Trends
in Computational and
Applied Mathematics

Formulation and Solution of an Inverse Reliability Problem to Simulate the Dynamic Behavior of COVID-19 Pandemic

F. S. LOBATO[1*], G. M. PLATT[2], G. B. LIBOTTE[3] and A. J. SILVA NETO[4]

ABSTRACT. Different types of mathematical models have been used to predict the dynamic behavior of the novel coronavirus (COVID-19). Many of them involve the formulation and solution of inverse problems. This kind of problem is generally carried out by considering the model, the vector of design variables, and system parameters as deterministic values. In this contribution, a methodology based on a double loop iteration process and devoted to evaluate the influence of uncertainties on inverse problem is evaluated. The inner optimization loop is used to find the solution associated with the highest probability value, and the outer loop is the regular optimization loop used to determine the vector of design variables. For this task, we use an inverse reliability approach and Differential Evolution algorithm. For illustration purposes, the proposed methodology is applied to estimate the parameters of SIRD (Susceptible-Infectious-Recovery-Dead) model associated with dynamic behavior of COVID-19 pandemic considering real data from China's epidemic and uncertainties in the basic reproduction number ($\mathscr{R}_0$). The obtained results demonstrate, as expected, that the increase of reliability implies the increase of the objective function value.

Keywords: inverse problem, reliability-based optimization, modeling, COVID-19.

1 INTRODUCTION

Since November 2019, the world has observed the spread of a new disease, the COVID-19 (Coronavirus disease 2019). On 11 March 2020, the World Health Organization (WHO) declared COVID-19 a pandemic. Since then, the scientific community have made efforts—in different areas—to understand and mitigate the effects of this disease. By July 27, 2020, more than 67,000

---

*Corresponding author: F. S. Lobato – E-mail: fslobato@ufu.br

[1]Chemical Engineering Faculty, Federal University of Uberlândia, Uberlândia, Brazil – E-mail: fslobato@ufu.br
https://orcid.org/0000-0002-7401-4718

[2]School of Chemistry and Food, Federal University of Rio Grande, Santo Antônio da Patrulha, Brazil – E-mail: gmplatt@furg.br https://orcid.org/0000-0003-0506-2561

[3]National Laboratory for Scientific Computing (LNCC/MCTI), Petrópolis, Brazil – E-mail: glibotte@lncc.br
https://orcid.org/0000-0002-4583-6026

[4]Polytechnic Institute, Rio de Janeiro State University, Nova Friburgo, Brazil – E-mail: ajsn@iprj.uerj.br
https://orcid.org/0000-0002-9616-6093

articles related to the COVID-19 have been published [12], many of them dealing with mathematical models related to the spread of the disease. To evaluate the individual behavioral response, governmental actions, zoonotic transmission and emigration of a large proportion of the population in a short period of time and vaccine administration related with the COVID-19, various mathematical model-based predictions have been proposed and studied [10]. These models are obtained through the formulation and solution of an inverse problem and aim to describe a state of infection (susceptible and infected) and a process of infection (the transition between these states) by using compartmental relations. In this case, the population is divided into compartments by taking assumptions about the nature and time rate of transfer from one compartment to another [2, 20].

The solution of inverse problems represents a significant challenge due to the complexity and non-linearity of the mathematical models frequently addressed. These models are typically associated with non-linear equations that represent mass, energy, and momentum balances, which are submitted to physical, constitutive, environmental, and design limitations. During the parameter estimation process, the vector of design variables and system parameters are usually considered as deterministic values; however, in a real-world scenario, we must consider some degree of uncertainty. In general, these uncertainties are related to [15]: *i*) geometrical and constitutive parameters, and *ii*) simplifications adopted in the model formulation.

Probabilistic methods have been widely used in optimization problems, in order to mitigate possible uncertainties inherent to mathematical models. An approach capable of providing results that incorporate the influence of uncertainties in the system is the reliability-based optimization. The problem is modeled using random variables and the reliability depends on satisfying the probabilistic constraints of the problem. In this context, the incorporation of uncertainties in the analysis of simulations of compartmental models can be a good alternative in an attempt to quantify uncertainties related to an epidemic. Taking into account the randomness related to contacts between individuals in a population, the parameters that describe the evolution of an epidemic over time can vary due to uncertainties.

In this contribution, the aim is to evaluate the influence of uncertainties related to the basic reproduction number ($\mathcal{R}_0$), during the solution of an inverse problem used to predict the dynamic behavior of COVID-19 pandemic considering real data from China's epidemic. In order to determine the parameters that characterizes the proposed mathematical model (compartmental SIRD - Susceptible-Infectious-Recovery-Dead model), a reliability inverse problem is formulated and solved considering the Differential Evolution Algorithm (DE) [18,19] as optimization tool. In order to evaluate the presence of uncertainties, the Inverse Reliability Analysis (IRA) strategy [11] approach is considered.

The analysis of the uncertainties in the value of $\mathcal{R}_0$ is promptly justified by the following aspects: (i) the values of $\mathcal{R}_0$ are intrinsically dependent on the methodology employed, even for a specific type of dynamic model [7] and (ii) there is a large range for values of $\mathcal{R}_0$ for the China's pandemic scenario (for instance, Riou and Althaus [14] reported $0.8 < \mathcal{R}_0 < 5$ in their computational experiments). In this sense, the computational framework developed here—based on a reliability

approach—provides a flexible tool to evaluate the effects of the commented uncertainties in the prediction of the spread of the disease.

This work is organized as follows. Section 2 presents the description of mathematical model considered to represent the evolution of COVID-19 pandemic. In Section 3 the general aspects regarding the Reliability-Based Optimization are presented. A review about DE to deal with mono-objective optimization is presented in Section 4. In Section 5, the proposed methodology is discussed. The results obtained from the application of the methodology are presented in Section 6. Finally, the conclusions are outlined in Section 7.

2   MATHEMATICAL MODELING IN EPIDEMIOLOGY

Traditionally, models based on compartments have been used to represent dynamic behavior of diseases [1,3,9] . These models are constituted by compartments–each of them representing fractions of the population–interconnected by specific terms. These terms are determined as functions of the hypotheses considered during the formulation of the particular model of disease.

2.1   The SIRD Model

As mentioned earlier, the aim of this contribution is to determine the parameters of an epidemiological model to predict the evolution of COVID-19 epidemic considering real data from China's pandemic. For this purpose, the SIRD (Susceptible-Infectious-Recovery-Dead) model is considered [1]. Here, among other possible formulations for the SIRD model [9], we employ the formulation described by Chatterjee et al. [3]. In SIRD model, an individual/host is susceptible to infection and the disease can be transmitted from any infected individual to any susceptible individual. For the susceptible individuals, the dynamics is described by:

$$\frac{dS}{dt} = -\beta SI/N, \quad S(0) = S_0 \tag{2.1}$$

where t is the time, β is the per capita transmission rate, S_0 is the initial condition for the susceptible population and N is the population size. Similarly, for the infected individuals, we have:

$$\frac{dI}{dt} = \beta SI/N - \alpha I - \gamma I, \quad I(0) = I_0 \tag{2.2}$$

where α is the effective per day recovery rate, γ is the per-capita death rate, and I_0 is the initial condition for the infected population. The individuals that dead due to disease are given by following relation:

$$\frac{dD}{dt} = \gamma I, \quad D(0) = D_0 \tag{2.3}$$

where D_0 is the initial condition for the dead population. The recovery individuals are modelled by the following differential equation:

$$\frac{dR}{dt} = \alpha I, \quad R(0) = R_0 \tag{2.4}$$

where R_0 is the initial condition for the recovered population.

The schematic representation of this model, with each compartment, is presented in Fig. 1.
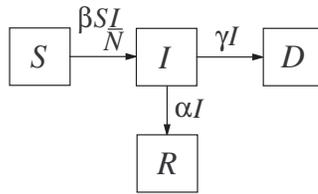

Figure 1: Compartments in the SIRD model [1, 3].

Here, we used a normalized version of the SIRD model, considering S_n, I_n, D_n and R_n, defined as $S_n \equiv S/N$, $I_n \equiv I/N$, $D_n \equiv D/N$ and $R_n \equiv R/N$. In this case, the constraint regarding the population size N is represented by $S_n(t) + I_n(t) + D_n(t) + R_n(t) = 1$. This normalized system is represented by the following set of differential equations:

$$\frac{dS_n}{dt} = -\beta S_n I_n, \quad S_n(0) = S_{n0} \tag{2.5}$$

$$\frac{dI_n}{dt} = \beta S_n I_n - (\alpha + \gamma) I_n, \quad I_n(0) = I_{n0} \tag{2.6}$$

$$\frac{dD_n}{dt} = \gamma I_n, \quad D_n(0) = D_{n0} \tag{2.7}$$

$$\frac{dR_n}{dt} = \alpha I_n, \quad R_n(0) = R_{n0} \tag{2.8}$$

Similarly, S_{n0}, I_{n0}, D_{n0} and R_{n0} represent the initial conditions for the susceptible, infected, dead and recovered populations, respectively, for the normalized model.

To predict the dissemination spread of infectious disease into a population, the basic reproduction number ($\mathscr{R}_0$) is usually considered as an epidemiological metric. As mentioned by Anastassopoulou and co-workers [1], this parameter represents the average number of secondary cases that result from the introduction of a single infectious case in a totally susceptible population during the infectiousness period.

It is important to mention that this parameter is affected by type of pathogen, environmental conditions and behaviour of infected population, among others [24]. In addition, when $\mathscr{R}_0 > 1$, the infection will be able to start spreading in a population, i.e., for an epidemic to occur in a susceptible population this parameter must be greater than 1 [1]. By using this value, the potential size of an outbreak can be evaluated and the proportion of the population that must be vaccinated can be estimated [7].

For the model described by Eqs. (2.5)-(2.8), the parameter $\mathscr{R}_0$ is given by following relation [1]:

$$\mathscr{R}_0 = \frac{\beta}{\alpha + \gamma} \tag{2.9}$$

As pointed out in Section 1, the uncertainties regarding the value of $\mathscr{R}_0$ justify the study of the reliability problem.

3   RELIABILITY-BASED OPTIMIZATION

Uncertainties are related to the observations of most phenomena that occur in nature. In general, this kind of problem can be modeled using random variables. In the case of optimization problems, reliable results can be interpreted as those that meet certain failure criteria. Specifically in terms of reliability-based optimization, probabilistic constraints are incorporated into the problem, so that optimizers calculated under these conditions satisfy certain levels of reliability and, therefore, represent reliable solutions.

Initially, let $\mathbf{X} = (X_1, \ldots, X_n)$ bet a set of n continuous random variables. The functional relationship between these variables is performed by means of a performance function, represented as $Y = g(\mathbf{X})$. The boundary between the safe and failure regions, which represents the state in which a system can no longer fulfill the function for which it is designed, is called the limit state function and denoted by $Y = 0$. Therefore, the probability of failure of a given random event, in terms of the performance function $g(\mathbf{X})$, is defined as

$$p_f = P[g(\mathbf{X}) \leq 0] = \int_{g(\mathbf{x}) \leq 0} f_{\mathbf{X}}(\mathbf{x}) \, d\mathbf{x}, \tag{3.1}$$

where $f_{\mathbf{X}}(\mathbf{x})$ represents the joint probability density function and the integration is performed over the failure region.

Equation (3.1) may be computationally expensive to be solved, since the number of random variables in practical applications is generally high. Additionally, the limit state can be described by a highly nonlinear function. Thus, there is rarely a closed-form solution for the integration of Eq. (3.1). One possible treatment is to use analytical approximations of such integrals, which are simpler to calculate. The following is an inverse approach to reliability analysis, capable of approximating reliable solutions through an iterative technique. For more details on reliability-based optimization, refer to Haldar and Mahadevan [6].

3.1   Inverse Reliability Analysis

Consider the case in which the failure region in the reliability problem is constrained by $g(\mathbf{X}) \leq \eta$, in such a way that $g(\mathbf{X})$ is highly nonlinear. In this case, $F_{\mathbf{X}}(\eta) = P[g(\mathbf{X}) \leq \eta]$, where $F_{\mathbf{X}}$ represents the cumulative distribution function. A generalized probability index Ω_g is a monotonically decreasing function, such that

$$F_{\mathbf{X}}(\eta) = \Phi\left(-\Omega_g\right), \tag{3.2}$$

where Φ represents the standard normal cumulative distribution function. In the specific case where the failure region is determined by $g(\mathbf{X}) \leq 0$, Eq. (3.2) is represented by

$$F_{\mathbf{X}}(0) \leq \Phi\left(-\Omega_t\right), \tag{3.3}$$

and Ω_t is a scalar that represents the target reliability index.

Equation (3.3) can take two distinct forms, taking into account the corresponding inverse transformations, given by

$$\Omega = -\Phi^{-1}\left(F_{\mathbf{X}}\left(0\right)\right) \geq \Omega_{\mathrm{t}} \tag{3.4a}$$

and

$$\eta^* = F_{\mathbf{X}}^{-1}\left(\Phi\left(-\Omega_{\mathrm{t}}\right)\right) \geq 0\,, \tag{3.4b}$$

where Ω is the reliability index and η^* is called the target probabilistic performance measure. One may notice that Eq. (3.4a) is the compact form of

$$\Omega = -\Phi^{-1}\left(\int_{g(\mathbf{x}) \leq 0} f_{\mathbf{X}}\left(x\right) \mathrm{d}\mathbf{x}\right)\,,$$

whose approximation can be calculated using the reliability index approach [13].

In turn, the inverse approach, which is the focus of this work, given by Eq. (3.4b), can be understood as the determination of a vector of design variables which satisfies the probabilistic constraint for a given target probability of failure, since $F_{\mathbf{X}}^{-1}\left(\bar{p}_{\mathrm{f}}\right) \geq 0$, where $\bar{p}_{\mathrm{f}} = \Phi\left(-\Omega_{\mathrm{t}}\right) = F_{\mathbf{X}}\left(0\right)$ is the limiting probability of failure. In this case, given a vector of variables $\mathbf{X}$, the probabilistic constraint is satisfied when $\eta^* \geq 0$, such that

$$\eta^* = F_{\mathbf{X}}^{-1}\left(\int_{g(\mathbf{x}) \leq \eta^*} f_{\mathbf{X}}\left(x\right) \mathrm{d}\mathbf{x}\right)\,.$$

Therefore, in the performance measurement approach [21], the value of the performance function $G\left(\mathbf{U}\right)$ must be minimized, while the distance between the origin of the standard normal space and the radius Ω_{t} remains constant, that is, $\mathbf{u}^* = \arg\min_{\mathbf{u}} G\left(\mathbf{u}\right)$, subject to $\|\mathbf{u}\| = \Omega_{\mathrm{t}}$. The approximate solution to this problem can be calculated using the Advanced Mean Value method, whose fundamental description is presented below.

3.1.1   Advanced Mean Value Method

One of the first algorithms designed to solve inverse reliability problems by means of the performance measurement approach, which stands out for its relative efficiency and simplicity, is the Advanced Mean Value (AMV) method [25, 26]. The method was originally proposed as a computationally efficient technique to calculate the cumulative distribution function of a response function. Later, the method was adapted to be applied in the reliability analysis.

In the first step of the algorithm, the random variables of the original problem must be transformed into the standard normal space using the Rosenblatt transformation [16]. Let $F_{\mathbf{X}}\left(\mathbf{x}\right)$ be the cumulative distribution function of the random variable $\mathbf{X}$. The transformation is expressed by $\mathbf{u} = \Phi^{-1}\left(F_{\mathbf{X}}\left(\mathbf{x}\right)\right)$, where Φ represents the standard normal cumulative distribution function. Transformations of this type are always possible for continuous random variables. A particularly simple transformation can be used when the random variables are mutually independent, that is, $\mathbf{u} = \left(\mathbf{x} - \mu_{\mathbf{X}}\right)/\sigma_{\mathbf{X}}$. It is important to emphasize that the transformation preserves the probability

and, therefore, such procedure does not introduce any type of potential error to the calculation of the probability of failure.

At each iteration, one must calculate the search direction

$$\mathbf{a}^{(k)} = \frac{\nabla G|_{\mathbf{u}^{(k)}}}{\left\| \nabla G|_{\mathbf{u}^{(k)}} \right\|} \tag{3.5}$$

and the estimate of the most probable point of failure in the iteration k + 1 is calculated using $\mathbf{u}^{(k+1)} = \Omega_t \mathbf{a}^{(k)}$. Since Ω_t is constant throughout the iterations, it is assumed that an approximation of the vector of design variables that satisfies the target reliability index is sufficiently accurate when $\mathscr{E}^{(k)} = \left\| \mathbf{u}^{(k)} - \mathbf{u}^{(k-1)} \right\| < \varepsilon$, where ε is the error threshold. When $k \to \infty$, the point $\mathbf{u}^*$ is expected to be obtained, which represents the approximate solution of the reliability-based optimization problem, that is, the point on the hypersphere of radius Ω_t, which satisfies the established level of reliability (defined in terms of the reliability index).

4   DIFFERENTIAL EVOLUTION

Differential Evolution (DE) is a metaheuristic algorithm proposed by Storn and Price [18], which has attracting attention in the last years, as a consequence of its simplicity and powerful performance [5]. The structure of the canonical version of the method is based on three mechanisms: mutation, crossover (or recombination) and selection. These steps are performed for each element of the population, formed by *NP* individuals, for a quantity of generations (represented by $G_{\max}$). In the first step, called mutation, two elements in the population are randomly selected, subtracted, multiplied by a factor (the perturbation rate *F*) and, then, added to a third random individual in the population, as follows:

$$rand/1 \to \mathbf{v} = \mathbf{x}_{r_1} + F\left(\mathbf{x}_{r_2} - \mathbf{x}_{r_3}\right) \tag{4.1}$$

where $\mathbf{x}_{r_1}$, $\mathbf{x}_{r_2}$ and $\mathbf{x}_{r_3}$ are candidate-solutions randomly selected in the population. This new vector ($\mathbf{v}$) is the mutant vector. The representation $rand/1$ indicates that the vector $\mathbf{x}_{r_1}$ is randomly selected and only one pair of vectors ($\mathbf{x}_{r_2}$ and $\mathbf{x}_{r_3}$) is used (other schemes are possible [18]). In the second step, the crossover, the mutant vector $\mathbf{v}$ is combined, coordinate to coordinate, to the current individual in the population ($\mathbf{x}$), following some stochastic rules (controlled by a crossover constant, *CR*), producing a candidate-vector (usually represented by $\mathbf{u}$). Finally, a greedy selection rule is employed to choose between the original individual ($\mathbf{x}$) and the candidate-vector ($\mathbf{u}$) to form the offspring in a new generation.

5   METHODOLOGY

In this section, the procedure adopted to solve both deterministic and reliable inverse problems are presented.

5.1   Inverse Deterministic Problem

In order to determine the SIRD parameters, it is necessary to formulate and solve an inverse problem considering the real data from China. In general, the identification procedure consists in obtaining the model parameters through the minimization of the difference between calculated and reported values. For this purpose, the SIRD parameters (β, γ and α) are considered as design variables in the proposed inverse problem. The non-dimensional value for the number of infectious individuals (I_{n0}) also is considered as an unknown variable. This is adopted in order to guarantee a coherent initial condition for the differential equation system, i.e., the initial condition for the normalized number of susceptible individuals is defined as a function of I_{n0}, that is, $(S_n(0), I_n(0), D_n(0), R_n(0)) = (1 - I_{n0}, I_{n0}, 0, 0)$. Thus, the proposed inverse (deterministic) problem considers the design variables γ, β, α and I_{n0}. In this case, the objective function is the sum of squared residuals related to reported data of dead and infected individuals. Thus, the objective is to minimize

$$OF = \frac{1}{\max\left(\mathbf{I}_n^{\text{rep}}\right)^2} \sum_{i=1}^{M} \left(I_{ni}^{\text{rep}} - I_{ni}^{\text{sim}}\right)^2 + \frac{1}{\max\left(\mathbf{D}_n^{\text{rep}}\right)^2} \sum_{i=1}^{M} \left(D_{ni}^{\text{rep}} - D_{ni}^{\text{sim}}\right)^2 \tag{5.1}$$

subject to Eqs. (2.5)–(2.8), where the superscripts "rep" and "sim" stand for reported and simulated data, respectively, and M is the number of reported data.

5.2   Inverse Reliability Problem

To formulate the inverse reliability problem considered in this work, it is necessary to define a set of random variables and their functional relationship. For this purpose, the probability of failure is based on the saturation of the parameter $\mathcal{R}_0$, defined by Eq. (2.9), in relation to a reference value. Mathematically, this deterministic constraint can be represented as $\mathcal{R}_0 - \mathcal{R}_{0,\text{ref}} \leq 0$, where $\mathcal{R}_{0,\text{ref}}$ is the upper limit of the basic reproduction number, defined by the solution of the inverse deterministic problem.

In the context of reliability-based optimization, β, α and γ are now random variables. Considering the Rosenblatt transformation (see Section 3.1.1), the probability of failure can be represented as

$$P\left[\frac{\mu_\beta + u_1\sigma}{\mu_\alpha + u_2\sigma + \mu_\gamma + u_3\sigma} - \mathcal{R}_{0,\text{ref}} \leq 0\right] \geq \Theta \tag{5.2}$$

where μ_β, μ_α and μ_γ represents the average values for the parameters β, α and γ, respectively, σ is the standard deviation and Θ is the probability of success.

It is important to emphasize that uncertainties are related to β, α and γ due to the random nature of epidemiological models. On the other hand, the variable I_{n0} is considered a deterministic value. The inverse reliability problem consists in calculating the deterministic and random variables in order to minimize Eq. (5.1), subject to the probabilistic constraint given by Eq. (5.2)) and to the SIRD model, Eqs. (2.5)–(2.8)).

To solve this problem, DE and AMV methods are employed together, using a double-loop structure. The main steps of the algorithm are the following: in the outer loop, the deterministic candidate solutions (vectors of design variables) are generated using DE; In the inner loop, for each candidate solution AMV runs in order to determine the value of the inequality constraint through the determination of the vector of random variables; Deterministic and random variables are used to evaluate the objective function and the inequality constraints. In this case, the objective function is penalized if any constraint is violated, by means of the Static Penalization Method [22].

6   RESULTS AND DISCUSSION

In this Section, we present the results for the inverse deterministic problem, also considering a deterministic constraint regarding the value of $\mathscr{R}_0$, in order to illustrate the deterioration of the unconstrained solution with respect to the constrained problems. The inverse reliability problem is also analyzed, and reliable simulated profiles are presented. For both problems, we adopt the number of daily infected individuals and the cumulative number of dead individuals to solve the inverse problems, reported from January 22, 2020 to May 1, 2020 [4, 8].

As pointed out previously, the inverse problems are solved by the DE algorithm, with the following parameters: $NP = 25$, $CR = 0.8$, $F = 0.8$, $G = 500$ and strategy rand/1 (see Eq. (4.1)). The evolutionary process is interrupted if a given number of generations is found (in this case, 500). The SIRD model is numerically solved by a 4-th order Runge-Kutta algorithm.

6.1   Inverse Deterministic Problem

This first case considers that all variables have no uncertainties, in a classical inverse problem approach. The objective function (to be minimized) is represented by Eq. (5.1). We used the following design space for the SIRD parameters: $0 \leq \beta \leq 1$; $0 \leq \alpha \leq 1$; $0 \leq \gamma \leq 1$; $0 \leq I_{n0} \leq 1$ (these values are defined after some preliminary runs). In order to produce a simple statistical analysis, each algorithm runs 20 times by using 20 different seeds for the random generation of the initial population.

Table 1 presents the results (best, average and worst) obtained by using DE, which obtained good estimates for the unknown parameters and, consequently, for the objective function. From the physical point of view, γ equal to 2.3927×10^{-3} implies a small *per capita* death rate, α equal to 7.5125×10^{-2} represent a moderate effective per day recovery rate. I_{n0} represents the initial fraction of cases that were reported (observed). Finally, the probability of disease transmission per contact of the China's population is equal to 0.3686, approximately.

Figure 2 present the non-normalized simulated and reported profiles considering the estimated parameters. In Figs 2(b) and 2(c) we can observe the quality of obtained profiles in comparison to real data. Figure 2(a) presents the evolution of susceptible population during the epidemic. As expected, after the maximum value observed for the number of infected population, the number of susceptible individuals decreases. Besides, both the number of dead and recovered individuals

Table 1: Results for the inverse problem (final time equal to 95 Days).

	Best	Average	Worst
β (Day^{-1})	0.3686	0.3686	0.3686
γ (Day^{-1})	2.3927×10^{-3}	2.3924×10^{-3}	2.3922×10^{-3}
α (Day^{-1})	7.5125×10^{-2}	7.5124×10^{-2}	7.5124×10^{-2}
I_{n0}	2.7806×10^{-3}	2.7806×10^{-3}	2.7805×10^{-3}
OF	0.8249	0.8249	0.8249

(a) Susceptible.

(b) Infectious.

(c) Dead (Cumulative).

(d) Recovered.
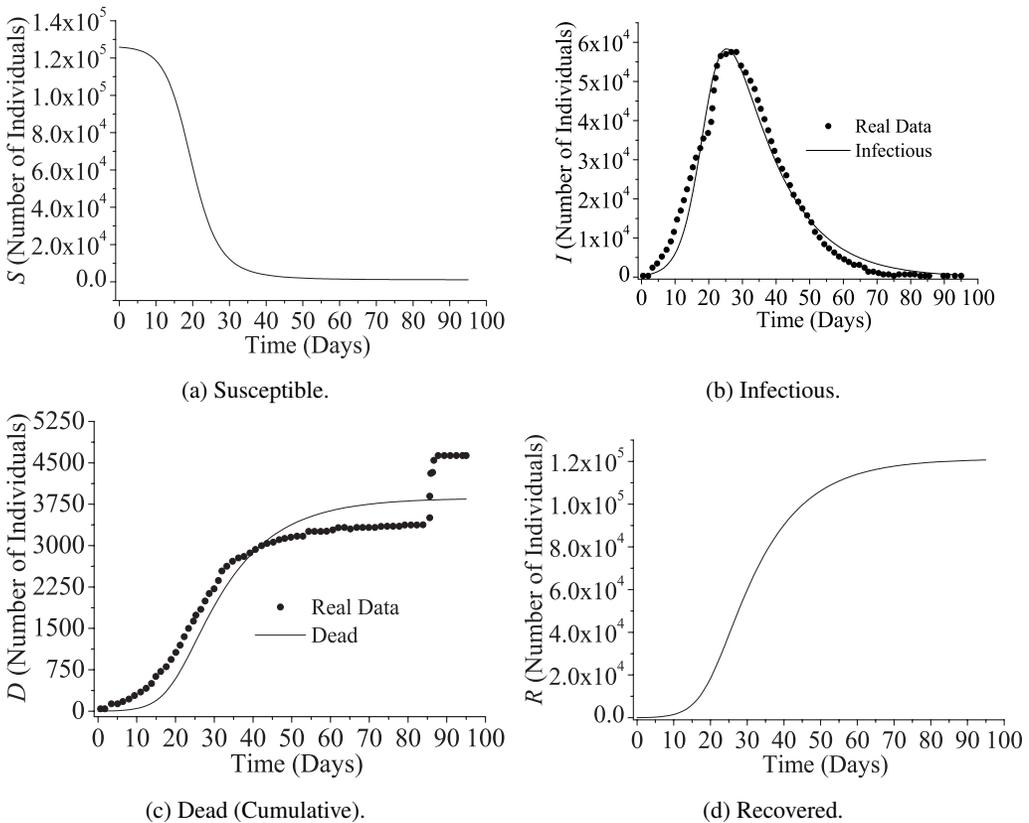
Figure 2: Simulated and reported profiles considering the best estimated parameters.

stabilize with the reduction in the number of infected individuals (see Figs. 2(c) and 2(d)). Finally, it is important to mention that the results are weighted in relation to the number of infected individuals estimated by using DE, i.e., the population size is a fraction of chinese population that, effectively, would have been diagnosed.

Considering the best values presented in the previous table, the basic reproduction number is equals to $\mathscr{R}_0 = 4.7554$. Recently, Sanche et al. [17] reported $\mathscr{R}_0 = 5.7$ for the situation of the

epidemic in China. One may notice that this value agrees with the results obtained in this work. Wearing et al. [23] demonstrated that the use of different methodologies to evaluate $\mathscr{R}_0$ may produce significant differences in this parameter.

As mentioned earlier, the basic reproduction number is used as the epidemiological metric. In order to evaluate the influence of the estimated parameters (β, γ, α and I_{n0}) with respect to $\mathscr{R}_0$, new simulations considering the inclusion of a deterministic inequality equation in the original inverse deterministic problem were performed. Table 2 presents the obtained parameters considering different values for the $\mathscr{R}_{0,\text{ref}}$. In this table we can observe that the increase of the parameter $\mathscr{R}_{0,\text{ref}}$ implies in an increase in the range that defines the inequality constraint. This allows more combinations among β, α, γ and, consequently, a better adjustment of the model in relation to reported data, as observed in Fig. 3. In summary, as the value of the parameter $\mathscr{R}_0$ is equal to 4.7554 (for the unconstrained problem), only for values of $\mathscr{R}_{0,\text{ref}}$ greater than 4.7554, the inequality constraint is, naturally, satisfied. An interesting result can be observed for $\mathscr{R}_{0,\text{ref}} = 1$, particularly in Fig. 3(c). We can note that the profiles for susceptible, infected and recovery individuals are quite different for $\mathscr{R}_{0,\text{ref}} = 1$ in comparison to other values; on the other hand, the behavior for dead individuals is, somehow, similar. This pattern can be explained by the high value for the parameter γ when $\mathscr{R}_{0,\text{ref}} = 1$ – in fact, the inverse problem is "trying" to match the real profile of dead individuals.

Table 2: Best results for the inverse problem considering different values for the parameter $\mathscr{R}_{0,\text{ref}}$ (final time equal to 95 days).

$\mathscr{R}_{0,\text{ref}}$	β (Day^{-1})	γ (Day^{-1})	α (Day^{-1})	I_{n0}	OF
1	0.2649	1.8787×10^{-1}	7.7165×10^{-2}	4.1938×10^{-3}	17.2903
2	0.2253	3.5523×10^{-2}	7.7164×10^{-2}	2.1340×10^{-2}	15.9483
3	0.2450	4.5079×10^{-3}	7.6254×10^{-2}	4.1302×10^{-3}	9.5994
4	0.3193	2.6826×10^{-3}	7.5835×10^{-2}	2.8668×10^{-3}	2.2746
5	0.3686	2.3927×10^{-3}	7.5125×10^{-2}	2.7806×10^{-3}	0.8249

6.2   Inverse Reliability Problem

In the inverse reliability problem, we consider the following design space: $0 \leq \mu_\beta \leq 1$; $0 \leq \mu_\alpha \leq 1$; $0 \leq \mu_\gamma \leq 1$ and $0 \leq I_{n0} \leq 1$.

The parameter $\mathscr{R}_{0,\text{ref}}$, necessary to formulate the inverse reliability problem in Eq. (5.2), was considered equal to 4.7554. It is important to mention that this value corresponds to the solution of the inverse deterministic problem and represents the limit value for the basic reproduction number.

To evaluate the influence of uncertainties, 25 values (equally spaced) considering the following range for the parameter Ω were studied: $1 \leq \Omega \leq 5$ (that corresponds to $84.14\% \leq \Theta \leq 99.99\%$). For each random variable (β, α and γ), the coefficient of variation (cov) equal to 0.05 and 0.1

(a) Susceptible.

(b) Infectious.

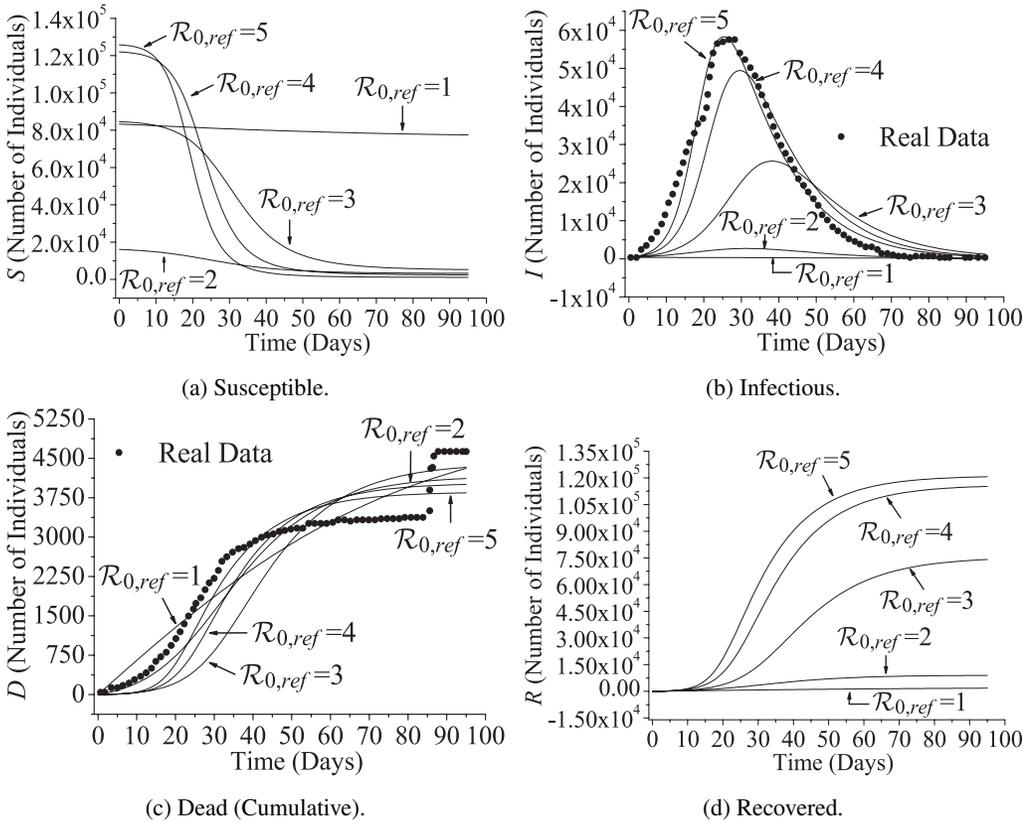(c) Dead (Cumulative).

(d) Recovered.

Figure 3: Simulated and reported profiles considering different values for the parameter $\mathscr{R}_{0,\text{ref}}$.

was adopted. Furthermore, we considered two types of distributions for each random variable (normal and lognormal).

The random variables **u** are considered equal to zero in the first iteration of the IRA strategy. The Euclidean norm of the vector **u** being smaller than 1.0×10^{-5} along two consecutive iterations was considered as the stopping criterion used to finish the IRA strategy. Table 3 shows the results considering different values for the parameter Ω and different types of distribution.

With respect to the deterministic solution, it is observed in this table that the increase of the parameter Ω implies on an increase of the objective function value, i.e., a more reliable solution represents a deteriorated solution in relation to the deterministic value. In addition, it is also possible to observe a reduction in the values of the design variable μ_β in relation to deterministic solution. On the other hand, both the parameters μ_γ and I_{n0} are increased by increasing the parameter Ω. In a similar way that occurs in the deterministic inverse problem, the increase in the value of μ_γ must promote an increase in the number of dead individuals – in order to assure that the constraint regarding the parameter $\mathscr{R}_0$ is not activated. The parameter μ_α has fell variation for each value of the parameter Ω and for each type of distribution.

Table 3: Results for the inverse reliability problem (final time equal to 95 days).

	Ω (Reliability), cov = 0.05 - Normal			Ω (Reliability), cov = 0.05 - LogNormal		
	1 (84.14%)	2.5 (99.38%)	5 (99.99%)	1 (84.14%)	2.5 (99.38%)	5 (99.99%)
μ_β^*	0.3531	0.3189	0.2701	0.3531	0.3192	0.2717
$\mu_\gamma^* \times 10^3$	2.4471	2.6955	3.6039	2.4470	2.6918	3.5374
$\mu_\alpha^* \times 10^2$	7.7165	7.7026	7.6974	7.7165	7.7023	7.6909
$I_{n0} \times 10^3$	2.7272	2.8757	3.5146	2.7272	2.8732	3.4672
OF	0.9628	2.3019	6.6220	0.9624	2.2868	6.4338
	Ω (Reliability), cov = 0.1 - Normal			Ω (Reliability), cov = 0.1 - LogNormal		
	1 (84.14%)	2.5 (99.38%)	5 (99.99%)	1 (84.14%)	2.5 (99.38%)	5 (99.99%)
μ_β^*	0.3299	0.2701	0.2265	0.3301	0.2718	0.2200
$\mu_\gamma^* \times 10^3$	2.5911	3.6039	16.5553	2.5889	3.5335	12.2093
$\mu_\alpha^* \times 10^2$	7.7165	7.7165	7.7167	7.7162	7.7161	7.7159
$I_{n0}^* \times 10^3$	2.8067	3.5146	12.5514	2.8052	3.4645	9.4883
OF	1.7090	6.6220	14.5111	1.6991	6.4181	14.1303

*Day^{-1}.

In terms of the average number of evaluations required, IRA (87525) required a value higher in relation to deterministic problem (11725). Obviously, this is an expected behavior due to the characteristics of the DE algorithm. It should be remembered that a population of candidates is proposed for each generation of DE, increasing the required number of evaluations of the objective function.

Figure 4 presents the simulated and reported profiles considering the best estimated parameters by using $\Omega = \{1, 2.5, 5\}$, cov = 0.05 and Normal distribution. In all these figures, we can observe the influence of reliability in each population profile in relation to deterministic solution. The increase of the parameter Ω imply on increase of distance between numerical solution and reported data. In practice, a more reliable solution avoid that small perturbations violate the probabilistic inequality constraint considered. On the other hand, any small perturbation in relation to obtained deterministic solution implies a greater chance of violating the inequality constraint considered. As pointed out previously, the more reliable solution implies a high quantity of dead individuals—as can be seen in the profiles depicted in Fig. 4(c)—as a consequence of the higher value for the parameter μ_γ, in order to maintain the constraint regarding the parameter $\mathscr{R}_0$ inactive.

Figure 5 presents the influence of the type of distribution and the coefficient of variation by using the IRA strategy. In these figures we can observe, for this application, small differences only for Ω greater than, approximately, 3.5 and 2.5, respectively for normal and lognormal distributions. In relation to coefficient of variation, it is possible to observe significant differences in each profile, i.e., the increase of this parameter implies in increase of the objective function value. This result is expected, since the increase in cov implies on an increase of the range considered for
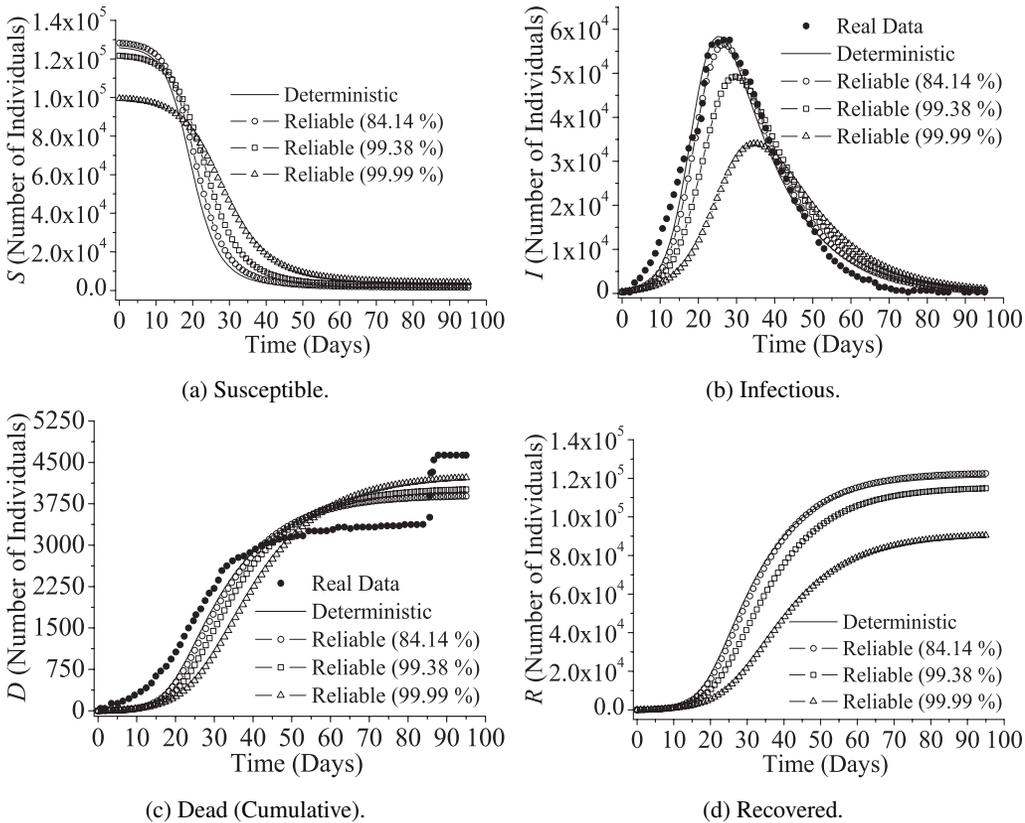
(a) Susceptible.

(b) Infectious.

(c) Dead (Cumulative).

(d) Recovered.

Figure 4: Simulated and reported profiles considering the best estimated parameters by using $\Omega = \{1, 2.5, 5\}$, cov = 0.05 and Normal distribution.

calculating the means of random variables and, consequently, more violations in the probabilistic inequality equations can be obtained.

7   CONCLUSIONS

This paper investigated the influence of uncertainties in the determination of the parameters of a SIRD (Susceptible-Infectious-Recovery-Dead) model used to predict the dynamic behavior of COVID-19 pandemic in China. For this purpose, the Inverse Reliability Analysis was considered. In the proposed methodology, two loops are performed. The inner optimization loop is used to find the solution associated with the highest probability value and the outer loop is the regular optimization loop used to determine the vector of design variables. The obtained results demonstrated that small quantities affected the objective function value, i.e., the increase of the reliability parameter implies on an increase of the objective function value. It is worth mentioning that the probability distribution of the system outputs should be known to apply the proposed approach, as required by other reliability based optimization methods.

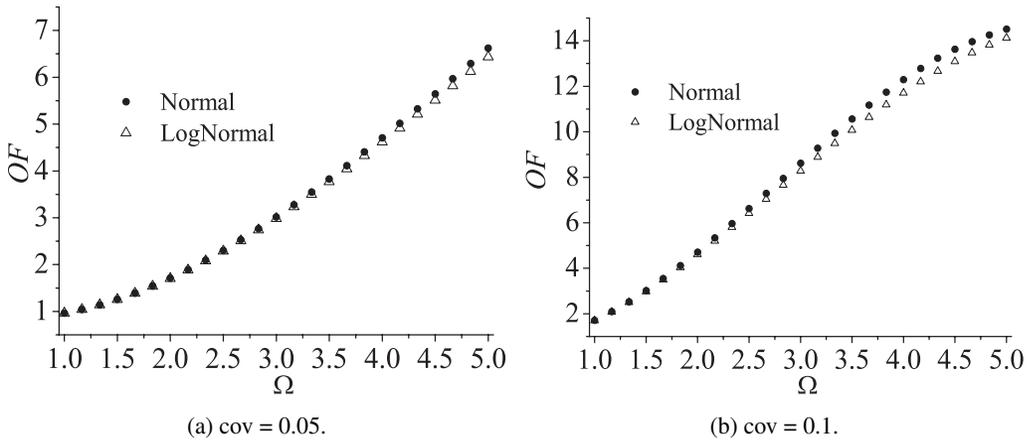(a) cov = 0.05.   (b) cov = 0.1.

Figure 5: Influence of the type of distribution and the coefficient of variation considering the solution obtained by IRA.

Finally, it is worthwhile to mention that the insertion of a reliability analysis technique in the formulation of the inverse problem implies increasing the value of the objective function, i.e., larger deviations between the simulated and reported profiles are observed. However, a reliable solution, unlike a deterministic solution, is less sensible to small perturbations. Thus, the study of uncertainties in the formulation and solution of inverse problems—especially in a core parameter such as $\mathscr{R}_0$—is justified in terms of this less sensibility of design variables.

Acknowledgement

RESUMO. Diferentes tipos de modelos matemáticos têm sido usados para prever o comportamento dinâmico do novo coronavírus (COVID-19). Muitos deles envolvem a formulação e solução de problemas inversos. Esse tipo de problema geralmente considera que o modelo, o vetor de variáveis de projeto e os parâmetros do sistema são todos valores determinísticos. Nesta contribuição, uma metodologia baseada em um processo de iteração de *loop* duplo é considerada para avaliar a influência de incertezas durante a resolução do problema inverso, proposto para estimar os parâmetros do modelo. No *loop* intermo, a solução associada ao maior valor de probabilidade associada as variáveis que apresentam incerteza é determinada. Já no *loop* externo, os valores das variáveis determinísticas é obtida. Para essa finalidade, a Análise de Confiabilidade Inversa é considerada como ferramenta de confiabilidade e o algoritmo de Evolução Diferencial é usado como

ferramenta de otimização. Para fins de ilustração, a metodologia proposta é aplicada para estimar os parâmetros do modelo SIRD (Susceptível- Infeccioso-Recuperado-Morto) associado ao comportamento dinâmico da pandemia de COVID-19, considerando dados reais da epidemia da China e incertezas no número básico de reprodução ($\mathscr{R}_0$). Os resultados obtidos demonstram que o aumento da confiabilidade implica no aumento do valor da função objetivo.

Palavras-chave: problema inverso, otimização baseada em confiabilidade, modelagem, COVID-19.

REFERENCES

[1] C. Anastassopoulou, L. Russo, A. Tsakris & C. Siettos. Data-based Analysis, Modelling and Forecasting of the COVID-19 Outbreak. *PLoS ONE*, **3**(15) (2020), 1–21.

[2] J.C. Blackwood & L.M. Childs. An Introduction to Compartmental Modeling for the Budding Infectious Disease Modeler. *Letters in Biomathematics*, **5**(1) (2018), 195–221. doi:10.1080/23737867. 2018.1509026.

[3] S. Chatterjee, A. Sarkar, S. Chatterjee, M. Karmakar & R. Paul. Studying the progress of COVID-19 outbreak in India using SIRD model. *Indian Journal of Physics and Proceedings of the Indian Association for the Cultivation of Science (2004)*, **23** (2020), 1–17.

[4] E. Dong, H. Du & L. Gardner. An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases*, **20**(5) (2020), 533–534. doi:10.1016/S1473-3099(20)30120-1.

[5] A. Ghosh, S. Das & A.K. Das. A simple two-phase differential evolution for improved global numerical optimization. *Soft Computing*, **24** (2020), 6151–6167.

[6] A. Haldar & S. Mahadevan. "Probability, Reliability, and Statistical Methods in Engineering Design", volume 1. John Wiley & Sons, Chichester, 1 ed. (2000).

[7] J.M. Heffernan, R.J. Smith & L.M. Wahl. Perspectives on the Basic Reproductive Ratio. *Journal of the Royal Society, Interface*, **2**(4) (2005), 281–293. doi:doi:10.1098/rsif.2005.0042.

[8] Johns Hopkins Resource Center. "Mapping 2019-nCoV" (2020). URL `https://systems.jhu.edu/research/public-health/ncov/`. Accessed October 22, 2020.

[9] M.J. Keeling & P. Rohani. "Modeling Infection Diseases in Humans and Animals". Princeton University Press (2005).

[10] G.B. Libotte, F.S. Lobato, G.M. Platt & A.J.S. Neto. Determination of an optimal control strategy for vaccine administration in COVID-19 pandemic treatments. *Computer Methods and Programs in Biomedicine*, **196**(November) (2020), 105664.

[11] F.S. Lobato, M.S. Gonçalves, B. Jahn, A.A. Cavalini Jr & V. Steffen Jr. Reliability-Based Optimization using Differential Evolution and Inverse Reliability Analysis for Engineering System Design. *Journal of Optimization Theory and Applications*, **1**(1) (2017), 1–33.

[12] Nature Index. "Coronavirus research publishing: The rise and rise of COVID-19 clinical trials" (2020). URL `https://www.natureindex.com/news-blog/the-top-coronavirus-research-articles-by-metrics`. Accessed August 1, 2020.

[13] E. Nikolaidis & R. Burdisso. Reliability Based Optimization: a Safety Index Approach. *Computers & Structures*, **28**(6) (1988), 781–788. doi:https://doi.org/10.1016/0045-7949(88)90418-X.

[14] J. Riou & C.L. Althaus. Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-nCoV), December 2019 to January 2020. *Eurosurveillance*, **25**(4) (2020), 1–5.

[15] T.G. Ritto, R. Sampaio & E. Cataldo. Timoshenko beam with uncertainty on the boundary conditions. *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, **30**(4) (2008), 295–303.

[16] M. Rosenblatt. Remarks on a Multivariate Transformation. *The Annals of Mathematical Statistics*, **23**(3) (1952), 470–472. doi:10.1214/aoms/1177729394.

[17] S. Sanche, Y. Lin-Ting, C. Xu, E. Romero-Severson, N. Hengartner & R. Ke. High Contagiousness and Rapid Spread of Severe Acute Respiratory Syndrome Coronavirus 2. *Emerging Infectious Diseases*, **26**(7) (2020). doi:10.3201/eid2607.200282.

[18] R. Storn & K. Price. Differential Evolution: a simple and efficient adaptive scheme for global optimization over continuous spaces. *Technical Report TR-95-012, International Computer Science Institute, Berkeley*, **2**(2) (1995), 1–17.

[19] R. Storn, K. Price & J.A. Lampinen. "Differential Evolution - A Practical Approach to Global Optimization". Springer - Natural Computing Series (2005).

[20] M. Trawicki. Deterministic SEIRs Epidemic Model for Modeling Vital Dynamics, Vaccinations, and Temporary Immunity. *Mathematics*, **5**(1) (2017), 7. doi:10.3390/math5010007.

[21] J. Tu, K.K. Choi & Y.H. Park. A New Study on Reliability-Based Design Optimization. *Journal of Mechanical Design*, **121**(4) (1999), 557–564. doi:10.1115/1.2829499.

[22] G.N. Vanderplaats. "Numerical Optimization Techniques for Engineering Design". Vanderplaats Research and Development, Inc., 3rd ed. (2001), 449 pp.

[23] H.J. Wearing, P. Rohani & M.J. Keeling. Appropriate Models for the Management of Infectious Diseases. *PLoS Medicine*, **2**(7) (2005). doi:10.1371/journal.pmed.0020174.

[24] World Health Organization. "Naming the Coronavirus Disease (COVID-19) and the Virus that Causes it" (2020). URL `https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it`. Accessed October 22, 2020.

[25] Y.T. Wu. Computational Methods for Efficient Structural Reliability and Reliability Sensitivity Snalysis. *AIAA Journal*, **32**(8) (1994), 1717–1723. doi:10.2514/3.12164.

[26] Y.T. Wu, H.R. Millwater & T.A. Cruse. Advanced Probabilistic Structural Analysis Method for Implicit Performance Functions. *AIAA Journal*, **28**(9) (1990), 1663–1669. doi:10.2514/3.25266.