Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Mathematical Biosciences 347 (2022) 108805

Contents lists available at ScienceDirect

Mathematical Biosciences

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

Original Research Article

A hybrid stochastic model and its Bayesian identification for infectious disease screening in a university campus with application to massive COVID-19 screening at the University of Liège

M. Arnst [a,*], G. Louppe [a], R. Van Hulle [a], L. Gillet [b], F. Bureau [c], V. Denoël [a]

[a] Faculty of Applied Sciences, Université de Liège, 4000 Liège, Belgium
[b] Immunology–Vaccinology, FARAH, Université de Liège, 4000 Liège, Belgium
[c] Laboratory of Cellular and Molecular Immunology, GIGA Institute, Université de Liège, 4000 Liège, Belgium

ARTICLE INFO

Keywords:
Infectious disease transmission
COVID-19
Screening testing
Stochastic dynamics
Bayesian inference
Simulation-based inference

ABSTRACT

Amid the COVID-19 pandemic, universities are implementing various prevention and mitigation measures. Identifying and isolating infectious individuals by using screening testing is one such a measure that can contribute to reducing spread. Here, we propose a hybrid stochastic model for infectious disease transmission in a university campus with screening testing and its surrounding community. Based on a compartmental modeling strategy, this hybrid stochastic model represents the evolution of the infectious disease and its transmission using continuous-time stochastic dynamics, and it represents the screening testing as discrete stochastic events. We also develop, in a Bayesian framework, the identification of parameters of this hybrid stochastic model, including transmission rates. These parameters were identified from the screening test data for the university population and observed incidence counts for the surrounding community. We implement the exploration of the Bayesian posterior using a machine-learning simulation-based inference approach. The proposed methodology was applied in a retrospective modeling study of a massive COVID-19 screening conducted at the University of Liège in Fall 2020. The emphasis of the paper is on the development of the hybrid stochastic model to assess the impact of screening testing as a measure to reduce spread. The hybrid stochastic model allows various factors to be represented and examined, such as interplay with the surrounding community, variability of the transmission dynamics, the rate of participation in the screening testing, the test sensitivity, the test frequency, the diagnosis delay, and compliance with isolation. The application in the retrospective modeling study suggests that a high rate of participation and a high test frequency are important factors to reduce spread.

1. Introduction

Amid the COVID-19 pandemic universities are implementing various prevention and mitigation measures. Screening tests for SARS-CoV-2 are intended to identify occurrence at the individual level, even if there is no reason to suspect infection, e.g., there is no known exposure: they are intended to identify infected individuals, without symptoms or prior to development of symptoms, who may be contagious, so that measures can be taken to prevent further transmission [1]. Screening of the university population can take different forms, such as one-time entry testing or periodic screening, and there can be different ways of selecting individuals who participate, such as mandatory participation of certain cohorts or testing randomly selected groups of individuals [2].

Mathematical modeling and numerical simulation can help investigate the efficacy of screening. A number of mathematical modeling and numerical simulation studies have already been carried out to investigate the effects of periodic screening on the spread of SARS-CoV-2 in university environments. Paltiel et al. [3] used a compartmental model to investigate effects of several periodic screening strategies, with different scenarios for the frequency and the sensitivity and specificity of the test, on the number of infections that could occur. Hill et al. [4] used a more detailed compartmental model on a network to capture details of household, study group, student association and sport club, and other contacts. Gressmann and Peck [5] used a stochastic agent-based model that captures patterns of academic, classroom, professional, environmental, and residential contacts. Using a compartmental model, Larremore et al. [6] found that periodic screening should

* Corresponding author.
E-mail addresses: maarten.arnst@uliege.be (M. Arnst), g.louppe@uliege.be (G. Louppe), romain.vanhulle@uliege.be (R. Van Hulle), l.gillet@uliege.be (L. Gillet), fabrice.bureau@uliege.be (F. Bureau), v.denoel@uliege.be (V. Denoël).

https://doi.org/10.1016/j.mbs.2022.108805
Received 6 July 2021; Received in revised form 25 February 2022; Accepted 2 March 2022
Available online 16 March 2022

prioritize accessibility, frequency, and turnaround time, more than the sensitivity of the test. Sturniolo et al. [7] proposed a methodology for incorporating representations of contact tracing and isolation in compartmental models. Enright et al. [8] bring together research carried out by multiple research groups on mathematical modeling of SARS-CoV-2 transmission in UK higher education settings. Using multiple mathematical modeling approaches, including statistical data analysis, compartmental modeling, and network modeling, they investigate within-institution spread, interplay with the wider community, and the impact of control measures, including repeat screening testing.

Paltiel et al. [3] had adopted for both the transmission dynamics and the periodic screening a continuous-time representation, according to which changes in infectiousness status happen and cases are detected continuously at rates that are constant in time. In previous work [9], we had extended Paltiel et al.'s model by including a larger surrounding community, which we had endowed with its own compartments to study the transmission in the surrounding population and its interaction with the transmission in the university population and the screening. However, important aspects of periodic screening, such as the temporally discrete nature of consecutive tests and the interplay between the duration of the infectious period and the period between consecutive tests, can be difficult to capture with a continuous-time representation of the periodic screening. In addition, Paltiel et al.'s model is essentially deterministic, and it does not take into account variability in the epidemic process (although Paltiel et al. ran several scenarios).

Here, we revisit the mathematical modeling and numerical simulation of infectious disease transmission in a university campus with periodic screening and in its surrounding community. We develop a hybrid stochastic model, which is *hybrid* in that it combines continuous-time and discrete-time parts. We model the evolution of the infectious disease and its transmission using continuous-time stochastic dynamics, and we model the effects of the screening as discrete stochastic events. More specifically, we develop a continuous-time deterministic compartmental model of the infectious disease and its transmission, with which, following [10,11], we associate a stochastic counterpart in the form of a stochastic jump Markov model, which we finally approximate with a stochastic diffusion model. Following [12,13], we subdivide the compartments of infectious individuals into multiple subcompartments, as a mathematical device to control the probability distribution of the duration of the infectious period and its amount of dispersion. We model the effects of the periodic screening as discrete stochastic events that occur periodically in time and result in changes whose probability distribution accounts for the test sensitivity and the participation rate. Finally, we introduce waiting compartments that serve to account for the diagnosis delay.

We also develop, in a Bayesian framework, the identification of parameters of the hybrid stochastic model, including transmission rates. These parameters were identified from the screening data for the university population and observed incidence counts for the surrounding community. We implement the exploration of the Bayesian posterior with a machine-learning simulation-based inference approach [14].

It should be noted that SARS-CoV-2 screening has also been considered for contexts other than university campuses, such as screening of very large populations, such as an entire country's population [15]. Libin et al. [16] have investigated the effectiveness of periodic screening for such very large populations by using an agent-based model. The way in which Libin et al. [16] incorporate periodic screening in an agent-based model shares some aspects with how we incorporate screening in a stochastic compartmental model, notably, Libin et al. [16] also model the effects of the screening as discrete stochastic events whose probabilistic description accounts for properties such as the test sensitivity (false negative rate), the participation rate (testing compliance), the diagnosis delay (reporting delay), and other properties.

The mathematical model described in this paper has been developed in the context of a massive COVID-19 screening that has been conducted at the University of Liège in Belgium since Fall 2020. We apply the proposed hybrid stochastic model in a retrospective modeling study of this massive COVID-19 screening conducted at the University of Liège, specifically using screening data collected during the Fall 2020 wave.

This paper is organized as follows. First, in Section 2, we present the context of the massive COVID-19 screening at the University of Liège. Then, in Sections 3 and 4, we describe the novel hybrid stochastic model and its Bayesian identification. Subsequently, in Section 5, we describe the implementation of the Bayesian identification using the simulation-based inference method. Finally, in Section 6, we provide the application to the retrospective modeling study of the massive COVID-19 screening at the University of Liège.

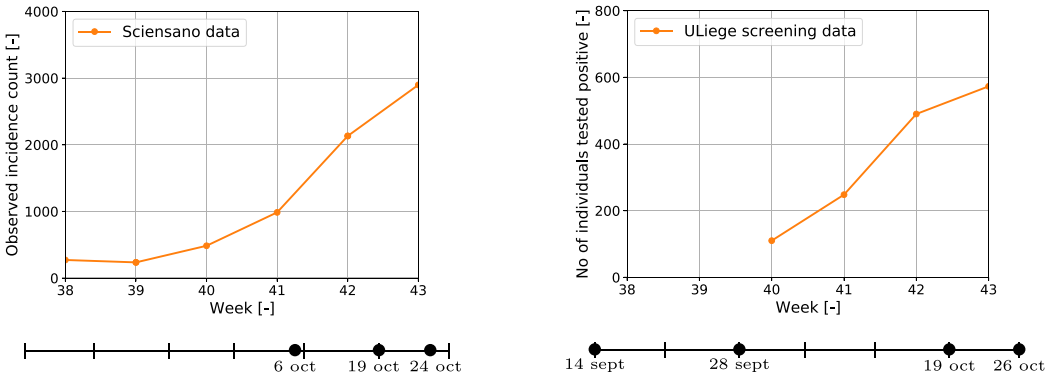
2. Context of massive COVID-19 screening at University of Liège

The University of Liège is a public and multidisciplinary university of the French Community of Belgium. The university has about 24 383 students and 5961 faculty and staff members. The university's activities span four campuses, two of which are located in or nearby the city of Liège, namely, the city center campus, home to the central administration, two faculties, and the management school, and the Sart-Tilman campus, the largest campus, home to 7 faculties and the university hospital, located in a wooded area on the outskirts of the city of Liège, about 10 km from the city center. The Gembloux campus, home to the Agro-Bio Tech faculty, and the Arlon campus, home to the department of environmental science and management, are located about 70 and 130 km from Liège, respectively.

Students returned to the university for the Fall 2020 term on Monday 14 September (the beginning of week 38) under a so-called *yellow* sanitary code, which involved measures such as a face mask requirement (in classrooms), increased physical distancing (distance of 1 m or occupation of 1 place out of 2 in classrooms), and increased ventilation. For many lectures, students had the choice to either attend in person or follow online though live streaming or podcasts. Confronted with rising new infections, the University of Liège adopted on Monday 19 October the *orange* sanitary code, and then on Monday 26 October (the beginning of week 44) the *red* sanitary code, which entailed all lectures to transition to an online format only, although certain laboratory sessions were allowed to continue on campus. Later in this paper, we will apply the proposed hybrid stochastic model in a retrospective modeling study relevant to the period from Monday 14 September (the beginning of week 38) (the return of students) to Monday 26 October (the beginning of week 44) (the transition to online lectures only). Please see also Fig. 1(b) for a schematic overview of the moments at which the aforementioned policy decisions were taken during this period of time.

In Belgium, several restrictions were in place in early September, including a mask requirement in shops and restrictions on social contacts defined in terms of household bubbles. Seeing rising infections, the government decided on 6 October that bars must close at 23:00, on 19 October a curfew from 0:00 to 6:00, a work-from-home mandate, the closure of bars and restaurants, and tightened limitations on social contacts, and on 24 October a curfew from 22:00 to 6:00 in the Wallonia region, in which Liège is located. Amid alarming signals of the hospital system and intensive care units reaching capacity, the country went into a new lockdown from November 2 onwards. Please see also Fig. 1(a) for a schematic overview of the moments at which the aforementioned policy decisions were taken in the studied period of time.

The University of Liège conducted a massive COVID-19 screening test program, which relied on pooled testing of saliva samples. This program stood out by its massive scope, with the capability to process about 10 000 tests per day (in pools of 3 samples, thus corresponding to about 3300 PCR evaluations). Starting on Monday 28 September, all students, faculty, and staff were offered the possibility to take a test, once a week, on a voluntary basis, anonymously, and with the



(a) Surrounding community. (b) University population.

Fig. 1. Context of massive COVID-19 screening at University of Liège: (a) weekly aggregated observed incidence counts for the city of Liège extracted from data made available by Belgium's public health institute Sciensano and (b) weekly aggregated screening counts (number of individuals tested positive) for the city center and Sart-Tilman campuses included in the University of Liège's massive COVID-19 screening test program. In the schematic representations of the time period under consideration, the bullets indicate the moments at which the policy decisions described in Section 2 were taken.

Table 1
Context of massive COVID-19 screening at University of Liège: Data for the period under study from the students returning to the university on Monday 14 September (the beginning of week 38) until the university adopted the red sanitary code and switched to fully online lectures (the end of week 43). The dash symbols in the table under (b) indicate that there was no screening in weeks 38 and 39.

(a) Observed incidence counts for the city of Liège.

Week	M	T	W	T	F	S	S	Weekly count
38	46	39	53	58	48	18	10	272
39	30	50	59	33	23	31	10	236
40	68	87	71	89	98	57	15	485
41	125	158	178	171	173	147	34	986
42	285	394	331	327	385	294	116	2132
43	394	486	467	525	586	356	84	2898

(b) Screening test data for the combination of the University of Liège's city center and Sart-Tilman campuses.

Week	Participation rate	Positivity rate
38	–	–
39	–	–
40	22.76%	1.70%
41	27.41%	3.19%
42	37.06%	4.66%
43	29.21%	6.92%

cost covered by the institution. All students, faculty, and staff were assigned two days in the week, on one of which they could retrieve on campus a testing kit with instructions on how to collect a saliva sample. They were asked to return their kit with their sample the next day on campus and asked to consult a website to obtain their test result about 12 to 36 h later. Individuals who tested positive for SARS-CoV-2 were recommended to contact a doctor to obtain medical advice and isolate themselves for 10 days. From Monday 26 October onwards, the program was suspended, and the test capacity was exploited to set up a screening test program in retirement homes. Table 1(b) and Fig. 1(b) provide summary statistics for the combination of the city center and Sart-Tilman campuses. The participation rates indicate that about 30% of the students and faculty and staff members participated, and the positivity rates indicate infections rising significantly in the university population.

Belgium built out significant capacity for nasopharyngeal swab testing with PCR evaluation, to which the University of Liège was also a significant contributor. Belgium's Public Health Institute Sciensano publishes daily observed incidence counts per municipality, among other information. With the system reaching its limits in the second half of October, the testing policy was changed, and, from 20 October onwards, testing was limited to symptomatic individuals only (individuals having had a high-risk contact or having traveled to a high-risk zone but not exhibiting symptoms could no longer take a test). Table 1(a) and Fig. 1(a) provide summary statistics for the city of Liège.

3. Hybrid stochastic model

In the context of the massive COVID-19 screening conducted at the University of Liège since Fall 2020, we have developed a mathematical model of the transmission dynamics of an infectious disease in a university campus with periodic screening situated within a larger surrounding community. Even though this development was motivated by the specific context of the massive COVID-19 screening at the University of Liège, the mathematical model is general and applicable to other contexts. It allows various factors to be represented and examined, such as interplay with the surrounding community, variability of the transmission dynamics, rate of participation in the screening test program, test sensitivity, test frequency, and diagnosis delay.

3.1. Hybrid deterministic model

The model describes the transmission dynamics of an infectious disease in a university campus with periodic screening situated within a larger surrounding community. The combined population is split into two subpopulations, namely, one subpopulation of n individuals that represents the university population with the screening test program, and another subpopulation of $\tilde{n}$ individuals that represents the surrounding community. Here, and in the following, the tilde is used to refer to the surrounding community.

The model adopts a compartmental modeling approach that further subdivides each subpopulation into disjoint compartments. Fig. 2 shows a graphical representation of the model. The population of the surrounding community is subdivided into compartments that track the number of susceptible ($\tilde{s}$), exposed, that is, infected but not yet infectious, ($\tilde{e}$), infectious ($\tilde{i}_1, \dots, \tilde{i}_{\tilde{m}}$), and removed due to recovery ($\tilde{r}$) individuals. The subdivision into $\tilde{m}$ subcompartments serves to control the shape of the probability distribution of the duration of the infectious period and its dispersion, namely, it serves to mimic an infectious period that is gamma distributed, as explained later in Section 3.3. The university population is subdivided into compartments that track the number of individuals categorized as susceptible (s), exposed (e), pre-symptomatically infectious ($i^{\mathrm{p}}_{11}, \dots, i^{\mathrm{p}}_{1m_{\mathrm{p}}}, \dots, i^{\mathrm{p}}_{c1}, \dots, i^{\mathrm{p}}_{cm_{\mathrm{p}}}$), symptomatically infectious ($i^{\mathrm{s}}_1, \dots, i^{\mathrm{s}}_{m_{\mathrm{s}}}$), asymptomatically infectious ($i^{\mathrm{a}}_{11}, \dots, i^{\mathrm{a}}_{1m_{\mathrm{a}}}, \dots, i^{\mathrm{a}}_{c1}, \dots, i^{\mathrm{a}}_{cm_{\mathrm{a}}}$), and removed ($r$). The compartments of pre-symptomatically and symptomatically infectious individuals concern infectious individuals who develop symptoms during the course of their disease. The pre-symptomatically infectious individuals are those who, early in the course of the disease, have already become infectious but have not developed symptoms yet, and the symptomatically infectious individuals are those who, later during the course of the disease, have developed symptoms and are still infectious. The compartments of asymptomatically infectious individuals concern infectious individuals who do not develop symptoms during the course of their disease. A number of additional compartments are created to track individuals with a positive screening test: the compartments $w^{\mathrm{p}}_{11}, \dots, w^{\mathrm{p}}_{1m_{\mathrm{p}}}, \dots, w^{\mathrm{p}}_{c1}, \dots, w^{\mathrm{p}}_{cm_{\mathrm{p}}}$ and $w^{\mathrm{a}}_{11}, \dots, w^{\mathrm{a}}_{1m_{\mathrm{a}}}, \dots, w^{\mathrm{a}}_{c1}, \dots, w^{\mathrm{a}}_{cm_{\mathrm{a}}}$ are "waiting" compartments that serve to account for diagnosis delay, and the compartments $q^{\mathrm{p}}_{11}, \dots, q^{\mathrm{p}}_{1m_{\mathrm{p}}}, \dots, q^{\mathrm{p}}_{c1}, \dots, q^{\mathrm{p}}_{cm_{\mathrm{p}}}$ and $q^{\mathrm{a}}_{11}, \dots, q^{\mathrm{a}}_{1m_{\mathrm{a}}}, \dots, q^{\mathrm{a}}_{c1}, \dots, q^{\mathrm{a}}_{cm_{\mathrm{a}}}$ are "quarantine" compartments that serve to track the number of positively tested pre-symptomatically and asymptomatically infectious individuals who have isolated themselves after receiving their positive diagnosis and who remain infectious. The subdivisions into m_{p}, m_{s}, and m_{a} subcompartments serve to control the shapes of the probability distributions of the durations of the pre-symptomatically, symptomatically, and asymptomatically infectious periods and their dispersion. The subdivisions into c subcompartments serve to allow for a possible subdivision of the university population into c cohorts with a different screening testing policy, such as a weekly screening test for the first cohort on Mondays, a weekly screening test for the second cohort on Tuesdays, and so forth, as explained later in this section.

The model is then obtained by defining transitions of individuals from one compartment to another. The model combines a continuous-time representation of the dynamics of the infectious disease and its transmission (arrows in the figure) with a discrete-time representation of the screening testing (dashed lines in the figure). The equations for the deterministic continuous-time dynamics are

$$
\begin{cases}
\dfrac{d\tilde{s}}{dt} = -\beta_{00}\,\tilde{s}\,\dfrac{\tilde{i}_1+\cdots+\tilde{i}_{\tilde{m}}}{\tilde{n}} - \beta_{01}\,\tilde{s}\,\left(\dfrac{i^{\mathrm{p}}_{11}+\cdots+i^{\mathrm{p}}_{cm_{\mathrm{p}}}+w^{\mathrm{p}}_{11}+\cdots+w^{\mathrm{p}}_{cm_{\mathrm{p}}}}{n} + \dfrac{i^{\mathrm{a}}_{11}+\cdots+i^{\mathrm{a}}_{cm_{\mathrm{a}}}+w^{\mathrm{a}}_{11}+\cdots+w^{\mathrm{a}}_{cm_{\mathrm{a}}}}{n}\right),\\
\dfrac{d\tilde{e}}{dt} = \beta_{00}\,\tilde{s}\,\dfrac{\tilde{i}_1+\cdots+\tilde{i}_{\tilde{m}}}{\tilde{n}} + \beta_{01}\,\tilde{s}\,\left(\dfrac{i^{\mathrm{p}}_{11}+\cdots+i^{\mathrm{p}}_{cm_{\mathrm{p}}}+w^{\mathrm{p}}_{11}+\cdots+w^{\mathrm{p}}_{cm_{\mathrm{p}}}}{n} + \dfrac{i^{\mathrm{a}}_{11}+\cdots+i^{\mathrm{a}}_{cm_{\mathrm{a}}}+w^{\mathrm{a}}_{11}+\cdots+w^{\mathrm{a}}_{cm_{\mathrm{a}}}}{n}\right) - \gamma\tilde{e},\\
\dfrac{d\tilde{i}_1}{dt} = \gamma\tilde{e} - \tilde{m}\tilde{\omega}\,\tilde{i}_1,\\
\dfrac{d\tilde{i}_\ell}{dt} = \tilde{m}\tilde{\omega}\,\tilde{i}_{\ell-1} - \tilde{m}\tilde{\omega}\,\tilde{i}_\ell, \quad \ell = 2,\dots,\tilde{m},\\
\dfrac{d\tilde{r}}{dt} = \tilde{m}\tilde{\omega}\,\tilde{i}_{\tilde{m}},
\end{cases}
\tag{1}
$$

$$
\begin{cases}
\dfrac{ds}{dt} = -\beta_{10}\,s\,\dfrac{\tilde{i}_1+\cdots+\tilde{i}_{\tilde{m}}}{\tilde{n}} - \beta_{11}\,s\,\left(\dfrac{i^{\mathrm{p}}_{11}+\cdots+i^{\mathrm{p}}_{cm_{\mathrm{p}}}+w^{\mathrm{p}}_{11}+\cdots+w^{\mathrm{p}}_{cm_{\mathrm{p}}}}{n} + \dfrac{i^{\mathrm{a}}_{11}+\cdots+i^{\mathrm{a}}_{cm_{\mathrm{a}}}+w^{\mathrm{a}}_{11}+\cdots+w^{\mathrm{a}}_{cm_{\mathrm{a}}}}{n}\right),\\
\dfrac{de}{dt} = \beta_{10}\,s\,\dfrac{\tilde{i}_1+\cdots+\tilde{i}_{\tilde{m}}}{\tilde{n}} + \beta_{11}\,s\,\left(\dfrac{i^{\mathrm{p}}_{11}+\cdots+i^{\mathrm{p}}_{cm_{\mathrm{p}}}+w^{\mathrm{p}}_{11}+\cdots+w^{\mathrm{p}}_{cm_{\mathrm{p}}}}{n} + \dfrac{i^{\mathrm{a}}_{11}+\cdots+i^{\mathrm{a}}_{cm_{\mathrm{a}}}+w^{\mathrm{a}}_{11}+\cdots+w^{\mathrm{a}}_{cm_{\mathrm{a}}}}{n}\right) - \gamma e,\\
\dfrac{di^{\mathrm{p}}_{k1}}{dt} = f_{\mathrm{s}} f_k \gamma e - m_{\mathrm{p}}\omega_{\mathrm{p}}\,i^{\mathrm{p}}_{k1}, \quad k = 1,\dots,c,\\
\dfrac{di^{\mathrm{p}}_{k\ell}}{dt} = m_{\mathrm{p}}\omega_{\mathrm{p}}\,i^{\mathrm{p}}_{k(\ell-1)} - m_{\mathrm{p}}\omega_{\mathrm{p}}\,i^{\mathrm{p}}_{k\ell}, \quad k = 1,\dots,c, \quad \ell = 2,\dots,m_{\mathrm{p}},\\
\dfrac{dw^{\mathrm{p}}_{k1}}{dt} = -m_{\mathrm{p}}\omega_{\mathrm{p}}\,w^{\mathrm{p}}_{k1}, \quad k = 1,\dots,c,\\
\dfrac{dw^{\mathrm{p}}_{k\ell}}{dt} = m_{\mathrm{p}}\omega_{\mathrm{p}}\,w^{\mathrm{p}}_{k(\ell-1)} - m_{\mathrm{p}}\omega_{\mathrm{p}}\,w^{\mathrm{p}}_{k\ell}, \quad k = 1,\dots,c, \quad \ell = 2,\dots,m_{\mathrm{p}},\\
\dfrac{dq^{\mathrm{p}}_{k1}}{dt} = -m_{\mathrm{p}}\omega_{\mathrm{p}}\,q^{\mathrm{p}}_{k1}, \quad k = 1,\dots,c,\\
\dfrac{dq^{\mathrm{p}}_{k\ell}}{dt} = m_{\mathrm{p}}\omega_{\mathrm{p}}\,q^{\mathrm{p}}_{k(\ell-1)} - m_{\mathrm{p}}\omega_{\mathrm{p}}\,q^{\mathrm{p}}_{k\ell}, \quad k = 1,\dots,c, \quad \ell = 2,\dots,m_{\mathrm{p}},\\
\dfrac{di^{\mathrm{s}}_1}{dt} = m_{\mathrm{p}}\omega_{\mathrm{p}}\left(i^{\mathrm{p}}_{11}+\cdots+i^{\mathrm{p}}_{cm_{\mathrm{p}}}+w^{\mathrm{p}}_{11}+\cdots+w^{\mathrm{p}}_{cm_{\mathrm{p}}}+q^{\mathrm{p}}_{11}+\cdots+q^{\mathrm{p}}_{cm_{\mathrm{p}}}\right) - m_{\mathrm{s}}\omega_{\mathrm{s}}\,i^{\mathrm{s}}_1,\\
\dfrac{di^{\mathrm{s}}_\ell}{dt} = m_{\mathrm{s}}\omega_{\mathrm{s}}\,i^{\mathrm{s}}_{(\ell-1)} - m_{\mathrm{s}}\omega_{\mathrm{s}}\,i^{\mathrm{s}}_\ell, \quad \ell = 2,\dots,m_{\mathrm{s}},\\
\dfrac{di^{\mathrm{a}}_{k1}}{dt} = (1-f_{\mathrm{s}}) f_k \gamma e - m_{\mathrm{a}}\omega_{\mathrm{a}}\,i^{\mathrm{a}}_{k1}, \quad k = 1,\dots,c,\\
\dfrac{di^{\mathrm{a}}_{k\ell}}{dt} = m_{\mathrm{a}}\omega_{\mathrm{a}}\,i^{\mathrm{a}}_{k(\ell-1)} - m_{\mathrm{a}}\omega_{\mathrm{a}}\,i^{\mathrm{a}}_{k\ell}, \quad k = 1,\dots,c, \quad \ell = 2,\dots,m_{\mathrm{a}},\\
\dfrac{dw^{\mathrm{a}}_{k1}}{dt} = -m_{\mathrm{a}}\omega_{\mathrm{a}}\,w^{\mathrm{a}}_{k1}, \quad k = 1,\dots,c,\\
\dfrac{dw^{\mathrm{a}}_{k\ell}}{dt} = m_{\mathrm{a}}\omega_{\mathrm{a}}\,w^{\mathrm{a}}_{k(\ell-1)} - m_{\mathrm{a}}\omega_{\mathrm{a}}\,w^{\mathrm{a}}_{k\ell}, \quad k = 1,\dots,c, \quad \ell = 2,\dots,m_{\mathrm{a}},\\
\dfrac{dq^{\mathrm{a}}_{k1}}{dt} = -m_{\mathrm{a}}\omega_{\mathrm{a}}\,q^{\mathrm{a}}_{k1}, \quad k = 1,\dots,c,\\
\dfrac{dq^{\mathrm{a}}_{k\ell}}{dt} = m_{\mathrm{a}}\omega_{\mathrm{a}}\,q^{\mathrm{a}}_{k(\ell-1)} - m_{\mathrm{a}}\omega_{\mathrm{a}}\,q^{\mathrm{a}}_{k\ell}, \quad k = 1,\dots,c, \quad \ell = 2,\dots,m_{\mathrm{a}},\\
\dfrac{dr}{dt} = m_{\mathrm{s}}\omega_{\mathrm{s}}\,i^{\mathrm{s}}_{m_{\mathrm{s}}} + m_{\mathrm{a}}\omega_{\mathrm{a}}\left(i^{\mathrm{a}}_{11}+\cdots+i^{\mathrm{a}}_{cm_{\mathrm{a}}}+w^{\mathrm{a}}_{11}+\cdots+w^{\mathrm{a}}_{cm_{\mathrm{a}}}+q^{\mathrm{a}}_{11}+\cdots+q^{\mathrm{a}}_{cm_{\mathrm{a}}}\right).
\end{cases}
\tag{2}
$$

In the surrounding community, the force of infection reads $\beta_{00}\,\frac{\tilde{i}_1+\cdots+\tilde{i}_{\tilde{m}}}{\tilde{n}} + \beta_{01}\left(\frac{i^{\mathrm{p}}_{11}+\cdots+i^{\mathrm{p}}_{cm_{\mathrm{p}}}+w^{\mathrm{p}}_{11}+\cdots+w^{\mathrm{p}}_{cm_{\mathrm{p}}}}{n} + \frac{i^{\mathrm{a}}_{11}+\cdots+i^{\mathrm{a}}_{cm_{\mathrm{a}}}+w^{\mathrm{a}}_{11}+\cdots+w^{\mathrm{a}}_{cm_{\mathrm{a}}}}{n}\right)$. The first term represents endogenous transmission and is given by the product of the transmission rate β_{00} and the prevalence $\frac{\tilde{i}_1+\cdots+\tilde{i}_{\tilde{m}}}{\tilde{n}}$ within the surrounding community, whereby the transmission rate β_{00} captures the combined effect of the probability of infection as a result of a contact and the rate of contacts of an individual in the surrounding community with other individuals within the surrounding community. The second term represents exogenous transmission due to contacts with the university population and is given by the product of the transmission rate β_{01} and the prevalence $\frac{i^{\mathrm{p}}_{11}+\cdots+i^{\mathrm{p}}_{cm_{\mathrm{p}}}+w^{\mathrm{p}}_{11}+\cdots+w^{\mathrm{p}}_{cm_{\mathrm{p}}}}{n} + \frac{i^{\mathrm{a}}_{11}+\cdots+i^{\mathrm{a}}_{cm_{\mathrm{a}}}+w^{\mathrm{a}}_{11}+\cdots+w^{\mathrm{a}}_{cm_{\mathrm{a}}}}{n}$ of non-isolated infectious individuals within the university population, whereby the factor β_{01} captures the combined effect of the probability of infection as a result of a contact and the rate of contacts of an individual in the surrounding community with individuals from the university population. Pre-symptomatically

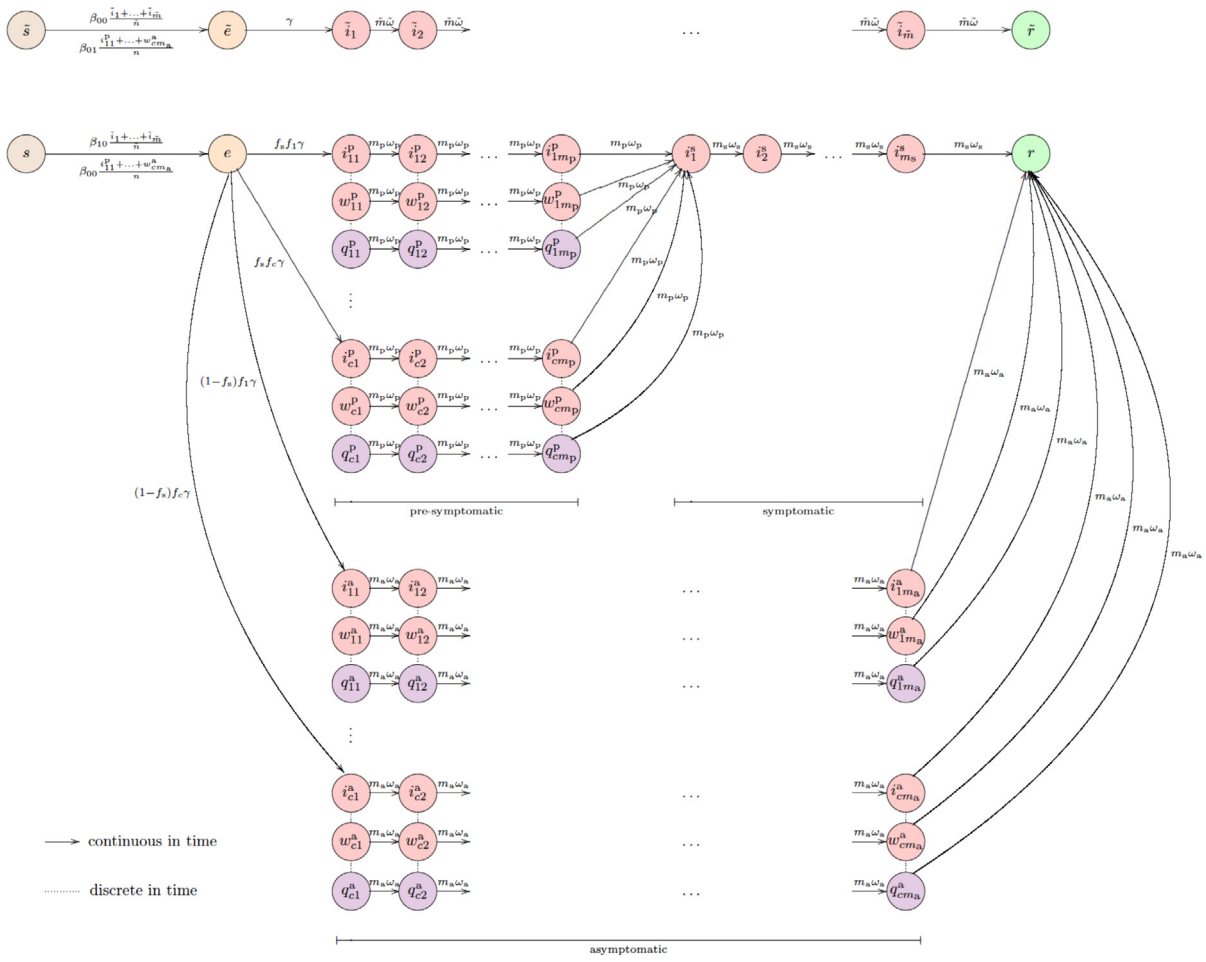


Fig. 2. Representation of the model as a directed graph.

infectious individuals are taken to contribute to transmission with the same transmission rate as infectious individuals who do not develop symptoms. Exposed individuals become infectious at rate γ (so that $1/\gamma$ can be interpreted as a characteristic duration of the latent period to the onset of infectiousness), and infectious individuals become removed at rate $\tilde{\omega}$ (so that $1/\tilde{\omega}$ can be interpreted as a characteristic duration of the infectious period).

In the university population, the force of infection is composed, analogously, of one term that represents endogenous transmission from contacts within the university population and another term that represents exogenous transmission due to contacts with the surrounding community. Among the exposed individuals, a fraction f_s develops symptoms during the course of the disease. At a rate of γ, these exposed individuals become pre-symptomatically infectious. At a rate of ω_p (so that ω_p^{-1} can be interpreted as a characteristic duration of the pre-symptomatically infectious period from the onset of infectiousness to the onset of symptoms), pre-symptomatically infectious individuals become symptomatically infectious. Symptomatically infectious individuals are taken to isolate spontaneously at the onset of symptoms and not contribute to transmission. We will comment on imperfect compliance with isolation of symptomatically infectious individuals later in Section 7. The complementary fraction $(1 - f_s)$ of the exposed individuals does not develop symptoms during the course of the disease. At a rate of γ, these exposed individuals become asymptomatically infectious.

As already mentioned, the partitioning into c cohorts is associated with different screening testing policies. The fractions $f_1, \ldots, f_c$, with $f_1 + \cdots + f_c = 1$, are the fractions of the total university population that these cohorts are composed of. Symptomatically and asymptomatically infectious individuals become removed at rates ω_s and ω_a, respectively.

Because contacts are reciprocal by nature, the transmission rates β_{01} and β_{10} should satisfy the reciprocity property

$$\beta_{01}\,\tilde{n} = \beta_{10}\,n. \tag{3}$$

These equations for the deterministic continuous-time dynamics of the infectious disease and its transmission can be written more concisely as follows. The compartment counts can be collected in a state vector

$$\boldsymbol{x}(t) = \big(\tilde{s}(t), \tilde{e}(t), \tilde{i}_1(t), \ldots, \tilde{i}_{\tilde{m}}(t), \tilde{r}(t), s(t), \ldots, r(t)\big). \tag{4}$$

The equations can be then described in terms of a matrix whose columns specify the changes in the compartment counts per transition,

$$[S] = \begin{array}{c} \\ \tilde{s} \\ \tilde{e} \\ \tilde{i}_1 \\ \vdots \end{array} \begin{array}{c} \begin{array}{cccc} \tilde{s}\to\tilde{e} & \tilde{e}\to\tilde{i}_1 & \cdots & \end{array} \\ \begin{bmatrix} -1 & 0 & 0 & \cdots \\ 1 & -1 & 0 & \cdots \\ 0 & 1 & \vdots & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \end{array}, \tag{5}$$

and a vector that specifies the propensities of these transitions:

$$\boldsymbol{g}(\boldsymbol{x}) = \begin{array}{c} \tilde{s}\to\tilde{e} \\ \tilde{e}\to\tilde{i}_1 \\ \vdots \end{array} \begin{bmatrix} -\beta_{00}\,\tilde{s}\,\dfrac{\tilde{i}_1+\cdots+\tilde{i}_{\tilde{m}}}{\tilde{n}} - \beta_{01}\,\tilde{s}\,\dfrac{i^{\mathrm{p}}_{11}+\cdots+i^{\mathrm{p}}_{cm_{\mathrm{p}}}+\cdots+w^{\mathrm{a}}_{11}+\cdots+w^{\mathrm{a}}_{cm_{\mathrm{a}}}}{n} \\ \gamma\tilde{e} \\ \vdots \end{bmatrix}. \tag{6}$$

With this state vector, matrix, and vector, the deterministic continuous-time dynamics of the infectious disease and its transmission read

concisely

$$\frac{d\boldsymbol{x}}{dt}(t) = \boldsymbol{f}\big(\boldsymbol{x}(t)\big) = [S]\boldsymbol{g}\big(\boldsymbol{x}(t)\big). \tag{7}$$

Let us now turn our attention to the periodic screening. The model considers that the screening results periodically in significant changes that happen over short periods of time, such as a large group of individuals taking a test over a short period of time, or a large group of individuals being informed of their diagnosis over a short period of time. Taking this behavior to the limit, the model considers that the screening results periodically in instantaneous changes in the state vector from a current value, denoted by $\boldsymbol{x}(t^-)$, to an updated value, denoted by $\boldsymbol{x}(t^+)$. In between these periodically recurring moments at which the effects of the screening testing manifest themselves, the evolution of the compartment counts consists of phases in which the state vector evolves continuously in accordance with (7).

As already mentioned, the university population is subdivided into c cohorts with a different screening testing policy. Let $t_{k1} = t_{k0} + \tau$, $t_{k2} = t_{k0} + 2\tau$, ... denote the times at which the individuals in the kth cohort take a test, with t_{k0} a reference time for the kth cohort and τ the period between any two consecutive tests. For each cohort, indexed by k, at each testing time t_{kj}, indexed by j, a fraction f_p of individuals is taken to take a test, and, from among the pre-symptomatically and asymptomatically infectious individuals, a fraction s_e of the tests is taken to return positive, with s_e the sensitivity of the test. Accordingly, at the testing times t_{kj}, with $k = 1, \dots, c$ and $j = 1, 2, \dots$, the compartment counts are changed instantaneously as follows:

$$\begin{cases} i^p_{k\ell}(t^+_{kj}) = i^p_{k\ell}(t^-_{kj}) - f_\mathrm{p} i^p_{k\ell}(t^-_{kj}) s_\mathrm{e}, \\ w^p_{k\ell}(t^+_{kj}) = f_\mathrm{p} i^p_{k\ell}(t^-_{kj}) s_\mathrm{e}, \end{cases} \quad \ell = 1, \dots, m_\mathrm{p}, \tag{8}$$

$$\begin{cases} i^a_{k\ell}(t^+_{kj}) = i^a_{k\ell}(t^-_{kj}) - f_\mathrm{p} i^a_{k\ell}(t^-_{kj}) s_\mathrm{e}, \\ w^a_{k\ell}(t^+_{kj}) = f_\mathrm{p} i^a_{k\ell}(t^-_{kj}) s_\mathrm{e}, \end{cases} \quad \ell = 1, \dots, m_\mathrm{a}. \tag{9}$$

The pre-symptomatically and asymptomatically infectious individuals testing positive are taken to still contribute to the transmission and their infectiousness status to evolve towards recovery while they await their positive diagnosis. In the model, this behavior is captured by moving these individuals from the compartments $i^p_{k1}, \dots, i^p_{km_\mathrm{p}}$ and $i^a_{k1}, \dots, i^a_{km_\mathrm{a}}$ of pre-symptomatically and asymptomatically infectious individuals to the associated "waiting" compartments $w^p_{k1}, \dots, w^p_{km_\mathrm{p}}$ and $w^a_{k1}, \dots, w^a_{km_\mathrm{a}}$. After the diagnosis delay, denoted by d, these individuals are taken to isolate themselves. Accordingly, at $t_{kj} + d$, $k = 1, \dots, c$, $j = 1, 2, \dots$, the compartment counts are changed instantaneously as

$$\begin{cases} w^p_{k\ell}\big((t_{kj}+d)^+\big) = 0, \\ q^p_{k\ell}\big((t_{kj}+d)^+\big) = q^p_{k\ell}\big((t_{kj}+d)^-\big) + w^p_{k\ell}\big((t_{kj}+d)^-\big), \end{cases} \quad \ell = 1, \dots, m_\mathrm{p}, \tag{10}$$

$$\begin{cases} w^a_{k\ell}\big((t_{kj}+d)^+\big) = 0, \\ q^a_{k\ell}\big((t_{kj}+d)^+\big) = q^a_{k\ell}\big((t_{kj}+d)^-\big) + w^a_{k\ell}\big((t_{kj}+d)^-\big), \end{cases} \quad \ell = 1, \dots, m_\mathrm{a}. \tag{11}$$

Individuals in the "quarantine" compartments $q^p_{k1}, \dots, q^p_{km_\mathrm{p}}$ and $q^a_{k1}, \dots, q^a_{km_\mathrm{a}}$ are taken to have isolated themselves and no longer transmit.

In summary, in the hybrid deterministic model, the trajectory of the compartment counts is partitioned into several time intervals. These time intervals are separated by the times at which the effects of the screening periodically manifest themselves and the state vector changes its value discontinuously in accordance with (9)–(11). Within these time intervals, the state vector evolves continuously in accordance with (7).

3.2. Stochastic Markov jump process model

Following [10,11], deterministic compartmental models are associated with stochastic compartmental models formulated in the framework of Markov jump processes. Markov jump processes are continuous-time stochastic processes that take their values in a finite or countably infinite set, see, for instance, [17–20], for reference texts. Because the set of values is finite or countably infinite, the trajectories must remain constant for a while in each new state, thus leading to jumps. For a homogeneous Markov jump process, that is, for a Markov jump process for which the probability laws governing the transitions are invariant under translation in time, the holding times, namely, the times between the consecutive jumps, are exponentially distributed. The correspondence between the deterministic compartmental model and its associated stochastic compartmental model is that the random fractions of individuals occupying the compartments in the stochastic compartmental model converge to the corresponding deterministic fractions in the deterministic compartmental model in the large-population limit.

These concepts are described for general compartmental models and applied to the classical SIR and other models in [10,11]. Here, we apply these concepts to our model. Thus, we associate the equations described above for the deterministic continuous-time dynamics of the infectious disease and its transmission with a Markov jump process

$$\{\boldsymbol{X}(t),\ t \geq 0\} \quad \text{with values in} \quad [0, 1, 2, \dots, \tilde{n}]^{\tilde{\kappa}} \times [0, 1, 2, \dots, n]^{\kappa}, \tag{12}$$

with

$$\boldsymbol{X}(t) = \big(\underbrace{\widetilde{S}(t), \widetilde{E}(t), \widetilde{I}_1(t), \dots, \widetilde{I}_{\tilde{m}}(t), \widetilde{R}(t)}_{\substack{\tilde{\kappa} = 3 + \tilde{m} \\ \text{compartments}}}, \underbrace{S(t), \dots, R(t)}_{\substack{\kappa = 3 + 3c(m_\mathrm{p} + m_\mathrm{a}) + m_\mathrm{s} \\ \text{compartments}}}\big), \tag{13}$$

in which uppercase regular letters denote scalar-valued random variables and uppercase boldface letters denote vector-valued random variables, a system of notation that we will seek to use throughout the remainder of the paper.

This Markov jump process is described by a probability distribution for the initial value and by a family of probability distributions governing the transitions defined in terms of the propensities in the equations above:

$$\mathbb{P}\big(\boldsymbol{X}(t+\Delta t) = \boldsymbol{x} + \boldsymbol{s}_{:i} \,\big|\, \boldsymbol{X}(t) = \boldsymbol{x}\big) = f_i(\boldsymbol{x})\Delta t + o(\Delta t), \quad i = 1, \dots, \zeta, \tag{14}$$

in which $\mathbb{P}$ denotes the probability and $\boldsymbol{s}_{:1}, \dots, \boldsymbol{s}_{:\zeta}$ are the ζ columns of the matrix $[S]$ defined in (5), namely,

$$\begin{aligned} &\mathbb{P}\Big(\big(\widetilde{S}(t+\Delta t), \widetilde{E}(t+\Delta t), \dots\big) = \big(\tilde{s}-1, \tilde{e}+1, \dots\big) \,\Big|\, \big(\widetilde{S}(t), \widetilde{E}(t), \dots\big) \\ &\quad = \big(\tilde{s}, \tilde{e}, \dots\big) \Big) \\ &\quad = \left(-\beta_{00}\, \tilde{s}\, \frac{\tilde{I}_1 + \dots + \tilde{I}_{\tilde{m}}}{\tilde{n}} - \beta_{01}\, \tilde{s}\, \frac{i^\mathrm{p}_{11} + \dots + i^\mathrm{p}_{cm_\mathrm{p}} + \dots + w^\mathrm{a}_{11} + \dots + w^\mathrm{a}_{cm_\mathrm{a}}}{n} \right) \Delta t \\ &\qquad + o(\Delta t), \\ &\quad \dots \end{aligned} \tag{15}$$

If the parameters involved in the propensities are independent of time, the Markov jump process is homogeneous, and an equivalent description is that it stays in state $\boldsymbol{x}$ during a holding time that is exponentially distributed with rate parameter $\sum_{i=1}^{\zeta} f_i(\boldsymbol{x})$ and then jumps to a new state selected from the $\boldsymbol{x}+\boldsymbol{s}_{:i}$ with probabilities $f_i(\boldsymbol{x}) / \sum_{j=1}^{\zeta} f_j(\boldsymbol{x})$.

3.3. Role of the subdivisions into $\tilde{m}$, m_p, m_s, and m_a subcompartments

If the subdivisions into multiple subcompartments are not used, that is, if $\tilde{m}$, m_p and m_s, and m_a are set equal to 1, then, in the stochastic model, the duration of the infectious period of infected individuals in the surrounding community, the durations of the pre-symptomatically and symptomatically infectious periods of infected individuals in the university population who develop symptoms, and the duration of the asymptomatically infectious period of infected individuals in the university population who do not develop symptoms are exponentially distributed with mean values $1/\tilde{\omega}$, $1/\omega_\mathrm{p}$ and $1/\omega_\mathrm{s}$, and $1/\omega_\mathrm{a}$, respectively. By contrast, with the subdivision into multiple subcompartments, these durations are obtained instead as the sum of

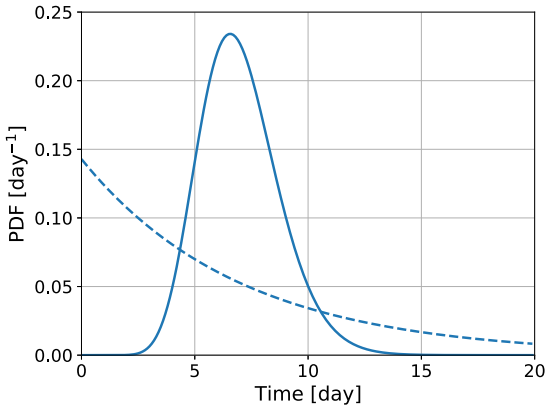


Fig. 3. Role of the subdivisions into $\tilde{m}$, m_p, m_s, and m_a subcompartments: Comparison of an exponential distribution with a mean value of 7 days (no subdivision) (dashed line) with a gamma distribution with a mean value of 7 days and a coefficient of variation of 25% (16 subcompartments) (solid line) for the duration of an infectious period.

$\tilde{m}$, m_p and m_s, and m_a exponentially distributed random variables with mean values $1/(\tilde{m}\tilde{\omega})$, $1/(m_\mathrm{p}\omega_\mathrm{p})$ and $1/(m_\mathrm{s}\omega_\mathrm{s})$, and $1/(m_\mathrm{a}\omega_\mathrm{a})$, respectively. Thus, with the well-known result that the sum of independent and identically distributed exponential random variables is gamma distributed, the durations of these periods are gamma distributed with mean values $1/\tilde{\omega}$, $1/\omega_\mathrm{p}$ and $1/\omega_\mathrm{s}$, and $1/\omega_\mathrm{a}$ and with coefficients of variation $1/\sqrt{\tilde{m}}$, $1/\sqrt{m_\mathrm{p}}$ and $1/\sqrt{m_\mathrm{s}}$, and $1/\sqrt{m_\mathrm{a}}$, respectively. Hence, through the subdivision into multiple subcompartments, the shape of the probability distributions of these durations of these infectious periods and their dispersion can be controlled [12,13].

Fig. 3 illustrates the effect of the subdivision on the distribution of the duration of an infectious period. We can observe that the exponential distribution (no subdivision) has a higher dispersion and assigns highest probabilities to very small values. By contrast, the gamma distribution (in the figure for a subdivision into 16 subcompartments) is more concentrated about its mean value and assigns negligible probabilities to very small values.

Thus, the role of the subdivisions into $\tilde{m}$, m_p, m_s, and m_a subcompartments is to serve as a mathematical device to strive to incorporate in the model an adequate representation of the dynamics of the infectiousness of the disease and its variability, as can be expected to be required to accurately assess the impact of control measures. Using adequate probability distributions for the durations of the infectious periods can be expected to be important to adequately capture the efficacy of periodic screening. Indeed, it can be expected to be important to adequately capture the interplay between the dynamics of the infectiousness of the disease and its variability, on the one hand, and the dynamics of the periodic screening, on the other hand. Notably, if the occurrence of very short durations of the pre-symptomatic or asymptomatic infectious periods is unrealistic, the subdivision into multiple subcompartments can be important to avoid that the model overestimates the number of individuals who develop symptoms or recover before they take a test and hence underestimates the efficacy of the screening.

3.4. Stochastic diffusion model

Following [10,11], Markov jump processes can be approximated with stochastic diffusion processes. The approximation is in that a measure of a distance between the random fractions of individuals occupying the compartments in the Markov jump process and the random fractions of individuals occupying the compartments in the approximating stochastic diffusion process decrease with an increase in the size of the population. Thus, for the approximation to be good, the population size must be sufficiently large. It is much more computationally efficient to numerically simulate trajectories from a stochastic diffusion process than from a Markov jump process [11].

As in the previous section, these concepts are described for general compartmental models and applied to the classical SIR and other models in [10,11], and, here, we apply these concepts to our model. Thus, we approximate the Markov jump process defined above with a stochastic diffusion process

$$\{\boldsymbol{X}(t),\ t \geq 0\} \quad \text{with values in} \quad \mathbb{R}^\kappa \times \mathbb{R}^\kappa, \tag{16}$$

that solves a stochastic differential equation written as

$$d\boldsymbol{X}(t) = \boldsymbol{b}\big(\boldsymbol{X}(t)\big)dt + \big[A\big(\boldsymbol{X}(t)\big)\big]d\boldsymbol{W}(t). \tag{17}$$

Please note that we do not distinguish in our system of notation between the Markov jump process $\{\boldsymbol{X}(t),\ t \geq 0\}$ in (12) and the stochastic diffusion process $\{\boldsymbol{X}(t),\ t \geq 0\}$ in (16). We will use the stochastic diffusion process throughout the remainder of this paper, so that the use of uppercase letters for the state vector and the compartment counts will always refer to the stochastic diffusion model throughout the remainder of this paper. In (17), $\{\boldsymbol{W}(t),\ t \geq 0\}$ is a normalized Wiener process indexed by $[0,+\infty[$ with values in $\mathbb{R}^\zeta$, the drift vector $\boldsymbol{b}(\boldsymbol{x})$ is given by

$$\boldsymbol{b}(\boldsymbol{x}) = \sum_{i=1}^{\zeta} f_i(\boldsymbol{x})\boldsymbol{s}_{:i} = [S]\boldsymbol{f}(\boldsymbol{x}), \tag{18}$$

and $[A(\boldsymbol{x})]$ is a square root of the diffusion matrix $[\Sigma(\boldsymbol{x})]$ given by

$$[\Sigma(\boldsymbol{x})] = \sum_{i=1}^{\zeta} f_i(\boldsymbol{x})\boldsymbol{s}_{:i} \otimes \boldsymbol{s}_{:i} = \big[S\big]\big[\mathrm{Diag}\big(f_1(\boldsymbol{x}),\dots,f_\zeta(\boldsymbol{x})\big)\big]\big[S\big]^\mathrm{T}; \tag{19}$$

here, we choose the square root

$$[A(\boldsymbol{x})] = \big[S\big]\big[\mathrm{Diag}\big(\sqrt{f_1(\boldsymbol{x})},\dots,\sqrt{f_\zeta(\boldsymbol{x})}\big)\big], \tag{20}$$

such that indeed $[\Sigma(\boldsymbol{x})] = [A(\boldsymbol{x})][A(\boldsymbol{x})]^\mathrm{T}$. Please note that $[A(\boldsymbol{x})]$ is rectangular, and the dimension of $\boldsymbol{W}(t)$ is different from the dimension of $\boldsymbol{X}(t)$.

The Euler–Maruyama method is a well-known method to numerically simulate trajectories of the solution to a stochastic differential equation. For a discretization of time with a constant time step Δt,

$$\sigma_0 = 0 < \sigma_1 = \Delta t < \sigma_2 = 2\Delta t < \cdots, \tag{21}$$

it leads to a time-stepping method of the form:

$$\boldsymbol{X}_{\sigma_{i+1}} = \boldsymbol{X}_{\sigma_i} + \boldsymbol{b}(\boldsymbol{X}_{\sigma_i})\Delta t + [A(\boldsymbol{X}_{\sigma_i})]\Delta\boldsymbol{W}_{\sigma_i}, \quad i = 0,1,\dots, \tag{22}$$

with an initial value that must be specified. In (22), the components of the random vectors $\Delta\boldsymbol{W}_{\sigma_0}$, $\Delta\boldsymbol{W}_{\sigma_1}$, ... are independent and identically distributed Gaussian random variables with zero mean and variance Δt.

3.5. Stochastic discrete events

In the stochastic model, at t_{kj}, $k = 1,\dots,c$, $j = 1, 2, \dots$, the numbers of pre-symptomatically and asymptotically infectious individuals identified by the screening are sampled from binomial probability distributions:

$$Y^\mathrm{p}_{kj} \sim \mathrm{Binomial}\left(\left\lfloor f_\mathrm{p}\sum_{\ell=1}^{m_\mathrm{p}} I^\mathrm{p}_{k\ell}(t^-_{kj})\right\rfloor, s_\mathrm{e}\right), \tag{23}$$

$$Y^\mathrm{a}_{kj} \sim \mathrm{Binomial}\left(\left\lfloor f_\mathrm{p}\sum_{\ell=1}^{m_\mathrm{a}} I^\mathrm{a}_{k\ell}(t^-_{kj})\right\rfloor, s_\mathrm{e}\right), \tag{24}$$

where $\lfloor\cdot\rfloor$ denotes the operation of rounding to the nearest integer below the argument, and the compartment counts are changed instantaneously as

$$\begin{cases} I^\mathrm{p}_{k\ell}(t^+_{kj}) = I^\mathrm{p}_{k\ell}(t^-_{kj}) - Y^\mathrm{p}_{kj} I^\mathrm{p}_{k\ell}(t^-_{kj})\Big/\sum_{\ell'=1}^{m_\mathrm{p}} I^\mathrm{p}_{k\ell'}(t^-_{kj}), \\ W^\mathrm{p}_{k\ell}(t^+_{kj}) = Y^\mathrm{p}_{kj} I^\mathrm{p}_{k\ell}(t^-_{kj})\Big/\sum_{\ell'=1}^{m_\mathrm{p}} I^\mathrm{p}_{k\ell'}(t^-_{kj}), \end{cases} \quad \ell = 1,\dots,m_\mathrm{p}, \tag{25}$$

$$\begin{cases} I^{\mathrm{a}}_{k\ell}(t^+_{kj}) = I^{\mathrm{a}}_{k\ell}(t^-_{kj}) - Y^{\mathrm{a}}_{kj} I^{\mathrm{a}}_{k\ell}(t^-_{kj}) \Big/ \sum_{\ell'=1}^{m_{\mathrm{a}}} I^{\mathrm{a}}_{k\ell'}(t^-_{kj}), \\ W^{\mathrm{a}}_{k\ell}(t^+_{kj}) = Y^{\mathrm{a}}_{kj} I^{\mathrm{a}}_{k\ell}(t^-_{kj}) \Big/ \sum_{\ell'=1}^{m_{\mathrm{a}}} I^{\mathrm{a}}_{k\ell'}(t^-_{kj}), \end{cases} \quad \ell = 1, \dots, m_{\mathrm{a}}. \tag{26}$$

Then, at $t_{kj} + d$, $k = 1, \dots, c$, $j = 1, 2, \dots$, the compartment counts are changed instantaneously as follows:

$$\begin{cases} W^{\mathrm{p}}_{k\ell}\big((t_{kj}+d)^+\big) = 0, \\ Q^{\mathrm{p}}_{k\ell}\big((t_{kj}+d)^+\big) = Q^{\mathrm{p}}_{k\ell}\big((t_{kj}+d)^-\big) + W^{\mathrm{p}}_{k\ell}\big((t_{kj}+d)^-\big), \end{cases} \quad \ell = 1, \dots, m_{\mathrm{p}}, \tag{27}$$

$$\begin{cases} W^{\mathrm{a}}_{k\ell}\big((t_{kj}+d)^+\big) = 0, \\ Q^{\mathrm{a}}_{k\ell}\big((t_{kj}+d)^+\big) = Q^{\mathrm{a}}_{k\ell}\big((t_{kj}+d)^-\big) + W^{\mathrm{a}}_{k\ell}\big((t_{kj}+d)^-\big), \end{cases} \quad \ell = 1, \dots, m_{\mathrm{a}}. \tag{28}$$

3.6. Hybrid stochastic model

In the hybrid stochastic model, the stochastic process of the compartment counts is partitioned into several intervals. These intervals are the time intervals separated by the times at which the effects of the screening periodically manifest themselves and the state vector changes its value discontinuously in accordance with (24)–(28). Within these time intervals, the state vector evolves continuously in accordance with (17).

4. Bayesian inference

We will now address the inference of parameters of the hybrid stochastic model from data (and prior information, if any) in a Bayesian framework.

4.1. Data

For the surrounding community, we consider that the data consist of observed incidence counts

$$\tilde{y}^{\mathrm{obs}}_{[t_0,t_1[}, \ \tilde{y}^{\mathrm{obs}}_{[t_1,t_2[}, \ \cdots \tag{29}$$

reported by a public health institute for periods of time

$$[t_{l-1}, t_l[, \quad l = 1, 2, \dots \tag{30}$$

The counts in (29) reflect a fraction of the number of (new) infections over the periods of time in (30). They represent only a fraction because of limitations of public health testing (limited detection rate; observation variability; ...). We consider the periods of time in (29) as daily intervals, and they will be in the case history in Section 6, but the representation here is sufficiently general to represent other types of time intervals.

For the university population, we consider that the data gathered from the screening consist of counts,

$$y^{\mathrm{obs}}_{kj}, \quad k = 1, \dots, c, \quad j = 1, 2, \dots, \tag{31}$$

of the numbers of individuals who test positive from the cohorts $1, \dots, c$, indexed by k, at the times t_{k1}, t_{k2}, indexed by j.

4.2. Coarsened temporal aggregation

Data gathered from observed incidence counts and periodic screening can exhibit patterns on finer time scales due to features of, and variations in, testing practices [21]. For example, lower incidence counts may be observed on weekends, followed by higher values on Mondays. In addition, the infection and transmission dynamics can itself exhibit patterns, such as work contacts dominating on weekdays and family contacts dominating on weekends.

We circumvent this issue by fitting the hybrid stochastic model to the data in such a way that the fit between the model predictions and the data is evaluated on coarser time scales that smoothen out such patterns. Thus, for the surrounding community, we define aggregate counts

$$\tilde{y}^{\mathrm{obs}}_{\mathcal{I}_1}, \ \tilde{y}^{\mathrm{obs}}_{\mathcal{I}_2}, \ \dots, \ \tilde{y}^{\mathrm{obs}}_{\mathcal{I}_s} \tag{32}$$

that sum the raw counts,

$$\tilde{y}^{\mathrm{obs}}_{\mathcal{I}_i} = \sum_{[t_{l-1},t_l[\subset \mathcal{I}_i} \tilde{y}^{\mathrm{obs}}_{t_l}, \tag{33}$$

over longer periods of time

$$\mathcal{I}_i = [\tau_{i-1}, \tau_i[, \quad i = 1, \dots, s. \tag{34}$$

Likewise, for the university population, we thus define aggregate counts

$$y^{\mathrm{obs}}_{\mathcal{I}_1}, \ y^{\mathrm{obs}}_{\mathcal{I}_2}, \ \dots, \ y^{\mathrm{obs}}_{\mathcal{I}_s} \tag{35}$$

that sum the raw counts,

$$y^{\mathrm{obs}}_{\mathcal{I}_i} = \sum_{k=1}^{c} \sum_{t_{kj} \in \mathcal{I}_i} y^{\mathrm{obs}}_{kj}, \tag{36}$$

over the periods of time defined in (34). We set the periods of time in (34) to weekly intervals in the case history in Section 6. For later reference, we collect the aggregate counts into vectors

$$\tilde{\boldsymbol{y}}^{\mathrm{obs}} = (\tilde{y}^{\mathrm{obs}}_{\mathcal{I}_1}, \ \tilde{y}^{\mathrm{obs}}_{\mathcal{I}_2}, \ \dots, \ \tilde{y}^{\mathrm{obs}}_{\mathcal{I}_s}) \quad \text{and} \quad \boldsymbol{y}^{\mathrm{obs}} = (y^{\mathrm{obs}}_{\mathcal{I}_1}, \ y^{\mathrm{obs}}_{\mathcal{I}_2}, \ \dots, \ y^{\mathrm{obs}}_{\mathcal{I}_s}). \tag{37}$$

4.3. Observation model

Because the data for the surrounding community are not directly related to a quantity in the hybrid stochastic model, we need an observation model to relate the observed variables to the evolution of the compartment counts. We consider that the incidences over the periods of time defined in (34) are linked to the evolution of the compartment counts as follows:

$$\underline{\tilde{y}}_{\mathcal{I}_i} = \int_{\tau_{i-1}}^{\tau_i} f_{\tilde{e}\to\tilde{i}}\big(\boldsymbol{x}(t)\big)\,dt = \int_{\tau_{i-1}}^{\tau_i} \gamma \tilde{e}(t)\,dt, \quad i = 1, \dots, s. \tag{38}$$

Alternatively, the incidences could be taken as a similar time integral of the transition rate from the susceptible to the exposed compartment. Only a fraction of the incidences are observed, and we consider this fraction to vary about a case detection rate of ρ, a parameter of the observation model. We model this variation in the observation process using a Poisson probability distribution. For a Poisson probability distribution, the variance is equal to the mean, so that variations are represented to be relatively more significant for smaller counts. Alternatively, other probability distributions could be used, such as a negative binomial probability distribution which introduces an overdispersion parameter that allows the variance to be larger than the mean [22]. Thus, assuming statistical independence of the variability in the observation process between observation intervals, the observed incidence counts are modeled to be distributed as

$$\tilde{\boldsymbol{Y}} | \boldsymbol{x} \sim \pi\big(\tilde{\boldsymbol{y}} \big| \boldsymbol{x}\big) = \prod_{i=1}^{s} \pi_{\mathrm{P}}\left(\tilde{y}_i; \ \rho \int_{\tau_{i-1}}^{\tau_i} \gamma \tilde{e}(t)\,dt\right), \tag{39}$$

in which $\pi_{\mathrm{P}}(\cdot;\lambda)$ denotes the Poisson probability mass function with mean λ. For the university population we have

$$\boldsymbol{Y} = \left(\sum_{k=1}^{c} \sum_{t_{kj} \in \mathcal{I}_1} (Y^{\mathrm{p}}_{kj} + Y^{\mathrm{a}}_{kj}), \dots, \sum_{k=1}^{c} \sum_{t_{kj} \in \mathcal{I}_s} (Y^{\mathrm{p}}_{kj} + Y^{\mathrm{a}}_{kj}) \right). \tag{40}$$

4.4. Bayesian posterior

A Bayesian identification begins with specifying for the unknown parameters a prior probability distribution that represents the state of knowledge about them before collecting the data. Then, the prior is updated by accounting for the data using a likelihood function in accordance with Bayes's formula. The result is a posterior probability distribution that represents the post-data state of knowledge of the plausibility of parameter values, as it arises from all the available information. Let us collect in a vector θ the parameters that must be identified. For a given set of parameter values and given initial conditions, we can use the hybrid stochastic model to generate sample paths of the compartment and screening counts. In turn, given a sample path of the compartment counts, we can use the observation model to generate samples of the observed incidence counts for the surrounding community. Thus, given a value θ of the parameters, the hybrid stochastic model, completed with the initial condition, and together with the observation model, determines a probability mass function for the aggregated counts, which we denote by $\pi_{(\tilde{\boldsymbol{Y}},\boldsymbol{Y})}(\cdot,\cdot|\theta)$. With this probability mass function, the Bayesian posterior is obtained as

$$\underbrace{\rho\big(\theta\big|\tilde{\boldsymbol{y}}^{\text{obs}},\boldsymbol{y}^{\text{obs}}\big)}_{\text{posterior}} \propto \underbrace{\pi_{(\tilde{\boldsymbol{Y}},\boldsymbol{Y})}(\tilde{\boldsymbol{y}}^{\text{obs}},\boldsymbol{y}^{\text{obs}}|\theta)}_{\text{likelihood}} \underbrace{\rho(\theta)}_{\text{prior}}. \tag{41}$$

5. Implementation of the Bayesian inference

In problems of Bayesian inference, the exploration of the posterior is implemented most often by using a Markov chain Monte Carlo method. However, Markov chain Monte Carlo methods require explicit evaluations of the likelihood function. In the problem under study, an expression for the likelihood function is not immediately available because the system is only partially observed. Here, as an alternative approach, we adopt for the exploration of the posterior a machine-learning simulation-based inference method that does not require explicit likelihood evaluations. This method has been recently introduced in [23], with related work, for instance, in [14,24–27]. Thus, we formulate the model inference as a Bayesian inference involving a likelihood function, but, in the subsequent numerical implementation of this Bayesian inference, we use a machine-learning simulation-based inference method that does not require explicit evaluations of this likelihood function. We describe in Appendix A in the Supplementary material how we implemented the exploration of the Bayesian posterior with a machine-learning simulation-based inference approach. Please note that the results described in the next section can be understood without reading these implementation details in Appendix A in the Supplementary material.

6. Application to massive COVID-19 screening at University of Liège

6.1. Formulation of the identification problem

We set up the model and the identification problem as follows:

- We used a population size of $\tilde{n} = 197\,355$ for the surrounding community (reflecting a recent population count for the city of Liège) and a population size of $n = 28\,371$ for the university population (reflecting the number of students, faculty members, and staff members for the combination of the university's city center and Sart-Tilman campuses).
- We simulated the period of 6 weeks from the students returning to the university on Monday 14 September (the beginning of week 38) until the university adopting the *red* sanitary code and transitioning to fully online lectures (the end of week 43) (please see also Fig. 1).
- We used for the parameters γ, $\tilde{\omega}$, ω_{p}, ω_{s}, ω_{a}, and f_{s}, characterizing the dynamics of the disease and states of infectiousness, the values in Table 2. We subdivided the compartments of infectious individuals into $\tilde{m} = 16$, $m_{\text{p}} = 8$, $m_{\text{s}} = 8$ and $m_{\text{a}} = 16$ subcompartments.
- We used for the case detection rate ρ for the surrounding community and the parameters τ, s_{e} and d characterizing the screening test program for the university population the values in Table 2. Whereas these parameters were constant (independent of time) in our simulations, we let f_{p}, the fraction of individuals participating in the screening test program, depend on time in our simulations. As per the screening test data in Table 1, we used constant values of 0, 0, 22.76%, 27.41%, 37.06%, and 29.21% for weeks 38, 39, 40, 41, 42, and 43, respectively. In the model, we used $c = 5$ cohorts of equal size, hence, $f_1 = \cdots = f_c = 1/5$, with the first cohort taking a weekly test on Monday, and so forth, until the fifth cohort taking a weekly test on Friday.
- We initialized the hybrid stochastic model with the compartment counts

$$\begin{aligned}&\Big(\tilde{s}(0),\ \tilde{e}(0),\ \tilde{i}_1(0),\dots,\tilde{i}_{\tilde{m}}(0),\ \tilde{r}(0)\Big)\\ &\quad=\Big(\tilde{n}-\tilde{f}_{\text{i}}\,\tilde{n}-\tilde{f}_{\text{i}}\,\frac{\tilde{\omega}}{\gamma}\,\tilde{n},\ \tilde{f}_{\text{i}}\,\frac{\tilde{\omega}}{\gamma}\,\tilde{n},\ \frac{1}{\tilde{m}}\tilde{f}_{\text{i}}\,\tilde{n},\ \dots,\ \frac{1}{\tilde{m}}\tilde{f}_{\text{i}}\,\tilde{n},\ 0\Big)\end{aligned} \tag{42}$$

for the surrounding community and

$$\begin{aligned}&\Big(\ s(0),e(0),\dots,i^{\text{a}}_{11}(0),\dots,i^{\text{a}}_{1m_{\text{a}}}(0),\\ &\qquad\dots,i^{\text{a}}_{c1}(0),\dots,i^{\text{a}}_{cm_{\text{a}}}(0)\dots,r(0)\ \Big)\\ &=\Big(\ n-f_{\text{i}}\,n-f_{\text{i}}\,\frac{\omega_{\text{a}}}{\gamma}\,n,\ f_{\text{i}}\,\frac{\omega_{\text{a}}}{\gamma}\,n,\ \dots,\ \frac{1}{c}\,\frac{1}{m_{\text{a}}}f_{\text{i}}\,n,\ \dots,\ \frac{1}{c}\,\frac{1}{m_{\text{a}}}f_{\text{i}}\,n,\\ &\qquad\dots,\ \frac{1}{c}\,\frac{1}{m_{\text{a}}}f_{\text{i}}\,n,\ \dots,\ \frac{1}{c}\,\frac{1}{m_{\text{a}}}f_{\text{i}}\,n,\ \dots,\ 0\ \Big)\end{aligned} \tag{43}$$

for the university population, thus assuming negligible immunity at the initial time. The parameters $\tilde{f}_{\text{i}}$ and f_{i} characterize the fractions of individuals infectious in the surrounding community and asymptomatically infectious in the university population at the initial time.
- We applied the Bayesian identification method described in Sections 4 and 5 to infer the parameters $\theta = (\tilde{f}_{\text{i}}, f_{\text{i}}, \beta_{00}, \beta_{11}, \beta_{10})$, that is, the fractions of individuals infectious in the surrounding community and asymptomatically infectious in the university population at the initial time and the transmission rates within and between the surrounding community and the university population. We took β_{00}, β_{11}, and β_{10} to be constant (independent of time) in our simulations.
- As data, we used as observed incidence counts $\tilde{y}^{\text{obs}}_{I_i}$ for the surrounding community the weekly aggregated observed incidence counts of 236, 485, 986, and 2132 for the city of Liège for weeks 39, 40, 41, and 42, respectively, and we used as weekly aggregated counts $y^{\text{obs}}_{I_i}$ of asymptomatic infectious individuals identified by the screening test program for the university population weekly aggregated counts of 0, 0, 110, 248, 490, 573 for the city center and Sart-Tilman campuses for weeks 38, 39, 40, 41, 42, 43, respectively (as per Table 1). We did not use the weekly aggregated observed incidence count for week 38 for the city of Liège because students had only just returned and it may thus not yet fully reflect the interaction of the university population and the surrounding community, and we did not use the weekly aggregated observed incidence count for week 43 due to the testing policy change.
- We used uniform prior probability distributions, with lower bounds of 0 and upper bounds sufficiently large so that they do not significantly influence the posterior probability distribution. This choice corresponds to a use of non-informative prior probability distributions, reflecting that no information other than the data is taken into account. Specifically, we used uniform priors restricted to $[0, 0.032]$ for $\tilde{f}_{\text{i}}$, $[0, 0.064]$ for f_{i}, $[0, 0.4]\,\text{day}^{-1}$ for β_{00}, $[0, 0.6]\,\text{day}^{-1}$ for β_{11}, and $[0, 0.8]\ \text{day}^{-1}$ for β_{10}.

Table 2
Formulation of identification problem: List of model parameters.

	Symbol	Description	Value	Source
Surrounding community	$\tilde{n}$	Population size	197 355	
	$\tilde{\gamma}^{-1}$	Characteristic duration of latency period	2 days	2 days in [28]
	$\tilde{\omega}^{-1}$	Characteristic duration of infectious period	7 days	4.5, 5–7, 9.3 days in [28]
	$\tilde{\rho}$	Case detection rate	30%	Estimate
University population	n	Population size	28 371	
	γ^{-1}	Characteristic duration of latency period	2 days	2 days in [28]
	ω_{p}^{-1}	Characteristic duration of pre-symptomatically infectious period for infected individuals who develop symptoms	3.5 days	3.2 days in [28]
	ω_{s}^{-1}	Characteristic duration of symptomatically infectious period for infected individuals who develop symptoms	3.5 days	3.5 days in [28]
	ω_{a}^{-1}	Characteristic duration of asymptomatically infectious period for infected individuals who do not develop symptoms	7 days	5, 6 days in [28]
	f_{s}	Fraction that develops symptoms and isolates spontaneously	30%	Assumption
	f_{p}	Fraction of individuals participating in the screening test program	See Table 1	From screening data
	τ	Duration between consecutive tests	7 days	
	s_{e}	Test sensitivity	85%	Estimate
	d	Diagnosis delay	1.5 days	Estimate

- In the numerical time integration of the stochastic dynamics, we used a time step of $\Delta t = 0.25\,\mathrm{day}$.

6.2. Bayesian posterior

We will begin the discussion of the results with discussing the identification of parameters $\tilde{f}_{\mathrm{i}}$, f_{i}, β_{00}, β_{11}, β_{10} from the data. Then, in the next sections, we will use the identified model to provide some insight into the hybrid stochastic modeling and interpretations for the case history.

To sample from the Bayesian posterior, we applied the machine-learning simulation-based inference method described in Appendix A in the Supplementary material. With reference to the implementation details provided in Appendix A in the Supplementary material, we obtained the results to follow with $\mu = 1\,000\,000$ training samples, the `sbi` toolbox's default architecture for the normalizing flow using piecewise polynomial transformations, and the `sbi` toolbox's default architecture for the deep neural network; once the training complete, we used the normalizing flow representation to draw $\nu = 100\,000$ samples from the Bayesian posterior.

From among the parameter samples drawn from the posterior, the most plausible parameter sample is the maximum-a-posteriori estimate. We found

$$\theta^{\mathrm{MAP}} = \big(\, \tilde{f}_{\mathrm{i}}^{\mathrm{MAP}} = 0.00052,\, f_{\mathrm{i}}^{\mathrm{MAP}} = 0.0057,\\ \beta_{00}^{\mathrm{MAP}} = 0.226\,\mathrm{day}^{-1},\, \beta_{11}^{\mathrm{MAP}} = 0.243\,\mathrm{day}^{-1},\, \beta_{10}^{\mathrm{MAP}} = 0.10\,\mathrm{day}^{-1} \,\big)\,. \tag{44}$$

The parameter samples drawn from the posterior provide not only this optimal parameter value, but also insight into the extent to which, even though less likely, other parameter values are still plausible and still fit the data well. Thus, the Bayesian identification provides insight into the extent to which the parameters are constrained by the data, that is, the extent to which uncertainty still remains in the parameter values after having accounted for the data. To convey such insight, Fig. 4 shows the one- and two-dimensional density plots of the parameter samples drawn from the posterior.

It can be observed that the parameter values carry significant uncertainty. For the transmission rates β_{00}, β_{11}, and β_{10}, the posterior is not concentrated specifically about the maximum-a-posteriori estimate, but spread out in a narrow region of the parameter space, including the maximum-a-posteriori estimate, with a clear dependence structure. This observation reflects a compensation effect: based on the available data, it is not possible to uniquely ascertain the rate β_{10} of the transmission between the university population and the surrounding community as compared with the rates β_{11} and β_{00} of the transmission within the university population and the surrounding community. Instead, the available data allow a range of parameter values to be identified that ranges from parameter values that combine smaller transmission rates within the university population and the surrounding community with larger transmission rates between them to values of the parameters that combine larger transmission rates within the university population and the surrounding community with smaller transmission rates between them.

It can also be observed that the parameter samples tend to combine a larger value of f_{i}, the initial fraction of asymptomatically infectious individuals in the university population, with a smaller value of $\tilde{f}_{\mathrm{i}}$, the initial fraction of individuals infectious in the surrounding community. Thus, it is considered more consistent with the data that the prevalence in the university population, predominantly students, was higher than the prevalence in the surrounding community at the beginning of the Fall term.

Fig. 5 compares the data with the corresponding predictions of the identified model. We ran the hybrid deterministic model using the best estimate of the parameter values, that is, the maximum-a-posteriori estimate in (44), and we show the predictions using the rounded markers connected with the thick solid lines. We also ran for each parameter sample drawn from the posterior a sample of the hybrid stochastic model, using, for each parameter sample, a new, independent, sample path of the Wiener process and new, independent, samples from the binomial distributions. Based on these simulations, we determined the 2.5% and 97.5% percentiles for the predictions, which we show as the darkly shaded regions. Thus, the darkly shaded regions quantify the combined impact of the parameter uncertainty, the variability in the epidemic process, and the variability in the case detection efficacy of the screening, on the predictions. For the surrounding community, the lightly shaded region adds the impact of the uncertainty contributed by the variability in the observation process. The lightly shaded and darkly shaded regions largely overlap, thus indicating that the added uncertainty contributed by the variability in the observation process is small compared with the other uncertainty contributions. As a manifestation of the aforementioned compensation effect, the uncertainty ranges are relatively smaller for the predictions (Fig. 5) than for the parameter values (Fig. 4).

6.3. Some insight into the hybrid stochastic modeling

We computed the evolution of the number of non-isolated pre-symptomatically and asymptomatically infectious individuals in the first cohort, that is, among those who, if they participate, take a weekly test on Mondays (Fig. 6). As in Fig. 5, the thick solid line was

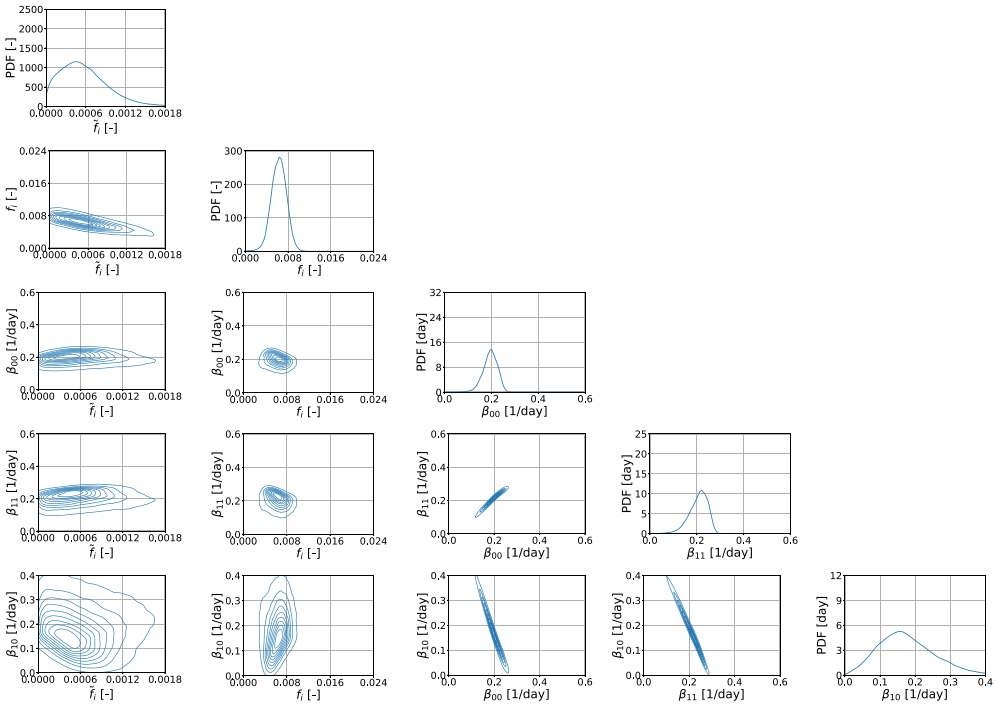
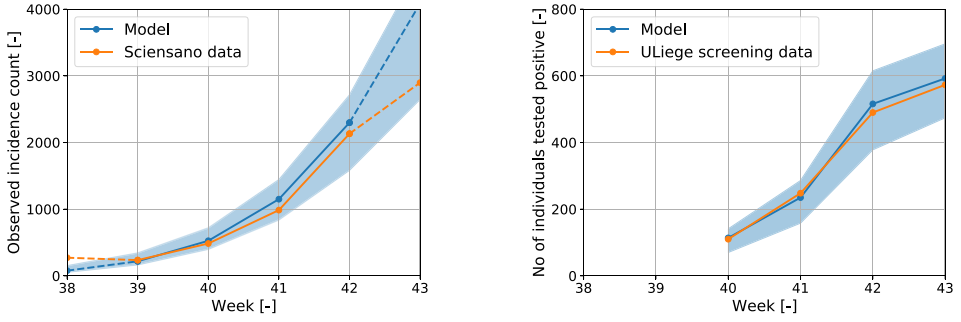


Fig. 4. Bayesian posterior: One- and two-dimensional density plots of the parameter samples drawn from the posterior.



(a) Surrounding community. (b) University population.

Fig. 5. Bayesian posterior: Comparison of the observed counts (orange rounded markers) with the predicted counts obtained with the identified model (best estimate provided by the hybrid deterministic model in blue rounded markers, uncertainty margins provided by the hybrid stochastic model in blue shaded regions; please see also the text in Section 6.2 for more details) for (a) the surrounding community and (b) the university population. The dashed lines serve to recall that the identification does not take into account the observed incidence counts for the city of Liège for weeks 38 and 43.

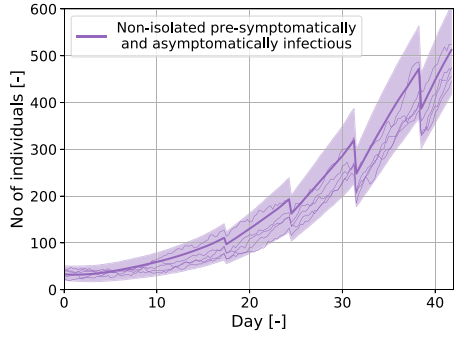


Fig. 6. Some insight into the hybrid stochastic modeling: Evolution of the number of non-isolated pre-symptomatically and asymptomatically infectious individuals in the first cohort, that is, among the individuals, who, provided that they participate, take a weekly test on Mondays (best estimate provided by the hybrid deterministic model in thick solid line, a few sample paths provided by the hybrid stochastic model in thin solid lines, and uncertainty margins provided by the hybrid stochastic model in shaded region; please see also the text in Section 6.3 for more details).

obtained with the hybrid deterministic model using the maximum-a-posteriori estimate in (44) and is the best estimate; the shaded region was obtained with the hybrid stochastic model and quantifies the combined impact of the parameter uncertainty, the variability in the epidemic process, and the variability in the case detection efficacy of the screening; and the thin solid lines are a few sample paths simulated with the hybrid stochastic model. This figure highlights that the hybrid stochastic model represents the evolution of the infectious disease and its transmission as a continuous-time stochastic dynamics, with the screening periodically resulting in a number of infectious individuals being removed.

6.4. *Illustration of the identified model informing a retrospective study*

We used the identified model to simulate aspects of the transmission dynamics that were not amenable to direct observation.

We computed the daily number of new infections in the university population due to contacts within the university population, the daily number of new infections in the university population due to

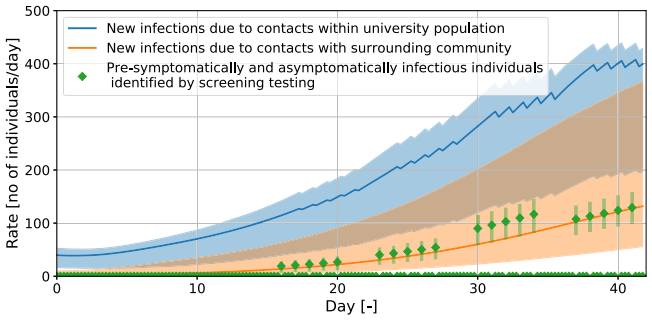


Fig. 7. Illustration of the identified model informing a retrospective study: Daily number of new infections in the university population due to contacts within the university population (blue), daily number of new infections in the university population due to contacts with the surrounding community (orange), and daily number of individuals who test positive in the screening test program (green) predicted by the hybrid stochastic model (the thick solid lines and the diamond markers are the best estimate provided by the hybrid deterministic model; the shaded regions are the uncertainty margins provided by the hybrid stochastic model; please note that the green diamond markers and shaded regions are predictions and not the observations; please see also the text in Section 6.4).

contacts with the surrounding community, and we compared these daily numbers with the daily number of individuals predicted to test positive in the screening test program (Fig. 7). Here too, the thick solid lines represent the best estimate, and the shaded regions quantify the combined impact of the parameter uncertainty, the variability in the epidemic process, and the variability in the case detection efficacy of the screening. The results suggest that exogeneous infections due to contacts with the surrounding community may have been a significant factor, even though, again, as a manifestation of the aforementioned compensation effect, the assessment of their significance is subject to significant uncertainty. For parameter samples that combine a larger value of the rate β_{10} of the transmission between the university population and the surrounding community with smaller values of the rates β_{11} and β_{00} of the transmission within the university population and the surrounding community, the number of new infections due to contacts within the university population is smaller, and the number of new infections due to contacts with the surrounding community is larger.

The results suggest that the screening test program had an important and valuable impact, having allowed a significant number of asymptomatically infectious individuals to be identified and isolated and thus prevent further transmission. However, the results indicate that the daily number of infectious individuals identified by the screening test program was significantly smaller than the daily number of new infections, thus leading to the assessment that the transmission rates were too high and the participation rate of about 30% and the test frequency of once-a-week tests not enough to contain the spread.

We also used the identified model to simulate two hypothetical scenarios in which the initial condition and the transmission rates were those that we identified, but we used different values for the participation rate, the test frequency, and the diagnosis delay. The use of the identified transmission rates represents, in the hybrid stochastic model, contact patterns and transmissibility similar to those present in the period in which the data were collected. In the first hypothetical scenario, there is no screening testing (Fig. 8(a)). The comparison of the results obtained for the real scenario (Fig. 8(b)) with those for this hypothetical scenario (Fig. 8(a)) leads to the estimate that the screening testing allowed about 628 infections to be avoided. In the second hypothetical scenario, the participation rate is higher (90%), the test frequency higher (twice-a-week tests), and the diagnosis delay shorter (1 day), and, rather than after two weeks as in the real scenario, the screening testing starts at the beginning of the semester (Fig. 8(c)). The comparison of the results obtained for the hypothetical scenario without screening testing (Fig. 8(a)) with those for this hypothetical scenario with intensified screening testing (Fig. 8(c)) leads to the estimate that with the higher participation rate, the higher test frequency of twice-a-week tests, the shorter diagnosis delay, and the start at the beginning of the semester, the screening test program might have allowed significantly more, namely, about 7597 infections to be avoided, thus suggesting key importance of a high participation rate, a high test frequency, and a short diagnosis delay.

7. Discussion of limitations and hypotheses

The proposed model, the proposed Bayesian identification, and their application to the massive COVID-19 screening at the University of Liège involve simplifications and hypotheses. Here, we will discuss some of the main simplifications and hypotheses.

7.1. Imperfect compliance with isolation in the university population

The model does not account explicitly for potentially imperfect compliance with isolation in the university population.

To address this limitation, the model can be enriched. Such a model enrichment can be achieved by modifying the model by replacing the equations in the first two lines in (1) with

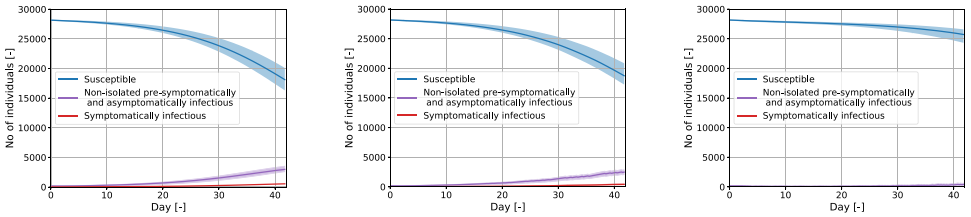
$$\frac{d\tilde{s}}{dt} = -\beta_{00}\,\tilde{s}\,\frac{\tilde{i}_1+\cdots+\tilde{i}_{\tilde{m}}}{\tilde{n}} - \beta_{01}\,\tilde{s}\,\Bigg(\frac{i^{\mathrm{p}}_{11}+\cdots+i^{\mathrm{p}}_{cm_{\mathrm{p}}}+w^{\mathrm{p}}_{11}+\cdots+w^{\mathrm{p}}_{cm_{\mathrm{p}}}}{n} + \frac{i^{\mathrm{a}}_{11}+\cdots+i^{\mathrm{a}}_{cm_{\mathrm{a}}}+w^{\mathrm{a}}_{11}+\cdots+w^{\mathrm{a}}_{cm_{\mathrm{a}}}}{n}\Bigg) - f_{\mathrm{ic}}\,\beta_{01}\,\tilde{s}\,\Bigg(\frac{i^{\mathrm{s}}_{1}+\cdots+i^{\mathrm{s}}_{m_{\mathrm{s}}}}{n} + \frac{q^{\mathrm{p}}_{11}+\cdots+q^{\mathrm{s}}_{cm_{\mathrm{p}}}}{n} + \frac{q^{\mathrm{a}}_{11}+\cdots+q^{\mathrm{a}}_{cm_{\mathrm{a}}}}{n}\Bigg), \tag{45}$$

$$\frac{d\tilde{e}}{dt} = \beta_{00}\,\tilde{s}\,\frac{\tilde{i}_1+\cdots+\tilde{i}_{\tilde{m}}}{\tilde{n}} + \beta_{01}\,\tilde{s}\,\Bigg(\frac{i^{\mathrm{p}}_{11}+\cdots+i^{\mathrm{p}}_{cm_{\mathrm{p}}}+w^{\mathrm{p}}_{11}+\cdots+w^{\mathrm{p}}_{cm_{\mathrm{p}}}}{n} + \frac{i^{\mathrm{a}}_{11}+\cdots+i^{\mathrm{a}}_{cm_{\mathrm{a}}}+w^{\mathrm{a}}_{11}+\cdots+w^{\mathrm{a}}_{cm_{\mathrm{a}}}}{n}\Bigg) + f_{\mathrm{ic}}\,\beta_{01}\,\tilde{s}\,\Bigg(\frac{i^{\mathrm{s}}_{1}+\cdots+i^{\mathrm{s}}_{m_{\mathrm{s}}}}{n} + \frac{q^{\mathrm{p}}_{11}+\cdots+q^{\mathrm{s}}_{cm_{\mathrm{p}}}}{n} + \frac{q^{\mathrm{a}}_{11}+\cdots+q^{\mathrm{a}}_{cm_{\mathrm{a}}}}{n}\Bigg) - \gamma\tilde{e} \tag{46}$$

and by modifying the equations in the first two lines in (2) analogously. Thus, in this modified model, instead of not contributing to transmission, symptomatically infectious individuals and isolated pre-symptomatically and asymptomatically infectious individuals have a reduced contribution to transmission, with a significance controlled by a compliance factor f_{ic}. We set $f_{\mathrm{ic}} = 20\%$. By comparing results obtained with the modified model (Fig. 9) with the corresponding results in Section 6 (Figs. 4 and 8(b)), we can observe that accounting for imperfect compliance with isolation in the university population mainly results in a lower value being identified for the transmission rate β_{10}, because less exogenous introductions are needed to explain the growth in infections when more (now also symptomatic and isolated pre-symptomatic and asymptomatic) individuals contribute to transmission and the imperfect compliance renders the screening test program less impactful.

7.2. Other limitations and hypotheses

- The interplay of the university population with the broader community is complex. The choice of which population size to use for the surrounding community and which data to use as observed incidence counts for it in the Bayesian identification is difficult (city, agglomeration, province, ...). We chose the population size of the surrounding community after the one of the city of Liège, and we used as observed incidence counts the ones reported by Belgium's Public Health Institute Sciensano for the city of Liège. Our choice was based on the consideration that many students live in student housing in the city of Liège.



(a) Hypothetical scenario without screening testing.

(b) Real participation rates of 0, 0, 22.76 %, 27.41 %, 37.06 %, and 29.21 % and test frequencies of 0, 0, $1/7\,\mathrm{day}^{-1}$, $1/7\,\mathrm{day}^{-1}$, $1/7\,\mathrm{day}^{-1}$, and $1/7\,\mathrm{day}^{-1}$ for weeks 38, 39, 40, 41, 42, and 43, respectively, and diagnosis delay of 1.5 days.

(c) Hypothetical scenario with a participation rate of 90 %, a test frequency of $2/7\,\mathrm{day}^{-1}$ and a diagnosis delay of 1 day for the entire period under study from week 38 to week 43.

Fig. 8. Illustration of the identified model informing a retrospective study: Evolution of the number of individuals who remain susceptible (blue), are non-isolated pre-symptomatically or asymptomatically infectious (purple), and are symptomatically infectious (red) in the university population in (a) a hypothetical scenario without screening testing, (b) the real scenario, and (c) a hypothetical scenario with a higher participation rate of 90%, a higher test frequency of $2/7\,\mathrm{day}^{-1}$, and a shorter diagnosis delay of 1 day (best estimate provided by the hybrid deterministic model in thick solid lines and uncertainty margins provided by the hybrid stochastic model in shaded regions; please see also the text in Section 6.4 for more details).

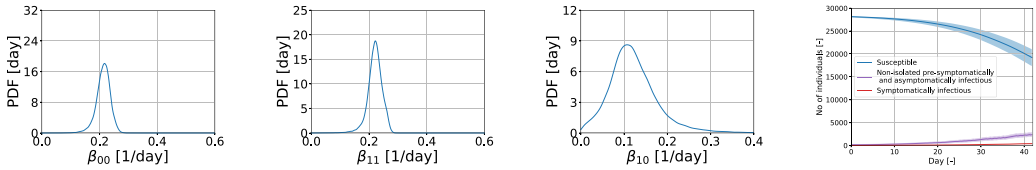


Fig. 9. Discussion of limitations and hypotheses: Results obtained with the model modified as in (45)–(46) to provide some insight into the impact of accounting explicitly for imperfect compliance with isolation in the university population.

- The way we formulated the model and its Bayesian identification entails an assumption of statistical homogeneity. Loosely speaking, the formulation will tend to extrapolate from the positivity rate observed among the individuals who participated in the screening test program that those who did not participate were positive with a similar rate. However, this assumption may not be adequate. For example, individuals who did not participate in the on-campus screening test program may have isolated themselves more stringently and been less exposed, so that the identified model may overestimate the number of infections.
- In the case history, we took the transmission rates to be constant (independent of time). In doing so, we sought a model able to explain the data with a minimal number of parameters (there were 5 parameters to be identified). However, the transmission rates may have changed over time, notably towards the end of the period under study, as more interventions were taken.
- In the case history, we used deterministic, fixed values for the parameters characterizing the dynamics of the disease and the states of infectiousness and for the parameters characterizing the screening test. Although their values were selected with reference to the state of the art, these parameters carry their own uncertainty, whose impact adds to the uncertainty ranges that we obtained for the parameters and the predictions. Uncertainty in the parameter f_s, which describes for the university population the fraction of infected individuals who develop symptoms, and uncertainty in the sensitivity s_e of the screening test are particularly notable. If the sensitivity of the screening test was higher than what we assumed (s_e = 85%), the identified model would overestimate the number of infected individuals, since a smaller total number of infections would have sufficed to explain the numbers of asymptotically infectious individuals who tested positive.
- In the case history, because the screening test program was set up in a way whereby testing kits were retrieved and returned anonymously, it is possible that some tests may not have been used for their intended purpose. Individuals who tested positive may have used a testing kit of an acquaintance a short while later to know whether they still tested positive. And weakly symptomatic individuals from the surrounding community suspecting infection may have used a testing kit of an acquaintance from the university population to know whether they were infected. Due to such possible unintended use of the screening tests, the identified model may overestimate the number of infections.
- In the case history, we found that owing to a compensation effect, the transmission rates, inferred only from the screening test data and observed incidence counts, carry significant uncertainty (data limitation). Information from contact surveys, demographic and other social data could help constrain the parameter values. Such information could be integrated through the prior in the Bayesian identification.
- The transmission rates identified in the case history are relevant to the restrictions and the variants prevalent at the time the data was collected. Transmission rates are likely to change in accordance with the restrictions and the prevalent variants.

8. Conclusion

We proposed a hybrid stochastic model for infectious disease transmission in a university campus with periodic screening situated within

a larger surrounding community. This hybrid stochastic model represents the evolution of the infectious disease and its transmission using continuous-time stochastic dynamics, and it represents the screening testing as discrete stochastic events. The stochastic model of these discrete stochastic events takes into account the number of individuals who participate in the screening test program, the frequency of the periodic screening, and the properties of the test, including the sensitivity and the diagnosis delay. We proposed a Bayesian methodology for the identification of parameters of the hybrid stochastic model, along with a characterization of their uncertainty, from the screening test data for the university population and observed incidence counts for the surrounding community. We applied the proposed methodology to a massive COVID-19 screening conducted at the University of Liège in Fall 2020. The application in the retrospective modeling study suggests that a high rate of participation and a high test frequency are important to reduce spread.

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 thank the members of the University of Liège's Risk Analysis Group, established in the context of the COVID-19 pandemic, for many relevant discussions and for providing the data. The first author would also like to acknowledge Romin Tomasetti for helpful discussions relevant to the machine-learning simulation-based inference approach.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.mbs.2022.108805.

References

[1] Interim guidance for use of pooling procedures in SARS-CoV-2 diagnostic, screening, and surveillance testing, 2020, https://www.cdc.gov/coronavirus/2019-ncov/lab/pooling-procedures.html, Accessed: 2020-02-06.

[2] Testing, screening, and outbreak response for institutions of higher education (IHEs), 2020, https://www.cdc.gov/coronavirus/2019-ncov/community/colleges-universities/ihe-testing.html, Accessed: 2020-02-06.

[3] A. Paltiel, A. Zheng, R. Walensky, Assessment of SARS-CoV-2 screening strategies to permit the safe reopening of college campuses in the United States, JAMA Netw. Open 3 (7) (2020) e2016818, http://dx.doi.org/10.1001/jamanetworkopen.2020.16818.

[4] E. Hill, B. Atkins, M. Keeling, M. Tildesley, L. Dyson, Modelling SARS-CoV-2 transmission in a UK university setting, Epidemics 36 (2021) 100476, http://dx.doi.org/10.1016/j.epidem.2021.100476.

[5] P. Gressman, J. Peck, Simulating COVID-19 in a university environment, Math. Biosci. 328 (2020) 108436, http://dx.doi.org/10.1016/j.mbs.2020.108436.

[6] D. Larremore, B. Wilder, E. Lester, S. Shehata, J. Burke, J. Hay, M. Tambe, M. Mina, R. Parker, Test sensitivity is secondary to frequency and turnaround time for COVID-19 screening, Science 7 (1) (2021) eabd5393, http://dx.doi.org/10.1126/sciadv.abd5393.

[7] S. Sturniolo, W. Waites, T. Colbourn, D. Manheim, J. Panovska Griffiths, Testing, tracing and isolation in compartmental models, PLOS Comput. Biol. 17 (3) (2021) e1008633, http://dx.doi.org/10.1371/journal.pcbi.1008633.

[8] J. Enright, E. Hill, H. Stage, K. Bolton, E. Nixon, E. Fairbanks, M. Tang, E. Brooks-Pollock, L. Dyson, C. Budd, R. Hoyle, L. Schewe, J. Gog, M. Tildesley, SARS-CoV-2 infection in UK university students: lessons from September–December 2020 and modelling insights for future student return, R. Soc. Open Sci. 8 (8) (2021) 210310, http://dx.doi.org/10.1098/rsos.210310.

[9] V. Denoël, O. Bruyère, G. Louppe, F. Bureau, V. D'Orio, S. Fontaine, L. Gillet, M. Guillaume, E. Haubruge, A. Lange, F. Michel, R. Van Hulle, M. Arnst, A. Donneau, C. Saegerman, Decision-based interactive model to determine re-opening conditions of a large university campus in belgium during the first covid-19 wave, Arch. Public Health 80 (2022) http://dx.doi.org/10.1186/s13690-022-00801-w.

[10] T. Britton, E. Pardoux, Stochastic Epidemic Models with Inference, Springer, Cham, Switzerland, 2019.

[11] L. Allen, A primer on stochastic epidemic models: Formulation, numerical simulation, and analysis, Infect. Dis. Model. 2 (2017) 128–142, http://dx.doi.org/10.1016/j.idm.2017.03.001.

[12] A. Lloyd, Realistic distributions of infectious periods in epidemic models: Changing patterns of persistence and dynamics, Theor. Popul. Biol. 60 (1) (2001) 59–71, http://dx.doi.org/10.1006/tpbi.2001.1525.

[13] H. Wearing, P. Rohani, M. Keeling, Appropriate models for the management of infectious diseases, PLOS Med. 2 (7) (2005) e174, http://dx.doi.org/10.1371/journal.pmed.0020174.

[14] K. Cranmer, J. Brehmer, G. Louppe, The frontier of simulation-based inference, PNAS 117 (48) (2020) 30055–30062, http://dx.doi.org/10.1073/pnas.1912789117.

[15] J. Peto, Covid-19 mass testing facilities could end the epidemic rapidly, BMJ 368 (2020) m1163, http://dx.doi.org/10.1136/bmj.m1163.

[16] P. Libin, L. Willem, T. Verstraeten, A. Torneri, J. Vanderlocht, N. Hens, Assessing the feasibility and effectiveness of household-pooled universal testing to control COVID-19 epidemics, PLOS Comput. Biol. (2021) http://dx.doi.org/10.1371/journal.pcbi.1008688.

[17] E. Parzen, Stochastic Processes, Holden-Day, San Francisco, California, 1962.

[18] J. Norris, Markov Chains, Cambridge University Press, Cambridge, United Kingdom, 1997.

[19] B. Bercu, D. Chafaï, Modélisation Stochastique et Simulation, Dunod, Paris, France, 2007.

[20] E. Pardoux, Markov Processes and Applications, Dunod and John Wiley & Sons, Chichester, West Sussex, United Kingdom, 2008.

[21] A. Bergman, Y. Sella, P. Agre, A. Casadevall, Oscillations in U.S. COVID-19 incidence and mortality data reflect diagnostic and reporting factors, mSystems 5 (4) (2020) http://dx.doi.org/10.1128/mSystems.00544-20.

[22] J. Fintzi, J. Wakefield, V. Minin, A linear noise approximation for stochastic epidemic models fit to partially observed incidence counts, Biometrics (2021) http://dx.doi.org/10.1111/biom.13538.

[23] D. Greenberg, M. Nonnenmacher, J. Mackle, Automatic posterior transformation for likelihood-free inference, in: K. Chaudhuri, R. Salakhutdinov (Eds.), Proceedings of the 36th International Conference on Machine Learning, PMLR, 2019, pp. 2404–2414.

[24] D. Rezende, S. Mohamed, Automatic posterior transformation for likelihood-free inference, in: F. Bach, D. Blei (Eds.), Proceedings of the 32nd International Conference on Machine Learning, PMLR, 2015, pp. 530–1538.

[25] G. Papamakarios, E. Nalisnick, D. Rezende, S. Mohamed, B. Lakshminarayanan, Normalizing flows for probabilistic modeling and inference, J. Mach. Learn. Res. (2021) 1–64.

[26] P. Gonçalves, J. Lueckmann, M. Deistler, M. Nonnenmacher, K. Öcal, G. Bassetto, C. Chintaluri, W. Podlaskiand S. Haddad, T. Vogels, D. Greenberg, J. Macke, Training deep neural density estimators to identify mechanistic models of neural dynamics, eLife 9 (2020) e56261, http://dx.doi.org/10.7554/eLife.56261.

[27] J. Hermans, N. Banik, C. Weniger, G. Bertone, G. Louppe, Towards constraining warm dark matter with stellar streams through neural simulation-based inference, Mon. Not. R. Astron. Soc. 507 (2) (2021) 1999–2011, http://dx.doi.org/10.1093/mnras/stab2181.

[28] S. Abrams, J. Wambua, E. Santermans, L. Willem, E. Kuylen, P. Coletti, P. Libin, C. Faes, O. Petrof, S. Herzog, P. Beutels, N. Hens, Modeling the early phase of the belgian COVID-19 epidemic using a stochastic compartmental model and studying its implied future trajectories, Epidemics 35 (2021) 100449, http://dx.doi.org/10.1016/j.epidem.2021.100449.