Chaos, Solitons and Fractals 164 (2022) 112671

Contents lists available at ScienceDirect

Chaos, Solitons and Fractals

journal homepage: www.elsevier.com/locate/chaos

Stochastic SIR model predicts the evolution of COVID-19 epidemics from public health and wastewater data in small and medium-sized municipalities: A one year study

Manuel Pájaro [a,b,*], Noelia M. Fajar [a], Antonio A. Alonso [a], Irene Otero-Muras [a,c,**]

[a] BioProcess Engineering Group, IIM-CSIC. Spanish National Research Council, Eduardo Cabello 6, 36208, Vigo, Spain
[b] Universidade da Coruña, CITIC research center, Department of Mathematics, Campus Elviña s/n, A Coruña, 15071, Spain
[c] Institute for Integrative Systems Biology I²SysBio (UV, CSIC) Spanish National Research Council, 46980, València, Spain

ARTICLE INFO

Keywords:
SIR model
COVID-19
SARS-coV-2
Stochastic mechanistic model
Chemical master equation
Stochastic simulation algorithm

ABSTRACT

The level of unpredictability of the COVID-19 pandemics poses a challenge to effectively model its dynamic evolution. In this study we incorporate the inherent stochasticity of the SARS-CoV-2 virus spread by reinterpreting the classical compartmental models of infectious diseases (SIR type) as chemical reaction systems modeled via the Chemical Master Equation and solved by Monte Carlo Methods. Our model predicts the evolution of the pandemics at the level of municipalities, incorporating for the first time (i) a variable infection rate to capture the effect of mitigation policies on the dynamic evolution of the pandemics (ii) SIR-with-jumps taking into account the possibility of multiple infections from a single infected person and (iii) data of viral load quantified by RT-qPCR from samples taken from Wastewater Treatment Plants. The model has been successfully employed for the prediction of the COVID-19 pandemics evolution in small and medium size municipalities of Galicia (Northwest of Spain).

1. Introduction

Since the emergence of the COVID-19 pandemics caused by the SARS-CoV-2 virus, great efforts have been made for the purposes of virus detection and epidemics forecasting. COVID-19 modeling approaches generally fall into one of the following categories, see [1]: statistic models for short-term forecasts; and mechanistic models, whether they are based on differential equations (like compartmental models) or agent-based [2,3], for analyzing the spread dynamics of SARS-CoV-2, investigating future possible scenarios and/or simulate interventions to control virus spreading. In this work, we introduce a stochastic mechanistic model valid for both testing scenarios and short-term forecasting. The model has been developed to study and predict the evolution of the pandemics at the level of municipalities, and it has been calibrated and tested during a one year study in Galicia (northwest of Spain) using measurements from the health system and viral load from wastewater samples. The model developed for SARS-CoV-2 can be easily adapted to the surveillance of other pathogens and therefore, the methodology presented makes a significant contribution to wastewater based epidemiology [4].

Epidemiological compartmental deterministic models, like the Susceptible–Infected–Recovered (SIR) model firstly described by [5] (and extended versions of it) have been employed to predict COVID-19 spread [e.g. 6–10]. However, predictability issues arise and models (whether they are phenomenological, mechanistic, or agent-based) are not efficient to predict the COVID-19 pandemics in the long term [e.g. 11,12]. Model predictive control approaches have been proposed to efficiently deal with the uncertainty and predict the effects of mitigation and suppression strategies [13].

The unpredictable nature of the pandemic spread has been tackled, on the one hand from the perspective of deterministic chaos [e.g. 14–16] and, on the other hand, using stochastic models [e.g. 17,18]. Dynamic stochastic models for COVID-19 spread prediction can be broadly categorized into: (i) stochastic differential equations based in classical SIR models [8,17], and (ii) compartmental models combined with Mote Carlo methods [6,19–21].

Here we use a reinterpretation of the classical compartmental modeling of infectious diseases (SIR type models) as chemical processes, which are inherently stochastic and governed by the Chemical Master Equations (CME). The CME describes the evolution in time of probability distributions [22,23], and the Stochastic Simulation Algorithm (SSA) by [24] can be used to compute exact realizations (time

* Corresponding author at: Universidade da Coruña, CITIC research center, Department of Mathematics, Campus Elviña s/n, A Coruña, 15071, Spain.
** Corresponding author at: Institute for Integrative Systems Biology I²SysBio (UV, CSIC) Spanish National Research Council, 46980, València, Spain.
E-mail addresses: manuel.pajaro@udc.es (M. Pájaro), irene.otero.muras@csic.es (I. Otero-Muras).

https://doi.org/10.1016/j.chaos.2022.112671
Received 10 June 2022; Received in revised form 24 August 2022; Accepted 4 September 2022
Available online 7 September 2022

course trajectories) of the CME. Data series of new infected persons, as provided by the public health systems, are indeed realizations of a stochastic process or random walks on the positive integers. Therefore, approaches based on a CME equation solved by the SSA algorithm are particularly convenient to develop COVID-19 predictive models [25–27].

Data from the health system have an inherent delay (from the time of infection until the positive case is reported). Moreover, the significant percentage of asymptomatic cases characteristic of the COVID-19 pandemics, hampers the prompt detection by the health systems (efforts including screening or contact tracing have been implemented to overcome this difficulty). In this regard, wastewater have been proven to be a good complementary tool for COVID-19 surveillance. The analysis of SARS-CoV-2 viral load in sewage at wastewater treatment plants can be interpreted as one pooled test for the area where they are located. Pooled or group testing is an excellent tool for surveillance of diseases spread in animals and humans [28], indicated also to expand COVID-19 surveillance [29]. Different studies confirmed wastewater monitoring as a convenient complementary approach to COVID-19 surveillance and testing strategies [30–32] and, in fact, viral RNA detection in Wastewater Treatment Plants (WWTPs) has shown an anticipative capacity with respect to the cases reported by the health system in several studies [e.g. 33–36].

In this work, we developed a stochastic SIR model with good predictive capacity which incorporates the data from health system and wastewater analysis. The model has been calibrated and tested during a one year study with health system data and wastewater analysis data from five different small-medium size municipalities in Galicia (Northwest of the Iberian peninsula) [36]. We implemented the model in a software package that can be used for predictions in different scenarios and forecasting.

- The model is a stochastic SIR: importantly, in stochastic mechanistic models uncertainty increases as we move into the future (like in the case of statistical models and unlike in the case of mechanistic deterministic models).
- The model is simple in terms of number of states and parameters, and it shows a good predictive capacity with very few parameters. Simplest SIR models were previously reported to perform better than other models with greater complexities [9].
- The model integrates, for the first time to the best of authors knowledge, SARS-CoV-2 viral load data from WWTPs within a mechanistic model with predictive capacity.
- The model is very robust (we obtain similar values of the parameters for the WWTPs tested).
- The model is capable to predict superspread events.
- The model can be used to test different scenarios and for forecasting in small and medium size municipalities with WWTPs in time horizons of 7–10 days.

In the next section we introduce the model with its different extensions. First, we develop a simple stochastic SIR (which proves to correctly predict the evolution of infected individuals from health system in small and medium-size municipalities). Then, we extend the model with the capability of incorporating variable degradation rates (SIRv). In order to be able to predict superspread events, we also introduce a SIR model with jumps (SIRj) allowing for the infection of various persons at the same time from one infected individual. Finally, we incorporate the WWTP data together with the public health system data into an integrated model (SIRO).

The SIR model and its extensions (SIRv, SIRj and SIRO) have been developed and tested in the context of COVID-19 surveillance of small and medium size municipalities in the Northwest of the Iberian peninsula [36] from June 2020 to December 2021.

2. Methods

In this section we introduce the models developed to predict the evolution of the COVID-19 pandemics. Our aim was to obtain a good predictive capability with the simplest possible model. Importantly, our model is intended to predict the evolution at the local level (municipalities), and for short term time horizons (7–10 days maximum). First, we start with a stochastic SIR model with standard rate expressions.

2.1. SIR model

We consider a total population of N individuals, distributed in the following types (or compartments):

- **Susceptible**, S, susceptible to acquire the disease.
- **Infected**, I, capable of spreading the disease to a susceptible person.
- **Recovered**, R, recuperated/dead from disease.

Considering the three categories S, I and R, we can represent their interactions as follows:

$$\begin{aligned} S + I &\xrightarrow{\beta \frac{S}{N} I} 2I \\ I &\xrightarrow{\alpha I} R \end{aligned} \tag{1}$$

with β and α being parameters (infection and recovery rate constants, respectively). The equation governing the dynamics of this process taking into account its inherent stochasticity is the Chemical Master Equation (CME) [22]. Let $\boldsymbol{x} = (S, I, R)$ be the states of the system and $\mathcal{P} : [0\ N]^3 \times \mathbb{R}^+ \to [0\ 1]$ be a probability density function of the state $\boldsymbol{x}$ at time $t \geq 0$. The CME governing the process (1) reads:

$$\begin{aligned} \frac{\mathrm{d}\mathcal{P}(\mathbf{x},t)}{\mathrm{d}t} &= \beta \frac{S+1}{N} I \mathcal{P}(S+1, I-1, R, t) - \beta \frac{S}{N} I \mathcal{P}(\mathbf{x}, t) \\ &\quad + \alpha (I+1) \mathcal{P}(S, I+1, R-1, t) - \alpha I \mathcal{P}(\mathbf{x}, t). \end{aligned} \tag{2}$$

We use the Stochastic Simulation Algorithm (SSA) [24] to simulate realizations of the SIR process (1). A trajectory of a single SSA simulation (or realization) is an exact sample from the probability function that is the solution of the CME (2), therefore, the solution of the CME can be approximated by a reasonable number of realizations.

In this way, the number of infected persons reported daily by the health system can be considered as single realization of the SSA for the SIR process (1). To calibrate with data, make predictions or forecasting for a given time horizon, we use the mean and standard deviation of $10^3 - 10^4$ SSA realizations.

2.1.1. SIR with variable infection rate

The infection rate β for the SIR model is constant. The SIR model with constant parameters cannot capture a turning point in the evolution of the pandemics unless the number of susceptible remains a limiting factor (which was never the case at least in the first two years of the COVID-19 pandemics). However, in many territories (including the municipalities under study) a series of policies (including lockdowns, travel bans, capacity limitations for social gatherings and other restrictions) have been applied by the authorities, at different stages of the pandemics, to decrease the infection rate. Therefore, time course data of infected people used for calibration might show changes of tendencies (turning points) that cannot be captured correctly by the standard SIR. Here we define a turning point (T_{TP}) as a point in time in which a (sustained) change in the sign of the slope of the total number of infected individuals is detected.

For calibration purposes, when the data show a turning point, we propose an extended SIR model starting from (1) with a variable infection rate β defined as:

$$\beta = \begin{cases} \beta_0 & \text{if } t < T_{TP} \\ \beta_1 & \text{if } t \geq T_{TP}, \end{cases} \tag{3}$$

where the t is the simulation time, and T_{TP} is a new parameter of the SIR model representing the turning point date. Note that, if $\beta_0 = \beta_1$, SIRv is equivalent to the SIR model.

2.1.2. SIR with jumps

Standard SIR models consider the infection of one person at a time (the stoichiometric S/I ratio is 1 : 1). This is the reason why SIR models cannot capture the jumps observed in time course data from SARS-CoV-2 infected individuals. One infected individual can infect more than one susceptible person as a result of the same interaction event (in this case the stoichiometric S/I ratio is n : 1 with $n > 1$, n being a positive integer). In order to consider these bursting (superspreading) events, we generalize the reactions associated with the SIR model (1) as follows:

$$\begin{aligned} nS + I &\xrightarrow{\beta \frac{\prod_{i=0}^{n-1}(S-i)}{N^n} I} (n+1)I \\ I &\xrightarrow{\alpha I} R \end{aligned} \tag{4}$$

In this work we consider $n = 5$ which is compatible with the mass gathering restrictions established by the health authorities in the period and geographic area under study. Note that, for $n = 1$, SIRj is equivalent to the simplest SIR model proposed.

Let $\boldsymbol{x} = (S, I, R)$ be the system states and $\mathcal{P} : [0\ N]^3 \times \mathbb{R}^+ \to [0\ 1]$ be a probability density function of being in state $\boldsymbol{x}$ at time $t \geq 0$. The associated CME to the reaction set (4) reads:

$$\begin{aligned} \frac{\mathrm{d}\mathcal{P}(\mathbf{x},t)}{\mathrm{d}t} &= \beta \frac{S+1}{N} I\mathcal{P}(S+1, I-1, R, t) - \beta \frac{S}{N} I\mathcal{P}(\mathbf{x},t) \\ &\sum_{k=2}^{n} \Big(\beta \frac{\prod_{i=1}^{k}(S+k)}{N^k} I\mathcal{P}(S+k, I-k, R, t) \\ &-\beta \frac{\prod_{i=0}^{k-1}(S-i)}{N^k} I\mathcal{P}(\mathbf{x},t) \Big) \\ &+\alpha(I+1)\mathcal{P}(S, I+1, R-1, t) - \alpha I\mathcal{P}(\mathbf{x},t). \end{aligned} \tag{5}$$

Note that this CME is an extension of the one obtained for the simplest SIR model, Eq. (2), by adding the new terms related to the new reactions (more than one infected at the same time) in the second line of expression (5).

2.2. SIRO model

The SIRO model is an adaptation of the SIR formulation described in Section 2.1 to incorporate the number of infected people (I) observed through the viral load monitoring in WWTPs, together with the subset of the total infected people (O) observed by the public health system, through the following set of reactions:

$$\begin{aligned} S + I &\xrightarrow{\beta \frac{S}{N} I} 2I \\ I &\xrightarrow{\alpha_I I} R \\ I &\xrightarrow{\gamma I} I + O \\ O &\xrightarrow{\alpha_O O} R_O \end{aligned} \tag{6}$$

where S, I and R are the susceptible, total infected and total recovered individuals in a given municipality. The infected individuals detected and reported by the health system are denoted by O, whereas R_O are recovered individuals that had been previously reported as infected. Let $\boldsymbol{x} = (S, I, R, O, R_O)$ be the states of the system and $\mathcal{P} : [0\ N]^5 \times \mathbb{R}^+ \to [0\ 1]$ be a probability density function of the state $\boldsymbol{x}$ at time $t \geq 0$. The CME governing the dynamics of the process described by the set of reactions in (6) reads:

$$\begin{aligned} \frac{\mathrm{d}\mathcal{P}(\mathbf{x},t)}{\mathrm{d}t} &= \beta \frac{S+1}{N} I\mathcal{P}(S+1, I-1, R, O, R_O, t) - \beta \frac{S}{N} I\mathcal{P}(\mathbf{x},t) \\ &+\alpha_I(I+1)\mathcal{P}(S, I+1, R-1, O, R_O, t) - \alpha_I I\mathcal{P}(\mathbf{x},t) \\ &+\gamma I\mathcal{P}(S, I, R, O-1, R_O, t) - \gamma I\mathcal{P}(\mathbf{x},t) \\ &+\alpha_O(O+1)\mathcal{P}(S, I, R, O+1, R_O-1, t) - \alpha_O O\mathcal{P}(\mathbf{x},t). \end{aligned} \tag{7}$$

We consider two observables of the model, denoted by y_1, y_2; on the one hand, the subset of the total infected people being reported by the health system is $y_1 = O$, and on the other hand the viral number of gene copies detected in wastewater is:

$$y_2 = C_w = \frac{I \cdot C_f \cdot \xi}{1000 \cdot Q \cdot \rho} \tag{8}$$

where C_w, C_f, Q and ρ are defined in Table 1. This formula is considered to properly normalize the wastewater data. In weeks where the number of total infected detected by viral load in WWTP at first day (I) is lower than the infected number reported by the health system at first day (O), we re-normalize as follows:

$$y_2 = y_1(day\ 1)\frac{y_2}{y_2(day\ 1)}. \tag{9}$$

3. Results and discussion

In this section, we first describe the scripts implemented for the proposed SIR and SIRO models. Then, we illustrate how the simplest SIR model has the capacity to represent the data provided by the public health system. After that, we show the capacity of the SIRO model to predict in a week time horizon the evolution of the number of infected people, starting from the WWTP viral load levels detected during the previous week.

3.1. Data and scripts

The models proposed in the previous section have been implemented in MATLAB and are available at https://github.com/manuelpajaro/stochasticSIR_O inside the folder `stochasticSIR_O`. The main programs are `SIR_ppal.m` for the simplest SIR model, `SIRv_ppal.m` and `SIRj_ppal.m` for the SIRv and SIRj model extensions respectively, and `SIRO_ppal.m` for the SIRO model. The data used for this study are saved in the `DATA.mat` archive which is available within the folder `sotchasticSIR_O`. The data of new infected persons per day and municipality were provided by the Galician Health System (Servizo Galego de Saúde SERGAS). The data from SARS-CoV-2 viral load in the WWTPs under study were obtained within the DIMCoVAR project consortium [36]. Specifically, we provide the number of infected persons detected by the health system per day (`I_Locality` variables) and the corresponding cumulative infected cases for fourteen days (`Icum14_Locality` variables). The variable I in (8) computed from the measurements of viral load in sewage is stored in `WTP_Locality`.

The main programs (`SIR_ppal.m`, `SIRv_ppal.m`, `SIRj_ppal.m`, and `SIRO_ppal.m`) share the same structure. The user can choose the municipality by modifying the value of the Li variable in the following code:

```
%% Locality selection one from [Ares,Baiona,Gondomar,
    Melide,Nigran]
localities = {'Ares','Baiona','Gondomar','Melide','Nigran'
    };
Li = 1; % 1 -> Ares; 2 -> Baiona; 3 -> Gondomar; 4 ->
    Melide; 5 -> Nigran
locality = localities{Li}; % to select one of the previous
     localities
```

where currently $Li = 1, \ldots, 5$ to select Ares, Baiona, Gondomar, Melide or Nigrán, respectively (of course the list can be extended to other municipalities of interest if access to data is provided). The starting date of the simulation can be chosen by assigning to variable f the date in the format year month day, [*yyyy mm dd*], as follows:

```
%% Select a date with format year month day f= [yyyy mm dd
    ] from 01/03/2020 to 27/06/2021
f= [2021 03 21];
```

Table 1
Notation and units.

Notation	Units	Description
C_w	copies/L	Copies per Liter of wastewater [37,38]
Q	m^3/day	Inlet flow by GESECO Aguas https://www.gesecoaguas.es/
ρ	g faeces/mL faeces	Faeces density [39]
C_f	copies/mL faeces	Average number of copies per mL of faeces in infected person [37,38]
ξ	g faeces/person/day	Average weight of faeces per person per day [39]

Finally, the free parameters for each model are indicated next. In the SIR model (`SIR_ppal.m`) the parameters α (recovery rate constant) and β (infection rate constant) are free *a priori*. To avoid autocorrelation problems, we fix the parameter $\alpha = 1/14$ (cumulated incidence is calculated for a time interval of 14 days). Importantly, the mitigation policies do not affect the value of the recovery rate constants of individuals (which is coherent with fixing the parameter α), but the infection rate constant β (which should be therefore calibrated from data).

```
%% Model Parameters
beta  = 0.14; % beta in {0.03,0.07,0.14,0.22}
alpha = 1/14; % it can be 1/7 or 1/10 for strong decrease
    in infected persons (beta = 0.03)
```

For the SIRv model (`SIRv_ppal.m`), the parameters β_0, β_1 and T_{change} can be calibrated. T_{change} is the point in time where a change in the sign of the slope of the dynamics occurs (i.e. a turning point as defined in the previous section).

```
%% Model Parameters (Pajaro et al 2022)
% Choose the ones given for each week (f) in Table 3
beta0   = 0.14;
beta1   = 0.03;
Tchange = 3;
alpha   = 1/14; % fixed
```

For the SIRj model (`SIRj_ppal.m`) the parameter β can be used for calibration purposes.

```
%% Model Parameters (Pajaro et al 2022)
% Choose the ones given for each week (f) in Table 4
beta  = 0.02; % beta in {0.02,0.03}
alpha = 1/14; % fixed
```

In the SIRO model (`SIRO_ppal.m`), the parameters β and γ can be used for calibration. Generally, the recovery rate constant should be equivalent, i.e. $\alpha_I = \alpha_O$.

```
%% Model Parameters (Pajaro et al 2022)
% Choose the ones given for each week (f) in Table 5
beta   = 0.1;
gamma  = 0.06;
alphaI = 1/14;
alphaO = 1/14;
```

3.2. SIR predictions

The SIR model is calibrated from the data of infected persons provided by the public health system. The parameter β for the SIR model described in Section 2.1 is estimated from the data of infected persons provided by the public health system in order to fit the trajectory of fourteen days cumulative infected cases. Calibrations are done per week. As justified in the previous section, the recovering rate constant is fixed at $\alpha = \frac{1}{14}$.

Remarkably, the evolution of the number of infected persons detected by the health system could be accurately predicted for most weeks (more than 70% of the total of 265) during the frame of the study with the simplest stochastic SIR model with a fixed $\alpha = \frac{1}{14}$ and only four different values of the infection rate constant β. Specifically, $\beta \in \beta_{set} = \{0.03, 0.07, 0.14, 0.22\}$, see Fig. 1 and Table 2. This has important implications for assessment and forecasting. First, the model can be used for quantitative assessment of COVID-19 mitigation policies, i.e. to quantify the impact of restrictions on the infection rate constant. Second, this facilitates the use of the model for forecasting purposes, to the point of needing only the one parameter (which can be fixed to the value obtained for the previous week) and the initial condition (number of infected reported by the health system) at the starting date, in order to obtain predictions at one week time horizon.

Table 2
Total number and percentage of weeks for which the specified model (left column) provided an accurate prediction.

Model	Ares	Baiona	Gondomar	Melide	Nigrán	Total	%
SIR ($\alpha = \frac{1}{14}$)	40	39	36	33	39	187	70.6
SIR ($\alpha < \frac{1}{14}$)	11	11	9	13	10	54	20.4
SIRv	1	1	2	2	5	11	4.1
SIRj	1	1	4	4	3	13	4.9
Total	53	52	51	52	57	265	100

As it is indicated Table 2, the SIR model accurately predicts the evolution in time of the infected cases detected by the health system (90% of the weeks by the simplest stochastic SIR with fixed $\beta \in \beta_{set}$ and fixed α, 4.1% of the weeks showed a variable β, and 4.9% of the weeks showed jumps).

In Fig. 2 we show two exceptional cases in which the best fits were obtained with $\alpha \neq 14$ (this might happen for example for those weeks in which many infected persons are recovered in the first days) as it happened in Melide for the week of 18 April 2021.

3.2.1. SIRv predictions

Those exceptional weeks for the overall time of the study showing turning points are specified in Table 3. Two selected examples are depicted in Fig. 3 where the predictions obtained by SIR and SIRv are compared. We show the 36th week of Ares and the 43rd week of Nigrán. As it can be seen in Fig. 3, the SIR model with two infection rates, β_0 and β_1, generates much more precise predictions of the cumulative level of infection. Note that SIRv can be used for calibration, and to analyze the conditions under which the turning point is produced, but, without additional information on the time of the turning, it cannot be used for forecasting (in the conclusion we propose as future work the implementation of real time calibration and machine learning methods to overcome this limitation).

3.2.2. SIRj predictions

The SIR model fails to predict superspreading events. Using the SIRj model developed in the Methods section, we not only reproduce the dynamics of cumulative cases, but also obtain better realizations than those obtained using the simplest SIR model. In order to measure the accuracy of the realizations for new infected cases we define the following metrics:

$$ERROR = \sum_{i=1}^{D} |Inew_i - Data_i| \qquad \text{with } D = 7, \tag{10}$$

where $Data$ is a vector of real data provided by the health system whereas $Inew$ is one realization obtained from SSA simulation. We define E_i as the number of realizations with $ERROR = i$, where E_0

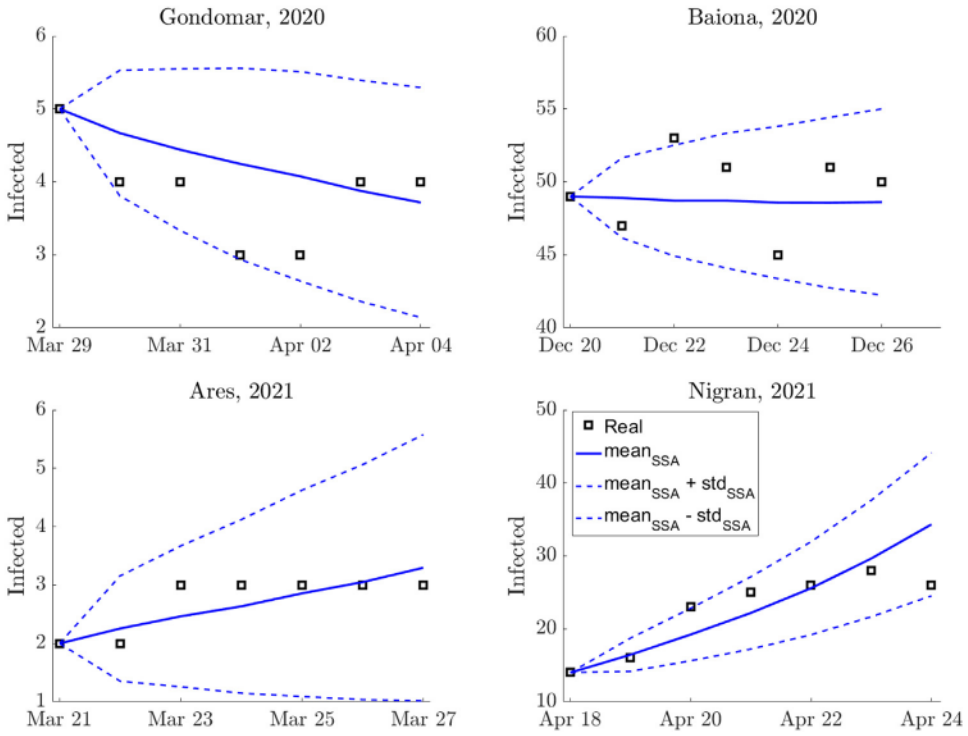

Fig. 1. From 10^3 SSA realizations of the SIR model we compute the infected mean (blue line) and standard deviation (dotted blue lines). Black squares represent the accumulated infected cases reported by the health system. $\alpha = 1/14$ for all cases. The infection rate constants are: $\beta = 0.03$ for Gondomar, $\beta = 0.07$ for Baiona, $\beta = 0.14$ for Ares and $\beta = 0.22$ for Nigrán.

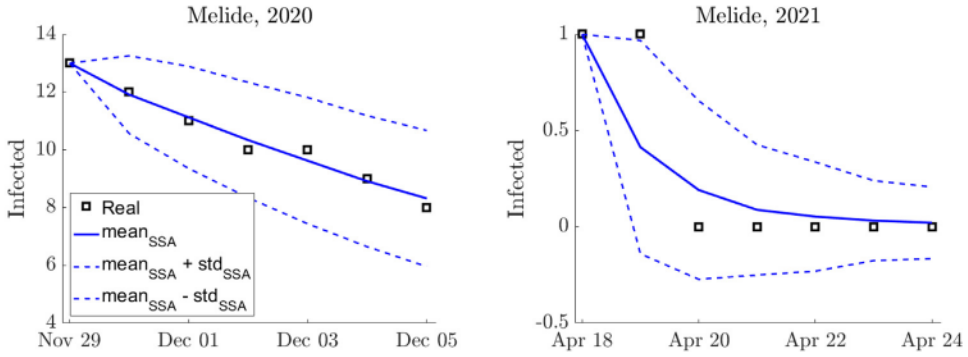

Fig. 2. From 10^3 SSA realizations of the SIR model we compute the infected mean (blue line) and standard deviation (dotted blue lines). Black squares represent the accumulated infected cases reported by health system. The infection rate is $\beta = 0.03$ for both cases. The recovered parameters are $\alpha = 1/10$ and $\alpha = 1/2$ for the 40th and 60th weeks of Melide, respectively.

Table 3
Parameters for the SIR model with 2 different beta, β_0, β_1.

Week	Locality	Date	β_0	β_1	α	T_{change}
36	Ares	01/11/20	0.14	0.03	1/14	3
36	Baiona	01/11/20	0.14	0.03	1/14	4
35	Gondomar	25/10/20	0.14	0.03	1/14	3
46	Gondomar	10/01/21	0.22	0.03	1/14	5
46	Melide	10/01/21	0.22	0.14	1/14	3
48	Melide	24/01/21	0.03	0.07	1/14	3
31	Nigrán	27/09/20	0.14	0.03	1/14	2
35	Nigrán	25/10/20	0.14	0.03	1/14	3
36	Nigrán	01/11/20	0.14	0.03	1/14	3
43	Nigrán	20/12/20	0.14	0.03	1/14	4
60	Nigrán	18/04/21	0.22	0.03	1/14	5

is the number of exact realizations (see for example the last plot in Fig. 4) and E_1 is the number of realizations which have exact number of infected for all days except one for which there is only a difference of one (see for example the first plot of the second row in Fig. 4).

In Fig. 4 we compare the results obtained after 10^4 SSA realizations for the SIR and the SIRj models. We selected the 35th week of Melide which stars at 25 November 2020 for which the SIR model does not capture the infection spread using the four values of β proposed as reference. So, we use $\beta = 0.35$ and $\alpha = 1/14$ for the simplest SIR, and for the model with jumps the parameters are given in Table 4, $\beta = 0.02$ and $\alpha = 1/14$. For the parameters chosen, both models reproduce the real cumulative cases accurately, as it can be seen in the first row of Fig. 4. However, when we observe the number of new infected persons per day, the SIRj model generates the best realizations with three exact realizations, $E_0 = 3$ (last plot in Fig. 4) and several realizations with $ERROR = 1$ from the 10^4 computed, $E_1 = 85$. For the SIR model there

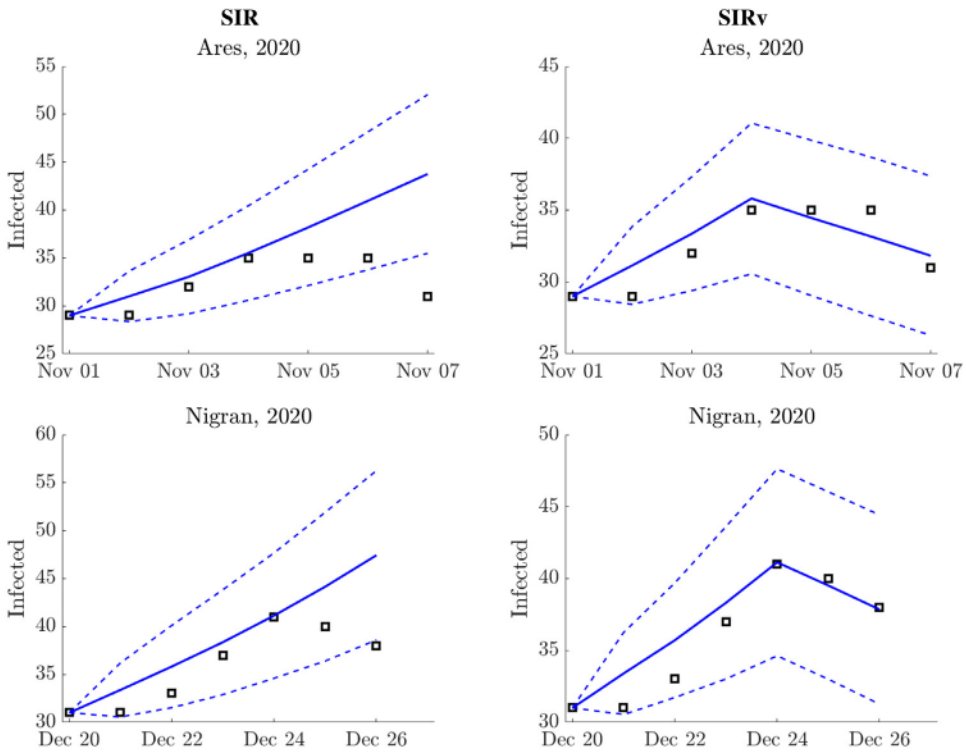

Fig. 3. From 10^3 SSA realizations of the SIR model we compute the infected mean (blue line) and standard deviation (dotted blue lines). Black squares represent the accumulated infected people reported by the health system. The results of SIR and SIRv models are shown in the first and second columns, respectively. $\alpha = 1/14$ for all cases. The infection parameters for SIR is $\beta = 0.14$, for both localities $\beta_0 = 0.14$ and $\beta_1 = 0.03$ for the SIRv model. T_{change} is respectively 3 and 4 for the weeks under study.

Table 4
Parameters for the SIR with jumps model.

Week	Locality	Date	β	α
34	Ares	18/10/20	0.03	1/14
55	Baiona	14/03/21	0.02	1/14
23	Gondomar	07/08/20	0.02	1/14
33	Gondomar	11/10/20	0.02	1/14
39	Gondomar	22/11/20	0.02	1/14
55	Gondomar	14/03/21	0.02	1/14
3	Melide	20/03/20	0.03	1/14
25	Melide	16/08/20	0.02	1/14
35	Melide	25/10/20	0.02	1/14
45	Melide	03/01/21	0.02	1/14
2	Nigrán	12/03/20	0.02	1/14
23	Nigrán	02/08/20	0.03	1/14
55	Nigrán	14/03/21	0.02	1/14

are not exact realizations ($E_0 = 0$) and only 7 of the 10^4 realizations are obtained with $ERROR = 1$ ($E_1 = 7$), one of them is shown in the first plot in the second row in Fig. 4. Again, SIR models cannot predict superspreader events in absence of *a priori* information, but we can use SIRj model to detect and quantify *a posteriori* the occurrence of a superspreader event.

3.2.3. Probability density distributions

The realizations of the SSA algorithm are used to compute the mean number of infected cases and the standard deviation to assess the accuracy of the predictions of the stochastic model. Besides this, for all SIR models we also obtain the probability density distribution of (new and accumulated) infected cases from the 10^3 SSA realizations which is an approximation of the solution of the associated CME (2). In Fig. 5 we depict the probabilities of a given number of cumulative cases for a specific week (we chose the 36th week of Ares, already discussed, see first row of Eq. (3)). As it can be observed in Fig. 5 the cases reported by the health system are close to the mode of the obtained probability distributions, which means that the proposed model captures well the evolution of the epidemics. Moreover, the SSA algorithm can be used to estimate the new infected cases per day. The probability distribution of these new cases is shown in Fig. 6 where the data provided by the health system (vertical black dotted lines) fall within the most probable cases predicted.

3.3. SIRO predictions

The SIRO model can be used to predict, via SSA simulations, the cumulative cases reported by the public health system (observed cases O) and those reported by wastewater treatment plant samples (we assume they are a proxy of the total infected cases I). The viral load in the WWTP samples was quantified using by RT-qPCR [36]. We consider periods of ten days as depicted in Table 5 together with the parameters obtained for the SIRO model. The mathematical model shows a good predictive capacity allowing us to forecast the evolution of infected persons, both total (I) and observed (O) by the health system in the municipalities within a horizon of 10 days. In Fig. 7 we show, as a representative illustration of the model outcome, the predictions of the SIRO model using 10^3 SSA realizations for the five municipalities: Melide, Nigrán, Baiona, Gondomar and Ares. The model predictions, mean and standard deviations (continuous and dashed lines, respectively) for the total number of infected (blue lines) and the observed number of infected (black lines) are depicted in Fig. 7 together with the real data obtained from viral load in sewage (blue circles) and health system (black squares).

4. Conclusions

In this work we present a stochastic model based in the classical compartmental models (SIR type) for which we have incorporated the stochastic character of viral spread. We consider the transitions

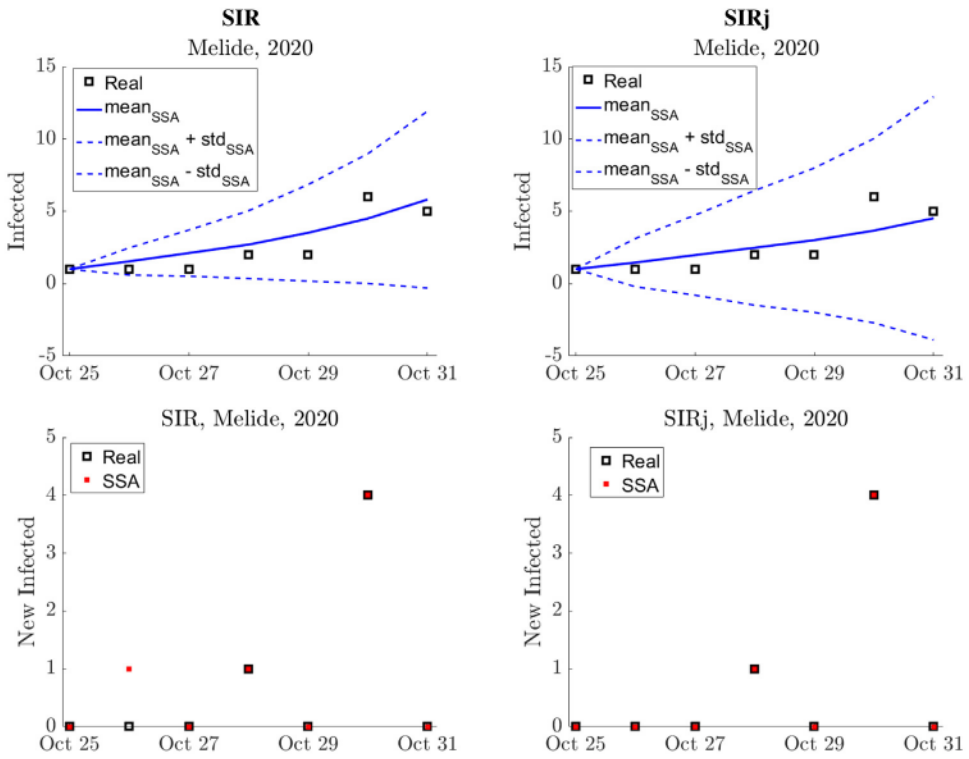

Fig. 4. From 10^4 SSA realizations we compute the infected mean (blue line) and standard deviation (dotted blue lines). Black squares represent the accumulated infected (first row) or the new infected cases (second row) reported by the health system. Red squares are the new infected cases obtained from the best SSA realization. The results obtained with SIR and SIRj models are shown in the first and second columns, respectively. The parameters are, $\beta = 0.35$ and $\alpha = 1/14$ for SIR and $\beta = 0.02$ and $\alpha = 1/14$ for SIRj.

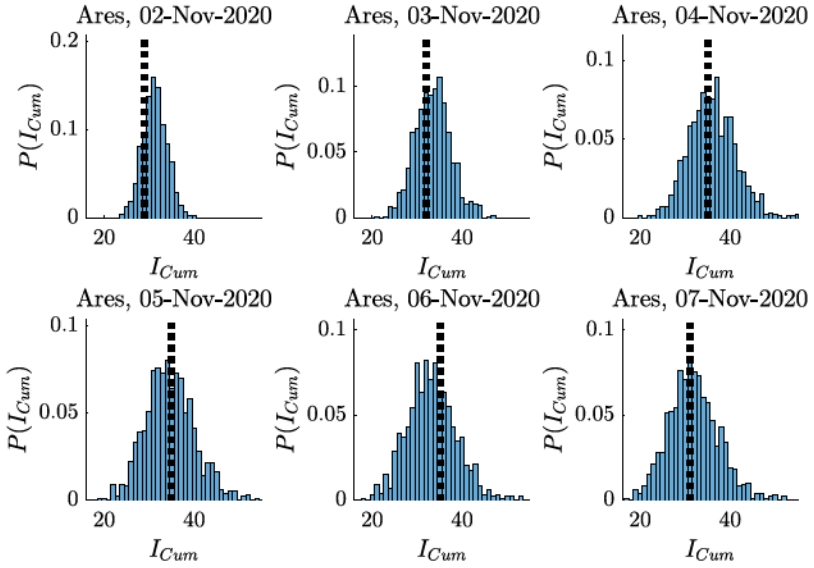

Fig. 5. Probability density distribution for accumulated cases obtained from 10^3 SSA realizations of the SIR with variable infection rates for Ares starting the 1 November 2020. The vertical black dotted line represent the real cases reported by the health system. The parameters considered are, $\beta_0 = 0.14$, $\beta_1 = 0.03$, $\alpha = 1/14$ and $T_{change} = 3$.

between each group of persons in which the total population is subdivided (for example, Susceptible, Infected and Recovered) as reactions of chemical species. The Chemical Master Equation (CME) is the model that incorporates the inherent stochasticity of chemical reaction systems and therefore, by using this new formulation and proposing the corresponding CME, we are able to incorporate the noise of the viral infection propagation in a natural form. Moreover, in this article we solve the different models proposed using the Stochastic Simulation Algorithm (SSA) of Gillespie, which allow us to obtain the solution of the CME (the time evolution of the probability density function of infected persons) together with realizations (possible trajectories of the time evolution of the number of infected persons). Whereas deterministic SIR models only capture a unique trajectory for a set of parameters, their stochastic versions produce a high number of realizations providing us automatically with a measure of the noise in the epidemics spread. We can observe how the uncertainty grows with time as the standard deviations reported or the tails of the distributions obtained are higher as we move forward in time.

The stochastic version of the classical SIR model presented in this work has been developed with the aim to analyze and predict the

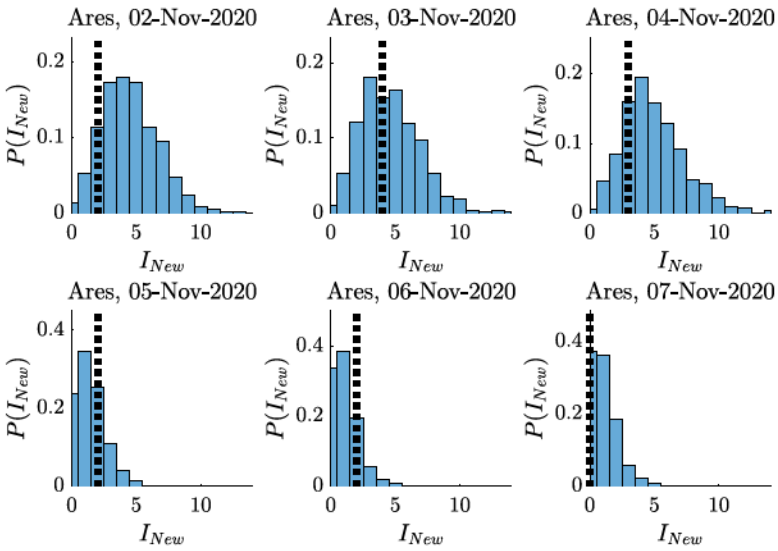

Fig. 6. Probability density distribution for new cases obtained from 10^3 SSA realizations of the SIR with variable infection rates for Ares starting the 1 November 2020. The vertical black dotted line represent the real cases reported by the health system. The parameters considered are, $\beta_0 = 0.14$, $\beta_1 = 0.03$, $\alpha = 1/14$ and $T_{change} = 3$.

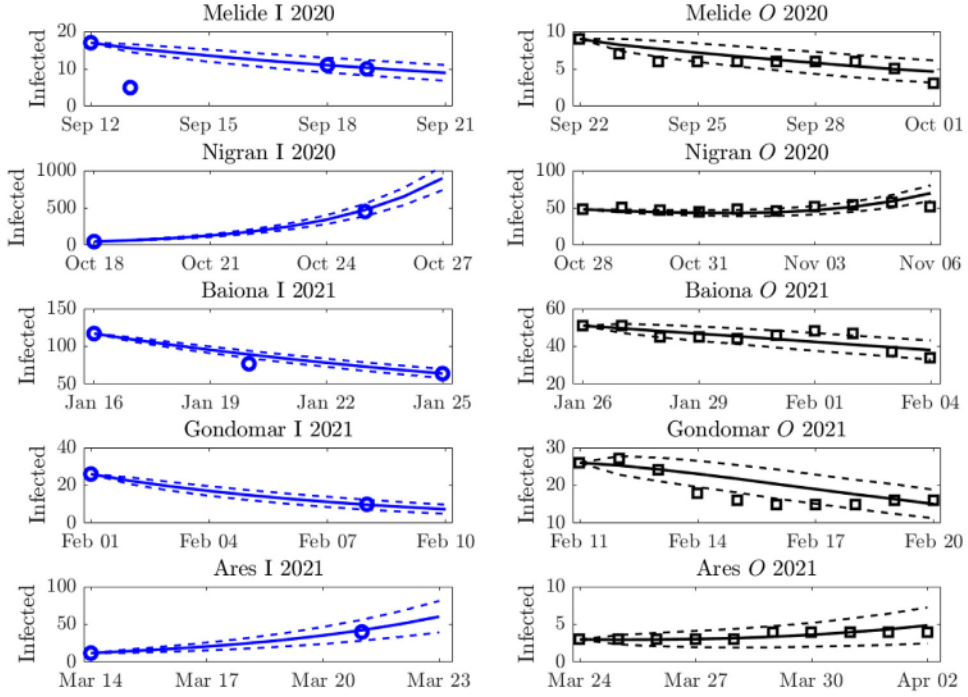

Fig. 7. Predictions of the SIRO model using 10^3 SSA realizations for each locality, Melide (first row), Nigrán (second row), Baiona (third row), Gondomar (fourth row) and Ares (fifth row). The estimated infected from viral load in sewage (blue points) and mean number of infected individuals, I, obtained using the SIRO model (blue lines) are shown at the left column. The number of infected persons reported by the health system (black points) and the SIRO model prediction, O, (black lines) are shown at the right column. The standard deviations obtained for I and O from the SIRO model simulations are represented with blue and black dashed lines, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

evolution of the COVID-19 epidemics at the level of municipalities in Galicia (Northwest of Spain) integrating data from wastewater and the public health system in Galicia. The model has been calibrated and successfully applied for SARS-CoV-2 surveillance in different municipalities [36], using data from the galician public health system (SERGAS), and RT-qPCR measurements from wastewater samples. The SIRO model presented is fed by the total number of infected persons estimated using the viral load in sewage to predict the number of positives observed by the health system during the following week.

The software with the model implementation is provided, such that it can be calibrated and used to make analysis and predictions in other municipalities (with WWTP data available), and for different purposes. The model is easy to calibrate (only 2 parameters for calibration from time course data) and, as it has been shown in Section 3.2, it is very robust, as it allows estimating the number of infected persons in all the municipalities studied during more than one year by using only four different infection rate parameters. Remarkably, we have obtained a set of 4 parameters for the stochastic SIR model which are enough to describe the viral infection in the hole period of time for the five municipalities considered. These parameters can be explained by the restrictions imposed (lockdowns) during the pandemic. Therefore, we provide evidences of the usefulness of this kind of stochastic models,

Table 5
Parameters obtained for the SIRO model.

	Locality	Date	β	γ	α_I	α_O
1	Ares	29/08/20	0.005	0.5	1/5	1/14
2	Ares	10/09/20	0.005	0.05	1/7	1/14
3	Ares	17/10/20	0.35	0.035	1/14	1/14
4	Ares	24/10/20	0.005	0.08	1/5	1/5
5	Ares	31/10/20	0.03	0.005	1/14	1/14
6	Ares	10/01/21	0.005	0.5	1/5	1/14
7	Ares	16/01/21	0.2	0.08	1/14	1/14
8	Ares	24/01/21	0.02	0.12	1/10	1/10
9	Ares	31/01/21	0.005	0.17	1/5	1/5
10	Ares	07/02/21	0.02	0.005	1/7	1/7
11	Ares	28/02/21	0.005	0.0005	1/5	1/5
12	Ares	07/03/21	0.005	0.025	1/7	1/7
13	Ares	14/03/21	0.25	0.015	1/14	1/14
14	Ares	21/03/21	0.02	0.0005	1/7	1/7
1	Baiona	16/08/20	0.3	0.02	1/14	1/14
2	Baiona	11/10/20	0.3	0.1	1/14	1/14
3	Baiona	18/10/20	0.12	0.1	1/14	1/14
4	Baiona	25/10/20	0.12	0.01	1/14	1/14
5	Baiona	02/11/20	0.005	0.025	1/7	1/7
6	Baiona	08/11/20	0.08	0.1	1/14	1/14
7	Baiona	17/11/20	0.05	0.17	1/14	1/14
8	Baiona	25/11/20	0.07	0.12	1/14	1/14
9	Baiona	01/12/20	0.03	0.17	1/14	1/14
10	Baiona	05/12/20	0.1	0.07	1/14	1/14
11	Baiona	14/12/20	0.05	0.05	1/14	1/14
12	Baiona	21/12/20	0.22	0.025	1/14	1/14
13	Baiona	29/12/20	0.12	0.05	1/14	1/14
14	Baiona	04/01/21	0.05	0.18	1/5	1/14
15	Baiona	16/01/21	0.005	0.02	1/14	1/14
16	Baiona	25/01/21	0.005	0.02	1/14	1/14
17	Baiona	30/01/21	0.005	0.25	1/5	1/14
18	Baiona	06/02/21	0.1	0.05	1/14	1/14
19	Baiona	20/02/21	0.1	0.005	1/14	1/14
20	Baiona	01/03/21	0.005	0.12	1/10	1/10
21	Baiona	07/03/21	0.005	3.2	1/7	1/14
22	Baiona	13/03/21	0.25	0.07	1/14	1/14
23	Baiona	20/03/21	0.08	0.005	1/10	1/10
1	Gondomar	11/01/21	0.3	0.02	1/14	1/14
2	Gondomar	18/01/21	0.005	0.03	1/7	1/7
3	Gondomar	25/01/21	0.3	0.0005	1/14	1/14
4	Gondomar	01/02/21	0.005	0.12	1/7	1/7
1	Melide	23/08/20	0.005	0.07	1/5	1/5
2	Melide	30/08/20	0.02	0.01	1/14	1/14
3	Melide	07/09/20	0.05	0.01	1/14	1/14
4	Melide	12/09/20	0.003	0.001	1/14	1/14
5	Melide	17/10/20	0.4	0.05	1/14	1/14
6	Melide	24/10/20	0.02	0.03	1/10	1/10
7	Melide	31/10/20	0.02	0.03	1/14	1/14
8	Melide	07/11/20	0.4	0.01	1/14	1/14
9	Melide	21/12/20	0.1	0.025	1/14	1/14
10	Melide	27/12/20	0.05	0.05	1/10	1/14
11	Melide	03/01/21	0.05	0.1	1/7	1/14
12	Melide	10/01/21	0.25	0.025	1/14	1/14
13	Melide	17/01/21	0.02	0.06	1/7	1/7
14	Melide	25/01/21	0.005	0.07	1/5	1/7
15	Melide	31/01/21	0.005	0.07	1/5	1/5
1	Nigrán	15/08/20	0.06	0.25	1/14	1/14
2	Nigrán	19/09/20	0.16	0.005	1/7	1/7
3	Nigrán	12/10/20	0.17	0.1	1/14	1/14
4	Nigrán	18/10/20	0.4	0.02	1/14	1/14
5	Nigrán	25/10/20	0.0005	0.25	1/5	1/14
6	Nigrán	01/11/20	0.2	0.03	1/14	1/14
7	Nigrán	21/11/20	0.1	0.03	1/14	1/14
8	Nigrán	28/11/20	0.01	0.22	1/14	1/14
9	Nigrán	05/12/20	0.05	0.13	1/14	1/14
10	Nigrán	14/12/20	0.35	0.005	1/14	1/14
11	Nigrán	21/12/20	0.13	0.04	1/14	1/14
12	Nigrán	09/01/21	0.12	0.07	1/14	1/14
13	Nigrán	18/01/21	0.005	0.15	1/7	1/7
14	Nigrán	25/01/21	0.17	0.005	1/14	1/14
15	Nigrán	01/02/21	0.02	0.02	1/14	1/14
16	Nigrán	08/02/21	0.13	0.005	1/14	1/14
17	Nigrán	15/02/21	0.06	0.005	1/7	1/7
18	Nigrán	22/02/21	0.1	0.01	1/14	1/14
19	Nigrán	01/03/21	0.1	0.06	1/14	1/14
20	Nigrán	22/03/21	0.17	0.06	1/14	1/14

(*continued on next page*)

based in CME, to predict the viral spread of infectious diseases, in particular for the case of COVID-19 spread.

Due to the robustness of the model, its easy calibration, its flexibility to be adapted to the surveillance of other pathogens, we believe it makes a significant contribution to wastewater based epidemiology. As a future work, we propose to incorporate artificial intelligence techniques for automated real time calibration of the parameters, which can significantly facilitate the forecasting of turning points.

Software availability

The scripts for the models used are available under GPLv3 license at https://github.com/manuelpajaro/stochasticSIR_O.

CRediT authorship contribution statement

Manuel Pájaro: Conceptualization, Methodology, Software, Data curation, Writing – original draft, Writing – review & editing, Supervision. **Noelia M. Fajar:** Data curation, Validation, Writing – original draft. **Antonio A. Alonso:** Conceptualization, Methodology, Funding acquisition. **Irene Otero-Muras:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Supervision.

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.

Acknowledgments

The authors gratefully acknowledge the support of this work through the project "DIMCoVAR" funded in the program "Fondo Supera COVID" of the CRUE (Spanish Universities)-Santander Foundation. The authors also thank Aguas de Galicia and Consellería de Sanidade - Xunta de Galicia for their support and funding of this study. MP acknowledges support from grant FJC2019-041397-I funded by MCIN/AEI/10.13039/501100011033. Funding for open access charge: Universidade da Coruña/CISUG

References

[1] Holmdahl I, Buckee C. Wrong but useful — What Covid-19 epidemiologic models can and cannot tell us. N Engl J Med 2020;383(4):303–5.
[2] Hoertel N, Blachier M, Blanco C, Olfson M, Massetti M, Sánchez Rico M, Limosin F, Leleu H. A stochastic agent-based model of the SARS-CoV-2 epidemic in France. Nat Med 2020;26(9):1417–21.
[3] Reguly IZ, Csercsik D, Juhász J, Tornai K, Bujtár Z, Keömley-Horváth B, Kós T, Cserey G, Iván K, Pongor S, Szederkényi G, Röst G, Csikász-Nagy A. Microsimulation based quantitative analysis of COVID-19 management strategies. PLoS Comput Biol 2022;18(1):e1009693.
[4] Snip LJP, Flores-Alsina X, Plósz BG, Jeppsson U, Gernaey KV. Modelling the occurrence, transport and fate of pharmaceuticals in wastewater systems. Environ Modell Softw 2014;62:112–27.
[5] Kermack WO, McKendrick AG. A contribution to the mathematical theory of epidemics. Proc R Soc Lond Ser A Math Phys Eng Sci 1927;115(772):700–21.
[6] Calleri F, Nastasi G, Romano V. Continuous-time stochastic processes for the spread of COVID-19 disease simulated via a Monte Carlo approach and comparison with deterministic models. J Math Biol 2021;83(4):34.

[7] Campos EL, Cysne RP, Madureira AL, Mendes GLQ. Multi-generational SIR modeling: Determination of parameters, epidemiological forecasting and age-dependent vaccination policies. Infect Dis Model 2021;6:751–65.
[8] El Kharrazi Z, Saoud S. Simulation of COVID-19 epidemic spread using stochastic differential equations with jump diffusion for SIR model. In: 2021 7th international conference on optimization and applications (ICOA). 2021, p. 1–4.
[9] Roda WC, Varughese MB, Han D, Li MY. Why is it difficult to accurately predict the COVID-19 epidemic? Infec Dis Model 2020;5:271–81.
[10] Cumsille P, Rojas-Díaz Ó, de Espanés PM, Verdugo-Hernández P. Forecasting COVID-19 Chile' second outbreak by a generalized SIR model with constant time delays and a fitted positivity rate. Math Comput Simul 2022;193:1–18.
[11] Moein S, Nickaeen N, Roointan A, Borhani N, Heidary Z, Javanmard SH, Ghaisari J, Gheisari Y. Inefficiency of SIR models in forecasting COVID-19 epidemic: a case study of isfahan. Sci Rep 2021;11(1):4725.
[12] Castro M, Ares S, Cuest JA, Manrubia S. The turning point and end of an expanding epidemic cannot be precisely forecast. Proc Natl Acad Sci USA 2020;117(42):26190–6.
[13] Peni T, Csutak B, Szedérkenyi G, Röst G. Nonlinear model predictive control with logic constraints for COVID-19 management. Nonlinear Dyn 2020;102(4):1965–86.
[14] Jones A, Strigul N. Is spread of COVID-19 a chaotic epidemic? Chaos Solitons Fractals 2021;142:110376.
[15] Mangiarotti S, Peyre M, Zhang Y, Huc M, Roger F, Kerr Y. Chaos theory applied to the outbreak of Covid-19: An ancillary approach to decision-making in pandemic context. Epidemiol Infect 2020;148:E95.
[16] Elnawawy M, Aloul F, Sagahyroon A, Elwakil AS, Sayed Wafaa S, Said Lobna A, Mohamed SM, Radwan Ahmed G. FPGA realizations of chaotic epidemic and disease models including Covid-19. IEEE Access 2021;9(9337855):21085–93.
[17] Cai Y, Kang Y, Wang W. A stochastic SIRS epidemic model with nonlinear incidence rate. Appl Math Comput 2017;305:221–40.
[18] Großmann G, Backenköhler M, Wolf V. Importance of interaction structure and stochasticity for epidemic spreading: A COVID-19 case study. In: Gribaudo Marco, Jansen David N, Remke Anne, editors. Quantitative evaluation of systems. Cham: Springer International Publishing; 2020, p. 211–29.
[19] Ball F, Sirl D, Trapman P. Analysis of a stochastic SIR epidemic on a random network incorporating household structure. Math Biosci 2010;224(2):53–73.
[20] Mahapatra DP, Triambak S. Towards predicting COVID-19 infection waves: A random-walk Monte Carlo simulation approach. Chaos Solitons Fractals 2022;156:111785.
[21] Nakamura GM, Cardoso GC, Martinez AS. Improved susceptible-infectious-susceptible epidemic equations based on uncertainties and autocorrelation functions. R Soc Open Sci 2020;7(2):191504.
[22] Pájaro M, Alonso AA, Otero-Muras I, Vázquez C. Stochastic modeling and numerical simulation of gene regulatory networks with protein bursting. J Theoret Biol 2017;421:51–70.
[23] Pájaro M, Otero-Muras I, Vázquez C, Alonso AA. Transient hysteresis and inherent stochasticity in gene regulatory networks. Nature Commun 2019;10(1):4581.
[24] Gillespie DT. A general method for numerically simulating the stochastic time evolution of couple chemical reactions. J Comput Phys 1976;22:403–34.
[25] Amadei A, Aschi M. A general model for Covid-19 epidemic kinetics: application to italian and german data. Theor Biol Forum 2020;113(1–2):19–30.
[26] Sánchez-Taltavull D, Castelo-Szekely V, Candinas D, Roldán E, Beldi G. Modelling strategies to organize healthcare workforce during pandemics: Application to COVID-19. J Theoret Biol 2021;523:110718.
[27] Winkelmann S, Zonker J, Schütte C, Conrad ND. Mathematical modeling of spatio-temporal population dynamics and application to epidemic spreading. Math Biosci 2021;336:108619.
[28] McLure A, O'Neill B, Mayfield H, Lau C, McPherson B. PoolTestR: An R package for estimating prevalence and regression modelling for molecular xenomonitoring and other applications with pooled samples. Environ Modell Softw 2021;145:105158.
[29] Sunjaya AF, Sunjaya AP. Pooled testing for expanding COVID-19 mass surveillance. Dis Med Public Health Prep 2020;14(3):e42–3.
[30] Foladori P, Cutrupi F, Segata N, Manara S, Pinto F, Malpei F, Bruni L, La Rosa G. SARS-CoV-2 from faeces to wastewater treatment: What do we know? A review. Sci Total Environ 2020;743:140444.
[31] Kitajima M, Ahmed W, Bibby K, Carducci A, Gerba CP, Hamilton KA, Haramoto E, Rose JB. SARS-CoV-2 in wastewater: State of the knowledge and research needs. Sci Total Environ 2020;739:139076.
[32] Peccia J, Zulli A, Brackney DE, Grubaugh ND, Kaplan EH, Casanovas-Massana A, Ko AI, Malik AA, Wang D, Wang M, Warren JL, Weinberger DM, Arnold W, Omer SB. Measurement of SARS-CoV-2 RNA in wastewater tracks community infection dynamics. Nat Biotechnol 2020;38(10):1164–7.
[33] Medema G, Heijnen L, Elsinga G, Italiaander R, Brouwer A. Presence of SARS-coronavirus-2 RNA in sewage and correlation with reported COVID-19 prevalence in the early stage of the epidemic in the netherlands. Environ Sci Technol Lett 2020;7(7):511–6.
[34] Randazzo W, Truchado P, Cuevas-Ferrando E, Simón P, Allende A, Sánchez G. SARS-CoV-2 RNA in wastewater anticipated COVID-19 occurrence in a low prevalence area. Water Res 2020;181:115942.
[35] Vallejo JA, Trigo-Tasende N, Rumbo-Feal S, Conde-Pérez K, López-Oriona Á, Barbeito I, Vaamonde M, Tarrío-Saavedra J, Reif R, Ladra S, Rodiño-Janeiro BK, Nasser-Ali M, Cid Á, Veiga M, Acevedo A, Lamora C, Bou G, Cao R, Poza M. Modeling the number of people infected with SARS-COV-2 from wastewater viral load in northwest Spain. Sci Total Environ 2022;811:152334.
[36] Novoa B, Rios-Castro R, Otero-Muras I, Gouveia S, Cabo A, Saco A, Rey-Campos M, Pájaro M, Fajar N, Arenguren R, Romero A, Panebianco A, Valdés L, Payo P, Alonso AA, Figueras A, Cameselle C. Wastewater and marine bioindicators surveillance to anticipate COVID-19 prevalence and to explore SARS-CoV-2 diversity by next generation sequencing: one-year study. Sci Total Environ 2022;833:155140.
[37] Zheng S, Fan J, Yu F, Feng B, Lou B, Zou Q, Xie G, Lin S, Wang R, Yang X, Chen W, Wang Q, Zhang D, Liu Y, Gong R, Ma Z, Lu S, Xiao Y, Gu Y, Zhang J, Yao H, Xu K, Lu X, Wei G, Zhou J, Fang Q, Cai H, Qiu Y, Sheng J, Chen Y, Liang T. Viral load dynamics and disease severity in patients infected with SARS-CoV-2 in Zhejiang province, China, January-March 2020: Retrospective cohort study. BMJ 2020;369:m1443.
[38] Giacobbo A, Rodrigues MAS, Zoppas Ferreira J, Bernardes AM, de Pinho MN. A critical review on SARS-CoV-2 infectivity in water and wastewater. what do we know? Sci Total Environ 2021;774:145721.
[39] Penn R, Ward BJ, Strande L, Maurer M. Review of synthetic human faeces and faecal sludge for sanitation and wastewater research. Water Res 2018;132:222–40.