Prediction of invasion from the early stage of an epidemic. Supplementary methods and results

Francisco J. Pérez-Reche[1,2], Franco M. Neri[3], Sergei N. Taraskin[4], and Christopher A. Gilligan[3]

[1]SIMBIOS Centre, University of Abertay Dundee, Dundee, UK

[2]Department of Chemistry, University of Cambridge, Cambridge, UK

[3]Department of Plant Sciences, University of Cambridge, Cambridge, UK

[4]St. Catharine's College and Department of Chemistry, University of Cambridge, Cambridge, UK

CONTENTS

I. CONTINUOUS-TIME EPIDEMIOLOGICAL MODEL FOR STEP (II)

In this section, we present the three-parameter continuous-time (CT) model used in methods E and F (Table 1, main text) to address step (ii) for prediction. In principle, this model is more realistic than the Reed-Frost (RF) model used in methods A-D and allows the existence of possible effects caused by the discrete-time character of the RF description to be explored. We use the fungal invasion experiment as a benchmark for the comparison between different models. In the CT, the spread (transmission) of the fungal colony between two neighbouring dots is treated as a time-inhomogeneous Poisson process [1]. The waiting time distribution $f(t)$ for each transmission event can be modelled by a Weibull distribution multiplied by the transmissibility T:

$$f(t) = T\frac{t^{k-1}}{\tau_0^k}\, e^{-\left(\frac{t}{\tau_0}\right)^k}, \tag{S.1}$$

where τ_0 is the characteristic time scale of the process, and k is a shape parameter of the distribution. Given the cumulative probability function $P(t) = \int_{u=0}^{t} f(u)\mathrm{d}u$, the survival function $S(t)$ (giving the probability that transmission did not occur by time t) obeys the following relation:

$$S(t) = 1 - P(t) = 1 - T\left(1 - e^{-\left(\frac{t}{\tau_0}\right)^k}\right), \tag{S.2}$$

The rate of the transmission process, $\phi(t)$, is a function of the time since colonisation of the donor dot and is given by the expression:

$$\phi(t) = \frac{f(t)}{S(t)} = -\frac{\mathrm{d}\ln\left(S(t)\right)}{\mathrm{d}\,t}\,.$$

In the limit $k \to \infty$, the CT model reduces to the RF model, with the same value of T and infectious period $\tau = \tau_0$. Indeed, for $k \to \infty$, the survival probability given by Eq. (S.2) becomes a step function,

$$S(t) = \begin{cases} 1, & t < \tau_0 \\ 1 - T, & t > \tau_0, \end{cases}$$

which corresponds to a RF model, in which infection can be transmitted from an infected host to a susceptible neighbour with probability T only once the infectious period τ_0 has passed.

II. METHODS FOR FITTING MODELS TO DATA IN STEP (III)

The aim of this section is to give a detailed description of the different methods used for fitting models $\mathcal{M}$ to data $\mathcal{D}$ described in step (iii) of the methods for prediction proposed in the main text. The ideas presented in the main text can be framed within a Bayesian approach which assumes that

the parameters $\boldsymbol{\theta}$ describing $\mathcal{M}$ are random variables. The aim of step (iii) is to evaluate $\pi(\boldsymbol{\theta}|\mathcal{D})$, the probability density that $\mathcal{M}$ with parameters $\boldsymbol{\theta}$ describes the data. According to Bayes' rule:

$$\pi(\boldsymbol{\theta}|\mathcal{D}) \propto \mathbb{P}(\mathcal{D}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}), \tag{S.3}$$

where $\pi(\boldsymbol{\theta})$ is the prior distribution of the parameters, reflecting our initial belief in their values, and $\mathbb{P}(\mathcal{D}|\boldsymbol{\theta})$ is the likelihood (the probability of the data given the parameters).

Several challenges arise when using this Bayesian approach in the analysis of epidemic spread. The first difficulty is associated with inherent limitations in the observations. A complete spatio-temporal data set would contain the precise time of colonisation t_j of each host j in the population, i.e. $\mathcal{D} = \{t_j\}$. Unfortunately, it is often the case that observations do not provide such detailed information. In the particular experimental data set considered in the main text, the status of each dot is only recorded at discrete (1-day) time intervals. Hence, the actual dataset is $\mathcal{D} = \{d_j\}$, where d_j is the day when dot j was first observed as colonised. We have explored two possible ways to deal with the lack of precise information. The option used in methods A-E (Table 1, main text) involves identifying descriptors for the evolution of the epidemic that are suitable for the prediction of the catastrophic event. The lower-dimensional descriptors used in this work are the cumulative incidence, $C(t)$, and the shell-evolution function, $F(l,t)$. Another option is to use data augmentation [2] that treats the unobserved colonisation times as parameters to be estimated. This is the procedure followed in method F.

Another source of difficulties is due to the fact that the analytical calculation of the posterior $\pi(\boldsymbol{\theta}|\mathcal{D})$ is, in general, impossible [3]. Therefore, it is common to resort to numerical methods to sample from $\pi(\boldsymbol{\theta}|\mathcal{D})$. In order to do this, we propose a new approximate method (denoted as MD), which is based on the calculation of the minimal distance between two datasets and does not require knowledge of the likelihood $\mathbb{P}(\mathcal{D}|\boldsymbol{\theta})$. In addition, we have used two known methods:

- a method belonging to the class of Approximate Bayesian Computation (ABC) [4], which calculates an *approximate* posterior, and which shares several basic features with the MD method.
- Markov-chain Monte Carlo (MCMC) with data augmentation [2, 5], which relies on the exact analytical form of $\mathbb{P}(\mathcal{D}|\boldsymbol{\theta})$ to sample from the exact posterior $\pi(\boldsymbol{\theta}|\mathcal{D})$.

The ABC method was used with multiple purposes:

(a) to test the new MD method (involving the *minimisation* of a given distance between observed and simulated data) against an already-known method (ABC, which involves a *cutoff* on the same distance).

(b) to test different choices in step (ii), by comparing estimations obtained using the RF model with estimations obtained by means of the CT model.

The MCMC method was employed to test the new MD method against a widely-used technique that uses both a different level of description (site-level vs. shell or MF level) *and* a different model $\mathcal{M}$ (CT dynamics instead of RF dynamics).

A. Minimum distance (MD) method

Let $\mathcal{D}_{\text{obs}}$ be the observed data, $\boldsymbol{\theta}$ a set of candidate parameters, and $\mathcal{D}_{\text{sim}}$ a simulated dataset generated using $\boldsymbol{\theta}$. Then, if $\mathcal{D}_{\text{obs}} = \mathcal{D}_{\text{sim}}$, the vector of parameters, $\boldsymbol{\theta}$, is drawn from $\pi(\boldsymbol{\theta}|\mathcal{D}_{\text{obs}})$. In practice, obtaining an exact match between observed and simulated dataset is often computationally unfeasible, and one has to resort to an approximate match. To this end, we define a metric $d^2(\mathcal{D}_1, \mathcal{D}_2)$ that measures the distance between two datasets $\mathcal{D}_1$ and $\mathcal{D}_2$. The aim of the MD method is to calculate the distribution of parameters $\boldsymbol{\theta}$ that minimise $d^2(\mathcal{D}_1, \mathcal{D}_2)$ and gives an approximate posterior. The choice of the metric $d^2\left(\mathcal{D}_{\text{sim}}, \mathcal{D}_{\text{obs}}\right)$ is problem-specific and, in general, not unique. We have tested two different metrics, corresponding to different descriptors of the data (cf. step (i) of the main text):

1. For the shell-based description, we used $d_f^2 = \sum_{l,\,t} \left(F_{\text{sim}}(l,\,t) - F_{\text{obs}}(l,\,t)\right)^2$, where l enumerates the shells, t is the discretised time (observation times in days), and F is the shell-evolution function.

2. Another option is to ignore any spatial information from the data (mean-field (MF) description) and consider only the total fraction of colonised sites at time t, $c(t) = C(t)/N$. In this case, we chose the distance function $d_c^2 = \sum_t \left(c_{\text{sim}}(t) - c_{\text{obs}}(t)\right)^2$, where t is again the discretised time.

The algorithm to implement the MD method considers two indexes, n, counting the parameter vectors resulting from the minimisation procedure, and r, counting the iterations. The values of n and r are in the range $n, r \in \mathbb{N}$ with maximum values $n_{\max}$ and R, respectively, and proceeds as follows:

MD.1 Set $n = 0$

MD.2 Set $r = 0$

MD.3 Chose a value $\boldsymbol{\theta}_n^{(r)}$ for the parameter vector sampled from the prior $\pi(\boldsymbol{\theta})$.

MD.4 Generate a data set $\mathcal{D}_{\text{sim}}^{(r)}$ from the model $\mathcal{M}$ with parameters $\boldsymbol{\theta}_n^{(r)}$.

MD.5 Calculate $d^2(\mathcal{D}_{\rm obs},\, \mathcal{D}^{(r)}_{\rm sim})$ and

- If $r < R$, set $r = r + 1$ and return to MD.3 or
- If $r = R$, go to MD.6

MD.6 Among all the parameters $\{\boldsymbol{\theta}_n^{(r)};\, r = 0, 1, \ldots, R\}$, chose the set of parameters $\boldsymbol{\theta}_n \in \{\boldsymbol{\theta}_n^{(r)}\}$ giving the closest simulated data, $\mathcal{D}_{\rm sim}$ to observations, i.e. $d^2(\mathcal{D}_{\rm obs},\, \mathcal{D}_{\rm sim}) = \min_{r=0,\ldots,R}\{d^2(\mathcal{D}_{\rm obs},\, \mathcal{D}^{(r)}_{\rm sim})\}$.

MD.7 Set $n = n + 1$ and return to MD.2 until $n = n_{\rm max}$.

The result of this algorithm is a set of parameter vectors $\{\boldsymbol{\theta}_n;\, n = 0, 1, \ldots, n_{\rm max}\}$ that give simulated data with minimum distance to observations. The normalised histogram for the obtained parameter vectors $\{\boldsymbol{\theta}_n\}$ defines a p.d.f. $\rho(\boldsymbol{\theta})$ that approximates the posterior $\pi(\boldsymbol{\theta}|\mathcal{D})$.

For all the results presented in the paper and obtained with the MD algorithm, we set $R = 5000$. For some fungal epidemics, we have checked that larger values of R do not lead to smaller values of $d^2(\mathcal{D}_{\rm obs},\, \mathcal{D}_{\rm sim})$ in step [MD.6].

B. Approximate Bayesian Computation (ABC)

In common with the MD method, the ABC approximate Bayesian method we have used also relies on the definition of a metric, d^2. If the distance between observed and simulated datasets is less than a given tolerance parameter, ϵ, i.e. $d^2(\mathcal{D}_{obs},\, \mathcal{D}_{sim}) \leq \epsilon$, then $\boldsymbol{\theta}$ is drawn from the *approximate* posterior $\pi\left(\boldsymbol{\theta}|d^2(\mathcal{D}_{obs},\, \mathcal{D}_{sim}) \leq \epsilon\right)$. The accuracy of the approximation increases as $\epsilon \rightarrow\ 0$.

For our estimations, we used a Markov-chain Monte Carlo algorithm that can be summarised as follows:

ABC.1 Set $n = 0$ and choose the initial value $\boldsymbol{\theta}_0$ of the parameter vector.

ABC.2 Generate a candidate vector, $\boldsymbol{\theta}'$, from a proposal distribution $q(\boldsymbol{\theta}'|\boldsymbol{\theta}_n)$.

ABC.3 Generate a data set $\mathcal{D}_{\rm sim}$ from the model $\mathcal{M}$ with parameters $\boldsymbol{\theta}'$.

ABC.4 Calculate $d^2(\mathcal{D}_{\rm obs},\, \mathcal{D}_{\rm sim})$ and

- if $d^2(\mathcal{D}_{\rm obs},\, \mathcal{D}_{\rm sim}) \leq \epsilon$, go to ABC.5;
- if $d^2(\mathcal{D}_{\rm obs},\, \mathcal{D}_{\rm sim}) > \epsilon$, set $\boldsymbol{\theta}_{n+1} = \boldsymbol{\theta}_n$ and go to ABC.7.

Rep.	D	B	E
1	0.03	1.5	0.5
2	0.03	1.5	0.7
3	0.03	1.5	1.5
4	0.03	1.5	0.5
5	0.03	1.5	1
6	0.03	1.5	1

TABLE I. Values of ϵ used for methods B, D, and E (Table 1, main text) using the ABC inference method in step (iii) for prediction.

ABC.5 Calculate the probability of acceptance:

$$p_{\rm acc} = \min\left(1, \frac{\pi(\boldsymbol{\theta}')q(\boldsymbol{\theta}_n|\boldsymbol{\theta}')}{\pi(\boldsymbol{\theta}_n)q(\boldsymbol{\theta}'|\boldsymbol{\theta}_n)}\right),$$

ABC.6 Set $\boldsymbol{\theta}_{n+1} = \boldsymbol{\theta}'$ with probability $p_{\rm acc}$, or $\boldsymbol{\theta}_{n+1} = \boldsymbol{\theta}_n$ with probability $1 - p_{\rm acc}$.

ABC.7 Set $n = n + 1$ and return to ABC.2 until the chain has converged and the required number of samples has been collected.

In our estimations, either a uniform distribution with the same support as the prior (see below) or normal distribution, $\mathcal{N}(0, \sigma^2)$ were used for the proposal distribution $q(.)$ (cf. steps ABC.2 and ABC.5). The criterion for the choice between these two distributions was to minimize the average number of simulations needed to generate a dataset with $d^2(\mathcal{D}_{\rm obs}, \mathcal{D}_{\rm sim}) \leq \epsilon$ in step [ABC.4]. The value of σ for the normal distribution was chosen according to the same criterion, and was typically a fraction between 0.05 and 0.1 of the support of the prior distribution.

Every chain was run for 5×10^4 steps, discarding an initial burn-in period of 5×10^3 steps. Since very small values of the tolerance ϵ imply very low acceptance rates, the final choice of ϵ was the result of a tradeoff between the accuracy of the approximation and CPU time available. The values of ϵ that were finally used in our analyses are shown in Table SI.

For both MD and ABC approaches, we assumed independent priors for all the parameters: $\pi(\boldsymbol{\theta}) = \pi(T)\pi(\tau_{\rm exp})$ (RF model) and $\pi(\boldsymbol{\theta}) = \pi(T)\pi(\tau_0)\pi(k)$ (CT model). For the RF model, all the priors were uniform: $\pi(T) = U(0, 1)$ and $\pi(\tau_{\rm exp}) = U(\tau_{\rm min}, \tau_{\rm max})$, where $\tau_{\rm min} = 1$d and $\tau_{\rm max}$ was changed between treatments (increasing with the lattice spacing, from $\tau_{\rm max} = 6$d for the lattice with $a = 8$mm spacing to $\tau_{\rm max} = 12$d for the lattice with $a = 16$mm spacing). For the CT model, we used two sets of priors: (i) noninformative uniform priors for all the parameters ($\pi(T) = U(0, 1)$, $\pi(\tau_0) = U(0, 20)$,

$\pi(k) = U(0,\, 20)$) and (ii) noninformative prior for the transmissibility ($\pi(T) = U(0,\, 1)$) and exponential priors for the other two parameters ($\pi(\tau_0) = \mathrm{Exp}(1)$, $\pi(k) = \mathrm{Exp}(1)$ as for the augmented-data MCMC estimations). Only results for noninformative priors from set (i) are presented below, and compared with results from the RF model. The different choice for the priors from set (ii) has an effect on the posterior distributions, but does not affect significantly the predicted incidence curves (see Section 3) and, for the sake of brevity, results from set (ii) are not presented.

The ABC and MD methods share several common features. Both of them rely on the simulation of epidemics using the current parameters, and on the calculation of a distance between simulated epidemics and the observed data. However, the ABC method considers any distance below the cutoff ϵ, while the MD method seeks to minimise the distance over a given number of simulations R. The iterative procedure used in MD method allows the use of an additional parameter analogous to ϵ in ABC method to be avoided. In both cases, the exact posterior is recovered in the proper limit ($\epsilon \to 0$ and $R \to \infty$, respectively).

C. Markov Chain Monte-Carlo (MCMC) method with data augmentation

In order to implement the data-augmented MCMC method, it is necessary to calculate the explicit form of the likelihood $\mathbb{P}(\mathcal{D}|\boldsymbol{\theta})$ (with $\boldsymbol{\theta} = (T,\, \tau_0,\, k)$). We sketch here the main steps of the calculation. Let $\mathcal{I}$ be the set of the dots that are colonised before the end of the experiment (at time $t_{\mathrm{end}} = 21$ days), and $\mathcal{U}$ be the set of those that are still uncolonised at $t = t_{\mathrm{end}}$. Assume first that we know the times of colonisation t_j of each dot $j \in \mathcal{I}$. The data then consist of the vector $\mathbf{t}$ of colonisation times, plus the set of uncolonised dots, i.e. $\mathcal{D} = (\mathbf{t},\, \mathcal{U})$. The nearest neighbours of dot j form the set $\mathcal{N}_j$, and the potential donors of j form the subset $\mathcal{S}_j \subseteq \mathcal{N}_j$. If $j \in \mathcal{U}$, then $\mathcal{S}_j$ contains the colonised neighbouring dots, i.e. $\mathcal{S}_j = \{i :\, i \in \mathcal{N}_j \cap \mathcal{I}\}$. If in contrast $j \in \mathcal{I}$, $\mathcal{S}_j$ contains the neighbouring dots that are colonised before j, i.e. $\mathcal{S}_j = \{i :\, i \in \mathcal{N}_j \cap \mathcal{I},\, t_i < t_j,\}$.

Given these definitions, the likelihood function can be written as the product of the contributions from individual dots j:

$$\mathbb{P}(\mathcal{D}|\boldsymbol{\theta}) = \prod_{j \in \mathcal{I}} f_j^{\mathcal{I}}(t_j) \prod_{j \in \mathcal{U}} \mathcal{P}_j^{\mathcal{U}}(t_{\mathrm{end}}), \tag{S.4}$$

where $f_j^{\mathcal{I}}(t_j)$, is the p.d.f. for the colonisation times t_j, and $\mathcal{P}_j^{\mathcal{U}}(t_{\mathrm{end}})$ is the probability for dot j to be uncolonised by the end of the experiment. These contributions can be calculated explicitly as follows. The probability that a dot $j \in \mathcal{I}$ has not been colonised by a given neighbour $i \in \mathcal{S}_j$ by time t_j is given by the survival function $S(t_j - t_i)$ (see Eq. (S.2)). Hence, the probability $\mathcal{P}_j^{\mathcal{I}}(t_j)$ that dot j is still

uncolonised at time t_j is given by the product over all $i \in \mathcal{S}_j$:

$$\mathcal{P}_j^{\mathcal{I}}(t_j) = \prod_{i \in\ S_j} S(t_j - t_i) = \prod_{i \in\ S_j} \exp\left(-\int_0^{t_j - t_i} \phi(t)\,\mathrm{d}t\right) = \exp\left(-\sum_{i \in\ S_j} \int_0^{t_j - t_i} \phi(t)\,\mathrm{d}t\right), \tag{S.5}$$

where we used the relation $S(t) = \exp(-\int_0^t \phi(u)\mathrm{d}u)$. The p.d.f. for t_j is then given by:

$$f_j^{\mathcal{I}}(t_j) = -\frac{d\mathcal{P}_j^{\mathcal{I}}(t_j)}{dt_j} = \left(\sum_{i \in\ S_j} \phi(t_j - t_i)\right) \exp\left(-\sum_{i \in\ S_j} \int_0^{t_j - t_i} \phi(t)\,\mathrm{d}t\right), \tag{S.6}$$

Likewise, a dot $j \in \mathcal{U}$ is still uncolonised by time t_{end} when transmission did not occur from any of its neighbours $i \in \mathcal{S}_j$, yielding the probability:

$$P_j^{\mathcal{U}}(t_{\mathrm{end}}) = \begin{cases} \prod\limits_{i \in\ \mathcal{S}_j} S(t_{\mathrm{end}} - t_i)\ , & \text{if} \quad \mathcal{S}_j \neq \emptyset\ , \\ 1\ , & \text{if} \quad \mathcal{S}_j = \emptyset\ . \end{cases} \tag{S.7}$$

In general, we are interested in obtaining the marginal distribution of a single parameter (in particular, T) from Eq. (S.3), and thus all other parameters entering the expression for $\mathbb{P}(\mathcal{D}|\boldsymbol{\theta})$ have to be integrated out. This is, in general, unfeasible analytically. In the fungal invasion experiment, the calculation is further complicated by *censoring*, i.e., by the fact that experimental observations are made at discrete (1-day) time intervals. As a consequence, if d_j is the day when dot j was first recorded as colonised, then the colonisation time t_j is constrained to lie in the interval $t_j \in (d_j - 1,\, d_j)$, but its exact value is unknown. The actual dataset is hence $\mathcal{D} = (\mathbf{d},\, \mathcal{U})$, and the likelihood has to be calculated from Eqs. (S.4-S.7) by integrating out the unobserved colonisation times,

$$\mathbb{P}(\mathbf{d},\, \mathcal{U}|\boldsymbol{\theta}) = \int_{\mathbf{T}(\mathbf{d})} \mathbb{P}(\mathbf{t},\, \mathcal{U}|\boldsymbol{\theta})\mathrm{d}\mathbf{t}\ ,$$

where the integral is performed over the (high-dimensional and, in general, very complex) space $\mathbf{T}(\mathbf{d})$ compatible with the observed data.

Since the high-dimensional integrals introduced above are analytically intractable, numerical techniques are commonly used to sample values from the posterior distribution $\pi(\boldsymbol{\theta}|\mathcal{D})$. The MCMC method consists in implementing a Markov chain for $\boldsymbol{\theta}$ that has $\pi(\boldsymbol{\theta}|\mathcal{D})$ as stationary distribution. A large literature exists on this subject (see e.g. Ref. [5]), to which the reader is referred for details. In our case, a Metropolis-Hastings algorithm has been used to build the Markov chain.

The algorithm can be summarised as follows:

MCMC.1 Set $n = 0$ and choose the initial value $\boldsymbol{\theta}_0$ of the parameter vector.

MCMC.2 Generate a candidate value of the vector $\boldsymbol{\theta}'$ from a proposal distribution $q(\boldsymbol{\theta}'|\boldsymbol{\theta}_n)$.

MCMC.3 Calculate the probability of acceptance:

$$p_{\rm acc} = \min\left(1, \frac{\pi(\boldsymbol{\theta}'|\mathcal{D})}{\pi(\boldsymbol{\theta}_n|\mathcal{D})}\right) = \min\left(1, \frac{\mathbb{P}(\mathcal{D}|\boldsymbol{\theta}')\pi(\boldsymbol{\theta}')q(\boldsymbol{\theta}_n|\boldsymbol{\theta}')}{\mathbb{P}(\mathcal{D}|\boldsymbol{\theta}_n)\pi(\boldsymbol{\theta}_n)q(\boldsymbol{\theta}'|\boldsymbol{\theta}_n)}\right),$$

MCMC.4 Set $\boldsymbol{\theta}_{n+1} = \boldsymbol{\theta}'$ with probability $p_{\rm acc}$, or $\boldsymbol{\theta}_{n+1} = \boldsymbol{\theta}_n$ with probability $1 - p_{\rm acc}$.

MCMC.5 Set $n = n + 1$ and return to MCMC.2 until the chain has converged and the required number of samples has been collected.

In order to deal with the unobserved colonisation times $\mathbf{t}$, we used data augmentation [2]. Their values were treated as parameters to estimate, i.e. the original parameter vector $\boldsymbol{\theta}$ was expanded (augmented) to $(\boldsymbol{\theta}, \mathbf{t})$.

New colonisation times were then proposed and accepted/rejected within the same Metropolis-Hastings algorithm. Additional care had to be taken in this case, since at each step a pathway of transmission must exist between the dot inoculated at time $t = 0$ and all the other colonised dots in the system (see the discussion in [2] and [6]).

Independent priors were used for all the parameters, so that $\pi(\boldsymbol{\theta}) = \pi(T)\pi(\tau_0)\pi(k)$. A noninformative uniform prior was used for the transmissibility, i.e. $\pi(T) = U(0, 1)$. For the other two parameters, exponentially distributed priors were used, i.e. $\pi(\tau_0) = \text{Exp}(1)$, $\pi(k) = \text{Exp}(1)$. Such choice was made in order to exploit our prior knowledge, i.e., respectively, that the typical time scale of the fungal spread in our system is of the order of days, and that the waiting-time distribution is not step-like (i.e., k is not too large; note that such assumption is opposite to that of the RF model). We checked that the posterior distribution was not too sensitive to changes up to a factor 2 in the parameters of the exponential priors. A uniform prior was used for each augmented colonisation time.

Every chain was run for 10^5 MCMC steps discarding an initial burn-in period of 10^3 steps. We checked that the final posterior distribution was robust with respect to the choice of the inital point $\boldsymbol{\theta}_0$.

D. Distribution for the estimated transmissibility, $\rho(\hat{T})$

This section complements the results presented in the main text (Figs. 2 and 3) for the p.d.f. $\rho(\hat{T})$ obtained with different methods for parameter estimation in the fungal invasion experiment. Within the Bayesian framework, the analogous of the distribution $\rho(\hat{T})$ introduced in the main text for the MD method is the marginal p.d.f of the posterior $\pi(\boldsymbol{\theta}|\mathcal{D})$ integrated over the variable $\tau_{\rm exp}$ for the RF model and over the variables τ_0 and k for the CT model.

Fig. S1 shows the comparison of the probability density functions $\rho(\hat{T})$ obtained for the agar dot experiment by all the methods summarised in Table 1 of the main text. The posteriors are shown for

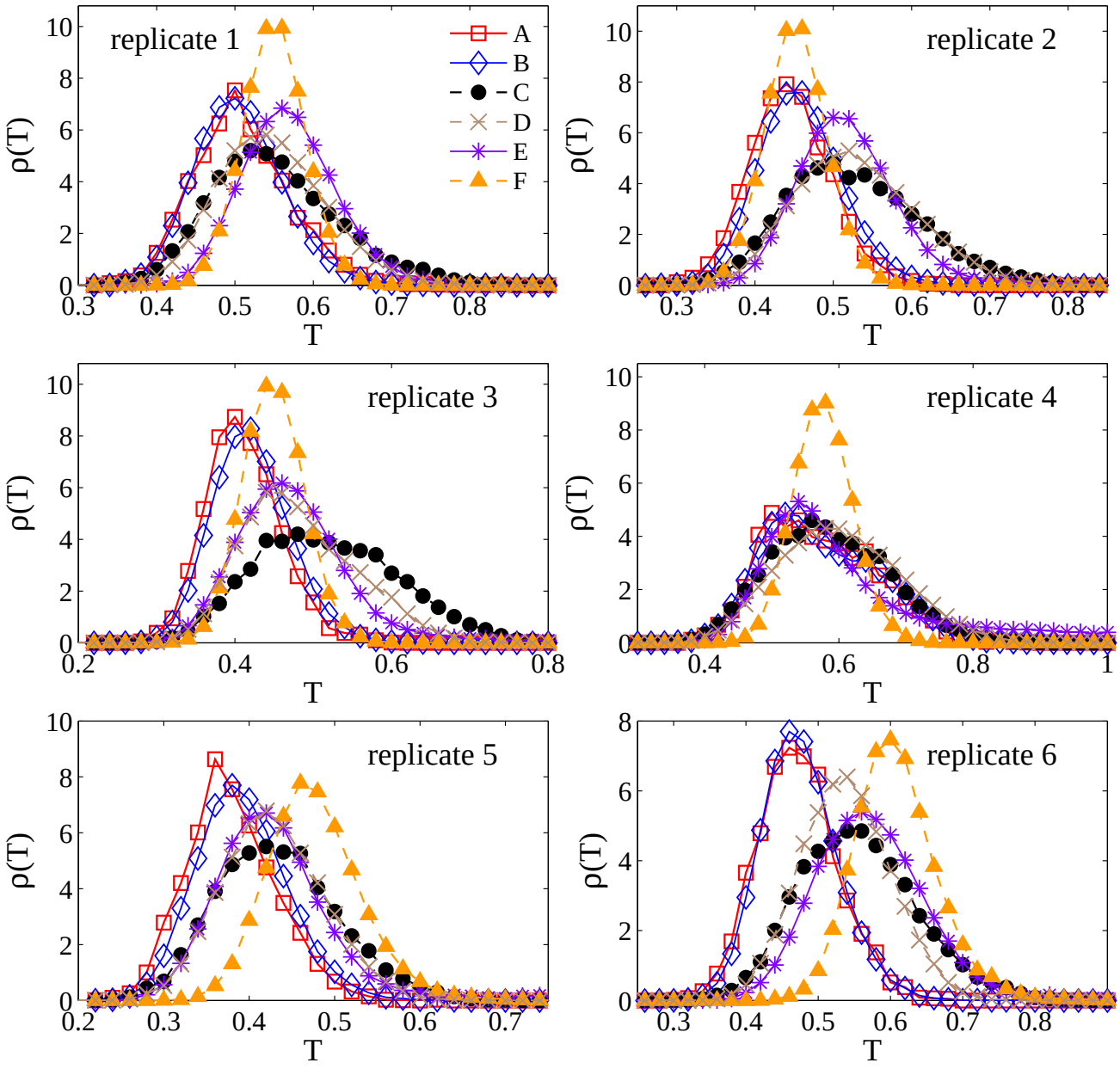

Fig. S1. Comparison of marginal posterior distributions $\rho(\hat{T})$ obtained with methods listed in Table 1 of the main text. Different symbols and lines correspond to different methods, as indicated by the legend. All the replicates correspond to invasion in populations with lattice spacing $a = 10$ mm.

experiments in populations with lattice spacing $a = 10$mm. This particular set of results was chosen because it summarised well all the main effects of the methodologies used. All the estimations correspond to observations over the complete duration of the experiment, i.e. $t \leq t_{\text{obs}} = 21$ days.

III. ADDITIONAL COMPARISONS OF FITTED MODELS WITH EXPERIMENTAL DATA

In the main text, we presented a test of the goodness of fit of models to data based on the mean squared distances d_c^2 and d_f^2. The purpose of this section is to give a more visual comparison between fitted models and observations based on the cumulative incidence.

For every model used, we obtain the statistics for the incidence corresponding to fitted models by sampling the parameters $\boldsymbol{\theta}$ from the joint (exact or approximate) posterior, $\pi(\boldsymbol{\theta}|\mathcal{D})$. This procedure gives a p.d.f. $\rho(C|t)$ for the incidence C at any given time t. The dispersion of $\rho(C|t)$ is associated both with the stochasticity in the simulated model for each value of the parameters, $\boldsymbol{\theta}$, and the dispersion for the values of these parameters given by $\pi(\boldsymbol{\theta}|\mathcal{D})$. Strictly speaking, the comparison of the experimental incidence with that obtained by methods A-D based on the RF model makes sense only when the stochastic nature of the process is taken into account. Indeed, the fine details of the two types of processes are different: while the experimental curve $C(t)$ corresponds to a discrete sampling of a *continuous-time process* (discrete set of observations), $\rho(C|t)$ gives the statistics of effective *discrete-time processes* with random values of transmissibility corresponding to $\boldsymbol{\theta}$ being drawn from $\pi(\boldsymbol{\theta}|\mathcal{D})$.

Figs. S2 and 3 show the comparison of the experimental incidence for systems with lattice spacing $a = 10$ mm with the estimations $\rho(C|t)$ obtained from observations during time $t \leq t_{\rm obs} = 21$ days. Comparisons are presented for methodologies B, C, E, and F. Results for methodology A (based on the MD method in step (iii) for prediction) are comparable to those obtained by methodology B (based on the ABC method) and have not been plotted in Figs. S2 and 3 for clarity. Similarly, the comparison for methodology D (based on the ABC inference method) is similar to that plotted for methodology C (based on the MD method) (not shown in Figs. S2 and 3). These results imply that, given a level of description of data (step (i)) and model in step (ii), results are quite independent of whether MD or ABC approaches are used to address step (iii).

As can be seen from the figures, the p.d.f $\rho(C|t)$ obtained by methods C, D, and E give a better description of the observed data for most of the replicates. On the other hand, the p.d.f. obtained by methodology F are often not able to capture the global trend of the incidence (see, e.g., replicate 3). This suggests that the focus of methodology F on finer details of the evolution may prevent the predictability of the invasive properties of the system. As discussed in the main text, this might be due to the negative interplay between the simplicity of the CT model and the individual-based description of the data-augmented MCMC method for inference.

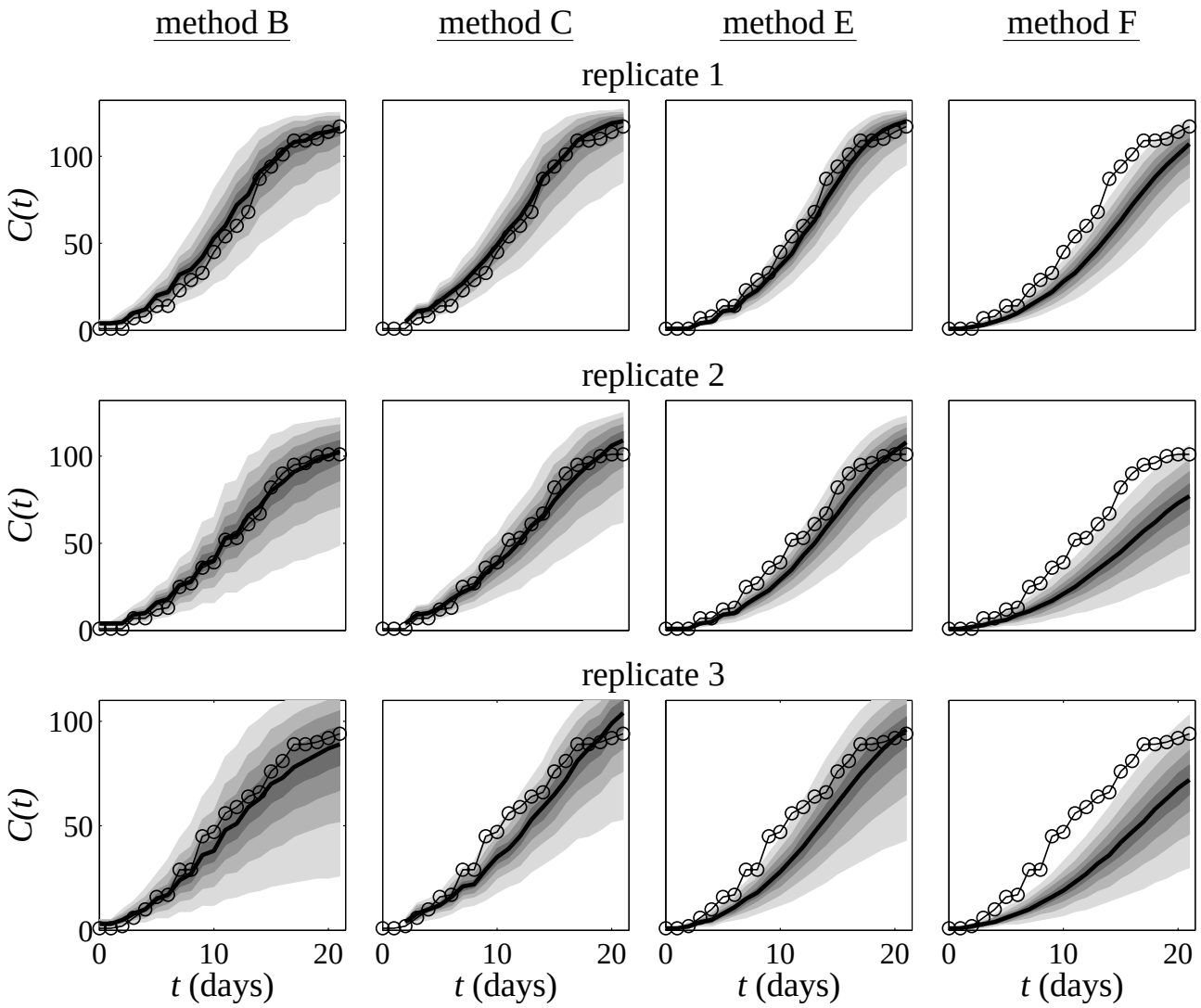

Fig. S2. Incidence (line with circles) for fungal invasion of a set of agar dots arranged on a triangular lattice with spacing $a = 10$ mm [7]. Replicates 1-3 are shown. For each replicate, four panels are shown, with the p.d.f. $\rho(C|t)$ for the incidence C at any time t obtained by means of methods B, C, E, and F (as described in Table 1 of the main text). In all the panels, the ridge (bold solid line) corresponds to the median of $\rho(C|t)$. The grey-scale shaded areas are the 20% (darker), 40%, 60%, and 80% (lighter) percentiles around the median.

IV. FORECAST OF THE INCIDENCE

In the main text, we quantified the differences between predictions and observations for all methods and fungal invasion experiments in terms of the quantities Δc and ΔF (cf. Fig. 6 of the main text). In this section, we give a more visual comparison between observations and predictions based on the temporal incidence, $C(t)$.

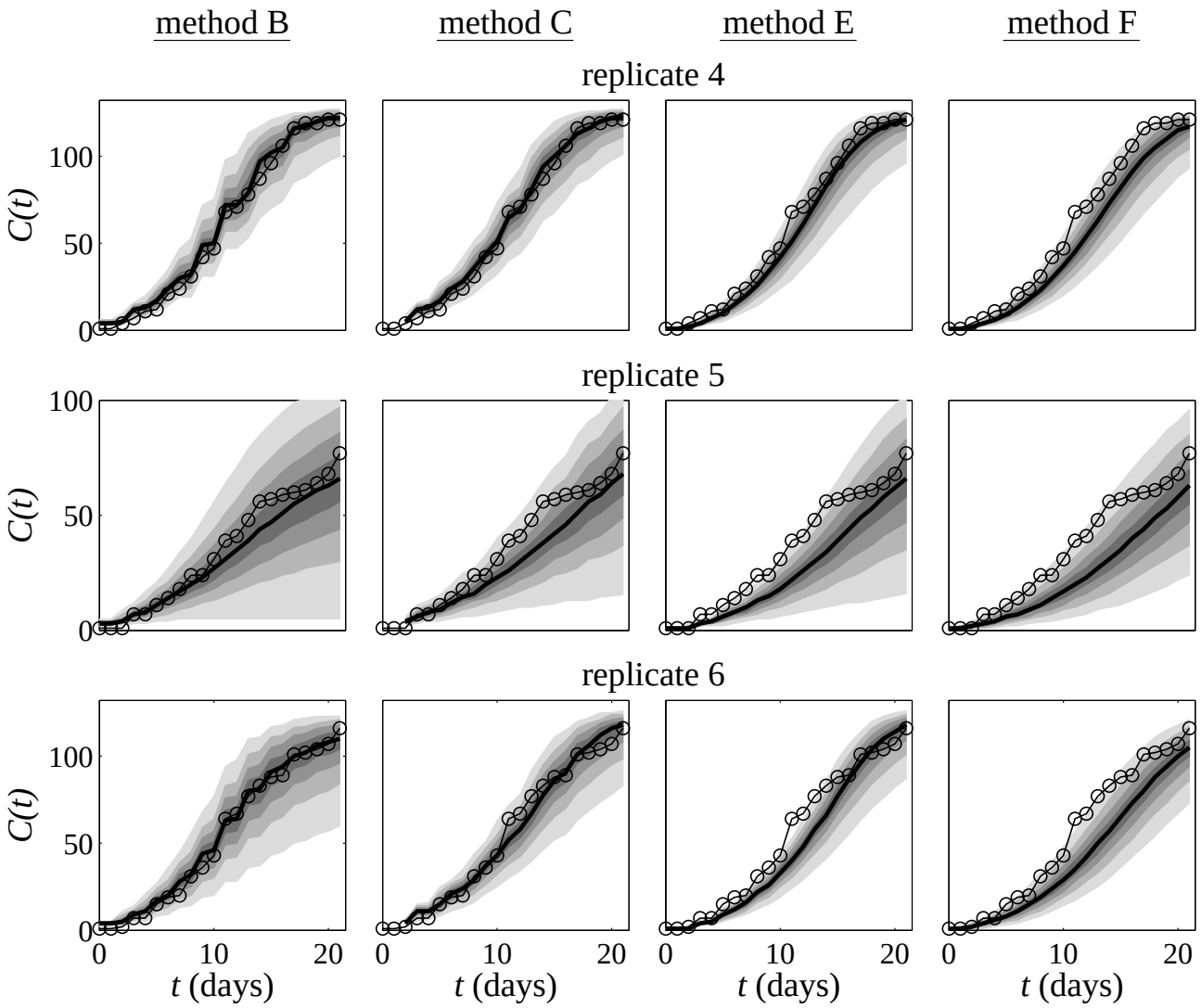

Fig. S3. Same as in Fig. S2, for replicates 4-6.

Figs. S4 and S5 show the comparison for the six replicates available in the experiments with $a =$ 10 mm and $a = 12$ mm, respectively. Predictions of the incidence between days 11 and 21 are based on estimates of the transmission parameters made during the first 10 days with methodology C. As can be seen, the estimated incidence provides a reasonable statistical description for the observed incidence for all the replicates. The quantity Δc quantifies the rms distance between the observed incidence and the predicted incidence obeying the p.d.f $\rho(C|t)$ (given by the grey-scale shaded area in Figs. S4 and S5).

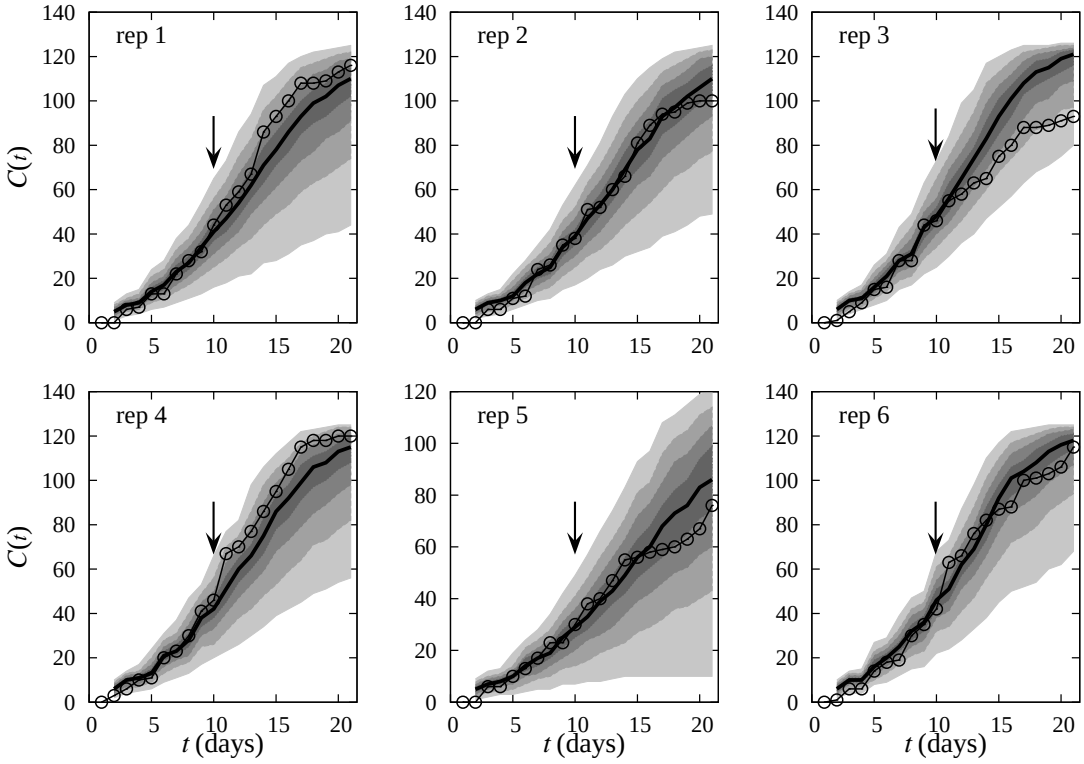

Fig. S4. Forecast of the incidence for fungal invasion in a lattice of agar dots with $a = 10$ mm. Each panel corresponds to a different experimental replicate of the epidemic. The grey-scaled shaded area shows the p.d.f. $\rho(C|t)$ for the numerically extrapolated incidence based on observations over time $t \leq t_{\rm obs} = 10$ days, as marked by arrows. The grey-scale shaded areas are the 20% (darker), 40%, 60%, and 80% (lighter) percentiles around the median (bold solid line).

V. ADDITIONAL RESULTS FOR THE NUMERICAL EXPERIMENTS

A. Homogeneous transmission of infection

Here, we give numerical support to the claim made in the main text that the most probable estimate for transmissibility, $\widehat{T}_\star$, corresponding to the maximum of the probability density function (p.d.f.) $\rho(\widehat{T})$, does not differ significantly from the actual transmissibility, T. This is shown in Fig. 6 where the estimate for the transmissibility, $\widehat{T}$, is plotted as a function of the transmissibility T. The estimates have been obtained for many different SIR numerical epidemics ($\sim 10^4$) with $T \in [0, 1]$. The analysis has been restricted to epidemics with final size $N_{\rm R}$ (i.e. the number of removed hosts) greater than a certain cut-off, N_0, in order to avoid estimates for small epidemics giving a poor estimator for transmissibility. Excluding small epidemics from the analysis also makes sense from a practical point of view because they are not a threat in terms of invasion. We have checked that the statistics for $\widehat{T}$ do not depend

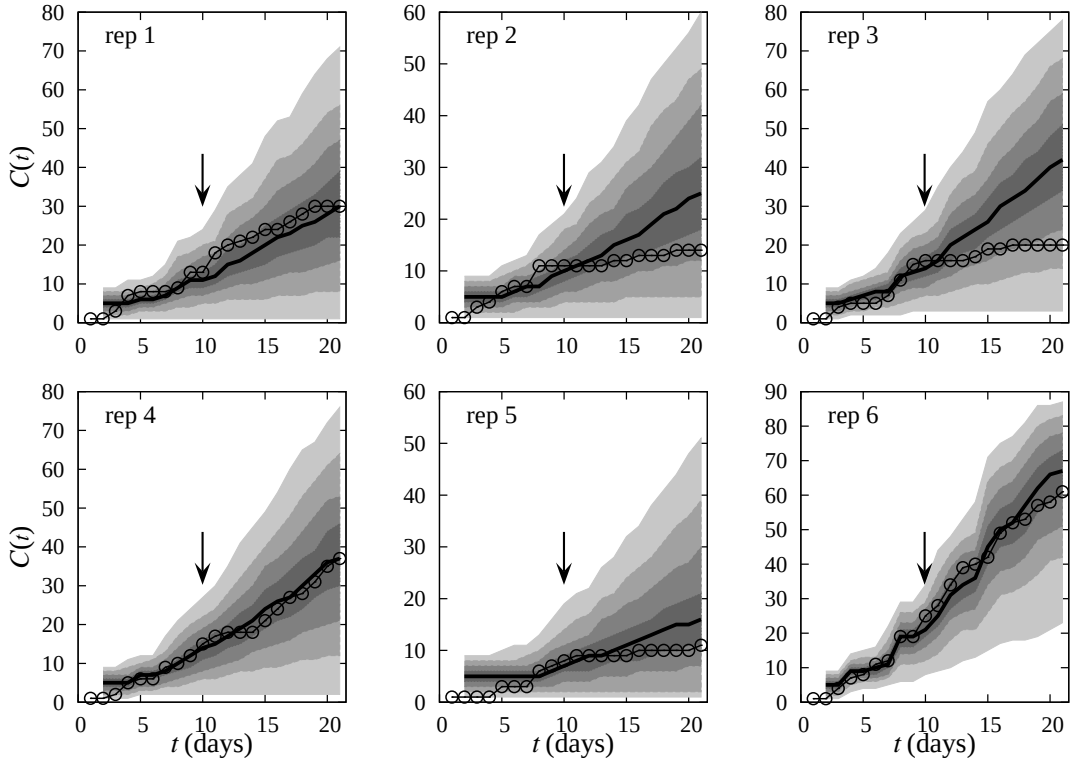

Fig. S5. Similar representation as in Fig. S4 for fungal epidemics in populations of agar dots with lattice spacing $a = 12$ mm.

significantly on N_0 for $N_0 \gtrsim 5$. For each epidemic i with given T, we obtain the p.d.f. $\rho_i(\widehat{T}|T)$ for the effective transmissibility, $\widehat{T}$, from observations of the initial stage (for $t \leq t_{\rm obs} = 7\tau$, where τ is the infectious period for infected hosts which is taken as the unit of time, $\tau = 1$). Then, the mean p.d.f., $\langle \rho(\widehat{T}|T) \rangle_{\rm e}$, is calculated by averaging $\rho_i(\widehat{T}|T)$ over $N_{\rm e}(T)$ different stochastic realisations of epidemics with given value of T:

$$\langle \rho(\widehat{T}|T) \rangle_{\rm e} = \frac{1}{N_{\rm e}(T)} \sum_{i=1}^{N_{\rm e}(T)} \rho_i(\widehat{T}|T) \ .$$

The first moment of $\langle \rho(\widehat{T}|T) \rangle_{\rm e}$ gives an estimate for the mean $\langle \widehat{T}_\star \rangle_{\rm e}$ averaged over stochastic realisations. The dependence of $\langle \widehat{T}_\star \rangle_{\rm e}$ on T is shown by the continuous line in Fig. S6. The dispersion of $\langle \rho(\widehat{T}|T) \rangle_{\rm e}$, shown by the shaded region in Fig. S6 as a function of T, contains contributions from both the width of each individual distribution, $\rho_i(\widehat{T}|T)$, and dispersion of the maxima for different replicates. In particular, the standard deviation, $\sigma(T)$, of $\langle \rho(\widehat{T}|T) \rangle_{\rm e}$ is given by

$$\sigma(T) = \left[\langle \sigma_i^2 \rangle_{\rm e} + \sigma_\star^2 \right]^{1/2} \ ,$$

where

$$\langle \sigma_i^2 \rangle_{\rm e} = \frac{1}{N_{\rm e}(T)} \sum_{i=1}^{N_{\rm e}(T)} \left[\int_0^1 \widehat{T}^2 \rho_i(\widehat{T}|T) {\rm d}\widehat{T} - \left(\int_0^1 \widehat{T} \rho_i(\widehat{T}|T) {\rm d}\widehat{T} \right)^2 \right]$$

is the average over stochastic realisations of the variance σ_i^2 of $\rho_i(\widehat{T}|T)$. The quantity $\sigma_\star^2$ is the variance of $\widehat{T}_{\star,i}$ over stochastic realisations, calculated as

$$\sigma_\star^2 = \frac{\sum_{i=1}^{N_\mathrm{e}(T)} \widehat{T}_{\star,i}^2}{N_\mathrm{e}(T)} - \left(\frac{\sum_{i=1}^{N_\mathrm{e}(T)} \widehat{T}_{\star,i}}{N_\mathrm{e}(T)} \right)^2 .$$

As can be seen from Fig. S6 the actual value for the transmissibility is statistically well described by the distribution $\langle \rho(\widehat{T}|T) \rangle_\mathrm{e}$. The mean for the most probable estimate for the transmissibility, $\langle \widehat{T}_\star \rangle_\mathrm{e}$, is in good agreement with the actual transmissibility. In particular, $\langle \widehat{T}_\star \rangle_\mathrm{e}$ provides an excellent estimate for T in the most interesting situations with $T \gtrsim 0.3$ where invasion is more likely. The deviations of $\langle \widehat{T}_\star \rangle_\mathrm{e}$ from T are larger for small values of T because epidemics are typically small and the deviations are large. However, it is important to note that $\langle \widehat{T}_\star \rangle_\mathrm{e}$ overestimates T in these situations and thus provides a safe bound for invasion.

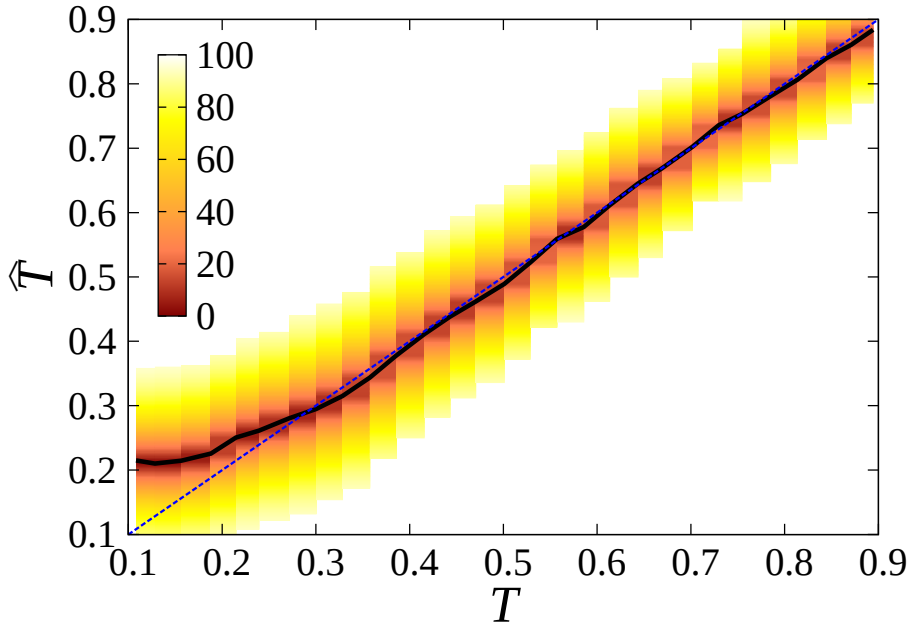

Fig. S6. Estimates of the transmissibility for numerical SIR epidemics with homogeneous transmissibility. The shaded yellow-brown region shows the levels of confidence as percentage of the p.d.f. $\langle \rho(\widehat{T}|T) \rangle_\mathrm{e}$ around the most probable mean transmissibility, $\langle \widehat{T}_\star \rangle_\mathrm{e}$ (continuous line), corresponding to each value of T. The dashed line representing the ideal situation (i.e. exact prediction) with $\hat{T} = T$ is given for comparison with the actual prediction shown by the continuous line.

B. Heterogeneous transmission of infection

In the main text, we have analysed the epidemics in model systems with homogeneous transmissibility for all pairs of connected hosts. Realistic populations of hosts exhibit inherent heterogeneity in transmissibility and it is crucial to understand its effect on the prediction method introduced in the main text.

In order to study the predictability of invasion for epidemics with heterogeneity in transmission of infection, we consider a simple but generic situation in which transmission is heterogeneous due to variability in the infectivity, $\mathcal{I}$, and susceptibility, $\mathcal{S}$, of hosts. As a first approximation, the rate of infection from an infected donor host with infectivity $\mathcal{I}_{\rm d}$ to a susceptible recipient host with susceptibility $\mathcal{S}_{\rm r}$ is defined as $\beta_{\text{d-r}} = \mathcal{I}_{\rm d}\mathcal{S}_{\rm r}$ [8]. We assume that $\mathcal{I}_{\rm d}$ and $\mathcal{S}_{\rm r}$ are independent random variables distributed according to truncated normal distributions, $\mathcal{N}(\bar{\mathcal{I}}, \sigma_I^2)$ and $\mathcal{N}(\bar{\mathcal{S}}, \sigma_S^2)$, respectively, which are the same for each host[9]. The mean values, $\bar{\mathcal{I}}$ and $\bar{\mathcal{S}}$, provide an effective measure of the mean strength of the transmissibility while the standard deviations, $\sigma_{\mathcal{I}}$ and $\sigma_{\mathcal{S}}$, characterize the degree of heterogeneity. The multiplicative form of the infection transmission rate $\beta_{\text{d-r}} = \mathcal{I}_{\rm d}\mathcal{S}_{\rm r}$ brings correlations in transmissibilities, $T_{\text{d-r}} = 1 - e^{\tau\beta_{\text{d-r}}}$ [10]. Indeed, all the transmissibilities from a donor are affected by the value of $\mathcal{I}_{\rm d}$ and thus they are not independent. Similarly, all the transmissibilities to a recipient are influenced by its susceptibility $\mathcal{S}_{\rm r}$ and thus also correlated. Such correlations make the invasion probability for heterogeneous system to be dependent on the whole set of transmissibilities $\{T_{\text{d-r}}\}$. In spite of that, the method based on a single effective transmissibility, T, and Eq. [1] in the main text is still applicable and useful. In realistic situations, the transmissibility is often assumed to be homogeneous because the precise degree of heterogeneity in transmission of infection is unknown and difficult to infer in detail.

In order to test our methodology in heterogeneous systems with different strengths of transmissibility, we perform numerical experiments for epidemics with $\sigma_{\mathcal{S}} = \sigma_{\mathcal{I}} = 0.2$, mean infectivity set to $\bar{\mathcal{I}} = 0.4$ and variable $\bar{\mathcal{S}}$. Again, observations are made over an initial interval of time $t \leq t_{\rm obs} = 7\tau$ for estimation of the effective transmissibility, $\widehat{T}$. Similarly to the analysis given in the previous section for epidemics with homogeneous transmission, Fig. S7 shows a comparison between the p.d.f. $\langle \rho(\widehat{T}|\langle T\rangle)\rangle_{\rm e}$ averaged over stochastic realisations of epidemics (shaded region) with the ideal situation giving exact prediction of the spatially averaged transmissibility used in the simulations, i.e. $\hat{T} = \langle T\rangle$ (dashed line). As can be seen, the estimates are statistically consistent with $\langle T\rangle$. In fact, the mean of the most probable transmissibility, $\langle \widehat{T}_\star\rangle_{\rm e}$, gives a good description for $\langle T\rangle$ (compare the continuous and dashed lines in Fig. S7). We proceed further as in the case with homogeneous transmission by calculating $\widehat{P}_{\rm inv}(L)$ for

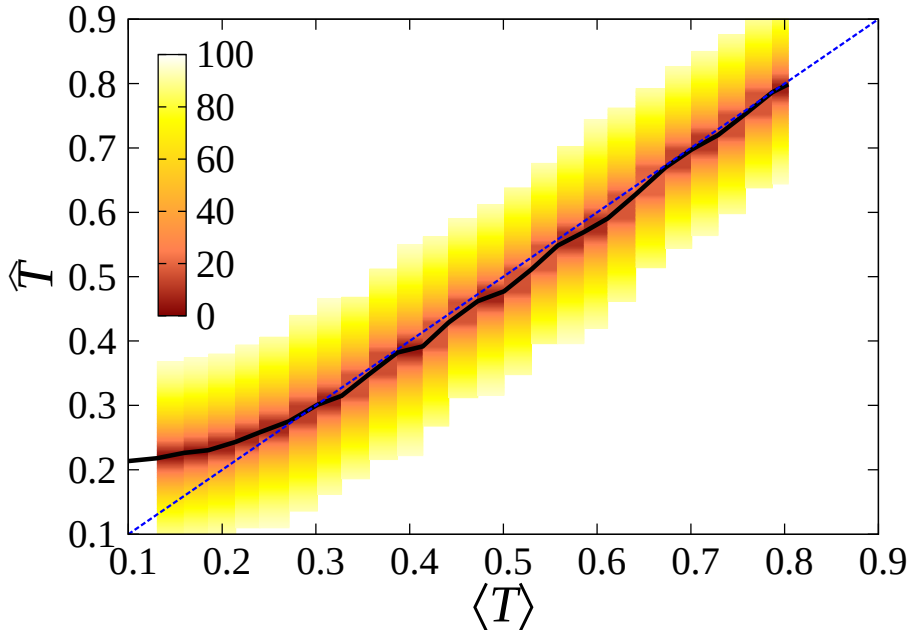

Fig. S7. Estimates for the effective transmissibility in numerical SIR epidemics with heterogeneous transmissibility introduced by Gaussian randomness in infectivity $\mathcal{I}$ and susceptibility $\mathcal{S}$ (with $\sigma_{\mathcal{I}} = \sigma_{\mathcal{S}} = 0.2$, mean infectivity set to $\bar{\mathcal{I}} = 0.4$, and variable mean susceptibility $0.1 \leq \bar{\mathcal{S}} \leq 3.5$). The shaded yellow-brown region shows the levels of confidence as percentage of the p.d.f. $\langle \rho(\widehat{T}|\langle T\rangle)\rangle_{\rm e}$ around the most probable mean transmissibility, $\langle \widehat{T}_\star \rangle_{\rm e}$ (continuous line). The dashed line representing the ideal situation (i.e. exact prediction) with $\widehat{T} = \langle T \rangle$ is given for comparison with actual prediction shown by the continuous line.

each of the numerical epidemics in a system of size $L = 51$ (see Fig. 2(a) in the main text) by using Eq. [1] in the main text, the estimated $\rho(\widehat{T})$, and the probability of invasion $P_{\rm hom}(\widehat{T}; L)$ for systems with homogeneous transmissibility equal to $\widehat{T}$. Note that here $P_{\rm hom}(\widehat{T}; L)$ corresponds to the function denoted as $P_{\rm inv}(\widehat{T}; L)$ in the main text. The notations have been changed in order to distinguish between the probability of invasion in homogeneous systems and the probability of invasion in the presence of heterogeneity, denoted as $P_{\rm het}(\widehat{T}; L)$. The results of our estimations are shown in Fig. S8. The relation between $\widehat{P}_{\rm inv}(L)$ and $P_{\rm hom}(\widehat{T}; L)$ is very similar to the relation reported in the main text for epidemics with homogeneous transmission. Indeed, for epidemics with low transmissibility, $\widehat{P}_{\rm inv}(L)$ typically overestimates $P_{\rm inv}$. In contrast, for more invasive epidemics, $\widehat{P}_{\rm inv}(L)$ underestimates $P_{\rm inv}$ for most of the possible effective transmissibilities. Although this comparison has some interest, in the current situation it makes more sense to compare the estimated probability of invasion with the actual probability of invasion in heterogeneous system, $P_{\rm het}(\widehat{T}; L)$, that can be calculated numerically and

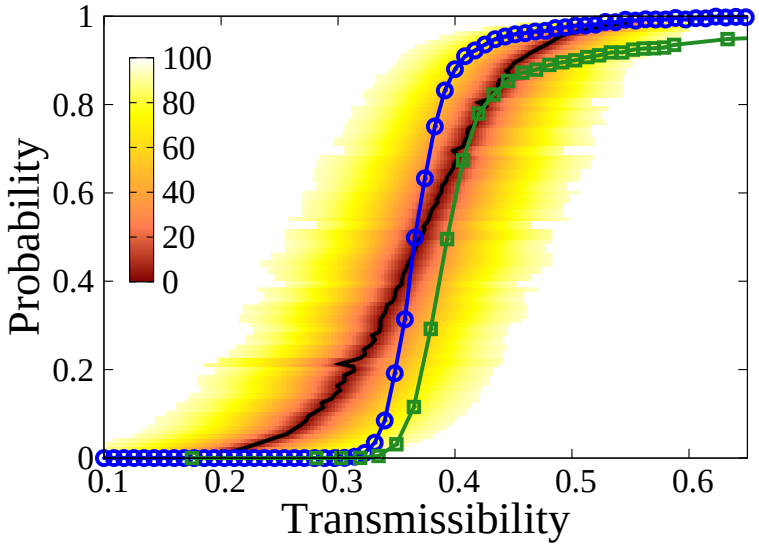

Fig. S 8. Numerical experiments of SIR epidemics with heterogeneous transmissibility induced by Gaussian randomness in infectivity $\mathcal{I}$ and susceptibility $\mathcal{S}$ (with $\sigma_{\mathcal{I}} = \sigma_{\mathcal{S}} = 0.2$, mean infectivity set to $\bar{\mathcal{I}} = 0.4$, and variable mean susceptibility $0.1 \leq \bar{\mathcal{S}} \leq 3.5$). The lines marked by circles and squares correspond to the probabilities of invasion $P_{\rm hom}$ and $P_{\rm het}$ plotted *vs* $\langle T \rangle$ for homogeneous and heterogeneous systems of size $L = 51$, respectively,. Estimates for the probability of invasion, $\widehat{P}_{\rm inv}(L)$, have been evaluated for each epidemic (out of $\sim 10^4$) using the p.d.f. $\rho(\widehat{T})$ obtained by observing the initial evolution during time $t \leq t_{\rm obs} = 7\tau$ and then fitting the observed spatio-temporal map by the shell-evolution function. Horizontal slices of the yellow-brown shaded area corresponding to a fixed value of $\widehat{P}_{\rm inv}(L)$ represent the distribution $\rho(\widehat{T})$ averaged over realisations of epidemics with the same value of $\widehat{P}_{\rm inv}(L)$. The points on the ridge (black solid curve) of the shaded area correspond to the most probable transmissibility, $\widehat{T}_*$, averaged over epidemics with a certain value of $\widehat{P}_{\rm inv}(L)$.

shown by the line with squares in Fig. S8. The comparison of this line with the shaded region reveals that the estimated $\widehat{P}_{\rm inv}(L)$ overestimates the actual probability of invasion in most of the situations both for cases with low and high transmissibility. Therefore, the estimates of $\widehat{P}_{\rm inv}$ typically provide safe bounds for the probability of invasion. In fact, the larger the heterogeneity in susceptibility and/or infectivity, the safer the bound is. This is a consequence of the inequality $P_{\rm inv}(\langle T \rangle) \geq P_{\rm het}(\langle T \rangle)$ that holds for any given value of $\langle T \rangle$ under quite general conditions due to the existence of correlations in transmission induced by heterogeneity in the transmission rates [8, 11]. Indeed, Fig. S8 shows that the inequality holds for the numerical experiments considered here with $\langle T \rangle = \widehat{T}$ (i.e. the line corresponding to the heterogeneous case marked by the squares is below the line marked by the circles for homogeneous system). These results are particularly encouraging for analysis of realistic epidemics in which a certain

degree of heterogeneity in susceptibility and infectivity of hosts is expected to be ubiquitous.

VI. DIFFERENCES BETWEEN $P_{\rm inv}$ AND $\hat{P}_{\rm inv}$

In this section, we discuss the origin of the difference between the estimated probability of invasion, $\hat{P}_{\rm inv}(L)$, evaluated at the most probable transmissibility, $\widehat{T}_\star$, and the actual probability of invasion $P_{\rm inv}(\widehat{T};L)$ one would obtain if the transmissibility of the epidemic was known exactly.

To start with, recall that the relation between $\hat{P}_{\rm inv}(L)$ and $P_{\rm inv}(\widehat{T};L)$ is given by Eq. [1] in the main text. As we have seen, $\rho(\widehat{T})$ is a peak-shaped function approximately symmetric about the peak position at $\widehat{T}_\star$. Therefore, the peak region will mainly contribute to the integral in Eq. [1] of the main text. If $\widehat{T}_\star$ is not too close to the inflection point of $P_{\rm inv}(\widehat{T};L)$, the Taylor series expansion of $P_{\rm inv}(\widehat{T};L)$ in $(\widehat{T}-\widehat{T}_\star)$ to second order, i.e.,

$$P_{\rm inv}(\widehat{T};L) = P_{\rm inv}(\widehat{T}_\star;L) + P'_{\rm inv}(\widehat{T}_\star;L)(\widehat{T}-\widehat{T}_\star) + \frac{1}{2}P''_{\rm inv}(\widehat{T}_\star;L)(\widehat{T}-\widehat{T}_\star)^2, \tag{S.10}$$

is sufficient to estimate the deviation of $\widehat{P}_{\rm inv}$ from $P_{\rm inv}$. Indeed, substitution of Eq. (S.10) into Eq. [1] of the main text gives:

$$\widehat{P}_{\rm inv}(L) = P_{\rm inv}(\widehat{T}_\star;L) + P'_{\rm inv}(\widehat{T}_\star;L)\int_0^1 \rho(\widehat{T})(\widehat{T}-\widehat{T}_\star){\rm d}\widehat{T} + \frac{1}{2}P''_{\rm inv}(\widehat{T}_\star;L)\int_0^1 \rho(\widehat{T})(\widehat{T}-\widehat{T}_\star)^2{\rm d}\widehat{T} \tag{S.11}$$

The distribution of $\widehat{T}$ is approximately symmetric around the maximum, i.e. $\rho(\widehat{T}-\widehat{T}_\star) \simeq \rho(\widehat{T}_\star-\widehat{T})$ (see Fig. 2(b) in the main text), and thus the term containing $P'_{\rm inv}$ in Eq. (S.11) is negligible in comparison with other terms in the sum. This means that

$$\begin{aligned} \widehat{P}_{\rm inv}(L) &> P_{\rm inv}(\widehat{T}_\star;L) \text{ if } P''_{\rm inv}(\widehat{T}_\star;L) > 0 \\ \widehat{P}_{\rm inv}(L) &\leq P_{\rm inv}(\widehat{T}_\star;L) \text{ if } P''_{\rm inv}(\widehat{T}_\star;L) \leq 0 \ , \end{aligned} \tag{S.12}$$

where we have taken into account that the last integral in Eq. (S.11) is positive. The above inequalities demonstrate that the value and sign of $\widehat{P}_{\rm inv}(L) - P_{\rm inv}(\widehat{T}_\star;L)$ depends on the curvature of $P_{\rm inv}(\widehat{T};L)$ around $\widehat{T}=\widehat{T}_\star$ which is given by $P''_{\rm inv}$.

---

[1] D. R. Cox and V. Isham, *Point Processes*, Monographs on Applied Probability and Statistics 12 (Chapman & Hall, London, 1980)

[2] G. J. Gibson and E. Renshaw, IMA Journal of Mathematics Applied in Medicine and Biology **15**, 19 (1998)

[3] P. Marjoram, J. Molitor, V. Plagnol, and S. Tavaré, Proc. Natl. Acad. Sci. USA **100**, 15324 (2003)

[4] P. Marjoram, J. Molitor, V. Plagnol, and S. Tavaré, Proc Natl Acad Sci USA **100**, 15324 (2003)

[5] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis* (Chapman and Hall/CRC, 2004)

[6] G. J. Gibson, W. Otten, J. A. N. Filipe, A. Cook, G. Marion, and C. A. Gilligan, Stat Comput **16**, 391 (2006)

[7] D. J. Bailey, W. Otten, and C. A. Gilligan, New Phytol. **146**, 535 (2000)

[8] J. Miller, J. Appl. Probab. **45**, 498 (2008)

[9] The support of the normal distributions has been restricted to $[0, \infty)$ to ensure that both $\mathcal{I}_i$ and $\mathcal{S}_i$ are positive.

[10] P. Grassberger, Math. Biosc. **63**, 157 (1983)

[11] J. T. Cox and R. Durrett, Stoch. Proc. Appl. **30**, 171 (1988)