

Finding needles (or ants) in haystacks: predicting locations of invasive organisms to inform eradication and containment

DANIEL SCHMIDT,¹ DANIEL SPRING,^{2,6} RALPH MAC NALLY,² JAMES R. THOMSON,² BARRY W. BROOK,³ OSCAR CACHO,⁴
AND MICHAEL MCKENZIE⁵

¹*Faculty of Information Technology, Monash University, Clayton, Victoria 3800 Australia*

²*Australian Centre for Biodiversity, School of Biological Sciences, Monash University, Clayton, Victoria 3800 Australia*

³*Environment Institute and School of Earth and Environmental Sciences, University of Adelaide, Adelaide, South Australia 5005 Australia*

⁴*School of Business, Economics and Public Policy, University of New England, Armidale, New South Wales 2351 Australia*

⁵*Discipline of Finance, Faculty of Economics and Business, The University of Sydney, Sydney, New South Wales 2006 Australia*

Abstract. To eradicate or effectively contain a biological invasion, all or most reproductive individuals of the invasion must be found and destroyed. To help find individual invading organisms, predictions of probable locations can be made with statistical models. We estimated spread dynamics based on time-series data and then used model-derived predictions of probable locations of individuals. We considered one of the largest data sets available for an eradication program: the campaign to eradicate the red imported fire ant (*Solenopsis invicta*) from around Brisbane, Australia. After estimating within-site growth (local growth) and inter-site dispersal (saltatory spread) of fire ant nests, we modeled probabilities of fire ant presence for >600 000 1-ha sites, including uncertainties about fire ant population and spatial dynamics. Such a high level of spatial detail is required to assist surveillance efforts but is difficult to incorporate into common modeling methods because of high computational costs. More than twice as many fire ant nests would have been found in 2008 using predictions made with our method rather than those made with the method currently used in the study region. Our method is suited to considering invasions in which a large area is occupied by the invader at low density. Improved predictions of such invasions can dramatically reduce the area that needs to be searched to find the majority of individuals, assisting containment efforts and potentially making eradication a realistic goal for many invasions previously thought to be ineradicable.

Key words: *Bayesian models; Queensland, Australia; red imported fire ant; Solenopsis invicta; spread models; surveillance.*

INTRODUCTION

As humans increasingly dominate natural ecosystems, the invasive organisms people facilitate continue to establish and spread into new areas, causing large economic and environmental losses and human health problems (Mack et al. 2000). If detected early, eradication of invasive species may be possible (Veitch and Clout 2002). In circumstances in which eradication is not feasible due to late initial detection, much of the damages of an uncontrolled invasion can be avoided by maintaining invading populations at a low density (Simberloff 2009). Both eradication and successful control are facilitated by developing methods to find more invasive organisms with available resources. This makes improved prediction of organism locations a particularly high management priority.

In broad terms, there are two main approaches to model invasion dynamics. One is to gather detailed life-history data and use this information to simulate the spread process (Hastings et al. 2005). This approach is predicated on selecting the appropriate model form for spread dynamics and in deriving meaningful parameter estimates for the models. Use of life-history data from the species' native range may not well represent changes in attributes expressed or encountered in an organism's introduced range, especially when natural enemies are lost (Broennimann and Guisan 2008). Important aspects of a species' dynamics, such as its dispersal behavior, often cannot be observed or are hard to estimate reliably (e.g., large "jumps" that are rarely observed directly; Buchan and Padilla 1999).

An alternative approach is to use predictive statistical models to represent dynamics from data collected during the course of monitoring or eradication and control programs (Hastings et al. 2005). Estimating spread dynamics and predicting invader locations based on a sequence of positive and negative occurrence data is challenging because there are many sources of uncer-

Manuscript received 11 May 2009; revised 24 September 2009; accepted 30 September 2009. Corresponding Editor: T. J. Stohlgren.

⁶ Corresponding author.

E-mail: Daniel.Spring@sci.monash.edu.au

tainty. When some places (“sites”) that might contain individuals are not surveyed immediately, uncertainty arises about whether a given infestation arose from a surveyed or from an unsurveyed site. More uncertainty arises from imperfect detectability (Royle et al. 2007). If control actions are not completely effective, the repeat detections at a given site might occur because initial treatment was unsuccessful or because organisms recolonized. It is crucial to quantify the effects of such uncertainty on predictions of spread of invasive species if the predictions are to be used as efficient and effective guides to direct on-ground control actions.

It is likely that many biological invasions occupy a large area at a low density, with a relatively small number of clusters of invading organisms separated by large distances (Leung et al. 2004). Clustering often occurs as the result of local dispersal, and large distances between clusters occur when some organisms make long-distance “jumps” mediated by anthropogenic vectors or rare natural events (Suarez et al. 2001). In such circumstances, there may be many unsurveyed sites that might contain individuals despite surrounding areas having none, meaning that a large total area needs to be surveyed at a high cost. This is the form of invasion we consider here.

Although methods currently exist that can consider invasions of this form and quantify the effects of multiple sources of uncertainty based on a sequence of positive and negative occurrence data, there have been few applications. This reflects that resources for eradicating or controlling biological invasions usually are insufficient to be applied over a large area and a long period. Almost all such programs have failed to achieve their goals and detailed data typically are not available on failed eradication programs (Simberloff 2009). In deciding which of the many available methods to use to assist surveillance in practice, each method’s predictive accuracy and computational cost need to be considered. Hierarchical Bayesian methods (Hooten and Wikle 2008) have a high computational cost that increases rapidly with both number of sites considered and the proportion of those sites that have not been surveyed. Boosted regression trees (Broennimann and Guisan 2008) and maximum entropy models (Phillips and Dudík 2008) have been found to have relatively high predictive accuracy (Elith 2006), but do not all produce information required for targeting surveillance efforts. Although boosted regression trees have been used with success to identify potential distributional areas for invasive species (Broennimann and Guisan 2008), the approach is poorly suited to considering factors such as directional biases in dispersal because the approach is inherently nonspatial.

Here, we consider a large data set for an eradication program, the campaign to exterminate the red imported fire ant (*Solenopsis invicta*) from near Brisbane, Australia. The project is one of the larger eradication efforts to have been attempted, in terms of its

spatiotemporal extent and amount of data collected, reflecting that almost AU\$250 million has been spent on the project to date. Fire ants are one of the world’s 100 worst invaders (Lowe et al. 2000) and have the potential for extensive invasion worldwide (Morrison et al. 2004). Our model was applied to a study region comprising 657386 “sites” (each of 1 ha), many more than have been considered with other spatially explicit stochastic models of invasion spread. The data are in the form of point locations where detection of fire ant nests occurred and where active surveillance activity took place by Biosecurity Queensland Control Centre (BQCC). In addition to detections resulting from active surveillance, many detections were made incidentally by private citizens, without active searching for nests. Such “passive detections” are most common in urban areas, so we distinguished between urban and rural areas in our model.

Our approach, which is Bayesian, has two components: inter-site dispersal and intra-site growth. To represent longer-range spread, a suitable kernel (a function describing the relationship between dispersal distance and probability of dispersal) is centered on each infested site found in a previous year. The probability of a site being infested is determined by a base probability coupled with the proximity of the site to nests found in the previous year and parameters to account for human population density (urban vs. rural) and habitat suitability.

The intra-site growth component accounts for the increase in the number of nests within an infested site over time (i.e., immediate local spread of nests). We assessed the usefulness of model predictions by estimating occupancy probabilities from data for 2004–2007. The sites of highest predicted probability of occupancy then were compared with sites known to be occupied in 2008. We also compared the priority order of sites that would have been searched under two relatively naïve search strategies, one of which is similar to the strategy used by the invasion manager in the case study we consider (proximity-based search). Predictive validation and comparison with alternative spread dynamic models and associated search patterns is required to explore whether use of our model could lead to improved management (Mac Nally and Fleishman 2004).

*The red imported fire ant
in southeast Queensland, Australia*

Fire ant nests may contain a single queen (monogyne) or multiple queens (polygyne) (Ross et al. 1996). Both forms of fire ants were introduced in the Brisbane area, with the monogyne form found originally at the container wharf in Moreton Bay and the polygyne form found over a large area of the southwestern suburbs of Brisbane. Individuals from monogyne colonies are territorial, which results in a wider spacing and, therefore, lower density of colonies than in the non-territorial polygyne form of the species (Macom and

Porter 1996). Monogyne queens have stronger wing muscles, which allow them to fly longer distances than the smaller polygyne queens (Ross and Keller 1995). However, these differences do not necessarily have clear implications for spread dynamics. For example, it has not been established which form of fire ant has the higher rate of long-distance jumps. Although polygyne queens tend to fly shorter distances than monogyne queens, they probably are more likely to be transported long distances by humans (King et al. 2009). This reflects the higher density of polygyne colonies and the ability of new colonies to form from part of a polygyne nest, both of which make it more likely that soil picked up from an infested site will contain viable propagules (King et al. 2009). Both the monogyne and polygyne forms are able to be dispersed by flowing water, facilitating increased spread during the wet season in places where seasonal flooding occurs (Morrill 1974). For these reasons, it is possible that fire ant dispersal dynamics differ between monogyne and polygyne colonies, but in practice, we could not model the two forms separately because the available detection data did not distinguish between them.

Data set

A site represented a 1-ha square block of land within the area under consideration. A site was deemed infested if it contained one or more fire ant nests. A site was regarded as surveyed if it was either searched or treated (i.e., poisoned) during a year. The area under consideration was divided into a two-dimensional grid with $U = 707$ rows and $V = 935$ columns, covering a rectangular area of roughly 71×94 km (Fig. 1).

As we are interested in the dynamic behavior of an invading species, we further extend this grid in time so that one can refer to a particular site (u, v) at time t . Using this notation, if a quantity is indexed by three variables (u, v, t) , the first two (u, v) denote its position within the grid and the third, t , denotes the year, while quantities indexed by only two variables are static. We also introduce the following shorthand notation:

$$\sum_{uv}(\cdot) \equiv \sum_{u=1}^U \sum_{v=1}^V (\cdot).$$

The data consisted of fire ant nest counts, $p_{uv,t}$, which were the number of nests discovered at site (u, v) at year t and binary survey information $S_{uv,t} \in \{0, 1\}$ which indicate whether site (u, v) had been examined (searched, treated, or both) during a particular year t .

1) Let x_{uv}, y_{uv} denote the x and y coordinates of each site.

2) Let $h_{uv} \in [0, 1]$ denote the habitat suitability of site (u, v) . Habitat suitabilities range from 0 (least suitable for establishment of fire ant nests) to 1 (most suitable), which have been previously determined statistically by associations between detected fire ant nests and habitat attributes (R. George, *unpublished manuscript*).

3) Let $d_{uv} \in \{0, 1, 2\}$ represent the human population density at site (u, v) . This is a categorical variable, where $d_{uv} = 0$ means a completely unsuitable area (such as rivers and ocean), $d_{uv} = 1$ means a rural area, and $d_{uv} = 2$ means an urban area.

4) Let m_t be the number of surveyed sites in year t , i.e.,

$$m_t = \sum_{uv} S_{uv,t}$$

and n_t is the number of infested sites in year t , i.e.,

$$n_t = \sum_{uv} I\{p_{uv,t} > 0\}$$

and $I\{\cdot\}$ is the “indicator function,” returning a value of 1 if the condition inside the function is satisfied and 0 otherwise.

Intra-site growth model

An intra-site growth model was developed to represent the increase in the number of nests over time within an infested site. While the proposed model will not directly affect the dispersal prediction, it allows insight into rates of reproduction and spread once a site has been infested. It may also be used to provide an estimate of how many years a newly discovered infestation had existed. The only assumption we make (a weak one) is that once a site has been infested, the average number of nests within the site increases from year to year:

$$E[P_{uv,t+1}] > E[P_{uv,t}] \tag{1}$$

where $P_{uv,t}$ is a particular site at year t under the assumption that it has previously been infested by at least one propagule and $E[\cdot]$ denotes expectation. When an infested site is found, it is destroyed, so that sampling from the growth process removes the nests. If one has access to a large body of nest cluster counts from within surveyed infested sites and it is assumed that these nest counts were sampled from a process satisfying Ineq. 1, then one would expect to find several modes within the empirical distribution of the nest counts, each corresponding to $E[P_{uv,t}]$ for different values of t .

If the search-and-destroy method is effective, one would expect the mode corresponding to one year’s worth of growth to be much higher than the mode that corresponds to two years of growth, and so on. This partly reflects that most nests are discovered in urban areas, where detection occurs relatively quickly after initial establishment. Nest clusters detected soon after establishment usually contain fewer nests than clusters detected after a longer period because there is less time for propagules to be produced and for new nests to form. More generally, the probability of detecting at least one nest in a cluster increases with time because cluster size increases and because there has been more time to find nests. Explicit definition of plausible forms for the process (Ineq. 1) is difficult given the paucity of biological information, so a nonparametric approach

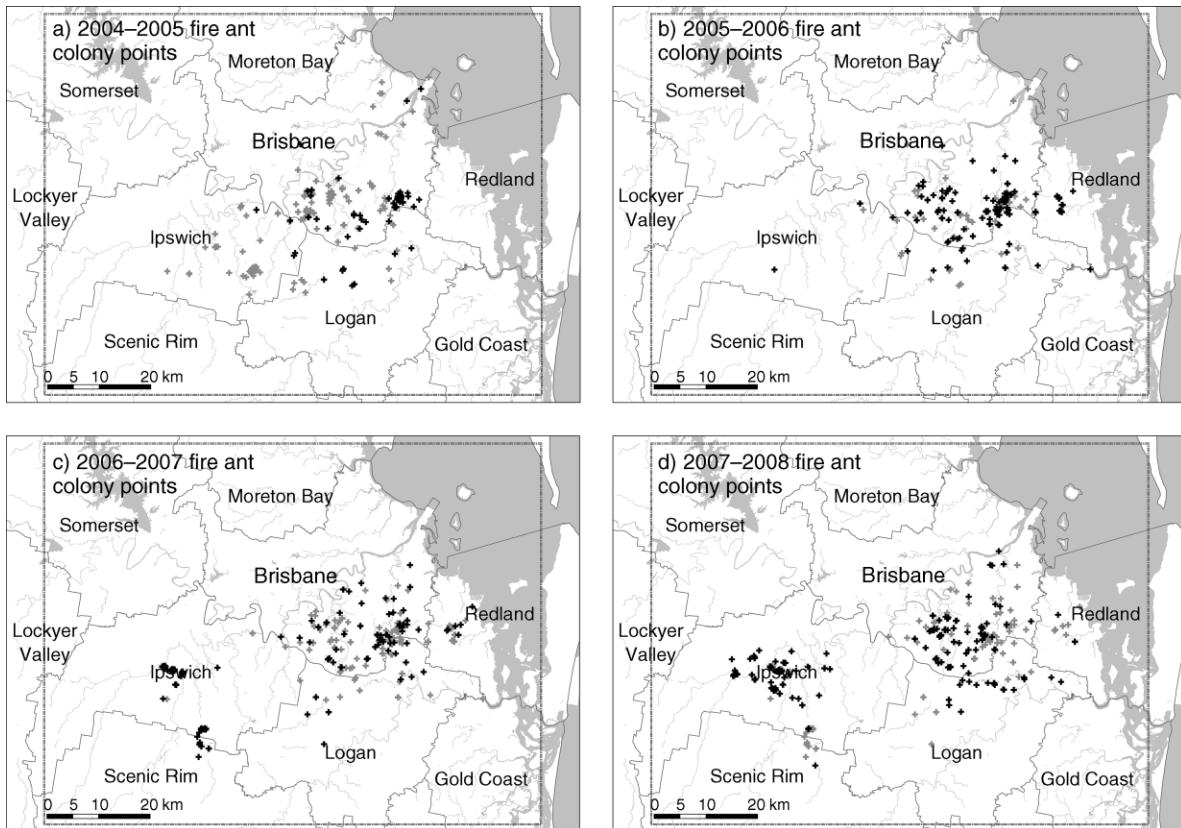


FIG. 1. Detections of the red imported fire ant (*Solenopsis invicta*) in selected years in the vicinity of Brisbane, Australia. The invasion is believed to have originated at the port of Brisbane (top right) and to have made an early long distance “jump” to (a) a central location, after which it spread (b–d) to nearby areas. New nest detections are denoted by black-shaded crosses; previously detected nests are denoted by gray-shaded crosses.

was taken. The process was approximated by a weighted mixture of K Poisson distributions

$$\Pr(P_j) = \sum_{k=1}^K w_k \text{Poi}(P_j | \lambda_k) \quad (2)$$

where w_k are the mixture weights satisfying

$$\sum_{k=1}^K w_k = 1.$$

$\text{Poi}(\cdot | \lambda_k)$ is a Poisson distribution with rate parameter λ_k . The mixture rate parameters λ and number of components K were estimated by a variant of the Snob unsupervised mixture modeler based on the minimum message length (MML) principle (Wallace and Dowe 2000, Wallace 2005). The mixture modeler was based on work by D. Schmidt (*unpublished manuscript*), which offers both an improved coding scheme and class parameter estimates. The mixtures would correspond to average growth rates expected for various t .

Statistical dispersal model

Although the data consist of nonnegative integers (i.e., the numbers of nests in each site per year), the

infested sites were sparse and generally distant from one another. For this reason, we divided growth into two components: an inter-site dispersal model that represented the spread of fire ant nests due to the dispersal of propagules beyond the immediate survey location and an intra-site growth model that represented the increase in the number of nests within an infested 1-ha site. The dispersal component of the model need only concern predicting whether a site will contain at least one nest in the next time period, which greatly simplifies prediction. An important issue was to decide which of the data were to be treated as “known.” Only those sites visited for search or for treatment within a given year (i.e., those sites for which $S_{uvi} = 1$) were regarded as “known.” Sites that were not visited were considered to be “missing.” The rationale and justification for this choice is discussed below in *Treatment of missing data*.

Inter-site dispersal model

While nonparametric “black-box” models, such as artificial neural networks or boosted classification trees, may offer superior predictive performance, these do not provide biologically informative interpretation. In contrast, our estimated parameters may be compared with

expert opinion or existing knowledge to verify plausibility. We chose to construct a model in which each parameter is related to a biologically important property of spread.

The main assumption underlying the dispersal model is that infested sites found in year $t + 1$ will be near to infested sites discovered in year t . This is modeled using radially decreasing kernels centered on the infested sites discovered in the previous year. The probability that a new site will be infested is determined by the probability that at least one propagule from the infested sites in the previous year successfully establishes, as well as being influenced by habitat suitability and human population density. Letting $\theta = (k, \sigma_1, \sigma_2, \eta, \theta_h, \theta_p)$ denote the vector of model parameters and $\Omega_t = \{u \in 1, \dots, U, v \in 1, \dots, V; p_{uv} > 0\}$ be the set of tuples containing indices of all infested sites in year t , the probability that site (u, v) will be infested in year $(t + 1)$, conditioned on the discovered sites of previous year t is

$$\Pr(p_{uv} > 0 | \Omega_t) = \underbrace{\left(1 - (1 - \eta)\Phi(u, v | \Omega_t)\right)}_{\text{Previous year nests}} \times \underbrace{\left(1 - \theta_h(1 - h_{uv})\right)}_{\text{Habitat effect}} \times \underbrace{\left(1 - \theta_p I\{d_{uv} = 2\}\right)}_{\text{Population effect}} I\{d_{uv} \neq 0\} \quad (3)$$

where η is a base prevalence rate of the fire ant infestations, $(\cdot)_+ = \max(\cdot, 0)$, θ_h controls the effect of habitat suitability, and θ_p controls the effect of human population density. The expression

$$\begin{aligned} \Phi(u, v | \Omega_t) &= \prod_{ij \in \Omega_t} \left(1 - \phi(x_{uv}, y_{uv} | x_{ij}, y_{ij}, \kappa)\right) \\ &= \prod_{ij \in \Omega_t} \left(1 - \frac{\kappa}{C(1 + \delta_{ijuv}^2)}\right) \end{aligned} \quad (4)$$

measures the contribution of sites thought to be infested in the previous year to the probability of infestation of a site in the current year. The quantity

$$\delta_{ijuv}^2 = \frac{(x_{ij} - x_{uv})^2}{\sigma_1^2} + \frac{(y_{ij} - y_{uv})^2}{\sigma_2^2} \quad (5)$$

is the weighted Euclidean distance from (x_{ij}, y_{ij}) to (x_{uv}, y_{uv}) . The parameters σ_1^2 and σ_2^2 control the amount of east–west and north–south spread, respectively, while the parameter κ represents the reproductive rate of the fire ants. The base rate parameter η has been included to ensure that the spread model is robust to outlying nests that arose by unpredictable long-distance dispersal. Given that the population density covariate d_{uv} is categorical, including an additional explicit effect for $d_{uv} = 1$ would lead to model identifiability problems because the effect is absorbed into the base rate η .

Expression 3 is multiplied by the term $I(d_{uv} \neq 0)$ to remove sites that were completely uninhabitable by fire ants, which primarily were large bodies of water, such as rivers and oceans.

The constant c is chosen so that the expected number of infested sites produced by a single nest within the span of a single year is κ , helping with interpretation of the final fitted model. We chose the Cauchy distribution for the kernel function because it is radially decreasing and has very heavy tails, allowing it to represent longer-distance spread more robustly than Gaussian or Student t distributions. The Cauchy has the added advantage of not requiring evaluation of transcendental functions such as $\exp(\cdot)$, and in addition to fitting the data well, the Cauchy kernel model was substantially quicker to compute than alternative kernels, such as the Gaussian. The calibrating constant c was found: for a suitably large area (which was ≥ 5 times the area infested), c is given by

$$c = \sum_{uv} \phi(x_{uv}, y_{uv} | x^*, y^*, \kappa = 1)$$

where x^* and y^* are the coordinates of the center of the area under consideration. At each site, p_{uv} is a Bernoulli random variate. This choice of c assists the model interpretation because the parameter κ corresponds to the expected number of new infested sites:

$$\sum_{uv} E[p_{uv}]$$

produced by a single infested site within the span of a year, in the absence of effects of habitat quality and human population density.

Likelihood function

The dispersal model can be used to form a likelihood function for the parameters θ by noting that the probability distribution over each site p_{uv} given θ is a set of Bernoulli variates. The dispersal model is designed to model establishment of new infested sites for the year $t + 1$, conditioned on the discovered infestations at year t . Thus, discovery of large groups of nests that have existed for ≥ 1 yr can seriously bias fitted parameters. We wish to avoid including large numbers of sites that are older than several years, which may arise through the discovery and follow-up search around a single infestation in an area distinct from those areas covered by the previous year’s search. Removing these sites from the list of sites to be predicted reduces bias in estimates. This reflects that these large groups of newly discovered, well-established infestations generally are distinct from the infested sites found in the previous year and thus would be highly unlikely to have arisen from propagules from these previously discovered infestations. All infested sites are used to form the conditioning set Ω_t , irrespective of age: i.e., all sites are always considered potential sources of propagules. The age of the nest

simply is used to determine whether they should be treated as data in the computation of the likelihood.

To estimate the age of a site, we used the nest count, $p_{uv,t}$, in conjunction with the intra-site growth model described below (see *Estimation of the inter-site dispersal model*). Following the logic driving this model, we need only to test whether the nest count belongs with highest probability to the class with the smallest rate parameter in our mixture model (Eq. 2) to determine whether we believe it has been in existence for a year. Practically, this will translate to a simple test of whether $p_{uv,t}$ is smaller than some threshold, ϵ , which is estimated from the data when we fit the intra-site growth mixture model (see *Results*). Given the assumptions that the observation (surveying) process is perfect and that unsurveyed sites are uninfested, the likelihood for data $\mathbf{P} = (p_{uv,t})$, $u = 1, \dots, U$, $v = 1, \dots, V$, from years $t = 1, \dots, T$, given parameters θ is

$$f(\mathbf{P} | \theta) = \prod_{t=1}^T \prod_{u,v=1}^{U,V} \Pr([p_{uv,t} > 0 | \Omega_t])^{I_{uv,t}} \cdot \Pr(p_{uv,t} = 0 | \Omega_t)^{1-I_{uv,t}}. \tag{6}$$

This yields a negative-log likelihood of

$$-\log f(\mathbf{P} | \theta) = -\sum_{t=1}^T \sum_{u,v=1}^{U,V} \left[I_{uv,t} \log \Pr(p_{uv,t} > 0 | \Omega_t) + \left((1 - I_{uv,t}) \log \Pr(p_{uv,t} = 0 | \Omega_t) \right) \right] \tag{7}$$

where $I_{uv,t} = I\{p_{uv,t} > 0 \wedge p_{uv,t} < \epsilon\}$ indicates whether a site (u, v) at year t is infested.

Estimation of the inter-site dispersal model

A Bayesian approach (Box and Tiao 1973) was taken to estimate the model from the data. The complete model prior $\pi(\theta)$, with $\theta = (\kappa, \sigma_1, \sigma_2, \eta, \theta_h, \theta_p)$ was composed of the following densities chosen for the model parameters:

$$\pi(\sigma_1, \sigma_2) \propto \frac{1}{\sigma_1 \sigma_2} \tag{8}$$

$$\pi(\eta, \theta_h, \theta_p) \propto 1 \tag{9}$$

$$\pi(\kappa) \propto \exp(-\alpha \kappa). \tag{10}$$

The priors for the two scale parameters σ_1 and σ_2 were chosen to be scale invariant, while uniform priors were deemed suitable for the base rate, habitat, and human population density parameters. The kernel scalar (reproductive capability) parameter was given an exponential prior with $\alpha = 1/50$, yielding an a priori expectation that a single infested site had the potential to produce many new infestations within a year. All priors were chosen to be near uninformative.

Using the likelihood function 6 and priors 8–10, we formed the posterior distribution $\pi(\theta | \mathbf{P}) \propto f(\mathbf{P} | \theta)\pi(\theta)$ of θ given the observed data \mathbf{P} and priors $\pi(\theta)$. The normalization constant is difficult to compute analytically, so a Markov chain Monte Carlo (MCMC) approach was employed. Sampling from the posterior distribution was performed using the Metropolis-Hastings algorithm (Robert and Casella 1999). The proposal distribution chosen was a multivariate normal distribution centered at the maximum a posteriori estimate, $\hat{\theta}_{\text{MAP}}$ of θ :

$$\hat{\theta}_{\text{MAP}} = \arg \max_{\theta} \{f(\mathbf{P} | \theta)\pi(\theta)\} \tag{11}$$

with covariance matrix proportional to the inverse of the Hessian of Eq. 7 evaluated at $\hat{\theta}_{\text{MAP}}$, the constant of proportionality being adjusted to achieve a 40% acceptance ratio.

Model selection

Model complexity can be controlled by fixing various parameters. For example, both population and habitat effects can be removed by setting $\theta_h = 0$ or $\theta_p = 0$, and the model can be reduced to isotropic spread by forcing $\sigma_1 = \sigma_2$. This property is very helpful if one were performing model selection or were to test a hypothesis such as whether habitat suitability plays a significant role in the spread of the invasive species.

The penalized fit of the model may be used to guide whether one should include or remove parameters in the dispersal model. The Bayes factor is a common method to compare two models, which accounts for both model fit and model complexity. Bayes factors often are difficult to compute from posterior samples, but an approximation to the marginal probability is given by the Bayesian information criterion (BIC) (Schwarz 1978):

$$-\log \int f(\mathbf{P} | \theta)\pi(\theta) d\theta = -\log f(\mathbf{P} | \hat{\theta}_{\text{MAP}}) + \frac{k}{2} \log n + O(1) \tag{12}$$

where $\hat{\theta}_{\text{MAP}}$ is the MAP estimator (Eq. 11), k is the number of free parameters, n is the number of data points, and $O(1)$ denotes terms of order unity (and hence, independent of n). Once a model structure has been chosen by minimizing the BIC score, one may use posterior sampling to derive Bayesian credible sets characterizing model parameters.

Treatment of missing data

Justification.—The model was estimated from the data with unsurveyed sites, i.e., those for which $S_{uv,t}$ were treated as uninfested. That is, the fire ant is treated as being absent from the site, while for sites that are surveyed, we used the observation obtained for the site. This allowed us to model the entire area currently delimited by the BQCC for management of the fire ant invasion at a sufficiently fine spatial scale (1 ha) to inform surveillance efforts. Had we taken a fully

Bayesian approach in which occupancies of unsurveyed sites were treated as random variables, the computational cost would have been prohibitive. Over the five years used in model building, only 437 536 of a possible 3 305 225 sites were surveyed by the BQCC. Even if surveying were perfect, the status of only approximately one-sixth of the potential sites was known. While it may seem disingenuous to treat unsurveyed sites as “uninfested,” there are several reasons for this choice. These areas were not surveyed because they were deemed unsuitable habitat (lakes, dense forest) or because there were no passive detections. Passive surveillance in urban areas has a high probability of finding fire ant nests when present (BQCC, *unpublished data*), and such areas occupy approximately half of the region. If no nests are detected in such areas, there probably are no nests there. This, coupled with a low occurrence rate of infested sites in comparison to total sites, suggests that there were very few infested sites within unsurveyed urban areas.

Another reason is the surveying procedure implemented by the BQCC. Every time an infested site was discovered, all nearby sites within a radius of 500 m or 1000 m were surveyed. If new infestations were discovered, a radius was drawn around the newly discovered infestation and the process continued. Thus, in the BCQQ data set, there was consistently a large buffer of uninfested sites between any discovered infestation and the unsurveyed sites. Given that we can only expect to learn about local (i.e., short-to-medium range) dispersal, even if there were unknown populations of fire ants in the unsurveyed areas, knowledge of this would be of little use because these populations must have either (1) arrived by unpredictable, rare, long dispersals or (2) be from short-range dispersal completely within the unsurveyed sites. In either of these scenarios, knowing (or suspecting) that the populations exist in these unsurveyed sites, which are, by the nature of the surveying process, a large distance from any known infestation, will contribute almost nothing to the estimation of the short-to-medium range dispersal behavior of the populations that were discovered. The simple assumption of ignorance should lead to a negligible loss of information, while taking the steps to construct a fully Bayesian model will lead to manifold increased complexity for marginal expected gains.

Impact on inferences.—Does treating unsurveyed sites as uninfested have a large impact on inference? Again, the nature of the data and the surveying process means that there is essentially no effect on estimation of the parameters, irrespective of how large the modeled area is taken to be and, hence, how much missing data there are. It is undesirable and unrealistic if inferences were affected by the amount of missing data included. The spread parameters, κ , σ_1 , and σ_2 , are invariant to the amount of missing data because of the large buffer of surveyed empty cells surrounding any infested site. The habitat and population parameters, θ_h and θ_p , are also largely invariant to the amount of missing data if one

TABLE 1. Estimated intra-site growth in the number of red imported fire ant (*Solenopsis invicta*) nests per year, Brisbane, Australia.

Class	λ_k	w_k	P
1	1.73 (0.001)	0.9275	1–5
2	7.46 (0.07)	0.0569	6–13
3	15.29 (0.662)	0.0123	14–24
4	33.72 (5.403)	0.0033	25+

Note: Abbreviations are: λ_k , mean number of nests per class; w_k , proportion of sites in each class; k , a manually selected number of years that partitions the data into two classes, one from year 2001 to year 2001 + k and the other from year 2001 + k to year 2007; P , the number of nests that are classified as being in a particular class, and thus, of a particular age. Values for λ_k are expressed as mean with SE in parentheses.

considers the behavior of the model (Eq. 3) in the unsurveyed cells. Any unsurveyed cell will be a large distance from a surveyed, infested site, and so the term due to the radially decreasing spread kernels, $\Phi(\cdot)$, $\cong 1$. For simplicity, we let $d_{uv} = 2$ at the cell of interest. If we consider the partial derivative of Eq. 3 with respect to θ_h , we have

$$\frac{\partial \Pr(p_{uvt} = 0 | \Omega_t)_+}{\partial \theta_h} = - \left(1 - (1 - \eta) \Phi(u, v | \Omega_t) \right) (1 - \theta_p) \times \frac{\partial \left(1 - \theta_h (1 - h_{uv}) \right)}{\partial \theta_h}. \quad (13)$$

As noted, due to the distance from the unsurveyed cells to the closest member of Ω_t (i.e., the closest surveyed site), $\Phi(u, v | \Omega_t) \approx 1$, and the overall gradient will be ~ 0 . A similar expression is obtained by differentiating with respect to the population effect θ_p . Thus, setting the missing data to empty does not lead to bias in estimation of the population and habitat effect parameters because they only have an effect for those sites for which $\Phi(\cdot)$ is not too near unity. Due to the buffer, these will generally only correspond to the surveyed sites.

RESULTS

Intra-site growth model

The Poisson mixture model for intra-site growth, Eq. 2, was estimated from a data set formed by aggregating all 1882 nonzero values of p_{uvt} for all years. Intra-site growth was assumed independent of the behavior of inter-site dispersal so the difference in dispersal patterns in years $t = 1$ and $t = 2$ from $t = 3, \dots, 5$ was unimportant.

Parameter estimates (and approximate standard errors) for the selected four-component mixture model are summarized in Table 1, the components being ordered by decreasing class weight w_k . The class weights decrease with increasing class rate parameter, with most of the weight assigned to the class with the smallest rate parameter. The hypothesis that the rate parameter of class k (the classes being sorted by decreasing relative abundance w_k) represented average growth expected

TABLE 2. Dispersal component of the model: details of posterior probability distributions for model parameters.

Parameter	Interpretation	Mode	Mean	SD
σ_1	east–west spread	4.97	2.34	0.32
κ	reproductive rate	1.91	2.01	0.12
η	outliers	0.000010	0.000009	0.000005
θ_h	habitat effect	1.40	1.39	0.015
σ_2	north–south spread	2.35	4.74	0.67
θ_p	urban effect	0.24	0.24	0.03

Notes: North–south and east–west are expressed in the same units and are directly comparable. The reproductive rate is the number of new nests produced per nest.

after $t = k$ years of growth in a site was thought to be plausible by experts at BQCC so the model was deemed suitable. Therefore, simulation of growth in a site in year $(t + 1)$ may be done by sampling from $p_{uv,t+1} \sim \text{Poi}(\lambda_{t+1})$. The intra-site growth model allows one to estimate the number of years that a site has been infested based on the nest count for that particular site. The rightmost column of Table 1 (P), which shows the range of nest counts that would classify a site as being in a particular age group, provides an estimate of the number of years that a newly discovered site has been infested.

Preprocessing of the data

While the data provided by the Queensland government agency covered seven years, 2001–2007, we only made use of data from years 2003–2007 when estimating the dispersal model parameters. This was done because the data for the first two years exhibited dissimilar dynamics to those observed from year 2003 onward. The data were divided into two segments, one from year 2001 to year $2000 + k$ and one from year $2000 + k$ to year 2007. Models were fitted to both sets of data and we chose a value of k of 2 because that yielded the minimum combined BIC score (penalized negative log-likelihood) of the two models. Parameter estimates when fitted to years 2003–2007 were relatively stable but changed considerably when years 2001 and 2002 were included. Thus, the choice was made to focus on the data for years 2003–2007 to better represent dynamics of the controlled invasion, with years 2001 and 2002 being more representative of unchecked invasion dynamics. For convenience, we calibrate the t variable such that $t = 1$ denotes year 2003, $t = 2$ year 2004, and so on.

The second preprocessing step was to determine which sites were likely to have been infested for a single year. Given the discovered intra-site growth model, classifying a site to be one year old may be done by setting $\epsilon = 6$, i.e., testing whether $p_{uv,t+1} < 6$, as one year of growth was expected to produce between one and five colonies.

Fitting the dispersal model

The statistics of the posterior distribution are shown in Table 2. The data suggested a greater tendency for spread to occur in the east–west than in the north–south directions. The habitat effect and the urban effect are on similar scales, so the habitat suitability index score had a greater effect on fire ant spread than did the urban or

rural classification of the site. The posterior standard deviations for the urban effect, reproductive rate, and the habitat effect are small compared to the maximum a posteriori (MAP) point estimates, found by solving Eq. 11. This indicates that there is small uncertainty about these parameters. An illustration of the spread of the fire ant invasion is given in Fig. 1. There appears to be an east–west bias of spread identified by the model (Table 2, Fig. 1).

Future validation for 2008 and comparison against BQCC strategy

To test the effectiveness of the estimated model, a simulated targeted search was undertaken on the 2008 data, conditioned on the nests found in 2007. The 2008 data were obtained after the model was developed and estimation was completed and therefore represent newly collected validation data. To evaluate the practical utility of our model we compared the number of nests found using the model with the number found using the current method of the BQCC, which we refer to as “proximity search.” That method is to search a predetermined radius around each detected nest (whether it be detected with passive surveillance or by BQCC staff). This can be described as a naïve method because it does not account for dispersive jumps, as our model does. Under the simulated search strategy carried out by BQCC in 2008, an equal area is searched around each of the nests detected in 2007. We considered 15 different resource levels, ranging from 1975 to 85 316 ha. Each total resource level corresponds to a given search radius around each nest detected in 2007. For example, a total search area of 1975 ha corresponds to a search radius of two cells (~20 ha), while a total search area of 85 316 ha corresponds to a search radius of 30 cells (~2900 ha). To compare the BQCC search strategy with our model, we ranked sites according to their probability of being infested and “searched” the sites with the highest probabilities, subject to the constraint on the total search area.

In 2008, BQCC searched 43 811 ha and 187 infested sites were found. The majority of those discoveries were due to passive detection in which members of the public reported ant infestations, and most of the remaining discoveries were made in follow-up searches conducted in the area of reported infestations.

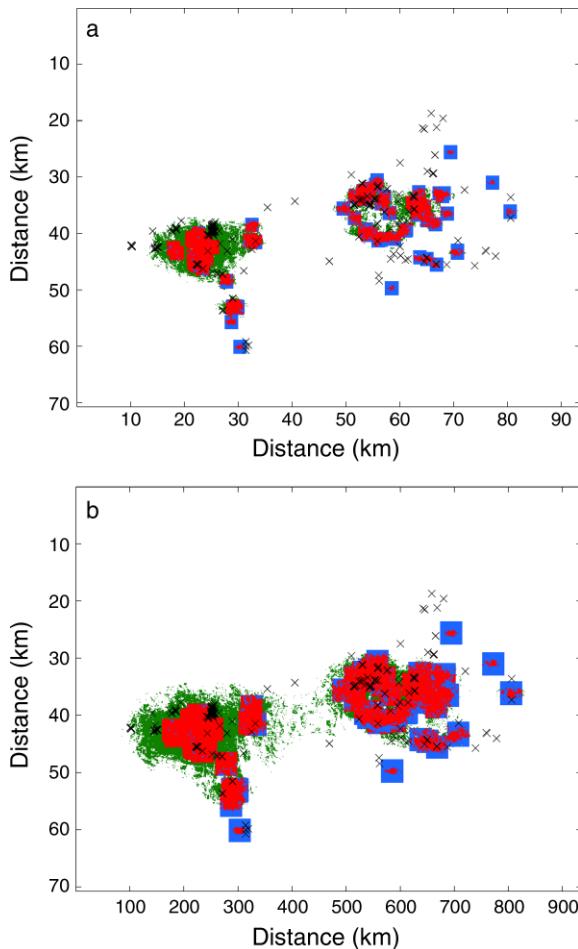


FIG. 2. Sites selected under a standard proximity-based search (blue shaded) and sites selected using our jump dispersal model (green shaded). Sites selected under both strategies are highlighted in red, and nests actually detected in 2008 are marked with × symbols. Total search areas were (a) 20 846 ha and (b) 53 290 ha.

Simulated search was undertaken by ignoring passive detections and performing a targeted search of 43 800 ha, guided by the probability map produced by the dispersal model. The simulation revealed that 148 of the 187 infested sites would have been discovered by targeted searching guided by the probability map. Thus, ~80% of nests discovered in 2008, largely by passive detection, would have been found using our model (Fig. 2). In comparison, a random search of the area under consideration would be expected to find ~11 infested nests, with the BQCC’s proximity-based search method (described in *Future validation for 2008...*) expected to find ~90 nests. Within the range of search effort typically available to BQCC (20 000–40 000 ha/yr), there is a large advantage in using our model to choose where to search for fire ant nests, with the BQCC search protocol finding less than half the number of nests (Fig. 3).

The BQCC search strategy can be interpreted as search over a probability map generated by a variant of our model. In that model variant, one ignores the effect of habitat and population density and chooses a kernel function that is constant inside a square centered on each of the previously infested sites in the set Ω_t , with probability zero outside. The primary difference between the BQCC version of our model and our own variant is that such square “kernels” do not interact in the former version. Rather, intersection between square kernels that may arise due to close proximity of several sites discovered in the previous year does not increase the probability of infestation in the areas in which the kernels overlap. This is in contrast to our model, in which proximity of several infested sites in the previous year will increase the probability of infestation in the areas that they overlap and thus raise the search priority of these areas.

DISCUSSION

Predicting locations of individuals of a given species is critical to many areas of ecology and conservation biology (e.g., invasive species, Peterson 2003; habitat suitability models, Fleishman et al. 2001; endangered species management, Lindenmayer and Possingham 1996; biodiversity assessment, van Jaarsveld et al. 1998; restoration ecology, Thomson et al. 2009). The prediction problem for invasive organisms is difficult because invader locations are generally in a state of relatively high flux, so models are nonstationary in character (Suarez et al. 2001). Many existing (statistical) models consider multiple sources of uncertainty, includ-

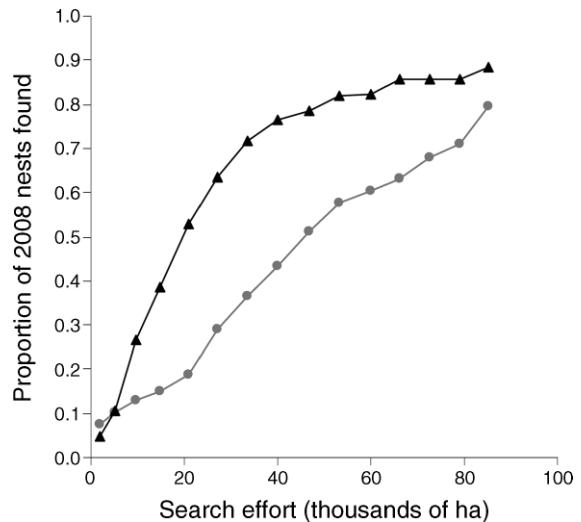


FIG. 3. Within the range of search effort that typically is available to Queensland Biosecurity Control Centre (20 000–40 000 ha/yr), there was a large advantage in using our model (triangles) compared with the standard approach (circles) to choose where to search for fire ant nests, with almost twice the detection rate.

ing process and state uncertainty, but are not well suited to representing invasions occupying a large area at low density, in which many sites have not yet been scrutinized.

The key characteristics of the fire ant invasion were: (1) an invasion in its relatively early stages and (2) a comparatively rich data set with a sequence of yearly information. We also assumed that the biology of the focal species involves intra-site population growth (accretion to nearby nests) and saltatory inoculation of distant sites, but is otherwise poorly understood. Under these circumstances, our model was demonstrably much more effective at predicting locations of fire ant locations than a simple proximity-based model, which had been the approach relied upon by management until the creation of our model. Because most invasion processes probably include both local dispersal and long distance “jumps” and many invaders have poorly understood biological attributes, our modeling structure has wide potential application.

Our method considered multiple sources of uncertainty in invasions occupying a large area at low density. The latter condition allowed us to make a simplifying assumption that there were no nests in unsearched areas. Given that the probability of occurrence at any randomly selected site was $\hat{\eta}_0 = 0.0016$ (probably a substantial overestimate given that this estimate excludes unsearched sites), this simplification was reasonable and offered huge benefits in the extent and grain (*sensu* Wiens et al. 1987) of the problem one could address within computational limits. Moreover, parameter inference should be invariant to inclusion of missing data about which nothing is known (i.e., an arbitrarily large overall area of consideration, the spatial extent). In our formulation, only known data contributed to parameter estimates. We also specified that the effects of habitat suitability and human population density enter in a multiplicative fashion, scaling the probability of infestation determined by the spread kernel model. Given the average distance of the unsurveyed sites from the set of known infested sites, the probabilities determined for the spread kernels are very low and the habitat and human population density parameters have little influence in probability estimates for the unsurveyed sites. This is advantageous because the habitat and human population density values at the sites of unsurveyed data will not contribute to the estimation of the habitat and human population density parameters.

One of the reasons for the great superiority of our model over the proximity-based search method was that many of the sites ranked relatively highly by our model did not fall within the search radius used in the latter, currently used approach. For example, in the leftmost cluster of infested sites (Figs. 1 and 2), there are “sub-clusters” separated by a distance greater than the search radius under proximity-based search. Those sub-clusters were not searched under that strategy. It seems likely

that there would be some infestations between such sub-clusters, assuming that nests are dispersed according to a Cauchy kernel (as we built into our model).

While our model clearly is much superior to the existing method used for this program of eradication, it is not clear whether the gains from using the model would be sufficient to increase the probability of eradication. Fig. 3 illustrates that if the area searched is between 20 000 and 40 000 ha, large savings can be made in search effort by using our model instead of the current search method. When resources are limited such savings can be of critical importance in controlling, rather than eradicating the invasion, because the savings would help the invasion manager to maintain population densities at a low level with available resources. However, eradication may still be infeasible. In our case study, a small proportion of fire ant nests result from unpredictable long-distance jumps. Such nests would not readily be found using a predictive model such as ours and may only be found by conducting an exhaustive search of a much larger area (Fig. 3). Such an exhaustive broad-area search effort would probably have a prohibitive cost when conventional search methods are used, implying that alternative surveillance methods need to be considered to achieve eradication. Using the model presented here, BQCC decided to change its eradication strategy to include improved predictions made with our model and a new search method, remote sensing, to find nests whose locations are difficult to predict. Remote sensing can cover a much larger area than can be searched with the current ground surveillance methods but has lower sensitivity. It is not yet known whether the combination of remote sensing and improved ground surveillance using our model would increase eradication probability. This important question is left to future research.

Our modeling structure offers much potential benefit in identifying areas of high probability of occurrence of individuals in systems experiencing rapid flux, including not only biological invasions but also shifts in species ranges under anthropogenic change. Improving the efficiency of detection of invasive or threatened organisms at low densities, when management intervention is often most crucial but relevant information is sparse or expensive to collect, will greatly enhance our capacity to manage sensitive systems so as to mitigate negative impacts in environmentally sensitive areas.

ACKNOWLEDGMENTS

This work was supported by the Australian Research Council Discovery Grant scheme (DP0771672). The authors acknowledge the financial and other support provided by the National Red Imported Fire Ant Eradication Program and participants at workshops held at the Biosecurity Queensland Control Centre and Monash University. Data were provided by Bob Bell, Biosecurity Queensland Control Centre. Dennis O'Dowd provided timely advice and many helpful points in discussion. This is publication number 189 from the Australian Centre for Biodiversity.

LITERATURE CITED

- Box, G., and G. Tiao. 1973. Bayesian inference in statistical analysis. Addison-Wesley, Reading, Massachusetts, USA.
- Broennimann, O., and A. Guisan. 2008. Predicting current and future biological invasions: both native and invaded ranges matter. *Biology Letters* 4:585–589.
- Buchan, L. A. J., and D. K. Padilla. 1999. Estimating the probability of long-distance overland dispersal of invading aquatic species. *Ecological Applications* 9:254–265.
- Elith, J., et al. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29: 129–151.
- Fleishman, E., R. Mac Nally, J. P. Fay, and D. D. Murphy. 2001. Modeling and predicting species occurrence using broad-scale environmental variables: an example with butterflies of the Great Basin. *Conservation Biology* 15: 1674–1685.
- Hastings, A., et al. 2005. The spatial spread of invasions: new developments in theory and evidence. *Ecology Letters* 8:91–101.
- Hooten, M. B., and C. K. Wikle. 2008. A hierarchical Bayesian non-linear spatio-temporal model for the spread of invasive species with application to the Eurasian Collared Dove. *Environmental and Ecological Statistics* 15:59–70.
- King, J. R., W. R. Tschinkel, and K. G. Ross. 2009. A case study of human exacerbation of the invasive species problem: transport and establishment of polygyne fire ants in Tallahassee, Florida, USA. *Biological Invasions* 11:373–377.
- Leung, B., J. M. Drake, and D. M. Lodge. 2004. Predicting invasions: propagule pressure and the gravity of Allee effects. *Ecology* 85:1651–1660.
- Lindenmayer, D. B., and H. P. Possingham. 1996. Ranking conservation and timber management options for Leadbeater's possum in southeastern Australia using population viability analysis. *Conservation Biology* 10:235–251.
- Lowe, S., M. Browne, S. Boudjelas, and M. De Poorter. 2000. 100 of the world's worst invasive alien species: a selection from the Global Invasive Species Database. International Union for the Conservation of Nature and Natural Resources, Gland, Switzerland.
- Mack, R. N., D. Simberloff, W. M. Lonsdale, H. Evans, M. Clout, and F. A. Bazzaz. 2000. Biotic invasions: causes, epidemiology, global consequences, and control. *Ecological Applications* 10:689–710.
- Mac Nally, R., and E. Fleishman. 2004. A successful predictive model of species richness using 'indicator' species. *Conservation Biology* 18:646–654.
- Macom, T. E., and S. D. Porter. 1996. Comparison of polygyne and monogyne red imported fire ant (Hymenoptera: Formicidae) population densities. *Annals of the Entomological Society of America* 89:535–543.
- Morrill, W. L. 1974. Dispersal of red imported fire ants by water. *Florida Entomologist* 57:39–42.
- Morrison, L. W., S. D. Porter, E. Daniels, and M. D. Korzukhin. 2004. Potential global range expansion of the invasive fire ant, *Solenopsis invicta*. *Biological Invasions* 6: 183–191.
- Peterson, A. T. 2003. Predicting the geography of species' invasions via ecological niche modeling. *Quarterly Review of Biology* 78:419–433.
- Phillips, S. J., and M. Dudík. 2008. Modeling of species distributions with Maxent: new extensions and comprehensive evaluation. *Ecography* 31:161–175.
- Robert, C., and G. Casella. 1999. Monte Carlo statistical methods. Springer-Verlag, New York, New York, USA.
- Ross, K. G., and L. Keller. 1995. Ecology and evolution of social organization: insights from fire ants and other highly eusocial insects. *Annual Review of Ecology and Systematics* 26:631–656.
- Ross, K. G., E. L. Vargo, and L. Keller. 1996. Social evolution in a new environment: the case of introduced fire ants. *Proceedings of the National Academy of Sciences USA* 93: 3021–3025.
- Royle, J. A., M. Kéry, R. Gautier, and H. Schmid. 2007. Hierarchical spatial models of abundance and occurrence from imperfect survey data. *Ecological Monographs* 77:465–481.
- Schwarz, G. 1978. Estimating the dimension of a model. *Annals of Statistics* 6:461–464.
- Simberloff, D. 2009. We can eliminate invasions or live with them: successful management projects. *Biological Invasions* 11:149–157.
- Suarez, A. V., D. A. Holway, and T. J. Case. 2001. Patterns of spread in biological invasions dominated by long-distance jump dispersal: insights from Argentine ants. *Proceedings of the National Academy of Sciences USA* 98:1095–1100.
- Thomson, J. R., A. Moilanen, P. Vesik, A. F. Bennett, and R. Mac Nally. 2009. Where and when to revegetate: a quantitative method for scheduling landscape reconstruction. *Ecological Applications* 19:817–828.
- van Jaarsveld, A. S., S. Freitag, S. L. Chown, C. Muller, S. Koch, H. Hull, C. Bellamy, M. Krüger, S. Enrödy-Younga, M. W. Mansell, and C. H. Scholtz. 1998. Biodiversity assessment and conservation strategies. *Science* 279:2106–2108.
- Veitch, C., and M. Clout. 2002. Turning the tide: the eradication of invasive species. International Union for the Conservation of Nature and Natural Resources, Gland, Switzerland.
- Wallace, C. S. 2005. Statistical and inductive inference by minimum message length. Springer-Verlag, New York, New York, USA.
- Wallace, C. S., and D. L. Dowe. 2000. MML clustering of multi-state, Poisson, von Mises circular and Gaussian distributions. *Statistics and Computing* 10:73–83.
- Wiens, J. A., J. T. Rotenberry, and B. Van Horne. 1987. Habitat occupancy patterns of North American shrubsteppe birds: the effects of spatial scale. *Oikos* 48:132–147.