Estimation of multiple transmission rates for epidemics in heterogeneous populations

Alex R. Cook, Wilfred Otten, Glenn Marion, Gavin J. Gibson, Christopher A. Gilligan

Abstract

In this paper we consider spatially heterogeneous epidemic systems, in which pathogen spread occurs through a landscape comprising favorable (e.g. susceptible or untreated) and less favorable (e.g. resistant or treated) sites. An important example is the deployment of resistant crop varieties in mixed species populations. It is well recognized that heterogeneity arising from the presence of multiple species or spatial variation in the interactions between individuals can endow population processes with a far more complex range of dynamics than would be exhibited in homogeneous settings. However, even though stochastic models for heterogeneous systems can be readily formulated, these models can only inform our understanding of any particular system if they can be parameterized for that system. We focus on processes and models with short-range interactions using data describing the spatio-temporal spread of Rhizoctonia solani, a soil-borne fungal plant pathogen, in mixed species populations as a convenient experimental system in which heterogeneity can be controlled. Such epidemics are frequently driven by an external source of infection (primary infection, the initiator of an epidemic and often associated with inoculum present in the soil), and secondary spread between infected and neighboring susceptibles in a population (secondary infection). We develop Bayesian methods to fit spatio-temporal percolation-based models, incorporating such features, to biological processes that evolve in homogeneous populations.

1 Introduction

There is a large body of literature dealing with mathematical spatio-temporal models for biological or physical processes that evolve through spatially structured populations. Examples include forest fires (He and Mladenoff, 1999; Hargrove et al., 2000), soil-borne epidemics (Gilligan, 1987; Gilligan, 2002), social networks (Newman, 2003), and computer viruses (Newman et al., 2002). Many of these processes occur in heterogeneous environments where the heterogeneity arises from the presence of multiple species or spatial variation in the interactions between individuals. It is well recognized that such heterogeneity can endow systems with a far more complex range of dynamics, relating to species and population persistence (Levin, 2000; Ettema et al., 2002; Gilligan, 2002; Park et al., 2002) or evolutionary processes (Shea et al., 2000), than would be exhibited in homogeneous settings. Therefore, even though stochastic models for heterogeneous systems can be readily formulated, these models can only inform our understanding of any particular system if they can be parameterized for that system. Sound parameterization of models is therefore essential if they are to be used, for example, in the development of control strategies such as the deployment of resistant varieties that minimize pathogen invasion with low risk of evolution of new virulent pathotypes.

In this paper we consider spatially heterogeneous epidemic systems, in which pathogen spread occurs through a landscape comprising favorable (e.g. susceptible or untreated) and less favorable (e.g. resistant or treated) sites. An important example is the deployment of resistant crop varieties in mixed species populations which has been studied extensively both mathematically (Brachet et al., 1999; Garrett et al., 1999; Finckh et al., 2000; Jeger, 2000) and empirically (Burdon et al., 1977; Garrett et al., 2000; Zhu et al., 2000; Mundt, 2002). In such populations, the spatial distribution of sites of each type is important in determining the likelihood of invasion and persistence of disease (Gilligan, 2002). Invasion of disease in such a landscape is affected by the dynamics of disease within each sub-population and the connectivity between favorable and unfavorable sites, which in turn is determined by the areas covered, the clustering of sites and the scale of dispersal (Perry, 2002). As connected pathways between favorable sites are explored preferentially by an invading epidemic, the contacts between favorable and unfavorable may become progressively more important as the epidemic progresses.

We focus on processes and models with short-range interactions. Soil-borne epidemics are exemplars of such processes. They are important determinants in the dynamics of plant populations in natural environments (Packer et al., 2000; van der Putten, 2000) and in epidemics in agricultural environments (Shea et al., 2000; Gilligan, 2002). Such epidemics are frequently driven by an external source of infection (primary infection, the initiator of an epidemic and often associated with inoculum present in the soil), and secondary

spread between infected and neighboring susceptibles in a population (secondary infection). Percolation-based modeling may be appropriate for epidemics with short-range contacts, as it can identify if disease is expected to invade in relation to the frequency and distribution of favorable and unfavorable sites in a population (Stauffer et al., 1991; Sander et al., 2002; Sander et al., 2003). In a recent paper, Gibson et al (2006) developed Bayesian methods to fit percolation-based, spatio-temporal models to biological processes that evolve in homogeneous populations. The method is based on a generalization of a standard epidemiological *S-I* model with primary and secondary infection transmission between nearest neighbors on a lattice married with a model for the time-varying susceptibility of the host. This can be fitted to observations of disease spread through time and space in replicated populations using powerful statistical techniques such as Markov chain Monte Carlo.

The model of Gibson et al. can be readily generalized to represent a mixed-species population to give the following. The heterogeneity of the population is described by assigning the covariate h_j to each member j of the population, where h_j is a binary variable representing sites in the population that are either favorable or unfavorable for disease transmission. Let $Y_j(t)$=1 if j is infected by time t and 0 if still susceptible. In common with most stochastic epidemiological models (e.g. Bailey, 1975; Sellke, 1983; Renshaw, 1991; Gibson et al, 2004; Höhle et al, 2005), we assume that $\Pr\left(Y_s(t+dt)=1 \middle| Y_s(t)=0\right)=\phi_s(t)dt$ as $dt \to 0$, where $\phi_s(t)$ is the rate of infection of s. In the specific model used here, $\phi_s(t)$ is comprised of terms representing the rate of *primary* infection from inoculum at time t if s is inoculated, denoted $\alpha[h_s](t)$; the rate of *secondary* infection from each infected neighbor i, denoted $\beta[h_i,h_s](t)$; and the rate of *tertiary* infection from background sources, at rate $\gamma[h_s](t)$. We restrict in our analysis the transmission of primary and secondary infection to nearest neighboring sites only, to accommodate for the limited dispersal commonly found for soil-borne pathogens and for which data for model testing were available. Using the indicator function $\mathbf{1}\{A\}$ = 1 if A is true and 0 otherwise, this can be written:

$$\phi_s(t)=\alpha[h_s](t)\mathbf{1}\{s\in X\}+\left(\sum_i \beta[h_i,h_s](t)\mathbf{1}\{Y_i(t)=1\}\mathbf{1}\{i\in N_s\}\right)+\gamma[h_s](t)$$

where X and N_s are the sets of inoculated hosts and nearest-neighbors of s, respectively. Various functional forms for the transmission rates can be used to represent how plants vary in their response to infective challenge with age. The total rate of infection $\phi_s(t)$ of a host s therefore depends on the time-varying transmission rates as well as on localized conditions (presence of inoculum, neighboring infectious plants) which evolve with time and are different for each host (**figure 1**). This complex interaction between transmission rates, host spatial structure, and the spatial pattern of disease presence means the behavior

of an epidemic cannot be predicted from a single component alone, but requires an integrated understanding of all three (see **figure 1**).

This paper applies a framework for analysis of spread of epidemics in heterogeneous environments within which experimentation, mathematical modeling and statistical inference are integrated. Central to such a framework are methods for parameter estimation including uncertainty associated with these, as well as criteria for formal comparison of models, testing for differences of transmission rates and understanding the relative importance of each phase in driving the epidemic. We illustrate the methods for the spread of *Rhizoctonia solani*, a fungal plant pathogen, in mixed species populations as a convenient experimental system in which heterogeneity can be controlled. The methods, however, are readily generalized to other scenarios. Our specific objectives are:

- to develop methods to fit percolation based, spatio-temporal models for the spread of epidemics that evolve in heterogeneous environments with local dispersal and time-varying transmission rates within and between species,
- to use formal criteria to test for differences in transmission rates within and between species,
- to infer the relative importance of different sources of infection for any particular combination of host spatial structure and transmission rates.

Finally, using parameter estimates, we quantify the effect of changes to environmental heterogeneity. Specifically we address the effect of large-scale (typified by the percentage of area covered by favorable sites) and small-scale heterogeneity (typified by the degree of clustering of favorable sites) on the connectivity within and between favorable and unfavorable sites, and how connectivity and transmission rates determine disease levels, by controlling the way an epidemic invades its environment.

2 Results

Models with various functional forms for the time-varying transmission rates were considered, including constant with time. These were fitted to spatio-temporal data (D) on the spread of damping-off epidemics through mixed populations of radish and mustard seedlings, hereafter referred to as favorable (F) and unfavorable (U) sites respectively. Formal comparison using the deviance information criterion (DIC), resulted in the following model:

$$\alpha[h_s](t) = a[h_s]\exp\left(-r[h_s]t\right),$$

$$\beta[h_i,h_s](t) = b[h_i,h_s]\kappa[h_i,h_s]\left\{\left(t/\mu[h_i,h_s]\right)^{\kappa[h_i,h_s]}\right\}\exp\left\{-\left(t/\mu[h_i,h_s]\right)^{\kappa[h_i,h_s]}\right\}/t,$$

$$\gamma[h_s](t) = \varepsilon[h_s].$$

in which the rate of infection from inoculum decays exponentially and the rate of neighbor-to-neighbor infection is proportional to a Weibull density function. Also included is background infection at a constant rate, which accommodates any infection that may have arisen from spread beyond nearest neighbor (Gibson et al., 2006).

The joint posterior density of this model's parameters has been sampled using MCMC whence *a posteriori* parameter means and 95% credible intervals are derived for the model's parameters (**table** 2). Posterior means and marginal 95% credible intervals for the primary and secondary transmission rates were subsequently derived from the joint posterior distribution, further supporting the rise and fall dynamics in secondary transmission for all species combinations **(figure** 2).

Comparison of transmission rates

An important feature of our method is that it enables testing for differences in primary and secondary transmission rates between favorable and unfavorable sites. Both parameters governing primary infection differed between species ($\Pr(a_U>a_F|D)<0.0001$, $\Pr(r_U>r_F|D)<0.01$). As both species were challenged by the same source of inoculum, this indicates that although the unfavorable sites were initially more susceptible to inoculum, they became resistant quicker (**figure** 2a,b).

Interpretation of the individual parameters for secondary infection is more complicated, since each set of three parameters contributes jointly to the shape of the secondary transmission rate (**figure** 2c-f). All four within- and between-species transmission rates displayed similar temporal dynamics: the rate of transmission initially increased, followed by a decrease. The absolute values for each of the four transmission rates were however different (**figure** 2c-f). The statistical significance of these differences was confirmed by comparing DIC scores for aggregated models, in which some secondary

transmission rates were assumed to be identical so that, for example, they depend only on donor or recipient sites. For all tested aggregated models, the DIC scores were substantially worse than those of the full model (**table** 1), showing that all four transmission rates were distinct.

The largest of the secondary transmission rates was found to be transmission between two favorable sites, followed by the transmission of infection from unfavorable to favorable. The rate at which unfavorable sites became infected was lower, in particular for transmission between two unfavorable sites. Overall, however, there was an appreciable rate for between-species transmission of infection (**figure** 2d,e). Note the difference in ranking: unfavorable hosts are more resistant to secondary infection but less to primary infection than the favorable species.

Goodness of fit

We plot the predicted and measured daily distribution of new infections as a measure of goodness of fit (**figure** 3). It is striking that with a single set of parameters, the prediction of the number of new infections for each day agrees well with the measured data for all population structures, representing a wide range of heterogeneity. Although the predicted distributions follow the central trend of the observations, the amount of variability in the number of new infections is under-estimated. This may indicate environmental differences between replicates that we have not modeled. Possible remedies are addressed in the discussion.

Identification of transmission pathways and dominant pathways of infection

A further novel feature of our approach is that it enables the identification of the posterior distribution of how many plants in a mixed population became infected by primary infection, how many by secondary infection from the same sites (e.g. favorable-favorable or unfavorable-unfavorable), and how many from transmission between-sites (**table 3)** . For example, that in the pure favorable populations, the majority of sites (58%) became infected by secondary infection, compared with 5% by primary infection. In a pure unfavorable population, the origin of infection is more balanced, with 6% and 8% of plants infected by primary and secondary infection, respectively. In heterogeneous mixed populations, doubling the fraction of the unfavorable sites from 25% to 50% of the total population significantly increased the infections caused by unfavorable hosts (**table** 3). Only through fitting of spatial models is it possible to obtain estimates of the most likely pathways of transmission of infection.

Effect of heterogeneity

Heterogeneity can vary at the large scale (e.g. the area covered by favorable

sites) or at the small scale (e.g. the clustering of favorable sites within an area). We used our model with the estimated parameters to understand the effect of environmental heterogeneity upon dynamics of epidemics, and in particular to test how this is affected by presence or absence of transmission between favorable and unfavorable sites. Presence or absence are selected here as examples to demonstrate the importance of knowledge of the four transmission rates. The predicted levels of disease changed substantially with an increase in area of favorable sites in the population (and concomitantly reduction in area of unfavorable sites). Not surprisingly, as the connectivity between favorable sites increases with the proportion (**figure 4a**) or clustering (**4c**), the disease level increased. As transmission involving unfavorable sites occur at a lower rate (**figure 3**), this pattern was the same regardless of whether transmission between favorable and unfavorable sites can occur.

For unfavorable sites the results are perhaps surprising. The response to an increase in area covered or an increase in clustering depended on the absence or presence of transmission between favorable and unfavorable sites (**figures 4b and 4d**). For example, as the proportion of unfavorable sites in the population declines, the expected disease level *decreases* if no transmission occurs between favorable and unfavorable sites, but *increases* if transmission can occur. The reason for this is apparent from **figures 4e and 4f**, which show that with increasing proportion of favorable sites, the unfavorable sites no longer form a connected network, but do have more contacts with neighbors that are favorable for spread. Additional simulations (not shown) showed that this pattern does not depend on the level of primary infection, or on the time-dependency of the transmission rates, but does depend on the relative value of these. The contrasting response to heterogeneity can therefore only be understood if all transmission rates that can occur within a heterogeneous population are known.

3 Discussion

Approach: Stochastic spatial models combined with recent advances in computational methods for statistical inference provide a powerful tool to analyze and understand epidemics in heterogeneous environments. We have presented and tested a framework within which experimentation, parameter estimation and modeling are integrated with implications for analysis of spatio-temporal data. We demonstrated the framework on damping-off epidemics and showed how it enables: (i) estimation of multiple transmission rates from spatio-temporal data of disease presence in heterogeneous environments, (ii) formal comparison of models and tests for significant differences between transmission rates, (iii) identification of the main sources and pathways of infection, (iv) analysis, using parameter estimates, of how epidemics evolve in response to changes to the heterogeneity of their environment. The framework is an essential tool in optimizing the deployment of various control strategies to prevent invasion of epidemics in heterogeneous environments.

Contact structure: The effect of spatial contact patterns of homogeneous hosts on epidemics has been the subject of considerable research effort in recent years, in response to the growing recognition that spatial structure plays a critical role in understanding and predicting the risk of disease outbreaks. Much of this research has been theoretical, and has shown how models of epidemics behave differently when spatial structure is accounted for (Dushoff and Levin (1995), Keeling and Eames (2005)) with important implications for disease management and control (Smith et al (2004), Watts and Strogartz (1998), Pastor-Satorras and Vespignani, 2001 & 2002; Dybiec et al, 2004) and the evolution of virulence in pathogens (Buckee et al (2004), Read and Keeling (2003)). Other work has involved describing the effects of heterogeneity in contact structures using moment closure (Bolker and Pacala, 1997; Keeling, 1999; Boots and Sasaki, 1999), or time-delays and functional responses (Keeling *et al.* (2000)).

Fitting models with heterogeneous host structure: In contrast, relatively little research has been done on fitting process-based models that take account of the spatial structure of environmental or host heterogeneities to observed disease-incidence data. Exceptions include the spread of rabies in a spatially heterogeneous Connecticut raccoon population (Smith et al (2002)). Similarly, Cook et al (to appear) show that the predicted spread of the alien weed *Heracleum mantegazzianum* in Great Britain is grossly overestimated if landscape heterogeneities are not taken into account. A key difficulty in fitting these models to data is that the contact-network structure is often unknown. Assumptions need to be made, and techniques for formal model selection

presented in this paper form an essential step in this process. The structure of heterogeneity also has implications for experimental design, with spatial arrangements differing in the information yielded from the system.

Distinguishing transmission rates: We developed our methods for a generic model for plant epidemics (Gilligan, 2002) where plants are classified as either susceptible (*S*) or infected (*I*). The model does not include any hidden classes, such as latently infected plants, that would have complicated considerably the estimation of the transmission parameters (Gibson et al., 2004). An important feature of the model is that the transmission rates from inoculum and infective plants are allowed and found to vary with time. Transmission rates are notoriously difficult to quantify (Dwyer et al., 1993). Time-dependency of transmission rates is often not known a priori, yet plays a crucial role in the dynamics of epidemics (Kleczkowski et al., 1996; Filipe et al., 2003) especially when it leads to rapid quenching of disease spread. In this paper we described a method that through parameter estimations will enable formal comparison of models to identify appropriate functions for time-dependency of transmission rates. In our specific example, the infection rate of both species from inoculum declined with time, while plant-to-plant infection rose and fell as the plants aged. This is consistent with analysis of epidemics in similar systems, and biological grounds for this have been described elsewhere (Otten et al., 2003). We used the DIC to test for further model simplification relating to specific biological hypotheses, in particular this allowed us to reject the hypotheses that inter-specific transmission rates depended only on donor or recipient type in favor of an interaction between the two. A range of simpler, commonly-used functional forms including *a priori* constant rates of infection were similarly rejected by means of the DIC (not shown).

Model selection: Use of a post-hoc measure of relative model goodness of fit, such as the DIC, allows additional candidate models to be considered in a convenient, modular fashion. Alternatives to the DIC for model selection within a Bayesian framework were considered, but do not have this flexibility. For example, models as well as priors can be assigned prior probability distributions and the model with highest *a posteriori* support used for analysis (see Green, 1995 and King et al, 2006), but in practice this requires the use of reversible jump MCMC to sample the space of all candidate models, increasing computation times substantially, and requiring the routine to be rerun each time new models are devised. Another post-hoc alternative is to extend the analysis of the distribution of stochastic residuals (Gibson et al, 2006) to multiple host types, which would have the benefit of giving an absolute rather than relative measure of model goodness of fit.

Future work: Although the fitted model characterized the observed pattern of new infections well, there was additional variability in the data (**figure** 3), in

common with studies in homogeneous populations (Gibson et al, 2006). Adapting the model used here to characterize such additional variability requires more complexity in the structure of the model than is used here; further work is underway to assess the feasibility of such analyses. For example, hierarchical modeling, in which parameters vary between replicates according to a hyper-distribution, would inflate the variance and allow any small environmental differences between replicated epidemics to be quantified. A further generalization is to incorporate synergism, or non-independence of multiple sources of infection, which has been observed in simpler systems. In both cases, the apparent simplicity of the model would be lost, and interpretation would become abstruse.

The model and approach readily generalize to multiple phases or host types. If h_j, the heterogeneity covariate of host j, remains categorical, the extension is obvious, although the number of parameters increases quadratically with the number of phases. If h_j represents a continually varying trait, then some functional form must be imposed for the rate $\beta[h_i,h_s](t)$ of infection from i to s: for example, that $\beta[h_i,h_s](t)$ is proportional to h_s and constant with time (O'Neill and Becker, 2001).

Importance of approach: The main feature of the framework is that it allows for analysis of epidemics in heterogeneous environments whilst accounting for two crucial aspects underlying the epidemics, namely (i) the contact between sites in the population, and (ii) real estimates of multiple transmission rates that operate in heterogeneous systems. We have shown that both factors affect the dynamics of epidemics and hence the effectiveness of disease control strategies. Such strategies could include a spatial distributions of control by shielding susceptible hosts or fields, for example by a spatial deployment of resistant varieties, or a local deployment of a chemical or biological control agent (Brophy et al., 1991; Garrett et al., 1999). The effectiveness is largely determined by the relative magnitude of the transmission rates. If transmission rates between favorable and unfavorable sites are intermediate or high, we showed that the levels of disease in sites that are less favorable for spread are dominated by the disease pressure from the favorable sites in the population (**figure** 3). Hence the heterogeneity of the population determines the underlying landscape within which contacts between infected and susceptible sites, either favorable or unfavorable, are subsequently dynamically generated by the pathogen as it explores specific contacts preferentially. The latter is mainly determined by the relative magnitude of the transmission rates. Knowledge of multiple transmission rates as estimated in this paper is therefore essential in addressing epidemiologically important questions such as: '*what is the minimum spatial coverage of a resistant variety required to reduce the risk of invasion?*'; or '*can we safely introduce a susceptible crop or cropping system (e.g. organic farms) without enhancing the risk of invasive spread at the regional scale?*'. These questions can only be addressed within spatial models.

Our method ensures that answers to such questions are statistically sound, and fully integrated with experimental trials.

4 Theory and approaches

Bayesian model fitting

Following Gibson et al. (2006), we fit the model by deriving the form for the likelihood function of the parameter vector (call this $\boldsymbol{\theta}$) conditioned on the sets of non-observed infection times ($\mathbf{t} = \{\tau_s = \min t : Y_s(t) = 1\}$) and sources of infection ($\boldsymbol{\sigma}$). The posterior distribution for $\boldsymbol{\theta}$ given the observed replicate data (D) is $f(\boldsymbol{\theta}|D) \propto f(\boldsymbol{\theta}) \iint f(\mathbf{t}, \boldsymbol{\sigma}|\boldsymbol{\theta}) d\boldsymbol{\sigma} d\mathbf{t}$ with the region of integration being limited to the set of $(\boldsymbol{t}, \boldsymbol{\sigma})$ consistent with D. The $f(\boldsymbol{t}, \boldsymbol{\sigma} \mid \boldsymbol{\theta})$ term is defined by the model and a result in Cox and Isham (1980) to be:

$$f(\mathbf{t}, \boldsymbol{\sigma}|\boldsymbol{\theta}) = \prod_{s:Y_s(t_{end})=1} \lambda_s(\tau_s) \exp\left\{-\int_0^{\tau_s} \phi_s(u)du\right\} \prod_{s:Y_s(t_{end})=0} \exp\left\{-\int_0^{\tau_{end}} \phi_s(u)du\right\}$$

where λ_s is the rate of infection exerted on s by the infecting source, σ_s. Details for a related model with homogeneous hosts can be found in Gibson et al. (2006). It is preferable to use functional forms for $\phi_s(t)$ that have analytical integrals. The prior, $f(\boldsymbol{\theta})$, is a subjective description of how likely different parameter values are before the experiment is undertaken. We take relatively broad, independent priors, namely exponential with mean 10 for all parameters except those for tertiary infection, which are exponential with mean 0.1. Evaluation of the posterior is hampered by the presence of the integrals over the space of parameters and the unknown times and natures of events. The integral has over 10 000 dimensions. Nonetheless, Markov chain Monte Carlo techniques (see Gilks et al., 1996, for a review) are capable of drawing a sample from the joint distribution of ($\boldsymbol{\theta}$, $\boldsymbol{t}$, $\boldsymbol{\sigma}$) that then provides an estimate of the posterior distribution of $\boldsymbol{\theta}$ from which summary statistics can be made (e.g. Gibson and Renshaw, 1998, 2001; O'Neill and Roberts, 1999). The sample also allows the distribution of any function of the parameters to be inferred, such as $\beta[h_i,h_s](t)$. Posterior distributions of times and sources of infection can also be inferred, the latter allowing us to find average proportions of infection by transmission pathway.

The sample was also used to generate a posterior predictive distribution of the stochastic process through Monte Carlo simulation to explore hypothetical scenarios such as the effects of varying heterogeneity. Additionally, the predictive distribution of the daily increment in infection conditional on the existing (observed) pattern of disease was used to assess the goodness-of-fit of the model.

Model selection

The DIC was used to select the best model from a set of competing models with

varying functional forms for primary and secondary infection; it was also used to test for differences in secondary infection rates within and between species. The definition of the DIC used here was given by Celeux et al. (2006) (DIC_8), and is obtained by first sampling the posterior before using the parameter means to estimate the augmented variables and hence the effective number of parameters of the model. Following the logic in ref (Spiegelhalter et al, 2002; Burnham and Anderson, 1998), we reject models with a DIC greater than 2 of the model with the lowest DIC score.

Experimental data

We demonstrate our model for epidemics in heterogeneous plant populations in replicated microcosm experiments. Details can be found in Otten et al. (2005). In summary, dynamics of damping-off epidemics were recorded in populations comprising 414 seedlings of a favorable (radish, *Raphanus sativus* L., *Cherry Belle*) or an unfavorable species (mustard, *Sinapsis alba* L.), planted in a square lattice. At the densities used, spread of disease occurs predominantly between nearest neighbors. Populations comprised either 100% favorable, 100% unfavorable, a mixture with 75% favorable and 25% unfavorable, or a 50–50% mixture, with up to 6 replicates per treatment. The host species at each point on the lattice was randomly selected, and in each tray 32 randomly selected plants were challenged by inoculum of the soil-borne fungal pathogen *Rhizoctonia solani*. The position of damped-off plants was recorded daily for 16 days following inoculation. The model was fitted to all replicates jointly.

Analysis of environmental heterogeneity

Once the posterior distribution of the parameters has been estimated, we simulate the role of environmental heterogeneity. Large-scale heterogeneity is investigated by changing the proportion of favorable and unfavorable sites in the population, allocating these randomly. For a selected large-scale heterogeneity (50% of each), small-scale heterogeneity is investigated by dividing the population into square sub-plots of side {1,…,6,9} (plus remainders at edges) and allocating these equiprobably to the two favorable or unfavorable sites. A measure of clustering K is defined to be the proportion of neighboring pairs of hosts of the same type. For each scenario, the estimated parameters are used to analyze the effect of heterogeneity on disease level. To demonstrate the importance of the transmission rates, we consider an alternative scenario in which between-species infection is not possible (representing for example specialized pathogens). In all cases, 10 000 simulations of each scenario and spatial arrangement are used to generate the posterior predictive distribution of disease levels.

Acknowledgements
AC and GM gratefully acknowledge financial support from the Scottish Executive Environment and Rural Affairs Department (SEERAD).

References
Brachet S, Olivieri I, Godelle B, Klein E, Frascaria-Lacoste N, Gouyon PH (1999) J Theor Biol 198: 479-495.
Brophy LS, Mundt CC (1991) Agr Ecosyst Environ 35: 1-12.
Burdon JJ, Chilvers GA (1977) Oecologia 28: 141-146.
Burnham KP, Anderson DR (1998) Model selection and inference (Springer, New York).
Dwyer G, Elkinton JS (1993) J Anim Ecol 62: 1-11.
Ettema CH, Wardle DA (2002) Trends Ecol Evol 17: 177-183.
Filipe JAN, Gilligan CA (2003) Phys Rev E 67: 021906-021908.
Finckh MR, Gacek ES, Goyeau H, Lannou C, Merz U, Mundt CC, Munk L, Nadziak J, Newton AC, de Vallavieille-Pope C, Wolfe MS (2000) Agronomie 20: 813-837.
Garrett KA, Mundt CC (1999) Phytopathology 89: 984-990.
Garrett KA, Mundt CC (2000) Phytopathology 90: 1313-1321.
Gibson GJ, Gilligan CA, Kleczkowski A (1999) Proc R Soc London Ser B 266: 1743-1753.
Gibson GJ, Kleczkowski A, Gilligan CA (2004) Proc Natl Acad Sci USA 101: 12120-12124.
Gibson GJ, Otten W, Filipe JAN, Gilligan CA, Cook A, Marion G (2006) Stat Comput 16: 391-402.
Gilligan CA (1987) in Populations of Plant Pathogens: Their Dynamics and Genetics, eds Wolfe MS, Caten CE (Blackwells, Oxford) pp 119-133.
Gilligan CA (2002) Adv Botanical Res 38: 1-64.
Gilligan CA, Kleczkowski A (1997) Phil Trans R Soc London Ser B 352: 591-608.
Green PJ (1995) Biometrika 82: 711-732.
Jeger MJ (2000) Plant Pathol 49: 651-658.
King R, Brooks SP, Morgan BJT, Coulson T (2006) Biometrics 62: 211-220.
Kleczkowski A, Bailey DJ, Gilligan CA (1996) Proc R Soc London Ser B 263: 777-783.
Kleczkowski A, Gilligan CA, Bailey DJ (1997) Proc R Soc London Ser B 264: 979-984.
Levin SA (2000) Ecosystems 3: 498-506.
Mundt CC (2002) Annu Rev Phytopathol 40: 381-410.
Newman MEJ (2003) Phys Rev E 67, art. no.-026126.
Newman MEJ, Forrest S, Balthrop J (2002) Phys Rev E 66, art. no.-035101.
O'Neill PD, Becker NG (2001) Biostatistics 2: 99-108.
Otten W, Filipe JAN, Gilligan CA (2005) Ecology 86: 1948-1957.
Otten W, Filipe JAN, Bailey DJ, Gilligan CA (2003) Ecology 84: 3232-3239.

Packer A, Clay K (2000) Nature 404: 278-281.
Park AW, Gubbins S, Gilligan CA (2002) Ecol Lett 5: 747-755.
Sander LM, Warren CP, Sokolov IM (2003) Physica A 325: 1-8.
Sander LM, Warren CP, Sokolov IM, Simon C, Koopman J (2002) Math Biosci 180: 293-305.
Shea K, Thrall PH, Burdon JJ (2000) Ecology Letters 3: 150-158.
Stauffer D, Aharony A (1991) Introduction to Percolation Theory (Taylor & Francis, London).
van der Plank JE (1965) Science 147: 120-124.
van der Putten WH (2000) Nature 404: 232-233.
Zhu YY, Chen HR, Fan JH, Wang YY, Li Y, Chen JB, Fan JX, Yang SS, Hu LP, Leung H, Mew TW, Teng PS, Wang ZH, Mundt CC (2000) Nature 406: 718-722.

Tables:

Table 1: Difference in deviance information criteria (DIC) relative to the best-fitting model, used to confirm absence of common parameters for secondary transmission rates, associated with the donor, the recipient, or both.

Model		DIC
Full		0
Donors same	$\beta[FF](t)=\beta[UF](t); \beta[FU](t)=\beta[UU](t)$	25
Recipients same	$\beta[FF](t)=\beta[FU](t); \beta[UF](t)=\beta[UU](t)$	482
Cross-infection same	$\beta[FU](t)=\beta[UF](t)$	50
All rates same	$\beta[FF](t)=\beta[FU](t)=\beta[UF](t)=\beta[UU](t)$	921

Table 2: Posterior means and 95% credible intervals for the model parameters, fitted to 19 replicates of population comprising different ratios of hosts favorable (*F*) and unfavorable (*U*) for damping-off infection caused by the soil-borne fungal plant pathogen *R. solani*.

Θ		posterior	
		mean	95% CI
Primary scale	*a[F]*	0.16	(0.11 , 0.17)
	a[U]	0.26	(0.20 , 0.32)
Primary resistance	*r[F]*	0.06	(0.01 , 0.12)
	r[U]	0.16	(0.11 , 0.22)
Secondary scale	*b[FF]*	1.4	(1.3 , 1.4)
	b[FU]	0.5	(0.4 , 0.6)
	b[UF]	1.2	(0.9 , 1.6)
	b[UU]	0.4	(0.4 , 0.6)
Secondary variability	κ*[FF]*	3.5	(3.3 , 3.8)
	κ*[FU]*	3.8	(2.9 , 4.6)
	κ*[UF]*	2.9	(2.1 , 3.6)
	κ*[UU]*	2.4	(1.7 , 2.9)
Secondary timing of peak	μ*[FF]*	9.3	(9.1 , 9.5)
	μ*[FU]*	9.5	(9.0 ,10.0)
	μ*[UF]*	10.2	(9.0 ,13.1)
	μ*[UU]*	9.4	(8.3 ,12.1)
Tertiary	ε*[F]*	0.012	(0.011, 0.014)
	ε*[U]*	0.004	(0.003, 0.005)

Table 3: Identification of the most likely pathways and sources of transmission of infection in mixed populations, estimated by fitting the model to replicated epidemics of mixed populations with different ratios of favorable (F) and unfavorable (U) plants. The values show the posterior mean (expressed as % of F or U present in the population) of plants that became infected by primary infection, secondary infection from favorable plants, or secondary infection from an unfavorable plant. Tertiary infection accounted for the infection of a further 11–15% of favorable and 4–5% of unfavorable. Standard deviations in parentheses.

	100% F	75% F		50% F		100% U
species	F	F	U	F	U	U
Primary infection	4.6 (0.2)	4.5 (0.3)	5.8 (0.3)	4.2 (0.3)	5.8 (0.2)	5.7 (0.2)
Secondary infection from F	58.4 (0.5)	45.4 (0.7)	27.1 (0.9)	25.7 (0.7)	14.5 (0.6)	-
Secondary infection from U	-	5.0 (0.5)	2.0 (0.5)	13.1 (0.6)	9.3 (0.5)	7.9 (0.3)

Figures:

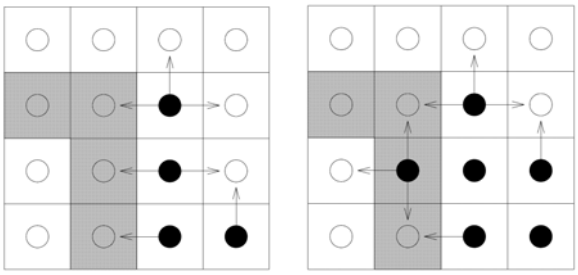

Figure 1: Dynamics of infection in a heterogeneous host population comprising unfavorable (U, grey squares) and favorable (F, white) sites. The state of an epidemic is shown for two time points, with solid circles indicating infected and hollow circles susceptible hosts. The arrows represent transmission routes between infected and neighboring uninfected sites. The figure shows how the localized conditions change with time, and how the disease load depends on the relative strength of each of the four possible transmission rates. In the left picture, the total infective challenge on F is $4\beta[F,F]$ and $3\beta[F,U]$ on U. As more hosts become infected (right), this changes to $3\beta[F,F]+\beta[U,F]$ on F and $2\beta[F,U]+2\beta[U,U]$ on U. Even though the number of transmission pathways has increased from seven to eight, the combined effective challenge may be greater or less than in the left-hand picture depending on the exact relative strength of each of the four transmission rates.

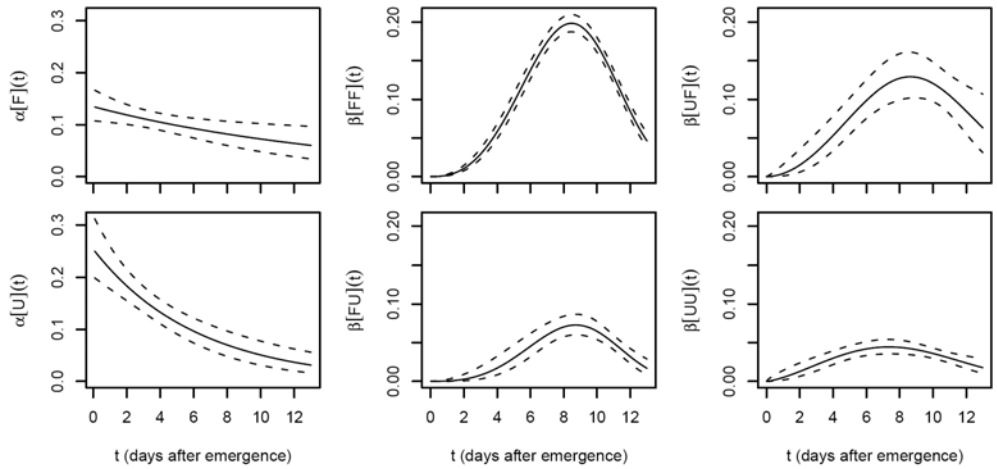

Figure 2: posterior mean and 95% credible interval for primary (*α[F](t), α[U](t)*) and each of the four secondary transmission rates (*β[FF](t), β[FU](t), β[UF](t), β[UU](t)*) against time.

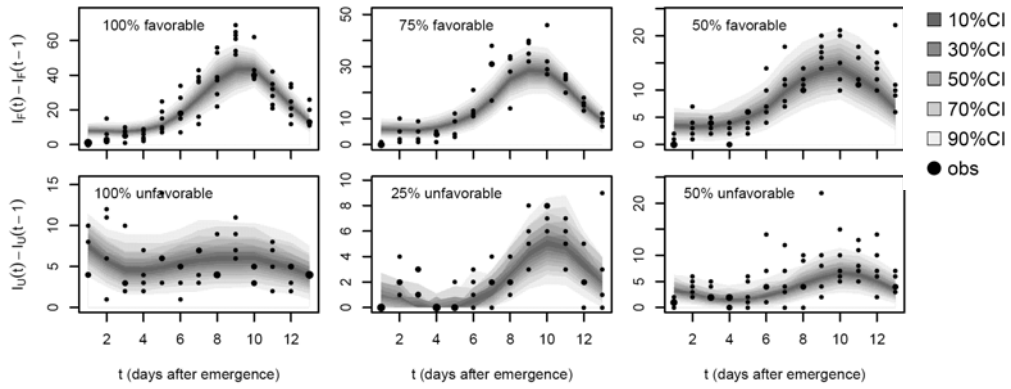

Figure 3: Posterior predictive new infections of daily increments (shaded region) with observations (points, area of symbols are proportional to the number of observations) for various global heterogeneity schemes. Predictions take the form of mixture distributions with each component conditional on the previous spatial observation of disease presence in an experimental replicate. Predictions take account of both parametric uncertainty and population stochasticity. Denote by $I_F(t)$ and $I_U(t)$ the number of infected favorable and unfavorable hosts at time t, respectively.

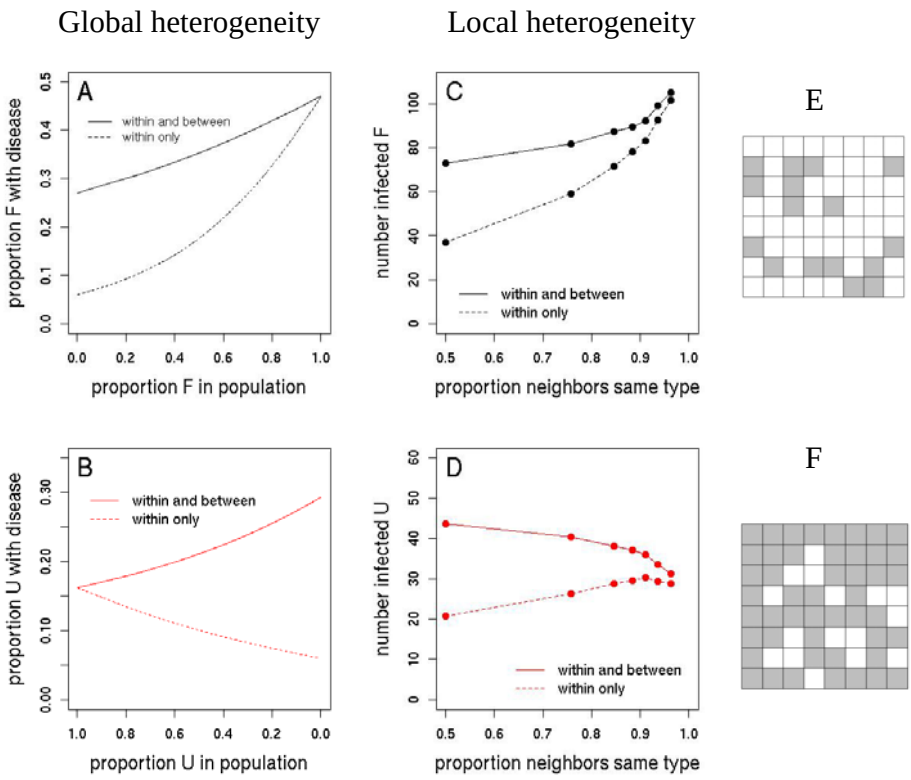

Figure 4: Effect of proportion of favorable (A) or unfavorable (B) sites in the population and the proportion of neighbors of the same type (measure of clustering) on the levels of disease (C&D), with (continuous lines) or without (dashed lines) inclusion of transmission of disease between favorable and unfavorable sites in the population. At 75% of unfavorable sites in the population (E, white squares), they form a well connected network. At 25%, the unfavorable sites form isolated patches, and the fraction of contacts with neighboring sites favorable to spread (F, grey squares) has significantly increased, leading to an increased disease pressure.