Predicting the spread of epidemiological diseases by using a multi-objective algorithm

Cite as: AIP Conference Proceedings **2070**, 020016 (2019); https://doi.org/10.1063/1.5089983
Published Online: 12 February 2019

M. R. Ferrández, B. Ivorra, J. L. Redondo, A. M. Ramos, and P. M. Ortigosa

Predicting The Spread Of Epidemiological Diseases By Using A Multi-Objective Algorithm

M. R. Ferrández[1,a)], B. Ivorra[2,b)], J. L. Redondo[1,c)], A. M. Ramos[2,d)] and P. M. Ortigosa[1,e)]

[1]*Dept. of Informatics, Universidad de Almería, ceiA3, Ctra. Sacramento, La Cañada de San Urbano, 04120 Almería, Spain.*
[2]*Instituto de Matemática Interdisciplinar (IMI) & Dept. de Matemática Aplicada, Universidad Complutense de Madrid, Plaza de las Ciencias, 3, 28040 Madrid, Spain.*

[a)]Corresponding author: mrferrandez@ual.es
[b)]ivorra@mat.ucm.es
[c)]jlredondo@ual.es
[d)]angel@mat.ucm.es
[e)]ortigosa@ual.es

Abstract. The epidemiological models are able to predict the spread of diseases, but a previous work on calibrating some involved parameters must be done. In this work, we propose a methodology to adjust those parameters based on solving a multi-objective optimization problem whose objective functions measure the accuracy of the model. More precisely, we have considered the Between-Countries Disease Spread model because it involves a set of countries taking into account the migratory movements among them. As a result, using some real data about the number of detected cases and the number of deaths for the Ebola virus disease, we have shown that the methodology is able to find a set of values for the parameters so that the model fits the outbreak spread for a set of countries.

INTRODUCTION

In 2014, the Ebola virus outbreak led to a serious concern about the authorities capacity for predicting and controlling the epidemic diseases and their spread between countries. In this context, some new approaches for modelling those situations emerged, as the Between-Countries Disease Spread (Be-CoDiS) proposed at [1]. In spite of the accurate predictions achieved with this model, it highlighted the challenge of adjusting the values for the involved epidemiological parameters. Those parameters depend on the characteristics of each country as, for instance, its economic development and its demography. Furthermore, they may be time dependent.

This work aims to achieve a global fitting methodology, where a set of countries linked by their migratory movements are considered. To do that, the Be-CoDiS model [1] is used and a multi-objective problem is defined. When the spread of the disease is numerically simulated according to this model, it returns the evolution of the infection. Since we want that this evolution meets as close as possible the real one, the objective functions are formulated as the differences between those predicted and real data.

THE EPIDEMIOLOGICAL MODEL: Be-CoDiS

The deterministic model called Between-Countries Disease Spread (Be-CoDiS) [1] describes the evolution of the outbreak in a group of $N_{\text{co}} \in \mathbb{N}$ countries taking into account the migratory fluxes between them. It is a compartmental-based model, such that the population of each country $i \in \{0, 1, \ldots, N_{\text{co}}\}$ is classified into the following disjoint states: susceptible, infected, infectious, hospitalized, recovered, death and buried. The number of people in the country i at time t belonging to each of those states are denoted, respectively, by $S(i,t)$, $E(i,t)$, $I(i,t)$, $H(i,t)$, $R(i,t)$, $D(i,t)$ and

Proceedings LeGO – 14th International Global Optimization Workshop
AIP Conf. Proc. 2070, 020016-1–020016-4; https://doi.org/10.1063/1.5089983
Published by AIP Publishing. 978-0-7354-1798-4/$30.00

$B(i,t)$. Considering $\mathrm{NP}(i,t)$ as the total number of persons in the country i at time t, then $\mathrm{NP}(i,t) = S(i,t) + E(i,t) + I(i,t) + H(i,t) + R(i,t)$.

This Be-CoDiS model, which was proposed and validated in [1, 2], consists in the following equations:

$$
\begin{aligned}
\frac{\mathrm{d}S(i,t)}{\mathrm{d}t} =& -\frac{S(i,t)\Big(m_I(i,t)\beta_I(i)I(i,t) + m_H(i,t)\beta_H(i)H(i,t) + m_D(i,t)\beta_D(i)D(i,t)\Big)}{\mathrm{NP}(i,t)} \\
& - \mu_m(i)S(i,t) + \mu_n(i)\Big(S(i,t) + E(i,t) + I(i,t) + H(i,t) + R(i,t)\Big) \\
& + \sum_{i\neq j} m_{\mathrm{tr}}(j,i,t)\tau(j,i)S(j,t) - \sum_{i\neq j} m_{\mathrm{tr}}(i,j,t)\tau(i,j)S(i,t), \\
\frac{\mathrm{d}E(i,t)}{\mathrm{d}t} =& \frac{S(i,t)\Big(m_I(i,t)\beta_I(i)I(i,t) + m_H(i,t)\beta_H(i)H(i,t) + m_D(i,t)\beta_D(i)D(i,t)\Big)}{\mathrm{NP}(i,t)} \\
& - \mu_m(i)E(i,t) + \sum_{i\neq j} m_{\mathrm{tr}}(j,i,t)\tau(j,i)\chi_{\epsilon_{\mathrm{fit}}}(E(j,t)) \\
& - \sum_{i\neq j} m_{\mathrm{tr}}(i,j,t)\tau(i,j)\chi_{\epsilon_{\mathrm{fit}}}(E(i,t)) - \gamma_E(i,t)\chi_{\epsilon_{\mathrm{fit}}}(E(i,t)), \\
\frac{\mathrm{d}I(i,t)}{\mathrm{d}t} =& \gamma_E(i,t)\chi_{\epsilon_{\mathrm{fit}}}(E(i,t)) - \Big(\mu_m(i) + \gamma_I(i,t)\Big)I(i,t), \\
\frac{\mathrm{d}H(i,t)}{\mathrm{d}t} =& \gamma_I(i,t)I(i,t) - \Big(\mu_m(i) + (1-\omega(i,t))\,\gamma_{HR}(i,t) + \omega(i,t)\gamma_{HD}(i,t)\Big)H(i,t), \\
\frac{\mathrm{d}R(i,t)}{\mathrm{d}t} =& (1-\omega(i,t))\,\gamma_{HR}(i,t)H(i,t) - \mu_m(i)R(i,t), \\
\frac{\mathrm{d}D(i,t)}{\mathrm{d}t} =& \omega(i,t)\gamma_{HD}(i,t)H(i,t) - \gamma_D(i,t)D(i,t), \\
\frac{\mathrm{d}B(i,t)}{\mathrm{d}t} =& \gamma_D(i,t)D(i,t),
\end{aligned}
\tag{1}
$$

where $\mu_n(i)$ is the birth rate (day^{-1}) and $\mu_m(i)$ is the mortality rate (day^{-1}) for the country i meaning the number of births and of deaths, respectively, per day and per capita. The disease effective contact rates $\beta_I(i), \beta_H(i)$, and $\beta_D(i)$ are constant parameters for each country i representing the mean number of contacts transmitting the disease of a person in states I, H, and D, respectively, per day before applying the control measures. The number of people moving from states E to I, I to H, H to D, H to R and D to B per day and per capita are denoted by $\gamma_E(i,t), \gamma_I(i,t), \gamma_{HD}(i,t), \gamma_{HR}(i,t)$, and $\gamma_D(i,t)$, respectively, as they are functions depending on the country but also on time because they varies with the application of the control measures. For describing the efficiency of these control measures, the following decreasing functions have been considered:

$$
m_I(i,t) = m_H(i,t) = m_D(i,t) = \exp\Big(-\kappa(i)\max(t-\lambda(i), 0.0)\Big).
$$

Notice that these functions multiply the disease contact rates in the System 1, such that the number of effective contacts is reduced as the control measures efficiency is improved. This reduction is managed through the parameter $\kappa(i) \in [0.0, +\infty)$ (day^{-1}), while the parameter $\lambda(i)$ refers to the first day of application of those control measures in the country i.

Furthermore, in System 1, $\omega(i,t)$ is the disease fatality rate representing the percentage of people who do not survive the disease for each country i at each time t. Finally, the migratory movements have been considered by means of the matrix $(\tau(i,j))_{i,j=1}^{N_{\mathrm{co}}}$, that is composed by the transfer rates (day^{-1}) of persons from one country i to another j expressed in % of population in i per unit of time. Those rates can be also diminished by the progressive application of the control measures, then they are multiplied by the function $m_{\mathrm{tr}}(i,j,t)$, which is the product of the efficiency of the control measures in both i and j involved countries: $m_{\mathrm{tr}}(i,j,t) = m_I(i,t) \cdot m_I(j,t)$.

Notice that the following filter function has been used in System 1 in order to avoid artificial spread of the disease due to negligible values of $E(i,t)$: $\chi_{\epsilon_{\mathrm{fit}}}(x) = x$ if $x \geq \epsilon_{\mathrm{fit}}$, $\chi_{\epsilon_{\mathrm{fit}}}(x) = 2x - \epsilon_{\mathrm{fit}}$ if $\epsilon_{\mathrm{fit}}/2 \leq x \leq \epsilon_{\mathrm{fit}}$, and 0 otherwise, where $\epsilon_{\mathrm{fit}} \geq 0$ is a small tolerance parameter.

OPTIMIZATION FOR ESTIMATING EPIDEMIOLOGICAL PARAMETERS

The World Health Organization (WHO) provides periodical reports about diseases of interest (such as the Ebola Virus Disease) concerning the cumulative number of cases and the cumulative number of deaths in the affected countries. Thus, assuming that those real data are known for the N_{co} considered countries at N_h dates, we denote them by $\{\text{CC}_{\text{real}}(i,t_j)\}_{j=0}^{j=N_h}$ and by $\{\text{CD}_{\text{real}}(i,t_j)\}_{j=0}^{j=N_h}$, respectively, where $i \in \{1,\ldots,N_{\text{co}}\}$ refers to each country. This information is also managed to obtain the initial conditions for the epidemiological model. Then, when the model is used for simulating the outbreak with a set ϕ of values for configuring the epidemiological parameters of interest, it returns a predicted evolution for the cumulative number of cases $CC^{\phi}(i,t)$ and for the cumulative number of deaths $CD^{\phi}(i,t)$. In particular, for the Be-CoDiS model, they are computed as:

$$CC^{\phi}(i,t) = CC(i,0) + \int_0^t \gamma_I(i,t)\cdot I^{\phi}(i,t)\mathrm{d}t, \qquad CD^{\phi}(i,t) = CD(i,0) + \int_0^t \omega(i,t)\cdot\gamma_{HD}(i,t)\cdot H^{\phi}(i,t)\mathrm{d}t, \tag{2}$$

where $CC(i,0)$ and $CD(i,0)$ are the initial number of cases and of deaths available in the disease reports.

Since the goal is that this evolution fits as close as possible the real one, the objective functions are formulated as the differences between those predicted and real data for each country $i \in \{1,\ldots,N_{\text{co}}\}$ as follows:

$$f_i(\phi) = \frac{\left\|\text{CC}_{\text{real}}(i,t_{\text{f}}) - CC^{\phi}(i,t_{\text{f}})\right\|_{L^2}}{\|\text{CC}_{\text{real}}(i,t_{\text{f}})\|_{L^2}}, \qquad f_{2i}(\phi) = \frac{\left\|\text{CD}_{\text{real}}(i,t_{\text{f}}) - CD^{\phi}(i,t_{\text{f}})\right\|_{L^2}}{\|\text{CD}_{\text{real}}(i,t_{\text{f}})\|_{L^2}}, \tag{3}$$

where the errors have been computed using the L^2 norm: $\|g(T)\|_{L^2} = \left(\int_0^T (g(t))^2\,\mathrm{d}t\right)^{1/2}$.

Therefore, the considered multi-objective problem is:

$$\begin{cases} \min & f_i(\phi), \quad \forall i \in \{1,\ldots,N_{\text{co}}\} \\ \min & f_{2i}(\phi), \quad \forall i \in \{1,\ldots,N_{\text{co}}\} \\ \text{s.t.} & \phi \in \Phi, \end{cases} \tag{4}$$

where ϕ denotes the set of parameters of the model to be estimated, Φ is the feasible set or search space given by the ranges of those parameters.

For solving the multi-objective optimization problem, we use the parallel version [3] of the algorithm called Weighting Achievement Scalarizing Function Genetic Algorithm (WASF-GA) [4]. It allows obtaining a good set of trade-off solutions focusing on a region of the objective space in a reduced computational time due to the use of High-Performance Computing. As a genetic algorithm, WASF-GA explores the search space by using both crossover and mutation methods to generate new points. In particular, in this work, the Simulated Binary Crossover (SBX) and the polynomial mutation have been considered. Among the Preference-Based Multi-Objective Evolutionary Algorithms (PMOEAs) family, to which WASF-GA belongs to, it has demonstrated to be competitive to deal with three or more objectives.

COMPUTATIONAL EXPERIMENTS AND RESULTS

The proposed methodology for fitting epidemiological parameters has been applied to the Ebola virus outbreak in 2014/2015. Several versions of the Problem 4, varying the number of objectives and decision variables, have been solved focusing on the prediction of the spread in those countries with the highest number of infected cases.

Despite the model can handle a set of countries, the first experiments were devoted to testing the fitting methodology when the target is to determine the parameters for a single country. In those simpler cases, Problem 4 involves only two objectives. Solving it with WASF-GA, we obtain a set of trade-off points among which the epidemiologists can select the point that better satisfies their interests. In particular, we have distinguished three strategies: (i) selecting the solution that provides the lower value for the first objective, i.e., the one that better fits the cumulative number of cases, (ii) choosing the one that exhibits the best compromise among the two objectives (with the Euclidean norm), and (iii) taking the point giving the lower value for the second objective, i.e., the one which better fits the cumulative number of deaths. Since WASF-GA is a meta-heuristic algorithm, several executions have been carried out in order to study the robustness of the results. In particular, 30 repetitions have been performed for the one-country cases. For each run, the parameters' values corresponding to the point selected with each strategy have been registered. Then, the

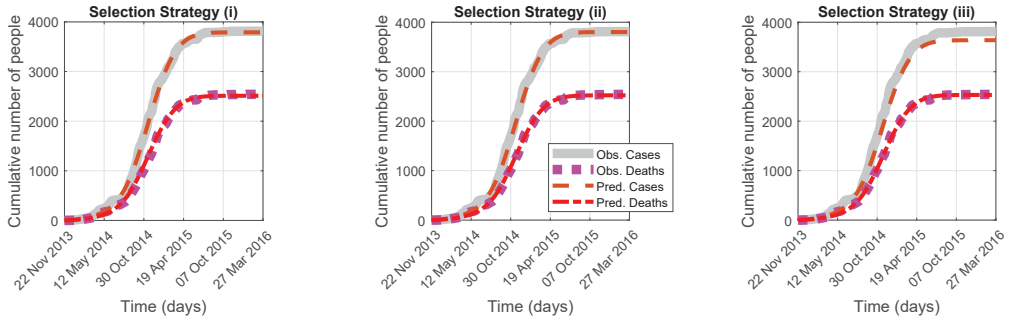

FIGURE 1. Evolution of the cumulative numbers of Ebola cases and deaths in Guinea: observed real data (thick lines) and Be-CoDiS predicted data (thin lines).

average and the standard deviation values for each parameter have been computed. From this study, we have obtained that the variation of those values is not significant and, therefore, our methodology is robust. For instance, the average of Guinea's effective contact rate for the infected people considering the strategy (ii) is $\beta_I = 0.1907$ and its standard deviation is 4.99E-04. Furthermore, considering all the WASF-GA repetitions, the parameters' values providing the worst objective values for each strategy have been identified. In Figure 1, the real evolution of the Ebola disease in Guinea is compared to the evolution predicted by the Be-CoDiS model using those parameters' values. As can be seen, even considering the worst WASF-GA outcome, our methodology is able to configure the epidemiological parameters such that the model fits quite accurate the real outbreak. In fact, for the strategy (ii), the relative errors computed by (2) are 1.92E-2 for the number of cases and 1.70E-2 for the number of deaths.

Next, the fitting methodology has been applied to a set of 176 countries. For this case, the Problem 4 has been considered with the objective functions given by (2) for the three countries where the Ebola virus had more virulence: Guinea, Liberia and Sierra Leone. Two more objective functions have been included for the cumulative cases and the cumulative deaths consisting in the average of their absolute error for the remaining countries. Additionally, to avoid failures in the detection of infected countries, one objective function counting the number of infected countries that the model predicts as non-infected and another one for the number of non-infected countries predicted as infected have been added to complete the Problem 4. The obtained results have been analyzed as in the one-country experiments and we can conclude that the methodology succeeds at finding a set of values for the Be-CoDiS model parameters which allows to accurately describe the spread in all the considered countries.

ACKNOWLEDGMENTS

This research has been funded by grants from the Spanish Ministry of Economy and Competitiveness (TIN2015-66680-C2-1-R and MTM2015-64865P); Junta de Andalucía (P12-TIC301), in part financed by the European Regional Development Fund (ERDF). Juana L. Redondo is a fellow of the Spanish "Ramón y Cajal" contract program, co-financed by the European Social Fund. This work used EPCC's Cirrus HPC Service (https://www.epcc.ed.ac.uk/cirrus) under the Project HPC-EUROPA3 (INFRAIA-2016-1-730897), with the support of the EC Research Innovation Action under the H2020 Programme.

REFERENCES

[1] B. Ivorra, D. Ngom, and A. M. Ramos, Bulletin of Mathematical Biology **77**, 1668–1704Sep (2015).
[2] B. Ivorra, D. Ngom, and A. M. Ramos, Biomath Communications Supplement **4** (2017).
[3] M. R. Ferrández, S. Puertas-Martín, J. L. Redondo, B. Ivorra, A. M. Ramos, and P. M. Ortigosa, The Journal of Supercomputing 1–16 (2018).
[4] A. B. Ruiz, R. Saborido, and M. Luque, Journal of Global Optimization **62**, 101–129 (2015).