Contents lists available at ScienceDirect

Science of the Total Environment

Effect of SARS-CoV-2 shedding rate distribution of individuals during their disease days on the estimation of the number of infected people. Application of wastewater-based epidemiology to the city of Thessaloniki, Greece

M. Kostoglou [a], M. Petala [b], Th. Karapantsios [a,*], Ch. Dovas [c], V. Tsiridis [b], E. Roilides [d], A. Koutsolioutsou-Benaki [e], D. Paraskevis [e,f], S. Metalidis [g], E. Stylianidis [h], A. Papa [i], A. Papadopoulos [j], S. Tsiodras [f], N. Papaioannou [c]

[a] Laboratory of Chemical and Environmental Technology, Dept. of Chemistry, Aristotle University of Thessaloniki, Thessaloniki 54124, Greece
[b] Laboratory of Environmental Engineering & Planning, Department. of Civil Engineering, Aristotle University of Thessaloniki, Thessaloniki 54 124, Greece
[c] Faculty of Veterinary Medicine, School of Health Sciences, Aristotle University of Thessaloniki, Thessaloniki 54124, Greece
[d] Infectious Diseases Unit and 3rd Department of Pediatrics, Aristotle University School of Health Sciences, Hippokration Hospital, Thessaloniki 54642, Greece
[e] Department of Environmental Health, Directory of Epidemiology and Prevention of Non Communicable Diseases and Injuries, National Public Health Organization, Athens, Greece
[f] National and Kapodistrian University of Athens, Athens, Greece
[g] Department of Haematology, First Department of Internal Medicine, Faculty of Medicine, AHEPA General Hospital, Aristotle University of Thessaloniki, Thessaloniki 54636, Greece
[h] School of Spatial Planning and Development, Faculty of Engineering, Aristotle University of Thessaloniki, 54124, Greece
[i] Department of Microbiology, Medical School, Aristotle University of Thessaloniki, Thessaloniki 54124, Greece
[j] EYATH S.A., Thessaloniki Water Supply and Sewerage Company S.A., Thessaloniki 54636, Greece

HIGHLIGHTS

- Leveraging SARS CoV-2 load in WW to evaluate Omicron variant spread in Thessaloniki.
- Impact of virus shedding rate distribution patterns in human stool to WBE models.
- Day of maximum shedding rate with respect to the initial infection day is critical.
- WBE data coupled with disease transmission data improve epidemiological models.

GRAPHICAL ABSTRACT
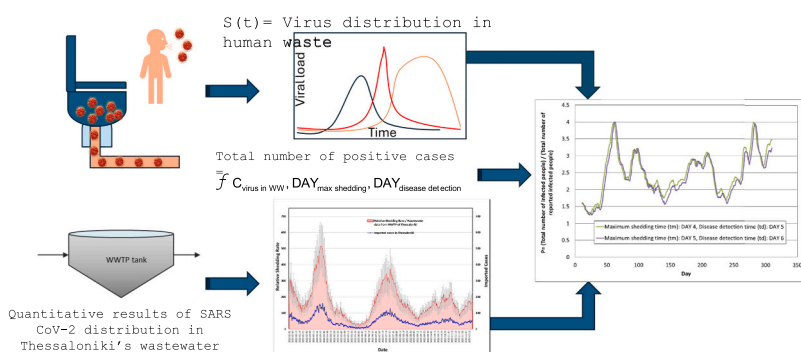
ARTICLE INFO

Editor: Lidia Minguez Alarcon

ABSTRACT

During the COVID-19 pandemic, wastewater-based epidemiology has proved to be an important tool for monitoring the spread of a disease in a population. Indeed, wastewater surveillance was successfully used as a

* Corresponding author.
  E-mail address: karapant@chem.auth.gr (Th. Karapantsios).

complementary approach to support public health monitoring schemes and decision-making policies. An essential feature for the estimation of a disease transmission using wastewater data is the distribution of viral shedding rate of individuals in their personal human wastes as a function of the days of their infection. Several candidate shapes for this function have been proposed in literature for SARS-CoV-2. The purpose of the present work is to explore the proposed function shapes and examine their significance on analyzing wastewater SARS-CoV-2 shedding rate data. For this purpose, a simple model is employed applying to medical surveillance and wastewater data of the city of Thessaloniki during a period of Omicron variant domination in 2022. The distribution shapes are normalized with respect to the total virus shedding and then their basic features are investigated. Detailed analysis reveals that the main parameter determining the results of the model is the difference between the day of maximum shedding rate and the day of infection reporting. Since the latter is not part of the distribution shape, the major feature of the distribution affecting the estimation of the number of infected people is the day of maximum shedding rate with respect to the initial infection day. On the contrary, the duration of shedding (total number of disease days) as well as the exact shape of the distribution are by far less important. The incorporation of such wastewater surveillance models in conventional epidemiological models - based on recorded disease transmission data- may improve predictions for disease spread during outbreaks.

1. Introduction

The recent outbreak of the COVID-19 epidemic brought in the foreground the subject of wastewater based epidemiology (Hegazy et al., 2022; Schill et al., 2023; Wade et al., 2022). The accurate quantification of the concentration of virus fragments in wastewater proved a useful tool, cheap and efficient, to measure the spread of the disease, complementary to conventional reporting of the medically tested infected individuals. The basic idea is that wastewater content of virus fragments mirrors the state of the whole population, symptomatic and asymptomatic infected people. This is an advantage compared to conventional medical surveillance, through rapid tests or PCR screening, of individuals since the latter capture only people who seek medical advice either because they have clinical symptoms or at least have serious suspicion of being infected. Typically, medical and wastewater data are analyzed and compared in scientific literature of wastewater based epidemiology. Apart from identifying persistent correlations between these two types of data, a yet more ambitious goal is to interrogate them in order to estimate the total number of infected people and, from that, even appraise the number of unreported infected individuals (those not registered by medical surveillance either because of having no or mild symptoms or lacking access to medical services). Success in such analysis can endorse the predictive power of wastewater data for the spread and evolution of a disease. Several types of mathematical models have been used in this context. A comprehensive review of the recent modeling works in wastewater based epidemiology can be found in (Ciannella et al., 2023). The interest here focuses on phenomenological models. The key in such models is the equation that relates the rate of shedding virus parts to wastewaters with the convolution integral of the number of infected people with age of infection τ (e.g., expressed in days) and the function S of individual shedding rate distributed on τ.

Let us examine some phenomenological models describing viral parts shedding to wastewater. In the study of Proverbio et al. (Proverbio et al., 2022) a typical SEIR epidemiological model is augmented by a compartment A accounting for the viral dynamics in a population. The model has been extended to include stochastic terms. The concentration of virus parts in wastewaters is assumed proportional to the value of A. The cumulative viral load in wastewater is estimated in the study of Phan et al. (Phan et al., 2023) by employing a simple SEI model and integrating in time the variable I multiplied by a set of constant coefficients. An innovative SI_NI_DR model has been developed in the study of Mattei et al. (Mattei et al., 2023) accounting separately for the number of detected I_D and not detected I_N infected individuals. The viral load in wastewater results through the typical convolution equation discussed above. The above models have been used to fit actual data. A more sophisticated approach to calculate wastewater virus shedding is part of the new age of infection models such as the model developed by Kostoglou et al. (Kostoglou et al., 2022). These models have a high degree of complexity requiring management of big data of individuals'

responses captured by distinct property distributions. Although the complexity restricts at present their applicability, these models represent a very promising tool to engage personalized responses to wastewater based epidemiology for public health decision making. In brief, each individual conveys a vector of properties related to the disease by the function S among others. In this case, the convolution integral can be expanded to include all the components of the characteristic vector. There are some cases in which the total number of infected people is known, that is, only reported infections exist (e.g., in a hospital or in a jail). In these cases, the data can be employed to estimate directly the function S (Cavany et al., 2022; Hoffmann and Alsing, 2021).

Typically, the function S is prescribed based on literature information and on observations, and parameters of multiplicative nature are determined by fitting the model to the data. These parameters contain also information on the decay of virus load in wastewater caused by physicochemical and biological parameters of wastewater (Proverbio et al., 2022). A systematic approach to account for this decay has been developed in literature (Kostoglou et al., 2021; Petala et al., 2021). The shortcoming of the proposed approach is that the corrected values of virus shedding are normalized by an unknown reference value. This problem has been overcome by an approach developed by (Petala et al., 2022) to find the number of unreported infected people in a population using directly the wastewater data, i.e., bypassing the use of a phenomenological model for the spread of the disease.

In this work, the method is slightly modified, and is applied to new data obtained at the city of Thessaloniki (Greece) -about 1,000,000 inhabitants, acquired in a period of 2022 where the omicron variant of SARS-CoV-2 was dominant in the city. It must be stressed that it is important to perform the analysis for a period of a single dominant variant of the virus since in other cases more than one forms of S(t) distributions must simultaneously be considered. In this context, a parametric analysis is performed focusing on the significance of the shape of function S(t) considering several shapes proposed in the relevant literature.

The structure of the present work is the following: At the next section, the proposed procedure of combining medical surveillance data with wastewater data to extract the ratio of unreported to reported infected people is described. Then, several shapes of distributions proposed in literature for S(τ) are reviewed and properly normalized. Finally, sensitivity of the analysis procedure with respect to the parameters of the distributions for a set of data from Thessaloniki during 2022 is presented and discussed.

2. Proposed procedure

An approach was proposed by (Petala et al., 2022) for cross-examining and exploiting medically reported infections data and wastewater virus concentration data in order to infer the evolution of the ratio of unreported to reported patients for a specific period of time.

The continuous form of the governing equations in case of disease detection at time t_d (number of days after the infection day) and disease duration at time t_e (days) is:

$$F(t) = \int_{t+t_d-t_e}^{t+t_d} f(x)dx \tag{1}$$

$$g(\tau,t) = \frac{f(t+t_d-\tau)U(t_e-\tau)}{F(t)} \tag{2}$$

$$R(t) = F(t) \int_0^{t_e} g(\tau,t)S(\tau)d\tau \tag{3}$$

where f(t) is the number of infections reported at time t, F(t) is the number of infected people at time t, $g(\tau,t)$ is the normalized distribution of infected people over time from the day of infection τ and R(t) is the normalized (see below)shedding rate of virus to the wastewater. The function U is the step function being unity for positive and zero for negative argument. The function $S(\tau)$ includes the dependence of viral shedding rate of an individual on the time since infection τ. The ratio of the virus shedding rates R(t) calculated from the number of reported infected people over the wastewater measured shedding rates $R_{exp}(t)$ allows estimation of the ratio of the reported infected people over the total (reported and unreported) infected people. The experimental data $R_{exp}(t)$ are normalized with respect to their values in a specific period in time and $S(\tau)$ is given in arbitrary units. Considering that the period chosen for normalization is a "quiet" period of low disease prevalence and as such there are approximately only reported infections, the ratio $R_{exp}(t)/R(t)$ normalized with its values in the reference period gives the ratio of total to reported patients (Petala et al., 2022). Since the reference period in that work referred to the existence of a different than omicron variant of the virus, a different procedure is followed here. At first, the ratio of normalized $R_{exp}(t)$ based on wastewater analysis to normalized R(t) based on detected infections is computed. Then the assumption is made that at the minimum value of this ratio there are no unreported infections. In this way, the evolution of the total to reported individuals been infected, is estimated. The uncertainty imported by this assumption can be straightforwardly assessed. If the ratio of total to reported cases at the time of minimum is a value larger than unity, this value must simply multiply the evolution in time of this ratio. The method used here is actually a simplification (regarding normalization of the results) of the method described in detail in Petala et al. (2022). The main difference is that several shapes of the distribution function S(τ) for the shedding rate are tested.

3. Proposed shapes for S(t)

3.1. General shapes

In what follows, the time (or age of infection) in functions with one parameter is denoted as t whereas in functions with two parameters t is the actual time and τ is the age of infection. Only in the Eqs. (2) and (3) the symbol τ is used for the age of infection. Several proposed in literature shapes of the function S(t) are presented here. The time t is shifted such as to take the value 0 at the time of onset of virus shedding at wastewater. An important aspect of the function S is the duration of virus shedding. There are three types of function S behavior with respect to duration time t_e. Type I behavior: The definition of t_e is necessary because the form of S is such that its integral up to infinity does not converge. Type II behavior: The definition of t_e is optional. In that case S(t) tends to zero as t increases in a way that allows integration of S(t) over time. Type III behavior: The value t_d is hardwired in the expression of S(t). Let us see one by one the available shapes of function S employed in literature.

Shape (a): This is a simple rational function having the form $S(t) = \frac{Ct}{b+t^2}$ (Proverbio et al., 2022). It follows type I behavior regarding t_e which means that the value of t_e must be defined together with the parameters C and b.

Shape (b): This is the shape of the well-known Gamma distribution $S(t) = Ct^\alpha e^{-bt}$ (Kostoglou et al., 2022). It follows type II behavior with respect to duration time which means that the definition of a value for t_e is optional. The parameters of the function are C, α, b and optionally t_e.

Shape (c): This shape is called double exponential. and it is given as $S(t) = C(1 - e^{-\alpha t})e^{-bt}$ (Miura et al., 2021). It follows type II behavior with respect to duration time. The parameters of the function are C, α, b and optionally t_e. It is noticed that this model is the only one of phenomenological nature based on a system of differential equations for the shedding rate. A detailed biological description can be found in (Teunis et al., 2015).

Shape (d): This shape consists also of two exponential functions (Petala et al., 2022). However, unlike shape (c) the two exponentials are defined in different domains of the independent variable t. The form of this distribution is.

$$S(t) = Ce^{-\alpha(t_m-t)} \quad 0 \leq t \leq t_m$$

$$S(t) = Ce^{-b(t-t_m)} \quad t_m \leq t \leq t_e$$

The parameter t_m denotes the time of occurrence of the maximum value of the distribution. This shape belongs to type II with respect to duration time. Its definition is optional which means a value of infinity t_e can be set in the above equation. The parameters of this shape are C, α, b, t_m and optionally t_e.

Shape (e): This is the standard Beta distribution (Wu et al., 2022). It has the form $S(t) = C\left(\frac{t}{t_e}\right)^\alpha \left(1 - \frac{t}{t_e}\right)^b$. It belongs to the type III behavior regarding duration time. It is clear that the duration type is part of the equation for the function S(t) and the parameter t_e must necessarily take a finite value. The parameters of this shape are C, α, b, t_e.

3.2. Normalization of S(t)

It is important to bring all the above proposed shapes of the function S(t) in a form which allows comparison to each other. The parameter C in all cases multiplies the shape, implying that it denotes in some way the intensity of the viral shedding process. However, this parameter has different quantitative meaning for each shape (even its units differs among shapes). To have a basis of comparison between shapes it is fruitful to renormalize the functions to bring them to the form S(t) = Ns(t) where N is the total amount of virus parts shed by an individual and s(t) has units of inverse time (i.e. days^{-1}). In practice the distribution s(t) can be computed by the following processing of S(t):

$$s(t) = \frac{S(t)}{\int_0^{t_e} S(\tau)d\tau} \tag{4}$$

After some algebra, the following relations can be found for the function s(t):

Shape (a):

$$s(t) = \frac{2t}{b+t^2} \frac{1}{ln\left(\frac{b+t_d^2}{b}\right)} \tag{5}$$

The parameters are b and t_e. The function s(t) appears in Fig. 1a for four combinations of b and t_e. In particular, these combinations are $t_e = 10$, b = 1; $t_e = 10$, b = 5; $t_e = 20$, b = 1; $t_e = 20$, b = 5. The effect of t_e is obvious in the figure. The effect of b is to make the distribution broader as b increases. The sharp cut off imposed by t_e can be clearly observed. The integral in the denominator of Eq. (4) does not converge as t_e goes to
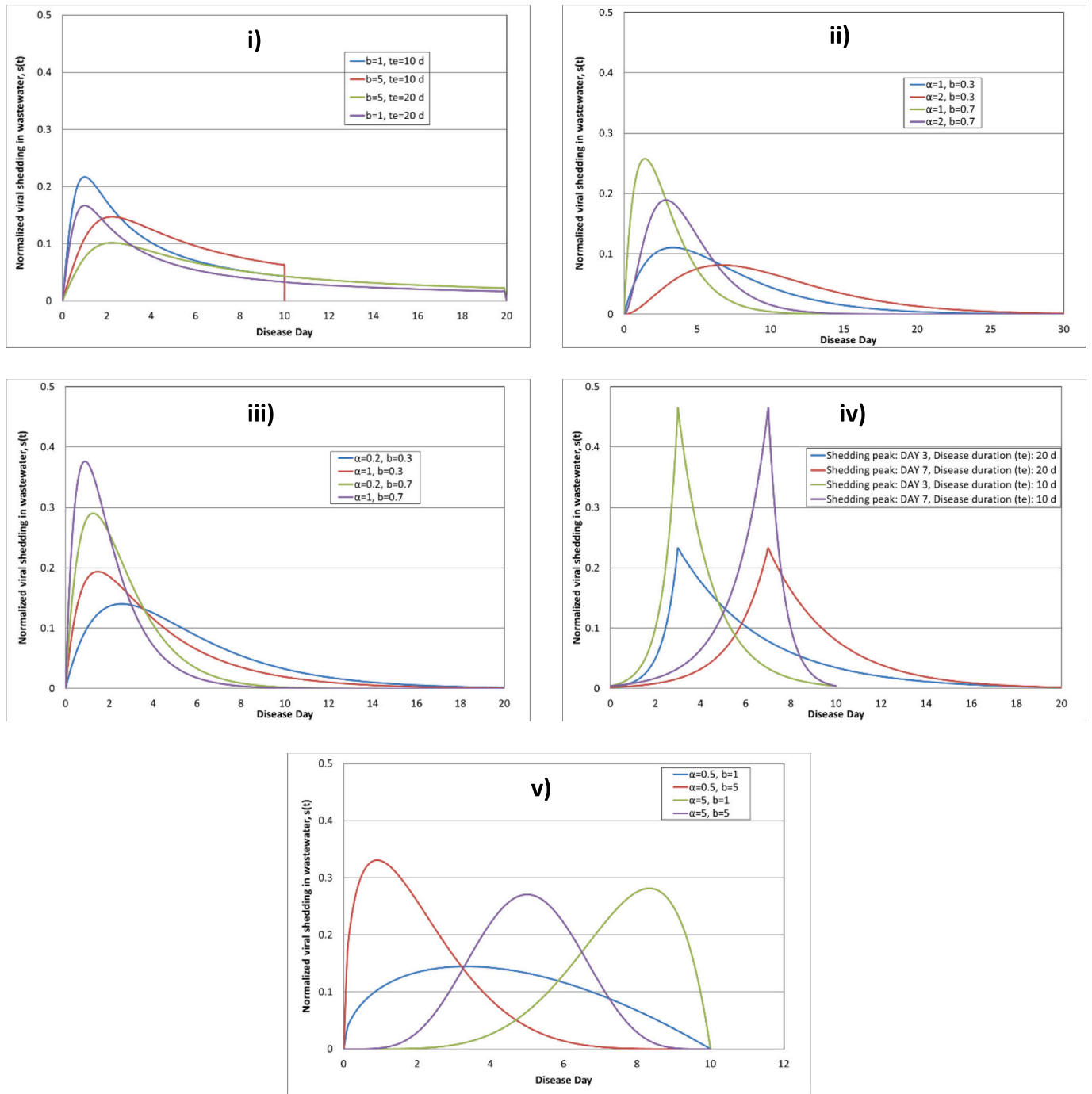
Fig. 1. Typical shapes of the function s(t) (normalized viral shedding vs time since infection) for several sets of their parameters. (i) shape (a), (ii) shape (b), (iii) shape (c), (iv) shape (d), (v) shape (e).

infinity as this is why the definition of t_e is necessary for this type of distribution.

Shape (b):

$$s(t) = \frac{b^{\alpha+1} t^\alpha e^{-bt}}{\Gamma(\alpha+1) - \Gamma(\alpha+1, t_d)} \tag{6}$$

where $\Gamma(x)$ and $\Gamma(x,y)$ are the Gamma and the incomplete Gamma functions, respectively, defined as $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt, \Gamma(x,y) = \int_y^\infty t^{x-1} e^{-t} dt$. The definition of t_e is optional. If a value for t_e is not

defined, the incomplete Gamma function term goes to zero. The necessary parameters of the distribution are α and b. The function s(t) for four combinations of the parameters $\alpha = 1$, b = 0.3; $\alpha = 2$, b = 0.3; $\alpha = 1$, b = 0.7; $\alpha = 2$, b = 0.7 is shown in Fig. 1b. The effect of α and b cannot be isolated. It appears that the distribution becomes narrower as α decreases or b increases. The curves shown refer to the absence of a value for t_e. Such a value can be introduced leading to the truncation of the corresponding curve (and to increased values imposed by renormalization required by the Eq. (4).

Shape (c):

$$s(t) = (1 - e^{-\alpha t})e^{-bt}\left(\frac{1 - e^{-\alpha t_e}}{b} - \frac{1 - e^{-(\alpha + b)t_e}}{\alpha + b}\right)^{-1} \tag{7}$$

The definition of t_e is optional. If it is not defined, the corresponding exponential terms simply goes to zero at this value. The function s(t) for four combinations of parameters α, b (α = 0.2, b = 0.3; α = 1, b = 0.3; α = 0.2, b = 0.7; α = 1, b = 0.7) in the absence of t_e appears in Fig. 1c. The distribution becomes narrower as α or b increases.

Shape (d): In this case a completely different parametrization of the function will be introduced for convenience. Instead of α, b the new parameter ε is introduced. This parameter denotes the ratio of the viral shedding between the inception and end of the shedding period to the maximum value of shedding. The modified equation takes the form:

$$s(t) = \varepsilon e^{-\zeta t/t_m}\frac{(\varepsilon t_m + t_e - t_m)(1 - e^{-\zeta})}{\zeta} \quad 0 \leq t \leq t_m \tag{8a}$$

$$s(t) = e^{-b(t - t_m)/(t_d - t_m)}\frac{(\varepsilon t_m + t_e - t_m)(1 - e^{-\zeta})}{\zeta} \quad t_m \leq t \leq t_d \tag{8b}$$

where $\zeta = \ln(1/\varepsilon)$.

In this form (unlike in the original form) the definition of the parameter t_e is necessary since it is used for the reparameterization. The new parameters are ε, t_m, t_e. Typically, a small value is assigned to parameter ε (e.g. ε = 0.01 in (Petala et al., 2022)]. It can be assumed fixed since the distribution has a small sensitivity to ε for such small values of the parameter. The other parameters t_m and t_e denote directly the position of the maximum values and the duration of viral shedding, respectively. Typical results for four combinations of the parameters t_m and t_e appear in Fig. 1d. It is interesting that the position of the maximum is free to move along the period of shedding.

Shape (e):

$$s(t) = \frac{1}{t_e}\frac{(t/t_e)^{\alpha}(1 - t/t_e)^b}{B(\alpha + 1, b + 1)} \tag{9}$$

where $B(x, y) = \int_0^1 t^{x-1}(1 - t)^{y-1}dt$ is the so called Beta function. In this case, the time t_e is hardwired in the equation and cannot be ignored. The function s(t) for t_e = 10 and four combinations of parameters α, b (α = 0.5, b = 1; α = 5, b = 1; α = 5, b = 1; α = 5, b = 5) is shown in Fig. 1e. As α and b increase the distribution becomes narrower. The position of the maximum value depends on the relative value of α with respect to b. This position can be freely move along the period of shedding.

3.3. Comparison between proposed distributions

It appears that the shape (a) has a sharp cutoff, and shapes (b), (c) have the location of maximum always at the first half period of shedding. The maximum location can be at any place along the shedding period for the shapes (d) and (e). The question is on what basis the comparison between distributions has to be made. Of course, it cannot be made through the direct parameter choice since parameters act in a different way for each shape. The requirement here is to construct an hierarchy of the features of the distribution. The most important features are (i) the total shed viral quantity N (ii) the duration of the viral shedding and (iii) the time of maximum shedding. The shape comparison of the distributions is meaningful for distributions having the same major features (i), (ii) and (iii). In order to compute the parameters of the distributions having the same features, analytical relations for the maximum location in terms of parameters are derived. The time of maximum shedding rate is given as: shape (a) $t_m = b^{0.5}$; shape(b) $t_m = \alpha/b$; shape (c) $t_m = \ln(1 + \alpha/b)\cdot 1/\alpha$; shape (d) t_m is a direct parameter; shape (e) $t_m = \alpha/(b-\alpha)$. The shape (d) distribution with the parameters t_m = 6, t_e = 32 employed in the data analysis of (Petala et al., 2022) is considered here. Then the parameters of the other shapes of the

distributions are chosen in order to match the values for t_m = 6 and t_e = 32 of shape (d). The resulting distributions are shown in the two Fig. 2a and b for clarity of presentation. The sharp cut off of shape (a) is obvious. According to Fig. 2b, the shapes (b), (c) and (d) are rather similar to each other with shape (e) having somewhat broader distribution whereas shapes (b) and (c) are almost identical. According to Fig. 2a, the shape (d) appears to be different than the shape (c), having a narrower form. Summarizing, the function S(t) is characterized hierarchically by the total viral shedding amount, the shedding duration and the time from shedding inception to maximum shedding. The exact shape of the distribution can be assumed as a fourth feature. A preliminary analysis showed that the shapes (a), (b), (c), (e) of the distribution are similar to each other and only the shape (b) is kept as a representative one to analyze their differences to the shape (d).

3.4. Discretization

Up to this point, we worked with continuous shapes of the distribution S(t). The corresponding equation is typically discretized on a daily basis. This implies that a detailed finite volume representation of the problem includes an integration over the period of a day. However, this is not the way of practical implementation. In practice a single value of S(t) at the end of a day is assumed to express the daily shedding rate. The two approaches are not equivalent. In Fig. 3 the resulting daily shedding rate based on continuous and discrete approaches are presented. As expected, (due to the smoothing effect of daily integration) the discrete approach leads to narrower distributions. Even the location of the maximum differs by one day between discrete and continuous approaches. So, it is necessary to define together with the shape of the distribution S(t) the way of its handling i.e. discrete or continuous.

4. Results and sensitivity analysis

The shape (d) of the function S(t) is considered as base case. The system of Eqs. (1)–(3) is discretized using single day intervals and the discrete version of S(t) is employed. It is noticed that in the discrete version the reporting happens at day t_d + 1. The procedure is applied to medical surveillance (https://eody.gov.gr) and wastewater data for the city of Thessaloniki between February 1st and December 15th, 2022, Fig. 4. Imported cases refer to daily cases reported by the National Health Organization (NHO), while the relative shedding rate is the normalized and rationalized SARS CoV-2 concentration (Petala et al., 2021) that was also announced daily. During that period everyday sampling, analysis and reporting to NHO and the public was implemented to allow for integrated COVID-19 transmission estimations. Samples are acquired every day from the entrance of the Wastewater Treatment Plant of the city. Normalization of viral load measurements with respect to wastewater flow rate and city population as well as rationalization with respect to specific physicochemical parameters of wastewater, e.g. suspended solids, organic load, dissolved oxygen etc. was performed according to a standard procedure explained before (Petala et al., 2021). Although normalization of viral load with respect to flow rate and population harmonizes adequately wastewater and medical data, it is notable that rationalization with respect to wastewater physicochemical enhances further the agreement between wastewater and medical data in Fig. 4. A detailed discussion on the latter is beyond the scope of the present work and is the subject of another publication which is underway. For the examined period, the analysis of clinical and wastewater data indicates a clear domination of the Omicron variant in the infected population (Chassalevris et al., 2022). This is a prerequisite for applying below a single S(t) shedding rate distribution. Furthermore, the first and last ten days of the examined period are not employed to avoid edge effects. The non-smoothed infections raw data are used in the model. Nevertheless, the computed ratio P of total (reported and unreported) to reported infected people is smoothed using a 7-days smoothing to account for the weekly periodicity of reported
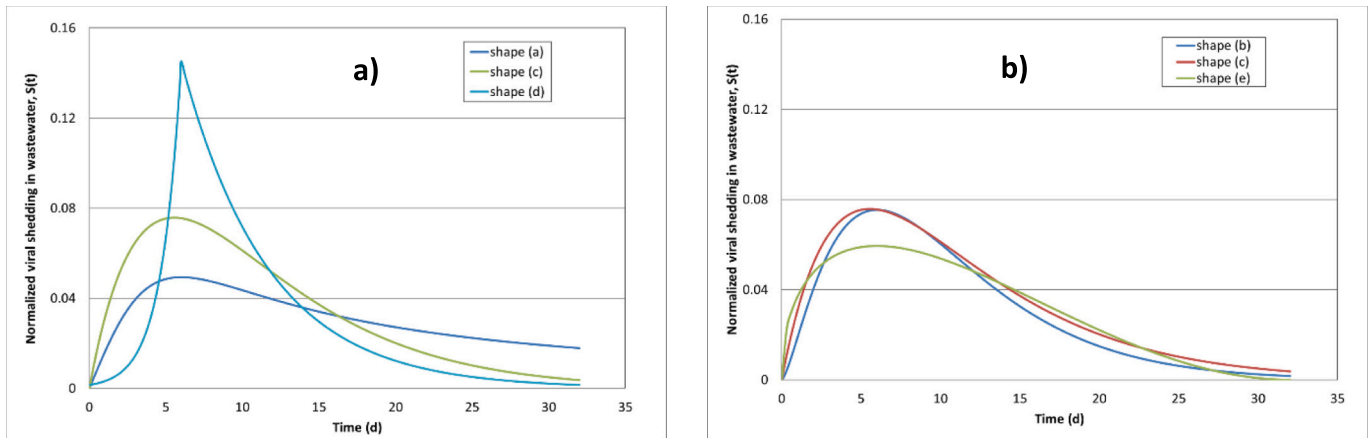
Fig. 2. Comparison between s(t) (normalized viral shedding vs time since infection) of several shapes with equivalent sets of parameters (resulting by matching main features of the s(t)). The set of parameters is the one used in (Petala et al., 2022).
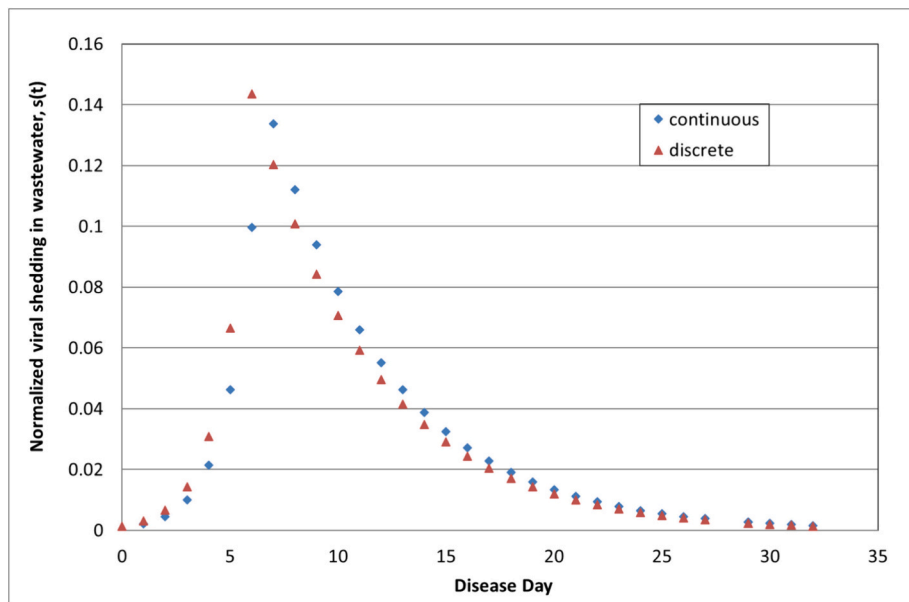

Fig. 3. Comparison between continuous and discrete distribution s(t) (normalized viral shedding vs time since infection) of shape (c). Same parameters as those used in Fig. 2.

infections (i.e. always lower during weekends due to limitations in clinical reporting, less medical tests, etc).

Based on literature work (e.g. (Puhach et al., 2023; Takahashi et al., 2022)) for the Omicron variant of the SARS-CoV-2 virus, the characteristic parameters values used in the model shape (d) are $t_e = 14$, $t_m = 4$. As a base case, the detection of infection is assumed to occur at the sixth day from the inception of viral shedding (i.e. $t_d = 5$). This is somewhat faster from what was considered in the previous period of COVID-19 (Petala et al., 2022) but it is justified by the higher awareness of the public about the infectiousness of the Omicron variant of the virus.

Below, the evolution of the ratio of the total infected (reported and unreported) people to reported infected people P during the period of study is presented and a sensitivity analysis with respect to the parameter values chosen follows. It is noticed that all the parameters are related to the function S(t) except t_d, the day of detection of the infection. The minimum value of P is not unity (as it should be due to the assumption of no unreported individuals at the minimum value of P) because of the smoothing procedure. The scenario $t_e = 14$, $t_m = 4$, $t_d = 5$ is the basic case. At first, the sensitivity with respect to t_m, the day of the maximum shedding rate in the distribution counted from the day of

infection, is studied. Fig. 5a presents the base case together with its modifications using $t_m = 3$ and $t_m = 5$. The average value of P goes from 2.38 to 2.66 as t_m goes from 3 to 5. The effect of t_m is not uniform during the period of consideration. The next parameter studied is the one related to the day of detection (i.e. t_d). The base case and its modifications for $t_d = 4$ and $t_d = 6$ appear in Fig. 5b. The average value of P goes from 2.39 to 2.68 as t_d goes from 6 to 4. It appears that the effect on P of increasing t_m is similar to the effect of decreasing t_d. This behavior supports the hypothesis that the difference between t_d and t_m is the main parameter affecting the evolution of P. The hypothesis is verified by comparing the base case (t_d-$t_m = 1$) with the one having $t_m = 5$, $t_d = 6$ (again t_d-$t_m = 1$). This comparison appears in Fig. 6. The deviation between the two distributions is very small verifying the argument that the P evolution is primary determined by the difference between the time of the maximum viral shedding to the time of infection detection. The next examined parameter is the duration of the shedding. The base case and its modifications for $t_e = 11$ and $t_e = 17$ appear in Fig. 7. The average P value goes from 2.5 to 2.55 as t_e goes from 11 to 17. It might be claimed that the effect of t_e on P is quite limited. The last investigated parameter is the shape of the function S(t). As an alternative to shape (d), the shape
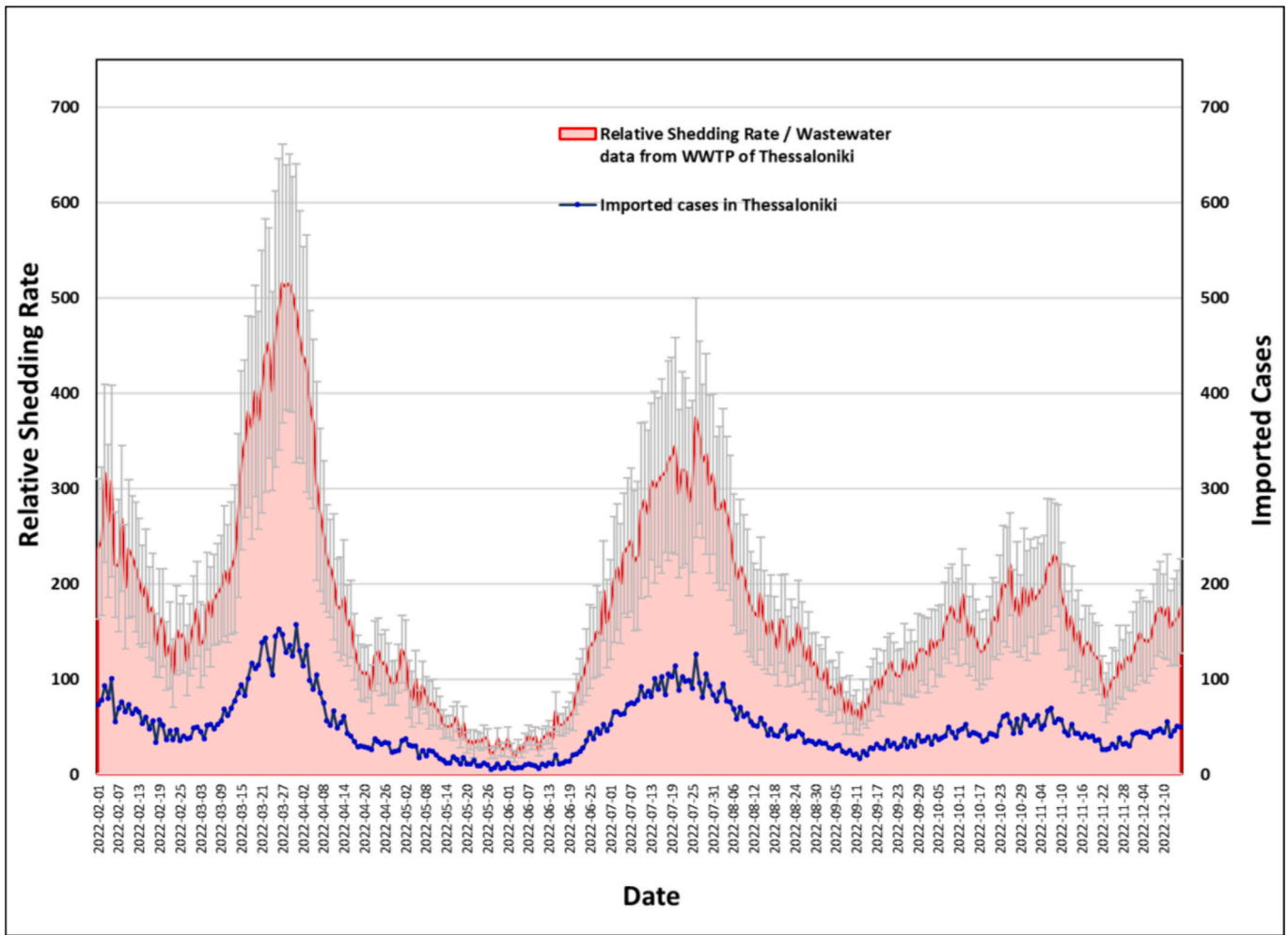
Fig. 4. Medical surveillance and wastewater data Rexp for the examined period.
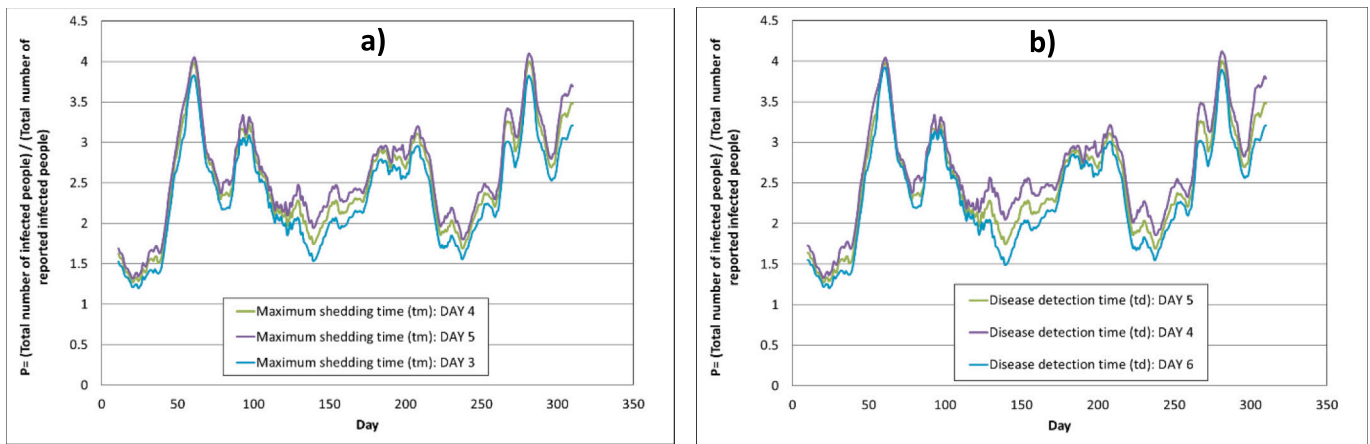

Fig. 5. Ratio of total (reported and unreported) to reported infected people P for shape (d) function s(t) (normalized viral shedding vs time since infection) with disease duration $t_e = 14$ d: (a) disease detection time $t_d = 5$ and examination of the effect of maximum shedding time t_m and (b) maximum shedding time $t_m = 4$ and examination of the effect of disease detection time t_d.

(b) as representative of the other shapes (as discussed in the previous section) is considered. The parameters of shape (b) are chosen to match the basic features of the base case shape (d) following the procedure explained in the previous section (resulting in $\alpha = 3$, $b = 0.75$). The comparison of the base case to its modification using shape (b) distribution appears in Fig. 8. The average *P* values are 2.54 (base case) and

2.45 (shape (b). One may argue that the effect of the shape of the distribution is small. In summary, it is shown that the main factor affecting the P evolution is the difference $t_d - t_m$. The absolute values of t_d and t_m, the duration t_e and the detailed shape of the S(t) have a small effect on the determination of P. It is noted that the above analysis is based on the assumption that shedding starts from the day of infection. The
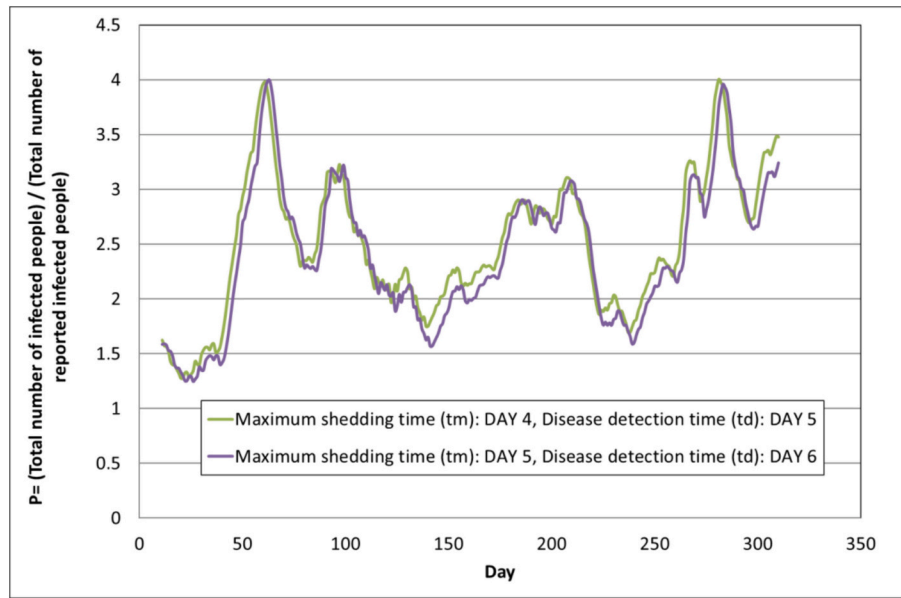
Fig. 6. Ratio of total (reported and unreported) to reported infected people P for shape (d) function s(t) (normalized viral shedding vs time since infection) with disease duration $t_e = 14$. The values t_d (disease detection time) and t_m (maximum shedding time) are chosen to have the same difference t_d-t_m.
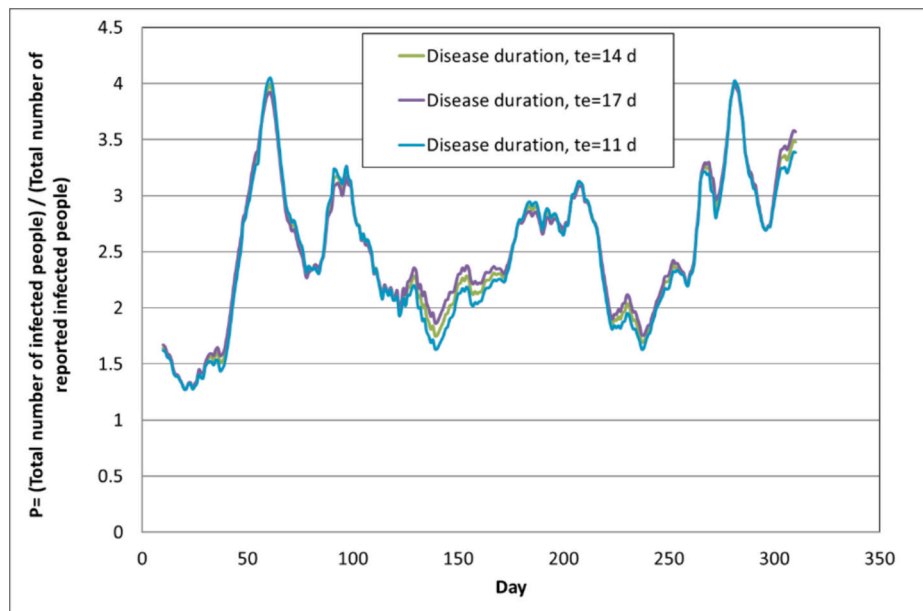

Fig. 7. Ratio of total (reported and unreported) to reported infected people P for shape (d) function s(t) (normalized viral shedding vs time since infection) with disease detection time $t_d = 5$, maximum shedding time $t_m = 4$. Examination of the effect of disease duration t_e.

possibility cannot be excluded that there may be a delay between infection and start of viral shedding. Such a generalization would not affect the results of the present work except that the definition of t_d should be different, i.e. the time of infection reporting after the inception of viral shedding. It is noted that there are no multiple sets of data to perform a statistical significance analysis. The only sensitivity analysis is performed with respect to the average value of P. It is clearly shown that this values is more sensitive to the parameter t_d-t_m (i.e. to the difference and not to the independent values of t_d and t_m) than to parameter t_e and to the shape of the distribution S. This information is vital for the derivation of an approximated model of reduced complexity compared to the present one.

The method employed here can be applied to any case where data of virus load in wastewater and of medically reported patients are available. The application of the suggested method over a period of time yields the temporal evolution of the ratio of the total (reported and unreported) patients over the reported patients. Apparently, the output of the method depends on the assumptions made. A basic assumption that permits mathematical treatment of the problem is that the viral shedding rate of infected individuals during the days of their disease is not constant but follows a distribution which starts from zero, passes through a peak/max value at a specific day of the infection and then gradually drops to practically zero at the end of the disease days. The scope of the present work is to compare different shedding rate distributions suggested in literature for SARS-CoV-2 in our model and extract the weight of the assumptions/parameters on the results of the method. It is found that the main parameter (maximum weight) is the time delay between reporting day and maximum shedding rate day.
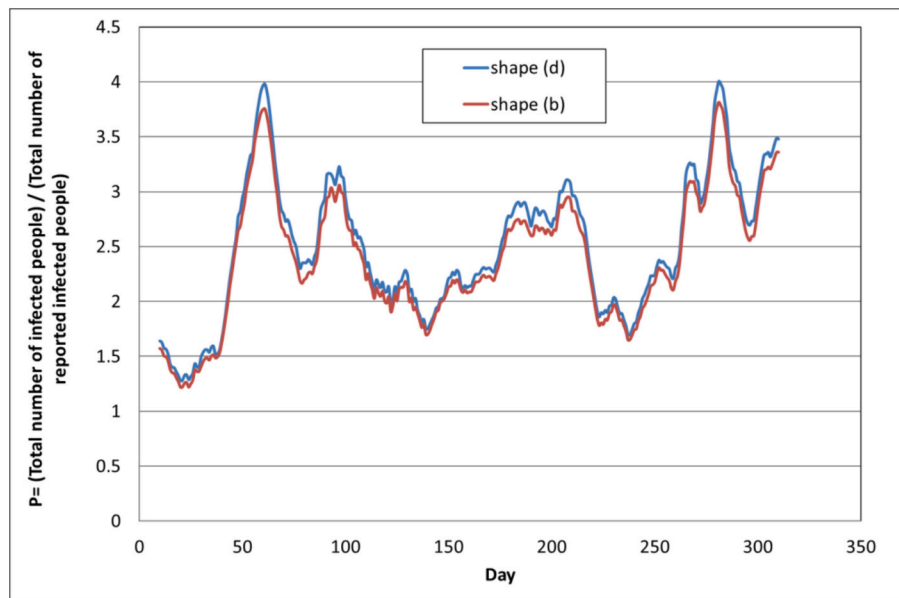
Fig. 8. Ratio of total (reported and unreported) to reported infected people P for shape (d) function s(t) (normalized viral shedding vs time since infection) with disease detection time $t_d = 5$, maximum shedding time $t_m = 4$, disease duration $t_e = 14$. Comparison is made with shape (b) function having equivalent parameters.

5. Conclusions

A simple model for calculating the total (reported and unreported) to reported infected people ratio based on wastewater viral content is introduced. A basic part of the model is the shedding function S(t) denoting how the shedding of virus parts by infected people evolves during infection. At least five different shapes for this function have been proposed in literature. These shapes are reviewed here and carefully normalized and rewritten in order to share an hierarchical set of features. Four of them are found to be quite similar to each other with only the fifth one differing from the others. So, there are two fundamentally different shapes. These two fundamental shapes are combined to the model and applied to medical surveillance and wastewater shedding rate data for the city of Thessaloniki for a period of 2022 dominated by the Omicron variant of the SARS-CoV-2 virus. It is found that the main parameter that determines the results of the model is the difference in time (i.e., days) between the maximum shedding and the infections reporting day. The duration of the disease (in days) and the exact shape of the function S(t) exhibit quite limited influence on the result of the calculations. So, the main outcome of the work is that the details of function S(t) are not really important in wastewater based epidemiology calculations. The most important feature of S(t) is by far the day of maximum shedding rate during the disease days of individuals. This feature is characteristic of the viral strain, therefore specific for dominant variants of each outbreak. Hence, determination of the day of maximum viral shedding, which is more feasible and faster than approximating the individual shedding distribution, can optimize the approaches to estimate disease transmission and spreading using wastewater and medical data during outbreaks. Coupling of wastewater and medical data is needed to achieve valuable projections for public health protection issues.

CRediT authorship contribution statement

M. Kostoglou: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **M. Petala:** Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Th. Karapantsios:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization. **Ch. Dovas:** Writing – review & editing, Investigation. **V. Tsiridis:** Writing – review & editing, Visualization, Data curation. **E. Roilides:** Writing – review & editing, Investigation. **A. Koutsolioutsou-Benaki:** Writing – review & editing, Investigation. **D. Paraskevis:** Investigation. **S. Metalidis:** Investigation. **E. Stylianidis:** Funding acquisition, Conceptualization. **A. Papa:** Writing – review & editing, Investigation. **A. Papadopoulos:** Writing – review & editing, Investigation. **S. Tsiodras:** Writing – review & editing, Investigation. **N. Papaioannou:** Supervision, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

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

Data availability

Data will be made available on request.

Acknowledgements

References

Cavany, S., Bivins, A., Wu, Z., North, D., Bibby, K., Perkins, T.A., 2022. Erratum: inferring SARS-CoV-2 RNA shedding into wastewater relative to the time of infection (epidemiology and infection (2022) 150 (E21) DOI:10.1017/S0950268821002752). Epidemiol. Infect. 150, 1–8. https://doi.org/10.1017/S0950268822000322.

Chassalevris, T., Chaintoutis, S.C., Koureas, M., Petala, M., Moutou, E., Beta, C., Kyritsi, M., Hadjichristodoulou, C., Kostoglou, M., Karapantsios, T., Papadopoulos, A., Papaioannou, N., Dovas, C.I., 2022. SARS-CoV-2 wastewater monitoring using a novel PCR-based method rapidly captured the Delta-to-omicron BA.1 transition patterns in the absence of conventional surveillance evidence. Sci. Total Environ. 844, 156932 https://doi.org/10.1016/J.SCITOTENV.2022.156932.

Ciannella, S., González-Fernández, C., Gomez-Pastora, J., 2023. Recent progress on wastewater-based epidemiology for COVID-19 surveillance: a systematic review of analytical procedures and epidemiological modeling. Sci. Total Environ. 878 https://doi.org/10.1016/j.scitotenv.2023.162953.

Hegazy, N., Cowan, A., D'Aoust, P.M., Mercier, É., Towhid, S.T., Jia, J.J., Wan, S., Zhang, Z., Kabir, M.P., Fang, W., Graber, T.E., MacKenzie, A.E., Guilherme, S., Delatolla, R., 2022. Understanding the dynamic relation between wastewater SARS-CoV-2 signal and clinical metrics throughout the pandemic. Sci. Total Environ. 853 https://doi.org/10.1016/j.scitotenv.2022.158458.

Hoffmann, T., Alsing, J., 2021. Faecal shedding models for SARS-CoV-2 RNA amongst hospitalised patients and implications for wastewater-based epidemiology. medRxiv 2021.03.16.21253603.

Kostoglou, M., Petala, M., Karapantsios, T., Dovas, C., Roilides, E., Metallidis, S., Papa, A., Stylianidis, E., Papadopoulos, A., Papaioannou, N., Cadaval, T.R., 2021. SARS-CoV-2 adsorption on suspended solids along a sewerage network: mathematical model formulation, sensitivity analysis, and parametric study. Environ. Sci. Pollut. Res. 16 https://doi.org/10.1007/s11356-021-16528-0.

Kostoglou, M., Karapantsios, T., Petala, M., Roilides, E., Dovas, C.I., Papa, A., Metallidis, S., Stylianidis, E., Lytras, T., Paraskevis, D., Koutsolioutsou-Benaki, A., Panagiotakopoulos, G., Tsiodras, S., Papaioannou, N., 2022. The COVID-19 pandemic as inspiration to reconsider epidemic models: a novel approach to spatially homogeneous epidemic spread modeling. Math. Biosci. Eng. 19, 9853–9876. https://doi.org/10.3934/mbe.2022459.

Mattei, M., Pintó, R.M., Guix, S., Bosch, A., Arenas, A., 2023. Analysis of SARS-CoV-2 in wastewater for prevalence estimation and investigating clinical diagnostic test biases. Water Res. 242, 120223 https://doi.org/10.1016/j.watres.2023.120223.

Miura, F., Kitajima, M., Omori, R., 2021. Duration of SARS-CoV-2 viral shedding in faeces as a parameter for wastewater-based epidemiology: re-analysis of patient data using a shedding dynamics model. Sci. Total Environ. 769, 144549 https://doi.org/10.1016/j.scitotenv.2020.144549.

Petala, M., Dafou, D., Kostoglou, M., Karapantsios, T.T., Kanata, E., Chatziefstathiou, A., Sakaveli, F., Kotoulas, K., Arsenakis, M., Roilides, E., Sklaviadis, T., Metallidis, S., Papa, A., Stylianidis, E., Papadopoulos, A., Papaioannou, N., 2021. A physicochemical model for rationalizing SARS-CoV-2 concentration in sewage. Case study: the city of Thessaloniki in Greece. Sci. Total Environ. 755, 142855 https://doi.org/10.1016/j.scitotenv.2020.142855.

Petala, M., Kostoglou, M., Karapantsios, T., Dovas, C.I., Lytras, T., Paraskevis, D., Roilides, E., Koutsolioutsou-Benaki, A., Panagiotakopoulos, G., Sypsa, V., Metallidis, S., Papa, A., Stylianidis, E., Papadopoulos, A., Tsiodras, S., Papaioannou, N., 2022. Relating SARS-CoV-2 shedding rate in wastewater to daily positive tests data: a consistent model based approach. Sci. Total Environ. 807, 150838 https://doi.org/10.1016/J.SCITOTENV.2021.150838.

Phan, T., Brozak, S., Pell, B., Gitter, A., Xiao, A., Mena, K.D., Kuang, Y., Wu, F., 2023. A simple SEIR-V model to estimate COVID-19 prevalence and predict SARS-CoV-2 transmission using wastewater-based surveillance data. Sci. Total Environ. 857, 159326 https://doi.org/10.1016/j.scitotenv.2022.159326.

Proverbio, D., Kemp, F., Magni, S., Ogorzaly, L., Cauchie, H.M., Gonçalves, J., Skupin, A., Aalto, A., 2022. Model-based assessment of COVID-19 epidemic dynamics by wastewater analysis. Sci. Total Environ. 827 https://doi.org/10.1016/j.scitotenv.2022.154235.

Puhach, O., Meyer, B., Eckerle, I., 2023. SARS-CoV-2 viral load and shedding kinetics. Nat. Rev. Microbiol. 21, 147–161. https://doi.org/10.1038/s41579-022-00822-w.

Schill, R., Nelson, K.L., Harris-Lovett, S., Kantor, R.S., 2023. The dynamic relationship between COVID-19 cases and SARS-CoV-2 wastewater concentrations across time and space: considerations for model training data sets. Sci. Total Environ. 871, 162069 https://doi.org/10.1016/j.scitotenv.2023.162069.

Takahashi, K., Ishikane, M., Ujiie, M., Iwamoto, N., Okumura, N., Sato, T., Nagashima, M., Moriya, A., Suzuki, M., Hojo, M., Kanno, T., Saito, S., Miyamoto, S., Ainai, A., Tobiume, M., Arashiro, T., Fujimoto, T., Saito, T., Yamato, M., Suzuki, T., Ohmagari, N., 2022. Duration of Infectious Virus Shedding by SARS-CoV-2 Omicron Variant–Infected Vaccinees. Emerg. Infect. Dis. 28, 998–1001. https://doi.org/10.3201/eid2805.220197.

Teunis, P.F.M., Sukhrie, F.H.A., Vennema, H., Bogerman, J., Beersma, M.F.C., Koopmans, M.P.G., 2015. Shedding of norovirus in symptomatic and asymptomatic infections. Epidemiol. Infect. 143, 1710–1717. https://doi.org/10.1017/S095026881400274X.

Wade, M.J., Lo Jacomo, A., Armenise, E., Brown, M.R., Bunce, J.T., Cameron, G.J., Fang, Z., Farkas, K., Gilpin, D.F., Graham, D.W., Grimsley, J.M.S., Hart, A., Hoffmann, T., Jackson, K.J., Jones, D.L., Lilley, C.J., McGrath, J.W., McKinley, J.M., McSparron, C., Nejad, B.F., Morvan, M., Quintela-Baluja, M., Roberts, A.M.I., Singer, A.C., Souque, C., Speight, V.L., Sweetapple, C., Walker, D., Watts, G., Weightman, A., Kasprzyk-Hordern, B., 2022. Understanding and managing uncertainty and variability for wastewater monitoring beyond the pandemic: lessons learned from the United Kingdom national COVID-19 surveillance programmes. J. Hazard. Mater. 424 https://doi.org/10.1016/j.jhazmat.2021.127456.

Wu, F., Xiao, A., Zhang, J., Moniz, K., Endo, N., Armas, F., Bonneau, R., Brown, M.A., Bushman, M., Chai, P.R., Duvallet, C., Erickson, T.B., Foppe, K., Ghaeli, N., Gu, X., Hanage, W.P., Huang, K.H., Lee, W.L., Matus, M., McElroy, K.A., Nagler, J., Rhode, S.F., Santillana, M., Tucker, J.A., Wuertz, S., Zhao, S., Thompson, J., Alm, E. J., 2022. SARS-CoV-2 RNA concentrations in wastewater foreshadow dynamics and clinical presentation of new COVID-19 cases. Sci. Total Environ. 805, 150121 https://doi.org/10.1016/j.scitotenv.2021.150121.