Research paper

A multi-objective approach to identify parameters of compartmental epidemiological models—Application to Ebola Virus Disease epidemics

Miriam R. Ferrández [a],[*],[1], Benjamin Ivorra [b],[1], Juana L. Redondo [a],[1], Ángel M. Ramos [b],[1], Pilar M. Ortigosa [a],[1]

[a] *Department of Computer Science, University of Almería, ceiA3, Ctra. Sacramento, La Cañada de San Urbano, 04120 Almería, Spain*
[b] *Interdisciplinary Mathematics Institute (IMI) and Department of Applied Mathematics and Mathematical Analysis, Complutense University of Madrid, Plaza de las Ciencias, 3, 28040 Madrid, Spain*

ARTICLE INFO

ABSTRACT

In this work, we propose a novel methodology to identify parameters of compartmental epidemiological models. It is based on solving a multi-objective optimization problem that consists in fitting some of the model outputs to real observations. First, according to the available data of the considered epidemic, we define a multi-objective optimization problem where the model parameters are the optimization variables. Then, this problem is solved by considering a particular optimization algorithm called ParWASF-GA (Parallel Weighting Achievement Scalarizing Function Genetic Algorithm). Finally, the decision maker chooses, within the set of possible solutions, the values of parameters that better suit his/her preferences. In order to illustrate the benefit of using our approach, it is applied to estimate the parameters of a deterministic epidemiological model, called Be-CoDiS (Between-Countries Disease Spread), used to forecast the possible spread of human diseases within and between countries. We consider data from different Ebola outbreaks from 2014 up to 2019. In all cases, the proposed methodology helps to obtain reasonable predictions of the epidemic magnitudes with the considered model.

1. Introduction

The COVID-19 pandemic has highlighted the importance of epidemiology, and in particular mathematical epidemiology, in the current globalized world. According to [1], epidemiology mainly addresses on (i) describing the distribution of the disease, (ii) identifying the risk factors that cause it, (iii) building and testing theories, and (iv) planning, implementing and evaluating prevention, detection and control programs. Inside the area of epidemiology, mathematical modeling has demonstrated to be a useful tool to simulate the spatial and temporal evolution of diseases, evaluate the effect of control measures and, therefore, design preventive plans [2–4]. Epidemiological models are able to estimate epidemics behavior and to make interesting predictions of outbreaks by using some known information about the disease and/or the affected areas (such as economical or ecological situation) [5–7]. Thus, they can be part of the decision process of authorities,

---

* Corresponding author.
 E-mail addresses: mrferrandez@ual.es (M.R. Ferrández), ivorra@mat.ucm.es (B. Ivorra), jlredondo@ual.es (J.L. Redondo), angel@mat.ucm.es (Á.M. Ramos), ortigosa@ual.es (P.M. Ortigosa).
 [1] All authors contributed equally to the work.

such as the World Health Organization (WHO), to elaborate and coordinate response strategies in case of epidemiological emergencies [8,9].

However, before applying an epidemiological model, some previous work must be done, in particular, to adjust the model parameters [1,7]. In addition to features proper to the areas and disease of interest, this calibration stage must be based on data and observations which are already available about the disease dynamics. Nowadays, this information includes data that are periodically-reported by national and/or international authorities [10,11]. As a consequence, the model should be frequently updated by recalibrating those parameters with recent data. However, adjusting parameters could not be a trivial and automatized task because the solutions may highly depend on the precision, validity or representativeness of available data, and they can also deeply variate during the epidemic [12,13].

In the literature, to the best of our knowledge, there is not a common general procedure to obtain parameters values for epidemiological models. For instance, in [14], some parameters concerning the Hepatitis C Virus Epidemic in France are estimated using a weighted least-squares criterion to fit the number of deaths observed during historical outbreaks. In [15], a non-linear least-squares fitting is performed to obtain the optimal parameters values of a compartment epidemiological model used to study the spread of rumors on Twitter. In [16,17], the authors propose an optimization approach based on the epidemic simulation to estimate the optimal parameters for an influenza disease model. More precisely, they minimize the difference between the simulated and observed daily number of infected people by using a direct search optimization method. Additionally, in some papers, Uncertainty Quantification methods are applied to determine the values of epidemiological model parameters [18–21]. In general, those methods employ probability distributions to represent the uncertainty in the model parameters and to estimate model input parameters from observation data. They also use Sensitivity Analysis techniques to obtain confidence intervals of the parameters and study which factors most influence the model output.

In this work, we propose an original general methodology to adjust parameters of epidemiological compartmental models. This methodology is based on solving a multi-objective optimization problem defined as the minimization of several functions that measure the errors between the model outputs and real data. As a result, we obtain a set of solutions offering different valid commitments between the objectives. Finally, as the last step, the most suitable configuration of parameters among the whole set of solutions is selected, according to the preferences of the *Decision Maker* (DM). Namely, the person (or persons) in charge of analyzing the solutions and choosing the preferences. In our context, the DM can be, for instance, the epidemiologist or the authorities interested in modeling the epidemic.

To carry out the performance analysis of the proposed methodology, we have selected a parallel version of an optimization method, the preference-based multi-objective algorithm WASF-GA [22,23]. It has demonstrated to be reliable and robust in previous studies as the ones accomplished for designing food treatments [23–25]. As epidemiological model, we have considered Be-CoDiS (Between-Countries Disease Spread) [7], which simulates the spread of some human diseases within and between countries. Finally, as the optimization objective, we have defined a minimization problem for calibrating Be-CoDiS to real data of several Ebola Virus Disease (EVD) outbreaks occurring between 2014 and 2019.

The 2014–2016 outbreak in West Africa was the largest EVD epidemic to date. Indeed, it was declared as an international emergency by the WHO [11]. It started in 2014 spreading from Guinea to Liberia and Sierra Leone, with a large magnitude of cases, and to Nigeria, Mali, Senegal, the USA, the United Kingdom, Italy, and Spain, with a limited number of cases. Officially, the epidemic finished the 29th of March, 2016. Nevertheless, two EVD outbreaks have recently occurred in the Democratic Republic of the Congo (DR Congo) in 2018 and 2019.

In this paper, we estimate the parameters of Be-CoDiS, for those particular EVD outbreaks by using the proposed methodology. Results are compared to the ones reported in [7]. In that work, published in the middle of 2015, the epidemiological model Be-CoDiS was also used for the EVD epidemic, obtaining accurate forecasts. However, as reported in the paper, adjusting the large number of epidemiological parameters to calibrate the model and fit it to the observed data was very difficult. Note that in that work, a time-consuming trial and error procedure was used to perform this task.

We note that the multi-objective approach proposed here to estimate parameters of compartmental epidemiological models has been recently applied to the COVID-19 pandemic considering several θ-SEIHRD models [26–28]. Its successful application has allowed us to evaluate different scenarios of the COVID-19 epidemic in several countries, such as China and Italy, and to extract some valuable information about the impact on the epidemic dynamics of the undetected infections, the control measures, the vaccination campaigns and the new variants of the virus. However, in those works, our approach was used as a black-box tool and no detailed description of this methodology was given.

This work is organized as follows. First, in Section 2, basic concepts about multi-objective optimization and the considered preference-based optimization algorithm are briefly introduced. Then, in Section 2.3, the methodology proposed to adjust the parameters of epidemiological models is detailed. In Section 3, our approach is applied to estimate the parameters of the Be-CoDiS model for EVD outbreaks. Finally, in Section 4, the proposed methodology is validated by considering some computational experiments based on data observed during recent EVD epidemics.

2. A general method for parameter estimation

From a general point of view, we propose a methodology to estimate the parameters of epidemiological models, based on the formulation and resolution of a multi-objective optimization problem. Therefore, to settle the framework of this work, we first recall some essential multi-objective optimization concepts. Then, we describe the optimization algorithm used for solving the resulting multi-objective problems. Finally, we describe step by step our method, focusing on (i) the identification of the model parameters to estimate, (ii) the formulation of the multi-objective problem to solve, and (iii) the extraction of a suitable configuration for the model parameters from the obtained solutions.

2.1. Basic definitions and concepts

Multi-objective optimization deals with problems consisting in minimizing (or maximizing) concurrently several functions. Those problems are called *multi-objective problems* and are generally formulated as:

$$\begin{cases} \min f_1(\mathbf{x}) \\ \min f_2(\mathbf{x}) \\ \dots \\ \min f_m(\mathbf{x}) \\ \text{subject to } \mathbf{x} \in S \subset \mathbb{R}^n, \end{cases} \tag{1}$$

where $f_i : \mathbb{R}^n \longrightarrow \mathbb{R}$, for $i = 1, \dots, m$, are the *objective functions*, $S \subset \mathbb{R}^n$ is known as the *feasible region* and $\mathbf{x} = (x_1, \dots, x_n) \in S$ is the *decision vector*, with $n \in \mathbb{N}$. Additionally, the image vectors $\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), \dots, f_m(\mathbf{x}))$ are called the *objective vectors*. We recall that $\mathbb{R}$ denotes the set of real numbers and $\mathbb{N}$ is the set of natural numbers.

In most multi-objective problems, a single decision vector minimizing all the objective functions simultaneously does not exist because those functions contradict each (or some) other. Then, the solution to these problems consists of several decision vectors offering different optimal compromises among the objectives.

The set of decision vectors solving a multi-objective problem is called the *Pareto optimal set*. Those decision vectors are known as *efficient vectors* or *Pareto optimal solutions*. An efficient vector is defined as a decision vector $\mathbf{x} \in S$ such that none of its objective function values can be enhanced without worsening at least one of the others. In order to express this optimality relationship among the decision vectors, the concept of dominance is introduced.

Definition 1. Given two decision vectors $\mathbf{x}, \mathbf{x}' \in S$, $\mathbf{x}$ *dominates* $\mathbf{x}'$ if and only if $f_i(\mathbf{x}) \leq f_i(\mathbf{x}')$ for all $i = 1, \dots, m$, and there exists one $j \in \{1, \dots, m\}$ such that $f_j(\mathbf{x}) < f_j(\mathbf{x}')$. In that case, we also say that $\mathbf{f}(\mathbf{x})$ *dominates* $\mathbf{f}(\mathbf{x}')$.

Therefore, the solution of a multi-objective problem is the Pareto optimal set denoted by $S_E \subset S$, which is formed by all the efficient vectors (i.e., the set of nondominated decision vectors, meaning those such that there does not exist another vector in S dominating them). Additionally, the set of all the nondominated objective vectors $\mathbf{f}(S_E) \subset \mathbf{f}(S)$ is called the *Pareto optimal front*.

In most practical cases (see, e.g., [24,29]), an analytical expression describing the Pareto optimal set and its corresponding Pareto optimal front cannot be obtained. Thus, the most extended approach is the use of multi-objective optimization (MOO) algorithms that provide a *Pareto front approximation* formed by a finite number of points that aim to be close to the exact Pareto optimal front and well distributed to give an accurate description of this front.

In practice, the number of points composing the Pareto front approximation must be high enough to evenly cover the whole Pareto optimal front. However, managing such an amount of points is not always operative due to limitations in computational time. Then, in the last decades, a family of MOO algorithms focusing on the approximation of only a part of the Pareto optimal front has arisen. Indeed, in practice, the DM is commonly not interested in knowing the entire Pareto front but only a so-called *Region Of Interest* (ROI). These algorithms are called *Preference-based Multi-Objective Evolutionary Algorithms* (PMOEAs) [30,31]. In most PMOEAs, the information about the DM preferences is given by using a point in the objective space $\mathbf{f}(S)$ that indicates an estimation of the desired values for the objective functions. This point is called *Reference Point* (RP).

A detailed description of the concepts introduced here can be found in [32–35] for MOO and in [30,36] for PMOEAs.

2.2. Considered optimization algorithm

To solve the multi-objective optimization problems formulated later in Section 2.3, the PMOEA called *Weighting Achievement Scalarizing Function Genetic Algorithm* (WASF-GA) is applied [22]. Here, we briefly recall the basis of this particular algorithm.

From a general point of view, WASF-GA is a population-based evolutionary algorithm that explores the feasible region looking for efficient solutions using an iterative procedure. At each iteration, it performs reproduction and replacement mechanisms over a population of candidate solutions, called individuals. In the reproduction stage, a new offspring population is created by applying some mechanisms inspired by the Darwinian theory, namely selection, crossover and mutation operators (see [37] for more details). The replacement is performed to compose the population of the next generation, among the individuals of the current population and the new ones obtained after the reproduction phase. In most algorithms, this replacement method is based on dominance, i.e., it prioritizes choosing solutions that are nondominated by the other individuals of the population. However, in WASF-GA, its replacement mechanism selects the closest individuals to a given Reference Point (RP), which represents the preferences of the DM. As a result, WASF-GA obtains a set of solutions covering a region of the Pareto optimal front (ROI), which is induced by the RP. The DM must specify, as an input parameter, the number of required points in the ROI. See [22–24] for a detailed description of the algorithm and its parameters.

The computational time of this optimization procedure depends not only on some WASF-GA input parameters, such as the number of required points in the ROI and the number of iterations, but also on the time spent in the evaluation of the

objective functions. When the number of variables or compartments (i.e. countries) involved in the outbreak is large, it might be convenient to distribute the objective function evaluations among different processing units using High-Performance Computing (HPC) techniques. Thus, to accelerate the optimization procedure of WASF-GA, we propose to use its parallel version called ParWASF-GA, which is described in [23]. ParWASF-GA manages several threads that access shared memory to perform the creation and evaluation of individuals of a new population concurrently.

2.3. A general method to estimate model parameters

Now, we introduce a novel and general methodology based on formulating and solving a multi-objective optimization problem, to adjust some parameters of epidemiological models with real observations. This approach includes three main steps:

Step 1: Determine the model parameters to estimate.
  Epidemiological models simulate the dynamics of diseases. They receive information about a particular occurrence of a disease as input and return its possible evolution as output.
  One of the first steps consists in collecting and analyzing the input information. Depending on the particular situation in which the model is applied or the approach taken by the specialists, obtaining those data can be challenging. Some kinds of data are not always easily obtained due to their nature or availability; others are not reliable because they could be distorted, adulterated, miss-collected, or badly reported; even some data are not completely representative, as it comes from other occurrences of the disease or similar diseases indeed, which correspond to different settings, and hence it may not be extrapolated. For instance, in this work, data related to the mean duration of the latent period or the population density of a geographical area were easily obtained. However, it was difficult to estimate or describe parameters such as the disease effective contact rate of a particular area or the efficiency of the control measures.
  In this step, we propose to identify the unknown parameters and solve a multi-objective optimization problem to estimate them.
  For the sake of generalization, let $\phi = (\phi_1, \ldots, \phi_n) \in S$, with $n \in \mathbb{N}$ and $S \subset \mathbb{R}^n$, be the vector corresponding to n parameters to be estimated.

Step 2: Formulate the multi-objective optimization problem.
  Given a vector of values for $\phi \in S$, the epidemiological model can compute outputs denoted by $\mathbf{O}(\phi) = (O_1(\phi), \ldots, O_k(\phi))$, $k \in \mathbb{N}$, and generally related to the evolution of the disease (e.g., the number of cases or deaths).
  In many cases, those outputs can be compared to real observations, provided, for instance, by national or international authorities (such as the WHO, https://www.who.int/). These observed data are denoted by $\mathbf{RO} = (RO_1, \ldots, RO_k)$.
  Aimed to find the set of parameters that minimizes the differences between the observed and the predicted data, we formulate the following multi-objective problem:

$$\begin{cases} \min f_1(\phi) \\ \min f_2(\phi) \\ \ldots \\ \min f_m(\phi) \\ \text{subject to } \phi \in S \subset \mathbb{R}^n, \end{cases} \tag{2}$$

  where $\mathbf{f}(\phi) = \mathcal{E}(\mathbf{O}(\phi), \mathbf{RO})$ and $\mathcal{E} = (\mathcal{E}_1, \ldots, \mathcal{E}_m)$ denotes a vector of metrics $\mathcal{E}_i : \mathbb{R}^k \times \mathbb{R}^k \longrightarrow \mathbb{R}$, $i = 1, \ldots, m$, which quantify some different errors between the outputs of the model $\mathbf{O}(\phi) \in \mathbb{R}^k$ and the observed data $\mathbf{RO} \in \mathbb{R}^k$.

Step 3: Solve the multi-objective optimization problem and extract the best configuration according to the DM preferences.
  Once the optimization problem is defined, it is solved by using ParWASF-GA, which provides a set of optimal solutions with different compromises among the objectives. The best solution of the set will depend on the concerns of the DM. Depending on the particular context, he/she can be interested in prioritizing the minimization of a particular objective function or extracting the most balanced solution, for instance, the one giving the minimum value for the Euclidean distance to the reference point (RP). Our methodology includes, as a final stage, a decision tool to select a single point that satisfies the current interests of the DM.

3. Implementation of the method by considering the Be-CoDiS model

In order to illustrate the performance of our new methodology, it is applied to estimate the parameters of a deterministic epidemiological model used to study EVD outbreaks. The epidemiological model considered has been Be-CoDiS, which simulates the evolution of human diseases within and between countries. In Section 3.1, the model is described. Later, in Section 3.2, we detail the implementation of the proposed parameter estimation methodology for the case at hand.

3.1. The Be-CoDiS model

The Be-CoDiS model stands out for its ability to characterize the spread of human diseases at two levels: locally, i.e., within the country, and globally, i.e., between countries.

To describe the local behavior inside a particular country, Be-CoDiS is based on a deterministic compartment model [38]. More precisely, the population is classified into the following disjoint states of the disease [7,39–42]:

- Susceptible (denoted by S): People free of the disease but susceptible to be infected.
- Infected (denoted by E): People who are infected by the disease pathogen but who have no visible clinical signs or symptoms, i.e., they still are in the incubation period. In this state, they are neither contagious nor have visible clinical signs (e.g., fever or hemorrhages). After this incubation period, the person passes to the Infectious state.
- Infectious (denoted by I): People who can spread the disease and start developing clinical signs. Once they are detected by the authorities, they pass to the Hospitalized state.
- Hospitalized (denoted by H): The person can still infect other people, but he/she is interned at a sanitary center to receive medical care. When this Hospitalized state finishes, the person can move either to the Recovered state or to the Dead state.
- Recovered (denoted by R): The person has survived the disease and is no longer contagious. Indeed, the person has developed a natural immunity to the disease pathogen.
- Dead (denoted by D): The person has not survived the disease. However, for a disease such as the EVD, the cadavers of infected people can still spread the pathogen until they are buried.
- Buried (denoted by B): The cadaver of an infected person is buried and cannot infect other people anymore.

Focusing on the global behavior of the disease, Be-CoDiS assumes that migratory movements between countries spread the epidemic from one country to another. This migratory flow is simulated through a matrix $(\tau_{i,j})_{i,j=1}^{N_{co}}$, where $\tau_{i,j} \in [0, 1]$ is the transfer rate (day^{-1}) of people from country i to country j, and $N_{co} \in \mathbb{N}$ is the total number of considered countries. This transfer rate is expressed in percentage of population of the origin country i per day. In order to fit with the real situation, those movements only affect people in susceptible (i.e., S) and infected (i.e., E) states because they are the only ones that can travel, as they do neither have clinical signs nor remain in quarantine. Additionally, Be-CoDiS takes into account the influence of measures applied by the authorities to control the disease spread [3], such as isolation, quarantine, tracing, vaccination and the increase of sanitary resources (see [7]).

For the sake of simplicity, the following hypotheses are assumed: (i) the population inside a country is constant and homogeneously distributed, such that the spatial distribution of the epidemic inside a country can be omitted; (ii) all newborns are susceptible to be infected. Considering these assumptions, the Be-CoDiS model is used to evaluate the spread of a human disease within and between countries during a fixed time interval $[0, T_{\max}]$, with $T_{\max} \in \mathbb{N}$, by means of the following system of ordinary differential equations [7]:

$$
\frac{dS_i}{dt}(t) = -\frac{S_i(t)\Big(m_{I,i}(t)\beta_{I,i}I_i(t) + m_{H,i}(t)\beta_{H,i}H_i(t) + m_{D,i}(t)\beta_{D,i}D_i(t)\Big)}{NP_i(t)}
$$

$$
- \mu_{m,i}S_i(t) + \mu_{n,i}\Big(S_i(t) + E_i(t) + I_i(t) + H_i(t) + R_i(t)\Big)
$$

$$
+ \sum_{i\neq j} m_{\mathrm{tr},j,i}(t)\tau_{j,i}S_j(t) - \sum_{i\neq j} m_{\mathrm{tr},i,j}(t)\tau_{i,j}S_i(t),
$$

$$
\frac{dE_i}{dt}(t) = \frac{S_i(t)\Big(m_{I,i}(t)\beta_{I,i}I_i(t) + m_{H,i}(t)\beta_{H,i}H_i(t) + m_{D,i}(t)\beta_{D,i}D_i(t)\Big)}{NP_i(t)}
$$

$$
- \mu_{m,i}E_i(t) + \sum_{i\neq j} m_{\mathrm{tr},j,i}(t)\tau_{j,i}\mathcal{X}_{\epsilon_\tau}(E_j(t))
$$

$$
- \sum_{i\neq j} m_{\mathrm{tr},i,j}(t)\tau_{i,j}\mathcal{X}_{\epsilon_\tau}(E_i(t)) - \gamma_E\mathcal{X}_{\epsilon_E}(E_i(t)),
$$

$$
\frac{dI_i}{dt}(t) = \gamma_E\mathcal{X}_{\epsilon_E}(E_i(t)) - (\mu_{m,i} + \gamma_{I,i}(t))I_i(t),
$$

$$
\frac{dH_i}{dt}(t) = \gamma_{I,i}(t)I_i(t) - \Big(\mu_{m,i} + (1 - \omega_i(t))\gamma_{HR,i}(t) + \omega_i(t)\gamma_{HD,i}(t)\Big)H_i(t),
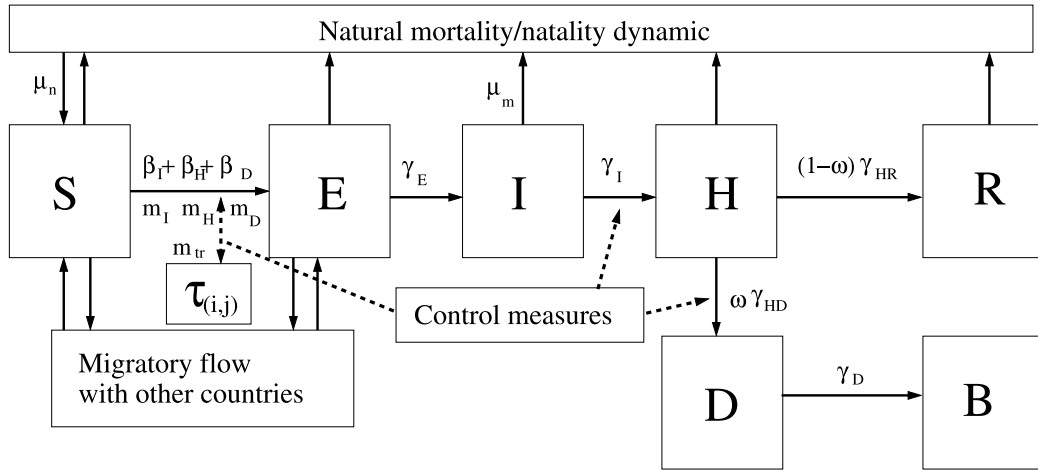$$

Fig. 1. Diagram summarizing the Be-CoDiS model.

$$\frac{dR_i}{dt}(t) = (1 - \omega_i(t))\gamma_{HR,i}(t)H_i(t) - \mu_{m,i}R_i(t),$$

$$\frac{dD_i}{dt}(t) = \omega_i(t)\gamma_{HD,i}(t)H_i(t) - \gamma_D D_i(t), \tag{3}$$

$$\frac{dB_i}{dt}(t) = \gamma_D D_i(t),$$

where:

- $i \in \{1, \ldots, N_{co}\}$ is the index of each country;
- $N_{co} \in \mathbb{N}$ is the number of countries;
- $NP_i(t) = S_i(t) + E_i(t) + I_i(t) + H_i(t) + R_i(t) + D_i(t)$ is the number of people (alive and also died because of the disease) in country i at time t;
- $\mu_{n,i} \in [0, 1]$ is the birth rate (day^{-1}) in country i: the number of births per day and per capita;
- $\mu_{m,i} \in [0, 1]$ is the mortality rate (day^{-1}) in country i: the number of deaths per day and per capita not caused by the disease;
- $\omega_i(t) \in [0, 1]$ is the disease fatality rate in country i at time t: the proportion of people who do not survive the disease;
- $\beta_{I,i}$, $\beta_{H,i}$, $\beta_{D,i} \in (0, +\infty)$ are the disease effective contact rates (day^{-1}) of a person in state I, H or D, respectively, in country i: the mean number of effective contacts (i.e., contacts sufficient to transmit the disease) of a person in state I, H or D, respectively, per day before applying control measures;
- γ_E, $\gamma_{I,i}(t)$, $\gamma_{HR,i}(t)$, $\gamma_{HD,i}(t)$, $\gamma_D \in (0, +\infty)$ denote the transition rate (day^{-1}) from state E, I, H (surviving), H (not surviving) or D to state I, H, R, D or B, respectively: the number of people per day and per capita passing from one state to the other. We note that $\gamma_{I,i}(t)$, $\gamma_{HR,i}(t)$ and $\gamma_{HD,i}(t)$ are time and country dependent, since, due to the applied control measures in country i, their value could vary in time.
- $m_{I,i}(t)$, $m_{H,i}(t)$, $m_{D,i}(t) \in [0, 1]$ are functions representing the efficiency of the control measures applied to non-hospitalized infectious people, hospitalized people and infected cadavers respectively, in country i at time t to eradicate the epidemic.
- $(\tau_{i,j})_{i,j=1}^{N_{co}}$ is a matrix composed by the transfer rates (day^{-1}) of persons from one country i to another j expressed in proportion of population in i per unit of time. Those rates can be also diminished by the progressive application of the control measures in countries i and j. Then, they are multiplied by the function $m_{tr,i,j}(t) = m_{I,i}(t) \cdot m_{I,j}(t)$.
- $\mathcal{X}_{\epsilon_{fit}}(x) = x$ if $x \geq \epsilon_{fit}$, $\mathcal{X}_{\epsilon_{fit}}(x) = 2x - \epsilon_{fit}$ if $(\epsilon_{fit}/2) \leq x \leq \epsilon_{fit}$, and 0 elsewhere, with $\epsilon_{fit} \geq 0$ being a small tolerance parameter. This function, with ϵ_{fit} small enough, is a filter used to avoid artificial spread of the epidemic due to negligible values of x.

Finally, to complete System (3) the initial data $S_i(0)$, $E_i(0)$, $I_i(0)$, $H_i(0)$, $R_i(0)$, $D_i(0)$, $B_i(0) \in [0, +\infty)$, for $i \in \{1, \ldots, N_{co}\}$, must be provided.

A general flow diagram describing Be-CoDiS is depicted by Fig. 1.

From a numerical point of view, in this work, System (3) is implemented by applying a fully explicit forward Euler numerical scheme considering a fixed time step of $h = 1$ day, as done in [7]. Following this previous article, this value

Table 1
Country indicators used to estimate some of the Be-CoDiS parameters.

Notation	Description
NAT_i	Natality Rate (day^{-1})
MLE_i	Mean Life Expectancy (years)
GNI_i	Gross National Income (US\$ persons^{-1} year^{-1})
SAN_i	Mean Health Expenditure (US\$ persons^{-1} year^{-1})
DEN_i	Population Density (persons km^{-2})

Table 2
Known parameters of the Ebola disease.
Source: [7,11,39,42].

Notation	Value (days)	Description
d_E	11.4	Mean duration of a person in the state E
d_I	5.0	Mean duration of a person in the state I
d_{HR}	5.0	Mean duration of a person from the state H to R
d_{HD}	4.2	Mean duration of a person from the state H to D
d_D	2.0	Mean duration of a person in the state D
d_g	3.0	Maximum reduction of the duration of a person in the state I due to control measures

gives a good balance between computational time and precision of results. We note that the computational time required for evaluating the model should be reasonable as it will be evaluated several times during the parameters calibration process.

3.2. Implementation of the general methodology

Now, the general methodology explained in Section 2.3 is applied to estimate some of the epidemiological parameters of the Be-CoDiS model in the case of EVD epidemics.

Step 1: Determine the model parameters to estimate.
The Be-CoDiS model requires the definition of the following parameters to perform simulations:

1. **Demography dynamic:** The birth and mortality rates of country i (i.e., $\mu_{n,i}$ and $\mu_{m,i}$) are computed from country indicators that are known and freely available on the World Data Bank website [43]. More precisely, the birth rate corresponds to the natality rate (denoted by NAT_i), i.e., $\mu_{n,i} = NAT_i$, and the mortality rate is computed as the inverse of the mean life expectancy (denoted by MLE_i), i.e., $\mu_{m,i} = 1/(365 \cdot MLE_i)$. In Table 1, some country indicators useful to estimate some of the Be-CoDiS parameters are reported.

2. **Transition rates:** They allow estimating the number of people passing from one state to the other per unit of time. They are obtained as the inverse of the mean duration of people in the considered state. Thus, for all $t > 0$, the transition rates are defined by:

$$
\begin{aligned}
\gamma_{E,i}(t) &= 1/d_E, \\
\gamma_{D,i}(t) &= 1/d_D, \\
\gamma_{I,i}(t) &= 1/(d_I - g_i(t)), \\
\gamma_{HR,i}(t) &= 1/(d_{HR} + g_i(t)), \\
\gamma_{HD,i}(t) &= 1/(d_{HD} + g_i(t)),
\end{aligned}
\tag{4}
$$

where $d_E, d_I, d_{HR}, d_{HD}, d_D$ denote the mean duration of one of those people in state E, I, H (surviving), H (not surviving) or D, respectively, without the application of control measures. Moreover, we assume that if the duration of a person in state I decreases because it is detected earlier thanks to more exhaustive control measures, then he/she will enter earlier into one of the hospitalized states HR or HD, so the duration in those states will increase. Thus, to model those changes in the duration, we introduce in the formula related to the transition rates $\gamma_{I,i}(t)$, $\gamma_{HR,i}(t)$ and $\gamma_{HD,i}(t)$, the term $g_i(t) = d_g \cdot (1 - m_{I,i}(t))$, where d_g is the maximum reduction of the duration of a person in the state I in days due to control measures.

Those mean durations $d_E, d_I, d_{HR}, d_{HD}, d_D$, as well as this maximum reduction of the I-state duration d_g, are data specific to the considered disease and are usually available in epidemiological reports. For the EVD, they have been taken from [7,11,39,42] and summarized in Table 2.

3. **Contact rates:** According to [7], the disease effective contact rate of a person in state I in country i, denoted by $\beta_{I,i}$, is assumed to be related to the density and economic level of the country through the following function:

$$
\beta_{I,i} = \beta_I(i) = \beta_I\left(\frac{DEN_i}{GNI_i}\right) = b_3 \cdot \arctan\left(b_1 \cdot \frac{DEN_i}{GNI_i} + b_2\right) + b_4,
\tag{5}
$$

where DEN_i is the population density (persons km^{-2}) and GNI_i is the gross national income (US$ persons^{-1} year^{-1}) of the country i. Both of them are known data (see Table 1).

However, parameters b_1 (km^2 US$ persons^{-2} year^{-1}), b_2 (non-dimensional), b_3 (day^{-1}) and b_4 (day^{-1}) in Eq. (5) are unknown and they must be adjusted to achieve a global law that applies to all countries. Therefore, they are considered as decision variables in the multi-objective formulation.

For the Ebola outbreak, the probability of being infected by contact with persons in the state H is 25 times lower than the probability of being infected by contact with persons in the state I, and the infection probability with persons in the state D is the same as the one with persons in the state I [7]. Thus, the contact rates of states H and D are computed by $\beta_{H,i} = \beta_{I,i}/25$ and $\beta_{D,i} = \beta_{I,i}$, respectively.

4. **Control measures:** The effects of some control measures is implemented in the Be-CoDiS model (see System (3)) by means of the functions $m_{I,i}(t)$, $m_{H,i}(t)$ and $m_{D,i}(t)$. They have a double purpose. On the one hand, as previously said, they appear in the definition of the transition rates because we assume that one of the effects of the control measures is the increase of surveillance that produces a decrease in the duration of the infected state. On the other hand, they multiply $\beta_{I,i}$, $\beta_{H,i}$ and $\beta_{D,i}$ to simulate the reduction of the number of effective contacts as the efficiency of control measures is improved over time (e.g. due to educational campaigns, increase of sanitary resource...). As done in [7], they are decreasing functions defined as follows [7,44]:

$$m_{I,i}(t) = m_{H,i}(t) = m_{D,i}(t) = \exp\left(-\kappa_i \max(t - \lambda_i, 0)\right), \qquad (6)$$

where $\kappa_i \in [0, +\infty)$ (day^{-1}) represents the efficiency of the control measures (greater value implies lower values of disease contact rates) and $\lambda_i \in \mathbb{R} \cup \{+\infty\}$ (day) denotes the first day of application of those control measures. In some cases, the values of λ_i can be taken from the literature, using the dates when the territories implement some specific control measures (see, for instance, [26]). However, in many situations as the one considered here, it is difficult to estimate when those measures are effective in the population. In those cases, the parameters λ_i must be included in the decision vector to be estimated by the multi-objective methodology.

Moreover, the efficiency of the control measures in country i (i.e., κ_i) is assumed to be related to the sanitary resources and the density of this country through the following formula (see [7]):

$$\kappa_i = \kappa(i) = \kappa\left(\frac{SAN_i}{DEN_i}\right) = k_3 \cdot \arctan\left(k_1 \cdot \frac{SAN_i}{DEN_i} + k_2\right) + k_4, \qquad (7)$$

where SAN_i is the mean health expenditure (US$ persons^{-1} year^{-1}) and DEN_i is the population density (persons km^{-2}) of the country i. Both are known data (see Table 1).

In (7), parameters k_1 (persons^2 year km^{-2} US^{-1}), k_2 (non-dimensional), k_3 (day^{-1}) and k_4 (day^{-1}) must be included in the decision vector to be estimated with the multi-objective methodology.

5. **Fatality rate:** The fatality rate $\omega_i(t)$ depends on the country i and time t. Here, the following expression, proposed in [7], is used:

$$\omega_i(t) = \omega(i, t) = \delta\left((1 - C_m \cdot m_{I,i}(t))\underline{\omega} + C_m \cdot m_{I,i}(t)W_i\right) + (1 - \delta)W_i, \qquad (8)$$

where $\delta \in [0, 1]$ denotes the proportion of the fatality rate that can be reduced due to the application of control measures, $C_m \in [0, 1]$ is a constant used to decrease the effect of control measures on the fatality rate reduction, and W_i denotes the disease fatality rate of country i when no control measures are applied. More specifically, the value of W_i has been computed as:

$$W_i = \frac{SAN_i}{C_{SAN}}\underline{\omega} + \left(1 - \frac{SAN_i}{C_{SAN}}\right)\overline{\omega} \qquad (9)$$

where $\underline{\omega}, \overline{\omega} \in [0, 1]$ are the minimum and the maximum disease fatality rates, respectively, and C_{SAN} is a constant for regulating the effect of the sanitary country indicator SAN_i in the fatality rate. For the Ebola virus disease, it can be considered $\underline{\omega} = 0.25$ and $\overline{\omega} = 0.728$ (see [7]).

In this case, the unknown parameters to be estimated as decision variables are δ, C_m and C_{SAN}.

6. **Migratory dynamics:** The transfer rates $\tau_{i,j}$ of persons from one country i to another j, expressed in proportion of the population in i per unit of time, have been obtained applying the following formula [7]:

$$\tau_{i,j} = \tau(i, j) = C_\tau \tilde{\tau}(i, j)/(5 \cdot 365 \cdot NPA_i(0)),$$

where $\tilde{\tau}(i, j)$ is the number of persons moving from i to j in a specific five-year period, $NPA_i(0)$ is the total number of persons alive in country i at the beginning of the simulation, and C_τ is a constant parameter used to increase or decrease the effect of the movement fluxes. Data about the migrations between 2005 and 2010, taken from the following sources [45–47], have been used to set the value of $\tilde{\tau}(i, j)$.

The parameter C_τ is considered as a decision variable to be also estimated.

7. **Numerical filters:** In System (3), two numerical filters are used to avoid the spread of the disease when the number of infected persons is not significant, i.e., $E_i(t)$ is below a certain threshold value. One of the filters modulates the influence of the movements of infected people over the spread of the disease among countries. More precisely, if $E_i(t) < \epsilon_\tau$, then the disease does not spread from country i to other countries. The other filter controls the spread process inside a country, in such a way that infected individuals in E_i do not evolve into infectious ones in I_i if $E_i(t) < \epsilon_E$.

Both thresholds ϵ_τ and ϵ_E are also included in the decision vector ϕ to estimate their values.

Thus, the decision vector, which is composed of the parameters to be adjusted with the proposed multi-objective methodology, is given by:

$$\phi = (\epsilon_\tau, \epsilon_E, C_\tau, k_1, k_2, k_3, k_4, b_1, b_2, b_3, b_4, \delta, C_m, C_{\text{SAN}}). \tag{10}$$

Step 2: Formulate the multi-objective optimization problem.

According to Section 2.3, the multi-objective problem is aimed at finding a set of values for ϕ, such that the disease behavior provided by the model fits, as close as possible, real observations.

For the Ebola outbreak, situation reports are periodically published by the WHO (see [11]). Those documents provide the cumulative number of cases and the cumulative number of deaths observed in each affected country $i \in \{1, \ldots, N_{\text{co}}\}$, denoted by $\text{RC}_i(t)$ and by $\text{RD}_i(t)$, respectively. This information is also used to set the initial conditions of the epidemiological model (see [7]).

Furthermore, given specific values for ϕ, Be-CoDiS estimates the evolution of the population in the different compartments denoted by $S_i^\phi(t)$, $E_i^\phi(t)$, $I_i^\phi(t)$, $H_i^\phi(t)$, $R_i^\phi(t)$, $D_i^\phi(t)$ and $B_i^\phi(t)$ and returns the predicted evolution of the cumulative number of cases $CC_i^\phi(t)$ and the cumulative number of deaths $CD_i^\phi(t)$ at time t, computed as follows:

$$CC_i^\phi(t) = CC_i(0) + \int_0^t \gamma_{I,i}(s) \cdot I_i^\phi(s)\mathrm{d}s, \tag{11}$$

$$CD_i^\phi(t) = CD_i(0) + \int_0^t \omega_i(s) \cdot \gamma_{HD,i}(s) \cdot H_i^\phi(s)\mathrm{d}s, \tag{12}$$

where $CC_i(0)$ and $CD_i(0)$ are the initial amounts of cases and deaths at the beginning of the simulation and available in situation reports.

Now, according to the general formulation (2), the multi-objective problem must be defined by expressing the objective functions that quantify the errors between real and predicted data. To do so, for each country i, the objective functions $f_{C,i}$ and $f_{D,i}$ that evaluate the relative errors committed in the cumulative number of cases and deaths, respectively, are considered:

$$f_{C,i}(\phi) = \frac{\left\| \text{RC}_i - CC_i^\phi \right\|_{L^2([0,t_f])}}{\|\text{RC}_i\|_{L^2([0,t_f])}}, \tag{13}$$

$$f_{D,i}(\phi) = \frac{\left\| \text{RD}_i - CD_i^\phi \right\|_{L^2([0,t_f])}}{\|\text{RD}_i\|_{L^2([0,t_f])}}, \tag{14}$$

where t_f is the final time of simulation, which corresponds to the date of the latest situation report, and $\|\cdot\|_{L^2([0,t_f])}$ denotes the L^2 norm in $[0, t_f]$ $\left(\text{i.e., } \|g\|_{L^2([0,t_f])} = \left(\int_0^{t_f} (g(s))^2 \, \mathrm{d}s\right)^{1/2}\right)$. We note that, in practice, data are discrete so we must consider a numerical approximation of (13) and (14) to deal with such kind of data. In particular, in this work, the integral operators are approximated by considering a trapezoidal rule with a time step equal to 1 day (which corresponds to the case of daily reported data).

Notice that the objective functions $f_{C,i}$ and $f_{D,i}$ are considered only for countries where the epidemic has higher impacts (i.e., the number of cases is greater than a fixed threshold, which depends on the considered outbreak). For all remaining countries, two other objective functions are included in the multi-objective formulation to jointly take into account possible errors in both their predicted number of cases and deaths. Those functions, denoted by f_C and f_D, are defined as the average values of the absolute errors in the cumulative number of cases and deaths, respectively, considering all the countries except the most affected ones. Thus, they correspond to:

$$f_C(\phi) = \frac{1}{N_{\text{co}} - |M|} \sum_{\substack{i=1,\ldots,N_{\text{co}} \\ i \notin M}} \left\| \text{RC}_i - CC_i^\phi \right\|_{L^2([0,t_f])}, \tag{15}$$

$$f_D(\phi) = \frac{1}{N_{\text{co}} - |M|} \sum_{\substack{i=1,\ldots,N_{\text{co}} \\ i \notin M}} \left\| \text{RD}_i - CD_i^\phi \right\|_{L^2([0,t_f])}, \tag{16}$$

where $M \subset \{1, \ldots, N_{\text{co}}\}$ is the subset containing the indexes of the most affected countries and $|M|$ is the number of countries in this subset M. As explained above, the objective functions $f_{C,i}$ and $f_{D,i}$ are only considered for the countries

whose indexes are $i \in M$. To determine the subset of most affected countries, we consider the countries for which the cumulative number of cases reported by the WHO at any time of the available dates is greater than a fixed threshold.

Finally, we have included two more objective functions to minimize (i) the number of false positive countries, i.e., countries detected as affected by the model but free of disease in the reality; and (ii) the false negative countries, that is countries free of disease in the simulation but detected as affected in the reality. They are denoted by f_{FP} and f_{FN}, respectively:

$$
\begin{aligned}
f_{FP}(\phi) & = \text{Number of non-affected countries} \\
& \quad \text{that the model predicts as affected;} \\
f_{FN}(\phi) & = \text{Number of affected countries} \\
& \quad \text{that the model predicts as non-affected.}
\end{aligned}
\tag{17}
$$

We note that we assume that a country is classified as affected when it has at least one cumulative case at the end of the simulation.

The multi-objective problem to fit epidemiological parameters can be formulated as:

$$
\begin{cases}
\min \ \left\{ \left\{ f_{C,i}(\phi), \forall i \in M \right\}, \left\{ f_{D,i}(\phi), \forall i \in M \right\}, f_C(\phi), f_D(\phi), f_{FP}(\phi), f_{FN}(\phi) \right\} \\
\text{subject to } \phi \in S \subset \mathbb{R}^n,
\end{cases}
\tag{18}
$$

where ϕ is the decision vector (10) containing the parameters to be estimated.

Step 3: Solve the multi-objective optimization problem and extract the best configuration according to the DM preferences.

According to our methodology, Problem (18) is now solved using the multi-objective optimization algorithm ParWASF-GA (see Section 2.2). As a consequence, a set of optimal points is obtained, i.e. an approximation of the Pareto front. Then, we have to select the best solution from the set according to the criteria established by the DM. For the case at hand, we have considered that he/she prefers the closest point (with respect to the Euclidean distance) to the RP in the objective space. This solution will be composed of an adjusted value for each of the epidemiological parameters included in the decision vector ϕ.

Furthermore, we establish the RP at the origin. This point corresponds to the ideal (but generally not feasible) solution for which the model perfectly fits the observation.

4. Numerical experiments and discussion

Here, we want to illustrate and validate our methodology by considering data from three real EVD epidemics. Two of them emerged in 2018 and affected the Democratic Republic of the Congo (DR Congo). The other outbreak occurred in 2014–2016 and involved several countries, mainly from West Africa. Focusing on the cumulative numbers of cases and deaths observed during the epidemics, we consider data provided by the WHO (https://www.who.int/). Since those data are generally not reported every day, we interpolate them by using cubic Hermite polynomials to obtain a daily estimation.

With the collected information, we have conducted two kinds of experiments: (i) focusing on a single country and (ii) including a set of countries. In Sections 4.1 and 4.2, we summarize the experimental framework, according to the methodology proposed in Section 3.2, and discuss the main results obtained.

All the computational experiments have been performed on a shared memory cluster composed of eight 16-core processors Intel Xeon E7 8860v3 with 2.3 TB of RAM. Thus, ParWASF-GA [23] has been used with 128 threads.

4.1. Experiments considering one single country

We focus here on two recent EVD outbreaks that emerged in the Democratic Republic of the Congo (DR Congo) in 2018. The first one affected the Équateur Province and was declared ended on the 24th of July, 2018 [48]. The second one affected the North Kivu, South Kivu and Ituri Provinces and ended on the 14th of May, 2020 [49]. In both cases, the virus spread only in DR Congo (except for four sporadic cases quickly contained in Uganda [49]). However, due to noticeable differences in the socio-political situation of the affected geographical areas (separated by more than 1.000 km) and different strains of the virus, the dynamic of those epidemics was quite different. Thus, we decided to apply our methodology by considering each case as an individual experiment. Sections 4.1.1 and 4.1.2 are devoted, respectively, to the first and second DR Congo outbreaks, which we label as **DRC1** and **DRC2**.

For both **DRC1** and **DRC2** epidemics, we calibrate the parameters of the model considering real observations. Additionally, since **DRC2** was still active during the development of this work, we include in Section 4.1.2 a real-time forecast (performed in July 2019) of its evolution and compare the predictions with the real situation at the end of the outbreak in May 2020.

For these experiments, the Be-CoDiS model is adapted to simulate the evolution of a disease in a single country (see System (3)). Then, we set $N_{co} = 1$. Furthermore, notice that we do not need Eqs. (5) and (7) to describe the parameters β_I and κ. Indeed, they can be estimated directly by considering them as decision variables instead of b_1, b_2, b_3, b_4 and

Table 3

Description and range of the parameters to optimize for the **DRC1** and **DRC2** one-country experiments presented in Sections 4.1.1 and 4.1.2, respectively.

Notation	Range	Description
β_I	[0.0, 0.5]	Disease contact rate of I (day^{-1} persons^{-1})
λ	[0.0, 700.0]	First day of application of control measures (day)
κ	[0.0, 1.0]	Efficiency of the control measures (day^{-1})
δ	[0.0, 1.0]	Proportion of the fatality rate that can be reduced by control measures
C_m	[0.0, 1.0]	Parameter multiplying the control measures value to decrease its effect on the fatality rate reduction
C_{SAN}	[50.0, 9000.0]	Parameter dividing the sanitary country indicator in Eq. (9)
λ_2*	[301, 400]	First day of restoration of the control measures (day). *Only considered in **DRC2**

Table 4

Summary of the main settings for the three different experiments: DR Congo outbreak in 2018 (**DRC1**), DR Congo epidemic in 2018–20 (**DRC2**), and West Africa outbreak in 2014–16 (**WEA**). We report the number of objectives and decision variables of the optimization problem, the number of individuals and iterations for the ParWASF-GA optimizer, the number of repetitions (Runs) of the optimization procedure, and the computational time (in hours) of a single run on average (Av(Time)) using 128 threads.

Outbreak		One country		Set of countries
		DRC1	**DRC2**	**WEA**
Number of countries	N_{co}	1	1	176
Optimization Problem	Objectives	2	2	10
	Variables	6	7	17
Optimizer Settings	Individuals	50	50	770
	Iterations	10 000	10 000	20 000
Runs	N_{runs}	30	30	10
Av(Time)	(h/run)	0.09	0.13	33.5

k_1, k_2, k_3, k_4. Therefore, the decision vector $\phi' \in S' \subset \mathbb{R}^{n'}$ is composed by $n' = 6$ parameters for the **DRC1** experiment (see Section 4.1.1) and $n' = 7$ for the **DRC2** case (see Section 4.1.2). Table 3 summarizes the set of parameters with their corresponding range of values. Those intervals have been set according to the literature (for instance, for β_I, see [7,50]), to the meaning of the parameters (it is the case of κ, δ, and C_m, which are proportions expressed between 0 and 1) or even on a combination of both to guarantee that the range is wide enough but realistic (for instance, λ, λ_2, and C_{SAN}).

Furthermore, when considering a single country, Problem (18) is solved by taking into account only the following two objective functions: (i) the relative error in the number of cases $f_{C,1}(\phi')$ (see Eq. (13)), and (ii) the relative error in the number of deaths $f_{D,1}(\phi')$ (see Eq. (14)) of the target country. More precisely, Problem (18) can be rewritten as:

$$\begin{cases} \min \; \{f_{C,1}(\phi'), f_{D,1}(\phi')\} \\ \text{subject to } \phi' \in S' \subset \mathbb{R}^{n'}. \end{cases} \tag{19}$$

According to the general methodology proposed in Section 3.2, the previous optimization problem is solved by using ParWASF-GA. The values for the set of input parameters of ParWASF-GA can be found in [22], except for the number of individuals and iterations, which have been adapted for the problem at hand. Additionally, notice that ParWASF-GA is a stochastic optimization algorithm, and hence it may obtain different Pareto front approximations for different executions of the same instance. Then, in order to ensure that the results and conclusions do not depend on the stochasticity of the algorithm, several runs of each experiment have been performed. Table 4 summarizes the number of objectives and decision variables of the optimization problem, the number of individuals and iterations considered in ParWASF-GA, the number of executions denoted by N_{runs}, and the averaged computational time of one execution (in hours) using 128 threads. For the sake of completeness, the same information is reported for the case when several countries are considered.

For each experiment, N_{runs} different Pareto set approximations $PS_1, \ldots, PS_{N_{runs}}$ in the decision space have been obtained. From each PS_r, $r = 1, \ldots, N_{runs}$, only one solution $\mathbf{x}_r \in PS_r$ is chosen by the DM according to the selection strategy (see Step 3 in Section 3.2). Here, we assume that the DM prefers the most balanced solution. Thus, the chosen solution is the one giving the minimum Euclidean distance to the RP. Then, to validate our methodology and prove its robustness no matter the execution, we consider the set of chosen solutions of all the runs $X = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{N_{runs}}\}$, and we extract the "best" and the "worst" result $\mathbf{x}_{best}, \mathbf{x}_{worst} \in X$. More precisely, the "best" result is the solution offering the minimum value for the Euclidean distance to the reference point, and the "worst" result is the one corresponding to the maximum value, i.e.:

$$\|\mathbf{f}(\mathbf{x}_{best}) - \mathbf{q}\|_2 = \min_{r=1,\ldots,N_{runs}} \{\|\mathbf{f}(\mathbf{x}_r) - \mathbf{q}\|_2\}, \tag{20}$$

$$\|\mathbf{f}(\mathbf{x}_{worst}) - \mathbf{q}\|_2 = \max_{r=1,\ldots,N_{runs}} \{\|\mathbf{f}(\mathbf{x}_r) - \mathbf{q}\|_2\}, \tag{21}$$

where $\mathbf{q}$ denotes the RP and $\|\cdot\|_2$ the Euclidean distance.

Table 5

Solutions obtained during the experiment DRC1 reported in Section 4.1.1. We present the values corresponding to the "best" ($\mathbf{x}_{best}$) and the "worst" ($\mathbf{x}_{worst}$) results in X, and the average ($Av(X)$) and standard deviation ($Dev(X)$) values. Recall that X is the set formed by one solution from each run according to the selection criterion.

DR Congo (2018 outbreak)				
Decision variables	$\mathbf{x}_{best}$	$\mathbf{x}_{worst}$	$Av(X)$	$Dev(X)$
β_I	3.54E−01	3.52E−01	3.52E−01	3.54E−03
λ	2.84E+01	2.72E+01	2.94E+01	1.32E+00
κ	2.65E−01	1.94E−01	3.59E−01	1.34E−01
δ	2.38E−01	4.64E−01	4.16E−01	7.16E−02
C_m	2.47E−01	1.50E−01	1.24E−01	7.81E−02
C_{SAN}	5.39E+01	8.36E+03	3.96E+03	3.09E+03

4.1.1. DRC1 - DR Congo outbreak in 2018: a posteriori adjustment and analysis

On the 8th of May, 2018, the WHO declared an outbreak of EVD in DR Congo [51]. It affected the towns of Bikoro, Iboko and Wangata, which are located in Équateur Province at the western side of the country. According to the WHO situation reports [51], it seems that the first cases reported by the sanitary authorities occurred on the 5th of April, 2018. On the 24th of July, 2018, the Ministry of Health of DR Congo officially confirmed the end of this outbreak, 42 days (i.e., a period corresponding to two incubation periods) after blood samples from the last confirmed Ebola patient tested negative for the disease.

Several forecasts of this outbreak based on the Be-CoDiS model were proposed from the 14th of May, 2018 up to the 13th of June, 2018 [12]. However, for those forecasts, the model parameters were adjusted through a time-consuming manual trial and error process. Later, we have developed and applied the parameter estimation methodology proposed in this work once the outbreak was already extinguished. Thus, the results presented below will be used as a posteriori analysis to validate our method.

To do that, we have used data reported by the WHO from the 5th of April up to the 17th of June, 2018 (i.e., the date of the last reported cases) [51]. Considering those data, the optimization problem (19) was solved with ParWASF-GA 30 times. Then, as explained above in Section 4.1, we have built the set X with one solution for each run (according to the DM preference), and we have selected the "best" and the "worst" result, $\mathbf{x}_{best}, \mathbf{x}_{worst} \in X$. The parameters associated with those two solutions are reported in Table 5. For the sake of completeness, in Table 5, we also show the average value and the standard deviation of each parameter in X.

In Fig. 2, we present the cumulative number of cases and deaths returned by the model considering the "best" and the "worst" optimal parameter set found with our methodology, and reported in Table 5. We also show observed data corresponding to the cumulative numbers of cases and deaths (see [51]). As can be seen in this figure, even considering the "worst" result, the parameters found for the one-country model are able to accurately adjust the behavior of the disease.

Focusing on Table 5, we observe that:

- The contact rate β_I is 3.52E−01 on average. During the 2014–16 Ebola epidemic, β_I was estimated to be between 1.91E−01 and 2.65E−01 for the most affected countries (see [7]). Thus, for this outbreak in DR Congo, the Ebola virus seems to spread faster than for the 2014–16 case. This quick behavior was also reported by authorities at the beginning of this epidemic and provoked a strict vigilance of this case by the international community [52].

- For this outbreak, the value of κ, the parameter measuring the efficiency of the control, is greater than 1.E−01. Thus, it is quite larger than the one reported for the 2014–16 epidemic, which was between 1.25E−03 and 2.70E−03 (see [7]). This fact is consistent with reality, because, for this case, national and international authorities quickly applied control measures that were known to be efficient during the 2014–2016 epidemic [53]. Additionally, some novel control measures were applied, such as the vaccination of the population at risk [54].

- The values of λ, which is the number of days between the start of the disease outbreak and the application of the control measures, seem to indicate that the control measures started to affect the epidemic dynamics between the 2nd and the 6th of May, 2018. According to situation reports [51], the control measures could have been applied around the 3rd and the 8th of May, 2018, when the local authorities, supported by WHO, started to monitor and analyze blood samples. However, the effect of those control measures is not immediate and estimating their impact on the outbreak is a complex task. Thus, the range of dates proposed by our method seems to be reasonable regarding the reported sanitary situation.

- The parameter C_{SAN} dividing the sanitary country indicator in Eq. (9) is lower than the one estimated for the 2014–16 outbreak, which was 9.06E+03 (see [7]). Then, according to Eq. (9), before the application of control measures, the fatality rate of the 2018 epidemic is lower than the one of the 2014–2016 outbreak. It is consistent with the data reported by WHO [51], which estimated that the fatality rate of the 2018 epidemic was between 42% and 61%, i.e., lower than the one of the 2014–16 outbreak, which was around 68% before applying control measures.

- The percentage δ of the fatality rate that can be reduced by control measures is 41.6% on average, with a standard deviation of 7.2%. Thus, it is lower than the one estimated for the 2014–16 Ebola outbreak, which was 53% (see [7]).
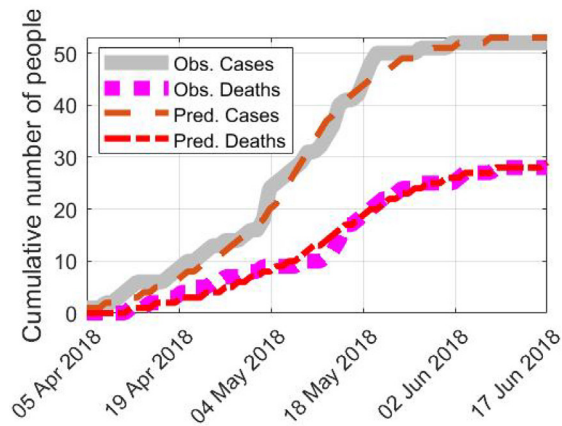
(a) Considering the "best" result ($\mathbf{x}_{\text{best}} \in X$)
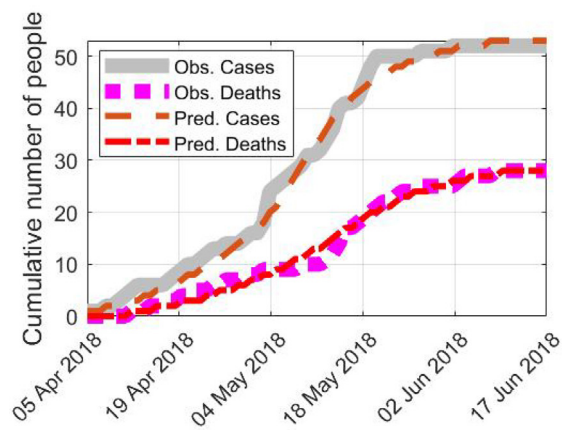
(b) Considering the "worst" result ($\mathbf{x}_{\text{worst}} \in X$)

Fig. 2. Comparison between the observed and the predicted cumulative number of cases and deaths considering the experiment **DRC1** described in Section 4.1.1.

This value is realistic because, according to WHO [11], in the 2014–16 epidemic, the fatality rate was reduced down to 40%. However, for the 2018 outbreak, the final fatality rate was 61% (see [51]).

• The parameter C_m multiplying the control measures value in Eq. (8) is 1.24E−01 on average, with a standard deviation of 7.81E−02. It is quite lower than the one considered for the 2014–16 epidemic, which was 1 (see [7]). It means that, for the 2018 outbreak, the control measures have a lower impact on the reduction of the fatality rate (see Eq. (8)). Again, this fits reality because in 2014–16 the mortality rate experienced a greater reduction than in the case of 2018 (see [11,51]).

Those results confirm that, although the 2018 outbreak was more aggressive, efficient control measures were applied, avoiding a sanitary emergency and leading to the eradication of the disease.

4.1.2. DRC2 - DR Congo outbreak in 2018–20: real-time adjustment and forecast

On the 1st of August, 2018, a new EVD outbreak was declared in DR Congo and reported to the WHO [51]. In this case, which ended on the 14th of May, 2020 [49], the affected provinces were, by order of infection, North Kivu, Ituri and South Kivu. According to the WHO situation reports [51], it seems that the first cases reported by the DR Congo sanitary authorities occurred on the 28th of July, 2018, but the disease could have started approximately around the 30th of April, 2018.

During this outbreak, we have performed real-time adjustments of the epidemiological parameters and forecasts of the possible evolution of the epidemic. Indeed, we have carried out exhaustive monitoring of the official situation reports and

Table 6

Solutions obtained during the experiment **DRC2** reported in Section 4.1.2. We present the values corresponding to the "best" ($\mathbf{x}_{best}$) and the "worst" ($\mathbf{x}_{worst}$) results in X, and the average ($\mathrm{Av}(X)$) and standard deviation ($\mathrm{Dev}(X)$) values. Recall that X is the set formed by one solution from each run according to the selection criterion.

DR Congo (2018–20 outbreak)				
Decision variables	$\mathbf{x}_{best}$	$\mathbf{x}_{worst}$	$\mathrm{Av}(X)$	$\mathrm{Dev}(X)$
β_I	2.13E−01	2.02E−01	2.10E−01	3.69E−03
λ	2.86E−01	3.72E+00	6.59E+00	2.06E+01
κ	8.21E−04	5.58E−04	7.94E−04	6.79E−05
δ	2.77E−01	9.99E−01	4.86E−01	1.71E−01
C_m	5.42E−01	9.73E−01	7.78E−01	1.82E−01
C_{SAN}	7.31E+03	8.77E+03	6.07E+03	2.62E+03
λ_2	3.53E+02	3.40E+02	3.53E+02	3.80E+00

have published several technical documents reporting the values of the parameters obtained by using our multi-objective methodology, as well as the prediction of the disease evolution [13].

Here, we present the estimated values of the parameters and the forecast based on data reported by the WHO and the DR Congo authorities from the 29th of April, 2018, up to the 21st of July, 2019.

We note that the geographical area where this outbreak occurs is conflictive, with many security problems and little confidence of the population in the outbreak response teams (see [55]). Indeed, according to the DR Congo authorities, between the 24th of February and the 30th of March, 2019, several sanitary centers of response were closed and some response teams canceled their work temporarily due to continuous attacks. Thus, for this particular case, during the parameter calibration process we have considered that the control measures were interrupted on the 24th of February. Moreover, the exact date for which the control measures were completely restored is not clear, so we have decided to estimate it by including a new parameter in the decision vector ϕ'. This new parameter, denoted by λ_2, is defined as the number of days between the start of the disease outbreak and the restoration of the control measures. To be consistent with the dates on which the attacks were produced, we assume $\lambda_2 \in [301, 400]$. The other decision variables composing ϕ' are those reported in Table 3 and the optimization problem to be solved is defined by (19) with $n' = 7$.

Analogously to the previous experiment, in Table 6, we report the parameter values for the "best" and the "worst" result found in the set X, i.e., $\mathbf{x}_{best}, \mathbf{x}_{worst} \in X$. We recall that X consists of 30 solutions, obtained by applying the DM preference criterion. We also include in Table 6 the average value and the standard deviation of each parameter considering those 30 solutions in X.

In Fig. 3, we present the cumulative number of cases and deaths returned by the model considering the "best" and the "worst" optimal parameter set found with our methodology, as well as real data reported by the WHO in [51].

We note that, for this case, the model was used to predict the future evolution of the disease. To do so, we ran the Be-CODiS model configured with the estimated parameters until the end of the simulated epidemic (i.e., until $t > 0$ such that $E_i(t) + I_i(t) + H_i(t) < 1$ for $i = 1, \dots, N_{co}$).

As can be seen in Fig. 3(a), considering the "best" result, the obtained forecast fits quite accurately with real data. According to this forecast published in [13], the epidemic should end around the 1st of May, 2020, and the final numbers of cases and deaths should be around 3584 and 2453, respectively. According to the last situation report published in [49], the epidemic ended on the 14th of May, 2020, and the final numbers of cases and deaths were 3470 and 2287, respectively. Considering this update, the final relative errors between our forecast and reported data are 3.11% for the number of cases and 7.08% for the number of deaths, which indicates a good accuracy of our forecast.

Focusing on the "worst" result, shown in Fig. 3(b), the obtained forecast estimates quite well real data up to November 2019. The overestimation of this forecast in the last months may be due to the application of a new vaccine that complemented the previous one and was introduced on the 14th of November, 2019 (see [56]). We note that this is a classical difficulty that occurs during forecasting tasks, when unpredictable changes affect the dynamics of the observed phenomenon (see, e.g., [57]).

Analyzing the results reported in Table 6, we observe that:

- The estimated initial EVD effective contact rate β_I is lower than the one of the previous DR Congo outbreak (see Table 5) but similar to the one of the 2014–16 EVD epidemic (see [7]).
- The estimated efficiency of the control measures κ is lower than the one of the two previous outbreaks (see Table 5 and [7]). The low efficiency of the control measures could be explained by the particular features of the DR Congo geographical area affected by the outbreak. More precisely, according to the WHO reports [55], it is a conflictive zone with security problems and deficient health system. Additionally, its population distrusts response teams and has high mobility. These difficulties could explain the fact that, despite the extra vaccination campaign carried out from the 14th of November, 2019, the outbreak remained active until the 14th of May, 2020.
- The estimated value for the number of days between the start of the outbreak and the application of the control measures is $\lambda = 6.59$ days on average. According to this, the control measures seem to start being applied approximately the 6th of May, 2018. It seems to indicate that some control measures were applied since the beginning of the outbreak by the local authorities, maybe because of the proximity with the previous 2018 Ebola outbreak in DR Congo studied in experiment **DRC1** [51].
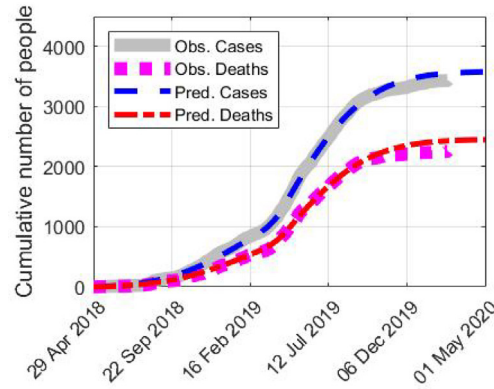
(a) Considering the "best" result ($\mathbf{x}_{\mathrm{best}} \in X$)
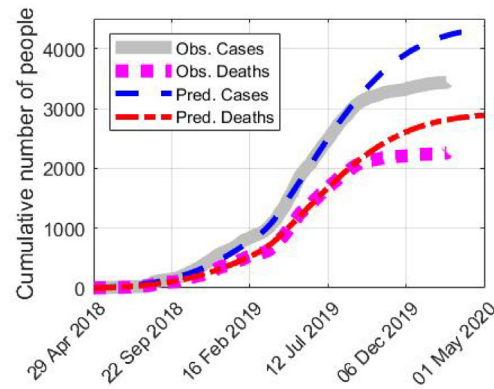
(b) Considering the "worst" result ($\mathbf{x}_{\mathrm{worst}} \in X$)

Fig. 3. Comparison between the observed and the predicted cumulative number of cases and deaths considering the experiment **DRC2** described in Section 4.1.2.

- The number of days between the start of the outbreak and the restoration of the control measures after the violent attacks is $\lambda_2 = 353$ days on average. Then, the control measures interrupted on the 24th of February, 2019, seem to be restored approximately on the 17th of April, 2019. According to situation reports [49], some Ebola treatment centers remained closed until the 30th of March, 2019, because of the damages provoked by the attacks. Moreover, the second International Health Regulation (IHR) Emergency Committee[2] on the EVD outbreak in North Kivu and Ituri Provinces of the DR Congo, was convened on 12 April 2019 [49]. Thus, the restoration date estimated by our method seems to be reasonable.
- The parameter dividing the sanitary country indicator in Eq. (9) is $C_{\mathrm{SAN}} = 6.07E+03$ on average. It is slightly lower than the one estimated for the 2014–16 outbreak, which was 9.06E+03 (see [7]). It is consistent with the data reported by the WHO [49], which estimated that the fatality rate of the 2018–20 epidemic is between 52% and 67% and, then, lower than the one of the 2014–16 outbreak before the application of control measures, which was around 68%.
- The estimated percentage δ of the fatality rate that can be reduced by control measures is 48.6% on average. It is slightly lower than the value used for the 2014–16 Ebola outbreak in [7], which was 53%. It is realistic because, according to the WHO [11], in the 2014–16 epidemic, the fatality rate was reduced until reaching a final percentage of 40%. However, for the 2018–20 outbreak, the current fatality rate is 66% (see [49]).
- The estimated value for the parameter that multiplies the effect of the control measures in Eq. (8) is $C_m = 7.78E-01$ on average. Then, the control measures for the 2018–20 outbreak seem to have a lower impact than the control measures applied for the 2014–16 epidemic (i.e., $C_m = 1$, see [7]). Again, this is coherent with the fact that in 2014–16 the fatality rate experienced a greater reduction than the one of 2018–20 (see [11,49]).

---

[2] The IHR Emergency Committee is an expert meeting convened by the General-Director of the WHO to seek appropriate recommendations about a public health emergency of international concern [58].

Table 7

Description and range of the parameters to be optimized for the experiment involving a set of countries and presented in Section 4.2.

Notation	Range	Description
ϵ_τ	[0.0, 10.0]	Numerical filter threshold to avoid transmission between countries with low number of individuals in E
ϵ_E	[0.0, 10.0]	Numerical filter threshold to avoid infection process in countries with low number of individuals in E
C_τ	[0.0, 10000]	Parameter multiplying the movement matrix $\tau_{i,j}$ to increase/decrease the movement fluxes
k_1 k_2 k_3 k_4	[0.0, 1.0] [−1.0, 1.0] [0.0, 1.0] [−1.0, 1.0]	Parameters involved in the control measures κ (see Eq. (7))
b_1 b_2 b_3 b_4	[0.0, 150.0] [−25.0, 25.0] [0.0, 1.0] [−1.0, 1.0]	Parameters involved in the transmission rate β_l (see Eq. (5))
δ	[0.0, 1.0]	Proportion of the fatality rate that can be reduced by control measures
C_m	[0.0, 1.0]	Parameter multiplying the control measures value to increase its effect on the fatality rate reduction
C_{SAN}	[50.0, 9000.0]	Parameter dividing the sanitary country indicator in Eq. (9)
λ_G λ_L λ_{SL}	[0.0, 700.0] [0.0, 700.0] [0.0, 700.0]	First day of application of control measures (day) in Guinea, Liberia and Sierra Leone, respectively

As can be seen in Table 6, most decision variables present a large deviation. This fact highlights the difficulty and complexity of estimating the global dynamics of the outbreak. In particular, the existence of a unique solution is not guaranteed and the objective functions could admit several local minima [59]. Thus, our multi-objective approach may return a variety of solutions with different compromises among the considered objectives. In this work, we propose a particular criterion to automatically select the suitable solution (i.e. the closest solution to the origin in the image space) but, of course, if the DM prefers another criterion, it could be easily implemented.

Again, according to the epidemic situation reported in the literature [49], the proposed methodology returned parameter values that have a reasonable biological and sociological meaning. Furthermore, considering those values, the model returns outputs that accurately fit real observations. Additionally, the forecast generated in July 2019 by using those parameter values predicted well real observations.

4.2. Experiments considering a set of countries

Here, the multi-objective estimation methodology is applied to find a set of values for the Be-CoDiS epidemiological parameters such that the model predictions fit the international spread of the Ebola Virus Disease (EVD) occurring during the 2014–2016 epidemic.

According to the formulas presented in Section 3.2 and used to describe some disease parameters depending on the country, the multi-objective optimization problem considered here has 17 decision variables. They are described in Table 7, where we also report their admissible ranges used in the optimization process. All these limits are based on values reported in [7].

During the 2014–16 EVD epidemic, the most affected countries were Guinea, Liberia and Sierra Leone (i.e., considering a threshold for the cumulative number of cases of 1000). Therefore, to guarantee an accurate description of the disease spread inside those countries, separate objective functions for each of them are considered in the formulation of the optimization problem (18) (see Step 3 in Section 3.2). More specifically, the following 10 objective functions have been considered:

- f_1 and f_2, which are the average of the absolute errors in the cumulative number of cases (i.e., Eq. (15)) and deaths (i.e., Eq. (16)), respectively, for all the countries except the most affected ones;
- f_3, f_4, f_5, f_6, f_7 and f_8, which are the relative errors in the number of cumulative cases (i.e., Eq. (13)) and in the number of deaths (i.e., Eq. (14)) for Guinea (f_3 and f_4), Liberia (f_5 and f_6), and Sierra Leone (f_7 and f_8);
- f_9 and f_{10}, which stand for the number of non-affected countries that the model predicts as affected and the number of affected countries predicted as non-affected, respectively (i.e., Eqs. (17)).

Table 8

Solutions obtained during the experiment that is reported in Section 4.2. We present the values corresponding to the "best" ($\mathbf{x}_{\text{best}}$) and the "worst" ($\mathbf{x}_{\text{worst}}$) results in X, and the average ($\text{Av}(X)$) and standard deviation ($\text{Dev}(X)$) values. Recall that X is the set formed by one solution from each run according to the selection criterion.

West Africa (2014–16 outbreak)				
Decision variables	$\mathbf{x}_{\text{best}}$	$\mathbf{x}_{\text{worst}}$	$\text{Av}(X)$	$\text{Dev}(X)$
ϵ_τ	6.90E+00	2.29E+00	5.06E+00	3.35E+00
ϵ_E	4.49E+00	2.17E+00	4.15E+00	2.24E+00
C_τ	1.41E+03	1.45E+03	1.62E+03	2.86E+02
k_1	1.69E−01	7.99E−01	2.30E−01	2.83E−01
k_2	−1.97E−01	2.99E−01	−6.20E−02	6.23E−01
k_3	6.07E−02	9.02E−01	5.63E−01	3.95E−01
k_4	7.10E−03	3.00E−01	2.58E−01	3.97E−01
b_1	2.41E+01	6.85E+00	6.55E+01	6.03E+01
b_2	9.36E+00	2.22E+01	9.07E+00	1.01E+01
b_3	7.00E−02	6.14E−01	2.20E−01	2.13E−01
b_4	1.29E−01	−7.04E−01	−9.55E−02	3.30E−01
δ	5.58E−01	3.32E−01	1.74E−01	2.02E−01
C_m	7.98E−01	4.31E−01	4.21E−01	3.55E−01
C_{SAN}	1.05E+02	1.23E+02	1.03E+02	8.92E+00
λ_G	2.60E+02	3.77E+02	3.56E+02	4.46E+01
λ_L	2.04E+02	2.65E+02	2.25E+02	4.06E+01
λ_{SL}	2.20E+02	2.82E+02	2.50E+02	4.71E+01

Since the number of decision variables and objectives is higher than in the previous one-country experiments, for solving this particular multi-objective problem, the ParWASF-GA optimizer has been configured with a population size of 770 individuals and 20 000 iterations. Moreover, ten independent runs have been performed (see Table 4).

Then, following the DM selection criterion for each one of the ten executions, we have obtained ten different solutions $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_{10}\}$ that better fit his/her preferences. From X, we have extracted the "best" and the "worst" results, $\mathbf{x}_{\text{best}}, \mathbf{x}_{\text{worst}} \in X$, as done in the previous experiments (see Section 4.1).

As in previous experiments, in Table 8, we report the parameter values for both the "best" and the "worst" solutions, and the average value and the standard deviation of each parameter. According to these results, it seems that the optimal decision vector can have significantly different values depending on the run. As the number of decision variables is large, there can exist different combinations of values providing a reasonable fitting for all the countries. Additionally, in Table 9, we report the values of the objective functions for those "best" and "worst" solutions. According to these values, it seems that the "worst" run provides a better estimation for the countries where the epidemic magnitude was low (i.e., lower values of f_1 and f_2). However, the "best" run result seems to better approach the spread of the disease in the most affected countries, Guinea, Liberia and Sierra Leone (i.e., lower values of functions f_3 to f_8). Furthermore, the "best" run predicts 8 countries as affected when they were not really affected (i.e., $f_9 = 8$), while the "worst" solution fails in 14 countries. Both the "best" and the "worst" solutions are able to detect all the affected countries (i.e., $f_{10} = 0$ in both cases).

Remark 2. This analysis highlights the fact that, in multi-objective problems, if there exist confronted objectives, then the problem may not admit a unique optimal solution improving simultaneously all the objectives, but a variety of solutions with different compromises among these objectives (see the standard deviation values in Table 8). Indeed, these solutions are non-comparable in the sense that we cannot establish which of them is better attending to the dominance criteria used in the multi-objective framework (see Section 2.1). So, it is important to remark that the terminology of "best" and "worst" is a subjective decision-making criterion.

In Fig. 4, the evolution of the cumulative number of cases and deaths predicted by Be-CoDiS with the "best" and "worst" sets of parameters is represented, as well as the real evolution of the Ebola virus disease, from the 6th of December, 2013, to the 27th of March, 2016. As can be seen, there is not a noticeable difference between the results of the "best" and the "worst" solutions for Guinea, Liberia, and Sierra Leone, even though the "best" solution seems to provide a slightly more accurate prediction. Therefore, even in the worst case, the estimation methodology achieves a set of parameter values for the Be-CoDiS model such that the prediction fits simultaneously all the countries. We note that the obtained results are not as precise as in the one-country experiments (see Fig. 2). In this sense, it is important to take into account that, here, a large set of countries was simultaneously adjusted, with many parameters to be optimized.

For the sake of completeness, the impact of the disease in the remaining countries has been analyzed. In Fig. 5, the final cumulative numbers of cases predicted by the Be-CoDiS model using the "best" and the "worst" parameter solutions are compared to the final numbers of cases observed in reality. Notice that only those countries which were really affected or predicted as affected by the model, have been included in this figure. As it was previously highlighted, the so-called "best" solution exhibits a lower number of countries predicted as affected than they actually are. At the same time, in general, it returns more cases per country than the "worst" solution. Indeed, the "worst" solution forecasts that the disease spreads to a greater number of countries many of them with two cases per country at maximum. In particular, it underestimates

Table 9

Solutions obtained during the experiment that is reported in Section 4.2. We present the objective function values ($\mathbf{f}(\mathbf{x}_{\text{best}})$ and $\mathbf{f}(\mathbf{x}_{\text{worst}})$) corresponding to the "best" and the "worst" results in X, respectively. Recall that X is the set formed by one solution from each run according to the selection criterion.

Objective functions	$\mathbf{f}(\mathbf{x}_{\text{best}})$	$\mathbf{f}(\mathbf{x}_{\text{worst}})$
f_1	2.68E−06	6.66E−07
f_2	1.25E−06	2.35E−07
f_3	9.17E−02	1.18E−01
f_4	8.63E−02	1.27E−01
f_5	8.91E−02	1.24E−01
f_6	9.18E−02	1.35E−01
f_7	9.99E−02	1.37E−01
f_8	8.96E−02	1.37E−01
f_9	8.00E+00	1.40E+01
f_{10}	0.00E+00	0.00E+00
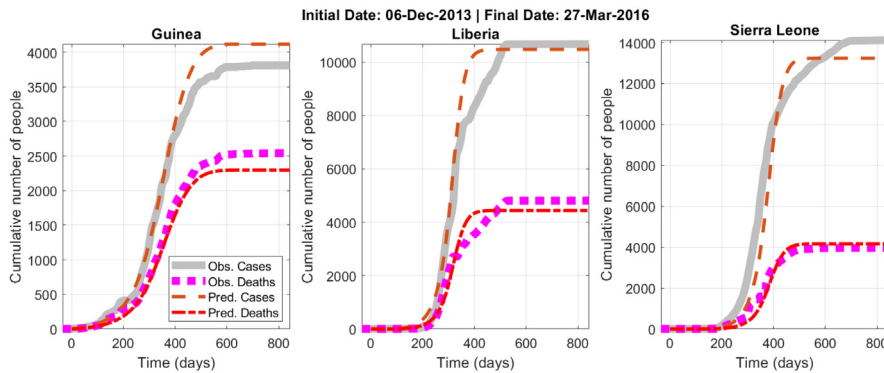
(a) Considering the "best" result ($\mathbf{x}_{\text{best}} \in X$)
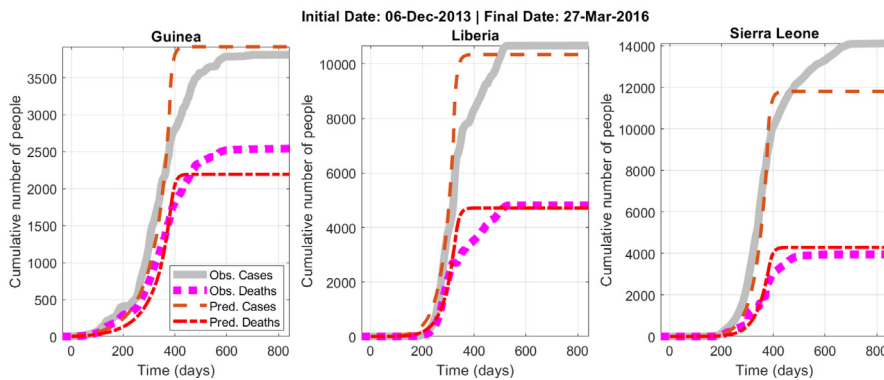
(b) Considering the "worst" result ($\mathbf{x}_{\text{worst}} \in X$)

Fig. 4. Comparison between the observed and the predicted cumulative number of cases and deaths for the most-infected countries: Guinea, Liberia and Sierra Leone.

the impact of the disease in Nigeria and Mali. Nevertheless, the most important conclusion of these results is that the predictions obtained by using the proposed estimation methodology seem to guess all the countries which were really affected. Detecting some extra affected countries is not always a negative point. Indeed, this information can be helpful in order to warn the authorities of those countries and to design preventive plans.

Finally, in order to highlight the advantages of the proposed estimation methodology, the current results have been compared to the previously published results of the Be-CoDiS model applied to the 2014–16 EVD epidemic (see [7]). In this previous work, some epidemiological parameters were calibrated by hand and others were obtained by using some regression techniques. Those regression methods were applied to adjust each parameter independently (i.e., setting the remaining ones to fixed values or even running the system without all the features). Then, achieving a global set of values for all the parameters was a difficult task because, in most cases, after fitting one particular country, the predictions for
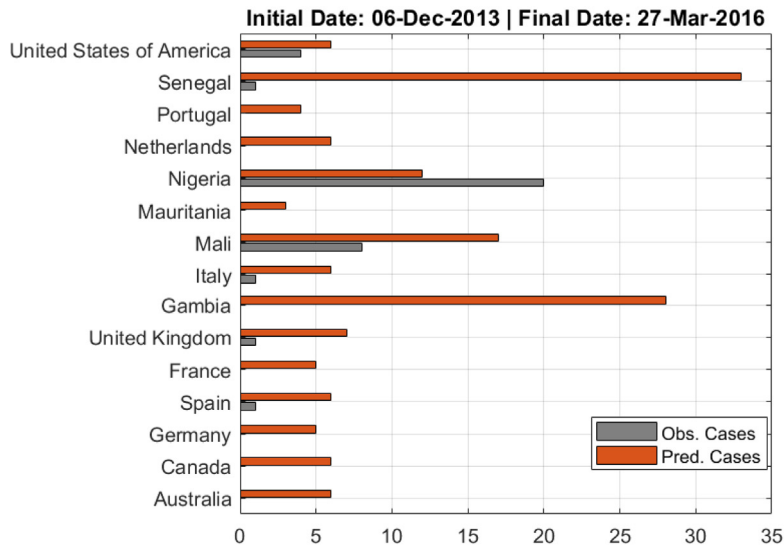
(a) Considering the "best" result ($\mathbf{x}_{\text{best}} \in X$)
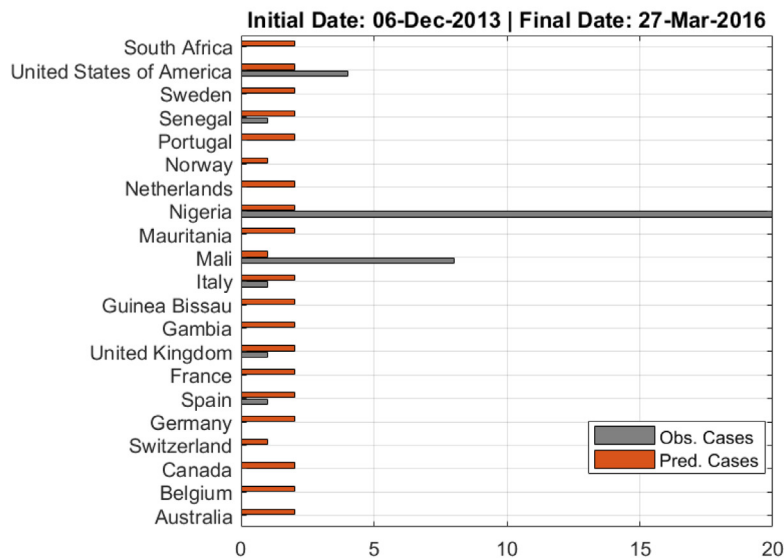
(b) Considering the "worst" result ($\mathbf{x}_{\text{worst}} \in X$)

Fig. 5. Comparison between the observed and the predicted cumulative number of cases for the countries where the EVD magnitude was low.

the remaining ones worsened. In spite of this, in [7], the authors achieved a successful calibration such that the model estimated a global magnitude of the outbreak similar to observed data. However, with those parameters, the Be-CoDiS model was not able to predict the times of infection and, thus, the global evolution of the simulated epidemic presents a delay when comparing it to the real situation (see Fig. 6, taken from [7]).

In Fig. 7, the global evolution obtained with the Be-CoDiS model using the novel estimation methodology is compared to the observed evolution of the Ebola outbreak. As previously, the results correspond to the "best" and the "worst" set of parameter values according to the decision-making strategy (see Sections 4.1 and 4.2). As can be seen, the multi-objective estimation methodology allows to achieve predictions that accurately fit the real behavior of the disease not only in
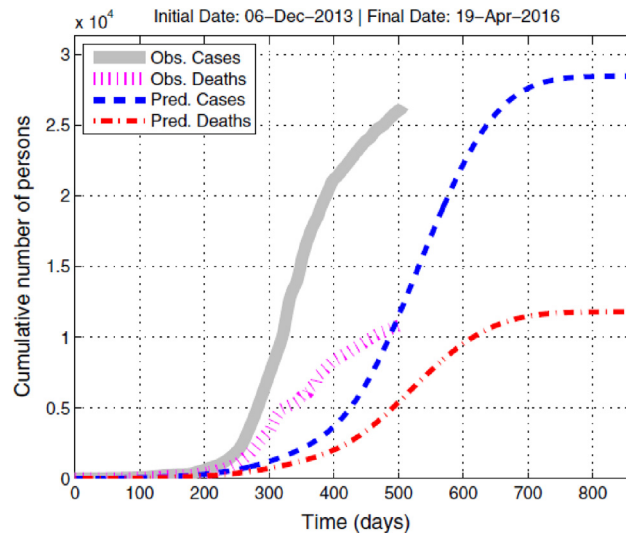
Fig. 6. Evolution of the observed and the predicted cumulative number of cases and deaths reported in [7]. This predicted evolution was obtained without using the multi-objective estimation methodology.

magnitude but also in time. Although we can obtain less accurate results (i.e., produced by the "worst" set of parameters), obtained results seem to succeed in guessing the critical time when the epidemic exploded or was controlled.

5. Conclusions

In this work, we have presented an original multi-objective approach to estimate the parameters of epidemiological models. To illustrate the performance of the proposed methodology, we have applied it to Be-CoDiS, which is a model that simulates the spread of human diseases inside and between countries. More precisely, we have estimated the Be-CoDiS parameters for different Ebola outbreaks from 2014 to 2020. Then, we have validated our approach by comparing the data obtained by the model, considering the estimated parameters, with real data reported by the WHO. Additionally, for the 2018–20 EVD outbreak in DR Congo, we have presented some forecasts of the disease evolution.

According to the obtained results, the proposed multi-objective estimation methodology is able to estimate the epidemiological parameters in such a way that the Be-CoDiS model describes accurately the disease behavior in both magnitude and time. Even in the worst outcomes of the optimization algorithm, the methodology seems to produce reasonable results.

The main advantages of the proposed methodology are that it is a general procedure which can be applied to any epidemiological model, and it works as an automatized tool that avoids the user to perform complex and time-consuming calibration processes. Furthermore, it allows to obtain an optimal set of solutions. These solutions provide different trade-off values for the parameters achieving diverse compromises among the objective functions. Thus, the DMs (epidemiologists or authorities) can deal with several scenarios with a single outcome from one execution. Then, they can choose the solution that they prefer attending to diverse selection criteria. However, they can also consider evaluating all the sets of parameters returned by the optimizer to see several possible scenarios of the disease evolution. This additional information may be valuable in order to apply the control measures.

Another interesting advantage of this general methodology is its versatility since the DM can customize: (i) the objective functions; and (ii) the selection criteria, to guide the parameter estimation process.

Since the methodology uses the parallel version of WASF-GA, the computational time may be reasonable, depending on the considered computer. Eventually, one can use the proposed methodology when new data are available to update the parameters and the predictions. Emphasizing again the versatility of the tool, users can also customize the optimization algorithm. Indeed, they can decide to implement any other competitive state-of-the-art algorithm that arises in the literature.

As a final remark, we highlight the fact that the methodology proposed here to estimate parameters of compartmental epidemiological models has been successfully applied to calibrate θ-SEIHRD models for studying the COVID-19 pandemic in several countries [26–28].

To conclude, we must underline the role played by mathematical modeling when we aim to use our calibrated model to anticipate changes in the disease dynamics, examine their causes, design suitable control strategies, and provide reliable forecasts. Even though the calibration process produces a good fitting, one should be aware of the underlying mechanisms of the phenomenon simulated by the model. It is important to note that mathematical modeling should be performed by
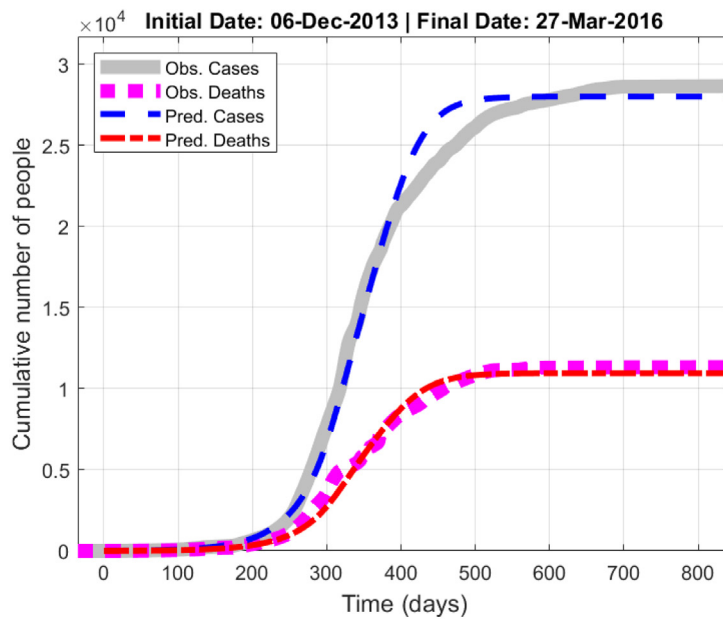
(a) Considering the "best" result ($\mathbf{x}_{\text{best}} \in X$)
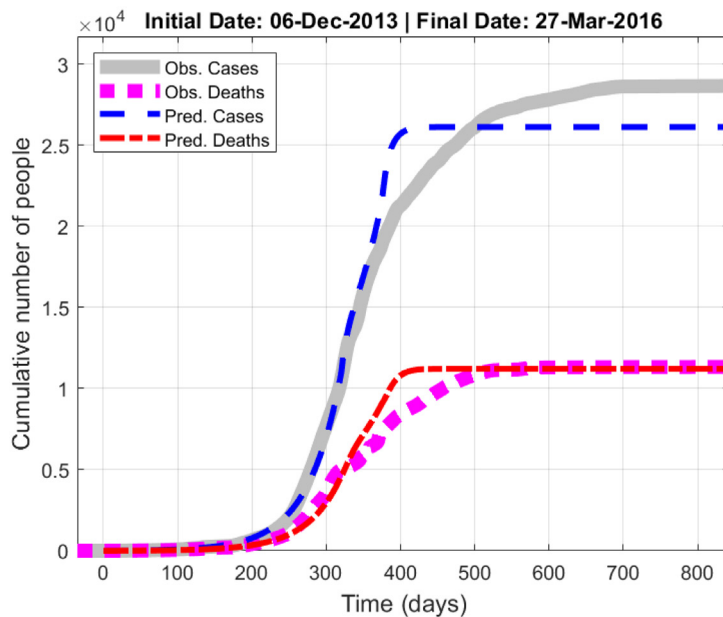
(b) Considering the "worst'" result ($\mathbf{x}_{\text{worst}} \in X$)

Fig. 7. Comparison between the observed and the predicted cumulative total (i.e. adding all countries) number of cases and of deaths.

taking into account all the relevant factors influencing the epidemiological dynamics (including external factors such as control measures), giving appropriately their place in the model, so that it can represent adequately the reality. Otherwise, the model could fit suitably the data using the calibration methodology but could not have biological meaning.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This research has been funded by grants from the Spanish "Ministerio de Ciencia e Innovación" (PID2021-123278OB-I00 and PID2019-106337GB-I00); "Junta de Andalucía" (UAL18-TIC-A020-B), in part financed by the European Regional Development Fund (ERDF); and the research group MOMAT (Ref. [910480]) supported by "Complutense University of Madrid".

References

[1] Ma S, Xia Y. Mathematical understanding of infectious disease dynamics. Lecture notes series, institute for mathematical sciences, World Scientific; 2009, URL: https://books.google.co.uk/books?id=r51pDQAAQBAJ.

[2] Brauer F, Castillo-Chavez C. Mathematical models in population biology and epidemiology, vol. 40. Springer; 2001.

[3] Hernández-Cerón N, Feng Z, Castillo-Chavez C. Discrete epidemic models with arbitrary stage distributions and applications to disease control. Bull Math Biol 2013;75(10):1716–46.

[4] Latha VP, Rihan FA, Rakkiyappan R, Velmurugan G. A fractional-order model for Ebola virus infection with delayed immune response on heterogeneous complex networks. J Comput Appl Math 2018;339:134–46.

[5] Ngwa GA, Shu WS. A mathematical model for endemic malaria with variable human and mosquito populations. Math Comput Modelling 2000;32(7–8):747–63.

[6] Yoshida N, Hara T. Global stability of a delayed SIR epidemic model with density dependent birth and death rates. J Comput Appl Math 2007;201(2):339–47.

[7] Ivorra B, Ngom D, Ramos AM. Be-CoDiS: A mathematical model to predict the risk of human diseases spread between countries—Validation and application to the 2014–2015 ebola virus disease epidemic. Bull Math Biol 2015;77(9):1668–704. http://dx.doi.org/10.1007/s11538-015-0100-x.

[8] Cvjetanović B, Grab B, Dixon H. Epidemiological models of poliomyelitis and measles and their application in the planning of immunization programmes 1982;60:405–422.

[9] Kruijshaar ME, Barendregt JJ, Hoeymans N. The use of models in the estimation of disease epidemiology 2002;80:622–628.

[10] Ministere de la Santé RDdC. Evolution de l'epidemie d'Ebola dans les provinces du Nord-Kivu et de L'Ituri. 2019, URL: https://mailchi.mp/sante.gouv.cd/ebola_kivu_21juil19.

[11] World Health Organization. Global Alert and Response: Ebola virus disease. 2016, URL: https://www.afro.who.int/health-topics/ebola-virus-disease.

[12] Ivorra B, Ramos AM. Application of the Be-CoDiS model to the 2018 ebola virus disease outbreak in the democratic republic of congo. Technical report, Complutense University of Madrid; 2018, URL: https://www.researchgate.net/publication/325035707_Application_of_the_Be-CoDiS_model_to_the_2018_Ebola_Virus_Disease_outbreak_in_the_Democratic_Republic_of_Congo.

[13] Ferrández MR, Ivorra B, Ortigosa PM, Ramos AM, Redondo JL. Application of the Be-CoDiS model to the 2018-19 ebola virus disease outbreak in the democratic republic of congo. Technical report, Complutense University of Madrid and University of Almería; 2019, http://dx.doi.org/10.13140/RG.2.2.13267.63521/2.

[14] Deuffic S, Buffat L, Poynard T, Valleron A-J. Modeling the hepatitis C virus epidemic in France. Hepatology 1999;29(5):1596–601.

[15] Jin F, Dougherty E, Saraf P, Cao Y, Ramakrishnan N. Epidemiological modeling of news and rumors on twitter. In: Proceedings of the 7th workshop on social network mining and analysis. ACM; 2013, p. 8.

[16] Nsoesie E, Mararthe M, Brownstein J. Forecasting peaks of seasonal influenza epidemics. PLoS Curr 2013;5.

[17] Nsoesie EO, Beckman RJ, Shashaani S, Nagaraj KS, Marathe MV. A simulation optimization approach to epidemic forecasting. PLoS One 2013;8(6):e67164.

[18] Capaldi A, Behrend S, Berman B, Smith J, Wright J, Lloyd AL. Parameter estimation and uncertainty quantification for an epidemic model 2012;9:553–576.

[19] Taghizadeh L, Karimi A, Heitzinger C. Uncertainty quantification in epidemiological models for the COVID-19 pandemic 2020;125:104011.

[20] Gugole F, Coffeng LE, Edeling W, Sanderse B, De Vlas SJ, Crommelin D. Uncertainty quantification and sensitivity analysis of COVID-19 exit strategies in an individual-based transmission model 2021;17:e1009355.

[21] Carpio A, Pierret E. Uncertainty quantification in covid-19 spread: lockdown effects 2022;35:105375.

[22] Ruiz AB, Saborido R, Luque M. A preference-based evolutionary algorithm for multiobjective optimization: the weighting achievement scalarizing function genetic algorithm. J Global Optim 2015;62(1):101–29.

[23] Ferrández MR, Puertas-Martín S, Redondo JL, Ivorra B, Ramos AM, Ortigosa PM. High-performance computing for the optimization of high-pressure thermal treatments in food industry. J Supercomput 2018;75:1187–202. http://dx.doi.org/10.1007/s11227-018-2351-4.

[24] Ferrández MR, Redondo JL, Ivorra B, Ramos AM, Ortigosa PM. Preference-based multi-objectivization applied to decision support for High-Pressure Thermal processes in food treatment. Appl Soft Comput 2019;79:326–40. http://dx.doi.org/10.1016/j.asoc.2019.03.050.

[25] Ferrández MR, Redondo JL, Ivorra B, Ramos AM, Ortigosa PM, Paechter B. Improving the performance of a preference-based multi-objective algorithm to optimize food treatment processes. Eng Optim 2019;1–18. http://dx.doi.org/10.1080/0305215X.2019.1618289, arXiv:https://doi.org/10.1080/0305215X.2019.1618289.

[26] Ivorra B, Ferrández MR, Vela-Pérez M, Ramos AM. Mathematical modeling of the spread of the coronavirus disease 2019 (COVID-19) taking into account the undetected infections. The case of China. Commun Nonlinear Sci Numer Simul 2020;88:105303. http://dx.doi.org/10.1016/j.cnsns.2020.105303.

[27] Ramos AM, Ferrández MR, Vela-Pérez M, Kubik AB, Ivorra B. A simple but complex enough θ-SIR type model to be used with COVID-19 real data. Application to the case of Italy. Physica D 2021;421:132839. http://dx.doi.org/10.1016/j.physd.2020.132839.

[28] Ramos AM, Vela-Pérez M, Ferrández MR, Kubik AB, Ivorra B. Modeling the impact of SARS-CoV-2 variants and vaccines on the spread of COVID-19. 2021, http://dx.doi.org/10.1016/j.cnsns.2021.105937.

[29] Redondo JL, Fernández J, Álvarez JD, Arrondo AG, Ortigosa PM. Approximating the Pareto-front of a planar bi-objective competitive facility location and design problem. Comput Oper Res 2015;62:337–49.

[30] Bechikh S, Kessentini M, Said LB, Ghédira K. Chapter four - preference incorporation in evolutionary multiobjective optimization: A survey of the state-of-the-art. Adv Comput 2015;98:141–207.

[31] Thiele L, Miettinen K, Korhonen PJ, Molina J. A preference-based evolutionary algorithm for multi-objective optimization. Evol Comput 2009;17(3):411–36.

[32] Coello CAC, Lamont GB, Van Veldhuizen DA. Evolutionary algorithms for solving multi-objective problems. Genetic and evolutionary computation, 2nd ed.. New York: Springer; 2007.

[33] Deb K. Multi-objective optimization using evolutionary algorithms. John Wiley & Sons; 2001.

[34] Deb K, Sindhya K, Hakanen J. Multi-objective optimization. In: Decision sciences: Theory and practice. CRC Press; 2016, p. 145–84.

[35] Zitzler E, Knowles J, Thiele L. Quality assessment of Pareto set approximations. In: Branke J, Deb K, Miettinen K, Slowinski R, editors. Multiobjective optimization. Interactive and evolutionary approaches, vol. 5252. Lecture notes in computer science, Berlin, Germany: Springer; 2008, p. 373–404.

[36] Branke J. MCDA and multiobjective evolutionary algorithms. In: Multiple criteria decision analysis. Springer; 2016, p. 977–1008.

[37] Goldberg DE. Genetic algorithms in search, optimization and machine learning. 1st ed.. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc.; 1989.

[38] Brauer F, Castillo-Chavez C. Mathematical models in population biology and epidemiology. Texts in applied mathematics, Springer; 2001.

[39] Legrand J, Grais R, Boelle P, Valleron A, Flahault A. Understanding the dynamics of Ebola epidemics. Med Hypotheses 2007;135(4):610–21.

[40] Meltzer MI, Atkins C, Santibanez S, Knust B, Petersen B, Ervin E, Nichol S, Damon I, Washington M. Estimating the future number of cases in the ebola epidemic - Liberia and Sierra Leone, 2014–2015. In: Centers for disease control. MMWR / early release, vol. 63. 2014.

[41] Peters CJ, Peters JW. An introduction to ebola: The virus and the disease. J Infect Dis 1999;179(Supplement 1):ix–xvi.

[42] Team WHOR. Ebola virus disease in west Africa - the first 9 months of the epidemic and forward projections. N Engl J Med 2014;371:1481–95.

[43] World Bank Group. World Bank Open Data: Free and open access to global development data. 2018, URL: http://data.worldbank.org.

[44] Lekone P, Finkenstädt B. Statistical inference in a stochastic epidemic SEIR model with control intervention: Ebola as a case study. Biometrics 2006;62(4):1170–7.

[45] Guy J, Nikola S. Quantifying global international migration flows. Science 2014;343(6178):1520–2.

[46] Nikola S, Guy JA, Bauer R. The global flow of people. 2014, URL: http://www.global-migration.info/.

[47] United Nations, Department of Economic and Social Affairs. Trends in international migrant stock: Migrants by destination and origin. 2012, URL: https://www.un.org/en/development/desa/population/migration/data/index.asp, United Nations database, POP/DB/MIG/Stock/Rev.2012, Table 7.

[48] World Health Organization. Ebola virus disease. Democratic Republic of the congo. External situation report 17: Declaration of the end of the outbreak. 2018, URL: https://apps.who.int/iris/bitstream/handle/10665/273348/SITREP_EVD_DRC_20180725-eng.pdf?ua=1.

[49] World Health Organization. Ebola health update - DRC, 2019. Situation reports. 2019, URL: https://www.afro.who.int/health-topics/ebola-virus-disease.

[50] Ivorra B, Ngom D, Ramos AM. Stability and sensitivity analysis of the epidemiological model Be-CoDiS predicting the spread of human diseases between countries 2020;2020. URL: http://ejde.math.txstate.edu.

[51] World Health Organization. Ebola situation reports: Democratic Republic of the Congo (Archive). 2019, URL: https://www.afro.who.int/health-topics/ebola-virus-disease.

[52] World Health Organization. Statement on the 1st meeting of the IHR emergency committee regarding the ebola outbreak in 2018. 2018, URL: https://www.who.int/news-room/detail/18-05-2018-statement-on-the-1st-meeting-of-the-ihr-emergency-committee-regarding-the-ebola-outbreak-in-2018.

[53] World Health Organization. Ebola virus disease – Democratic Republic of the Congo. 2018, URL: http://apps.who.int/iris/bitstream/handle/10665/272536/SITREP-EVD-DRC-20180514.pdf.

[54] World Heath Organization. Ebola virus disease – Democratic Republic of the Congo: Update on Ring Vaccination. 2018, URL: https://www.who.int/emergencies/crises/cod/DRC-ebola-disease-outbreak-response-plan-10august2018-1612-EN.pdf.

[55] World Health Organization. Ebola virus disease – Democratic Republic of the Congo. Disease outbreak news: Update 18 April 2019. 2019, URL: https://apps.who.int/iris/bitstream/handle/10665/312080/SITREP-EVD-DRC-20192404-eng.pdf.

[56] World Health Organization. Second Ebola vaccine to complement "ring vaccination" given green light in DRC. 2019, URL: https://www.who.int/news-room/detail/23-09-2019-second-ebola-vaccine-to-complement-ring-vaccination-given-green-light-in-drc.

[57] Gómez S, Ivorra B, Ramos AM. Optimization of a pumping ship trajectory to clean oil contamination in the open sea. Math Comput Modelling 2011;54(1):477–89.

[58] World Heath Organization. International health regulations committees and expert roster. 2020, URL: https://www.who.int/ihr/procedures/ihr_committees/en/.

[59] Ferrández MR. Modeling and optimization of health problems via high-performance computing (Ph.D. thesis), University of Almería; 2019, URL: https://www.educacion.gob.es/teseo/mostrarRef.do?ref=1750434.