



Research



Cite this article: Mursel S, Bocchini P, Buceta J.

2024 Stochastic analysis of Ebola infection in small zoonotic niches. *R. Soc. Open Sci.* **11**: 240298.

<https://doi.org/10.1098/rsos.240298>

Received: 20 February 2024

Accepted: 8 October 2024

Subject Category:

Mathematics

Subject Areas:

health and disease and epidemiology, ecology, mathematical modelling

Keywords:

stochastic SIR model, Ebola, zoonosis, mathematical epidemiology

Author for correspondence:

Javier Buceta

e-mail: javier.buceta@csic.es

Stochastic analysis of Ebola infection in small zoonotic niches

Sena Mursel^{1,2}, Paolo Bocchini^{1,2} and Javier Buceta³

¹Department of Civil and Environmental Engineering, and ²Catastrophe Modeling Center, Lehigh University, Bethlehem, PA, USA

³Theoretical and Computational Systems Biology Program, Institute for Integrative Systems Biology (I2SyBio), CSIC-UV, Paterna, Spain

SM, 0000-0002-1021-274X; PB, 0000-0002-5685-2283; JB, 0000-0003-1791-0011

The size of fruit bat colonies ranges from dozens to hundreds of thousands of individuals, depending on the species. While a deterministic modelling approach is appropriate for large colonies, the role of population fluctuations can be all-important for small colonies. From this perspective, we analyse the infection dynamics in small zoonotic niches due to filoviruses, e.g. Ebola. To this end, we perform stochastic numerical simulations and analytical calculations. The inherent stochasticity in ecological processes may play a significant role in driving small populations towards extinction. Here, we reveal that fluctuations can either lead to virus eradication or to sustain infection compared with the deterministic dynamics, depending on the size of the zoonotic niche. Altogether, our findings reveal non-trivial stochastic effects, which can shed light on the infection dynamics in small- and medium-sized bat colonies and help design preventive measures for zoonotic diseases.

1. Introduction

Infectious diseases exert a major burden on global public health and the global economy, especially in low-income countries [1,2]. In particular, diseases transmitted to humans from animal carriers, i.e. zoonoses, such as SARS, MERS, avian influenza, COVID-19 and Ebola virus disease (EVD), account for approximately 70% of the emerging infectious diseases [3]. In the case of EVD and other filovirus diseases (e.g. Marburg), fruit bats are believed to be one of the main reservoirs [4]. The first-ever recorded spillover of Ebola into humans occurred in Zaire (currently Democratic Republic of Congo) in 1976. Since then, and for more than four decades, a significant number of outbreaks have taken place in sub-Saharan Africa [5]. The

2013–2015 outbreak of EVD in West Africa has been the most severe outbreak so far and caused approximately 11 000 deaths [6]. Moreover, EVD is lethal among non-human primates like apes and monkeys, as well as other mammals, like antelopes, and during outbreaks a significant portion of the wildlife population is wiped out [7]. Thus, it is paramount to develop quantitative predictive tools to mitigate the devastating effects of zoonoses and, in particular, EVD.

Numerous models have been proposed in the context of infectious diseases to understand the conditions that lead to their propagation [8,9]. One particular framework that has been extensively used is compartmental models, where populations are divided into different subcategories depending on their status with respect to the disease, e.g. susceptible, infected, recovered and so on [10]. In that context, during the past years, several compartmental models have been proposed to understand the spreading of filoviruses from an ecological viewpoint, i.e. among the reservoir population [11–14]. However, these models usually consider that bat populations are ‘large’ and consequently the disease progression follows a deterministic behaviour. In other words, the common goal of these models is to capture the average dynamics and, for the law of large numbers, the overall behaviour of populations with a large number of individuals will be very similar to the average behaviour, which may be captured well by a deterministic model, especially if the problem is close to being linear.

The assumption of large enough colonies for the law of large numbers to apply is indeed well justified for some species of fruit bats (e.g. *Eidolon*) where the colonies comprise hundreds of thousands of individuals [15]. However, some species of fruit bats (e.g. *Hypsignathus monstrosus*) congregate in colonies that include as few as a dozen individuals [16]. In small colonies, the law of large numbers does not apply with the same force and the specific behaviour of each individual can significantly skew the whole population dynamics. For this reason, deterministic models are not appropriate, and stochastic models that can capture random fluctuations become necessary. As a matter of fact, the effect of internal fluctuations has been revealed to be of great significance in other zoonotic diseases, such as Hantavirus [17–22]. The size of what can be defined as a ‘small’ population of bats varies depending on the species and its ecological context. For example, the little brown bat roosts in hibernating colonies comprising up to 10^5 individuals, although the average colony size is approximately 10^4 [23,24]. Also, in the context of population recovery, a study mentioned approximately 10 individuals per cluster when almost approximately 10^3 bats were present in the hibernaculum [25]. In this study, for the sake of establishing a clear and convenient categorization of population sizes, we define the reference sizes for ‘small’, ‘medium’ and ‘large’ populations as 10, 100 and 1000 individuals, respectively.

Herein, we propose a stochastic model of EVD propagation in bat colonies of different sizes that we analyse by means of numerical simulations and analytical approximations. By comparing the results with a deterministic version of the model, we conclude that fluctuations can lead to virus eradication for some trajectories, and therefore delay on average the onset of infection. Contrary to intuition, we also reveal that stochasticity can sustain infection with respect to the deterministic system depending on the size of the zoonotic niche. Finally, our article elaborates on the consequences of our results for designing preventive measures and preparedness against EVD and other zoonotic diseases.

2. Methods

2.1. Deterministic and stochastic SIR models

In this work, we follow the main ideas presented in the studies by Buceta and Johnson [13] and Fiorillo *et al.* [14] to model EVD epidemiology in bats’ zoonotic niches. In the model, the carrying capacity, K , represents the maximum number of individuals that the available resources at a given location can support. The model is schematically represented in figure 1. The different processes considered in the model among susceptible (S), infected (I), recovered (R) states can be represented in terms of the following transitions:

$$\text{Birth: } Z \xrightarrow{b} Z + S, \quad (2.1)$$

$$\text{Death: } Z \xrightarrow{\mu} \emptyset, \quad (2.2)$$

$$\text{Infection: } S + I \xrightarrow{\beta} I + I, \quad (2.3)$$

$$\text{Recovery: } I \xrightarrow{\gamma} R, \quad (2.4)$$

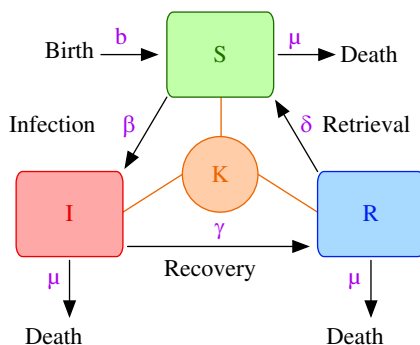


Figure 1. Schematic representation of the SIR model of EVD in a bat population. Bats are assumed to be born in a susceptible state (S) and after being infected (I) they can recover from the infection (R). Magenta letters indicated the nomenclature for the transition rates associated with the different transition events. Death rates, denoted as μ , are assumed to be the same regardless of the infection state. Competition for resources sets a carrying capacity K .

$$\text{Retrieval: } R \xrightarrow{\delta} S, \quad (2.5)$$

$$\text{Competition: } N + Z \xrightarrow{\frac{|b - \mu|}{K}} N, \quad (2.6)$$

where Z stands for S , I or R (i.e. the number of bats in compartment Z) and $N = S + I + R$ (i.e. the total number of individuals in compartments S , I and R). The transition rates (i.e. the probability of events per unit of time), e.g. b , are assumed to be constant and need to be calibrated [14], as discussed later. Assuming that the total number of bats, $N \gg 1$ (i.e. a large population) a mass-action law approach leads to the deterministic SIR compartmental model [26,27] described in equations (2.7)–(2.9).

$$\dot{S} = bN - \mu S - \frac{|b - \mu|}{K} NS - \beta SI + \delta R, \quad (2.7)$$

$$\dot{I} = -\mu I - \frac{|b - \mu|}{K} NI + \beta SI - \gamma I, \quad (2.8)$$

$$\dot{R} = -\mu R - \frac{|b - \mu|}{K} NR + \gamma I - \delta R. \quad (2.9)$$

Consequently, from a deterministic perspective, the total population experiences logistic growth,

$$\dot{N} = (b - \mu)N - \frac{|b - \mu|}{K} N^2, \quad (2.10)$$

and at the steady state the population disappears if the birth rate is smaller than the death rate ($N = 0$ if $b < \mu$) or the population reaches the carrying capacity if the birth rate is larger than the death rate ($N = K$ if $b > \mu$). As previously shown (see [13,14]), in the deterministic model the infection is sustained in the population if $K > K_c = \frac{b + \gamma}{\beta}$ or, alternatively, if the basic reproduction number, $\mathcal{R}_0 = \beta K - b + \mu/\gamma + \mu = \beta/\gamma + \mu(K - K_c) + 1 > 1$ [13,14]. That is, due to the logistic growth of the population, for a fixed set of parameter values, $\mathcal{R}_0$ increases linearly with K .

When the population is small and fluctuations cannot be neglected, the set of transitions in equations (2.1)–(2.6) can be described by the following forward Kolmogorov equation (i.e. master equation) [28,29]:

$$\begin{aligned}
\dot{P}_{(S,I,R)} = & b(N-1)P_{(S-1,I,R)} + (S+1) \\
& \times \left(\mu + \frac{|b-\mu|}{K}N \right) P_{(S+1,I,R)} \\
& + (I+1) \left(\mu + \frac{|b-\mu|}{K}N \right) \\
& \times P_{(S,I+1,R)} + (R+1) \\
& \times \left(\mu + \frac{|b-\mu|}{K}N \right) P_{(S,I,R+1)} \\
& + \beta(S+1)(I-1)P_{(S+1,I-1,R)} \\
& + \gamma(I+1)P_{(S,I+1,R-1)} \\
& + \gamma(R+1)P_{(S-1,I,R+1)} \\
& - \left\{ \left(b + \mu + \frac{|b-\mu|}{K}(N-1) \right) N + \beta SI \right. \\
& \left. + \gamma I + \delta R \right\} P_{(S,I,R)},
\end{aligned} \tag{2.11}$$

where $P_{(S,I,R)}$ is the probability of having S , I and R individuals at time t in the various compartments of the population.

As for the total population of bats, it is described by the transitions,

$$N \xrightarrow{b} N + N, \tag{2.12}$$

$$N \xrightarrow{\mu} \emptyset, \tag{2.13}$$

$$N + N \xrightarrow{\frac{|b-\mu|}{K}} N \tag{2.14}$$

and the master equation for the probability of having $N = S + I + R$ individuals at time t , P_N , reads

$$\begin{aligned}
\dot{P}_N = & b(N-1)P_{N-1} + (N+1) \\
& \times \left(\mu + \frac{|b-\mu|}{K}N \right) P_{N+1} \\
& - \left(b + \mu + \frac{|b-\mu|}{K}(N-1) \right) N P_N
\end{aligned} \tag{2.15}$$

2.2. Moment hierarchy breaking approach

There is no analytical solution for the master equations (2.11) and (2.15). However, the exact simulation of individual stochastic trajectories can be obtained by means of the Gillespie algorithm as discussed later. Unfortunately, the practical application of this approach sometimes leads to a high computational burden, because an accurate assessment of the probability distribution requires a very large number of samples to simulate.

Here, we implement the moment hierarchy breaking (MHB) approach, also referred to as the moment closure method, to obtain approximate analytical solutions for the first moments of the joint probability mass function $P_{(S,I,R)}$ and the probability mass function P_N [30–32]. Our approach is based on the fact the cumulants, κ , are a way to characterize the probability distribution alternative to the calculation of its moments. Given a random variable X characterized by the moment generating function $M_X(t)$, its cumulant-generating function $K_X(t)$ reads [33]

$$K_X(t) = \log(M_X(t)) = \log(e^{tX}).$$

Then, the cumulants κ_n are derived from a power series expansion of $K_X(t)$,

$$K_X(t) = \sum_{n=1}^{\infty} \kappa_n \frac{t^n}{n!} = \kappa_1 \frac{t}{1!} + \kappa_2 \frac{t^2}{2!} + \dots$$

Thus, the n th derivative of the previous expression evaluated at zero defines the n th-order cumulant,

$$\kappa_n = \left. \frac{d^n K_X(t)}{dt^n} \right|_{t=0}.$$

Cumulants encapsulate the most significant statistical information of the respective random variables through their moments [34]. In fact, two probability distributions whose moments are identical will have identical cumulants and vice versa [35,36]. The first moments, μ_i and cumulants, κ_i , satisfy the following relationships:

$$\begin{aligned}\kappa_1(X) &= \mu_1(X) = \langle X \rangle, \\ \kappa_2(X) &= \mu_2(X) - \mu_1^2(X) = \langle X - \langle X \rangle \rangle^2, \\ \kappa_3(X) &= \mu_3(X) - 3\mu_2(X)\mu_1(X) + 2\mu_1^3(X) = \langle X - \langle X \rangle \rangle^3,\end{aligned}$$

where $\langle \cdot \rangle$ denotes the statistical average or expectation operator. Thus, in the particular case of a Gaussian distribution with mean μ and variance σ^2 the following holds:

$$\kappa_1(X) = \mu, \quad \kappa_2(X) = \sigma^2, \quad \kappa_n(X) = 0, \quad n \geq 3.$$

In the multivariate scenario, a similar rationale applies. Thus,

$$\begin{aligned}\kappa_1(X) &= \kappa(X) = \langle X \rangle, \\ \kappa_2(X, Y) &= \kappa(X, Y) = \langle XY \rangle - \langle X \rangle \langle Y \rangle, \\ \kappa_3(X, Y, Z) &= \kappa(X, Y, Z) = \langle XYZ \rangle - \langle XY \rangle \langle Z \rangle \\ &\quad - \langle XZ \rangle \langle Y \rangle - \langle YZ \rangle \langle X \rangle \\ &\quad + 2\langle X \rangle \langle Y \rangle \langle Z \rangle,\end{aligned}$$

where in the context of this study, X , Y or Z stand for S , I or R and $\langle \cdot \rangle$ denotes in this case the multivariate statistical average or expectation operator,

$$\langle f(S, I, R) \rangle = \sum_{S, I, R} f(S, I, R) P_{(S, I, R)}.$$

When considering the total population, N , the definition of the cumulants and the procedure explained in the following are straightforward because in this case $X = Y = Z = N$.

We then build an approximate solution by assuming that there exist the probability P^{approx} for which the third-order cumulants are null, i.e. $\kappa(X, Y, Z) = 0$, and consequently third-order correlations can be approximated as

$$\begin{aligned}\langle XYZ \rangle &= \langle XY \rangle \langle Z \rangle + \langle XZ \rangle \langle Y \rangle \\ &\quad + \langle YZ \rangle \langle X \rangle - 2\langle X \rangle \langle Y \rangle \langle Z \rangle.\end{aligned}$$

This procedure, as shown in the appendix, allows us to break the hierarchy of moments and obtain closed-form ordinary differential equations for the moments $\langle X \rangle$ and $\langle XY \rangle$. Consequently, the relationship $\kappa(X, Y, Z) = 0$ implicitly assumes a Gaussian-like property for the probability P^{approx} [34]. Notice that pushing the approximation further and imposing a moment closure relationship at the level of the second-order cumulants, i.e. $\kappa(X, Y) = 0$, would oversimplify the probabilistic description, lead to the deterministic solution, and consequently, the probability density would be approximated in that case by $P^{\text{approx}} = \delta(X - \langle X \rangle) \delta(Y - \langle Y \rangle) \delta(Z - \langle Z \rangle)$. On the other hand, implementing the closure relationship at the level of the fourth-order cumulant makes the analytical calculations cumbersome and increases unnecessarily the complexity of the problem since, as shown below, the Gaussian approximation adequately captures the non-trivial stochastic effects. As a cautionary note, we stress that the Gaussian approximation, while accurate, is nonetheless unphysical since the described variables, i.e. populations, are implicitly allowed to take negative values. This fact can be particularly problematic when describing populations close to zero. Still, we note that when solving the MHB relationships for the statistical moments (see appendix), only non-negative solutions are allowed.

2.3. Numerical methods and parameters

Ordinary differential equations were solved using the fourth-order Runge–Kutta method as implemented in Matlab's embedded library [37–39]. The simulations of the stochastic dynamics were performed using the Gillespie algorithm, in particular, using the direct method [40,41]. Numerical trajectories were simulated for various time windows spanning from one month to 50 years (see §3.3). As for the initial condition of our simulations, for every value of the carrying capacity, K , we explored

10^3 random sets of initial conditions that satisfy the condition $S + I + R = K$ with S , I and R greater than zero. We concluded that, as expected, as long as a large number of trajectories (i.e. stochastic realizations) are explored (see below) the ergodicity of the system ensures that a single initial condition is enough to capture the stationary probability density in full. On the other hand, when characterizing transient behaviours, averaging of different sets of initial conditions is key. In our case, to examine steady-state results, we typically used as initial condition 70%, 10% and 20% for susceptible, infectious and recovered states, respectively. Given that the maximum average lifespan of some bat species (e.g. *Rousettus aegyptiacus*) can be as long as 25 years (see parameter list below), the 50-year simulation time frame is deemed sufficiently comprehensive to explore all relevant time scales in the model at the steady state [42,43]. To analyse statistical moments at the steady state, we computed averages over 10 000 stochastic trajectories. The probability of sustained infections was defined as the ratio of trajectories resulting in infected bats at the end of the analysed time window to the total number of trajectories. In our simulations, we used the following parameter set, based on previous studies [14,43–45]. In particular, taking into account the range of variability described in [14], we used the mean values: $b = 1/365 \text{ day}^{-1}$, $\mu = 1/(27 \times 365) \text{ day}^{-1}$, $\gamma = 10/365 \text{ day}^{-1}$, $\beta = 3.9 \times 10^{-4} \text{ day}^{-1}$ and $\delta = 11/365 \text{ day}^{-1}$.

In our simulations, we also explored the effect of the duration of the observation window, T . This parameter dictates the duration of the simulation and it relates to how long it is reasonable to consider a colony as an isolated system. For example, through migration, the influx and efflux of individuals can change the epidemiological conditions of a colony. In this study, we considered the sensitivity of the results to the values of the observation window, T , ranging from one month (transient effects) to 50 years (steady state) using different sets of initial conditions as discussed above.

3. Results

3.1. Small populations are vulnerable to extinction due to stochasticity

In order to check the applicability of the MHB approach (see §2 and appendix), we first examined its results for the whole bat population, $N = S + I + R$, which should display a logistic growth behaviour, as shown in equation (2.10). Also in this case, the stochastic model is described by the transitions in equations (2.12)–(2.14) that define the master equation (2.15).

We performed stochastic simulations (see §2) for different values of the carrying capacity, K . For K , we only considered integer values because the total population of bats is the number of individuals. When $K = 1000 \gg 1$, as expected, $\langle N \rangle_{K \gg 1} = K$, i.e. as the carrying capacity increases, the average number of bats tends to the stationary deterministic solution: $N = K$. Figure 2a shows 10 stochastic trajectories as well as the analytical values for $\langle N \rangle \pm \sigma_N$ and the probability distribution obtained in the simulations and by the MHB approach.

An analysis of the 10 000 trajectories reveals that the statistics for the first two moments obtained with the MHB method are indeed in excellent agreement with the analytical calculations for different values of K (table 1). In the table 1, the stochastic effects are quantified through the coefficient of variation (CoV), $c_V^N = \sigma_N / \langle N \rangle$, that according to the analytical calculations (appendix) reads

$$c_V^N = \sqrt{\frac{\left(1 + K - \sqrt{1 + K\left(2 - \frac{8b}{b-\mu} + K\right)}\right)}{\left(3 + 3K + \sqrt{1 + K\left(2 - \frac{8b}{b-\mu} + K\right)}\right)}}. \quad (3.1)$$

Thus, $\lim_{K \rightarrow \infty} c_V^N = 0$ and, as expected, the stochastic effects are negligible for large populations. As shown

in equation (3.1), c_V^N incorporates square roots and we emphasize that c_V^N needs to be a real and positive value by definition. Furthermore, K can only take integer values. Thus, equation (3.1) is explicitly defined for K values greater than or equal to a predetermined lower bound, K_{lower} . This lower limit,

K_{lower} is determined as $\lceil \frac{3b + \mu + \sqrt{8b(b + \mu)}}{b - \mu} \rceil$ where $\lceil \cdot \rceil$ represents the ceiling function. For instance, in our simulations and the specified values of b and μ (§2), the lower bound for K becomes 7. This indicates that the expression yields a complex value when K is less than 7. For the same values of b and μ in our simulations, c_V^N can approach values near 0.5 for K close to K_{lower} and the trend

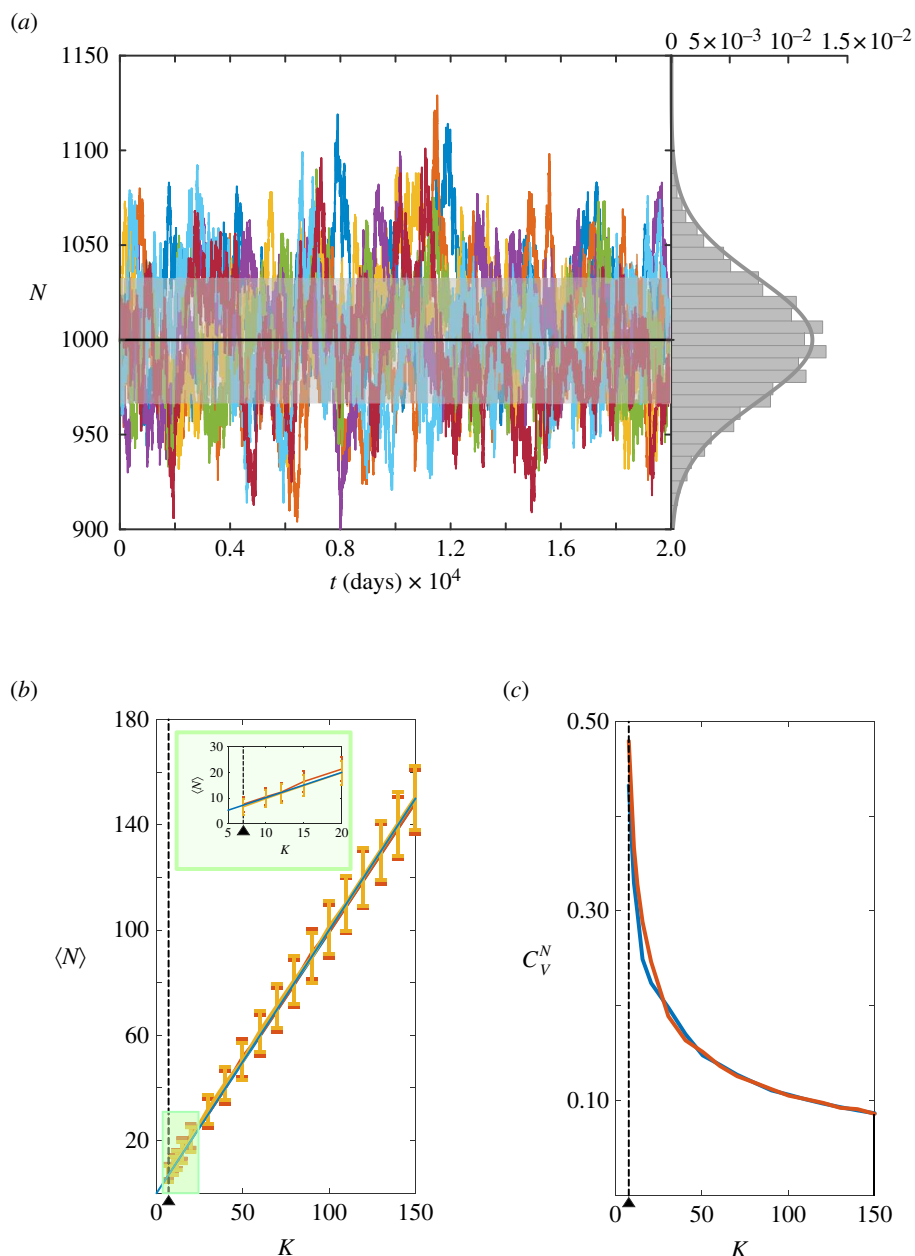


Figure 2. (a) The rugged colour lines show different stochastic trajectories for the total number of bats for $K = 1000$. The black line with the grey error band indicate the analytical calculation for $\langle N \rangle \pm \sigma_N = 1000 \pm 33$. On the right, we show the probability density of N as obtained in numerical simulations (grey bars) and the Gaussian distribution derived from the MHB theory. (b) Comparison between numerical simulations (yellow), analytical calculations (orange) and the deterministic solution (blue) about the expected value of $\langle N \rangle$ at the stationary state as a function of K . Simulated time 20 years. The error band accounts for the standard deviation, σ_N . (c) Coefficient of variation in numerical simulations (blue) and analytical calculations (orange) as a function of K . As K increases c_V^N decreases and it is kept at low values thus precluding extinction events. The vertical dotted lines and the black triangles indicate the K_{lower} value.

for c_V^N values consistently decreases as K increases. Consequently, for K values smaller than K_{lower} , even without an analytical expression, the observed trend suggests an increase in the coefficient of variation. A coefficient of variation of more than 50% for small populations is very substantial and these results, indeed, indicate that stochasticity has a notable effect over the whole population and that fluctuations can lead to systematic extinction events in small niches. As an illustration, when K is set to 12, representing a reasonable colony size for *H. monstrosus*, the CoV is calculated to be 31.4% [46]. In fact, based on our simulations, the observed CoV in this specific case is 30.2% (see table 1).

Table 1. Comparison between analytical calculations and the values for $\langle N \rangle$, σ_N and c_V^N obtained in numerical simulations for different values of K (10^4 stochastic samples).

K	analytical calculations			numerical simulations		
	$\langle N \rangle$	σ_N	c_V^N	$\langle N \rangle$	σ_N	c_V^N
7	6.60	3.04	0.46	7.55	2.90	0.38
10	9.79	3.44	0.35	11.80	3.44	0.29
12	11.83	3.72	0.31	12.45	3.75	0.30
50	49.94	7.28	0.15	51.51	7.19	0.14
100	99.95	10.24	0.10	101.45	10.18	0.10
500	499.95	22.81	0.045	501.61	22.66	0.045
1000	999.96	32.24	0.032	1001.80	31.96	0.032
5000	5000.00	72.06	0.014	5000	71.98	0.014

To evaluate the agreement between the analytical approximations and numerical results, we also ran simulations in small- to medium-sized niches: $K \in (7, 150)$, [figure 2b](#). The results reveal that (i) $\langle N \rangle$ follows the deterministic behaviour, (ii) the MHB approach and the simulations are in agreement, and (iii) the probability of extinction becomes essential for small populations.

3.2. Fluctuations result in a higher probability of virus eradication in zoonotic niches

From a deterministic viewpoint, an analysis of the transition towards the infective state indicates the existence of a threshold value of the carrying capacity, K_c (or equivalently as a function of the basic reproduction number $\mathcal{R}_0$, see §2) [13]. If $K < K_c = (b + \gamma)/\beta$, i.e. if $\mathcal{R}_0 < 1$, (where β is the infection rate) at steady state the population is free of infection and $S = K$. On the other hand, if the size of the zoonotic niche, K , is larger than K_c , i.e. if $\mathcal{R}_0 > 1$, the infection propagates and $S = \frac{b + \gamma}{\beta}$, $I = \frac{b + \delta}{b + \gamma + \delta}(K - K_c)$ and

$R = \frac{\gamma}{b + \gamma + \delta}(K - K_c)$. Notice that regardless of the infective state (i.e. free of infection or not) at steady state $N = S + I + R = K$.

In order to evaluate the effect of the population fluctuations that originate from a stochastic description, we first analyse the results of the MHB approach (see §2). Interestingly, the analytical calculations predict that the average population of susceptible bats $\langle S \rangle$ increases as K for values of K slightly larger than K_c , $K \gtrsim K_c$, as seen in [figure 3a](#). Consequently, since the total average bat population in the steady state is conserved also in the stochastic model (i.e. $\langle N \rangle = \langle S + I + R \rangle = K$), it can be concluded that the fluctuations delay, on average, the appearance of infection compared with the deterministic model. Such a delay can also be observed by studying the correlations between S , I and R (appendix, [figure 6](#)). Also, as expected, when K is higher, the stochastic solution approaches the deterministic solution, $\langle S \rangle \simeq (b + \gamma)/\beta = K_c \simeq 77$. As a result, the infection-buffering effect induces a non-monotonic behaviour of $\langle S \rangle$. Finally, we performed stochastic simulations to validate the aforementioned theoretical results, [figure 3a](#). The simulations confirm the MHB theoretical predictions qualitatively and quantitatively and we can conclude that the fluctuations of the bat population lead to a non-null probability of eradication of the Ebola infection in medium size populations.

In order to shed light on the observed behaviour, we computed the probability density of susceptible, infected and recovered bats for values of the carrying capacity lower, similar to and larger than the threshold K_c ([figure 3b](#)). The average values of S , I and R approach their deterministic values when K is either significantly smaller or larger than K_c . However, for K similar to K_c the probability density reveals the discrepancies shown in [figure 3a](#) and the reason underlying such a mismatch. Given the low numbers of infected and recovered bats, near K_c the fluctuations of the population drive their extinction and the population splits in two categories. Survival trajectories of I and R satisfy, on average, the deterministic behaviour. However, a number of their trajectories become extinct and consequently drive the average of the probability density towards lower values. Since $I = R = 0$ is an

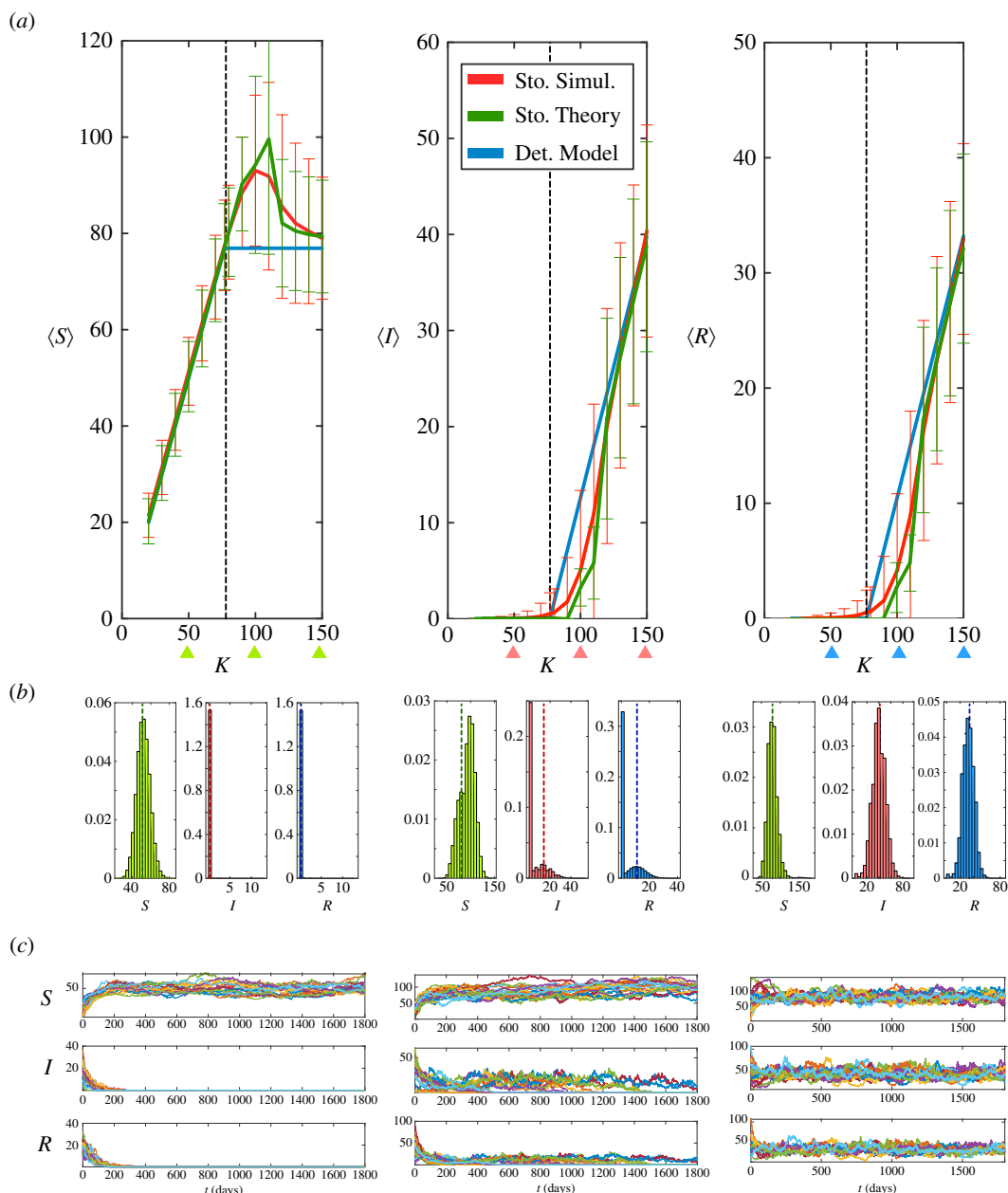


Figure 3. (a) From left to right, the panels show the stationary solution (solid lines) of the average susceptible $\langle S \rangle$, infected $\langle I \rangle$ and recovered $\langle R \rangle$ bats as a function of the carrying capacity (i.e. size of the zoonotic niche), K , as predicted by stochastic simulations (red curve), the MHB theoretical calculations (green curve) and the deterministic model (blue curve). The goodness of the MHB theory is also noticeable by the estimation of the second moments as reflected by the errors bars (standard deviation). The vertical black dashed line indicates the critical value of the carrying capacity, $K_c \approx 77$, as predicted by the deterministic model. Parameter values as indicated in the main text. In the stochastic simulations 10^3 realizations were considered (simulation time 50 years). (b) Normalized histograms (i.e. probability density) of susceptible, infected and recovered bats as obtained in the stochastic simulations for $K = 50$ (left), $K = 100$ (centre) and $K = 150$ (right) (triangles in panel (a)). The vertical dotted lines indicate the expected deterministic value. (c) Stochastic trajectories ($n = 20$, randomly selected) of S , I and R as a function of time and starting from different initial conditions. From left to right the values of K as indicated in panel (b).

absorbing state and the total average numbers of bats is conserved, the amount of susceptible bats necessarily increases when the infection is eradicated. As for the possible effect of the initial condition, figure 3c shows stochastic trajectories for the values of K analysed in figure 3b (50, 100 and 150). These trajectories were generated using random initial conditions (but ensuring $(S, I, R) > 0$ and $S + I + R = K$, see §2). As expected, the initial condition is irrelevant at the steady state since the system is ergodic.

3.3. The probability of sustained infections reveals stochastic effects as a function of the observation window

Given the above results, we explored deeper how infection in a colony is modulated by fluctuations. To that end, we computed the probability of sustained infections, $P_{\text{Inf.}}$. Namely, the probability that in a stochastic trajectory of duration T (i.e. an observation window of duration T), $I \neq 0$ at time T (see §2).

Figure 4a shows $P_{\text{Inf.}}$ for different values of T . On the one hand, the results reveal, in agreement with figure 3, that if $K > K_c$ the fluctuations lead to an increased probability of virus eradication. Namely, if $K > K_c$, $P_{\text{Inf.}} = 1$ in the deterministic system, but in the stochastic approach the infection is reduced and $P_{\text{Inf.}} < 1$. Also, as the observation window, T , increases the stochastic effects become more evident because larger values of the zoonotic niche, K , are needed for the infection to survive with respect to the deterministic behaviour for which $P_{\text{Inf.}} = 1$ if $K > K_c$ independently of the value of T .

Figure 4a also suggests that in the limit for $T \rightarrow \infty$ (i.e. at the steady state) there exist a stochastic critical threshold of the carrying capacity, $K_c^{\text{sto.}}$, below which the infection dies out. To test this hypothesis, we determined the stochastic transition points, K^* (symbols in figure 4b), as a function of T by fitting the curves $P_{\text{Inf.}}(K)$ to a Hill sigmoid function for each investigated value of T (equation (3.2)).

$$P_{\text{Inf.}}(K) = \frac{K^n}{(K^*)^n + K^n}. \quad (3.2)$$

As expected, figure 4b reveals that, as T increases, the value of K^* increases and approaches an asymptote that defines the stochastic critical threshold of the carrying capacity, $K_c^{\text{sto.}}$. In order to determine $K_c^{\text{sto.}}$, we fitted our data to the exponential behaviour given in equation (3.3),

$$K^*(T) = K_c^{\text{sto.}}(1 - e^{-\epsilon T^\zeta})^\eta \quad (3.3)$$

and found that the best fit is given by values $\epsilon = 13.94$, $\zeta = 0.0475$, $\eta = 3.5 \times 10^6$ and $K_c^{\text{sto.}} = 135$ (black solid line in figure 4b). Thus, our data indicate that, at the steady state, the stochasticity increases the size of the zoonotic niche required to sustain infection by approximately 75% with respect to the deterministic value. We also note that $K_c^{\text{sto.}}$ determines the value of the carrying capacity at which the stochastic and deterministic solutions for the means, $\langle S \rangle$, $\langle I \rangle$ and $\langle R \rangle$, begin converging.

On the other hand, we observed that if $K < K_c$ the fluctuations, depending on the value of T and K , may either help to sustain or to eradicate the infection with respect to the deterministic behaviour (figure 4a). In fact, the probability to sustain infections at a time T in the deterministic system can be estimated as (see details in appendix)

$$P_{\text{Inf.}}^{\text{det.}} \simeq \theta(K - K_c) + \theta(K_c - K)e^{-T \cdot \beta(K_c - K)}, \quad (3.4)$$

where $\theta(\cdot)$ is the Heaviside step function. When compared with the data from stochastic simulations, the results suggest that when T is small enough (e.g. $T =$ one month) the stochasticity sustains infection. That is $P_{\text{Inf.}}^{\text{sto.}} > P_{\text{Inf.}}^{\text{det.}}$. However, as the observation window, T , increases (e.g. $T = 1$ year) the fluctuations contribute to mitigate the infection as when $K > K_c$. We notice, however, that the deterministic estimate, equation (3.4), is based on a linear approximation, and consequently, a precise quantitative comparison is not possible (§4).

3.4. The extinction time for infection diverges as the carrying capacity approaches the stochastic critical threshold

To further assess the dynamics of infection extinction depending on the size of the zoonotic niche, K , we estimated the mean time for the infection to die $\langle \tau_{\text{ext.}} \rangle$, figure 5a. To that end, we use different observation windows of duration T . In other words, given a particular observation window, T , and a size of the zoonotic niche, K , we evaluated the average time it takes for a trajectory to reach $I = 0$. Thus,

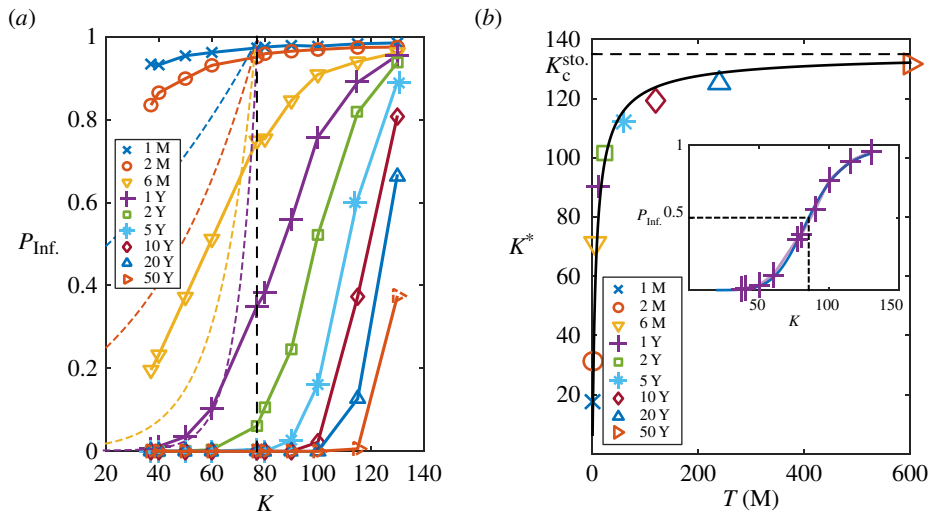


Figure 4. (a) Probability of sustained infections, P_{Inf} , as a function of K and different values of T (in the legend, M and Y stand for months and years, respectively). Solid lines with symbols represent stochastic simulations (based on 10^4 stochastic realizations while dashed lines show deterministic estimations. The dashed vertical black line indicates the population threshold (i.e. critical size of the zoonotic niche) to sustain the infection in the deterministic system: $K_c \approx 77$. (b) Estimation of the population threshold to sustain the infection in the stochastic system (see text), $K_c^{\text{sto}} = 135$. The inset shows an example of sigmoid fitting of P_{Inf} for the case $T = 1$ year, where $K^* \approx 90$ is determined by the value of K at which $P_{\text{Inf}} = 1/2$.

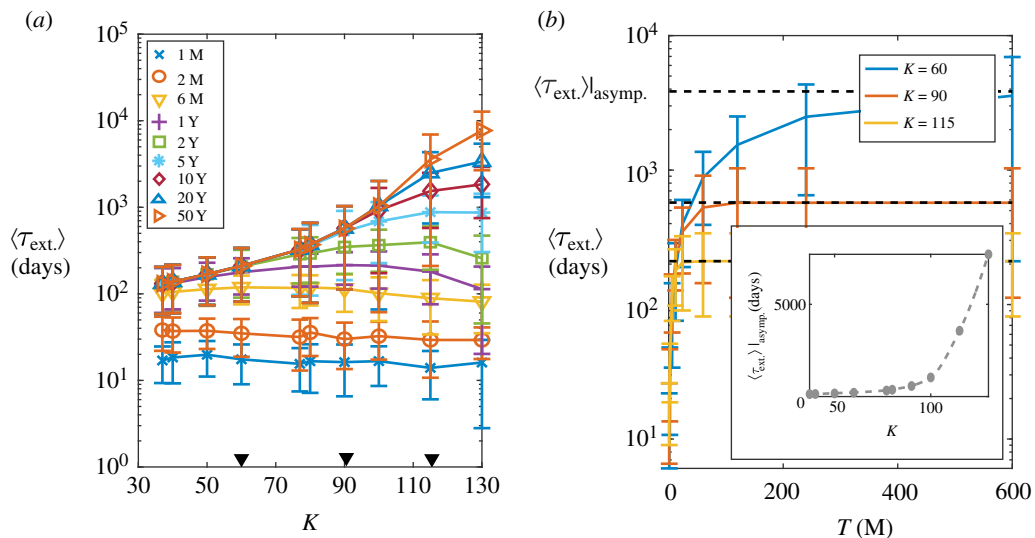


Figure 5. (a) Average time for infection extinction, $\langle \tau_{\text{ext}} \rangle$, as a function of the size of the zoonotic niche, K , for different observation window values, T (where M and Y stand for months and years respectively; 10^4 stochastic realizations). Error bars: standard deviation. (b) $\langle \tau_{\text{ext}} \rangle$ as a function of T for some values of K (represented by black solid triangles in panel (a)). Symbols denote stochastic simulations (error bars: standard deviation), solid lines indicate exponential fitting (see text) and dashed lines represent values of $\langle \tau_{\text{ext}} \rangle_{\text{asympt.}}$ as determined by the fitting. The inset shows $\langle \tau_{\text{ext}} \rangle_{\text{asympt.}}$ as a function of K .

for a given value of T , $\langle \tau_{\text{ext}} \rangle$ is calculated considering only trajectories where the number of infected bats reaches zero at times $t < T$. In this regard, we point out that if the parameters are such that the infection becomes extinct $\langle \tau_{\text{ext}} \rangle$ is independent of T . However, in practical terms, this requires to run long simulations to estimate $\langle \tau_{\text{ext}} \rangle$ accurately. To overcome this issue, we implement an approach that allows us to extrapolate the asymptotic behaviour from simulations by using different values of T . Our results indicate that if $K < K_c^{\text{sto}}$, there is a convergence as T increases (i.e. the asymptotic behaviour of $\langle \tau_{\text{ext}} \rangle$ is well defined). However, as K approaches K_c^{sto} , convergence requires larger values of T . This

observation, in agreement with figure 4b, suggests that infection is sustained if $K > K_c^{\text{sto}}$. These results are more clearly shown by plotting $\langle \tau_{\text{ext.}} \rangle$ as a function of T for different values of K , figure 5b. In that panel, to determine the asymptotic value of $\langle \tau_{\text{ext.}} \rangle$, we proceed as in the case of K^* (figure 4b) by fitting the curves to an exponential function (equation (3.5)),

$$\langle \tau_{\text{ext.}} \rangle(T) = \langle \tau_{\text{ext.}} \rangle|_{\text{asympt.}} \left(1 - e^{-\theta T^\lambda}\right)^\omega. \quad (3.5)$$

By using the asymptotic fitted values, as shown in the inset in figure 5b, we observed that $\langle \tau_{\text{ext.}} \rangle|_{\text{asympt.}}$ increases as K increases and shows a diverging behaviour as the size of the zoonotic niche approaches K_c^{sto} . These results highlight that depending on the colony size, i.e. on the value of K , an accurate estimation of the average time for an infection to die out requires very different observation times. Importantly, this fact should be considered in relation to the amount of time that the zoonotic niche can be considered isolated (§4).

Everything considered, from both the viewpoints of the probability of sustained infections and the extinction time, our results indicate that the fluctuations of the population act as a protective mechanism against infection such that the critical size of the zoonotic niche for which infection is sustained increases with respect to the expected deterministic behaviour.

4. Conclusions and discussion

The global challenge of epidemic disease transmission has sought the attention of decision-makers, as it poses significant threats to both human lives and economic stability. Researchers are driven by the urge to comprehend this issue and propose strategies to mitigate the spread of such diseases. The first generation of models emerged in the form of deterministic frameworks. However, these models exhibited certain limitations in their realism due to their disregard for the influence of intrinsic and extrinsic fluctuations. Recent advancements in the field have led to a more profound understanding of disease outbreaks by incorporating the intrinsic randomness inherent in such processes [47,48]. Herein, through rigorous theoretical analysis using an MHB approach, we present closed-form solutions that account for stochastic effects. Numerical results provide evidence that the outcomes derived from the theoretical framework align closely with the results obtained through numerical analysis. In particular, our results reveal that stochasticity leads to a non-null probability to eradicate infections, and as a result the size of the zoonotic niches that sustain infection is larger than the deterministic expectation. Moreover, our calculations also suggest that under some conditions, the fluctuations are able to sustain the infection for longer times than in a deterministic system. Nonetheless, in this case, a cautionary tale is needed when comparing the deterministic and stochastic models. By definition, the probability of infection in the deterministic model is either 1 or 0. To circumvent this issue and estimate a continuous value of the probability, i.e. $P_{\text{inf.}}^{\text{det.}} \in (0,1)$, we implemented a linear approach (appendix). Consequently, the comparison between the deterministic (linear) model and the stochastic (nonlinear) model must be carefully analysed.

It is also important to acknowledge the limitations of the MHB technique. The multivariate Gaussian approximation implemented through the MHB method explains very well the values of the two first moments obtained in the simulations but disregards correlations among variables beyond the second order. This introduces a level of approximation and is one of the reasons why there is some disparity between analytical and numerical results (another one being the numerical errors and the finite statistics of our simulations). Related to this, in our study, we have broken the moments' hierarchy at the third order, but we could have extended this approach to higher orders by implementing, for example, that the fourth-order cumulant is null. In that case, the equations to be solved increase from 9 to 19 and the analytical calculations become cumbersome; but we could expect even better results. However, given the excellent agreement, we have obtained in our study we have deemed this to be unnecessary. Altogether, while quantitative predictions from the theoretical analysis could be certainly improved, the stochastic effects that we report here will still be present and the MHB approach when assuming 'Gaussianity' capture them convincingly well.

Our modelling approach also includes some important assumptions. Namely, the independence of the carrying capacity from environmental factors and, consequently, the implicit assumption that those are constant, despite their time-dependent nature. Additionally, parameters such as the birth rate often exhibit temporal variations. These assumptions stress the importance of the analyses carried out as

a function of the observation window, T . This quantity informs about the amount of time that one could consider the zoonotic niche to be isolated (lack of influx/efflux of individuals) and subjected to invariable conditions. We have shown that, as expected, transient behaviours may differ significantly from the steady-state conditions (figures 4 and 5). The analysis of the infection probability in relation to the carrying capacity and the duration of the simulation time frame demonstrates a multi-faceted interplay between various variables that impact the dynamics of the disease. Particularly, in short-term simulations (e.g. one month), it appears that the solutions generated do not guarantee a steady-state outcome. As a result, the reliability of these short-term simulations tends to be comparatively lower. Nevertheless, these outcomes still hold important value, especially when devising strategies for short-term intervention mechanisms. Conversely, in longer time frames, in both deterministic and stochastic cases, the critical values dictating the onset of infection exhibit a significant discrepancy. This phenomenon diverges from the initial expectation that these critical values would gradually converge as the simulation duration extends. The observed divergence is not unique to this particular study but is a phenomenon that surfaces in the context of various diseases, including instances like Hantavirus modelling where stochasticity plays a role [17].

Our conclusions have direct implications. Vector control actions must continue not only when the critical values are attained but also for values nearby such values taking into account the potential presence of intrinsic fluctuations, especially when small and medium-sized niches are encountered. Standard procedures involve using $\mathcal{R}_0$, the basic reproduction number, to consider potential control and mitigation measures. That is, if the number is greater than 1, action needs to be taken to prevent the spread of the disease. This study, however, demonstrates that this approach alone might be insufficient. Stochastic analysis reveals the reverse of what the deterministic model, which corresponds to a single value of infection point, predicts. The theoretical framework shows that the potential variation due to bat dynamics might directly lead to a spark to start an epidemic, thus control measures must be integrated with stochasticity [49]. This study can help future Ebola epidemic preparedness and contingency plans, particularly in small and medium-sized zoonotic niches [50]. In particular, increased wildlife disease surveillance is required, and data from this surveillance must be incorporated into model development and analysis to hinder spillover [51,52].

Our framework has been applied in this study to address the EVD, but its applicability can extend to other zoonotic diseases, including avian influenza, Marburg and Zika viruses. While some parameters and the SIR model may vary, the theory still remains valid, as long as the infection dynamics are driven by the population density. Any necessary modifications can be readily implemented, enabling the calculation of the infected animal population across varying carrying capacities. Effects of intrinsic fluctuations are a function of different parameters leading to a spectrum of effects ranging from minimal to substantial. For example, for Marburg virus, if the model yields similar patterns as observed in the Ebola virus model presented in this manuscript, including zoonotic niches in the analysis can identify locations that may be more prone to disease spillover [53].

Altogether, herein we have shown important, non-trivial, stochastic effects that can hopefully help to understand the infection dynamics of zoonotic niches.

Ethics. This work did not require ethical approval from a human subject or animal welfare committee.

Data accessibility. The manuscript contains all the data, along with corresponding appendixes and electronic supplementary material. Access to the scripts/codes is available through the Figshare repository [54].

Declaration of AI use. We have not used AI-assisted technologies in creating this article.

Authors' contributions. S.M.: formal analysis, investigation, methodology, software, validation, visualization, writing—original draft, writing—review and editing; P.B.: funding acquisition, investigation, methodology, writing—original draft, writing—review and editing; J.B.: conceptualization, formal analysis, funding acquisition, investigation, methodology, software, supervision, visualization, writing—original draft, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We declare we have no competing interests.

Funding. P.B. and J.B. gratefully acknowledge support from Lehigh University through the 'Research Futures: Major Program Development' and the 'Research Futures: Special Seed Funding Opportunity' grants. P.B. and S.M. gratefully acknowledge support from NSF through award PIPP-2200066 'Dynamics of Pandemic Spread and Prevention in Indigenous Communities'. J.B. acknowledges support from grant PID2022-137436NB-I00 funded by MCIN/AEI/10.13039/501100011033 and by ERDF 'A way of making Europe' by the EU.

Acknowledgements. This work is part of the activities of the 'Center for Catastrophe Modeling' at Lehigh University and of the 'Catastrophe Modeling Coordination Network' (<https://www.catmodeling.org/>).

Appendix

Master equation: cumulants and moment hierarchy breaking approach

The following equation describes the master equation for entire bat population, $N = S + I + R$,

$$\frac{\partial P_N}{\partial t} = b(N-1)P_{N-1} + (N+1)\left(\mu + \frac{|b-\mu|}{K}N\right)P_{N+1} - \left(b + \mu + \frac{|b-\mu|}{K}(N-1)\right)NP_N.$$

The n -moment of the P_N distribution reads $\langle N^n \rangle = \sum_{N=1}^{\infty} N^n P_N$, and by manipulating the previous equation, we obtain the following dynamical equations for the first two moments:

$$\begin{aligned}\left\langle \frac{\partial N}{\partial t} \right\rangle &= \langle N \rangle (b - \mu) \left(1 + \frac{1}{K}\right) - \langle N^2 \rangle \left(\frac{b - \mu}{K}\right), \\ \left\langle \frac{\partial N^2}{\partial t} \right\rangle &= -2\langle N^3 \rangle \left(\frac{b - \mu}{K}\right) + \langle N^2 \rangle (b - \mu) \left(2 + \frac{3}{K}\right) + \langle N \rangle \left(b + \mu - \frac{b - \mu}{K}\right).\end{aligned}$$

By imposing the closure condition (see text) $\kappa_3(N) = \langle N^3 \rangle - 3\langle N^2 \rangle \langle N \rangle + 2\langle N \rangle^3 = 0$, we obtain the following stationary solutions for $\langle N \rangle$ and $\langle N^2 \rangle$:

$$\begin{aligned}\langle N \rangle &= \frac{1}{4} \left(3 + 3K + \sqrt{1 + K \left(2 - \frac{8b}{b - \mu} + K \right)} \right), \\ \langle N^2 \rangle &= \frac{1}{4} (K + 1) \cdot \left(3 + 3K + \sqrt{1 + K \left(2 - \frac{8b}{b - \mu} + K \right)} \right).\end{aligned}$$

As for the case where the whole population is split in terms of S , I and R , the master equation reads

$$\begin{aligned}\frac{\partial P_{(S,I,R)}}{\partial t} &= b(S + I + R - 1)P_{(S-1,I,R)} \\ &\quad + \left(\mu(S + 1) + \frac{|b - \mu|}{K}(S + I + R)(S + 1) \right) P_{(S+1,I,R)} \\ &\quad + \left(\mu(I + 1) + \frac{|b - \mu|}{K}(S + I + R)(I + 1) \right) P_{(S,I+1,R)} \\ &\quad + \left(\mu(R + 1) + \frac{|b - \mu|}{K}(S + I + R)(R + 1) \right) P_{(S,I,R+1)} \\ &\quad + \beta(S + 1)(I - 1)P_{(S+1,I-1,R)} + \gamma(I + 1)P_{(S,I+1,R-1)} \\ &\quad + \delta(R + 1)P_{(S-1,I,R+1)} \\ &\quad - b(S + I + R)P_{(S,I,R)} - \mu(S + I + R)P_{(S,I,R)} \\ &\quad - \frac{|b - \mu|}{K}(S + I + R)(S + I + R - 1)P_{(S,I,R)} \\ &\quad - \beta SIP_{(S,I,R)} - \gamma IP_{(S,I,R)} - \delta RP_{(S,I,R)}.\end{aligned}$$

Thus, the dynamical equations for the first two moments of the different species become:

$$\begin{aligned}\left\langle \frac{\partial S}{\partial t} \right\rangle &= b\langle S + I + R \rangle - \mu\langle S \rangle - \frac{|b - \mu|}{K}\langle S \cdot (S + I + R - 1) \rangle - \beta\langle S \cdot I \rangle + \delta\langle R \rangle \\ \left\langle \frac{\partial I}{\partial t} \right\rangle &= -\mu\langle I \rangle - \frac{|b - \mu|}{K}\langle I \cdot (S + I + R - 1) \rangle + \beta\langle S \cdot I \rangle - \gamma\langle I \rangle \\ \left\langle \frac{\partial R}{\partial t} \right\rangle &= -\mu\langle R \rangle - \frac{|b - \mu|}{K}\langle R \cdot (S + I + R - 1) \rangle + \gamma\langle I \rangle - \delta\langle R \rangle \\ \left\langle \frac{\partial S^2}{\partial t} \right\rangle &= b\langle S + I + R \cdot (2S + 1) \rangle - \mu\langle S \cdot (2S - 1) \rangle - \frac{|b - \mu|}{K}\langle S \cdot (S + I + R - 1) \cdot \\ &\quad \cdot (2S - 1) \rangle - \beta\langle S \cdot I \cdot (2S - 1) \rangle + \delta\langle R \cdot (2S + 1) \rangle \\ \left\langle \frac{\partial I^2}{\partial t} \right\rangle &= -\mu\langle I \cdot (2I - 1) \rangle - \frac{|b - \mu|}{K}\langle I \cdot (2I - 1) \cdot (S + I + R - 1) \rangle \\ &\quad + \beta\langle S \cdot I \cdot (2I + 1) \rangle - \gamma\langle I(2I - 1) \rangle \\ \left\langle \frac{\partial R^2}{\partial t} \right\rangle &= -\mu\langle R \cdot (2R - 1) \rangle - \frac{|b - \mu|}{K}\langle R \cdot (S + I + R - 1) \cdot (2R - 1) \rangle \\ &\quad + \gamma\langle I \cdot I \cdot (2R + 1) \rangle - \delta\langle R \cdot (2R - 1) \rangle\end{aligned}$$

and the cross terms of species S , I and R read:

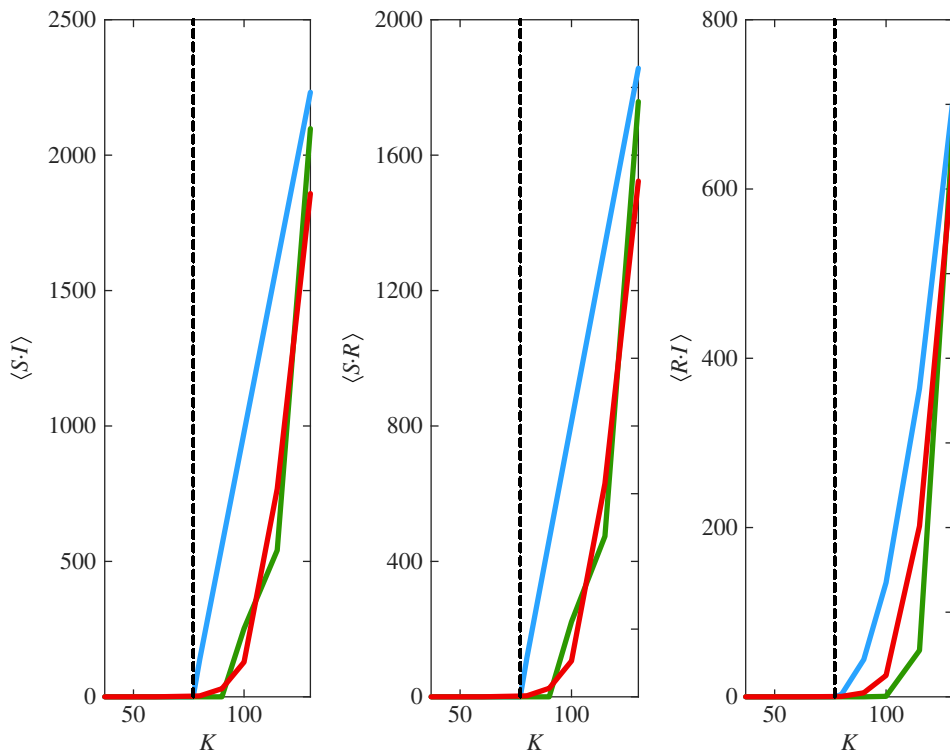


Figure 6. Correlations among species in the deterministic (blue line) and stochastic models (red line: numerical simulations; green line: MHB theory). The vertical dotted line indicates the transition points towards infection for the deterministic model ($K_c \approx 77$).

$$\begin{aligned}\left\langle \frac{\partial SI}{\partial t} \right\rangle &= b\langle I \cdot (S + I + R) \rangle - 2\mu\langle S \cdot I \rangle - 2\frac{|b - \mu|}{K}\langle S \cdot I \cdot (S + I + R - 1) \rangle \\ &\quad + \beta\langle S \cdot I \cdot (S - I - 1) \rangle + \delta\langle I \cdot R \rangle \\ \left\langle \frac{\partial SR}{\partial t} \right\rangle &= b\langle R \cdot (S + I + R) \rangle - 2\mu\langle S \cdot R \rangle - 2\frac{|b - \mu|}{K}\langle S \cdot R \cdot (S + I + R - 1) \rangle \\ &\quad + \beta\langle S \cdot I^2 \rangle + \gamma\langle S \cdot I \rangle + \delta\langle R \cdot (R - S - 1) \rangle \\ \left\langle \frac{\partial IR}{\partial t} \right\rangle &= -2\mu\langle I \cdot R \rangle - 2\frac{|b - \mu|}{K}\langle I \cdot R \cdot (S + I + R - 1) \rangle + \beta\langle S \cdot I \cdot R \rangle \\ &\quad + \gamma\langle I \cdot (I - R - 1) \rangle - \delta\langle I \cdot R \rangle\end{aligned}$$

On the other hand, the following equations describe the third-order cumulants of the multivariate probability functions $P_{(S, I, R)}$:

$$\begin{aligned}\kappa(S, I, R) &= \langle S \cdot I \cdot R \rangle - \langle S \cdot I \rangle \cdot \langle R \rangle - \langle S \cdot R \rangle \cdot \langle I \rangle - \langle I \cdot R \rangle \cdot \langle S \rangle + 2\langle S \rangle \cdot \langle I \rangle \cdot \langle R \rangle \\ \kappa(S, S, I) &= \langle S^2 \cdot I \rangle - \langle S^2 \rangle \cdot \langle I \rangle - 2\langle S \cdot I \rangle \cdot \langle S \rangle + 2\langle S \rangle^2 \cdot \langle I \rangle \\ \kappa(S, S, R) &= \langle S^2 \cdot R \rangle - \langle S^2 \rangle \cdot \langle R \rangle - 2\langle S \cdot R \rangle \cdot \langle S \rangle + 2\langle S \rangle^2 \cdot \langle R \rangle \\ \kappa(I, I, S) &= \langle I^2 \cdot S \rangle - \langle I^2 \rangle \cdot \langle S \rangle - 2\langle I \cdot S \rangle \cdot \langle I \rangle + 2\langle I \rangle^2 \cdot \langle S \rangle \\ \kappa(I, I, R) &= \langle I^2 \cdot R \rangle - \langle I^2 \rangle \cdot \langle R \rangle - 2\langle I \cdot R \rangle \cdot \langle I \rangle + 2\langle I \rangle^2 \cdot \langle R \rangle \\ \kappa(R, R, S) &= \langle R^2 \cdot S \rangle - \langle R^2 \rangle \cdot \langle S \rangle - 2\langle R \cdot S \rangle \cdot \langle R \rangle + 2\langle R \rangle^2 \cdot \langle S \rangle \\ \kappa(R, R, I) &= \langle R^2 \cdot I \rangle - \langle R^2 \rangle \cdot \langle I \rangle - 2\langle R \cdot I \rangle \cdot \langle R \rangle + 2\langle R \rangle^2 \cdot \langle I \rangle \\ \kappa(S, S, S) &= \langle S^3 \rangle - 3\langle S^2 \rangle \cdot \langle S \rangle + 2\langle S \rangle^3 \\ \kappa(I, I, I) &= \langle I^3 \rangle - 3\langle I^2 \rangle \cdot \langle I \rangle + 2\langle I \rangle^3 \\ \kappa(R, R, R) &= \langle R^3 \rangle - 3\langle R^2 \rangle \cdot \langle R \rangle + 2\langle R \rangle^3.\end{aligned}$$

The dynamical equations together with the closure relationships (null third-order cumulants) lead to a closed system of ordinary differential equations for nine variables ($\langle S \rangle$, $\langle I \rangle$, $\langle R \rangle$, $\langle S^2 \rangle$, $\langle I^2 \rangle$, $\langle R^2 \rangle$, $\langle S \cdot I \rangle$, $\langle S \cdot R \rangle$ and $\langle R \cdot I \rangle$).

Deterministic and stochastic correlations between species

The deterministic approach implies that correlations can be factorized, i.e. $\langle XY \rangle = \langle X \rangle \langle Y \rangle$. Thus, $\langle SI \rangle = \theta(K - K_c)K_c \frac{b + \delta}{b + \gamma + \delta}(K - K_c)$, $\langle SR \rangle = \theta(K - K_c)K_c \frac{\gamma}{b + \gamma + \delta}(K - K_c)$, and $\langle IR \rangle = \theta(K - K_c)K_c \frac{\gamma(b + \delta)}{b + \gamma + \delta}(K - K_c)^2$. Consequently, the transition towards infection is determined by non-null values of the correlations terms. Figure 6 shows the comparison between deterministic and stochastic results for $\langle SI \rangle$, $\langle SR \rangle$ and $\langle RI \rangle$. In agreement with the results shown in figure 3, the fluctuations acts as a barrier against infection with respect to the deterministic model. Figure 6 also highlights the agreement between stochastic simulations and the MHB analytical calculations.

Linear stability analysis of the deterministic model: characteristic decay time of infection

The steady-state solution that describes that infection-free system of equations (2.7)–(2.9) is $\mathbb{B} = (S = K, I = 0, R = 0)$. The Jacobian matrix that describes the behaviour of small perturbations around that steady-state reads (see [13,14] for details)

$$\mathbb{J}|_{(K, 0, 0)} = \begin{pmatrix} \mu - b & \mu - \beta K & \mu + \delta \\ 0 & -b - \gamma + \beta K & 0 \\ 0 & \gamma & -b - \delta \end{pmatrix}$$

that leads to the following eigenvalues: $\lambda_1 = \mu - b$, $\lambda_2 = -(b + \delta)$ and $\lambda_3 = -(b + \gamma) + \beta K = \beta(K - K_c)$. Thus, the stability of the infection-free system is determined by λ_3 : if $K < K_c$ then $\lambda_3 < 0$ ($\mathbb{N}$ is stable) and if $K > K_c$ then $\lambda_3 > 0$ ($\mathbb{N}$ is unstable). Thus, given an initial condition $I(t = 0) = I^0$, we expect that if $K < K_c$ the infected population will decay as $I(t) \simeq I^0 e^{-t \cdot \beta(K_c - K)}$ and, consequently, the probability of observing infected bats at time T can be estimated as $P_{\text{Inf.}} \simeq I(T)/I^0 = e^{-t\beta(K_c - K)}$. On the other hand, if $K > K_c$ the infection is sustained in the deterministic system, i.e. $P_{\text{Inf.}} = 1$. Altogether, the deterministic estimation of the probability of sustained infections reads

$$P_{\text{Inf.}} \simeq \theta(K_c - K) + \theta(K - K_c)e^{-T \cdot \beta(K_c - K)}.$$

References

1. Ram B, Thakur R. 2022 Epidemiology and economic burden of continuing challenge of infectious diseases in India: analysis of socio-demographic differentials. *Front. Public Health* **10**, 901276. (doi:10.3389/fpubh.2022.901276)
2. Shepard DS, Undurraga EA, Halasa YA, Stanaway JD. 2016 The global economic burden of dengue: a systematic analysis. *Lancet Infect. Dis.* **16**, 935–941. (doi:10.1016/S1473-3099(16)00146-8)
3. Dibble CJ, O'Dea EB, Park AW, Drake JM. 2016 Waiting time to infectious disease emergence. *J. R. Soc. Interface* **13**, 20160540. (doi:10.1098/rsif.2016.0540)
4. Centers for Disease Control and Prevention. *About Ebola virus disease*. See <https://www.cdc.gov/ebola/about/index.html> (accessed 25 October 2024).
5. Mursel S, Alter N, Slavitt L, Smith A, Bocchini P, Buceta J. 2022 Estimation of Ebola's spillover infection exposure in Sierra Leone based on sociodemographic and economic factors. *PLoS One* **17**, e0271886. (doi:10.1371/journal.pone.0271886)
6. Centers for Disease Control and Prevention. *History of Ebola virus disease outbreaks*. See <https://www.cdc.gov/ebola/outbreaks/index.html> (accessed 25 October 2024).
7. Leroy EM *et al.* 2004 Multiple Ebola virus transmission events and rapid decline of central African wildlife. *Science* **303**, 387–390. (doi:10.1126/science.1092528)
8. Anderson RM, May RM. 1992 *Infectious diseases of humans: dynamics and control*. Oxford, UK: Oxford University Press. (doi:10.1093/oso/9780198545996.001.0001)
9. Telford CT, Amman BR, Towner JS, Bowden S, Montgomery JM, Lessler J, Shoemaker T. 2023 A predictive model of Ebolavirus spillover incorporating change in forests and human populations across spatial and temporal scales. *medRxiv* 2023–08. (doi:10.1101/2023.08.29.23294795)
10. Nieddu GT, Forgoston E, Billings L. 2022 Characterizing outbreak vulnerability in a stochastic SIS model with an external disease reservoir. *J. R. Soc. Interface* **19**, 20220253. (doi:10.1098/rsif.2022.0253)
11. Leroy EM *et al.* 2005 Fruit bats as reservoirs of Ebola virus. *Nature* **438**, 575–576. (doi:10.1038/438575a)

12. Hayman DTS. 2015 Biannual birth pulses allow filoviruses to persist in bat populations. *Proc. R. Soc. B* **282**, 20142591. (doi:10.1098/rspb.2014.2591)
13. Buceta J, Johnson K. 2017 Modeling the Ebola zoonotic dynamics: interplay between enviroclimatic factors and bat ecology. *PLoS One* **12**, e0179559. (doi:10.1371/journal.pone.0179559)
14. Fiorillo G, Bocchini P, Buceta J. 2018 A predictive spatial distribution framework for filovirus-infected bats. *Sci. Rep.* **8**, 7970. (doi:10.1038/s41598-018-26074-4)
15. Fahr J, Abedi-Lartey M, Esch T, Machwitz M, Suu-Ire R, Wikelski M, Dechmann DKN. 2015 Pronounced seasonal changes in the movement ecology of a highly gregarious central-place forager, the African straw-coloured fruit bat (*Eidolon helvum*). *PLoS One* **10**, e0138985. (doi:10.1371/journal.pone.0138985)
16. Ortega J. 2016 *Sociality in bats*. Cham, Switzerland: Springer. (doi:10.1007/978-3-319-38953-0)
17. Escudero C, Buceta J, de la Rubia F, Lindenbergh K. 2004 Effects of internal fluctuations on the spreading of Hantavirus. *Phys. Rev. E* **70**, 461–470. (doi:10.1103/PhysRevE.70.061907)
18. Odatalla S. 2020 Stochastic modeling of zoonotic disease. Master's thesis, Montclair State University, Montclair, NJ. <https://digitalcommons.montclair.edu/etd/474>.
19. Case BKM, Young JG, Hébert-Dufresne L. 2023 Accurately summarizing an outbreak using epidemiological models takes time. *R. Soc. Open Sci.* **10**, 230634. (doi:10.1098/rsos.230634)
20. Nieddu G, Billings L, Forgoston E. 2014 Analysis and control of pre-extinction dynamics in stochastic populations. *Bull. Math. Biol.* **76**, 3122–3137. (doi:10.1007/s11538-014-0047-3)
21. Billings L, Forgoston E. 2018 Seasonal forcing in stochastic epidemiology models. *Ric. Mat.* **67**, 27–47. (doi:10.1007/s11587-017-0346-8)
22. Forgoston E, Moore RO. 2018 A primer on noise-induced transitions in applied dynamical systems. *SIAM Rev.* **60**, 969–1009. (doi:10.1137/17M1142028)
23. Langwig KE, Frick WF, Bried JT, Hicks AC, Kunz TH, Kilpatrick AM. 2012 Sociality, density-dependence and microclimates determine the persistence of populations suffering from a novel fungal disease, white-nose syndrome. *Ecol. Lett.* **15**, 1050–1057. (doi:10.1111/j.1461-0248.2012.01829.x)
24. Vonhof MJ, Russell AL, Miller-Butterworth CM. 2015 Range-wide genetic analysis of little brown bat (*Myotis lucifugus*) populations: estimating the risk of spread of white-nose syndrome. *PLoS One* **10**, e0128713. (doi:10.1371/journal.pone.0128713)
25. Martínková N, Baird SJE, Káňa V, Zima J. 2020 Bat population recoveries give insight into clustering strategies during hibernation. *Front. Zool.* **17**, 26. (doi:10.1186/s12983-020-00370-0)
26. Hethcote HW. 2000 The mathematics of infectious diseases. *SIAM Rev.* **42**, 599–653. (doi:10.1137/S0036144500371907)
27. Martinez K, Brown G, Pankavich S. 2022 Spatially-heterogeneous embedded stochastic SEIR models for the 2014–2016 Ebola outbreak in West Africa. *Spat. Spatiotemporal Epidemiol.* **41**, 100505. (doi:10.1016/j.sste.2022.100505)
28. Yang Q, Jiang D, Shi N, Ji C. 2012 The ergodicity and extinction of stochastically perturbed SIR and SEIR epidemic models with saturated incidence. *J. Math. Anal. Appl.* **388**, 248–271. (doi:10.1016/j.jmaa.2011.11.072)
29. Allen LJS. 2008 An introduction to stochastic epidemic models. *Lect. Notes Math.* **1945**, 81–130. (doi:10.1007/978-3-540-78911-6_3)
30. Näsell I. 2003 An extension of the moment closure method. *Theor. Popul. Biol.* **64**, 233–239. (doi:10.1016/s0040-5809(03)00074-1)
31. Singh A, Hespanha JP. 2011 Approximate moment dynamics for chemically reacting systems. *IEEE Trans. Automat. Contr.* **56**, 414–418. (doi:10.1109/TAC.2010.2088631)
32. Lee CH, Kim KH, Kim P. 2009 A moment closure method for stochastic reaction networks. *J. Chem. Phys.* **130**, 134107. (doi:10.1063/1.3103264)
33. Cornish EA, Fisher RA. 1938 Moments and cumulants in the specification of distributions. *Rev. Int. Stat. Inst.* **5**, 307–320. (doi:10.2307/1400905)
34. Rota GC, Shen J. 2000 On the combinatorics of cumulants. *J. Combin. Theory, Ser. A* **91**, 283–304. (doi:10.1006/jcta.1999.3017)
35. Davies P, Stuart A, Ord JK. 1988 Kendall's advanced theory of statistics. *J. R. Stat. Soc. Ser. A. Stat. Soc.* **151**, 240. (doi:10.2307/2982217)
36. Lloyd AL. 2004 Estimating variability in models for recurrent epidemics: assessing the use of moment closure techniques. *Theor. Popul. Biol.* **65**, 49–65. (doi:10.1016/j.tpb.2003.07.002)
37. Dormand JR, Prince PJ. 1980 A family of embedded Runge-Kutta formulae. *J. Comput. Appl. Math.* **6**, 19–26. (doi:10.1016/0771-050X(80)90013-3)
38. MATLAB. 2022 version 9.13.0 (R2022b). Natick, MA: The MathWorks Inc.
39. Abdennur N. 2022 Gillespie stochastic simulation algorithm. <https://github.com/nvictus/Gillespie>
40. Gillespie DT. 1977 Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* **81**, 2340–2361. (doi:10.1021/j100540a008)
41. Gillespie DT. 2000 The chemical Langevin equation. *J. Chem. Phys.* **113**, 297–306. (doi:10.1063/1.481811)
42. Kwiecinski GG, Griffiths TA. 1999 *Rousettus aegyptiacus*. *Mamm. Species* **611**, 1–9. (doi:10.2307/3504411)
43. Wilkinson GS, South JM. 2002 Life history, ecology and longevity in bats. *Aging Cell* **1**, 124–131. (doi:10.1046/j.1474-9728.2002.00020.x)
44. Amman BR *et al.* 2012 Seasonal pulses of marburg virus circulation in juvenile *Rousettus aegyptiacus* bats coincide with periods of increased risk of human infection. *PLoS Pathog.* **8**, e1002877. (doi:10.1371/journal.ppat.1002877)
45. Pourrut X, Souris M, Townner JS, Rollin PE, Nichol ST, Gonzalez JP, Leroy E. 2009 Large serological survey showing cocirculation of Ebola and Marburg viruses in Gabonese bat populations, and a high seroprevalence of both viruses in *Rousettus aegyptiacus*. *BMC Infect. Dis.* **9**, 159. (doi:10.1186/1471-2334-9-159)
46. Bradbury JW. 1979 Behavioural aspects of reproduction in Chiroptera. *Reproduction* **56**, 431–438. (doi:10.1530/jrf.0.0560431)

47. Pastor-Satorras R, Castellano C, Van Mieghem P, Vespignani A. 2015 Epidemic processes in complex networks. *Rev. Mod. Phys.* **87**, 925–979. (doi:10.1103/RevModPhys.87.925)
48. Kramer AM, Pulliam JT, Alexander LW, Park AW, Rohani P, Drake JM. 2016 Spatial spread of the West Africa Ebola epidemic. *R. Soc. Open Sci.* **3**, 160294. (doi:10.1098/rsos.160294)
49. McCormack RK. 2006 Multi-host multi-patch mathematical epidemic models for disease emergence with applications to hantavirus in wild rodents. PhD dissertation, Texas Tech University, Lubbock, TX. <http://hdl.handle.net/2346/17111>.
50. Pigott DM *et al.* 2014 Mapping the zoonotic niche of Ebola virus disease in Africa. *eLife* **3**, e04395. (doi:10.7554/eLife.04395)
51. Watsa M. 2020 Rigorous wildlife disease surveillance. *Science* **369**, 145–147. (doi:10.1126/science.abc0017)
52. Allen LJS, Brown VL, Jonsson CB, Klein SL, Lavery SM, Magwedere K, Owen JC, Van Den Driessche P. 2012 Mathematical modeling of viral zoonoses in wildlife. *Nat. Resour. Model.* **25**, 5–51. (doi:10.1111/j.1939-7445.2011.00104.x)
53. Pigott DM, Golding N, Mylne A, Huang Z, Weiss DJ, Brady OJ, Kraemer MUG, Hay SI. 2015 Mapping the zoonotic niche of Marburg virus disease in Africa. *Trans. R. Soc. Trop. Med. Hyg.* **109**, 366–378. (doi:10.1093/trstmh/trv024)
54. Figshare. 2024 *Stochastic Analysis of Ebola Infection in Small Zoonotic Niches*. See <https://figshare.com/s/20d9606cb253ef6fe7fb>.