

On discrete time epidemic models in Kermack-McKendrick form

Odo Diekmann¹, Hans G. Othmer², Robert Planqué^{3,†}, and
Martin C. J. Bootsma^{1,4}

¹*Mathematical Institute, Utrecht University, Utrecht, The Netherlands*

²*School of Mathematics, University of Minnesota, Minneapolis, USA*

³*Department of Mathematics, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands*

⁴*Department of Epidemiology, Universitair Medisch Centrum Utrecht, Utrecht, The Netherlands*

[†]*r.planque@vu.nl*

March 26, 2021

1 Abstract

Surprisingly, the discrete-time version of the general 1927 Kermack-McKendrick epidemic model has, to our knowledge, not been formulated in the literature, and we rectify this omission here. The discrete time version is as general and flexible as its continuous-time counterpart, and contains numerous compartmental models as special cases. In contrast to the continuous time version, the discrete time version of the model is very easy to implement computationally, and thus promises to become a powerful tool for exploring control scenarios for specific infectious diseases. To demonstrate the potential, we investigate numerically how the incidence-peak size depends on model ingredients. We find that, with the same reproduction number and initial speed of epidemic spread, compartmental models systematically predict lower peak sizes than models that use a fixed duration for the latent and infectious periods.

1 Introduction

The day-night cycle has a strong impact on the behaviour of humans, animals and plants. As a rule, the resulting time heterogeneity is ignored in epidemiological and ecological models. One simply pretends that the representation of time by a continuous quantity t , "flowing" at a constant rate, is suitable for bookkeeping of the time course of the relevant events.

Census data, on the other hand, are often collected at regular intervals, so on a discrete time basis. Indeed, as evidenced by the Covid-19 pandemic, incidence is usually reported in the form of the number of new cases on a particular day or in a specified week.

24 So even when the processes that we want to capture take place in continuous
25 time, one may want to consider discrete time bookkeeping schemes, in order to
26 relate directly to the data, as advocated in the pioneering paper [2]. Moreover,
27 a significant bonus of discrete time models is that numerical implementation
28 is straightforward and that, accordingly, simulations are easy to perform. In
29 sharp contrast, the numerical solution of continuous time renewal equations, as
30 discussed in [3], presents a substantial challenge to the uninitiated.

31 Counter to the practical advantages runs a modelling difficulty: the formu-
32 lation of discrete time models is subtle and hence error-prone. In infinitesimal
33 time intervals the effects of different mechanisms are independent. Consequently
34 one can add terms that describe contributions to the rate of change of a quan-
35 tity. When trying to capture (in one go, rather than by solving a differential
36 equation) change in a finite time interval, we do have to think about the order of
37 events and how one event may trigger or prevent another event. In the epidemic
38 context a key point is that, when a susceptible host is infected by an infectious
39 host, it cannot any longer be infected by another infectious host.

40 The first aim of this short note is to formulate the discrete time version
41 of the general Kermack-McKendrick epidemic model from 1927. As far as we
42 know, this has not been done before. And yet the model is, most likely, an ideal
43 tool for data driven analysis of infectious disease outbreaks since, as we shall
44 demonstrate, it is not only general and flexible, but also extremely user friendly.

45 A second (and admittedly somewhat pedantic) aim of this note is to point out
46 explicitly a frequent mistake in the formulation of discrete time epidemiological
47 and ecological models. We hope that by clearly exposing the underlying fallacy,
48 the mistake will have had its day.

49 To put some flesh on the bones, we show that the qualitative behaviour of
50 the discrete time model is the spitting image of the qualitative behaviour of the
51 continuous time model. In order not to ignore the popularity of compartmental
52 variants (which is unwarranted, in our opinion, as there is neither evidence that
53 the length of, for instance, the infectious period is exponentially/geometrically
54 distributed nor that infectiousness is constant during this period), we put them
55 in the spotlight. To conclude, we illustrate the relevance (in particular for public
56 health policy) of the lesser known members of the Kermack-McKendrick family,
57 by investigating numerically how the peak of the incidence curve varies among
58 members that are identical with respect to both the initial growth rate ρ and
59 the basic reproduction number R_0 , but differ in assumptions about the duration
60 of the exposed and the infectious period.

61 To avoid misunderstanding, we now clarify what is, and what is not, stochas-
62 tic in the models formulated and analyzed below. All models are deterministic
63 at the population level. (One can think of them, in Kurtz spirit, see [16], as
64 the large initial population size limit of a stochastic model for finitely many
65 individuals.)

66 When one assumes that the infectious period of all individuals, once infected,
67 has exactly the same length and their infectiousness during this period is one
68 and the same constant, there is no randomness at the individual level either.

69 But most models incorporate heterogeneity/stochasticity at the individual
70 level. For instance, in the familiar continuous time *SIR* compartmental model,
71 the length of the infectious period of a newly infected individual is exponen-
72 tially distributed, say with parameter γ . If during the infectious period all
73 infected individuals have the same constant infectiousness β , then the expected

74 infectiousness $A(\tau)$ at time τ after becoming infected equals $\beta e^{-\gamma\tau}$. This re-
75 flects that after time τ has elapsed, the probability to be still infectious equals
76 $e^{-\gamma\tau}$. At the population level, we translate this into: a fraction $e^{-\gamma\tau}$ of those
77 infected at time t is still infectious at time $t + \tau$. In the discrete time setting,
78 we should replace the exponential distribution by the geometric distribution.
79 As a final elucidation we mention that the parameter β and the function $A(\tau)$
80 implicitly incorporate information about the contact (between individuals) pro-
81 cess that underlies transmission. A key feature is that contacts are assumed
82 to be uniformly at random (so neither spatial nor age nor social structure are
83 incorporated).

84 2 The cumulative force of infection

85 First of all we want to motivate and promote the equation

$$S(t+1) = e^{-\hat{\Lambda}(t)} S(t) \quad (2.1)$$

86 as a building block for discrete time models of the spread of an infectious disease
87 in a host population when

- 88 • the disease generates permanent immunity,
- 89 • the host population is demographically closed (meaning that demographic
90 turnover happens at a much slower time scale than transmission of the
91 disease and is therefore ignored).

92 As usual, $S(t)$ denotes the size of the subpopulation of susceptibles at census
93 time t . Underlying (2.1) is a choice of the unit of time: it equals the length of the
94 interval between one census and the next. So the magnitude of $\hat{\Lambda}(t)$ is propor-
95 tional to this length and when this magnitude figures in our discussion below,
96 one may interpret the statements in terms of the length of the discretization
97 step. We call $\hat{\Lambda}(t)$ the cumulative force of infection in the time window $(t, t+1]$
98 for reasons that we now explain. The continuous time version of (2.1) reads

$$\frac{dS}{dt}(t) = -\Lambda(t)S(t) \quad (2.2)$$

99 where $\Lambda(t)$ is the force of infection at time t , i.e., the probability per unit of
100 time for a susceptible to become infected at time t . By integration we deduce
101 from (2.2) the relation

$$S(t+1) = e^{-\int_t^{t+1} \Lambda(\tau) d\tau} S(t) \quad (2.3)$$

102 The first factor at the right hand side of (2.1) and (2.3) is, in both cases, the
103 probability for a susceptible to escape from infection in the time window $(t, t+1]$.
104 The integral

$$\int_t^{t+1} \Lambda(\tau) d\tau$$

105 in (2.3) is replaced by $\hat{\Lambda}(t)$ in (2.1) and this, we hope, clarifies why we call $\hat{\Lambda}(t)$
106 the cumulative force of infection.

107 We insist that one should adjust the multiplicative factor (as was indeed
108 done in [2]) and *not* replace the differential equation (2.2) by the “additive”
109 difference equation

$$S(t+1) - S(t) = -\hat{\Lambda}(t)S(t)$$

110 i.e., by

$$S(t+1) = (1 - \hat{\Lambda}(t))S(t) \quad (2.4)$$

111 Of course (2.4) provides a good approximation of the “true” equation (2.1) for
112 small values of $\hat{\Lambda}(t)$. But (2.4) is not exact and it may fail dramatically for
113 not so small values of $\hat{\Lambda}(t)$, in particular since it may lead to negative values of
114 S . The reason is that (2.4) does not take into account that a host can become
115 infected only once. To stress this point, we now present a somewhat mechanistic
116 derivation of the multiplicative factor in (2.1), showing that (2.1) does take this
117 into account.

118 Assume that when a single infectious individual is present in a certain host
119 population (during a time interval of, say, length one) every susceptible host be-
120 comes infected with probability p . Then any susceptible escapes from becoming
121 infected with the complementary probability $1 - p$. Next assume that there are
122 N infectious individuals and that these make contacts with susceptibles inde-
123 pendently of each other. Then any susceptible escapes from becoming infected
124 with probability

$$(1 - p)^N = e^{N \ln(1-p)}$$

125 So a susceptible is infected with probability $1 - (1 - p)^N$ rather than with
126 “probability” pN .

127 From a numerical point of view, the exponential has the disadvantage of
128 being expensive in terms of calculation costs. It may therefore be tempting
129 to reduce the step size in order to work safely with the linear approximation.
130 We actually wonder whether solving the relevant ODEs with for instance a
131 Runge Kutta solver is not a more attractive alternative, especially in terms of
132 accuracy. Moreover, we maintain that choosing a time interval that matches
133 the data points has definite advantages.

134 **3 The general discrete time Kermack-McKendrick** 135 **model**

136 As already expressed in (2.2), the incidence at time t equals $\Lambda(t)S(t)$ with Λ the
137 force of infection. Common sense tells us that the current force of infection is
138 generated by individuals who were themselves infected some time ago. Following
139 earlier work by Ross and Hudson (see [10, 11] and references in there), Kermack
140 and McKendrick translate this observation into the constitutive equation

$$\Lambda(t) = \int_0^\infty A(\tau)\Lambda(t-\tau)S(t-\tau)d\tau \quad (3.1)$$

141 with $A(\tau)$ the *expected* contribution to the force of infection at time τ after
142 infection. So in this top-down approach the infinite dimensional parameter A
143 is introduced as a key model ingredient. For any specific disease one may, in
144 principle, use a within-host model of the struggle between pathogen and immune
145 system to provide bottom-up a quantitative specification. Alternatively, one

146 may use population level data to infer (certain characteristics of) A , cf. [19].
147 Often this is done after first restricting to a parameterized family of functions
148 A (but see [9] for an alternative methodology.) Note that, with N denoting the
149 population size, we have

$$R_0 = N \int_0^\infty A(\tau) d\tau$$

150 and that, in the initial phase of the epidemic, the distribution of the generation-
151 interval (see [17, 18], and the references given in there) has as density the renor-
152 malized (to have integral 1) function A .

153 The incidence at time t is given by $\Lambda(t)S(t)$. Under the “permanent immu-
154 nity and no demographic turnover” assumption, the incidence equals $-S'(t)$,
155 cf. (2.2). Substituting this into (3.1), we deduce by integration (and upon
156 changing the order of integration) the identity

$$\int_t^{t+1} \Lambda(\sigma) d\sigma = \int_0^\infty A(\tau)[S(t-\tau) - S(t+1-\tau)] d\tau.$$

157 The discrete time counterpart reads

$$\hat{\Lambda}(t) = \sum_{j=1}^\infty A_j[S(t-j) - S(t+1-j)] \quad (3.2)$$

158 where now A_j is the expected contribution to the cumulative force of infection
159 in $(t, t+1]$ of an individual who itself became infected in the time window
160 $(t-j, t-j+1]$, so j time steps earlier, and $S(t-j) - S(t+1-j)$ is the
161 incidence in $(t-j, t-j+1]$. So the key model ingredient is now the collection

$$\{A_j\}_{j=1}^\infty$$

162 of positive/non-negative numbers, which we assume to be such that

$$\sum_{j=1}^\infty A_j < \infty.$$

163 (Incidentally, note that control measures or seasonality may cause the A_k to
164 depend on calendar time. See the end of Section 6 for a somewhat concrete
165 example. In Section 7 we shall briefly indicate how one can easily implement
166 this generalization.)

167 Equation (2.1), with $\hat{\Lambda}(t)$ specified by (3.2), provides an updating scheme,
168 but to get started one needs to specify an “initial” condition in the form of the
169 history of S up to a certain point in time. The interpretation requires that
170 this prescribed history is a monotone non-decreasing (when looking back into
171 time) sequence, bounded from above by the total host population size. We shall
172 denote this total size by N .

173 As we show next, one can reformulate (2.1), (3.2) as the scalar higher order
174 recursion relation

$$s(t+1) = e^{-\sum_{k=1}^\infty (1-s(t-k+1))\bar{A}_k} \quad (3.3)$$

175 where

$$s(t) := \frac{S(t)}{N} \quad (3.4)$$

176 and

$$\tilde{A}_k := A_k N. \quad (3.5)$$

177 Equation (3.3) is the discrete time analogue of the nonlinear renewal equation

$$s(t) = e^{-\int_0^\infty (1-s(t-\tau))NA(\tau)d\tau} \quad (3.6)$$

178 that follows by combining (2.2) with (3.2) and incidence equal to $-S'$, see [3, 6].

179 Both (3.3) and (3.6) involve the additional assumption

$$S(-\infty) = N, \quad (3.7)$$

180 expressing that in the infinite past all host individuals were susceptible.

181 To derive (3.3), first note that iteration of (2.1) yields, if (3.7) holds, the
182 identity

$$S(t+1) = e^{-\sum_{i=0}^\infty \hat{\Lambda}(t-i)} N. \quad (3.8)$$

From (3.2) we deduce

$$\begin{aligned} \sum_{i=0}^\infty \hat{\Lambda}(t-i) &= \sum_{i=0}^\infty \sum_{k=1}^\infty A_k [S(t-i-k) - S(t-i-k+1)] \\ &= \sum_{k=1}^\infty A_k \sum_{i=0}^\infty [S(t-i-k) - S(t-i-k+1)] \\ &= \sum_{k=1}^\infty A_k [N - S(t-k+1)]. \end{aligned}$$

183 If we use this last identity in (3.8), divide both sides of (3.8) by N and adopt
184 the notation (3.4) and (3.5), we obtain (3.3).

185 If one copies (3.3), with $t+1$ replaced by t , and combines the two formulas,
186 one can derive the variant

$$s(t+1) = s(t)e^{-\sum_{k=1}^\infty (s(t-k) - s(t-k+1))\tilde{A}_k}. \quad (3.9)$$

187 This variant has the advantage that one can provide an initial condition, say
188 at time zero, by prescribing $s(0)$ and the (nonnegative) incidences $\dots, s(-3) -$
189 $s(-2), s(-2) - s(-1), s(-1) - s(0)$. We refer to Section 7 for a more pragmatic
190 formulation of the initial value problem.

191 We conclude that (3.3)/(3.9) is the mathematical form of the discrete time
192 Kermack-McKendrick model with, in principle, a countably infinite parameter
193 $\{A_k\}_{k=1}^\infty$, but in practice a finite dimensional parameter with an infinite tail of
194 zeros.

195 4 The initial phase and the final size

196 To capture the demographic stochasticity during the very early phase of the
197 introduction of an infectious disease in a host population, we need branching
198 processes, see e.g. [6]. But once there is a large number of infected individuals,
199 we can switch to a deterministic description. The large number may, of course,
200 still constitute only a rather small *fraction* of a very large host population. In
201 this situation we may put

$$s(t) = 1 - x(t) \quad (4.1)$$

202 into (3.3) and assume that x is so small that it makes sense to replace the
203 exponential by the zero'th and first order terms of its Taylor expansion. This
204 yields the linearized equation

$$x(t+1) = \sum_{k=1}^{\infty} \tilde{A}_k x(t-k+1). \quad (4.2)$$

205 We define

$$R_0 = \sum_{k=1}^{\infty} \tilde{A}_k = \sum_{k=1}^{\infty} A_k N, \quad (4.3)$$

206 and interpret, based on the last identity, R_0 as the expected number of secondary
207 cases caused by a primary case in a totally susceptible host population.

208 In order to show that positive solutions of (4.2) grow when $R_0 > 1$, but
209 decline when $R_0 < 1$, we make the Ansatz

$$x(t) = \lambda^t. \quad (4.4)$$

210 By substitution of (4.4) into (4.2) we find that x defined by (4.4) is indeed a
211 solution if and only if λ is a real root of the discrete time characteristic equation

$$1 = \sum_{k=1}^{\infty} \lambda^{-k} \tilde{A}_k \quad (4.5)$$

212 known as the Euler-Lotka equation. The non-negativity of \tilde{A}_k , $k = 1, 2, \dots$,
213 guarantees that (4.5) has at most one real root ρ and that it does indeed have a
214 real root when the right hand side assumes a value bigger than one for some real
215 λ , so in particular when $R_0 > 1$ (when $R_0 < 1$ and \tilde{A}_k has power-like behaviour
216 for $k \rightarrow \infty$, the value of the right hand side may jump from a value less than one
217 to infinity when λ is decreased; when $\tilde{A}_k = 0$ for large k , this cannot happen and
218 ρ exists). Readers who wonder (or even worry) about the potential importance
219 of complex roots can consult [6, Section 8.2] and the references given there, to
220 be eased.

221 So we see that a key point is that the right hand side of (4.5) is a monotone
222 decreasing function of real λ . And as a consequence we have

$$\text{sign}(\rho - 1) = \text{sign}(R_0 - 1). \quad (4.6)$$

223 (Incidentally, note that ρ corresponds to e^r , with r the Malthusian parameter
224 featuring in the continuous time theory.) General linear theory, cf. [1], guaran-
225 tees that positive solutions of (4.2) grow geometrically with rate ρ for $t \rightarrow \infty$
226 when $\rho > 1$ (and decline with rate ρ , when ρ exists and is less than one). Gen-
227 eral nonlinear theory, cf. [20], guarantees that the steady state solution $s(t) \equiv 1$
228 of (3.3) is asymptotically stable for $\rho < 1$ (hence for $R_0 < 1$), but unstable for
229 $\rho > 1$, i.e., for $R_0 > 1$ (here we refer to the Principle of Linearized Stability; the
230 more general Hartman-Grobman Theorem implies that the intersection of the
231 unstable manifold and the positive cone is one-dimensional; this means that,
232 modulo translation, there is exactly one positive solution of (3.3) that has limit
233 1 for $t \rightarrow -\infty$, see [4] for the continuous time version).

234 So when $R_0 > 1$, the introduction of the pathogen will, provided the pathogen
235 does not go extinct by bad (or good, depending on the point of view) luck when

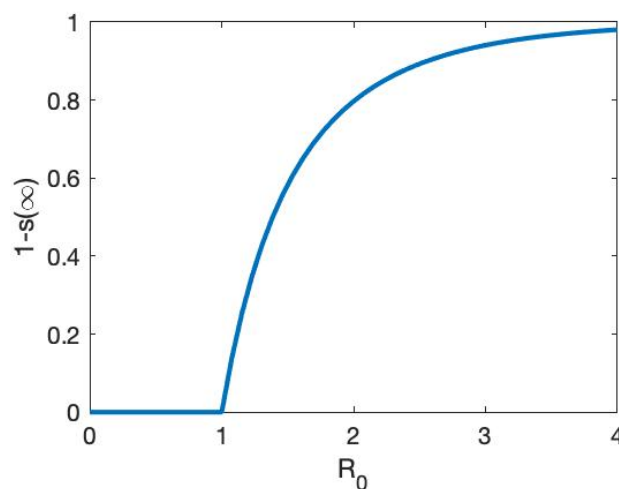


Figure 1: Graph of the final size $1 - s(\infty)$, i.e., the fraction of the population that gets infected in the course of the outbreak, as a function of the basic reproduction number R_0 .

236 still very rare, break through and cause s to decrease to below 1. The interpre-
 237 tation makes it obvious that s is a monotone decreasing function of time, and
 238 that it has a limit for $t \rightarrow \infty$. We denote this limit by $s(\infty)$. The equation

$$s(\infty) = e^{-R_0(1-s(\infty))} \quad (4.7)$$

239 is obtained by passing to the limit in (3.3), while using that the $\{\tilde{A}_k\}$ are
 240 summable. For $R_0 > 1$ this equation has a unique solution in $(0, 1)$, see Figure
 241 1, and [6, Exercise 1.19].

242 A comparison of the results in [13], [6, Chapter 1] and [3], with those above,
 243 establishes that when we compare the continuous time and discrete time for-
 244 mulations,

- 245 • there is only a formal difference in the expressions for R_0 ;
- 246 • if we put $\rho = e^r$, there is only a formal difference in the equations charac-
 247 terizing, respectively, ρ and r ;
- 248 • the equations specifying $s(\infty)$ on the basis of R_0 are identical (as already
 249 noted in [2]).

250 We conclude that at the level of theory, there is an exact parallel.

251 5 Compartmental formulation for some very spe- 252 cial cases

253 We shall use the standard notational convention (or should one say “ambigu-
 254 ity”?) that a compartment and its contents are denoted by the same symbol.
 255 We start with SIR and after that generalize to $SEIR$, hoping that these two

256 examples elucidate the general pattern how to construct discrete time models
257 in compartmental settings. See [12] for a more general set-up.

258 Assume that, upon infection, an individual is transferred from the com-
259 partment S to the compartment I of infectious individuals. Assume that every
260 following time step this infected individual stays in I with probability $1-\alpha$ while
261 being “removed” (i.e., losing infectiousness, either by way of the immune sys-
262 tem conquering the pathogen, or by death) with probability α . We put removed
263 individuals in a compartment R and assume that immunity is permanent (and
264 resurrection impossible). Finally, we assume that the cumulative force of infec-
265 tion equals βI , i.e., the per capita contribution to the force of infection equals
266 β . Note that β is proportional to the length of the discretization interval, i.e.,
267 the time between census points, and that $\alpha = 1 - e^{-\tilde{\alpha}}$ with $\tilde{\alpha}$ proportional to
268 this length.

These assumptions lead to the system of recurrence relations

$$\begin{aligned} S(t+1) &= e^{-\beta I(t)} S(t), \\ I(t+1) &= (1 - e^{-\beta I(t)}) S(t) + (1 - \alpha) I(t), \\ R(t+1) &= \alpha I(t) + R(t). \end{aligned} \quad (5.1)$$

269 We show that the system (5.1) may be reduced to the scalar recurrence (3.3)
270 by choosing the \tilde{A}_k appropriately:

$$\tilde{A}_k = \beta(1 - \alpha)^{k-1} N. \quad (5.2)$$

271 The first step corresponds to the derivation of (3.8): by iteration of the first
272 equation of (5.1) we obtain

$$S(t+1) = e^{-\beta \sum_{j=0}^{\infty} I(t-j)} N.$$

273 Rewriting the second equation of (5.1) as

$$I(t+1) = S(t) - S(t+1) + (1 - \alpha) I(t)$$

274 we obtain by summation the identity

$$\sum_{j=0}^{\infty} I(t-j) = N - S(t) + (1 - \alpha) \sum_{j=0}^{\infty} I(t-j-1),$$

275 and by substitution of this identity repeatedly at the right hand side,

$$\sum_{j=0}^{\infty} I(t-j) = N - S(t) + (1 - \alpha)(N - S(t-1)) + (1 - \alpha)^2 N(S(t-2)) + \dots$$

276 Finally, substitution of this last identity in the formula for $S(t+1)$ above yields
277 (3.8).

278 Conversely, starting from (3.3) with \tilde{A}_k given by (5.2), we easily recover
279 (5.1) by defining

$$I(t) := \sum_{k=1}^{\infty} [S(t-k) - S(t-k+1)] (1 - \alpha)^{k-1} \quad (5.3)$$

280 (note that the equation for $R(t)$ is just an appendix; it has no impact on the
281 dynamics of $S(t)$ and $I(t)$; it simply keeps track of individuals that are no longer
282 infectious).

283 We emphasize that if one replaces $e^{-\beta I(t)}$ by $1 - \beta I(t)$, the reduction to a
284 higher order scalar recursion relation fails (we invite readers to convince them-
285 selves of this fact)!

286 In order to capture a latent period, we next change the assumptions. Upon
287 infection, an individual now enters the compartment E of exposed (i.e., infected
288 but not yet infectious) individuals. When the length of the latent period is
289 geometrically distributed with parameter γ , we have to replace (5.1) by

$$\begin{aligned} S(t+1) &= e^{-\beta I(t)} S(t), \\ E(t+1) &= (1 - e^{-\beta I(t)}) S(t) + (1 - \gamma) E(t), \\ I(t+1) &= \gamma E(t) + (1 - \alpha) I(t), \\ R(t+1) &= \alpha I(t) + R(t). \end{aligned} \tag{5.4}$$

290 Do parameters \tilde{A}_k exist such that (5.4) can be condensed to (3.3)? It is helpful
291 to think in terms of a stochastic process in which an individual can be in the
292 states S , E , I and R . In fact, E and I suffice, since we start “looking” at the
293 individual when it is infected and stop “looking” when it loses infectiousness. If
294 we label E with index 1 and I with index 2, then the probability distribution
295 of the state-at-infection is represented by the vector

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

296 The state transitions are described by the matrix

$$M = \begin{pmatrix} 1 - \gamma & 0 \\ \gamma & 1 - \alpha \end{pmatrix}$$

297 and infectiousness by the vector

$$b = \begin{pmatrix} 0 & \beta \end{pmatrix}.$$

298 So the expected infectiousness k units of time after becoming infected is given
299 by

$$A_k = b M^{k-1} \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

300 and hence by

$$A_k = b(M^{k-1})_{2,1} = \beta \sum_{l=1}^{k-1} \gamma(1 - \gamma)^{l-1} (1 - \alpha)^{k-1-l} \tag{5.5}$$

301 (with the convention that the sum equals zero when the upper index does not
302 exceed or equal the lower index). The parameters \tilde{A}_k are again defined by (3.5).
303 And when \tilde{A}_k has the form defined by (3.5) and (5.5), then (5.4) follows from
304 (3.3) if we define

$$\begin{aligned} E(t) &= \sum_{j=1}^{\infty} [S(t-j) - S(t-j+1)] (1 - \gamma)^{j-1}, \\ I(t) &= \sum_{j=1}^{\infty} [S(t-j) - S(t-j+1)] \sum_{l=1}^{\infty} \gamma(1 - \gamma)^{l-1} (1 - \alpha)^{j-1-l}. \end{aligned} \tag{5.6}$$

305 We trust that our presentation above, in terms of two vectors and one matrix, all
306 having a well-defined interpretation, makes clear how one can in general relate
307 compartmental epidemic models to a scalar higher order recursion. See Section
308 9.3 of [7] for a detailed elaboration of the continuous time case.

309 We conclude that there is a multitude of compartmental models that corre-
310 spond to a special choice of the parameters A_k . To prove the results of Section
311 4 directly for a model with 27 compartments is a hell of a job, especially if one
312 does not recognize the underlying structure (and 27 components make recogni-
313 tion difficult). More importantly, it is an unnecessary job: one only needs to
314 observe that one deals with a special case of (3.3).

315 6 On the choice of parameters A_k

316 The ingredients $\{A_k\}$ subsume mechanistic properties of the process of contact
317 between hosts as well as physiological/immunological properties of within-host
318 dynamics. As a rule, information about such properties is scarce. One has to
319 make educated guesses. See [15] for a concrete example.

320 The *SIR* and *SEIR* formulations of the last section have the advantage of
321 involving just a few parameters. But, in our opinion, they have the disadvan-
322 tage of being wrongly-educated guesses: they result from the tendency to do
323 what others do, despite the fact that data, as a rule, do *not* support geometric
324 distributions for the length of the latent- and/or the infectious period.

325 A facilitating aspect is that A_k are averages (see [6, Section 2.1] for a detailed
326 exposition, including examples). If we “know” that at day six after infection
327 only 10% of the infected individuals is infectious, while at day seven this rises to
328 20%, we can use this information directly in our choice of A_k . If we know that
329 at days six and seven the degree of infectiousness differs among individuals, we
330 can still use the guestimated average.

331 A more theoretical example is the following. Assume that a fraction p of the
332 infected individuals is asymptomatic. Assume that a symptomatic individual
333 has at day j after infection a probability θ_j to be detected and next put into
334 quarantine. Assume that the intrinsic infectiousness and contact intensity of
335 symptomatic and asymptomatic cases is identical and given by $\{B_k\}$. Then we
336 choose

$$A_k = \left(p + (1 - p) \prod_{j=1}^k (1 - \theta_j) \right) B_k. \quad (6.1)$$

337 Note that (6.1) is based on the debatable assumption that at the day of its de-
338 tection an individual does not contribute to the force of infection. This weakness
339 is easily remedied, but at the cost of introducing yet another parameter.

340 The parameters θ_j can capture the effect of testing. During a serious out-
341 break, such as the Covid-19 pandemic, the testing policy and possibility depend
342 on calendar day. This introduces time dependence in the parameters θ_j . Sim-
343 ilarly, control measures that reduce contact opportunities affect the B_k in a
344 time-dependent multiplicative manner. In the next section we introduce a com-
345 putational scheme in which such time dependence is easily incorporated.

346 7 Reformulation as a first order system

347 As Section 4 shows, the scalar higher order recursion relation (3.3) is very convenient for theoretical purposes. But for doing computations, a first order system
348 of equations is more convenient.

349 For feasibility, we want a finite dimensional system. To achieve this, we make the very reasonable assumption that the indices j for which A_j is strictly
350 positive have a finite upper bound. In other words, we assume that an integer
351 m exists such that $A_j = 0$ for all $j \geq m + 1$. The relevant consequence is that
352 the history of S , that matters for determining the future, has finite length.
353

354 Define

$$355 X_j(t) := s(t + 1 - j), \quad j = 1, \dots, m. \quad (7.1)$$

356 Much of the dynamics of the vector X amounts to shifting:

$$X_j(t + 1) = X_{j-1}(t), \quad j = 2, \dots, m. \quad (7.2)$$

357 Combination of (2.1) and (3.3) yields the rule for extension

$$X_1(t + 1) = X_1(t)e^{-\sum_{j=1}^m \tilde{A}_j(X_{j+1}(t) - X_j(t))}. \quad (7.3)$$

358 In (7.3) it is harmless to allow \tilde{A}_j to depend on time t !

359 Alternatively we might start from (3.9) and choose as before

$$X_1(t) = s(t) \quad (7.4)$$

360 but for $j > 1$

$$X_j(t) = s(t + 1 - j) - s(t + 2 - j) \quad (7.5)$$

which corresponds to the incidence in time window $t + 1 - j$. This leads to the update rules

$$X_1(t + 1) = X_1(t)e^{-\sum_{k=1}^{\infty} X_{k+1}(t)\tilde{A}_k}, \quad (7.6)$$

$$X_2(t + 1) = X_1(t) - X_1(t + 1), \quad (7.7)$$

$$X_j(t + 1) = X_{j-1}(t) \text{ for } j > 2. \quad (7.8)$$

361 In this formulation too, we can allow \tilde{A}_k to depend on time t .

362 This seems a good moment to point out that the use of labels like ‘exposed’
363 or ‘infectious’ is perfectly possible within the general framework. For any such
364 label, say L , specify, on the basis of the choice of the parameters A_k as described
365 in Section 6, the probability π_j that an individual carries this label at time j
366 after becoming infected. Then the number of individuals carrying label L at
367 time t is given by

$$N_L(t) = \sum_{j=1}^N \pi_j X_{j+1}(t). \quad (7.9)$$

368 So all one needs to do to plot the time course of N_L , is to add to (7.6)-(7.8) the
369 equation (7.9) (with t replaced by $t + 1$, for consistency).

370 Note that (5.3) and (5.6) are examples of (7.9). In the very special situation
371 considered in Section 5, the labels actually correspond to states at the individual
372 level and as a consequence one can express $N_L(t + 1)$, for $L = S, E, I, R$ in
373 terms of these same quantities at time t , without reference to $X(t)$. In general

374 this is impossible. (Incidentally, note that probabilists often speak about non-
375 Markovian models when the labels refer to compartments and sojourn time
376 distributions are not exponential, while calling the labels ‘states’, even though,
377 strictly speaking, they do not qualify as such.)

378 8 About the peak of the incidence curve

379 An epidemic curve has many features, such as

- 380 • the initial growth rate ρ ;
- 381 • the height and timing of the peak;
- 382 • the final size.

383 For the first and last of these, it is well understood how they relate to the
384 parameters of simple models that ignore heterogeneity. For instance, the final
385 size is completely determined by R_0 , while ρ is a solution of the Euler-Lotka
386 equation, cf. (4.5).

387 At the start of an outbreak, one may observe the initial growth rate and
388 next use information about the generation interval to make inferences about
389 R_0 , see [17, 18] and the references given there. Next one may choose the model
390 parameters such that ρ and R_0 of the model correspond to the estimated values.

391 The ongoing outbreak of Covid-19 generates much interest in peaks, largely
392 because of concern that hospitals may be overwhelmed with patients, leading
393 to healthcare breakdown. As far as we know, there is no analytical method to
394 determine the height and timing of the peak from model parameters (except,
395 perhaps, in the oversimplified *SIR* system of differential equations). So one has
396 to rely on numerical calculations.

397 The key question addressed in this section is: how much is peak height
398 influenced by model details? Here, we systematically compare the discrete-time
399 *SEIR* model, described by (5.4) and corresponding to geometric distributions of
400 the length of the latent and infectious period, to a model with deterministic, i.e.,
401 fixed, duration of these periods and constant infectiousness during the infectious
402 period. Thus both types of model have three parameters. By restricting to
403 $R_0 = 2.5$ we fixed the infectiousness parameter in terms of the other two. We
404 calibrated the models by making sure that ρ and the mean length of the latent
405 period are the same, thus creating a one-to-one relationship between the two
406 parameters of one type of model and the two parameters of the other type of
407 model.

408 As initial condition we took a short history of decreasing fractions of sus-
409 ceptibles, reflecting an exponential increase in new cases at the rate ρ . We
410 computed the peak value of the incidence for both types as a function of the
411 two parameters and next their ratio. The results are depicted in Figure 2.

412 The main conclusions are:

- 413 • deterministic periods lead to higher peaks than geometrically distributed
414 periods;
- 415 • this is most prominent when the latent period is large and the infectious
416 period is small;

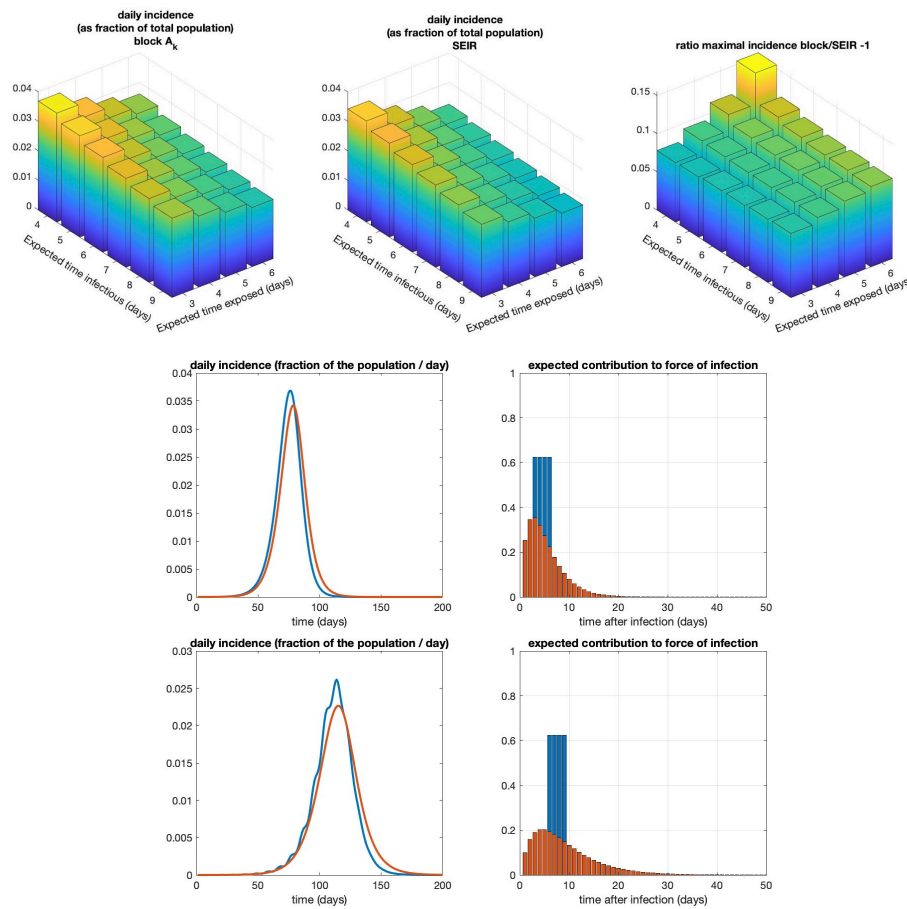


Figure 2: Comparing incidences between two types of models. In the block model, the lengths of the latent and infectious periods are deterministic (fixed for all individuals); in the $SEIR$ model, the lengths of these periods are stochastic (independently exponentially distributed with identical parameters, respectively, $1/T_E$ and $1/T_I$). See Appendix for details. Top row: maximum incidence of the block model (with deterministic periods; left), $SEIR$ model (with stochastic periods; middle) and the relative ratio between the two (i.e., $\frac{\text{block-}SEIR}{SEIR}$; right), as a function of T_E , the (actual resp. expected) time individuals are exposed and T_I , the (actual resp. expected) time individuals are infectious. Models were compared after ensuring that they have the same R_0 and initial speed ρ . Note that the incidence of the deterministic model always reaches a higher peak within the ranges of T_E and T_I considered, by about 8-15%, than the corresponding $SEIR$ model. Middle and bottom rows: example simulations with the deterministic model (blue) and corresponding $SEIR$ model (red) with the same R_0 and ρ . The middle row corresponds to the parameters at which the ratio of peak heights is minimal, $(T_E, T_I) = (3, 4)$; the bottom row to when this ratio is maximal, $(T_E, T_I) = (6, 4)$. One can clearly see that the incidence grows initially at the same rate.

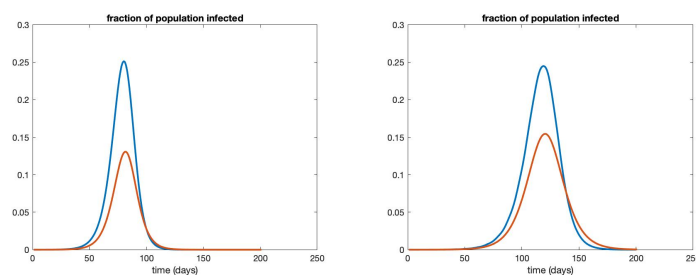


Figure 3: The fraction of the population that is either latently infected or infectious in the block model (blue) and the corresponding *SEIR* model. Left: $(T_E, T_I) = (3, 4)$; right: $(T_E, T_I) = (6, 4)$. The corresponding incidences can be found in Figure 2. Note that although the peak incidences are not so very different (see Fig. 2), there is a large difference in the fraction of infected-and-not-yet-removed individuals, due to the comparatively fatter tail in the expected future contribution to the force of infection in the *SEIR* model.

- the difference is, for reasonable parameter values, in the order of 10%.

(Note that, since compartmental models have fatter tails, they need, for given R_0 , to have an earlier peak of infectiousness in order to have the same ρ . This is clearly visible in Figures 2 and 4.)

After the first conclusion emerged, we aspired to find a somewhat mechanistic explanation. This led to the following observation. Roughly speaking, an outbreak reaches its peak when S is reduced to the level corresponding to $R_0 = 1$. How many more cases there will be after the peak depends largely on the number of individuals that are, or are on the way of becoming, infectious at the time the peak is reached. (In [6, Section 1.3.2] it is explained how the overshoot phenomenon corresponding to a large stock of recently infected individuals at the time of reaching the peak, causes the final size, as fraction of the population, to increase when R_0 increases.) For compartmental models, there is a relatively fat tail in the distribution of the time until becoming ‘removed’, i.e., having no future infectiousness. So when comparing models with the same R_0 , and hence the same final size, we should expect that for compartmental models the reservoir of latent and infectious individuals is less big, at peak-time, than for models in which expected future infectiousness reduces to zero after finite time. This is illustrated in Figure 3. And as reservoir size correlates with peak size, we should expect lower peaks for compartmental models, exactly as found in our numerical results.

Just to elucidate that the higher-peak-phenomenon matters in Covid-19 context, we chose, on the one hand, the parameters A_k as integrals over one day time-intervals of the Weibull generation-interval distribution as derived from data in [8] and, on the other hand, determined the one-parameter family of *SEIR* models that has both R_0 and ρ equal to these quantities for the Weibull. The results of a comparison are presented in Figure 4. The peak heights differ 5 to 10%.

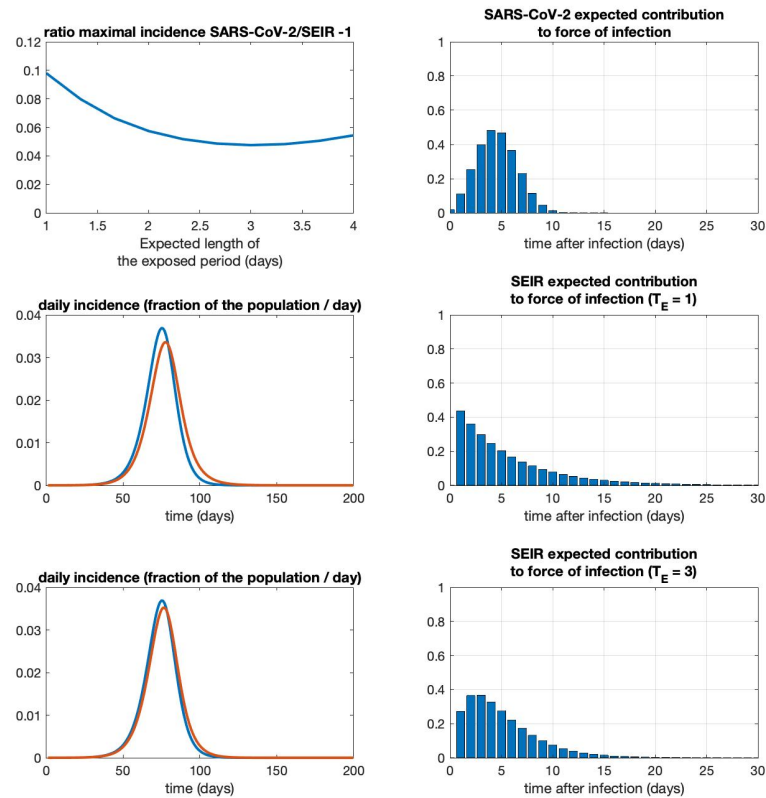


Figure 4: Comparing a model with parameters corresponding to the Weibull distribution derived from SARS-CoV-2 generation-interval data in [8] to the *SEIR* model with the same reproduction number R_0 and the same initial growth rate ρ . Top row: left: relative ratio between the maximal incidences (-1 , as in Figure 2) as a function of T_E , the expected length of the exposed period (note that the SARS-CoV-2 model does not depend on T_E , see the Appendix). right: expected contribution to force of infection for the SARS-CoV-2 model. Middle row: left: incidence of the SARS-CoV-2 model (blue) and the *SEIR* model (red) with $T_E = 1$. At this T_E , the relative difference in the peak incidences are maximal. right: the expected contribution of the *SEIR*-model for $T_E = 1$. Bottom row: same, but now for $T_E = 3$, at which the difference in peak incidence is minimal. For $T_E > 4.12$, the *SEIR*-model cannot be parameterized to have the same R_0 and ρ as the SARS-CoV-2 model. For further details, see the Appendix.

445 9 Conclusions

446 The success of the *SIR* and *SEIR* variants dwarfs the attention for the general
447 Kermack-McKendrick model from 1927, even though, in principle, the latter has
448 much on offer for a would-be modeller. We surmise that the reason is that the
449 general model is formulated as a renewal (or, Volterra integral) equation and
450 that for these unfamiliar equations there are no user friendly numerical tools
451 available.

452 Here we introduced a discrete time version that has many advantages:

- 453 • the generality and flexibility is retained;
- 454 • computing the epidemic time course is super easy;
- 455 • the time step can be adjusted to the time interval between data points
456 (e.g., one day or one week).

457 In Section 8 we showed that precise assumptions about the latent and infec-
458 tious period matter for predicting the peak of the incidence curve, a quantity of
459 interest from a public health perspective. So, we claim, the generality matters
460 for practical issues and is not just an academic fancy. Of course in a practical
461 context all kinds of heterogeneity (e.g., reflecting age) matter as well. These
462 have been neglected here, but in the text book [6] and in [3] they have received
463 ample attention in the continuous time setting, so it should not be too difficult
464 to incorporate them in the discrete time framework as well.

465 Our hope is that our pragmatic reformulation leads to, well-deserved and
466 long overdue [5], popularity of the true Kermack-McKendrick model.

467 Acknowledgement. It is a pleasure to thank Matthias Kreck for prompting
468 one of us to formulate the discrete time Kermack-McKendrick model. In turn,
469 it was the Covid-19 pandemic that sparked the interest of Matthias Kreck and
470 Erhard Scholz, see [14, 15].

471 Declaration of interest: none

472 References

- 473 [1] A. Bátkai, M. K. Fijavž, and A. Rhandi. *Positive Operator Semigroups*.
474 Birkhäuser Basel, 2017.
- 475 [2] F. Brauer, Z. Feng, and C. Castillo-Chavez. Discrete epidemic models.
476 *Math. Biosciences Eng. AIMS*, 7(1):1–15, 2010.
- 477 [3] D. Breda, O. Diekmann, W. F. de Graaf, A. Pugliese, and R. Vermiglio.
478 On the formulation of epidemic models (an appraisal of Kermack and McK-
479 endrick). *J. Biol. Dynamics*, 6(sup2):103–117, 2012.
- 480 [4] O. Diekmann. Limiting behaviour in an epidemic model. *Nonlin. Anal.*
481 *Theory and Appl.*, 1:459–470, 1977.
- 482 [5] O. Diekmann. The 1927 epidemic model of Kermack and McKendrick: a
483 success story or a tragicomedy? *Newsletter of the Japanese Society for*
484 *Mathematical Biology*, 92:8–11, 2020.

- 485 [6] O. Diekmann, H. Heesterbeek, and T. Britton. *Mathematical Tools for*
486 *Understanding Infectious Disease Dynamics*. Princeton University Press,
487 Princeton, NJ, 2012.
- 488 [7] O. Diekmann, M. Gyllenberg, and J. Metz. Finite dimensional state rep-
489 resentation of linear and nonlinear delay systems. *J. Dyn. Diff. Equat.*, 30:
490 1439–1467, 2018.
- 491 [8] L. Ferretti, C. Wymant, M. Kendall, L. Zhao, A. Nurtay, L. Abeler-Dörner,
492 M. Parker, D. Bonsall, and C. Fraser. Quantifying SARS-CoV-2 transmis-
493 sion suggests epidemic control with digital contact tracing. 368(6491):
494 eabb6936, 2020.
- 495 [9] P. Groeneboom and G. Jongbloed. *Nonparametric Estimation under Shape*
496 *Constraints*. Cambridge Series in Statistical and Probabilistic Mathematics.
497 Cambridge University Press, 2014.
- 498 [10] J. Heesterbeek. A brief history of R_0 and a recipe for its calculation. *Acta*
499 *Biotheoretica*, 50(3):189–204, 2002.
- 500 [11] J. A. P. Heesterbeek. The law of mass-action in epidemiology: a his-
501 torical perspective. In K. Cuddington and B. Beisner, editors, *Ecological*
502 *Paradigms lost: routes of theory change*, pages 81–105. San Diego, USA:
503 Academic Press, 2005.
- 504 [12] N. Hernandez-Ceron, Z. Feng, and C. Castillo-Chavez. Discrete epidemic
505 models with arbitrary stage distributions and applications to disease con-
506 trol. *Bull. Math. Biol.*, 75(10):1716–1746, 2013.
- 507 [13] W. O. Kermack and A. G. McKendrick. A contribution to the mathematical
508 theory of epidemics. *Proc. Roy. Soc. London A*, 115:700–721, 1927.
- 509 [14] M. Kreck and E. Scholz. Proposal of a recursive compartment model of
510 epidemics and applications to the covid-19 pandemic. arXiv:2009.00308v1,
511 2020.
- 512 [15] M. Kreck and E. Scholz. A discrete Kermack-McKendrick model and why
513 it is better adapted to Covid-19 than standard SIR. submitted, 2021.
- 514 [16] T. G. Kurtz. *Approximation of Population Processes*. Number 36 in CBMS-
515 NSF Regional Conference Series in Applied Mathematics. SIAM: Society
516 for Industrial and Applied Mathematics, 1981.
- 517 [17] S. W. Park, D. Champredon, J. S. Weitz, and J. Dushoff. A practical
518 generation-interval-based approach to inferring the strength of epidemics
519 from their speed. *Epidemics*, 27:12–18, 2019.
- 520 [18] S. W. Park, K. Sun, D. Champredon, M. Li, B. M. Bolker, D. J. D. Earn,
521 J. S. Weitz, B. T. Grenfell, and J. Dushoff. Forward-looking serial intervals
522 correctly link epidemic growth to reproduction numbers. *Proc. Nat. Acad.*
523 *Sciences USA*, 118(2):e2011548118, 2020.
- 524 [19] F. P. Pijpers. A non-parametric method for determining epidemiological
525 reproduction numbers. *J. Math. Biology*, 82:37, 2021.

526 [20] C. Robinson. *Dynamical Systems. Stability, Symbolic Dynamics and Chaos.*
527 CRC Press, 1995.

528 10 Appendix

529 10.1 Detailed description of numerical work displayed in 530 Figure 2

In Figure 2 we numerically compare two types of models that have the same initial rate of increase ρ and basic reproduction number R_0 but different expected contributions to the force of infection A_k . Both models are defined by prescribing the values of the rescaled expected contribution to the force of infection, \tilde{A}_k , where k is the number of days that have elapsed since becoming infected. Let us choose T_E , the number of days an individual is exposed but not yet infectious, and T_I , the number of days an individual is infectious. In the ‘block model’, in which these periods are assumed to be deterministic, so the same for all individuals, we set

$$\begin{aligned} \tilde{A}_k &= 0, & 1 \leq k \leq T_E, \quad k \geq T_E + T_I + 1, \\ \tilde{A}_k &= R_0/T_I, & T_E + 1 \leq k \leq T_E + T_I. \end{aligned}$$

531 In this way, $\sum_{k=1}^{\infty} \tilde{A}_k = R_0$. The initial exponential rate ρ with which the
532 epidemic spreads is the solution to (4.5).

533 With R_0 , ρ , and T_E , we can set up now the corresponding *SEIR* model.
534 Using the explicit expression of the A_k for the *SEIR* model (5.5), we can de-
535 termine the relation between α , β and γ on the one hand, and ρ on the other,
536 using (4.5),

$$\frac{1}{N} = \frac{\beta\gamma}{(\alpha + \rho - 1)(\gamma + \rho - 1)}. \quad (10.1)$$

537 From the expression (4.3) for R_0 , we obtain the familiar expression $R_0 = \frac{\beta N}{\alpha}$.
538 Substitution into (10.1) gives us

$$\alpha = \frac{(\rho + \gamma - 1)(\rho - 1)}{1 - \rho + \gamma(R_0 - 1)}. \quad (10.2)$$

539 Lastly, the link between γ and T_E is simply $\gamma = \frac{1}{T_E}$. So in all, T_E , ρ and R_0
540 define the following parameters for the *SEIR* model,

$$\begin{aligned} \alpha &= \frac{(\rho + \gamma - 1)(\rho - 1)}{1 - \rho + \gamma(R_0 - 1)}, \\ \beta &= \frac{R_0}{N} \frac{(\rho + \gamma - 1)(\rho - 1)}{1 - \rho + \gamma(R_0 - 1)}, \\ \gamma &= \frac{1}{T_E}. \end{aligned} \quad (10.3)$$

541 so that the \tilde{A}_k are given by

$$\tilde{A}_k = \frac{R_0}{T_E} \frac{(\rho + \gamma - 1)(\rho - 1)}{1 - \rho + \gamma(R_0 - 1)} \sum_{l=1}^{k-1} \left(1 - \frac{1}{T_E}\right)^{l-1} \left(1 - \frac{(\rho + \gamma - 1)(\rho - 1)}{1 - \rho + \gamma(R_0 - 1)}\right)^{k-l-1}.$$

542 (The reader may verify that with this choice $\sum_{k=1}^{\infty} \tilde{A}_k = R_0$, and that the
543 characteristic equation (4.5) is satisfied exactly at $\lambda = \rho$, as required.)

544 The numerical simulations were carried out using (3.9), using an initial con-
545 dition in which the epidemic has started increasing at rate ρ for six days:

$$s(t) = 1 - 0.00001\rho^{t+6}, \quad t = -1, -2, \dots, -6.$$

546 **10.2 Figure 3: Comparing the *SEIR* model with a SARS- 547 **CoV-2 model****

548 In [8], the generation interval distribution $g(\tau)$ is approximated by a Weibull
549 distribution with shape parameter 2.826, and scale parameter 5.665. We dis-
550 cretized this,

$$g_k = \int_{k-1}^k g(\tau) d\tau, k = 1, 2, \dots$$

551 Then we set $\tilde{A}_k = R_0 g_k$, $k = 1, 2, \dots$. The initial rate of increase is again
552 estimated using 4.5, and gives $\rho = 1.1919$.

553 The corresponding *SEIR*-model is now given by using α , β and γ as in
554 (10.3) and defining the \tilde{A}_k as before.

555 The initial condition is the same as in the previous illustration.