable, time to peak number of cases, within a geographic context. We design experiments for quantifying the impact of uncertainty of epidemiological model parameters on distribution of peak times for the state of California. We conduct our analysis at the county-level and perform a non-parametric, global sensitivity analysis to quantify interplay between the uncertainty of epidemiological parameters and design variables.

CCS Concepts

• **Information systems → Geographic information systems**.

Keywords

Sensitivity Analysis, Epidemiological Model, COVID-19, Spatial-temporal Analysis, Uncertainty

ACM Reference Format:

Zhongying Wang and Orhun Aydin. 2020. Sensitivity Analysis for COVID-19 Epidemiological Models within a Geographic Framework. In *1st ACM SIGSPATIAL International Workshop on Modeling and Understanding the Spread of COVID-19 (COVID-19), November 3, 2020, Seattle, WA, USA.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3423459.3430755

1 Introduction

COVID-19 is a severe acute respiratory syndrome (SARS) caused by the SARS-CoV-2 virus [7]. On March 11, 2020, COVID-19 is declared to be a pandemic with 12,552,795 infected persons and 561,617 deaths globally as of July 12, 2020 [12]. In the United States, the number of total cases is at 4,974,959 with 161,284 deaths [12], making the COVID-19 pandemic a national problem.

Spatial analysis has played an essential role during the COVID-19 crisis in terms of spatial analysis of transmission and the number of new cases [2, 6, 11, 13, 20, 21], and mapping susceptible populations [10]. The SARS-CoV-2 is contracted from person-to-person via respiratory droplets [18], making public places where people are in close contact likely places for high transmission rates [2, 6].

Epidemiological models are used within a geographic context to map the spread of the novel SARS-CoV-2 virus and to make decisions regarding state-wide interventions and allocating hospital resources [13]. Data required for epidemiological models are often incomplete, and biased, making uncertainty quantification a necessity for decision-making. The spatial resolution of currently available curated data for United States COVID-19 case and death statistics is at the county-level. Thus, it is important to understand the impact of epidemiological model parameters on actionable variables within a geographic setting.

In this research paper, a sensitivity analysis on decision variables is conducted, and implications of parameter uncertainty on decision variables used by public health officials are showcased for California. We use the Sobol sensitivity [17] to model the impact of epidemiological variables on the spatial and space-time patterns of new COVID-19 hospitalizations at the county level in the state of California. We use the CHIME model (COVID-19 Hospital Impact Model for Epidemics) [5] from University of Pennsylvania to define COVID-19 hospitalization projections. The spatiotemporal series for predicted new COVID-19 hospitalizations is summarized temporally and spatially with time to peak demand, and the Moran's I statistic, respectively. The impact of epidemiological parameters of the CHIME model on the space-time patterns of modeled hospital demand are quantified with the Sobol sensitivity.

2 Data & Methodology

2.1 Data

CHIME model requires several parameters, including population, the number of currently hospitalized COVID-19 patients, doubling time, social distancing effect, infectious days, and optional including hospital resource parameters (number of beds, intensive care units (ICUs) and ventilators) to forecast future COVID-19 hospitalizations and its impact on the hospital resources.

The population data used in the model is from ESRI's 2019 Updated Demographics [1]. This data updates annually based on several sources of data, including a full-time series of intercensal and vintage-based county estimates from the US Census Bureau and a time series of county-to-county migration data from the Internal Revenue Service. Projections are necessarily derived from current events and past trends, which is calculated from previous census counts provided by the American Community Survey (ACS). COVID-19 related data, including incidence, confirmed cases and death are all from JHU CSSE COVID-19 Data [2][4].

COVID-19, November 3, 2020, Seattle, WA, USA

[1] https://doc.arcgis.com/en/esri-demographics/reference/methodologies.htm
[2] https://coronavirus.jhu.edu/

2.2 The CHIME Model

Predictive Healthcare at Penn Medicine initiated the tool Hospital impact model to assist hospitals and public health officials with capacity planning, including daily increase, peak hospitalized census, ICU admissions, number of patients requiring ventilators and timeline prediction.

CHIME model is one of many customized models based on SIR (Susceptible, Infected, Recovered) model [1], which is a commonly used epidemiological model to forecast the number of infected people from a disease in a closed population over time. The main idea of this model is dividing the population into compartments throughout the progression of the disease, such as susceptible, infected and recovered population. The model dynamics are defined by the following equations:

$$(_{C+1} = (_C - V(_C \, _C \tag{1}$$

$$_{C+1} = \, _C + V(_C \, _C - W \, _C \tag{2}$$

$$'_{C+1} = '_C + W \, _C \tag{3}$$

where V represents the effective contact rate, which can be computed as the transmissibility g multiplied the average number of people exposed 2: $V = g \times 2$. The transmissibility is the basic property related to the virulence of the pathogen, but the number of people exposed is the parameter can be changed by policies, like social distancing or mask wearing. W is the inverse of the mean recovery time, and recovery time indicates the period of infection getting cleared and varies for the severity of the symptoms. For COVID-19, the average is normally considered as 1/14. The basic reproduction number ($'_0$) is an indicator of the contagiousness or transmissibility of infectious and parasitic agents and represents average number of people can be infected by any given infected person without immunity from past exposures or vaccination [3]. It is defined as $'_0 = V/W$. The disease is supposed to spread if $'_0$ is > 1 and the larger the number is, the faster it will spread. Since the transmissibility and social contact rates are hard to compute, this parameter can be replaced by doubling times. Since the rate of new infections in the SIR model 6 can be computed with doubling time $)_3$: $6 = V(- W$, V can be computed with the initial population size of susceptible individuals as $V = (6 + W)$.

ESRI has developed a toolbox for the CHIME model and parameters used in sensitivity analysis and their explanations are shown in **Table 1**.

2.3 Sobol Sensitivity

Sobol sensitivity analysis quantifies the impact of total-effect indices and higher-order interactions and has no limit for the preparation of the input sample, and such characters enable it to deal with auto-correlated spatial input [8].

The Sobol method is one of the variance-based methods, which can compute sensitivity indices regardless of the linearity or monotonicity, or other assumptions on the underlying model. In variance based method, the fractional contribution of each input to the variance V of the model is estimated and the total variance $+$ of the model output is decomposed to calculate the sensitivity indices for every independent $-_8$.

$$+ = \tilde{O}_8 +_8 + \tilde{O}_{8<9} +_{89} + \tilde{O}_{8<9<<} +_{89<} + \cdots + +_{12\,\prime\prime\prime\prime:} \tag{4}$$

where $+_8$ is the share of the output variance explained by the 8th model input, and indicates the sensitivity of $.$ to $-_8$. $+_{89}$ is the share of the output variance explained by the interaction of the 8th and 9th model inputs, and indicates the sensitivity of $.$ to the interaction of $-_8$ and $-_9$. $:$ is the total number of the model inputs.

The first-order sensitivity computes the contribution to the output variance of the main effect of $-_8$ and is defined with conditional variances as

$$/_8 = \frac{+_8}{+} = \frac{+OA[\ (.|-_8)]}{+OA(.)} \tag{5}$$

where the inner expectation of the numerator is conditional on $-_8$ taking a value $-_{8*}$ within its range of uncertainty, while the outer variance is calculated over all possible values of $-_8$. If the variance of the conditional expectation $(.|-_8 = G_{8*})$ for some particular value $-_8 = G_{8*}$ is relatively large when compared to the total variance, and all the effects of the $-_9$, $9 < 8$, then factor $-_8$ can be considered as an influential one. Similarly, $/_{89} = \frac{+_{89}}{+}$ indicates the sensitivity indices of the interaction effect of $-_8$ and $-_9$ [17].

According to $\sum_{8=1}^{:} /_8 + \sum_{8=1}^{:} /_{89} + \cdots + \sum_{8=1}^{:} /_{89\ldots:} = 1$, total-order index $/)_8$, which measures the contribution to the output variance of $-_8$ including all variance caused by its interactions, of any order, with any other input variables can be defined as

$$/)_8 = 1 - \frac{+_{-8}}{+} = 1 - \frac{+OA[\ (.|-_{-8})]}{+OA(.)} \tag{6}$$

Sobol sensitivity quantifies the contribution of variance from a set of explanatory variables on the variation of target variable of interest. Thus, it provides a statistical framework within which the impact of a model parameters can be assessed marginally and jointly.

2.4 Experimental Design

The method of Sobol sensitivity analysis computes the indices by using the decomposition of the output variance in Eq.1. Capturing representative variance requires rigorous design of experiments. In this work, experiments are designed using the Saltelli sampling scheme [15]. Steps of defining experiments are elaborated below:

(1) Choose an integer N as the size of the base sample.
(2) Generate a sample matrix ($\#\bullet 2:$) of the input factors by using the Saltelli sampler, where $:$ is the number of input factors. Divide the matrix into two and define each part as and , which contain half of the sample data.
(3) Duplicate the matrix and replace the 8-th column with the same column from matrix B, then define it as $_8$. The matrix $_8$ is the duplicate of matrix , except that the $8 - C$ column is replaced with the $8 - C$ column in matrix A.
(4) Compute the model output for all the input values in the sample matrices and then use the Eq.5 and 6 to compute the sensitivity indices.

Sampling scheme above defines experiments to model variance of response variables without increasing the computational load by employing full factorial design.

2.5 Spatiotemporal Sensitivity Analysis

The output response for every experiment is a spatiotemporal series of CHIME model output. We summarize the predicted number of hospitalizations time series at every county with time to peak hospitalizations.

Table 1: Parameters in the CHIME model

Parameter	Explanation
Doubling Time in Days	The number of days that the number of infected individuals to double without interventions.
Social Distancing	The quantitative estimation of social contact reduction in each catchment area.
Infectious Days	The number of days an infected person has the ability to infect others.
Hospitalization Rate	The percentage of all infected cases that will need hospitalization.
Average Days of Hospital Stay	The average number of days COVID-19 patients have needed to stay in a hospital.
ICU Rate	The percentage of all infected cases which will need to be treated in an ICU.
Average Days in ICU	The average number of days COVID-19 patients have needed ICU care.
Ventilated Rate	The percentage of all infected cases that need mechanical ventilation.
Average Days on Ventilator	The average number of days with ventilation needed for COVID-19 patients.

$$C^{(8)}_{?40:} = \max_{1 \le C \le C < 0G} {}^{(8)}_{=4F} \tag{7}$$

In Eq. 7, the time to peak hospitalizations at location 8, $C^{(8)}_{?40:}$, is the time at which the number of predicted hospitalizations ${}^{(8)}_{=4F}$ reaches its maximum value. In cases where projections decline, $C^{(8)}_{?40:}$ is assumed to be 0. $C^{(8)}_{?40:}$ reduces the time series into a spatial distribution of time to peak hospitalizations, denoted as $)_{?40:} = C^{(1)}_{?40:} \cdot C^{(2)}_{?40:} \cdot \text{""} \cdot C^{(\#)}_{?40:}$.

The spatial distribution of time to peak hospitalizations are summarized using the Moran's I statistic, that quantifies the spatial patterns of time to peak hospitalizations.

$$= \frac{\# \sum_{8=1} \sum_{9=1} F_{8\cdot 9}(C^{(8)}_{?40:} - \bar{)}_{?40:})(C^{(9)}_{?40:} - \bar{)}_{?40:})}{, \sum_{8=1}(C^{(8)}_{?40:} - \bar{)}_{?40:})^2} \tag{8}$$

In Eq. 8, w is the geographic weight, and , is the sum of all weights, , $= \sum_{8=1} \sum_{9=1} F_{8\cdot 9}$. A positive and statistically significant indicates spatial clustering, and a significant and negative indicates dispersion of $C^{(8)}_{?40:}$.

3 Results

The decision variable we investigate is the spatial clustering of time until the peak number of cases, $)_{?40:}$, are observed at a given county.A short time to peak indicates that the hospitals in that county are about to receive a high number of COVID-19 patients. Spatial clustering of $)_{?40:}$ indicates that similar volumes in hospital demand will exist at neighboring counties.

We conducted 800 simulations by varying three model parameters. Our choice behind these parameters are due to high uncertainty associated with them. According to epidemiology analysis, the ($'_0$) ranges from about 2 to 6 based on initial estimates of the early dynamics of the outbreak in Wuhan, China [14]. The doubling time is computed with this uncertainty range. According to CDC and other researches, 88% and 95% of specimens no longer yielded replication-competent virus after 10 and 15 days, but recovery of replication-competent virus between 10 and 20 days after symptom onset has been documented in some persons with severe symptoms [19]. The uncertainty of infectious days is then chosen from 10 to 20. Experiment parameters and their ranges are presented in Table 2.

Table 2: Epidemiological Experiment Variables

Epidemiological Experiment Variables	
Doubling Time in Days	[2.27,10.05]
Social Distancing (%)	[0,50]
Infectious Days	[10,20]

We present the response surface for the average number of days to peak in the state of California with respect to uncertain model parameters. The response surface is depicted in Figure 1.

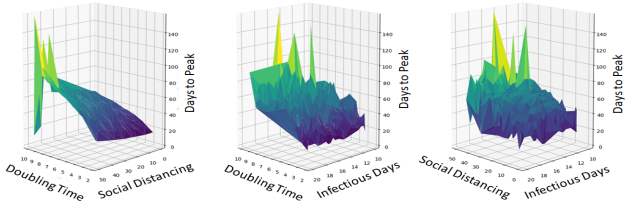

Figure 1: Surface Plot of Sensitivity Analysis

In some of our simulations, the peak is not observed within our simulation time span (180 days). These simulations correspond to peaks in the response surface. Our results indicate that for increasing doubling time and social distancing, the peak is delayed. Response surfaces with respect to infectious days indicate a more complex relationship that points to a high amount interaction between infectious days and other epidemiological variables.

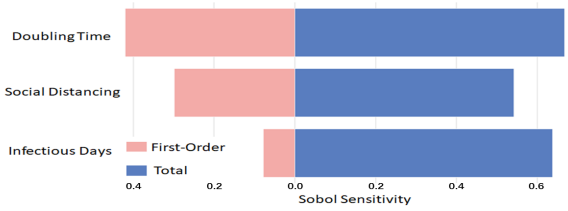

Figure 2: Tornado Plot for Sobol Sensitivity

Sensitivity of spatial patterns of $)_{?40:}$ is showcased in Figure 2. Doubling time is the most first-order sensitive variable, followed by social distancing and infectious days. The variables are ranked with respect to their first-order sensitivity. Note that all three model variables have high total sensitivity. This indicates strong interactions between these variables and the spatial patterns of hospital demand.