How Does Forets Fragmentation Relate to Epidemic Volatility in the Case of Sudden Oak Deathg?

December 16, 2017

1 Introduction

The dynamics of epidemic communities with heterogeneous host populations are fairly complex (so-called "diversity-disease relationships"), and even well-accepted hypotheses such as the dilution effect are turning out to be more nuanced than previously thought (Johnson, Ostfeld, and Keesing 2015; Randolph and Dobson 2012). One interesting wrinkle in diversity-disease relationships that has so far received relatively little attention is the so-called *variance reduction effect* (Fig. 1). This refers to a commonly observed phenomenon in disease ecology where disease risk tends to be less variable in communities with high species richness than in their less speciose counterparts (Fig. 1). Recently I performed a modelling analysis which suggests that, when disease is density dependent (ie. when transmission increases with host density), a key factor in producing a variance reduction is the level of pathogen transmission between host sub-populations. These results suggest that when disease communities exhibit a high degree of correlation in host population infection risk they may become volatile. From a community ecology perspective this idea is not new, the relationship between community complexity and stability has been pointed out in food-web dynamics (May 2006) as well as community biomass (Doak et al. 1998), but as far as I can tell relatively little work has been done exploring these ideas in relation to disease ecology. Epidemics in plant communities are situated at a unique intersection between landscape pattern and epidemic dynamics and I think they will provide an excellent venue for bringing these ideas to disease ecology (Holdenrieder et al. 2004; Plantegenest, Le May, and Fabre 2007).

The relationships between community-level factors and the variance or volatility of epidemic dynamics have important consequences for managers or policy-makers. More and more epidemic management strategies rely on statisical models for forecasting pathogen spread or comparing the efficacy of control strategies. However when epidemic dynamics are volatile, forecasting them over policy-relevant timescales becomes extremely difficult. Developing understanding about when the underlying systemic uncertainty is high is therefore crucial for using forecasting tools properly.

In this work I prototype a data analysis of the relationship between the the variability in disease risk between plant communities and the spatial fragmentation of those communities. Specifically I will be considering the case of Sudden Oak Death, which has a number of attractive features for this purpose. First, Sudden Oak Death infects a large number of hosts, with the consequences of infection ranging from leaf blight to rapid death. This degree of host diversity means that the disease is ideal for studying epidemic dynamics in heterogeneous host populations. Second, because management of the

disease is largely focused on preventation and containment, Sudden Oak Death is well-monitoredwith a lot of data and models for tracking and predicting its spread. Finally, the transmission of Sudden Oak Death has a strong spatial component (Rizzo, Garbelotto, and Hansen 2005) and forest fragmentation has been shown to have some effect on pathogen transmission (Condeso and Meentemeyer 2007). Since forest fragmentation is a suitable proxy for pathogen transmission between sub-communities (ie. "pockets") of the forest, then it also provides a good metric for assessing the relationship between interpopulation transmission and epidemic volatility.

This project has two components: constructing a synthetic data set and designing a statistical model to analyze its output. Synthetic data was generated by my implementation of a popular model of the spread of Sudden Oak Death, using parameter value determined from epidemic data (Meentemeyer et al. 2011). The statistical analysis used a hierarchical model of epidemic growth rates in forets communities with different levels of fragmentation. The hierarchical nature of this model is a key element of this analysis, as it is designed to assess the role of "hidden treatments" (Huston 1997). Typically these refer to a class of phenomena observed in community ecology where community features may appear to relate to species richness as a statistical artifact of the experimental design. In the case of landscape fragmentation, it may be that there is more or less variability between host communities with low amounts of fragmentation. Even if there was no interaction between fragmentation and epidemic volatility, the changing variance on the level of the landscape would produce an apparent variance effect in pathogen growth rates. By employing a hierarchical model we can rigorously seperate these effects from meaningful ecological relationships in the data.

2 Methods

2.1 Natural History

Sudden Oak Death (SOD) is caused by *Phytophthera Ramorum*, a water mold which is known to infect at least 40 genera of plant hosts (Rizzo, Garbelotto, and Hansen 2005). Since 1995 SOD has caused enormous amounts of damage to coastal forest in California and Oregon, and has since spread elsewhere in North America and Europe.

P. Ramorum is transmitted between hosts in a number of ways and over multiple scales. Aerial transmission occurs by the pathogen undergoing sporulation on the surface of infected hosts, and spore dispersion can range from 10-15m (common) to up to 3km (rare). In addition to aerial transmission, P. Ramorum produces zoospores, which are capable of swimming. Rain can therefore transmit the pathogen to adjacent hosts by splashing or runoff, or can introduce spores to large waterways allowing for longer distance tranmssion. Finally human or animal activity may also facilitate long-distance transmission by either the translocation of infected hosts, or by unwittingly carrying dormant spores to uninfected communities. These modes of transmission suggest that landscape and weather can play a large role in the spread of P. Ramorum, which was confirmed by Condeso and Meentemeyer 2007.

2.2 Generating Synthetic Data

My original goal was to use real data on Sudden Oak Death spread. However due to time-constraints, for this project I instead generated synthetic data of a SOD epidemic. This

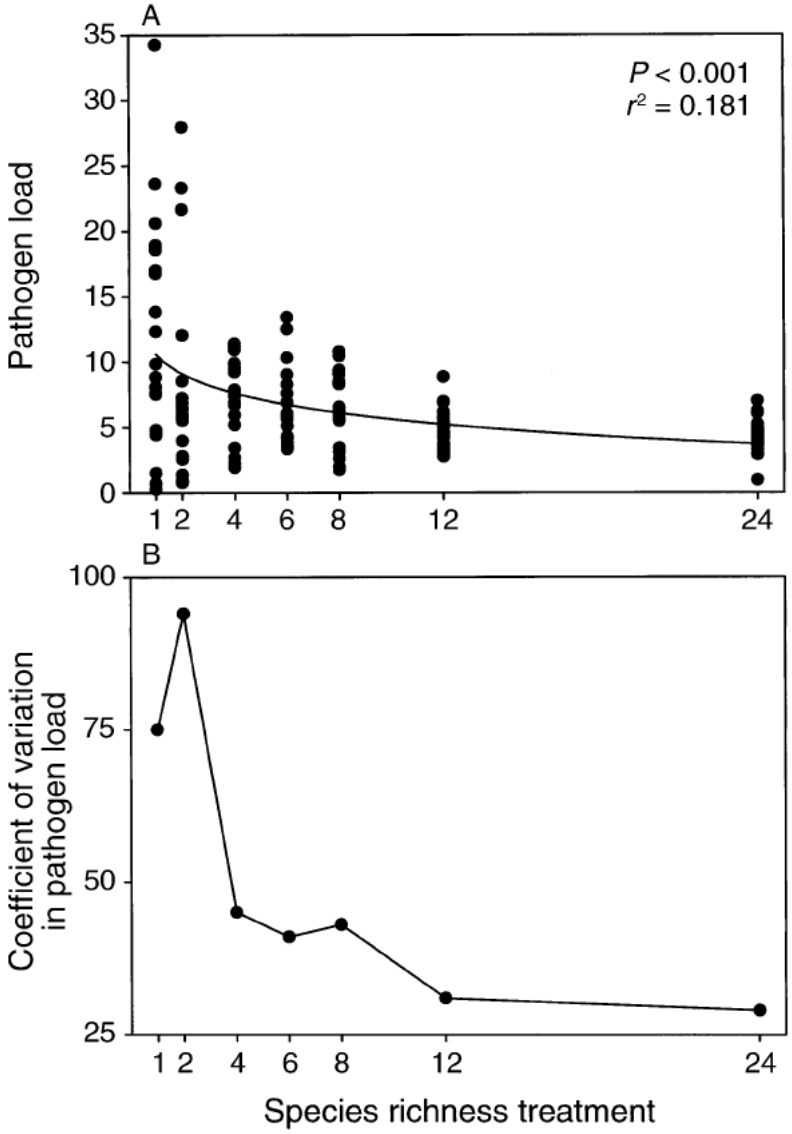
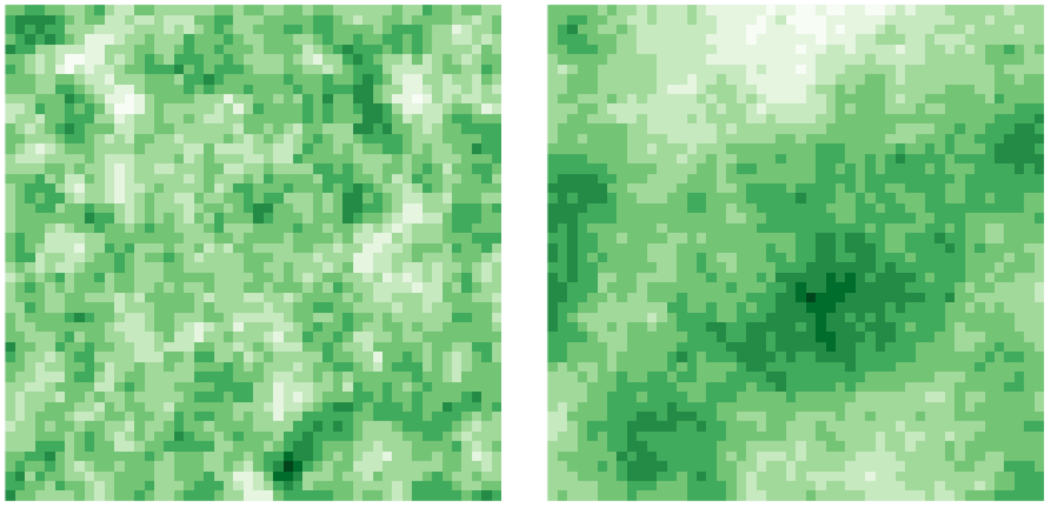


Figure 1: Results from a study of the relationship between plant pathogen load and species richness from Mitchell, Tilman, and Groth 2002 highlighting the dilution effect, but also the variance reduction effect. Note that as richness increases the variance of the pathogen prevalence sharply decreases.



(a) (b)

Figure 2: Two examples of host density data generated using a Gaussian process, dark green cells indicate high host density, lighter cells indicate low density. By adjusting the range of spatial autocorrelation I can control the amount of fragmentation; communities with more autocorrelation, such as (b), are expected to have less fragmentation and thus higher amounts of interpopulation transmission compared to the lower autocorrelation communities such as (a).

was done implementing a simplified model of Sudden Oak Death epidemics, following Meentemeyer et al. 2011. Their model is relatively straightforward, and in its complete form accounts for weather effects as well as heterogeneously susceptible host populations. Furthermore they provide parameter estimates estimated from real epidemic data from the California epidemics of the late 90's and early 00's. While Meentemeyer et. al. run their model on a gridded map of California with real host species data, for simplicity (and because I'm a novice at GIS) I will be using a square region with randomly generated host density and ignoring weather effects. The host density data will be generated using a Gaussian process (Fig. 2). See Appendix 1 for details of the model's operation.

Using this model I simulated an SOD epidemic in 100 different host communities at each of 7 levels of forest fragmentation (so 700 epidemics were simulated in total). Infections were randomly seeded and epidemics were run for 13 weeks, and at each week the total number of infected individuals was tallied and recorded giving $700 \times 13 = 9100$ data points.

2.3 Statistical Model

2.4 Volatility Model

Once the data was generated I disregarded any knowledge of the generation process, and treated the data as if it was "genuine". The aim is now to assess how forest fragmentation impacts the volitility of an epidemic. Let $y_t^{(n,k)}$ denote the number of infected individuals at time t in community $k = 1, ..., 100$ which has fragmentation level $n = 1, ..., 7$. We

assume that the epidemic follows a stochastic growth model, given as follows:

$$\begin{aligned} m_t^{n,k} &\sim N(\theta, \nu) \\ \ln(\frac{y_{y+1}^{n,k} - y_t^{n,k}}{y_t^{n,k}}) &= m_t^{n,k} \end{aligned} \tag{1}$$

The behavior of this model can be understood by rewriting:

$$y_{t+1}^{n,k} = y_t^{n,k}(1 + e^{m_t^{n,k}})$$

Hence $m_t^{n,k}$ is the log of the number of secondary infections produced by infectious hosts during one time-step of the epidemic. This means that ν, the variance of these (log) growth rates, reflects the inherent volatility of the epidemic, while θ gives average growth rate in a single time step.

In this model the $m_t^{n,k}$ are treated as exchangeable random variables; we assume that the passsage of time does not effect the distribution of these log growth rates. This assumption does not hold for an epidemic over long periods of time, the growth rates of an epidemic near its conclusion are generally smaller than those during its initial growth phase. However, over a sufficiently short time window this does not appear to be a particularly poor assumption, with one caveat: I omitted the first two growth rates $m_1^{n,k}$ and $m_2^{n,k}$ from my analysis, as these were almost always inconsistent with the other rates. These "burn-in" growth rates tended to reflect the idiosyncracies of the random infection seed rather than the true dynamics of the population. Furthermore, in a real analysis one does not typically have access to data from the very start of the epidemic, so it is reasonable to omit these data from a prototype analysis.

2.5 Hierarchical Modeling

A main goal aim of this project was to determine not only if forest fragmentation effected the variance of epidemic outcomes, but also whether these changes were "true" effects, or the result of "hidden treaments". Hidden treatments were originally proposed in the context of biodiversity-productivity studies and referred to cases where manipulating the biodiversity of a population inadvertently manipulated features of that population that effected community productivity (Huston 1997). This caused an apparent relationship between species richness and things like community biomass, where none existed. This concept extends fairly straightforwardly to landscape ecology. For example, in the context of this project a hidden treatment might be a relationship between spatial autocorrelation and the "path connectivity" of the host population. Path connectivity here refers to the number of paths connecting two points in the forest that one could walk without leaving the forest; it measures the redudancy of transmission pathways between two host trees in the forest. It seems plausible that path connectivity is more proximally causal of epidemic dynamics than autocorrelation, but because it is difficult to measure connectivity we use autocorrelation instead. It's possible, and indeed seems inuitively probable, that all highly autocorrelated forests have very similar levels of path connectivity, whereas low-autocorrelation forests have a wide variety of path connectivities. If path connectivity strongly determines epidemic severity, and the variance of path connectivity is inversely related to autocorrelation, then it would spuriously appear that the variance of epidemic severity is also inversely related to autocorrelation, even though no such relationship

exists. I would like my analysis methodology to allow for these kinds of relationships, because they have different implications for management strategies.

To this end I used a Bayesian Hierarchical Model, sometimes referred to as "multilevel modeling", "random effects models" or "partial pooling" (Gelman 2006). This treats the data as being drawn from a two-stage sampling process:

$$\begin{aligned} \theta^{n,k} &\sim N(\mu_n, \sigma_n) \\ m_t^{n,k} &\sim N(\theta^{n,k}, \nu_n) \end{aligned} \tag{2}$$

This model allows for variability to change at two levels. First we retain ν_n, the volatility parameter for autocorrelation treatment level n. If we see ν_n changing over different autocorrelation treatments then this implies a "true" effect, autocorrelation is controlling the volatility of epidemic dynamics. On the other hand, we may see variance changes occurring over the σ_n, which controls the variance at the level of the *treatments*. This would suggest that there is a hidden treatment that is being manipulated, rather than an actual effect on the growth rates themselves. Models of this type are fairly straightforward to fit using Bayesian methoods, which I will be performed using the popular statistical modeling languages Stan and R (**˙stan:˙???**),.

3 Results

The Bayesian approach to model fitting produces probability density functions, called "posteriors", over model parameters. An advantage of this approach is that credible intervals (the Bayesian equivalent of confidence intervals) are straightforward to interpret. A 90% credible interval means that we are 90% confident that the true parameter value lies within that interval; the frequentist confidence intervals cannot be interpreted in this way. Therefore I present boxplots which indicate both 50% and 90% credible intervals for each of the treatment-level parameters in the hierarchical model (2). When the boxplots for two parameters overlap significantly it suggests that the two parameters probably have the same value, and the opposite holds when they do not overlap.

Fitting the model given in (2) shows that decreased fragmentation increases the mean epidemic growth rate μ_n as well as the volatility of those growth rates ν_n, but has little to no effect on the treatment level variance σ_n (see Figs 3-5). This suggests that the hierarchical model was unnecessary, and that all variance changes in my synthetic data were due to changing epidemic volatility. I confirmed this by also fitting a non-hierarchical volatility model and comparing it to the fit of the hierarhical model. After penalizing for model complexity (through the use of the Watanabe-Akaike Information Criterion) I determined that the hierarchical model was indeed unecessary.

4 Discussion

The main takeaway of these results is that forest fragmentation does indeed impact epidemic volatility, and that furthermore this relationship is highly non-linear. In Fig. 3 we see two clusters of volatility parameter values, so it appears that the system is crossing some critical threshold of autocorrelation over which the epidemic dynamics go from stable to volatile. I suspect that this thresholding occurs when the autocorrelation causes the forest to switch from many small stands to a handful of larger ones. The

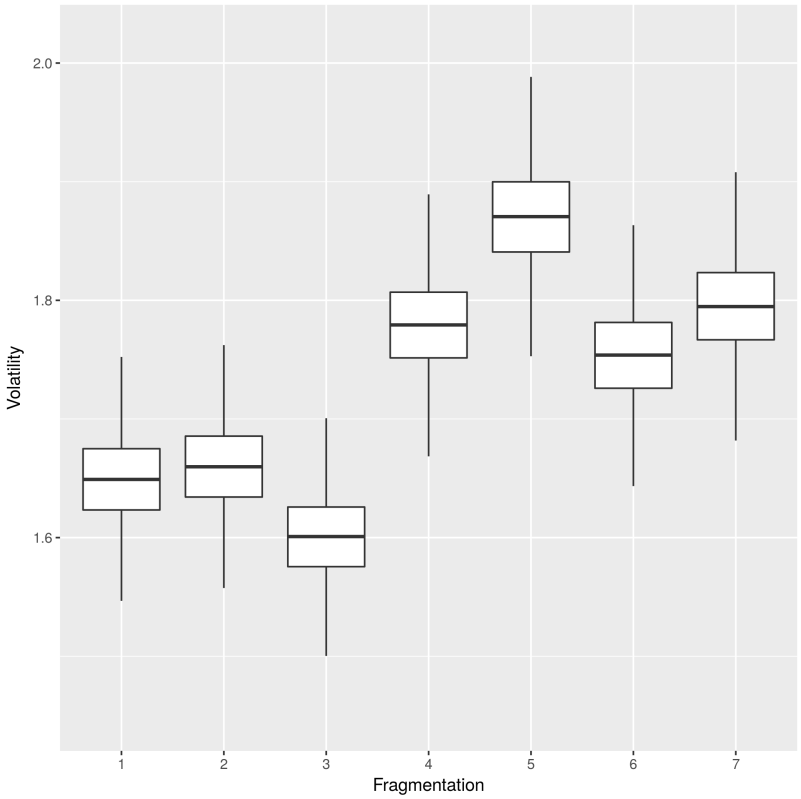


Figure 3: Boxplots of the posterior samples for the epidemic volatility parameter. Based on the 90% credible intervals shown here it seems reasonable to conclude that decreasing fragmentation increases epidemic volatility. Warning: the numbers along the axis are meaningless labels, fragmentation is decreasing to the right (so '7'=lowest fragmentation, '1'=highest fragmentation).

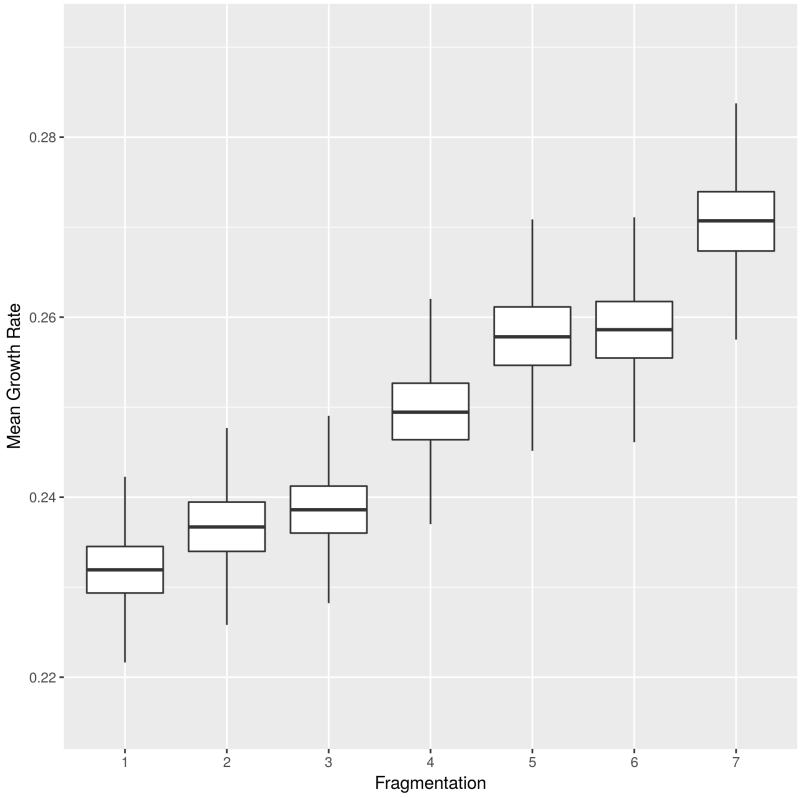


Figure 4: Boxplots of the posterior samples for the treatment-level average growth rate, μ_n. Warning: the numbers along the axis are meaningless labels, fragmentation is decreasing to the right (so '7'=lowest fragmentation, '1'=highest fragmentation).

nonlinearity of this transition is particularly interesting in light of the fact that the relationship between mean epidemic growth and fragmentation actually appears to be fairly linear (Fig. 4). The difference in trend between these two parameters suggests that the volatility changes are not simply due to the epidemic "scaling up", but rather that there is a substantative mechanistic relationship in play.

The substance of this relationship is also confirmed by the lack of change in σ_n. In Fig. 5 we see that the variance between the $\theta^{(n,k)}$ is basically uneffected by fragmentation, which implies that the hierarchical component of the statistical model is not necessary. I did not think this result was likely at the start of the project, but in my opinion it's the most intetersting outcome.

With that said, I am hesitant to put too much stock into these results. My synthetic data model notably lacked any weather effects, which play a major role in the transmission dynamics of P. Ramorum, so all of these results could be an artifact of that particular modeling choice. On the other hand the linear relationship between mean growth rate and fragmentation (Fig. 4) does agree with the Condeso paper, which suggests that the data model is behaving at least somewhat realistically, so it seems to reasonable to expect similar volatility trends to occur in analysis of real data. Another shortcoming of this analysis is that it focuses only on the early-time growth of the pathogen population, which may not capture all subtleties of the epidemic behavior. Going forward it may be better to compare the behavior over multiple time-scales to see if these volatility trends only hold in the initial growth-phase of the epidemic.

I'm not entirely sold on the hierarchical model being unecessary, so going forward I

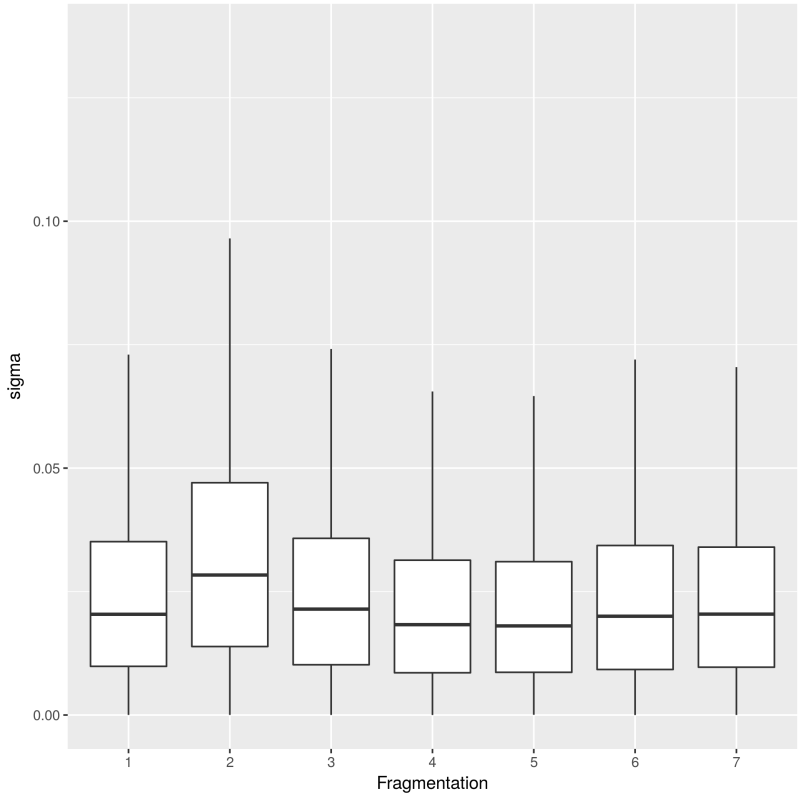


Figure 5: Boxplots of the posterior samples for the treatment-level variance, σ_n. These suggest that fragmentation has no effect on this parameter. Warning: the numbers along the axis are meaningless labels, fragmentation is decreasing to the right (so '7'=lowest fragmentation, '1'=highest fragmentation).

would like to tweak my synthetica data model a little to see if I can get that to change. It's possible that the structure of my synthetic host populations is unrealistic in some key way. I generated the Gaussian Processes with the R package `gstat`, which I found a little limiting, and it may be better to implement these by hand instead, which could allow for finer control of the spatial structure. I'd also like to see what happens if instead of autocorrelation I use some other metrics of interconnectedness, such as path-connectednes, Shannon Entropy, etc.

After sussing out the issues discussed above, my main goal is to put together an actual GIS data set of Sudden Oak Death epidemics and feed them into my analysis pipeline. I do suspect that the volatility relationships I see here will also show up in the data, but I think a hidden treatment effect is likely to also occur and I'm wondering if I won't see more nonlinearities in μ_n as well. Furthermore I think that the volatility might exhibit some temporal dependence that may have important management implications, so assessing how these results change over different time scales will be important. Finally, I would really like to look at how fragmentation interacts with the effect of weather and temperature on pathogen spread, although this would require a much larger data set and a refinement of the hierarchical model (2).

In this project I examined the behavior of a simple model of Sudden Oak Death spread in a randomized host forest with varying fragmentation. I concluded that increased forest fragmentation stabilizes epidemic dynamics, leading to less variability in pathogen prevalence at any time-step. Furthermore I was able to determine that, at least with the synthetic data, there is no hidden-treatment effect which leads me to believe that

Par'm	Def'n	Value
β	Pathogen transmission rate	4.4
S_{it}	Number of susceptible hosts in cell i at time t	-
I_{it}	Number of infect hosts in cell i at time t	-
$N_{\max}$	Maximum number of host units in any cell	-
$K(d;\alpha_1,\alpha_2,\gamma)$	Distance kernel	-
d_{ij}	Dist between cells i and j	-
α_1	Short-range dispersal scale	20.57
α_2	Long-range dispersal scale	$9.5 \times 10^3 4$
γ	Proportion of short range dispersal events	0.9947

Table 1: Parameter definitions and values (where applicable) for the model given in (A.1). Dashes indicate that the parameter had no fixed value.

auto-correlation of host competence may be a direct driver of epidemic dynamics in the case of Sudden Oak Death.

5 Appendix

5.1 Appendix 1

Meentemeyer et. al. treat used gridded GIS data of host plant locations to assign a "host susceptiblity index" to each grid cell in a map of California. This index effectively represents the number of susceptible hosts in the cell. In this project the index for every cell in a 50×50 forest grid was generated as the normalized output of a Gaussian Process. The number of these suscpetibles which become infected in any given week is treated as a Poission random variable, where the rate of this process for cell i at time step t is given as:

$$\psi_{it} = \sum_j \frac{\beta I_{jt} S_{it}}{N_{\max}} \frac{K(d_{ij};\alpha_1,\alpha_2,\gamma)}{d_{ij}} \tag{3}$$

Where parameter definitions and values are given in Table A.1

The distance kernel $K(d_{ij};\alpha_1,\alpha_2,\gamma)$ controls the range of pathogen transmission and is given by $K(d_{ij};\alpha_1,\alpha_2,\gamma) = \gamma(1+(d/\alpha_1)^2)^{-1} + (1-\gamma)(1+(d/\alpha_2)^2)^{-1}$.

References

Condeso, T. Emiko and Ross K. Meentemeyer (2007). "Effects of Landscape Heterogeneity on the Emerging Forest Disease Sudden Oak Death". In: *Journal of Ecology* 95.2, pp. 364–375. ISSN: 0022-0477. URL: http://www.jstor.org/stable/4495989 (visited on 12/10/2017).

Doak, D. F. et al. (1998). "The Statistical Inevitability of Stability Diversity Relationships in Community Ecology". In: *The American Naturalist* 151.3, pp. 264–276. ISSN: 0003-0147. DOI: 10.1086/286117. URL: http://www.jstor.org/stable/10.1086/286117 (visited on 12/14/2016).

Gelman, Andrew (2006). "Multilevel (Hierarchical) Modeling: What It Can and Cannot Do". In: *Technometrics* 48.3, pp. 432–435. ISSN: 0040-1706, 1537-2723. DOI: 10.1198/004017005000000661. URL: http://www.tandfonline.com/doi/abs/10.1198/004017005000000661 (visited on 11/12/2017).

Holdenrieder, Ottmar et al. (2004). "Tree diseases and landscape processes: the challenge of landscape pathology". In: *Trends in Ecology & Evolution* 19.8, pp. 446–452. ISSN: 0169-5347. DOI: 10.1016/j.tree.2004.06.003. URL: http://www.sciencedirect.com/science/article/pii/S0169534704001703.

Huston, Michael A. (1997). "Hidden treatments in ecological experiments: re-evaluating the ecosystem function of biodiversity". In: *Oecologia* 110.4, pp. 449–460. ISSN: 0029-8549, 1432-1939. DOI: 10.1007/s004420050180. URL: http://link.springer.com/article/10.1007/s004420050180 (visited on 12/14/2016).

Johnson, Pieter T. J., Richard S. Ostfeld, and Felicia Keesing (2015). "Frontiers in research on biodiversity and disease". In: *Ecology Letters* 18.10. Ed. by Eric Seabloom, pp. 1119–1133. ISSN: 1461023X. DOI: 10.1111/ele.12479. URL: http://doi.wiley.com/10.1111/ele.12479 (visited on 12/14/2016).

May, Robert M. (2006). "Network structure and the biology of populations". In: *Trends in Ecology & Evolution* 21.7, pp. 394–399. ISSN: 01695347. DOI: 10.1016/j.tree.2006.03.013. URL: http://linkinghub.elsevier.com/retrieve/pii/S0169534706001005 (visited on 11/12/2017).

Meentemeyer, Ross K. et al. (2011). "Epidemiological modeling of invasion in heterogeneous landscapes: spread of sudden oak death in California (1990–2030)". In: *Ecosphere* 2.2, art17. ISSN: 2150-8925. DOI: 10.1890/ES10-00192.1. URL: http://doi.wiley.com/10.1890/ES10-00192.1 (visited on 11/09/2017).

Mitchell, Charles E., David Tilman, and James V. Groth (2002). "Effects of Grassland Plant Species Diversity, Abundance, and Composition on Foliar Fungal Disease". In: *Ecology* 83.6, pp. 1713–1726. ISSN: 1939-9170. DOI: 10.1890/0012-9658(2002)083[1713:EOGPSD]2.0.CO;2. URL: http://onlinelibrary.wiley.com/doi/10.1890/0012-9658(2002)083[1713:EOGPSD]2.0.CO;2/abstract (visited on 12/16/2016).

Plantegenest, Manuel, Christophe Le May, and Frédéric Fabre (2007). "Landscape epidemiology of plant diseases". In: *Journal of the Royal Society Interface* 4.16, pp. 963–972. ISSN: 1742-5689. DOI: 10.1098/rsif.2007.1114. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2394555/.

Randolph, S. E. and A. D. M. Dobson (2012). "Pangloss revisited: a critique of the dilution effect and the biodiversity-buffers-disease paradigm". In: *Parasitology; Cambridge* 139.7, pp. 847–63. ISSN: 00311820. DOI: http://dx.doi.org.colorado.idm.oclc.org/10.1017/S0031182012000200. URL: http://search.proquest.com.colorado.idm.oclc.org/docview/1017536663/abstract/BC0121ED6FC04C59PQ/1.

Rizzo, David M., Matteo Garbelotto, and Everett M. Hansen (2005). "Phytophthora ramorum: Integrative Research and Management of an Emerging Pathogen in California and Oregon Forests". In: *Annual Review of Phytopathology* 43.1, pp. 309–335. DOI: 10.1146/annurev.phyto.42.040803.140418. URL: https://doi.org/10.1146/annurev.phyto.42.040803.140418 (visited on 12/09/2017).