Supporting Information

Keith and Spring 10.1073/pnas.1216146110

Software Validation

To ensure that the software implementing our Markov chain Monte Carlo (MCMC) method was correctly encoded, we used the software validation technique of Cook et al. (1). The method involves the following steps, performed repeatedly.

i) Draw the parameters of the model from the prior distribution.

ii) Simulate data by drawing from the likelihood model, given these parameters.

iii) Sample from the posterior distribution for the parameters, using MCMC.

iv) Estimate the posterior quantiles of the parameters drawn at step *i* (or functions of these parameters), using the sample drawn at step *iii.*

The posterior quantiles from multiple repetitions are then processed to generate a statistic z_θ for each parameter or function of parameters. If the software is implemented correctly, this statistic should have a standard normal distribution, and extreme values are detected graphically.

To keep the simulated datasets relatively small and ensure computational feasibility, we selected the following prior distributions. (Note these prior distributions differ from those used in the larger-scale analyses described in the main text and in *SI Materials and Methods* below.) The number of initial nests was selected uniformly in the range 1–10, and the positions of initial nests were distributed uniformly on the rectangle bounded by x coordinates 470,000 and 530,000 and y coordinates 6,920,000 and 6,970,000. The vector of mixture proportions γ (defined in *SI Materials and Methods* below, together with the other parameters discussed in this section) was selected from a uniform Dirichlet distribution, and the SDs of jump distances σ_{Xi} and σ_{Yi} (for $i = 1$, 2, 3, 4) were selected from an inverse gamma distribution with shape parameter 2.0 and rate parameter 98.0. (The latter prior effectively constrains the average founding distance in the range 1 m to 1 km.) Treatment and targeted search efficacies α_0 and α_1 were selected uniformly in the range (0.8, 1.0), and the efficacies of passive detection α_2 to α_5 were selected uniformly in the interval (0.01, 0.1), subject to the constraint that $\alpha_2 \geq \alpha_3 \geq \alpha_4 \geq \alpha_5$. Constraining the passive detection probabilities above 0.01 helps limit the rate of expansion, and constraining them below 0.1 prevents most simulated invasions from ending in rapid eradication. Relative establishment probabilities β_2 to β_4 were selected uniformly in the interval (0,1) subject to the constraint $\beta_2 \geq \beta_3 \geq \beta_4$. The reproductive rate (number of nests founded per nest per month) λ was drawn from a gamma distribution with shape parameter 5.0 and rate parameter 15.0. Again, this prior confines λ to a range that limits expansion but also limits the probability of rapid eradication.

After each set of parameter values was drawn from the prior distribution, an invasion was simulated, consisting of a 36-mo unhindered expansion phase without any detections or treatment, followed by a 60-mo period during which the simple eradication strategy described in *SI Materials and Methods* below was simulated. As in the larger analyses described in the main text and sections below, after each nest was founded there was a 6-mo period during which the nest was not detectable (but could still be destroyed by treatment), and a simultaneous 8-mo maturation period before the nest can itself found other nests.

We drew 100 independent sets of parameters from the prior distribution, and for each such draw we generated an MCMC sample of 1,000 draws from the posterior distribution, after discarding 1,000 draws as burn-in. Each draw from the posterior distribution was selected uniformly at random from a sequence of 300,000 updates.

Evaluations of the z_θ statistic are shown in Fig. S1. The rows labeled $\mathrm{Q_X}$ and $\mathrm{Q_Y}$ each contain ten z_θ statistics for the 5%, 15%, . . ., 95% percentiles of the distribution of founding distances in the X and Y directions, evaluated using γ and σ_X for $\mathrm{Q_X}$ and using γ and σ_Y for $\mathrm{Q_Y}$. Note that the parameters in each row are correlated, especially the percentiles of founding distances. No extreme values are apparent, and thus we find no evidence of errors in the software.

SI Results

In this section we present additional graphical summaries of results referred to in the main manuscript as Figs. S2–S7. These were estimated using the same MCMC samples of size 10,000 used in the main manuscript.

SI Materials and Methods

Data. For each detected nest $i = 1, \ldots, N,$ the available data consists of a position (x_i, y_i), a date of discovery t_i and a type of discovery d_i positions consist of an x and y coordinate, where x coordinates are integers in the range 377,638–561,437 and y coordinates are integers in the range 6,843,963–7,035,887. Units are meters, so the rectangle considered is ~184 km by 192 km, somewhat larger than the area in which nests have been found. The current model discretises time into months, but it could be easily modified for a finer time scale. The month of discovery is an integer in the range 61–191, corresponding to the months February 2001 to December 2011. Note that the model assumes a nest is killed immediately after discovery, so that t_i may alternatively be referred to as the date of death for nest i. The type of discovery d_i is 0 for a nest reported by the public and 1 for a nest found by targeted search.

It is known which regions were subject to targeted search in each month. This is codified using an indicator function $I_1(x, y, t)$ taking the value 1 if the position (x, y) was searched in month t, and zero otherwise. It is also known which regions were treated in each month. This is similarly codified using an indicator function $I_2(x, y, t)$ taking the value 1 if the position (x, y) was treated in month t, and zero otherwise. It is possible for a region to be both treated and searched in the same month.

The land type and habitat suitability at each location are also known. The land type is encoded by a function $S(x, y)$ and affects the probability that a nest is discovered by passive detection at position (x, y). The function $S(x, y)$ returns the value 1, 2, 3, or 4, representing Major Urban, Other Urban, Defined Boundary, and Rural Balance, respectively. The habitat suitability is encoded by a function $\mathrm{H}(x, y)$ and affects the probability that a nest will become established at position (x, y). The function $H(x, y)$ returns the value 1, 2, 3, or 4 in order of decreasing suitability for ant nests. Although in principle the functions S and H are functions of time t, it is assumed in practice that the land type and habitat suitability maps did not change during the period modeled.

Unknown Parameters of Nests. It is assumed that an unknown number U of undetected nests existed at some time during the period modeled. Each undetected nest had a position (x_i, y_i), where $i = N + 1, \ldots, N + U$.

Each nest $i = 1, \ldots, N + U$ (that is, including both detected and undetected nests) is assumed to have some unknown founding time f_i before (that is, strictly less than) its time of death. Founding dates can be in the range 0–190, corresponding to the months January 1996 through November 2011. Note that no detected nest was allowed a founding time of 191, because the founding date is required to be strictly less than the date of death. For convenience, no undetected nests were allowed a founding date of 191 either, effectively ignoring any nests founded in the final month of data collection. A founding time in the range 0–60 indicates that a nest was founded before the first nest discoveries. A founding date of −1 was also allowed, indicating that the nest was founded some time before January 1996; such nests are termed "initial nests." The number of months modeled before the first detection can easily be adjusted by the user, but in this study we allowed for roughly 5 y (or, more precisely, 61 mo) of undetected expansion. Note that in principle an undetected nest can also be an initial nest.

Undetected nests have an unknown time of death of t_i. This must correspond to a month in which the position (x_i, y_i) was treated unless the nest was still alive at the end of 2011, which is represented by setting $t_i = 9{,}999$. Natural deaths are not considered, so nests are effectively immortal unless either detected or killed by treatment. It is also convenient to define the type of discovery d_i for an undetected nest to be 2, regardless of whether it was destroyed by treatment or still living at the end of 2011.

The model also allows for nests that fail to "establish" after founding. Such nests are included so that the probability of a nest being founded at a particular location can depend on the suitability of the habitat (discussed in more detail in the following section). Nests that do not establish can be thought of as nests that died immediately after founding, without producing any other nests. The value e_i is defined for each nest to be 1 for a nest that established and 0 for a nest that did not establish. Detected nests automatically have e_i =1, whereas undetected nests may have $e_i = 0$ or $e_i = 1$, and thus e_i is another unknown.

All noninitial nests, detected and undetected, are assumed to have an unknown parent or founding nest p_i. Noninitial nests also have an unknown founding type J_i. It is possible to set the number of founding types to two, in which case $J_i = 0$ represents a local founding event and $J_i = 1$ represents a long-distance founding event (a "jump"). However, in this study we used a larger number of jump types to more accurately approximate the true distribution of jump distances. It was found convenient to allow four different jump types, corresponding to four different scales of founding event. The number of founding types can be changed by the user. More detail about the different jump types is given in the following sections.

Global Unknown Parameters. In addition to the unknown parameters for each nest, there are some global unknown parameters. These include the number U of undetected nests, already defined above. Another global unknown is the vector $\gamma = (\gamma_1, \gamma_2, \gamma_3, \gamma_4)$ of probabilities of the four jump types. Note these probabilities sum to 1.

For a noninitial nest i, the probability that its location is (x_i, y_i) depends on the location (x_{p_i},y_{p_i}) of its parent nest p_i. The distribution of the distance $| x_{p\_i} - x_i |$ is discussed in *Likelihood Model* below, but for now note that it depends on the jump type J_i and has unknown SD σ_{XJ_i}. Similarly, the distance $| y_{p\_i} - y_i |$ depends on J_i and has unknown SD σ_{YJ_i}.

The unknown probability that a specific nest will be killed if treated we refer to as the treatment efficacy and denote by α_0. Similarly, the unknown probability that a specific nest will be detected by targeted search we refer to as the targeted search efficacy, denoted α_1. The probability that the nest is detected by the public is α_2, α_3, α_4, or α_5 for land types $S(x, y)$ equal to 1, 2, 3, and 4, respectively. It is natural to constrain $1 \geq \alpha_2 \geq \alpha_3 \geq \alpha_4 \geq \alpha_5 \geq 0$. In this study, we also constrained $\alpha_0 \geq 0.8$ and $\alpha_1 \geq 0.8$, unless explicitly stated otherwise.

The unknown probability that a nest will become established at a given location is β_1, β_2, β_3, or β_4 for habitat types $H(x, y)$ equal to 1, 2, 3, and 4, respectively. Because it is not possible to infer the absolute values of the betas from the data provided, it is convenient to set $\beta_1 = 1$, so that β_2, β_3, and β_4 are relative establishment probabilities. It is natural to constrain $1 \geq \beta_2 \geq \beta_3 \geq \beta_4 \geq 0$.

Finally, the average number of nests founded per nest per month is an unknown parameter λ.

Likelihood Model. A conditional dependency graph showing all of the known and unknown parameters pertaining to a noninitial nest i is shown in Fig. 6 of the main text. The eight parameters for nest i are displayed inside the box labeled "Nest i." Six of the seven global unknowns are shown in a row at the bottom of the figure. (Note that the functions H, S, I_1, and I_2 are part of the data, and hence known.) The seventh global parameter is U, the number of unobserved nests at the end of the data collection period. At the top of the figure, some of the parameters of the parent nest p_i are shown, and the arrows indicate the parameters of nest i that depend on them.

Note that there are also a number of constraints on the allowed values that imply additional conditional dependencies. Specifically, a nest may not found another nest before or in the month it is itself founded, after or in the month it is discovered, or in any month if it is not established.

For an initial nest ($f_i = -1$), the much simpler conditional dependency graph is shown in Fig. S8.

Based on these figures, the likelihood can be written in the form

$$\begin{aligned} P\left(f,p,J,x,y,e,t,d|\lambda,\gamma,\sigma_X,\sigma_Y,\beta,\alpha,H,S'\right) = P(f,p|\lambda,t) \\ \times P(J|\gamma)P(x|p,J,\sigma_X)P(y|p,J,\sigma_Y) \\ \times P(e|f,x,y,\beta,H)P\left(t,d|f,x,y,\alpha,S'\right), \end{aligned}$$

where S' represents the triplet (S, I_1, I_2). Note that although this expression appears to have f conditional on t and t conditional on f, there is in reality no circular dependence, because the founding time for nest i depends on the time of discovery of its parent p_i, whereas the time of discovery for nest i depends on that nest's own founding time.

Each nest is assumed to take 8 mo to mature, during which time it is unable to found other nests. After maturation, the number of nests founded per nest per month is assumed to follow a Poisson distribution with parameter λ. The number of initial nests is assumed to have been selected uniformly at random between 0 and some large maximum value $N_{\max}$. Hence the vector of founding dates $f = (f_1, \ldots, f_{N+U})$ and the vector of parent nests $P = (p_1, \ldots, p_{N+U})$ have a joint distribution given by

$$P(f,p|\lambda,t) = \frac{1}{N_{\max}+1} \prod_{\{i:f_i \neq -1\}} \prod_{j=f_i+8}^{\min\{t_i=1,190\}} Pois\left(n_{ij}(f,p)|\lambda\right),$$

where $n_{ij}(f, p)$ is the number of nests founded by nest i in month j. This number can be obtained for each nest i by simply counting the nests with parent nest i and founding date j. Note this probability is instead set to 0 for impossible combinations of values, specifically if any noninitial nest i has a parent nest p_i that died in the month f_i or before, or was founded in the month t_i or after.

The vector of jump types $J = (J_1, \ldots, J_{N+U})$ has probability determined by the vector γ:

$$P(J|\gamma) = \prod_j \gamma_j^{m_j},$$

where m_j is the number of nests founded by jump type j. For each jump type j, the x_i and y_i coordinates are modeled independently and the distance of the nest from its founder along the x or y axis is assumed to be exponentially distributed with mean (and SD) σ_{Xj} and σ_{Yj}, respectively. Hence, for noninitial nests

$$P\left(x_i | x_{p_i}, J_i = j, \sigma_X\right) = \frac{1}{2\sigma_{Xj}} e^{-|x_i - x_{pi}|/\sigma_{Xj}},$$

with a similar expression for y_i. Note that the factor of one-half in this expression allows for both positive and negative values of $x_i - x_{pi}$. The probability of an initial nest i having x coordinate x_i is uniform over the range [438,000, 529,000] and similarly the probability of an initial nest i having y coordinate y_i is uniform over the range [6,920,000, 6,985,000]. This is a smaller range than is permitted for noninitial nests, but it is highly likely that the invasion was confined within this smaller range at the beginning of 1996.

Which nests are established is determined by the vector of establishment probabilities β as follows:

$$P(e | \beta, H, f) = \prod_{\{i: f_i \neq -1\}} \beta_{H(x_i y_i)}^{e_i} \left(1 - \beta_{H(x_i y_i)}\right)^{1-e_i}.$$

The probability of the search outcomes (discoveries and failures to discover) is given by one of the following expressions, depending on the value of d_i:

$$P\left(t_i, d_i = 0 | f_i, x_i, y_i, \alpha, S'\right) = \alpha_{1+S(x_i y_i)} \prod_{j = \max\{61, f_i + 6\}}^{t_i - 1} \left(1 - \alpha_{1+S(x_i, y_i)}\right)$$
$$\times (1 - \alpha_1)^{I_1(x_i, y_i, j)} \prod_{j = \max\{61, f_i + 1\}}^{t_i - 1} (1 - \alpha_0)^{I_2(x_i, y_i, j)}$$

$$P\left(t_i, d_i = 1 | f_i, x_i, y_i, \alpha, S'\right) = \alpha_1 \left(1 - \alpha_{1+S(x_i, y_i)}\right) \prod_{j = \max\{61, f_i + 6\}}^{t_i - 1}$$
$$\times \left(1 - \alpha_{1+S(x_i, y_i)}\right)(1 - \alpha_1)^{I_1(x_i, y_i, j)} \prod_{j = \max\{61, f_i + 1\}}^{t_i - 1} (1 - \alpha_0)^{I_2(x_i, y_i, j)}$$

$$P\left(t_i, d_i = 2 | f_i, x_i, y_i, \alpha, S'\right) = \alpha_0 \prod_{j = \max\{61, f_i + 6\}}^{t_i} \left(1 - \alpha_{1+S(x_i, y_i)}\right)$$
$$\times (1 - \alpha_1)^{I_1(x_i, y_i, j)} \prod_{j = \max\{61, f_i + 1\}}^{t_i - 1} (1 - \alpha_0)^{I_2(x_i, y_i, j)}$$

$$P\left(t_i = 9999, d_i = 2 | f_i, x_i, y_i, \alpha, S'\right) = \prod_{j = \max\{61, f_i + 6\}}^{t_i} \left(1 - \alpha_{1+S(x_i, y_i)}\right)$$
$$\times (1 - \alpha_1)^{I_1(x_i, y_i, j)} \prod_{j = \max\{61, f_i + 1\}}^{t_i} (1 - \alpha_0)^{I_2(x_i, y_i, j)}.$$

Note that nests are assumed to be undetectable until 6 mo after founding, hence searches performed during that period do not contribute to the likelihood. Treatments, however, are immediately effective. This probability is instead set to 0 for impossible combinations of values, for example, if an undetected nest has a time of death in a month that its position was not treated ($d_i = 2$ but $I_2(x_i, y_i, t_i) = 0$).

Note that these expressions imply that each month from 61 to 190 consisted of four distinct phases—public searches, targeted searches, treatments, and founding events—occurring in that order, without overlap. Only founding events occurred in the months 0–60, and only searches and treatments occurred in month 191. Note also that t_i and d_i are not defined for nonestablished nests.

Finally, it is convenient for sampling purposes to construct several lists of nests and include the probability of their specific ordering as part of the likelihood. A list of all undetected nests is maintained, and also separate lists of child nests founded by each individual nest in each individual month after maturity. All permutations within lists are considered equally likely, so the likelihood of a permutation of L elements is $(1/L!)$. However, a subtle point is that the above model already assumes positions and other distinguishing characteristics are assigned in a sequential manner to child nests founded by the same nest in the same month, and thus identifying a specific permutation of these lists contributes no additional term to the likelihood. This is not the case for the list of all undetected nests, which does therefore contribute an additional factor of $(1/N_u!)$ to the likelihood, where N_u is the number of undetected nests.

Prior Probabilities. To implement a Bayesian model, prior probabilities must be specified. These represent the state of knowledge about the unknown parameters of the model, before examining the data. Prior distributions are required for each of the global parameters γ, σ_X, σ_Y, α, β, and λ. In addition, for each initial nest ($f_i = -1$), prior probabilities are needed for the parameters x_i and y_i. The prior for the jump probabilities γ was a uniform Dirichlet prior, and the priors for the search parameters α and establishment parameters β were uniform on the interval [0,1]. The priors for the parameters σ_X and σ_Y were taken to be inverse gamma distributions with shape and scale parameters equal to 1. This is equivalent to setting priors for ${\sigma_X}^{-1}$, ${\sigma_Y}^{-1}$ to be gamma distributions with shape and scale equal to 1. The prior for λ was also taken to be a gamma distribution with shape and scale parameters λ_r and λ_s, respectively. The values of λ_r and λ_s were set to large values to constrain λ to be close to the prior mean λ_r/λ_s, which was adjusted to take values 0.15, 0.2, 0.25, or 0.3 (discussed in main text). However, for some runs λ was unconstrained, and for those cases we set $\lambda_r = \lambda_s = 1$. After applying Bayes' rule to obtain the posterior distribution, as discussed below, we integrated over λ. Because λ appears only in the prior and the Poisson distributions for each established nest and each month that nest is able to found, the posterior contains the following factor after integration:

$$\Gamma(n + \lambda_s)(z + \lambda_r)^{n+\lambda_s} \prod_{i=1}^{N+U} \prod_{k = f_i + 8}^{\min\{190, t_i - 1\}} \frac{1}{n_{ik}!},$$

where n is the number of noninitial nests, Γ represents the gamma function, n_{ik} is the number of nests founded by nest i in month k, and z is the total number of months in which nests can be founded, summed over established nests. That is,

$$z = \sum_{i=1}^{N+U} e_i [\min\{190, t_i - 1\} - (f_i + 8) + 1].$$

For the initial nests, the prior probabilities for the position coordinates x_i and y_i were taken to be uniform over the ranges specified above.

MCMC Sampling. Once the prior and likelihood models are specified, Bayes' rule can be invoked to obtain a posterior probability distribution for the unknown parameters of the model. This posterior distribution associates probabilities with all possible values of the unknowns, including the number and locations of undiscovered nests. Standard Bayesian procedure is then to use MCMC to draw a sample from the posterior distribution and use this sample to construct marginal distributions for each of the parameters. Here we used an MCMC technique known as the generalized Gibbs sampler (2). The sampler iteratively cycled through the following updates. (*i*) For each nest, update the phylogeny using subtree pruning and regrafting moves (similar to those described in ref. 3) using that nest as the root of the subtree to be moved. (*ii*) For each nest, consider deleting or inserting a new undetected child nest. (*iii*) Update all σ terms. (*iv*) Update all α terms. (*v*) Update all β terms. (*vi*) Update all γ terms. (*vii*) For each nest, consider swapping its position in the phylogeny with that of its parent. (*viii*) Permute the order of the list of undetected nests (which determines the order in which nests are considered for updates *i*, *ii*, and *vii*). The first two updates are described in detail in the main text; the rest are described below.

For each of the scenarios described in the main text, the sampler was used to generate 20,000 samples, each of which consisted of a complete set of values ascribed to all unknowns. Each of these samples was drawn at random from a sampling block of 100,000 individual updates. Convergence was assessed graphically by plotting the log-likelihood and the global parameters N, σ, α, β, γ, and λ as time series. Convergence was judged to have occurred within the first 10,000 samples in all cases; these were discarded as burn-in, leaving 10,000 for estimating marginal posterior probabilities.

Move Types. In the main text, two move types that are not conventional Gibbs updates are described: founder updates and insertion/deletion moves. These move types are available for all $N + U$ nests. After these, the following conventional Gibbs updates are performed.

Updating each σ_{Xj} is straightforward, and involves drawing a new value from the posterior conditional distribution for σ_{Xj}, which takes the form of an inverse gamma distribution:

$$\sigma_{Xj}|x,J,p \sim \mathrm{Inv}-\Gamma\left(1+n_j,1+z_j\right),$$

where $1+n_j$ is the shape parameter and $1+z_j$ the scale parameter. The value of n_j is the number of nests with founding type j. The value of z_j is

$$z_j=\sum_{\{i:J_i=j\}}|x_i-x_{p_i}|.$$

The conditional distribution for σ_{Yj} is similarly defined.

Each α_k is updated by drawing from the posterior conditional distribution for α_k (k = 0, 1, 2, 3, 4, 5) given by

$$p\left(\alpha_k|e,t,d,f,x,y,S'\right)\propto\alpha_k^{m_k+1}(1-\alpha_k)^{n_k+1}.$$

The value m_k for k = 2, 3, 4, 5 is the number of nests found by public searches in land type k. Similarly, m_1 is the number of nests found by targeted search and m_0 is the number of nests killed by treatment. The value n_k for k = 2, 3, 4, 5 is the total number of failed public searches in land type k, whereas n_1 is the number of failed targeted searches and n_0 is the number of failed treatments. Sampling from these distributions could be accomplished by drawing from a beta distribution if all values of α_k in the range [0,1] were allowed. However, we apply the constraints $\alpha_2 \geq \alpha_3 \geq \alpha_4 \geq \alpha_5$. For some of the scenarios described in the main text, α_0 and α_1 are allowed any value in [0,1], but for other scenarios we constrain α_0, $\alpha_1 \geq 0.8$ or α_0, $\alpha_1 \geq 0.85$. These constrained distributions for α_k are sampled using the slice sampler (4).

Each β_k is updated by drawing from the posterior conditional distribution for β_k (k = 2, 3, 4) given by

$$p(\beta_k|e,f,x,y,H)\propto\beta_k^{m_k+1}(1-\beta_k)^{n_k+1},\ \beta_{k+1}\leq\beta_k\leq\beta_{k-1},$$

where to simplify the expression of the inequality we define $\beta_1 = 1$ and $\beta_5 = 0$. The value m_k is redefined in this expression as the number of nests established in habitat type k. Similarly, the value n_k is redefined as the number of nests that failed to establish in habitat type k. Recall that all nests founded in habitat type 1 are automatically established. Again this conditional distribution for β_k is sampled using the slice sampler (4).

The founding type probabilities γ are updated by drawing from their posterior conditional distribution, which takes the form of a Dirichlet distribution:

$$\gamma|J\sim\mathrm{Dirichlet}(1+n_1,1+n_2,1+n_3,1+n_4),$$

where n_j is again redefined as the number of nests with founding type j.

Another move type that was implemented for all $N+U$ nests is a parent/child swap. This is a conventional Metropolis–Hastings-style update. The sampler scans through the ordered list of nests and for each nest i it considers swapping the children of nest i with the children of its parent p_i. The founding dates and founding types of nests i and p_i are also swapped, i becomes the parent of p_i, and the parent of p_i, if there is one, becomes the parent of i. For initial nests, no change is made and the swap is automatically considered a failure. Automatic failures also occur if nest i is not established, or if the swap would violate any natural constraints such as that a nest can found other nests only during its mature lifetime. Rejected swaps are also considered failures. After each failure, the next nest on the list after i is considered for swapping. However, after a successful swap, the next nest on the list after p_i is considered for swapping. This continues until a failed swap occurs for the last nest on the list.

The probability of accepting a swap depends on the ratio of

$$e^{-\left(|x_i-x_{g_i}|/\sigma_{XJ_{pi}}+|y_i-y_{g_i}|/\sigma_{YJ_{pi}}\right)}\prod_{\{m:p_m=i\}}e^{-\left(|x_m-x_{p_i}|/\sigma_{XJ_m}+|y_m-y_{p_i}|/\sigma_{YJ_m}\right)}$$
$$\times\prod_{\{m:p_m=p_i,p_m\neq i\}}e^{-\left(|x_m-x_i|/\sigma_{XJ_m}+|y_m-y_i|/\sigma_{YJ_m}\right)}\prod_{j=\max\left\{61,f_{p_i}+6\right\}}^{\min\{191,t_i,f_i+5\}}\left(1-\alpha_{1+S(x_i,y_i)}\right)$$
$$\times(1-\alpha_1)^{I_1(x_i,y_i,j)}\prod_{j=\max\left\{61,f_{p_i}+1\right\}}^{\min\{191,t_i-1,f_i\}}(1-\alpha_0)^{I_2(x_i,y_i,j)}(z_{\mathrm{new}}+\lambda_r)^{-(n+\lambda_s)}$$

to

$$e^{-\left(|x_{pi}-x_{g_i}|/\sigma_{XJ_{pi}}+|y_{pi}-y_{g_i}|/\sigma_{YJ_{pi}}\right)}\prod_{\{m:p_m=i\}}e^{-\left(|x_m-x_i|/\sigma_{XJ_m}+|y_m-y_i|/\sigma_{YJ_m}\right)}$$
$$\times\prod_{\{m:p_m=p_i,p_m\neq i\}}e^{-\left(|x_m-x_{pi}|/\sigma_{XJ_m}+|y_m-y_{pi}|/\sigma_{YJ_m}\right)}\prod_{j=\max\left\{61,f_{pi}+6\right\}}^{\min\left\{191,t_{pi},f_i+5\right\}}\left(1-\alpha_{1+S(x_{pi},y_{pi})}\right)$$
$$\times(1-\alpha_1)^{I_1(x_{pi},y_{pi},j)}\prod_{j=\max\left\{61,f_{pi}+1\right\}}^{\min\left\{191,t_{pi}-1,f_i\right\}}(1-\alpha_0)^{I_2(x_{pi},y_{pi},j)}(z_{\mathrm{old}}+\lambda_r)^{-(n+\lambda_s)}$$

Here g_i represents the grandparent of nest i, that is, the parent of p_i. However, if nest p_i is an initial nest and has no parent, then the first factor (the only factor containing g_i) in the above expressions is removed. The terms z_{old} and z_{new} represent the number of months after maturation summed over all nests, before and after the swap, respectively. Note that the number of months after maturation only changes for nests i and p_i.

At the end of each cycle through move types, the ordered list of undetected nests is permuted, with all permutations being equally likely. In other words, the conditional distribution over the space of permutations is uniform.

Run Times

The computational effort required to produce these results comprised the following analyses, performed in parallel on separate central processing units:

i) Two analyses of the simulated dataset, with and without the additional constraints on search and treatment efficacies and reproductive rate,

ii) Four analyses of the complete dataset collected 2001–2011 using the additional constraints with different values of λ, plus a fifth without the additional constraints,

iii) Five independent analyses of the 2001–2008 dataset with different random starts, plus a sixth with increased lower bounds on search and treatment efficacies.

Exact run times were not recorded, but each analysis required ~2–3 wk of continuous central processing unit time. However, the code has not been optimized, and significant improvements in efficiency may still be possible.

1. Cook SR, Gelman A, Rubin DB (2006) Validation of software for Bayesian models using posterior quantiles. *J Comput Graph Statist* 6(3):675–692.
2. Keith JM, Kroese DP, Bryant D (2004) A generalized Markov sampler. *Methodol Comput Appl Probab* 6(1):29–53.
3. Keith JM, Adams P, Ragan MA, Bryant D (2005) Sampling phylogenetic tree space with the generalized Gibbs sampler. *Mol Phylogenet Evol* 34(3):459–468.
4. Neal RM (2003) Slice sampling. *Ann Stat* 31(3):705–767.

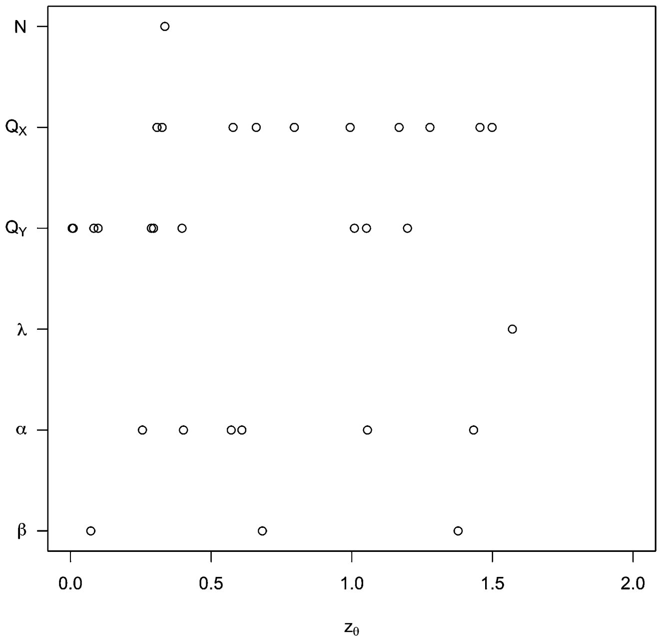

Fig. S1. Evaluations of statistic z_0 for the total number of nests *N*; the 5%, 15%, . . ., 95% percentiles of the distribution of founding distances in the *x* direction (Q_X) and *y* direction (Q_y); reproductive rate λ; treatment and search efficacies α_0 to α_5; and relative establishment probabilities β_2 to β_4. All statistics are based on 100 simulated datasets.

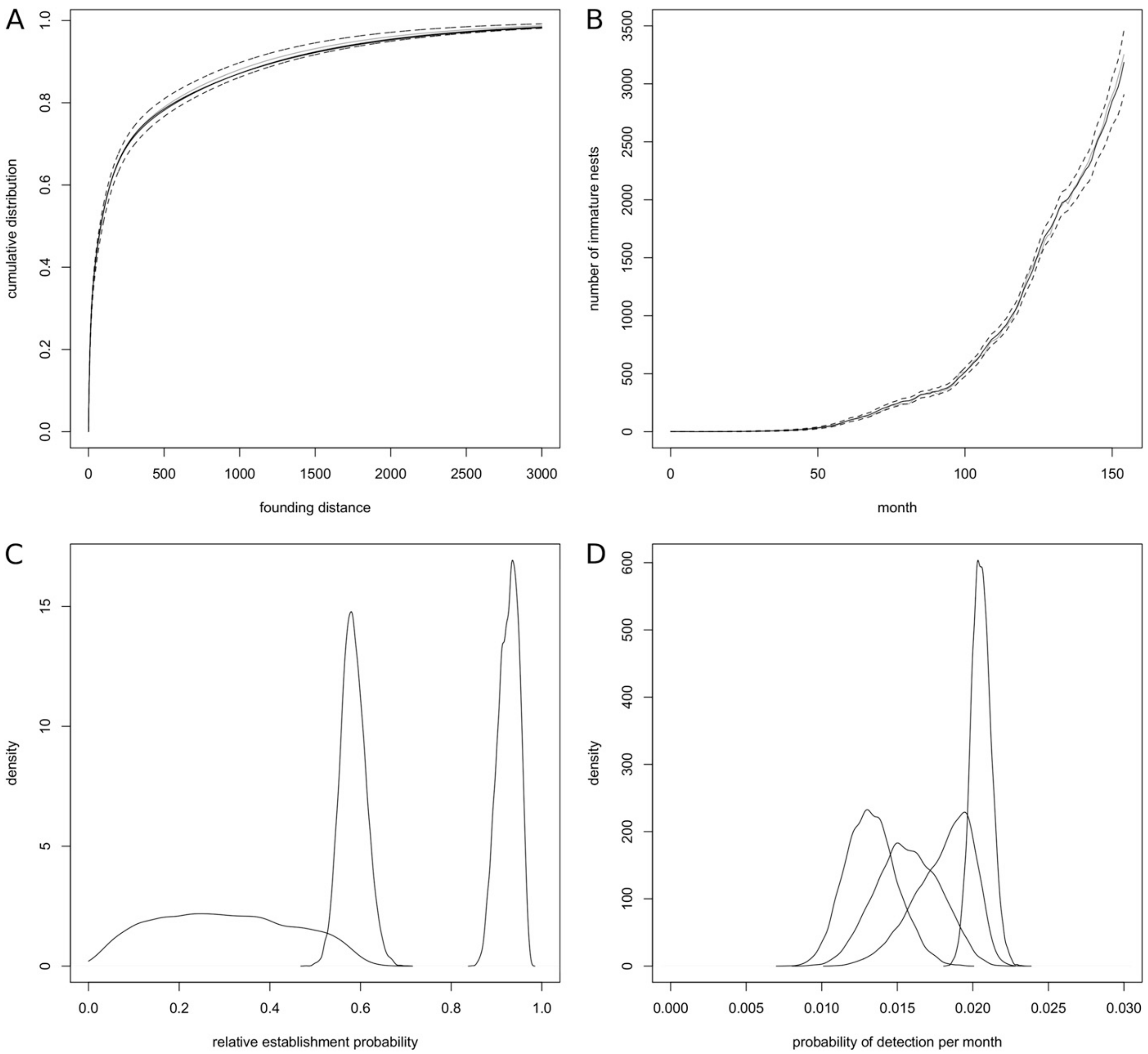

Fig. S2. (*A*) The true distribution of founding distances for the simulated dataset (gray) and posterior median inferred distributions of founding distances in the *x* and *y* directions (solid black lines, almost indistinguishable) with 95% credible intervals for the *x* direction (dashed black). (*B*) The number of immature nests existing in each month of a simulated invasion (light gray, solid line) and the posterior median number of immature nests estimated with additional constraints that α_0 and α_1 are both greater than or equal to 0.8 and $\lambda \approx 0.25$ (black, solid line) with 95% credible intervals (black, dashed lines). The posterior median number of immature nests estimated without these constraints is indistinguishable from the estimate made using them. (*C*) Posterior marginal densities of the relative establishment probabilities β_4, β_3, and β_2, in that order from left to right, inferred from simulated data. The true values used in simulation were 0.3, 0.6, and 0.9, respectively. (*D*) Posterior marginal densities of the probabilities of passive detection per month α_5, α_4, α_3, and α_2, in that order from left to right, inferred from simulated data. The leftmost density is for rural areas, and the rightmost for urban. The true values used in simulation were 0.01, 0.013, 0.016, and 0.02, respectively.

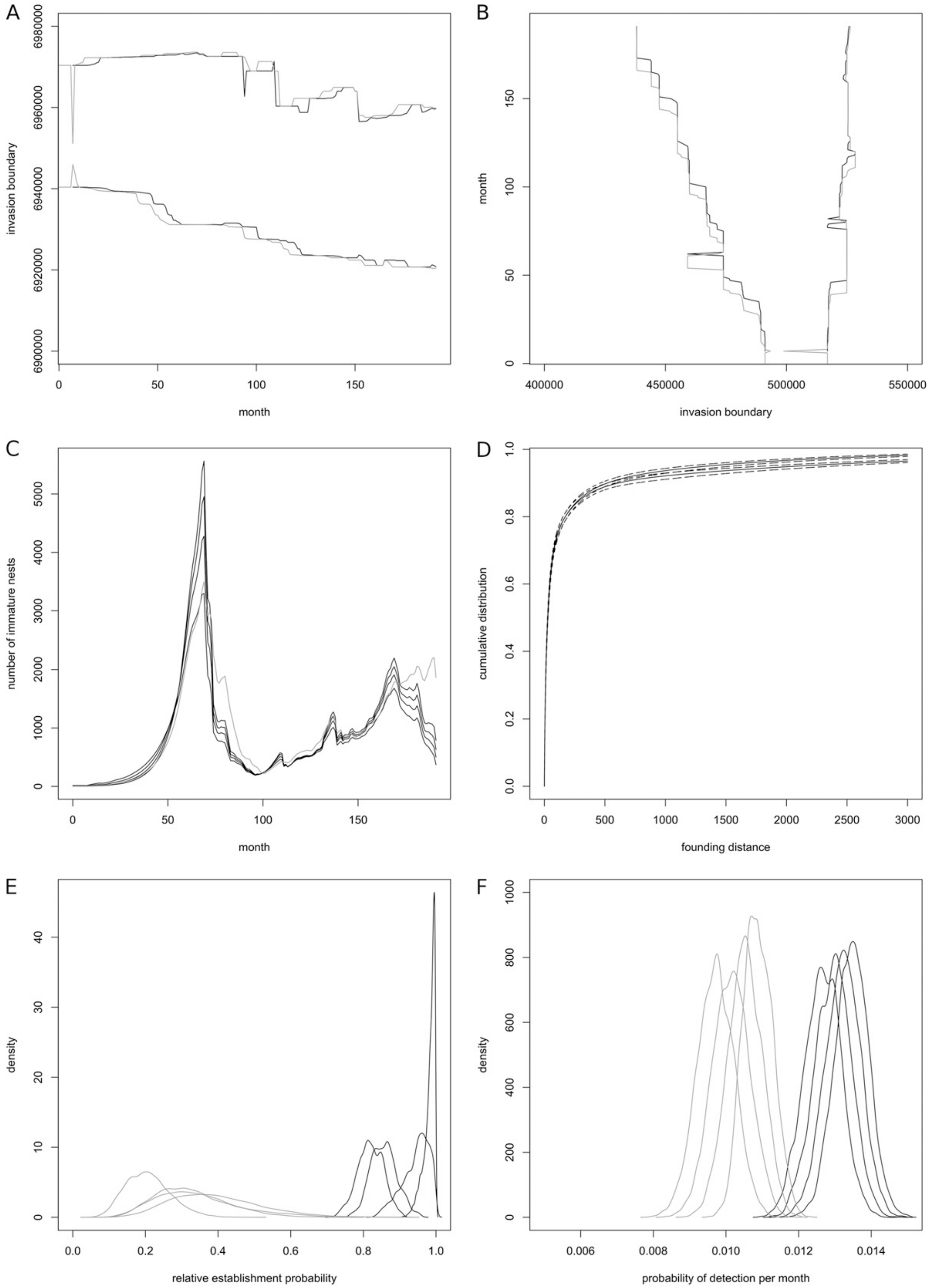

Fig. S3. (*A*) The posterior median northernmost and southernmost *y* coordinates of all mature nests (black lines) and immature nests (gray lines) in each month from January 1996 to December 2011 estimated using additional constraints that α_0 and α_1 are both greater than or equal to 0.8 and $\lambda \approx 0.25$. Corresponding inferences with different values of λ or without these constraints are almost identical. (*B*) The corresponding trajectories of the westernmost and easternmost *x* coordinates. (*C*) The estimated number of immature nests existing in each month from January 1996 to December 2011, using additional constraints that α_0 and α_1 are both greater than or equal to 0.8 and $\lambda \approx 0.15$, 0.2, 0.25, and 0.3 (four black lines, bottom to top, respectively) and corresponding estimates without the additional constraints (gray line). (*D*) Posterior median inferred distributions of founding distances in the *x* (lower solid line) and *y* (upper solid line) directions with

Legend continued on following page

95% credible intervals (dashed lines), using additional constraints that α_0 and α_1 are both greater than or equal to 0.8 and $\lambda \approx 0.25$. Corresponding inferences with different values of λ or without such constraints are almost identical. (*E*) Posterior marginal densities of relative establishment probabilities β_4 using additional constraints that α_0 and α_1 are both greater than or equal to 0.8 and $\lambda \approx 0.15$, 0.2, 0.25, and 0.3 (four gray lines, right to left, respectively) and corresponding densities for β_3 (four black lines, right to left, respectively). (*F*) Posterior marginal densities of the probabilities of passive detection per month α_5, α_4, α_3, and α_2, in that order from left to right, using additional constraints that α_0 and α_1 are both greater than or equal to 0.8 and $\lambda \approx 0.3$ (four black lines). The leftmost density is for rural areas, and the rightmost for urban. Also shown are the corresponding marginal densities without the additional constraints (four gray lines).

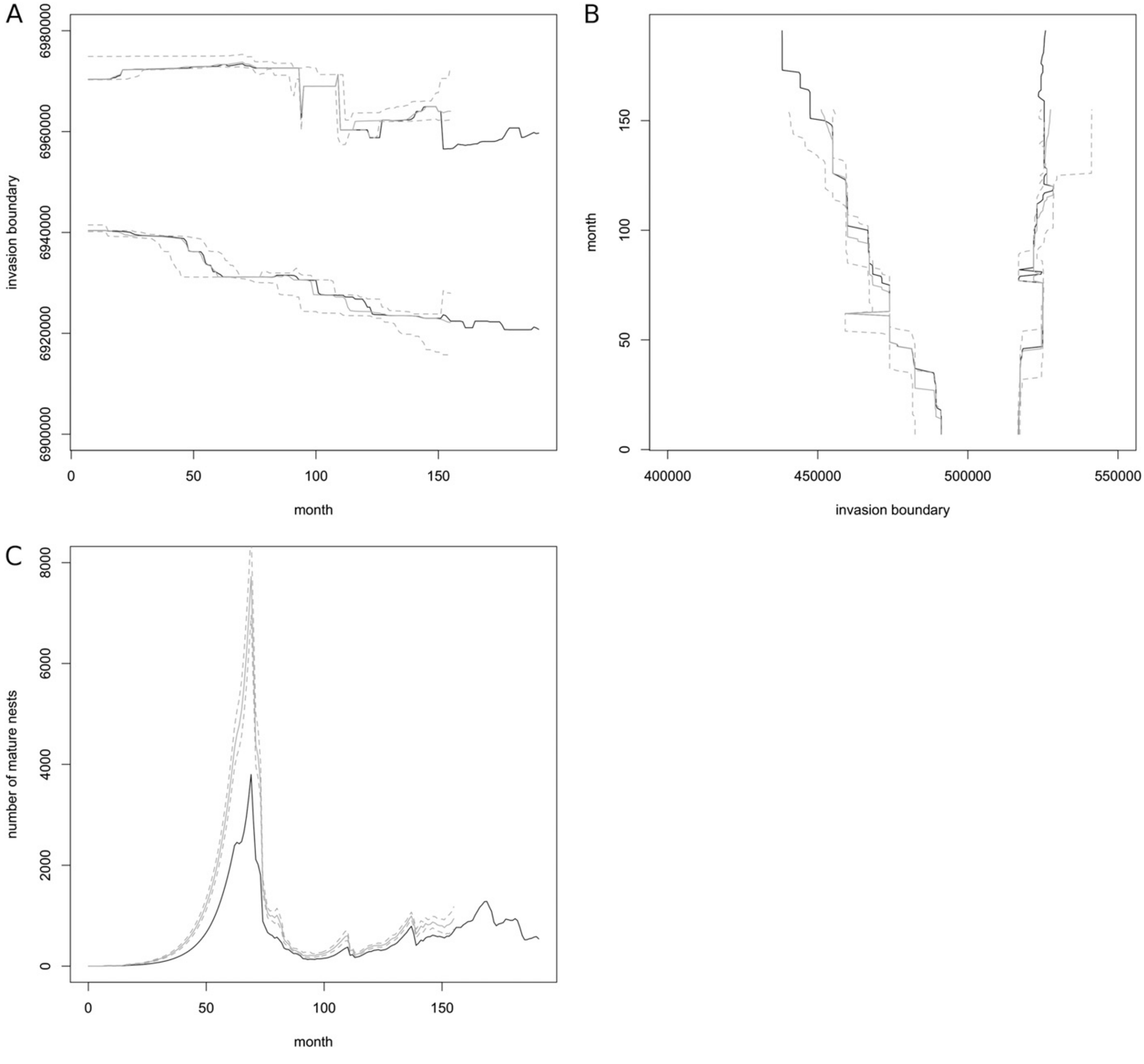

Fig. S4. (*A*) The posterior median northernmost and southernmost *y* coordinates of all mature nests (black lines) in each month from January 1996 to December 2011 estimated using additional constraints that α_0 and α_1 are both greater than or equal to 0.8 and $\lambda \approx 0.25$ and corresponding estimates for January 1996 to December 2008 using only the data available in December 2008 (gray, solid lines) with 95% credible intervals (gray, dashed lines). (*B*) The corresponding trajectories of the westernmost and easternmost *x* coordinates. (*C*) The posterior median number of mature nests existing in each month from January 1996 to December 2011, estimated using additional constraints that α_0 and α_1 are both greater than or equal to 0.8 and $\lambda \approx 0.25$ (black, solid line) with 95% credible intervals (black, dashed lines) and corresponding estimates using only the data available in December 2008 (gray lines, solid for median, dashed for 95% credible intervals).

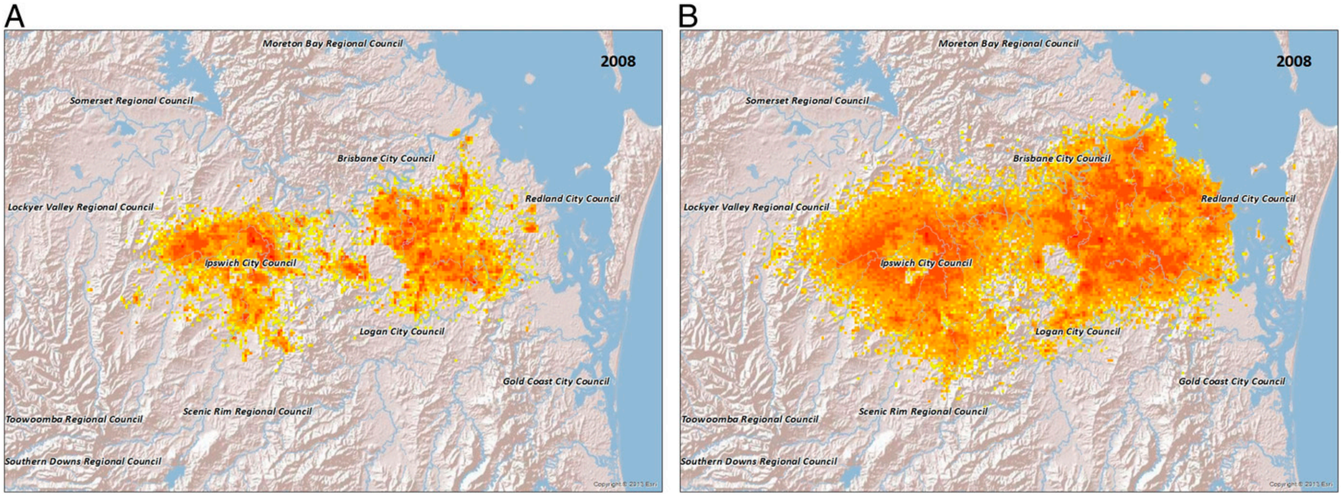

Fig. S5. Heat maps for the posterior expected number of nests in grid cells 500 m by 500 m in 2008 based on all of the data collected 2001–2011 (*A*) and on only the data from 2001 to 2008 (*B*). Brighter cells (yellows) have lower expected numbers of nests and darker cells (reds) have higher expected numbers. Color classes are on a logarithmic scale.

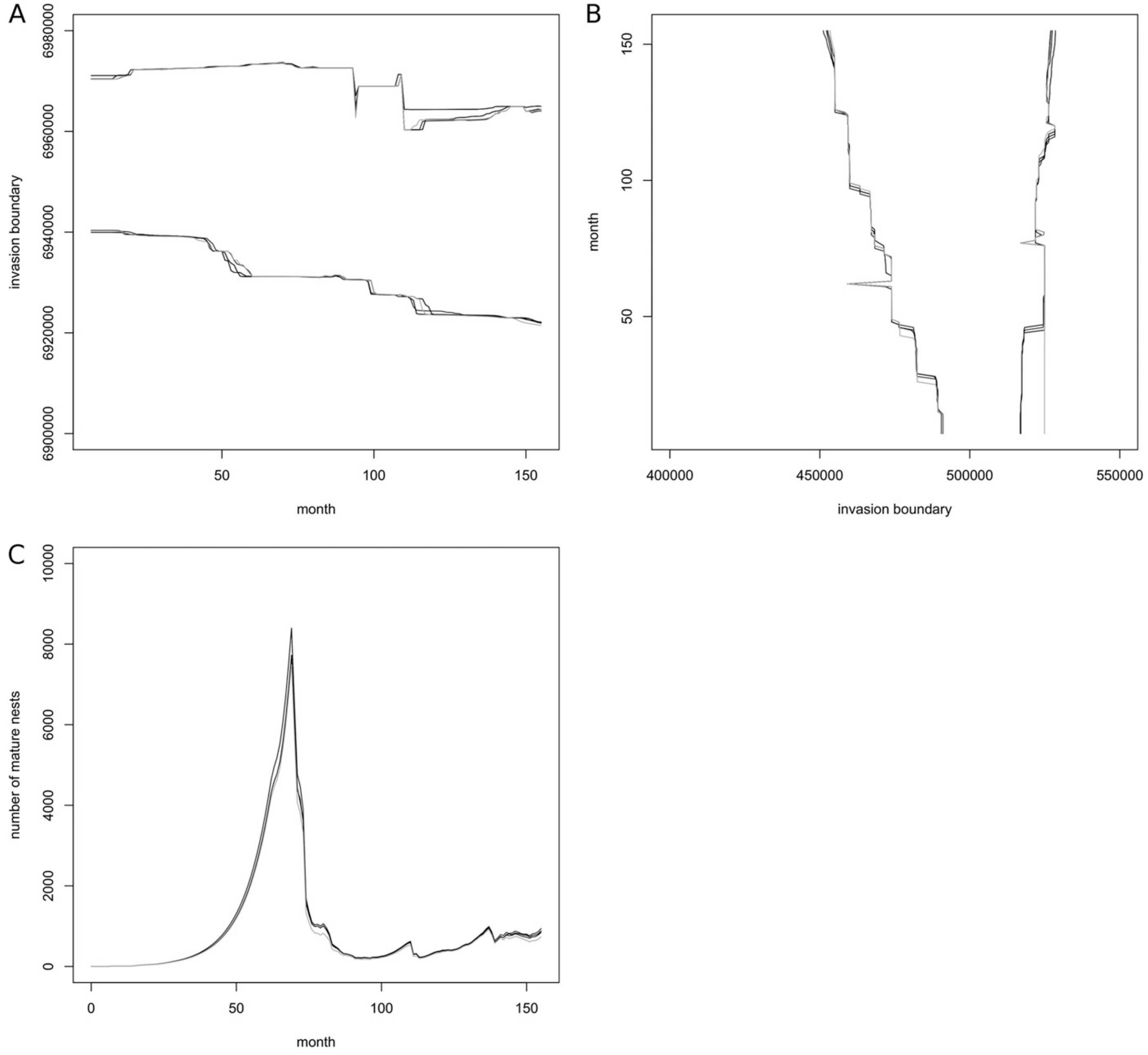

Fig. S6. (*A*) The posterior median northernmost and southernmost *y* coordinates of all mature nests in each month from January 1996 to December 2008 estimated using additional constraints that α_0 and α_1 are both greater than or equal to 0.8 and $\lambda \approx 0.25$, for five different random starts (five black lines) and corresponding estimates using an alternative constraint that α_0 and α_1 are both greater than or equal to 0.85 (gray line). (*B*) The corresponding trajectories of the westernmost and easternmost *x* coordinates. (*C*) The posterior median number of mature nests existing in each month from January 1996 to December 2008, estimated using additional constraints that α_0 and α_1 are both greater than or equal to 0.8 and $\lambda \approx 0.25$, for five different random starts (five black lines) and corresponding estimates using an alternative constraint that α_0 and α_1 are both greater than or equal to 0.85 (gray line).

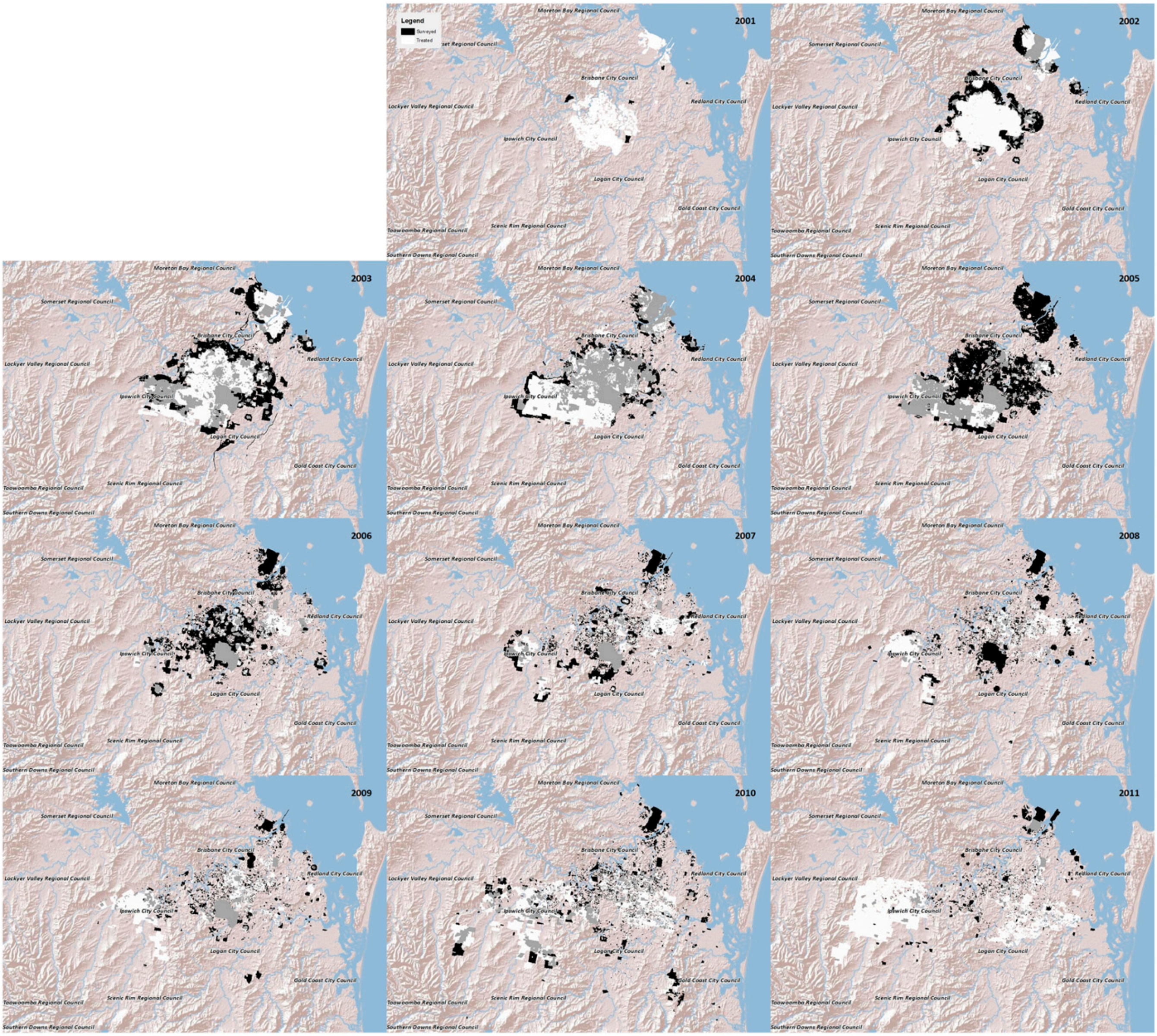

Fig. S7. Searched (black) and treated (white) areas 2001–2011. Areas that were both searched and treated are gray.

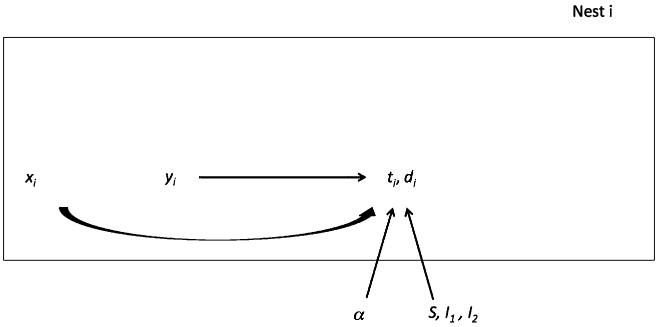

Fig. S8. Model parameters associated with initial nests.