

Epidemiological modeling of invasion in heterogeneous landscapes: spread of sudden oak death in California (1990–2030)

ROSS K. MEENTEMEYER,^{1,†} NIK J. CUNNIFFE,² ALEX R. COOK,³ JOAO A. N. FILIPE,² RICHARD D. HUNTER,¹
DAVID M. RIZZO,⁴ AND CHRISTOPHER A. GILLIGAN²

¹Department of Geography and Earth Sciences, University of North Carolina, Charlotte, 9201 University City Boulevard,
Charlotte, North Carolina 28223 USA

²Department of Plant Sciences, University of Cambridge, Downing Street, Cambridge, CB2 3EA United Kingdom

³Department of Statistics and Applied Probability, National University of Singapore, Singapore 117546

⁴Department of Plant Pathology, University of California, Shields Avenue, Davis, California 95616 USA

Abstract. The spread of emerging infectious diseases (EIDs) in natural environments poses substantial risks to biodiversity and ecosystem function. As EIDs and their impacts grow, landscape- to regional-scale models of disease dynamics are increasingly needed for quantitative prediction of epidemic outcomes and design of practicable strategies for control. Here we use spatio-temporal, stochastic epidemiological modeling in combination with realistic geographical modeling to predict the spread of the sudden oak death pathogen (*Phytophthora ramorum*) through heterogeneous host populations in wildland forests, subject to fluctuating weather conditions. The model considers three stochastic processes: (1) the production of inoculum at a given site; (2) the chance that inoculum is dispersed within and among sites; and (3) the probability of infection following transmission to susceptible host vegetation. We parameterized the model using Markov chain Monte Carlo (MCMC) estimation from snapshots of local- and regional-scale data on disease spread, taking account of landscape heterogeneity and the principal scales of spread. Our application of the model to Californian landscapes over a 40-year period (1990–2030), since the approximate time of pathogen introduction, revealed key parameters driving the spatial spread of disease and the magnitude of stochastic variability in epidemic outcomes. Results show that most disease spread occurs via local dispersal (<250 m) but infrequent long-distance dispersal events can substantially accelerate epidemic spread in regions with high host availability and suitable weather conditions. In the absence of extensive control, we predict a ten-fold increase in disease spread between 2010 and 2030 with most infection concentrated along the north coast between San Francisco and Oregon. Long-range dispersal of inoculum to susceptible host communities in the Sierra Nevada foothills and coastal southern California leads to little secondary infection due to lower host availability and less suitable weather conditions. However, a shift to wetter and milder conditions in future years would double the amount of disease spread in California through 2030. This research illustrates how stochastic epidemiological models can be applied to realistic geographies and used to increase predictive understanding of disease dynamics in large, heterogeneous regions.

Key words: computational biology; emerging infectious disease; GIS; landscape epidemiology; Markov chain Monte Carlo; *Phytophthora ramorum*; spatial heterogeneity; species distribution model.

Received 16 December 2010; **accepted** 11 January 2011; **published** 16 February 2011. Corresponding Editor: M. Wimberly.

Citation: Meentemeyer, R. K., N. J. Cunliffe, A. R. Cook, J. A. N. Filipe, R. D. Hunter, D. M. Rizzo, and C. A. Gilligan. 2011. Epidemiological modeling of invasion in heterogeneous landscapes: spread of sudden oak death in California (1990–2030). *Ecosphere* 2(2):art17. doi:10.1890/ES10-00192.1

Copyright: © 2011 Meentemeyer et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits restricted use, distribution, and reproduction in any medium, provided the original author and sources are credited.

† **E-mail:** rkmeente@uncc.edu

INTRODUCTION

Emerging infectious diseases (EIDs) continue to threaten the stability and persistence of plant and animal populations in natural and managed ecosystems (Anderson et al. 2004). While crop and livestock pathogens affect food security and the sustainability of agricultural production (e.g., Gilligan 2008), invasive pathogens in natural ecosystems pose substantial risks to biodiversity and ecosystem function (Daszak et al. 2000, Dobson and Foufopoulos 2001, Harvell et al. 2002, Anderson et al. 2004). As reports of EIDs and their impacts grow, we are increasingly faced with the challenge of predicting the spread of disease at landscape- to regional-scales in order to understand potential ecological consequences of invasion. Predictive models are also required to design control strategies that are capable of having impacts at appropriately large scales to manage or contain invasions (Holdenrieder et al. 2004, Ostfeld et al. 2005, Gilligan et al. 2007). However, our ability to predict whether or not an epidemic will arise in natural ecosystems often hinges on incomplete knowledge of complex ecological factors, such as the host range of the pathogen, spatio-temporal heterogeneity of host and environmental conditions, and the rates at which the pathogen can disperse through and among fragmented host populations (Holdenrieder et al. 2004). For example, some landscape features may function as inoculum reservoirs and dispersal corridors that foster spread, while others act as barriers to dispersal of infective propagules (Plantegenest et al. 2007). How landscape heterogeneity of host availability and environmental conditions affects the persistence and spread of disease is central to the emerging field of landscape epidemiology (Collinge et al. 2005, Condeso and Meentemeyer 2007, Ellis et al. 2010), and pertinent to prediction of epidemic outcomes over large heterogeneous regions.

To date, spatially-explicit models of disease spread have tended to use either a biogeographical, *pattern-based* approach (e.g., Vaclavik and Meentemeyer 2009) or an epidemiological, *process-based* approach (e.g., Gilligan 2008). The growing use of geographic information system (GIS) technologies in biogeographical models has provided insight into how real-world landscape heterogeneity affects spatial patterns of disease

risk, frequently leading to static maps of disease-risk geographies (Kelly et al. 2007, Meentemeyer et al. 2008a). However, the pattern-based nature of biogeographical models largely ignores key epidemiological processes that drive dynamics of disease, limiting the ability of this approach to examine epidemic progress over time and space. Epidemiological models, on the other hand, include the dynamics at a range of scales that drive disease spread, but commonly model spatial heterogeneity by resorting to simulated or artificial landscape data (e.g., Russell et al. 2005). This requires assumptions concerning the spatial scale, extent and heterogeneity of the system in question, which are often not fully justified. With notable exceptions largely confined to agriculture (e.g., Ferguson et al. 2001, Keeling et al. 2001), dynamic models of epidemic spread rarely incorporate landscape heterogeneity of host species populations. For plant pathogens, particularly in natural environments, there is also a need for epidemiological models to incorporate key abiotic drivers, such as the effects of weather variability (Davidson et al. 2005) or hydrological flow patterns (Kauffman and Jules 2006) on the dispersal and rate of infectious spread. One obvious restricting factor has been a paucity of high resolution, real-world landscape data. However, even when data are available or could be collected, modelers face the challenge of balancing epidemiological and landscape parameters to capture the dynamics of a particular host-pathogen system without (i) exceeding acceptable computational and data collection costs or (ii) over-fitting models that contain large numbers of free parameters and that are difficult to interpret.

In this paper, we describe an approach to model the geographical spread of the invasive pathogen *Phytophthora ramorum* through spatially heterogeneous host populations in Californian plant communities. Since it was first observed in the mid 1990s, this generalist plant pathogen—the causal agent of sudden oak death—has killed millions of oak (*Quercus* spp.) and tanoak (*Notholithocarpus densiflorus*) trees in coastal forests of the Western U.S. (Rizzo and Garbelotto 2003, Rizzo et al. 2005, Hansen et al. 2008; Fig. 1). The pathogen can infect dozens of native forest species across these heterogeneous coastal landscapes; individual hosts, however, may vary

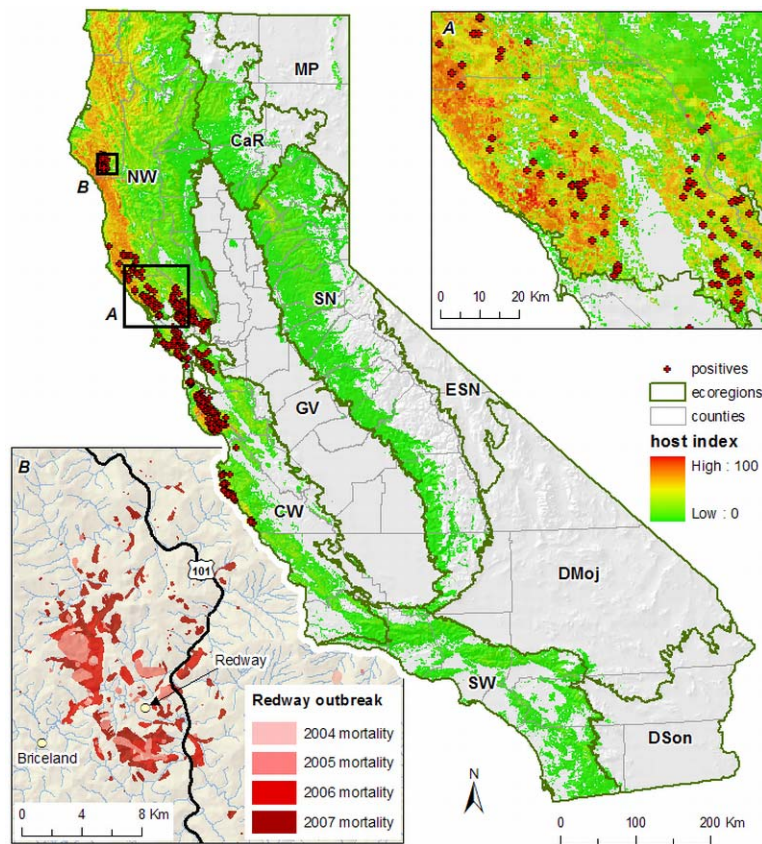


Fig. 1. The state of California colored according to the distribution of scaled (0–100) host index data, comprising joint measures of host abundance, weighted for relative susceptibility, within 250×250 m cells derived from data described in Meentemeyer et al. (2004). The main map also shows all known *P. ramorum* positive sites confirmed by the California Department of Food and Agriculture (CDFA), which were used to estimate long-distance dispersal (2001–2007). Inset A illustrates a magnified view of the CDFA data in Sonoma County and southern Mendocino County in the NW ecoregion. Inset B shows data on the spread of sudden oak death in the isolated outbreak in Humboldt County (Valachovic et al. 2008), which were used to estimate local-scale dispersal. Also illustrated are California's ten ecoregions (Hickman 1993); NW = Northwestern California, CW = Central Western California, GV = Great Central Valley, MP = Modoc Plateau, SW = Southwestern California, SN = Sierra Nevada, CaR = Cascade Ranges, ESN = Eastern Sierra Nevada, DMOj = Desert Mojave, DSon = Desert Sonoma).

dramatically in susceptibility and infectiousness (Rizzo et al. 2005, Cobb et al. 2010). This broad host range in Californian forests (Rizzo and Garbelotto 2003) along with favorable weather conditions and recent changes in host density (Meentemeyer et al. 2008b) are likely to have facilitated the rapid spread of this pathogen. The potential for further spread of sudden oak death has generated considerable concern as significant tree mortality is likely to change the structure of

plant, microbial, and aquatic communities, alter ecosystem function, and increase the risk of forest fire hazards (Rizzo and Garbelotto 2003, Rizzo et al. 2005). Moreover, the cost of regulatory monitoring, eradication, and containment is increasing each year, particularly for the nursery and garden industry (Osterbauer et al. 2004). Thus, knowing when and where the pathogen is likely to spread is essential for understanding its invasion dynamics and applying strategies to

control further spread.

We use a spatio-temporal, stochastic epidemiological model (Gilligan et al. 2007, Gilligan 2008), in combination with realistic geographical modeling to predict the spread of the pathogen through heterogeneous host populations in wildland forests subject to variable weather conditions. Our analysis addresses three objectives: (i) development of a parsimonious, epidemiological model for predicting the spread of an emerging forest disease over a large, and therefore heterogeneous geographic region while allowing efficient simulation of numerous replicate epidemics and analysis of model uncertainty; (ii) parameterization of a model from snapshots of incomplete data on host and pathogen distribution that take account of the principal scales of spread; and (iii) spatially-explicit prediction of likely epidemic outcomes through to 2030. The spatially-explicit predictions are used to identify high-risk locations in the absence of extensive disease control. They are also used to assess risks of invasion to critical ecoregions, such as forests of the North Coast and Sierra Nevada foothills, where susceptible hosts occur but the disease has not yet been recorded.

To model disease spread across wildlands of California, we estimate dispersal kernels for *P. ramorum* based upon regional- and local-scale data of disease distribution in forests, using Bayesian Markov chain Monte Carlo methods (Cook et al. 2008), with the effects of weather conditions on transmission parameters informed by published field and laboratory studies (Davidson et al. 2002, Davidson et al. 2003, Rizzo and Garbelotto 2003, Davidson et al. 2005). GIS spatial models of vegetation distribution and weather conditions (Meentemeyer et al. 2004) enable us to resolve California's heterogeneous landscapes into a detailed lattice of contiguous cells. Within each cell, distinction is made in the model between susceptibility and sporulation capacities of the location based on forest structure and variable weather conditions. In this respect our approach builds upon previous ecological niche-based models analyzed by Meentemeyer et al. (2004, 2008a) which we combine with dynamic epidemiological processes for inoculum build-up and transmission, subject to stochastic variation. Using this framework we model the spatial and temporal

dynamics of California's sudden oak death epidemic since its likely introduction and establishment in the late 1980s or early 1990s. We evaluate the accuracy of our models by comparing predictions of pathogen spread with an independent dataset of pathogen incidence in spatially distributed field plots in 2004–2006 (Meentemeyer et al. 2008a) to determine if we can successfully predict the patterns of spread since the introduction and establishment of the pathogen. We identify key parameters driving disease spread and assess the magnitude of stochastic variability in the predictions. Finally, we examine the degree to which local- versus non-local dispersal affects patterns of spread and we analyze how scenarios of future climate change would affect rates of disease spread in Californian forests.

METHODS

The model

We use a spatially-explicit, discrete-time, stochastic, SI model to represent the population dynamics of *P. ramorum* in forest landscapes, partitioning California into 250 m by 250 m cells containing multiple susceptible (S) and infected (I) host units (Fig. 1). The model incorporates the effects of five spatially heterogeneous factors on disease spread: forest community type, local weather conditions, heterogeneous host density, local transmission and intensification, and long-distance transmission of the pathogen. We chose the 250 m cell size to balance computational efficiency and spatial integrity of the vegetation data while matching the scale at which spatial heterogeneity of the host distribution affects disease establishment (Condeso and Meentemeyer 2007). The total number of *host units* in each grid cell (defined below) was determined from the composite abundance of *P. ramorum* host species, derived from detailed GIS maps of Californian vegetation weighted by susceptibility and capacity for inoculum production (Meentemeyer et al. 2004). The population dynamics of the pathogen depend upon local climatic conditions, host density and time of year. Dispersal is controlled by a kernel that describes the spatial spread. Stochasticity enters the model by influencing (1) production of inoculum at a given site, (2) the chance that inoculum is dispersed within

and between sites and (3) the probability of infection following transmission to susceptible host vegetation. To initiate the model, we introduced the pathogen at three susceptible forest locations in central coastal California (Fig. 1). We determined these introduction sites based on our best knowledge of invasion history, determined from genetically-identified founder populations and assessments of tree mortality (Rizzo et al. 2005, Mascheretti et al. 2008, Mascheretti et al. 2009, Prospero et al. 2009). The spread of the pathogen and disease is simulated over the 41 year period 1990 to 2030.

We assume that infection of site j during week t arises as a Poisson process with rate $\phi_{jt} \sum_i \psi_{ijt}$, where ψ_{ijt} is the rate of spread from an infected site i to site j during week t (note that the summation runs over all infected sites including j , meaning that an infected site can re-infect itself up to the local limit of susceptible host units). The contribution to the infectious pressure from site i to site j is modeled via:

$$\psi_{ijt} = \beta \left(\chi_t(f_i) m_{it} c_{it} I_{it} \right) \left(\chi_t(f_j) m_{jt} c_{jt} S_{jt} / N^{\max} \right) \times K(d_{ij}; \alpha_1, \alpha_2, \gamma) / d_{ij}, \quad (1)$$

where:

- $\chi_t(f_j)$ is a seasonal indicator variable, equal to 1 if cells of type f_j are able to infect and be infected at time t , and 0 otherwise;
- m_{jt} and c_{jt} are time- and space-dependent weather conditions of moisture and temperature, respectively;
- I_{it} and S_{it} are the numbers of infected and susceptible host units at time t in cells i and j respectively;
- $N^{\max}$ is the maximum number of host units in any site;
- $K(d_{ij}; \alpha_1, \alpha_2, \gamma)$ is a dispersal kernel for movement of inoculum over distance d_{ij} parameterized by scale parameters α_1 , governing the scale of short-distance spread; α_2 , governing the long-distance scale; and γ , governing the proportion of spore units that are locally (e.g., <1 km) dispersed; and
- β controls the overall rate at which spores are produced by infected host units within a given site per unit time.

Each of these components is explained more fully below and the state variables and parameters of the model are summarized in Table 1. Little is yet known about the long-term effects of sudden oak death on forest ecosystems. Maloney

et al. (2005) and Cobb et al. (2010) suggest that the abundant host tree bay laurel (a highly infectious host that experiences a non-lethal leaf blight) will gradually replace dead oak and tanoak trees and continue producing inoculum in affected forests indefinitely. For tractability, we assume constant host-composition in the landscape during the period of simulations. We hypothesize that this assumption will produce slightly conservative predictions for disease spread as within-cell density of bay laurel would increase somewhat over time. We also assume no geographical redistribution of the host trees in response to possible climate change in the future. To restrict these biases, we confine our predictions to roughly 20 years out to 2030.

Susceptible host vegetation

The spatial distribution and abundance of *P. ramorum* forest host species were mapped as a host competency index by Meentemeyer et al. (2004) from the CALVEG vegetation database (USDA Forest Service RSL 2003), supplemented by data from the California GAP Analysis project (Davis et al. 1995) in a few areas where CALVEG was incomplete. The estimated abundance (ha_k) and susceptibility (hs_k) of the main host species (indexed by species identity k) were calculated per cell (Meentemeyer et al. 2004); susceptibility combines host vulnerability to infection (Davidson et al. 2002, Garbelotto et al. 2003) with the ability of the pathogen to produce inoculum on susceptible hosts (Davidson et al. 2005, Davidson et al. 2008). Abundance and susceptibility are aggregated for each cell (i) over all host species to produce a localized host index $h_i = \sum_k (ha_k)_i (hs_k)_i$, which is rescaled according to the maximum value over the entire landscape and quantised to determine an integer number of host units in the range 0 to 100 ($=N^{\max}$). Including multiple host units per cell allows local intensification of infection within a cell. The observed seasonality of spore production in different vegetation types (Rizzo et al. 2005) is accounted for by a binary forest-type mask for each cell (f_i), which distinguishes cells dominated by redwood-tanoak (denoted R) from mixed-evergreen forests dominated by coast live oak and bay laurel (M). Following the start of the rainy season in California's Mediterranean climate, spore pro-

Table 1. Default and fitted parameter values for inoculum production, dispersal and infection processes used in Eq. 1.

Symbol	Meaning	Typical value or range
i	Index of a particular cell	—
t	Index of a particular week	—
N^{max}	Maximum number of host units in any cell	100
N_i	Number of host units in cell i	$[0, N^{max}]$
f_i	Forest type mask in cell i (R-Redwood-tanoak) (M- Mixed evergreen forest)	R or M
S_{it}	Number of susceptible host units in cell i at time t	$[0, N^{max}]$
I_{it}	Number of infected host units in cell i at time t	$[0, N^{max}]$
m_{it}	Moisture suitability index in cell i at time t	$[0, 1]$
c_{it}	Temperature suitability index in cell i at time t	$[0, 1]$
(χ_{fi})	Seasonality in infection due to forest type	0 or 1
$K(d_{ij})$	Dispersal kernel between cells i and j (depends on distance between them, d_{ij})	—
a_1	Scale parameter for short-range dispersal	20.57 m
a_2	Scale parameter for long-range dispersal	9.504 km
γ	Proportion of short range dispersal events	0.9947
β	Rate of inoculum produced (per host unit per week) that infect a host unit elsewhere	4.4 1/week
ψ_{ijt}	Infectious pressure on cell i from cell j at time t	—
ϕ_{jt}	Total infectious pressure on cell j at time t , where $\phi_{jt} = \sum_i \psi_{ijt}$	—
ρ	Odds ratio of presence/absence of infection as predicted by the model against observation	—

duction is typically initiated and peaks earlier in redwood-tanoak forests than in mixed-evergreen forests (Rizzo et al. 2005). This difference is reflected in the model by restricting infection in forests of type R to January–September, with corresponding restriction in forests of type M to February–September.

Weather conditions

Survival, sporulation and transmission of *P. ramorum* in forests are highly sensitive to fluctuations in temperature and moisture (Davidson et al. 2005, Fichtner et al. 2009). We introduce a local, time-specific, moisture suitability index, m_{it} , which increases linearly for each cell between 0 and 1 according to the number of days in week t with precipitation greater than 2.5 mm in cell i . A local, time-dependent, temperature suitability index, c_{it} , is also represented on a $[0, 1]$ scale, using replicated experimental data that quantified the mean production of *P. ramorum* zoospores under seven temperature treatments between 0 and 30°C (Davidson et al. 2005). We rescaled the mean of each temperature-dependent sporulation response relative to the maximal sporulation rate observed in the experiment to create an index of potential pathogen sporulation according to temperature. We then used regression to express temperature x as a suitability index y by fitting a cubic polynomial ($y = -0.066 + 0.056x - 0.0036(x - 15)^2 - 0.0003(x - 15)^3$) to these rescaled data ($n = 14$, $R^2 = 0.79$)

subject to the constraint $y \geq 0$, $\forall x$. This function was used to convert the weekly average temperature in cell i during week t to the cell's suitability index c_{it} .

We mapped the weather condition indices for each week at the 250m by 250m resolution based on spatially-interpolated estimates of daily precipitation and mean daily temperature (Hunter and Meentemeyer 2005). Using geostatistical kriging, these interpolation methods integrated daily weather data recorded at base stations with long-term average climate data from PRISM (Daly et al. 2001) to produce reliable daily weather estimates at the required resolution between 1990 and 2008. We also used these data to forecast future disease spread under different weather scenarios through to the year 2030 and to examine effects of weather variability on future epidemic outcomes. Specifically, we ranked each year (y) of weather conditions $\sum_{i,t \in y} m_{it}c_{it}$ based on favorability for inoculum production and infection (favorable = upper 50% and unfavorable = lower 50%). Three weather scenarios were simulated from 2009 through to 2030: (i) a *favorable weather scenario* where a sequence of weather years was randomly selected from the upper 50% ($m \times c$) conditions; (ii) an *unfavorable scenario* made up of randomly selected weather years from the lower 50%; (iii) a *random scenario* made up of randomly selected weather years between 1990–2008.

Estimation of transmission parameters

Estimation of scale parameters for long and short scale dispersal.—We identified two scales for dispersal kernels: one for short distance, typified by local dispersal of the pathogen in wind-driven rain and mist, the other for occasional long-distance dispersal effected by a range of mechanisms associated principally with anthropogenic activity such as movement of infected ornamentals and transport of infected organic material by hikers, vehicles or perhaps by birds (Rizzo et al. 2005, Cushman and Meentemeyer 2008). Note that we do not expect distance-parametric kernels to realistically represent inoculum introductions associated with movement in nursery trade because of the absence of spatial scale in this process and lack of conclusive data on nursery-to-wild transmission. While wind-driven rain is thought to be an important dispersal process at local scales (Rizzo et al. 2005), we necessarily assumed isotropic dispersal due to the complexity of including advection in a large-scale model and lack of data on local- to landscape-scale wind patterns.

Our subsequent estimations revealed typical ranges of the two types of dispersal to be ~0–1 km and ~1–100 km, respectively. Exploratory analysis of disease data across Californian forests indicated that both scales of dispersal should be considered in order to reproduce sufficient local intensification of disease while allowing for occasional long-distance spread (Mascheretti et al. 2008) (Fig. 1 and Appendix A: Fig A1). We estimated the scale parameters for both local and long-distance spread of *P. ramorum* in the wild using Markov chain Monte Carlo (MCMC) methods (e.g., Gilks et al. 1996 and references therein). We converted annual records of presence and absence of the disease in forests obtained via (i) aerial surveys (2004–2007) in Humboldt county (Fig. 1; Valachovic et al. 2008) to estimate local dispersal, and (ii) all known *P. ramorum* positive sites (2001–2007) confirmed by the California Department of Food and Agriculture (CDFA) to estimate long-distance dispersal, into annual sets of 250 m and 10 km square cells respectively (Fig. 1). The CDFA dataset documents locations of infection recorded by land managers, arborists, and scientists since 2000 and is actively maintained by the California Oak Mortality Task Force (COMTF) (Kelly and Tuxen

2003). At these scales, we are confident that significant amounts of infection will be detected fairly rapidly, and we therefore made the assumption that infection of a cell precedes observation of symptoms by up to one year (Rizzo et al. 2005), noting that each cell contains multiple host species, some of which have a short asymptomatic period.

We fitted a simplified form of the full model (Eq. 1), separately to each dataset. The model accounts for infectious host distribution but not the other heterogeneities (which are later re-incorporated), specifically:

$$\psi_{ijt}^{\text{inference}} = b_k I_{it} (S_{jt}/N^{\text{max}}) K_k(d_{ij}; \alpha_k)/d_{ij}, \quad (2)$$

where k is 1 for local and 2 for long-distance dispersal, and b_k is the amount of spores that are dispersed locally ($k = 1$) or globally ($k = 2$). The Humboldt data (Fig. 1) were used to estimate the local scale parameter α_1 under the assumption that only one long distance dispersal event reached the county. We used COMTF data (Fig. 1 and Appendix A: Fig A1) to estimate the long-distance scale parameter α_2 under the assumption that local dispersal does not contribute at the 10 km scale; the ratio $b_1/(b_1 + b_2)$ is equated to γ in the full model. The restriction of the dispersal analysis to California is addressed in the discussion.

Conditioning on the unobserved infection times, the likelihood function can be calculated directly, exploiting standard properties of Poisson processes. We therefore follow Gibson and Renshaw (1998) in augmenting the parameter vector to include these unobserved event times and then integrating over them using MCMC. This allows a posterior distribution for α_k and b_k to be sampled, whence the posterior mode can be found. We take flat, pseudo-objective prior distributions. For examples of this general approach and more details on the algorithm, see e.g., Cook et al. (2008).

A candidate set of functional forms encompassing exponential and power-law functions was considered for both components of the dispersal kernel. At both scales the best-fitting kernel was Cauchy, $K_k(d; \alpha_k) = (1 + (d/\alpha_k)^2)^{-1}$; this was strongly supported by the deviance information criterion (DIC) (Spiegelhalter et al. 2002) (DIC for local scale: Cauchy = 4851, Exponential = 5785, Gaussian = 7695; DIC for statewide scale:

Cauchy = 785, Exponential = 797, Gaussian = 812). The Cauchy dispersal kernel is a long-tailed distribution with power-law properties (Minogue 1989). The final form of the overall kernel was therefore

$$K(d; \alpha_1, \alpha_2, \gamma) = \gamma \left(1 + (d/\alpha_1)^2\right)^{-1} + (1 - \gamma) \left(1 + (d/\alpha_2)^2\right)^{-1}, \quad (3)$$

with $\alpha_1 = 20.57$ m (95% credible interval 7.19–49.50 m), $\alpha_2 = 9.50$ km (95% credible interval 6.19–52.55 km) and $\gamma = 0.9947$. Satisfactory convergence (at both spatial scales) was assessed by visual inspection of traceplots and by monitoring the acceptance rate of parameter proposals and the chain's autocorrelation. Additionally convergence was assessed quantitatively via the Geweke diagnostic of first and last tertiles (Geweke 1992).

Estimation of spore production rate.—The obvious estimate $\hat{\beta} = b_1 + b_2$ of the spore production parameter (β) that could be obtained via the MCMC approach would be biased by the absence of some of the heterogeneities, and would therefore be inappropriate for use in the full model. Instead we obtained the optimal value of β via a line search algorithm maximizing (over a range of values of β) the odds ratio (Fielding and Bell 1997), comparing model predictions to observed data on pathogen presence/absence (Meentemeyer et al. 2008a; Appendix A: Fig A1). Model predictions were scaled-up to 10×10 km cells to match the spatial resolution of the survey data, and probabilities of infection were categorized to presence/absence predictions to allow direct comparison with the survey. Full details are given in Appendix A. We limited ourselves to samples from 2003 to thus calibrate $\beta = 4.4 \text{ week}^{-1}$, retaining data from years 2004 to 2006 for validation of the best-fitting model.

Model implementation and computational efficiency

The model was routinely simulated ≥ 1000 times for a given scenario, such as weather conditions, in order to generate probability distributions for the spatial and temporal spread of infection resolved to 250×250 m cells throughout California. To overcome the computational burden in simulating large numbers of

replicated epidemics, we work with the discrete time approximation in the model described above, with time steps of *one week*. Processing time is substantially increased by using a particle-emission reformulation of the dispersal kernel: a Poisson number of spores is assumed to be emitted from the center of infected cells at rate B per week; these are sent in a random direction on $[0, 2\pi)$ by a distance distributed according to the dispersal kernel, and if the destination cell contains susceptible host units then they are challenged by the pathogen. Thus, computational burden scales with the number of infections rather than as a function of the total size of the landscape.

We further increase computing speed by considering “spore units” instead of spores themselves. A standard property of the Poisson process (Rényi 1964) is that if events arise at rate B and are each retained with probability p , the stream of retained events is a Poisson process with rate pB . This allows two savings: (i) we can work with spore units, created as a Poisson process at rate $\beta = pB$, rather than individual spores at rate B , where p is the probability that an individual spore infects a host unit it lands upon (note neither B nor p need be specified separately); (ii) we further scale the baseline spore unit production rate to the maximum expected number of successful infections the most infective host unit could create if all spores went to the most susceptible host unit; each spore unit is then retained according to the ratio of actual to baseline product of infectivity and susceptibility.

Model validation

We performed both qualitative and quantitative validations of the model. We first assessed qualitative performance by comparing risk maps of infection probability with the spatial pattern of known *P. ramorum* infections (Appendix A: Fig. A1). There was good visual correspondence between model results and both the presence-absence survey data from Meentemeyer et al. (2008a), and the COMTF dataset of *P. ramorum* positive locations (Kelly and Tuxen 2003). We then quantitatively assessed the goodness-of-fit of the model both at the statewide level, and independently for the two ecoregions (Northwestern and Central Western) containing the vast

majority of known *P. ramorum* infections. We followed a similar procedure to that used in estimating β , calculating the odds ratio (ρ) (Fielding and Bell 1997) that compares model results with the presence-absence data of Meentemeyer et al. (2008a), restricting ourselves to data from 2004–2006 (as data from 2003 were used in estimating β). We used Fisher's Exact test (Zar 2005) to examine the hypothesis $\rho = 1$; i.e., that there is no association between the infection status of samples in the survey and in the model. Detailed description of the validation procedure and associated statistical testing is given in Appendix A.

Risk predictions

The risk of infection in any cell was quantified by taking the number of infected host units in that cell over an ensemble of 1000 runs of the model from 1990–2030, and averaged relative to a hypothetical maximum ($I = 100$) corresponding to total infection in a cell with the maximum host availability. This form of averaging allows host availability to influence our measure of risk. Initial investigation revealed the level of replication was sufficient to give a reasonable approximation to the full range of epidemic outcomes, while adequately accounting for the dual sources of variability in the model (i.e., in both the model itself and in the forward projection of climatic conditions). The infected area was determined for each run by counting the number of cells with $I > 0$, with the distribution of these over the runs of the simulations examined to estimate the variability in this key indicator of spread. The infected area was examined for each ecoregion (Fig. 1) and for the different parameter and weather combinations (Table 1). This figure is likely an over-estimate of the actual area occupied by the pathogen, as a cell with only a fraction of the hosts infected is taken to contribute $2.5 \times 2.5 = 6.25$ ha to the area. However, in general, once a cell has become infected, saturation of that cell occurs rapidly (given locally favorable weather conditions within a couple of years).

Secondary infection risk

To examine and quantify the potential for local disease amplification and secondary spread after a hypothetical introduction of the pathogen in a

given region, we artificially seeded the model with a single infected host unit in 1990 at each of 1000 randomly chosen host-containing cells. For each initial location we independently ran the model 50 times, and averaged the number of infected cells after 5 years. This measure of local outbreak potential, an operational surrogate of the basic reproductive number (R_0) (Heffernan et al. 2005), is designed to quantify the potential for spread conditioned on an isolated introduction at any location. Using ordinary kriging (Cressie 1993), the 1000 sample locations were spatially interpolated over California to visualize geographical variation in the risk of secondary infection.

RESULTS

The MCMC estimation of dispersal parameters (α_1 , α_2 and γ) showed that over 99% of *Phytophthora ramorum* dispersal in forested landscapes is short range (Table 1), with approximately 93% of the short-range dispersal occurring within a 250 m grid cell and 95% of long-range dispersal occurring within 100 km of a parent source of inoculum (Fig. 2). The high correspondence between observed and simulated patterns of pathogen spread (Appendix A: Fig A1) was supported by statistical testing of the odds ratio (ρ): Statewide $\rho = 15.4$, Northwestern Ecoregion $\rho = 36.7$, Central Western Ecoregion $\rho = 7.6$; all p-values < 0.0001 (Appendix A).

Application of the model (Eq. 1) to spatially and temporally heterogeneous host availability and weather conditions predicted the progression of epidemic spread in Californian forests over the 40-year simulation period (1990–2030; Fig. 3: see online Appendix B for an animated version). Early stages of disease spread are concentrated around the three hypothesized introductions in the San Francisco Bay Area (Mascheretti et al. 2008). By 2020, the model predicts a rapid increase in the rate of epidemic spread and the stochastic variability in epidemic outcomes increases through time (Fig. 4). Stochastic variability largely depends on the chance timing of rare long-range dispersal events to the North Western (NW) ecoregion, which, once infected, experiences rapid pathogen spread due to large contiguous regions of host availability and suitable weather conditions (see online

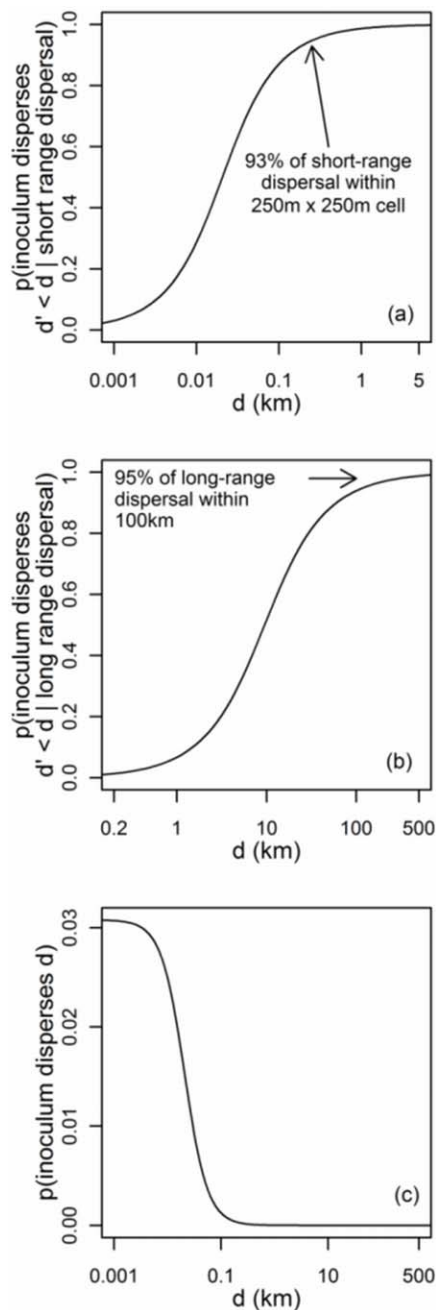


Fig. 2. Cumulative probability of (a) short-range dispersal (α_1) and (b) long-range dispersal (α_2). The proportion of spore units locally dispersed equals 99.4% (γ). The full mixed kernel is shown in (c).

Appendices C and E for animations of two realizations of the model, one where long-range spread establishes relatively early in the epidemic and one later; online Appendix D shows the

Appendix C simulation for a smaller extent focused on the southern portion of the NW ecoregion). As a result, the NW region is predicted to be the most impacted area in California by the year 2030.

The potential for the pathogen to spread and establish disease in forests of the Sierra Nevada (SN) and Southwest (SW) ecoregions has also generated considerable concern due to the presence of host species and moderately suitable weather conditions for sporulation and transmission (Meentemeyer et al. 2004, Rizzo et al. 2005). The probability of invasion in the SN and SW ecoregions increases from 30% and 5% in 2010 to nearly 100% and 30%, respectively, by 2030 (Fig. 4). However, once infected very little secondary infection occurs in these two regions, especially in the dryer, lower host density SW region. The rate of epidemic spread in the Central Western (CW) ecoregion, where the pathogen was originally introduced, slows between 2010 and 2030 as the supply of uninfected, susceptible forest vegetation decreases. In contrast, the NW ecoregion experiences an explosive rate of spread starting in 2015 through to 2030 characterized by considerable stochastic variability in the timing of long-distance introductions to this region.

Our risk assessment of the local potential for secondary spread shows considerable geographical variability throughout California in the amount of infection caused by a single infectious focus (Fig. 5). The highest risk of secondary cell-to-cell infection occurs in the northern portion of forests in the NW ecoregion. This region of Del Norte and northern Humboldt counties is currently not known to be infected by *P. ramorum*. Were the region to become infected, the model indicates a high risk of rapid disease spread due to highly suitable weather conditions ($\sum_{i,t} m_{it}c_{it}$) and high landscape connectivity and availability of non-infected hosts. Further inspection of the results (insets Fig. 5) shows that favorable weather or high connectivity of susceptible hosts alone are insufficient to promote a high risk of secondary cell to cell spread. For example, comparatively low risks occur in portions of the Sierra Nevada foothills where there may be relatively high amounts of susceptible host vegetation but unfavorable weather.

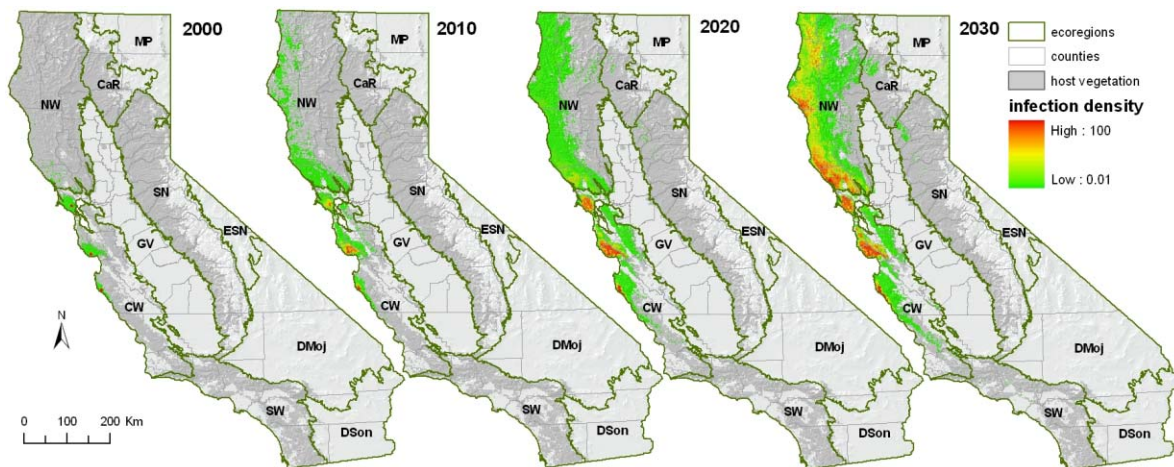


Fig. 3. Predicted spread of the sudden oak death epidemic through time (1990–2030) under an assumption of weather conditions like those experienced 1990–2008. The intensity of green to yellow to red shading is the mean infection density (I) of a 250 m grid cell over 1000 model simulations. Independent simulations were all initiated from the three sites of introduction in 1990 based on California Department of Food and Agriculture records (Kelly and Tuxen 2003). The sites are identifiable by the red regions in year 2000. Host vegetation with no infection is denoted in dark gray. Locations where host vegetation is absent are denoted in light gray.

Weather conditions exert a large influence on predicted rates of epidemic spread across California. A shift to more frequent, favorable weather conditions after 2008, characterized by a greater than average number of mild rainy days, would produce roughly five-fold more epidemic spread by 2030 compared with unfavorable conditions (Fig. 6a). In contrast, a random distribution of weather patterns typical of those experienced since the introduction of the pathogen (1990–2007) produces half as much epidemic spread by 2030 compared with favorable conditions but exhibits more stochastic variability in epidemic outcomes (Fig. 6a).

A reduction in the proportion of long-distance dispersal ($1 - \gamma$) via hypothetical management intervention after 2008 would decrease the amount of disease spread in California (Fig. 6b). However, approximately 8,000 km² of land area is predicted to be infected by the year 2030 even if long-range dispersal were to be nearly eliminated, presumably because the tail of the short-range kernel (α_2) is sufficiently long to contribute to landscape-scale spread (Fig. 2) and significant long-range dispersal had occurred before 2008 (Figs. 3 and 4).

DISCUSSION

Model predictions

Large-scale application of process-based epidemiological models to realistic landscape conditions is difficult to implement, but is an essential tool to predict disease dynamics and to improve our ability to manage invasive pathogens in heterogeneous ecosystems. We have analyzed and predicted the likely spread of the sudden oak death pathogen in California using a stochastic, spatio-temporal, epidemic model, parameterized by Bayesian MCMC methods, and supplemented by GIS data to account for the complicated effects of landscape heterogeneity on disease dynamics over large areas. The model predicts that there is a very high risk of continued disease spread in California during the next 20 years. Nearly all management of sudden oak death in California has occurred on an ad hoc basis at very local scales and is likely to have had limited impact on the overall spread of *P. ramorum*. In the absence of extensive disease management, by 2030 the pathogen is likely to have spread far beyond the current known extent of invasion. However, our model also predicts considerable spatial and temporal variation

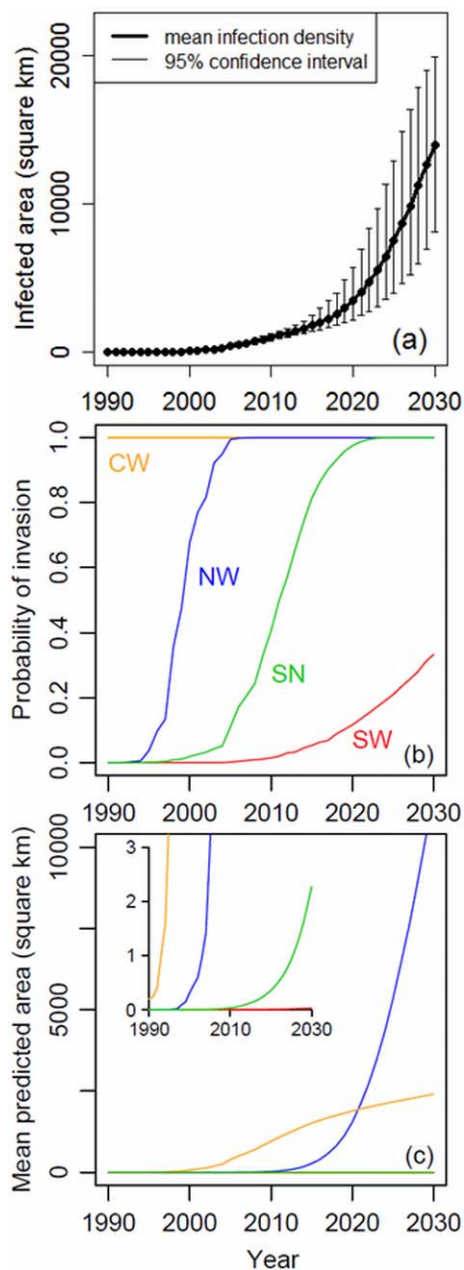


Fig. 4. Predicted progression of disease spread in California under an assumption of weather conditions for 2008–2030 typical of recent past (1990–2007) (means of 1000 independent simulations). (a) Rate of spread and stochastic variability increases through time (95% intervals are shown with bars). (b) Probability of invasion and (c) area affected by disease varies by ecoregion as the epidemic establishes. The inset in plot c is scaled to a small range (<3 km²) to contrast rates of spread between ecoregions during the early phase of invasion (<3 km²). CW = Central Western

across the Californian landscape in the potential establishment and subsequent spread of the pathogen. Knowledge of when and where the pathogen and disease is likely to spread is critical to setting disease management goals and priorities for managers and policy makers in an era of limited financial resources. In particular, our maps of epidemic spread (Fig. 3) identify regions on the epidemic front which could be subjected to intensive detection surveys, and our map of secondary infection risk (Fig. 5) could be used to prioritize locations for prevention management or control following new outbreaks.

The model predicts most spread of the pathogen to occur in coastal forests from the San Francisco Bay Area northward to the Oregon border (NW ecoregion; Figs. 3–5). Given the predicted rate of pathogen spread to this region, substantial tree mortality, particularly of tanoak, is likely to follow (Maloney et al. 2005, Cobb et al. 2010). While there is a high probability that sudden oak death spreads to the Sierra Nevada foothill region by 2020 (SN; Fig. 4), our results suggest that, in contrast to the north coast only small amounts of secondary spread will occur in this region (Figs. 3–5). The risk of continued southward spread through the central coast region from the heavily impacted Big Sur region is even smaller (Figs. 3–5), most likely due to low landscape connectivity of available non-infected hosts. In this area, susceptible forest types are inter-mixed with a wide range of non-susceptible plant communities such as chaparral and grassland.

Although long-distance dispersal is possible, our results suggest that the spread of *P. ramorum* is mostly local (<1 km; Fig. 2); this result is supported by both ecological (Condeso and Meentemeyer 2007, Ellis et al. 2010) and genetic (Mascheretti et al. 2008, Mascheretti et al. 2009)

← California, NW = Northwestern California, SN = Sierra Nevada, SW = Southwestern California ecoregions (geographical boundaries are shown in Fig. 1). Once infected the vast majority of spread occurs in the NW region. Long-distance dispersal events to the SN and SW regions occur infrequently with small amounts of secondary subsequent spread.

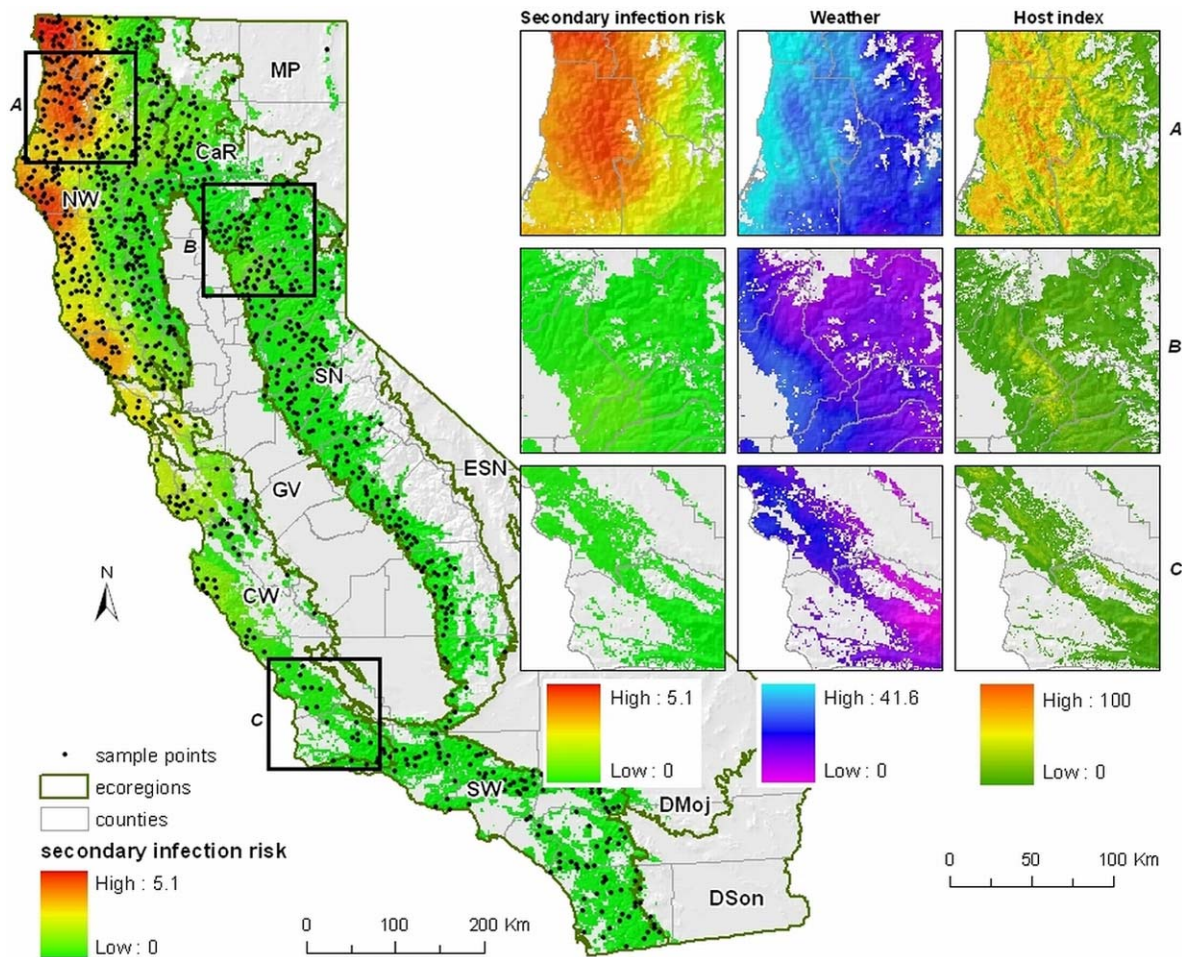


Fig. 5. Geographical variability of the number of secondary infections caused by a single infectious focus randomly located in California, based on 1000 randomly located sample points. We simulated 50 iterations of the model from 1990–1995 using each of the 1000 points as a single starting point for secondary infection and spread. The number of infected cells associated with each location was calculated and spatially interpolated across the pathogen's host range using ordinary kriging with the values here shown on a log scale. The 3×3 panel inset illustrates the range of host and weather interactions that lead to large and small amounts of secondary infection. Favorable weather or high connectivity of susceptible hosts alone is insufficient to promote a high risk of secondary cell to cell spread. For example, comparatively low risks occur in portions of the Sierra Nevada foothills where there may be relatively high amounts of susceptible host vegetation (center of inset) but unfavorable weather.

field studies. Dispersal of *P. ramorum* is therefore strongly dependent upon landscape connectivity of susceptible host vegetation and favorable weather conditions. For example, landscapes characterized by high host index and unfavorable weather conditions (e.g., Sierra Nevada foothills) do not lead to a large amount of secondary infection. Likewise, the combination

of favorable weather conditions and relatively low host connectivity in the Santa Barbara region in southern California also exhibits low potential for secondary infection (Fig. 5). This result corroborates other reports that spread of invasive plant pathogens may be limited both by the presence of suitable environmental conditions necessary for establishment (e.g.,

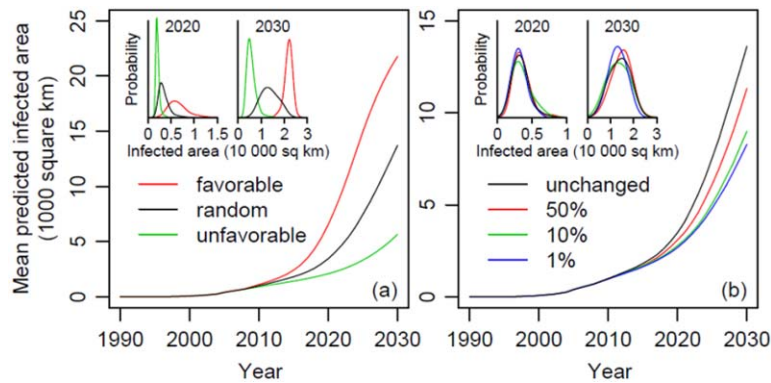


Fig. 6. Effects of (a) climate variation and (b) hypothetical reductions in long-distance dispersal ($1 - \gamma$; Table 1) on disease spread. (a) Favorable weather conditions starting in 2008 lead to a five-fold increase in infected area by the year 2030 compared with unfavorable conditions. A random distribution of weather conditions based upon weather for years 1990–2007 leads to almost half as much disease spread by 2030 as would be produced by favorable conditions. The inset shows that consistently favorable weather conditions also reduce variability in epidemic outcomes. (b) Reducing the proportion of long-distance dispersal after 2008 decreases disease spread. However, even if long-range dispersal were to be severely curtailed the model predicts approximately 8,000 km² of land area would be infected by 2030 due to the significance of short-range dispersal on spread rates. Surprisingly, variability by 2030 is largely unaffected by curtailing long distance spread after 2008; this counter-intuitive result is likely due to a significant number of isolated daughter foci having already been produced by long distance dispersal prior to the intervention.

Burdon 1993, Agrios 2005) and by long-distance dispersal (Aylor 2003). Finally, our results suggest that hypothetical climate variation toward more favorable weather conditions over the next 20 years could lead to almost twice as much disease spread by 2030 compared with a scenario of continued weather patterns similar to those experienced since introduction of the pathogen (1990–2007) (Fig. 6a). Future work to predict the effects of climate change on disease dynamics will require predictions from down-scaled global climate models (GCM). Although forest pathogens are substantially more sensitive to shifts in weather conditions than their host tree species (Woods et al. 2005), predictive epidemiological models should also begin considering the role that climate change may play in redistributing competent host species.

Model structure: host distribution in realistic landscapes

The complex epidemiology of invasive pathogens, like *P. ramorum*, with wide host ranges and long-range dispersal capabilities poses major challenges to developing a realistic model that

accurately predicts future spread through a heterogeneous landscape (Rizzo et al. 2005). The environmental conditions suitable for inoculum production and susceptibility differ substantially across *P. ramorum*'s diverse hosts (Meentemeyer et al. 2004, Davidson et al. 2005) and from one location to another within the same host population (Anacker et al. 2008). In addition, *P. ramorum* is thought to be dispersed by a variety of mechanisms including wind-blown rain splash of infected foliage (Davidson et al. 2005, Davidson et al. 2008, Fichtner et al. 2009), long-distance wind transport (Hansen et al. 2008, Mascheretti et al. 2008, Mascheretti et al. 2009), transport along streams (Davidson et al. 2005), and by human activity (Cushman and Meentemeyer 2008). Faced with considerable complexity and imprecise knowledge of epidemiological drivers, we have emphasized parsimony in our approach, seeking to construct a modeling framework that is simple enough but has sufficient geographical realism to capture the essential processes that drive and constrain spread. Our epidemiological model (Eqs. 1, 3) therefore incorporates only a few influences on

pathogen spread, including the type of forest community, weekly weather conditions, landscape structure of host density, and a dual-scale dispersal. While additional variables could be envisaged, such as variations in host resistance (e.g., Anacker et al. 2008), pathogen virulence and evolutionary lineage (e.g., Mascheretti et al. 2008, Goss et al. 2009, Mascheretti et al. 2009), or anisotropic spread via wind or streams (e.g., Kauffman and Jules 2006), this first model for the spread of *P. ramorum* has necessarily focused on the key dynamic factors that we could reliably model and map in a GIS.

Filipe et al. (2004) point out that the necessary landscape data are typically not available for applying epidemiological models to particular geographic regions and that calibrating models to such data can be technically challenging and computationally intensive if applied over large heterogeneous areas. Despite the relatively large cost of data collection and the use of a computationally-intensive procedure for model parameterization and forward prediction, our spatially-explicit mapping of host population densities and variable weather conditions allowed us to predict epidemic outcomes for particular locations and regions in California. Given the large amounts of abiotic and ecological heterogeneity that exist in California, ignoring this complexity would be likely to lead to significant underestimation of infection at locations with high host densities and overestimation at locations with low density or highly fragmented host populations. However, exactly how much landscape information is needed for developing regional-scale models of plant epidemics in natural ecosystems deserves further attention, as the cost of acquiring and processing such data can be high.

Previous models for the spread of plant pathogens have adopted simpler epidemiological approaches that do not take account of the natural and often heterogeneous structure of the host population. For example, Swinton and Gilligan (1996) adapted a mean-field approach to predict future outbreaks of Dutch elm disease in the UK over a long time scale of 50 years. Subsequent research on the deployment of control strategies, using hypovirulent strains of the pathogen, adopted a spatially-explicit stochastic approach involving dissection of the

landscape of susceptible trees in Southern England into a spatially-structured metapopulation (Swinton and Gilligan 1999). The sub-populations were large with arbitrarily defined mean-field dispersal within, and nearest-neighbor dispersal between sub-populations arranged on a lattice. Other authors, have also used metapopulation frameworks to analyze the spread of disease through natural plant populations, notably Burdon et al. (1999, 2002) for the spread of the rust *Melampsora lini* through populations of *Linum marginale* and Antonovics et al. (1997) on the spread of anther smut in populations of *Silene* spp. (also see Brooks et al. 2008, Soubeyrand et al. 2009). Our approach has close analogies with the so-called spatially-structured metapopulation models (Ovaskainen and Hanski 2001), with our cells (250×250 m) comprising sub-populations and transmission occurring within and between cells. Our use of two continuous dispersal kernels is a parsimonious formulation for capturing multi-scale disease spread within a metapopulation model. More importantly, our current approach combining GIS modeling with a spatially-explicit, epidemic model incorporates a realistic landscape that captures the local as well as the large scale heterogeneities in the distribution of the host and environmental drivers, in this case variable weather conditions.

Model structure: incorporating weather

Our predictions suggest that, along with host availability, local weather conditions play a major role in the spread of *P. ramorum*. In our model (Eq. 1), weather conditions affect two key epidemiological processes that vary in space and time: the probability of inoculum production at an infected site *i* and infection at a deposition site *j*. The distinction between sporulation and susceptibility in space is important for regional-scale models because, as is the case in California, weather conditions in one location may be very different just a few kilometers away. Therefore, for any weekly time step a large amount of inoculum dispersed from a favorable site could lead to little or no infection at another location and vice versa. Given the amount of heterogeneity in California, many dispersal events that escape a 250 m cell never lead to significant secondary infection. This is consistent with the

observational study of Condeso and Meentemeyer (2007) that landscape heterogeneity plays a major role in the establishment of *P. ramorum* through variation in host availability and microclimate conditions. Incorporation of weather conditions at high temporal resolution is relatively uncommon in plant epidemic models in agriculture and almost non-existent in natural ecosystems. Most modeling studies that include a meteorological component focus on annual or seasonally-averaged climatic variables, which may provide invaluable geographic information on climate suitability but cannot be used as a dynamic driver of pathogen spread in a process-based epidemiological model. We chose to implement our model on a weekly time step to balance the need to model dynamic epidemiological processes for the build-up and transmission of inoculum with the need for reasonable computational efficiency.

Model structure: incorporating stochasticity

Stochasticity enters our model at three distinct stages: (1) production of inoculum; (2) probability that inoculum is dispersed to a given location; and (3) chance of infection upon reaching a susceptible host. While the spread of *P. ramorum* appears to be predictable, our results indicate that stochastic processes play an important role in epidemic outcomes. In particular, the rate of disease spread and its stochastic variability increases through time, especially following the explosive growth in infection and disease that is predicted to occur around 2016.

Our simulations are consistent with the history of an unlikely long-range dispersal event to the north coast. In 2002, the California Department of Food and Agriculture confirmed a small, isolated infection in a redwood-tanoak forest near the town of Redway in Humboldt County (Fig. 1). Within only a few years the initial infection had spread substantially (>10 km) in response to high host densities and several seasons of favorable weather. Spread of the pathogen, however, slowed in 2007 and 2008 following three years of drought in northern California (Valachovic et al. 2008). Nevertheless, our model predictions indicate that unlikely outbreaks like these, once established, can rapidly accelerate the spread of sudden oak death in California (see Appendices C, D, and E

for online animations of two realizations of the model). The development, implementation and evaluation of disease management strategies across large landscapes requires an understanding of the potential magnitude of stochasticity in epidemic outcomes that is associated with chance long-distance dispersal and year-to-year variability in weather.

Pathogen dispersal

Although a hypothetical intervention starting in 2008 reducing the proportion of long-distance dispersal, $1 - \gamma$, slows the spread of disease (Fig. 6), the model predicts that almost 8,000 km² of land area (based on 250 m cells) would become infected by 2030 even if long-range dispersal were almost eliminated—presumably due to the significant contribution of shorter-range dispersal mechanisms, such as wind-driven rain (Davidson et al. 2005). However, we also found that variability in the amount of infected area by 2030 is generally unaffected by a reduction in long distance spread after 2008. This counter-intuitive result possibly reflects a substantial number of isolated daughter foci that established by long distance dispersal prior to 2008.

Spatio-temporal data on pathogen distribution are critical to the development of epidemiological models, but are often unavailable or imperfect for infectious diseases in natural ecosystems. Using successive but incomplete snapshots for disease occurrence, our MCMC methods allowed us to estimate parameters for the principal scales of pathogen spread, taking into account ecological heterogeneity and uncertainty. We estimated three parameters of the epidemiological model (Eqs. 1, 3): these comprised two scale parameters (α_1 governing the scale of short-distance spread and α_2 governing the long-distance scale) for the movement of inoculum over distance d_{ij} , and γ for control of the fraction of spore units that are locally dispersed per infected cell. Preliminary analyses showed that a model with a single scale of dispersal failed to reproduce the known characteristics of spread. For example, a dispersal kernel based on the field data for disease spread at the state-wide extent underestimated shorter-range dispersal (<1 km) whereas the kernel based on local- to landscape-scale spread underestimated long-range dispersal events that

may be relatively uncommon but critical to epidemic outcomes (Fig. 6). Where the complexity of human movement is critical, such as in the UK, models of sudden oak death spread may additionally need to account for human activities in a trade network (e.g., Harwood et al. 2009).

We did not include data on the isolated forest outbreak in Curry County, southwestern Oregon, originally reported in 2001 (Hansen et al. 2008, Vaclavik et al. 2010), in the estimation of the large-scale dispersal kernel principally because genetic and observational data (Prospero et al. 2007, Mascheretti et al. 2009) indicate that this may be a unique event that does not inform our understanding of natural spread in California. The origin of the inoculum that initiated the Curry County forest site is still unknown at this time. Recent genetic data strongly suggest that potential long distance movement (i.e., greater than 100 km) of *P. ramorum* from a forest in California to a forest in Oregon would most likely not be a natural dispersal event, but rather an anthropogenic event such as movement of nursery material (as discussed in Mascheretti et al. 2008, Prospero et al. 2009). The Oregon outbreak has been kept under tight control via an eradication program with no genetic evidence that inoculum from Oregon has contributed to the California epidemic (Hansen et al. 2008, Kanaskie et al. 2009, Prospero et al. 2009, Vaclavik et al. 2010). Therefore, modeling ongoing dispersal from Oregon to California was also not necessary at this time.

Concluding remarks

Globalization and widespread land use changes continue to accelerate the introduction of exotic pathogens to natural ecosystems, resulting in worldwide declines in biodiversity and habitat quality (Vitousek et al. 1996, Foley et al. 2005). Consequently, landscape- to regional-scale models of plant epidemics are increasingly required to predict large-scale impacts of disease and to assess practicable options for control. Our ability to predict epidemic outcomes in natural ecosystems, however, is frequently impaired by incomplete knowledge of the pathogen distribution and landscape heterogeneity, and in turn, the rate at which disease can spread through and among fragmented host populations. Our analysis ad-

dresses two related challenges. First is to develop a parsimonious, epidemiological model to predict the spread of disease over a large, heterogeneous region while allowing efficient simulation of numerous replicate epidemics and analysis of stochasticity. Second is to parameterize a model from snapshots of incomplete data on host and pathogen distribution, taking account of heterogeneity of environmental conditions and the principal scales of spread. The integration of a spatially-explicit epidemic model with GIS modeling allowed plausible predictions of the sudden oak death epidemic from 1990 through to 2030. The model's generality makes it readily adaptable for parameterization in other regions that have been impacted by *Phytophthora ramorum*, such as Oregon and the United Kingdom (Rizzo et al. 2005). Future work will use our model to examine the efficacy of control strategies that are being proposed and to determine economic impacts of this destructive biological invasion. With increasing numbers of emerging infectious diseases and limited resources for management, application of dynamic epidemiological models to realistic landscape conditions will not only increase our understanding of natural disease dynamics, but also increase our ability to devise and implement effective control strategies that consider the spatial scale and cost of a control strategy in the context of the scale of the epidemic.

ACKNOWLEDGMENTS

The authors gratefully acknowledge financial support from the National Science Foundation (EF-0622677) and the USDA Forest Service (Pacific Southwest Research Station). CAG was also supported by a BBSRC Professorial Fellowship and ARC by a BBSRC research grant. We thank Richard Cobb and Bethan Purse for valuable comments on our methods and interpretation of data and we thank Tomáš Václavík, Markéta Václavíková, Douglas Shoemaker and Matt Castle for vital assistance with geospatial modeling and preparation of figures.

LITERATURE CITED

- Agrios, G. 2005. Plant Pathology. Academic Press, New York, USA.
- Anacker, B. L., N. E. Rank, D. Huberli, M. Garbelotto, S. Gordon, T. Harnik, R. Whitkus, and R. Meentemeyer. 2008. Susceptibility to *Phytophthora ramorum*

- in a key infectious host: landscape variation in host genotype, host phenotype, and environmental factors. *New Phytologist* 177:756–766.
- Anderson, P. K., A. A. Cunningham, N. G. Patel, F. J. Morales, P. R. Epstein, and P. Daszak. 2004. Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. *Trends in Ecology and Evolution* 19:535–544.
- Antonovics, J., P. H. Thrall, and A. M. Jarosz. 1997. Genetics and the ecology of species interactions: the *Silene Ustilago* system. Pages 158–180 in D. Tilman and P. Kareiva, editors. *Spatial Ecology: The Role of Space in Population Dynamics and Interspecific Interactions*. Princeton University Press, Princeton, New Jersey, USA.
- Aylor, D. E. 2003. Spread of plant disease on a continental scale: Role of aerial dispersal of pathogens. *Ecology* 84:1989–1997.
- Brooks, C. P., J. Antonovics, and T. H. Keitt. 2008. Metapopulation dynamics and persistence are explained by temporal heterogeneity in spatial network structure. *American Naturalist* 149–159.
- Burdon, J. J. 1993. The Structure of Pathogen Populations in Natural Plant-Communities. *Annual Review of Phytopathology* 31:305–323.
- Burdon, J. J., P. H. Thrall, and A. H. D. Brown. 1999. Resistance and virulence structure in two *Linum marginale*–*Melampsora lini* host-pathogen metapopulations with different mating systems. *Evolution* 53:704–716.
- Burdon, J. J., P. H. Thrall, and G. J. Lawrence. 2002. Coevolutionary patterns in the *Linum marginale*–*Melampsora lini* association at a continental scale. *Canadian Journal of Botany-Revue Canadienne De Botanique* 80:288–296.
- Cobb, R. C., R. K. Meentemeyer, and D. M. Rizzo. 2010. Apparent competition in canopy trees determined by pathogen transmission rather than susceptibility. *Ecology* 91:327–333.
- Collinge, S. K., W. C. Johnson, C. Ray, R. Matchett, J. Grensten, J. F. Cully, K. L. Gage, M. Y. Kosoy, J. E. Loye, and A. P. Martin. 2005. Landscape structure and plague occurrence in black-tailed prairie dogs on grasslands of the western USA. *Landscape Ecology* 20:941–955.
- Condeso, T. E., and R. K. Meentemeyer. 2007. Effects of landscape heterogeneity on the emerging forest disease Sudden Oak Death. *Journal of Ecology* 95:364–375.
- Cook, A. R., G. J. Gibson, T. R. Gottwald, and C. A. Gilligan. 2008. Constructing the effect of alternative intervention strategies on historic epidemics. *Journal of the Royal Society Interface* 5:1203–1213.
- Cressie, N. A. C. 1993. *Statistics for spatial data*. Wiley-Interscience.
- Cushman, J. H., and R. K. Meentemeyer. 2008. Multi-scale patterns of human activity and the incidence of an exotic forest pathogen. *Journal of Ecology* 96:766–776.
- Daly, C., G. H. Taylor, W. P. Gibson, T. W. Parzybok, G. L. Johnson, and P. Pasteris. 2001. High-quality spatial climate data sets for the United States and beyond. *Transactions of the American Society of Agricultural Engineers* 43:1957–1962.
- Daszak, P., A. A. Cunningham, and A. D. Hyatt. 2000. Emerging infectious diseases of wildlife—Threats to biodiversity and human health. *Science* 287:443–449.
- Davidson, J. M., H. A. Patterson, and D. M. Rizzo. 2008. Sources of inoculum for *Phytophthora ramorum* in a redwood forest. *Phytopathology* 98:860–866.
- Davidson, J. M., D. M. Rizzo, and M. Garbelotto. 2002. *Phytophthora ramorum* and Sudden Oak Death in California: II. Pathogen transmission and survival. USDA Forest Service, Albany, California, USA.
- Davidson, J. M., S. Werres, M. Garbelotto, E. M. Hansen, and D. M. Rizzo. 2003. Sudden oak death and associated diseases caused by *Phytophthora ramorum*. *Plant Health Progress* [doi: 10.1094/PHP-2003-0707-01-DG]
- Davidson, J. M., A. C. Wickland, H. A. Patterson, K. R. Falk, and D. M. Rizzo. 2005. Transmission of *Phytophthora ramorum* in mixed-evergreen forest in California. *Phytopathology* 95:587–596.
- Davis, F. W., P. A. Stine, D. M. Stoms, M. I. Borchert, and A. D. Hollander. 1995. GAP analysis of the actual vegetation of California 1. The southwestern region. *Madrono* 42:40–78.
- Dobson, A., and J. Foufopoulos. 2001. Emerging infectious pathogens of wildlife. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* 356:1001–1012.
- Ellis, A. M., T. Vaclavik, and R. K. Meentemeyer. 2010. When is connectivity important? A case study of the spatial pattern of sudden oak death. *Oikos* 119:485–493.
- Ferguson, N. M., C. A. Donnelly, and R. M. Anderson. 2001. The foot-and-mouth epidemic in Great Britain: Pattern of spread and impact of interventions. *Science* 292:1155–1160.
- Fichtner, E. J., S. C. Lynch, and D. M. Rizzo. 2009. Survival, Dispersal, and Potential Soil-Mediated Suppression of *Phytophthora ramorum* in a California Redwood-Tanoak Forest. *Phytopathology* 99:608–619.
- Fielding, A. H., and J. F. Bell. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* 24:38–49.
- Filipe, J. A. N., W. Otten, G. J. Gibson, and C. A. Gilligan. 2004. Inferring the dynamics of a spatial epidemic from time-series data. *Bulletin of Math-*

- ematical Biology 66:373–391.
- Foley, J. A., R. DeFries, G. P. Asner, C. Barford, G. Bonan, S. R. Carpenter, F. S. Chapin, M. T. Coe, G. C. Daily, H. K. Gibbs, J. H. Helkowski, T. Holloway, E. A. Howard, C. J. Kucharik, C. Monfreda, J. A. Patz, I. C. Prentice, N. Ramankutty, and P. K. Snyder. 2005. Global consequences of land use. *Science* 309:570–574.
- Garbelotto, M., J. M. Davidson, and K. Ivors. 2003. Non-oak native plants are the main hosts for the sudden oak death pathogen in California. *California Agriculture* 57:18–23.
- Geweke, J. 1992. Evaluating the accuracy of sampling-based approaches to calculating posterior moments. in J. B. M. Bernardo, A. P. Dawid, and A. F. M. Smith, editors. *Bayesian Statistics 4*. Clarendon Press, Oxford.
- Gibson, G. J., and E. Renshaw. 1998. Estimating parameters in stochastic compartmental models using Markov chain methods. *Ima Journal of Mathematics Applied in Medicine and Biology* 15:19–40.
- Gilks, W. R., S. Richardson, and D. J. Spiegelhalter. 1996. *Markov Chain Monte Carlo in Practice: Interdisciplinary Statistics*. Chapman and Hall, London.
- Gilligan, C. A. 2008. Sustainable agriculture and plant diseases: an epidemiological perspective. *Philosophical Transactions of the Royal Society B-Biological Sciences* 363:741–759.
- Gilligan, C. A., J. E. Truscott, and A. J. Stacey. 2007. Impact of scale on the effectiveness of disease control strategies for epidemics with cryptic infection in a dynamical landscape: an example for a crop disease. *Journal of the Royal Society Interface* 4:925–934.
- Goss, E. M., I. Carbone, and N. J. Grunwald. 2009. Ancient isolation and independent evolution of the three clonal lineages of the exotic sudden oak death pathogen *Phytophthora ramorum*. *Molecular Ecology* 18:1161–1174.
- Hansen, E. M., A. Kanaskie, S. Prospero, M. McWilliams, E. M. Goheen, N. Osterbauer, P. Reeser, and W. Sutton. 2008. Epidemiology of *Phytophthora ramorum* in Oregon tanoak forests. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere* 38:1133–1143.
- Harvell, C. D., C. E. Mitchell, J. R. Ward, S. Altizer, A. P. Dobson, R. S. Ostfeld, and M. D. Samuel. 2002. Ecology—Climate warming and disease risks for terrestrial and marine biota. *Science* 296:2158–2162.
- Harwood, T. D., X. Xu, M. Pautasso, M. J. Jeger, and M. W. Shaw. 2009. Epidemiological risk assessment using linked network and grid based modelling: *Phytophthora ramorum* and *Phytophthora kernoviae* in the UK. *Ecological Modelling* 220:3353–3361.
- Heffernan, J. M., R. J. Smith, and L. M. Wahl. 2005. Perspectives on the basic reproductive ratio. *Journal of the Royal Society Interface* 2:281–293.
- Hickman, J. C., editor. 1993. *The Jepson Manual: higher plants of California*. University of California Press, Berkeley, California, USA.
- Holdenrieder, O., M. Pautasso, P. J. Weisberg, and D. Lonsdale. 2004. Tree diseases and landscape processes: the challenge of landscape pathology. *Trends in Ecology and Evolution* 19:446–452.
- Hunter, R. D., and R. K. Meentemeyer. 2005. Climatologically aided mapping of daily precipitation and temperature. *Journal of Applied Meteorology* 44:1501–1510.
- Jimenez-Valverde, A., and J. M. Lobo. 2007. Threshold criteria for conversion of probability of species presence to either-or presence-absence. *Acta Oecologica-International Journal of Ecology* 31:361–369.
- Kanaskie, A., E. Goheen, E. Hansen, N. Osterbauer, M. McWilliams, R. Schultz, S. Savona, W. Sutton, and P. Reeser. 2009. Early detection and eradication of *Phytophthora ramorum* (sudden oak death) in Oregon forests. *Phytopathology* 99:S61–S61.
- Kauffman, M. J., and E. S. Jules. 2006. Heterogeneity shapes invasion: Host size and environment influence susceptibility to a nonnative pathogen. *Ecological Applications* 16:166–175.
- Keeling, M. J., M. E. J. Woolhouse, D. J. Shaw, L. Matthews, M. Chase-Topping, D. T. Haydon, S. J. Cornell, J. Kappey, J. Wilesmith, and B. T. Grenfell. 2001. Dynamics of the 2001 UK foot and mouth epidemic: Stochastic dispersal in a heterogeneous landscape. *Science* 294:813–817.
- Kelly, M., Q. Guo, D. Liu, and D. Shaari. 2007. Modeling the risk for a new invasive forest disease in the United States: An evaluation of five environmental niche models. *Computers Environment and Urban Systems* 31:689–710.
- Kelly, M., and K. Tuxen. 2003. WebGIS for monitoring “sudden oak death” in California. *Computers, Environment and Urban Systems* 27.
- Maloney, P. E., S. C. Lynch, S. F. Kane, C. E. Jensen, and D. M. Rizzo. 2005. Establishment of an emerging generalist pathogen in redwood forest communities. *Journal of Ecology* 93:899–905.
- Mascheretti, S., P. J. P. Croucher, M. Kozanitas, L. Baker, and M. Garbelotto. 2009. Genetic epidemiology of the Sudden Oak Death pathogen *Phytophthora ramorum* in California. *Molecular Ecology* 18:4577–4590.
- Mascheretti, S., P. J. P. Croucher, A. Vettraino, S. Prospero, and M. Garbelotto. 2008. Reconstruction of the Sudden Oak Death epidemic in California through microsatellite analysis of the pathogen *Phytophthora ramorum*. *Molecular Ecology* 17:2755–2768.
- Meentemeyer, R. K., B. L. Anacker, W. Mark, and D. M.

- Rizzo. 2008a. Early detection of emerging forest disease using dispersal estimation and ecological niche modeling. *Ecological Applications* 18:377–390.
- Meentemeyer, R. K., N. E. Rank, B. L. Anacker, D. M. Rizzo, and J. H. Cushman. 2008b. Influence of land-cover change on the spread of an invasive forest pathogen. *Ecological Applications* 18:159–171.
- Meentemeyer, R. K., D. Rizzo, W. Mark, and E. Lotz. 2004. Mapping the risk of establishment and spread of Sudden Oak Death in California. *Forest Ecology and Management* 200:195–214.
- Minogue, K. P. 1989. Diffusion and Spatial Probability Models for Disease Spread. Pages 127–143 in M. J. Jeger, editor. *Spatial Components of Plant Disease Epidemics*. Prentice-Hal, New, Jersey, USA.
- Osterbauer, N. K., J. A. Griesbach, and J. Hedberg. 2004. Surveying for and eradicating *Phytophthora ramorum* in agricultural commodities. American Phytopathological Society, St. Paul, Minnesota, USA.
- Ostfeld, R. S., G. E. Glass, and F. Keesing. 2005. Spatial epidemiology: an emerging (or re-emerging) discipline. *Trends in Ecology and Evolution* 20:328–336.
- Ovaskainen, O., and I. Hanski. 2001. Spatially structured metapopulation models: Global and local assessment of metapopulation capacity. *Theoretical Population Biology* 60:281–302.
- Pearce, J., and S. Ferrier. 2000. Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling* 133:225–245.
- Plantegenest, M., C. Le May, and F. Fabre. 2007. Landscape epidemiology of plant diseases. *Journal of the Royal Society Interface* 4:963–972.
- Prospero, S., N. J. Grunwald, L. M. Winton, and E. M. Hansen. 2009. Migration patterns of the emerging plant pathogen *Phytophthora ramorum* on the west coast of the United States of America. *Phytopathology* 99:739–749.
- Prospero, S., E. M. Hansen, N. J. Grunwald, and L. M. Winton. 2007. Population dynamics of the sudden oak death pathogen *Phytophthora ramorum* in Oregon from 2001 to 2004. *Molecular Ecology* 16:2958–2973.
- Rényi, A. 1964. Sur les espaces des probability conditionnelles (On conditional probability spaces). *Annales de l'Institut Poincaré N Série, Section B* 1:3–21.
- Rizzo, D. M., and M. Garbelotto. 2003. Sudden oak death: endangering California and Oregon forest ecosystems. *Frontiers in Ecology and the Environment* 1:197–204.
- Rizzo, D. M., M. Garbelotto, and E. A. Hansen. 2005. *Phytophthora ramorum*: Integrative research and management of an emerging pathogen in California and Oregon forests. *Annual Review of Phytopathology* 43:309–335.
- Russell, C. A., D. L. Smith, J. E. Childs, and L. A. Real. 2005. Predictive spatial dynamics and strategic planning for raccoon rabies emergence in Ohio. *Plos Biology* 3:382–388.
- Soubeyrand, S., A. L. Laine, I. Hanski, and A. Penttinen. 2009. Spatiotemporal Structure of Host-Pathogen Interactions in a Metapopulation. *American Naturalist* 174:308–320.
- Spiegelhalter, D. J., N. G. Best, B. R. Carlin, and A. van der Linde. 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society Series B-Statistical Methodology* 64:583–616.
- Swinton, J., and C. A. Gilligan. 1996. Dutch elm disease and the future of the elm in the UK: A quantitative analysis. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* 351:605–615.
- Swinton, J., and C. A. Gilligan. 1999. Selecting hyperparasites for biocontrol of Dutch elm disease. *Proceedings of the Royal Society of London Series B-Biological Sciences* 266:437–445.
- USDA Forest Service RSL. 2003. CALVEG Vegetation Mapping Program. 1920 20th Street, Sacramento, CA. (<http://www.fs.fed.us/r5/rsl/projects/mapping/>)
- Vaclavik, T., A. Kanaskie, E. M. Hansen, J. L. Ohmann, and R. K. Meentemeyer. 2010. Predicting potential and actual distribution of sudden oak death in Oregon: Prioritizing landscape contexts for early detection and eradication of disease outbreaks. *Forest Ecology and Management* 260:1026–1035.
- Vaclavik, T., and R. K. Meentemeyer. 2009. Invasive species distribution modeling (iSDM): Are absence data and dispersal constraints needed to predict actual distributions? *Ecological Modelling* 220: 3248–3258.
- Valachovic, Y., C. Lee, J. Marshall, and H. Scanlon. 2008. Wildland management of *Phytophthora ramorum* in northern California forests. USDA Forest Service, Albany, California, USA.
- Vitousek, P. M., C. M. DAntonio, L. L. Loope, and R. Westbrooks. 1996. Biological invasions as global environmental change. *American Scientist* 84:468–478.
- Woods, A., K. D. Coates, and A. Hamann. 2005. Is an unprecedented dothistroma needle blight epidemic related to climate change? *Bioscience* 55:761–769.
- Zar, J. H. 2005. *Biostatistical analysis*. 4th edition. Pearson Education, Inc., New Delhi.

APPENDIX A

Model Rate Calibration and Validation

Model validation and part of the model calibration (of spore production rate β) was based on comparing model predictions with the presence/absence data from Meentemeyer et al. (2008a) (Fig. A1). The 2003 survey data were used to calibrate the model (i.e., to estimate the optimal value of β), with data from the 2004–2006 surveys retained for statistical validation and additional visualizations of model performance.

Calibration of Rate β

The optimal value of β was found by a line search algorithm, maximizing the average odds ratio ($\bar{p}(\beta)$) over a large ensemble of model simulations, 1000 for each value of β . To match the spatial scale of model predictions to that of the survey data, we overlaid a grid of 10×10 km cells over the spatial domain, and projected both the model results and the survey data onto this grid. The 2003 survey consisted of a number of

visits, N_+ of which had *P. ramorum* present (“survey positives”), and N_- of which had *P. ramorum* absent (“survey negatives”), where visit j is associated with a particular 10×10 km cell $i(j)$ by virtue of its location. We note that $i(j_1) = i(j_2)$ for a small number of distinct visits j_1 and j_2 , corresponding to a particular 10×10 km cell being visited multiple times (at different locations which mapped to the same cell on the large-scale grid). However, to make maximal use of the available data, we treated any such visits entirely independently in the following.

To facilitate direct comparison between the results of a single simulation and the results of the survey, we converted the model results to presence/absence predictions on the 10×10 km scale as follows. (i) Define $\pi_{i(j)}^s(\beta)$ as the proportion of host units infected by 2003 in $i(j)$ corresponding to survey visit j (according to simulation s with particular spore production rate β). (ii) Produce a presence/absence prediction for each survey visit j (for simulation s with

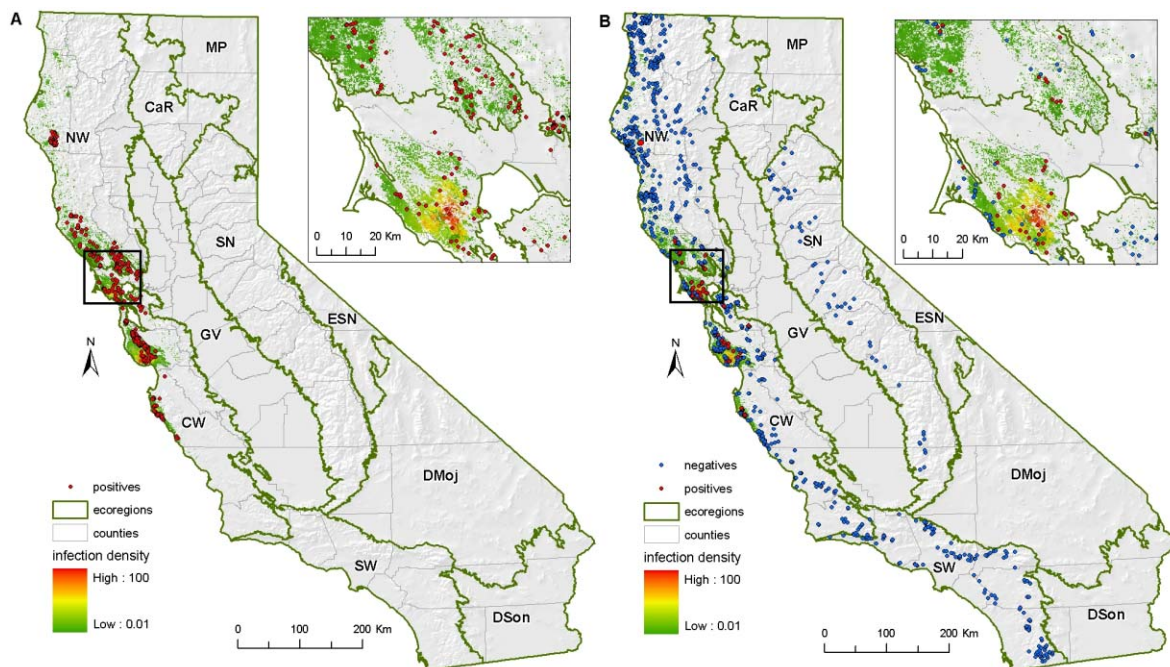


Fig. A1. Field data used for calibration and validation. Left panel shows simulated mean infection density through year 2007 in relation to COMTF positives (2001–2007) used in calibration. Right panel shows simulated infection density through 2006 in relation to independent field data described in Meentemeyer et al. (2008a).

Table A1. Confusion matrix.

Simulation status	Survey positive	Survey negative
Simulation positive	$H_+^s(\beta)$	$H_-^s(\beta)$
Simulation negative	$L_+^s(\beta)$	$L_-^s(\beta)$

spore production rate β) by sorting the probabilities $\pi_{i(j)}^s(\beta)$ and assigning the largest N_+ values to be “simulation positives” and the remaining N_- values to be “simulation negatives” (i.e., by using a data-driven threshold (Fielding and Bell 1997)).

We then constructed (for each simulation s and spore production rate β) a confusion matrix (Table A1; Fielding and Bell 1997) comparing the predictions of the model and the survey data where (for example) $H_+^s(\beta)$ was the number of positive visits that were correctly predicted by the model.

The objective function to be maximized was taken as the average odds ratio across the entire ensemble of simulations for each particular value of β

$$\bar{\rho}(\beta) = \sum_{s=1}^S \rho_s(\beta) / S,$$

where $S = 1000$, and $\rho_s(\beta)$, the odds ratio for an individual simulation s , was constructed from the confusion matrix via

$$\rho_s(\beta) = \frac{H_+^s(\beta)L_-^s(\beta)}{H_-^s(\beta)L_+^s(\beta)}.$$

Validation

Model validation was similar to calibration, with three differences: (i) The 2004–2006 survey data from Meentemeyer et al. (2008a) were used to validate the model, with calculations of $\pi_{i(j)}^s(\beta)$ to reflect the year in which a particular field observation occurred. (ii) To allow the odds ratio (corresponding to an entire ensemble of runs) to be examined with a single statistical test, we averaged the probabilities of infection across 1000 simulations for comparison to the positive/negative status of a field observation. (iii) In addition to validating at the scale of the entire state, we repeated the analysis for the Northwestern (NW) and Central Western (CW) ecoregions separately in order to compare accuracy by region and minimize the possibility of an over-optimistic assessment of model perfor-

mance which includes locations relatively far from the current known distribution of *P. ramorum*.

For the statewide scale and the specific NW and CW ecoregions, the odds ratio for the entire ensemble was given by

$$\rho(\beta) = \frac{H_+L_-}{H_-L_+}.$$

Goodness of fit was then assessed by testing the null hypothesis $\rho = 1$ (equivalent to “there is no association between the infection status of observations in the survey and in the model”). This hypothesis was assessed using Fisher’s exact test. Under the null hypothesis of no association, the number of correctly predicted positive observations, as they are drawn without replacement, should follow the hypergeometric distribution (Zar 2005).

Additional Assessment

In our validation procedure we used a standard thresholding method (Fielding and Bell 1997) to obtain binary presence/absence predictions involving a data-driven threshold on the probabilities of *P. ramorum* infection. Allowing the number of positive visits in the survey data to inform the partitioning of model results to presence/absence predictions avoids difficulties inherent in choosing optimal thresholds for binary classifiers (Jimenez-Valverde and Lobo 2007). However, we calculated Receiver Operating Characteristic (ROC) to provide assessment of model performance that is not constrained by the conversion to a binary prediction scale.

ROC curves visually summarize the accuracy of model prediction at the entire range of possible thresholds, and are shown for all three validation regions in Fig. A2. In all cases the ROC curve is skewed to the top left, indicative of good discrimination (Fielding and Bell 1997, Pearce and Ferrier 2000). Good model performance is supported by the Area Under the ROC Curve (AUC) at each scale; AUC(Statewide) = 0.89, AUC(NW) = 0.83, AUC(CW) = 0.76.

To further examine the performance of our model, we present box plots comparing the distributions of average probabilities of infection in 10×10 km squares according to their survey positive/negative status. Boxplots are shown for

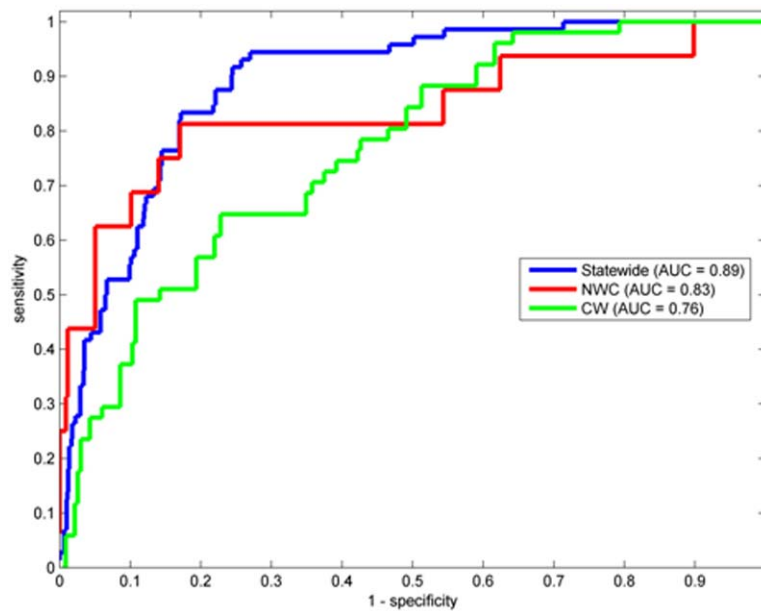


Fig. A2. Receiver Operating Characteristic (ROC) curves, showing model performance over all possible classification thresholds.

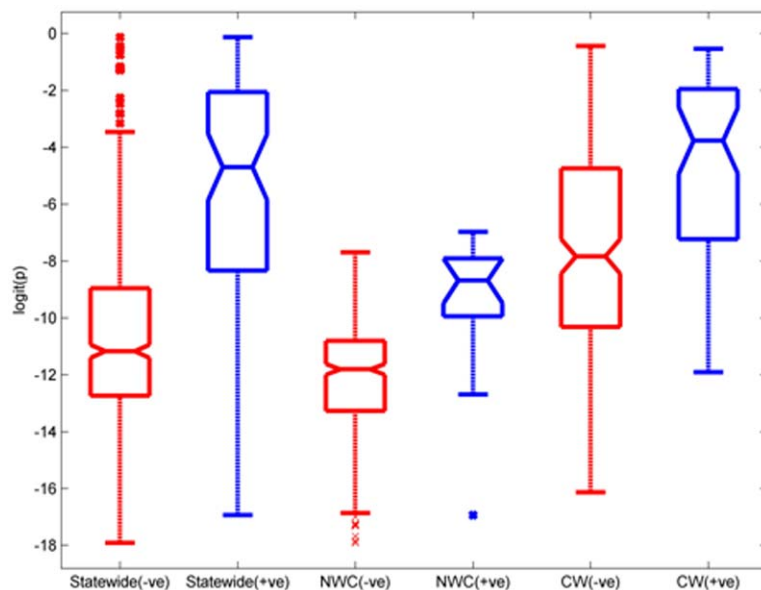


Fig. A3. Boxplots, showing model probabilities according to survey status of the corresponding 10×10 km cell for each validation region, statewide, Northwestern (NW), and Central Western (CW).

all three validation regions in Fig. A3; where probabilities are transformed to a logit scale. The distributions of model probabilities for each validation region are well-separated for positive

and negative surveys; this is strongly supported by Kruskal-Wallis tests (p values < 0.0001 for the whole state and for the NWC and CW ecoregions).

APPENDIX B

Animated version of Fig. 3 (mean infection density). [doi:10.1890/ES10-00192.2]

APPENDIX C

Animation of a single realization of the model, where long-range spread establishes relatively early in the epidemic. [doi:10.1890/ES10-00192.3]

APPENDIX D

Simulation of Appendix C, focused on the southern portion of the NW ecoregion. [doi:10.1890/ES10-00192.4]

APPENDIX E

Animation of a single realization of the model, where long-range spread establishes relatively late in the epidemic. [doi:10.1890/ES10-00192.5]