

RESEARCH ARTICLE

Inference for epidemic models with time-varying infection rates: Tracking the dynamics of oak processionary moth in the UK

Laura E. Wadkin¹  | Julia Branson²  | Andrew Hoppit³ | Nicholas G. Parker¹  |
Andrew Golightly^{1,4}  | Andrew W. Baggaley¹ 

¹School of Mathematics, Statistics and Physics, Newcastle University, Newcastle upon Tyne, UK

²GeoData, Geography and Environmental Science, University of Southampton, Southampton, UK

³Forestry Commission England, Nobel House, London, UK

⁴Department of Mathematical Sciences, Durham University, Durham, UK

Correspondence

Laura E. Wadkin, School of Mathematics, Statistics and Physics, Newcastle University, Newcastle upon Tyne NE1 7RU, UK.
Email: laura.wadkin@ncl.ac.uk

Funding information

Engineering and Physical Sciences Research Council, Grant/Award Number: EP/V048511/1

Abstract

1. Invasive pests pose a great threat to forest, woodland, and urban tree ecosystems. The oak processionary moth (OPM) is a destructive pest of oak trees, first reported in the UK in 2006. Despite great efforts to contain the outbreak within the original infested area of South-East England, OPM continues to spread.
2. Here, we analyze data consisting of the numbers of OPM nests removed each year from two parks in London between 2013 and 2020. Using a state-of-the-art Bayesian inference scheme, we estimate the parameters for a stochastic compartmental SIR (susceptible, infested, and removed) model with a time-varying infestation rate to describe the spread of OPM.
3. We find that the infestation rate and subsequent basic reproduction number have remained constant since 2013 (with R_0 between one and two). This shows further controls must be taken to reduce R_0 below one and stop the advance of OPM into other areas of England.
4. *Synthesis.* Our findings demonstrate the applicability of the SIR model to describing OPM spread and show that further controls are needed to reduce the infestation rate. The proposed statistical methodology is a powerful tool to explore the nature of a time-varying infestation rate, applicable to other partially observed time series epidemic data.

KEYWORDS

Bayesian inference, epidemics, Markov chain Monte Carlo, oak processionary moth, SIR model, stochastic differential equation, susceptible-infected-removed model

TAXONOMY CLASSIFICATION

Invasion ecology

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Ecology and Evolution* published by John Wiley & Sons Ltd.

1 | INTRODUCTION

Invasive pests, such as non-native insects, pose a threat to forest, woodland, and urban tree ecosystems by damaging and killing trees and reducing biodiversity (Freer-Smith & Webber, 2017; Kenis et al., 2009; Manchester & Bullock, 2000). This threat has increased in recent years due to growth in international travel and trade (Roy et al., 2014) coupled with a changing climate driving the migration of species into new ecosystems (Lenzner et al., 2020). The loss of biodiversity has a profound economic impact, through short- to long-term control measures and the impact on ecosystem services (Aukema et al., 2011; Boyd et al., 2013; Bradshaw et al., 2016).

The oak processionary moth (OPM), *Thaumetopoea processionea*, is an invasive and destructive pest of oak trees, causing defoliation and making trees vulnerable to other stressors and pathogens. The larvae of OPM have poisonous hairs, containing an urticating toxin (thaumetopoein) which is harmful to human and animal health (Gottschling & Meyer, 2006; Maier et al., 2003, 2004; Rahlenbeck & Utikal, 2015).

OPM was introduced to the UK through accidental imports on live oak plants, first reported in 2006. Up to 2010, the governmental policy was one of eradication (Mindlin et al., 2012; Tomlinson et al., 2015). However, in 2011 it was decided that OPM was fully established in the South-East England area and so the government moved to a containment strategy, aiming to contain the OPM infestations within this original outbreak area (Tomlinson et al., 2015). In 2018, legislation was introduced to curb continuing imports through the Plant Health Order (Plant Health (England) (Amendment) (no. 3) Order, 2018). Despite the containment strategies, the extent of OPM has continued to spread with recent analysis suggesting an expansion rate of 1.7 km/year for 2006–2014, with an increase to 6 km/year from 2015 onwards (Suprunenko et al., 2021). The regions surrounding the current infestation area are particularly climatically suitable (Godefroid et al., 2020), and so being able to predict and control the future dynamics of the OPM population is crucial to protect these areas.

Mathematical models provide a powerful tool for describing and predicting the spread of tree disease and pest infestation (Gertsev & Gertseva, 2004; Orozco-Fuentes et al., 2019; Wang & Song, 2008). For OPM, previous work has included using models from electric network theory to predict high-risk regions (Cowley et al., 2015) along with species distribution models to examine the spatial distributions of OPM (Scholtens, 2021) and the effects of climate change on its expansion (Godefroid et al., 2020). Bayesian inference can be used to inform and evaluate these ecological mathematical models (Ellison, 2004). Previously, Bayesian approaches have been used to estimate key parameters in the spatio-temporal invasion of alien species (Cook et al., 2007); however, the techniques have yet to be applied to data for the spread of OPM. Nevertheless, the Bayesian paradigm provides a natural mechanism for quantifying and propagating uncertainty in the model parameters and dynamic components. Consequently, Bayesian inference techniques have been ubiquitously applied in the broad area of epidemiology (see

e.g., Fuchs, 2013; Kypraios et al., 2017; McKinley et al., 2014 for an overview).

In this paper, we use data tracking the numbers and locations of OPM nests removed from oak trees as part of a control program in two parks in London. We illustrate the use of statistical inference techniques for estimating the parameters for a classic SIR compartmental model (Bartlett, 1949; Kermack & McKendrick, 1927) consisting of susceptible, infested, and removed states. To allow for intrinsic stochasticity in the spread of OPM, we use an Itô stochastic differential equation (Oksendal, 1995) representation of the SIR model. This is further modified via the introduction of a time-varying infestation rate, as is necessary to capture the effect of unknown influences such as preventative measures (Dureau et al., 2013). Bayesian inference for the resulting model is complicated by the intractability of the observed data likelihood, and subsequently, the joint posterior distribution of the key quantities of interest (model parameters and dynamic components). We overcome these difficulties via a linear Gaussian approximation of the stochastic SIR model, coupled with a Markov chain Monte Carlo scheme (Fearnhead et al., 2014) for generating samples from the joint posterior. These methods are outlined in Section 2 and detailed in Appendix S1, Sections S1 and S2, for use as a toolbox to apply to other ecological datasets. We use the parameters from the compartmental model to estimate a yearly R_0 measure for OPM, analogous to the basic reproduction number for a pathogen (Heesterbeek & Dietz, 1996), and estimate the OPM population in 2021.

2 | METHODS

In this section, we present the observational time series data with a summary of the data collection methods (Section 2.1), the details of the stochastic SIR model (Section 2.2), and an outline of our statistical inference methods (Section 2.3). Further statistical details including the relevant algorithms are set out in Appendix S1 (Sections S1 and S2).

2.1 | Data

The data in this paper are from Richmond and Bushy Park, collected and processed by The Royal Parks. This is shared with the Forestry Commission on an annual basis to inform the national Oak Processionary Moth Programme (Contingency Plan, 2021). The University of Southampton (GeoData) provide analysis, support, and hold the program data on behalf of the Forestry Commission.

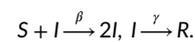
The data used in this study were obtained through the recording of OPM nest removals in Richmond and Bushy Parks in South-West London. For each of the years 2013–2020, it contains (i) the eastings and northings of trees which had nests removed and (ii) the number of nests removed from each tree. The dataset consists of 8470 unique tree locations, with 1767 in Bushy Park and 6703 in Richmond Park. The locations of the trees which had

nests removed are shown across the two parks in [Figure 1](#). There are no recordings of the locations of trees, which did not have any nests removed.

The raw and cumulative time series of the numbers of removed nests are shown in [Figure 2\(a,b\)](#). We count each tree in the year it first had nests removed as one “removal” in the SIR model (see [Section 2.2](#)), regardless of how many nests were recorded as removed from this location. The raw and cumulative time series for the number of unique trees which had nests removed are shown in [Figure 2\(c,d\)](#). We use the latter cumulative time series, $R(t)$, as our observational data in the following sections.

2.2 | Stochastic SIR model

We consider a SIR model ([Andersson & Britton, 2000](#); [Kermack & McKendrick, 1927](#)) in which a population of trees of fixed size N is



classified into compartments consisting of susceptible (S), infested (I), and removed (R) trees. Although most commonly used in epidemiology, SIR models have previously been used to describe the spread of tree diseases ([Parry et al., 2014](#); [Rodriguez-Quinones & Gordillo, 2019](#)) and invasive species through varying landscapes ([Ferrari et al., 2014](#); [Wildemeersch et al., 2019](#)). In our case, susceptible trees are those that have yet to ever be infested with OPM nests and are at risk from the currently infested trees. The time series data we use (see [Section 2.1](#)) are observations of the removed category, trees that have previously been infested with OPM nests and have now had these nests removed. A fixed population of trees is appropriate as over the timescale of interest the number of trees born into the S compartment will be sufficiently small to be negligible. Transitions between compartments can be summarized via two pseudo-reactions of the form

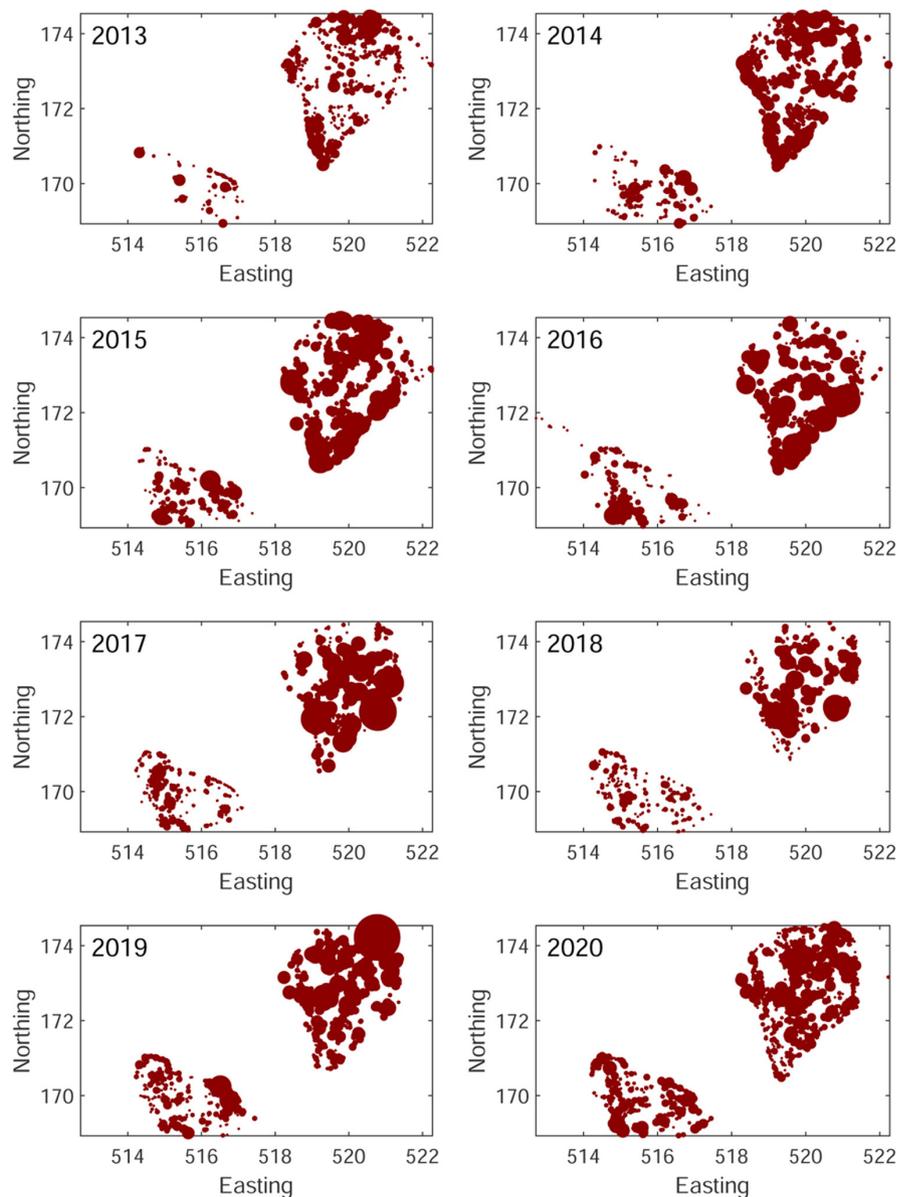


FIGURE 1 Map of the nests removed from Bushy (bottom left) and Richmond (top right) parks between 2013 and 2020. The area of the marker is proportional to the number of nests removed

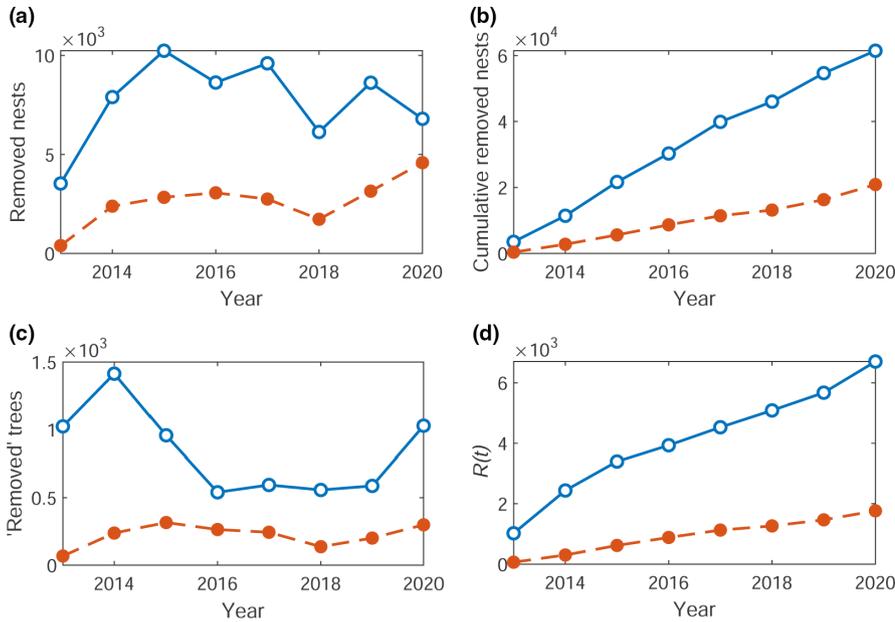


FIGURE 2 (a) Raw and (b) cumulative number of OPM nests removed from Richmond (blue) and Bushy (orange dashed) parks between 2013 and 2021. The number of (c) raw and (d) cumulative unique trees (described by their eastings and northings) which had nests removed between 2013 and 2021. The cumulative number of trees is the time series $R(t)$ corresponding to the “removed” category in the SIR model (see Section 2.2)

The first transition describes the “contact” of a currently infested tree with a susceptible tree (in this case, since trees have fixed locations, a probabilistic opportunity for the S to I transition to occur through the dispersal of OPM) and with the net effect resulting in one additional infested tree and one fewer susceptible. The second transition accounts for the currently infested trees moving into the R category as their nests are removed. The parameters β and γ govern the rate of infestation and removal, respectively. The infestation rate for the whole population (assuming one initial infested tree) of trees is thus βN to scale for the number of possible “contacts”. It is clear that transitions should result in discrete changes to the numbers of trees in each state. This most naturally leads to a continuous time, discrete valued Markov jump process (MJP) description of disease dynamics, as detailed in Appendix S1, Section S1 (Ho et al., 2018). We eschew the MJP formalism in favor of a continuous-valued approximation, formulated as a stochastic differential equation (SDE). This is a pragmatic choice, since the SDE model ultimately leads to a computationally efficient inference scheme, and the model can be easily augmented with additional components, such as time-varying parameters, which we now describe.

The SDE representation of the standard SIR model can be derived directly from the MJP (see Appendix S1, Section S1). Here, we extend this to include a time-varying infestation process. Let $X_t = (S_t, I_t, \tilde{\beta}_t)'$ where S_t and I_t denote the numbers in each of the states S and I at time $t \geq 0$ and $\tilde{\beta}_t = \log \beta_t$ is the (transformed) time-varying infestation process. Note that the fixed population size gives $R_t = N - S_t - I_t$ for all $t \geq 0$ so that the current state of the SIR model is completely described by X_t . We model $\tilde{\beta}_t$ as a generalized Brownian motion process so that

$$d\tilde{\beta}_t = \sigma dW_{3,t}$$

and $W_{3,t}$ is a standard one-dimensional Brownian motion process. Hence, we assume that the log infestation rate evolves according to a random walk in continuous time, with variability controlled by σ .

Combining this process with component SDEs describing the dynamics of S_t and I_t gives the complete SDE description of the SIR model with time-varying infestation rate as

$$dX_t = a(x_t, \theta)dt + \sqrt{b(x_t, \theta)}dW_t. \quad (1)$$

Here, $x_t = (s_t, i_t, \tilde{\beta}_t)'$ is the state of the system at time t , $\theta = (\gamma, \sigma)'$ is the vector of static parameter values, $W_t = (W_{1,t}, W_{2,t}, W_{3,t})'$ is a vector of uncorrelated standard Brownian motion processes, and the drift function $a(x_t, \theta)$ and diffusion coefficient $b(x_t, \theta)$ are given by

$$a(x_t, \theta) = \begin{pmatrix} -\exp(\tilde{\beta}_t)s_t i_t \\ \exp(\tilde{\beta}_t)s_t i_t - \gamma i_t \\ 0 \end{pmatrix}, \quad b(x_t, \theta) = \begin{pmatrix} \exp(\tilde{\beta}_t)s_t i_t & -\exp(\tilde{\beta}_t)s_t i_t & 0 \\ -\exp(\tilde{\beta}_t)s_t i_t & \exp(\tilde{\beta}_t)s_t i_t + \gamma i_t & 0 \\ 0 & 0 & \sigma^2 \end{pmatrix}. \quad (2)$$

Unfortunately, due to the nonlinear forms of $a(x_t, \theta)$ and $b(x_t, \theta)$, the SDE specified by (1) and (2) cannot be solved analytically. We, therefore, replace the intractable analytic solution with a tractable Gaussian process approximation, which is the subject of the next section. The resulting linear noise approximation is subsequently used as the inferential model.

2.2.1 | Linear noise approximation

The linear noise approximation (LNA) provides a tractable approximation to the SDE given by (1) and (2). In what follows we give a brief derivation; formal details can be found in Kurtz, (1972) (see also Kampen, 2001; Komorowski et al., 2009).

Consider a partition of X_t as

$$X_t = \eta_t + Z_t, \quad (3)$$

where $\{\eta_t, t \geq 0\}$ is a deterministic process satisfying the ODE

$$\frac{d\eta_t}{dt} = a(\eta_t, \theta), \quad \eta_0 = x_0, \quad (4)$$

and $\{Z_t, t \geq 0\}$ is a residual stochastic process. The residual process Z_t satisfies

$$dZ_t = \{a(x_t, \theta) - a(\eta_t, \theta)\} dt + \sqrt{b(x_t, \theta)} dW_t,$$

which will typically be intractable. The assumption that $\|X_t - \eta_t\|$ is “small” motivates a Taylor series expansion of $a(x_t, \theta)$ and $b(x_t, \theta)$ about η_t , with retention of the first two terms in the expansion of a and the first term in the expansion of b . This gives an approximate residual process $\{\hat{Z}_t, t \geq 0\}$ satisfying

$$d\hat{Z}_t = H_t \hat{Z}_t dt + \sqrt{b(\eta_t, \theta)} dW_t,$$

where H_t is the Jacobian matrix with (i, j) th element

$$(H_t)_{i,j} = \frac{\partial a_i(\eta_t, \theta)}{\partial \eta_{j,t}}.$$

For the SIR model in (1) and (2), we therefore have

$$H_t = \begin{pmatrix} -\exp(\tilde{\beta}_t) i_t & -\exp(\tilde{\beta}_t) s_t & -\exp(\tilde{\beta}_t) s_t i_t \\ \exp(\tilde{\beta}_t) i_t & \exp(\tilde{\beta}_t) s_t - \gamma & \exp(\tilde{\beta}_t) s_t i_t \\ 0 & 0 & 0 \end{pmatrix}.$$

Given an initial condition $\hat{Z}_0 \sim N(\hat{z}_0, \hat{V}_0)$, it can be shown that \hat{Z}_t is a Gaussian random variable (see Fearnhead et al., 2014). Consequently, the partition in (3) with Z_t replaced by \hat{Z}_t , and the initial conditions $\eta_0 = x_0$ and $\hat{Z}_0 = 0$ give

$$X_t \sim N(\eta_t, V_t), \quad (5)$$

where η_t satisfies (4) and V_t satisfies

$$\frac{dV_t}{dt} = V_t H_t' + b(\eta_t, \theta) + H_t V_t, \quad V_0 = 0. \quad (6)$$

Further details on the derivation of (6) are given in Appendix S1, Section S1. Hence, the linear noise approximation is characterized by the Gaussian distribution in (5), with mean and variance found by solving the ODE (ordinary differential equation) system given by (4) and (6). Although this ODE system will typically be intractable, a numerical scheme can be straightforwardly applied.

2.3 | Bayesian inference

We consider the case in which not all components of the stochastic epidemic model are observed. Moreover, we assume that data points are subject to measurement error, which accounts for mismatch

between the latent and observed process, due to, for example, the way in which the data are collected. Observations (on a regular grid) $y_t, t = 0, 1, \dots, n$ are assumed conditionally independent (given the latent process) with conditional probability distribution obtained via the observation equation,

$$Y_t \sim N(P'x_t, \sigma_e^2 P'x_t), \quad t = 0, 1, \dots, n \quad (7)$$

where $P = (1, 1, 0)'$. This choice of P is due to the data consisting of observations on the removed state R_t , which, for a known population size N , is equivalent (in information content) to observing the sum $S_t + I_t$. Note that the logarithm of the infestation rate process, $\tilde{\beta}_t$, is completely unobserved. Our choice of observation model is motivated by a Gaussian approximation to a Poisson $Po(P'x_t)$ distribution, with the role of σ_e^2 to allow a decoupling of the mean and variance. Moreover, the assumption of a Gaussian observation model admits a tractable observed data likelihood function, when combined with the LNA (see Section 2.2, Fearnhead et al., 2014; Golightly et al., 2015) as a model for the latent epidemic process X_t . Details on a method for the efficient evaluation of this likelihood function can be found in Section S2.3 of Appendix S1.

Given data $y = (y_0, y_1, \dots, y_n)$ and upon ascribing a prior density $\pi(\theta)$ to the components of $\theta = (\gamma, \sigma, \sigma_e)'$ (augmented to include σ_e), Bayesian inference proceeds via the joint posterior for the static parameters θ and unobserved dynamic process $x = (x_0, x_1, \dots, x_n)$. We have that

$$\pi(\theta, x|y) \propto \pi(\theta)\pi(y|\theta)\pi(x|y, \theta), \quad (8)$$

where $\pi(y|\theta)$ is the observed data likelihood and $\pi(x|y, \theta)$ is the conditional posterior density of the latent dynamic process. Although $\pi(y|\theta)$ and $\pi(x|y, \theta)$ can be obtained in closed form under the LNA, the joint posterior in (8) is intractable. In Appendix S1 Section S2, we describe a Markov chain Monte Carlo scheme for generating (dependent) samples from (8). Briefly, this comprises two steps: (i) the generation of samples $\theta^{(1)}, \dots, \theta^{(M)}$ from the marginal parameter posterior $\pi(\theta|y) \propto \pi(\theta)\pi(y|\theta)$ and (ii) the generation of samples $x^{(1)}, \dots, x^{(M)}$ by drawing from the conditional posterior $\pi(x|y, \theta^{(i)}), i = 1, \dots, M$.

Given inferences on the static parameters θ and the latent dynamic process x , we consider the following diagnostics for assessing model fit. The within-sample predictive density is

$$\pi(\tilde{y}|y) = \int \int \pi(\tilde{y}|x, \theta)\pi(\theta, x|y) dx d\theta, \quad (9)$$

and the one step ahead out of sample predictive density is

$$\pi(y_{n+1}|y) = \int \int \pi(y_{n+1}|x_{n+1}, \theta)\pi(x_{n+1}|x_n, \theta)\pi(\theta, x|y) dx_0 \dots dx_n d\theta. \quad (10)$$

Hence, in both cases we properly account for parameter and latent process uncertainty. Although the densities in (9) and (10) will be intractable, we may generate samples via Monte Carlo, see Appendix S1 Section S2 for further details.

3 | RESULTS

We assume the epidemic time series (see Section 2.1) for the cumulative number of trees with removed nests, $R(t)$, shown in Figure 2(c), can be described by the compartmental SIR model with a time-varying infestation rate (see Section 2.2). The aim is to estimate the key parameters through the Bayesian inference techniques described in Section 2.3. These are the time-varying infestation rate per tree, $\beta(t)$, with corresponding stochastic noise parameter σ describing $d\tilde{\beta}_t = d\log(\beta_t) = \sigma W_{3,t}$, the removal rate γ , and the observation error σ_e .

Section S2.4 of Appendix S1 provides details of the assumed population sizes for each park, initial numbers for the S (susceptible) and I (currently infected) tree categories and the initial infestation rate, along with the starting parameter values for the MCMC scheme and prior specification. Regarding the latter, we take an independent prior specification for the components of θ , so that $\pi(\theta) = \pi(\gamma)\pi(\sigma)\pi(\sigma_e)$. We then take lognormal $LN(1, 1)$ distributions for σ and σ_e , and a lognormal $LN(0, 0.5^2)$ distribution for γ . We assume that initial log infestation rate $\tilde{\beta}_0$ follows a Gaussian $N(-8.5, 0.5^2)$ distribution. These choices are motivated by the assumption of a median removal time of around 1 year (95% credible interval: (0.38, 2.66)), and a basic reproduction number at time 0 of $R_0 = \beta_0 N / \gamma$ covering a wide range of plausible values. For example, with $N = 5 \times 10^3$ the prior distributions for γ and $\tilde{\beta}_0$ lead to a 95% credible interval for R_0 of (0.25, 4.1). The initial conditions are chosen based on the

increase in the removal category in the first available year, for example, for Richmond Park there were 1414 new trees in the removal category between 2013 and 2014 (new trees that had nests removed in 2014), so we assume this was approximately the number of infested locations in 2013. We investigated several choices for initial conditions and find our results robust to these variations.

3.1 | Inference results

We ran the MCMC scheme for 10×10^3 iterations and monitored the resulting chains for convergence. Indicative trace plots can be found in Appendix S1, Figure S1 and suggest that the sampler has adequately explored the parameter space. Additional chains initialized at different starting values (not shown) further confirm convergence.

From the main MCMC run, we obtain the posterior within-sample means (with 50% and 95% credible intervals) for $R(t)$, $S(t)$, and $I(t)$, shown in Figures 3 and 4(a–c) for Bushy and Richmond Park, respectively. The logarithmic time-dependent infestation rate, $\tilde{\beta}_t = \log(\beta)$, is shown in Figures 3 and 4(d). For Bushy Park, the logarithmic infestation rate is plausibly constant (given *a posteriori* variance) at $\tilde{\beta}_t \approx -8$, corresponding to an approximate infestation rate of $\beta = 3.4 \times 10^{-4}$ and thus for the whole population of Bushy park an infestation rate of $\beta N = 1.7$. Similarly, for Richmond Park, the infestation rate is plausibly constant with $\tilde{\beta}_t \approx -10$, corresponding to an infestation rate of $\beta = 4.5 \times 10^{-5}$ and $\beta N = 1.8$. Reassuringly, samples from the

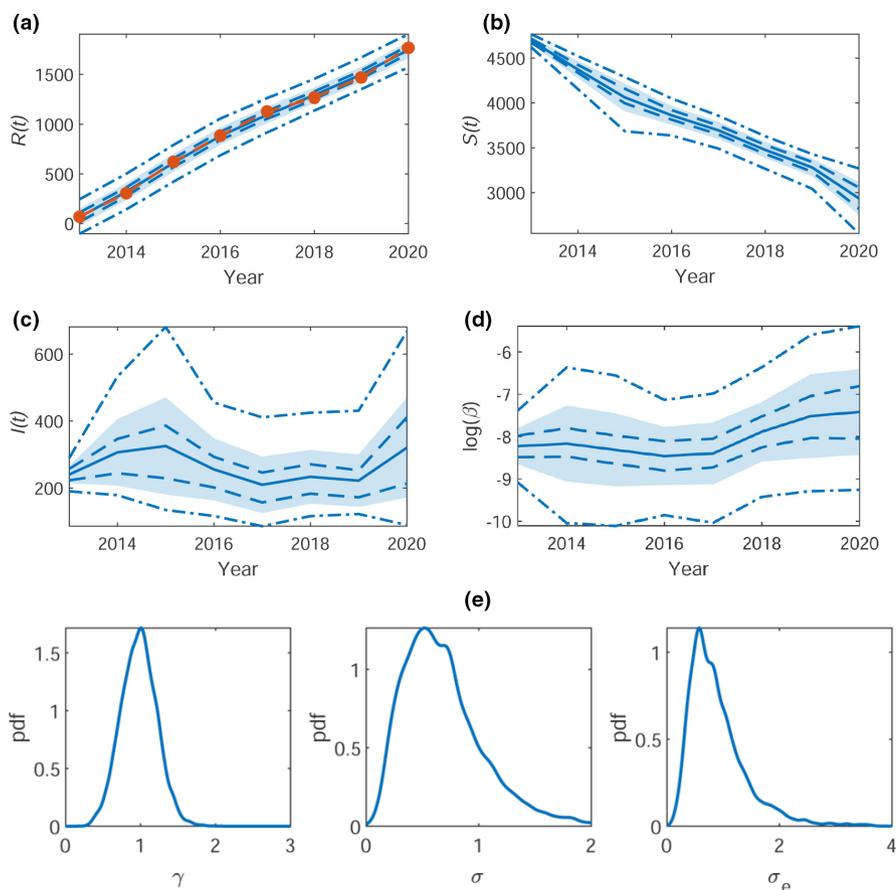
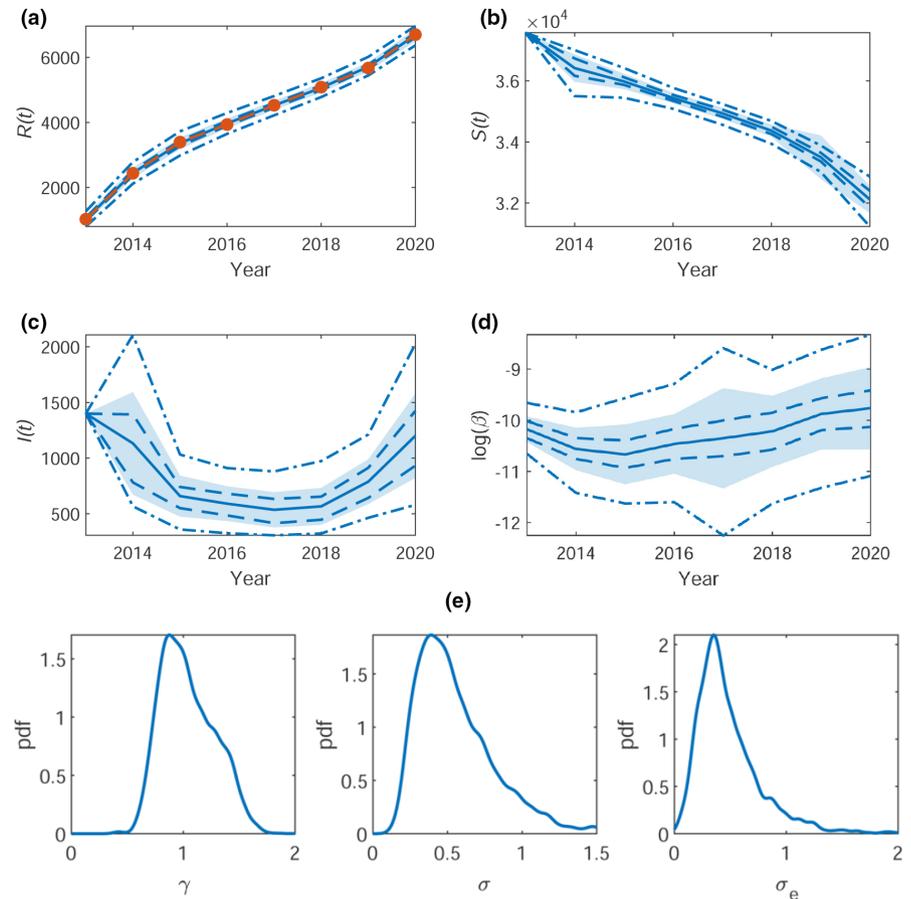


FIGURE 3 Bushy Park. The within-sample posteriors for (a) $R(t)$, (b) $S(t)$, (c) $I(t)$, and (d) $\log(\beta_t)$ with mean (blue solid line) \pm one standard deviation (shaded region), the 50% (blue dashed), and the 95% (blue dot-dashed) credible regions. The observed time series for $R(t)$ is overlaid in (a) (orange dashed). The corresponding (e) posterior densities for the inferred parameters γ (removal rate), σ (noise on $\tilde{\beta}_t$) and σ_e (observation error)

FIGURE 4 Richmond Park. The within-sample posteriors for (a) $R(t)$, (b) $S(t)$, (c) $I(t)$, and (d) $\log(\beta_t)$ with mean (blue solid line) \pm one standard deviation (shaded region), the 50% (blue dashed), and the 95% (blue dot-dashed) credible regions. The observed time series for $R(t)$ is overlaid in (a) (orange dashed). The corresponding (e) posterior densities for the inferred parameters γ (removal rate), σ (noise on $\tilde{\beta}_t$) and σ_e (observation error)



within-sample predictive for $R(t)$ are consistent with the data used to fit the model (see panel (a) of Figures 3 and 4).

The posterior density plots of the parameters $\theta = (\gamma, \sigma, \sigma_e)$ are shown in Figures 3 and 4(e), for Bushy and Richmond park, respectively. Pairwise joint posterior densities can be found in Appendix S1, Figure S2. The marginal posterior distribution of γ is centered around $\gamma \approx 1$ for both Bushy and Richmond. The marginal posterior for σ is centered around $\sigma \approx 0.75$ for Bushy and $\sigma \approx 0.5$ for Richmond. The observation error σ_e is centered around $\sigma_e \approx 1$ for Bushy and $\sigma_e \approx 0.5$ for Richmond.

3.2 | Estimation of R_0

From the posterior estimations of $\tilde{\beta}_t$ for each year and the parameter γ , we can estimate the basic reproduction number R_0 . This gives a measure of the strength of infectivity through the number of trees which become infested as a result of a single infested tree over its infested life time (i.e., the expected number of secondary infestations resulting from a single original infestation). This provides additional information over the parameter β , since R_0 takes into account the lifetime in which a tree can infest another tree (i.e., the removal rate γ). In a deterministic system, for an epidemic to die out, R_0 must be less than the threshold value of one. However, in the stochastic case, it is possible for R_0 to be above one but the epidemic still die out as a result of the stochastic fluctuations. Therefore, it is required that

$R_0 < 1$ for the epidemic to shrink, upon averaging over the stochasticity. In a SIR model with a constant infestation rate, β , the basic reproduction number is given by $R_0 = \beta N / \gamma$. Here, we adapt this to use the time variant infestation rate to get a reproduction number for each of the years between 2013 and 2020, $R_0 = \beta(t)N / \gamma$. Box plots showing the posterior distributions of R_0 for both parks are shown in Figure 5. For both parks R_0 has been stable, within errors, since 2013 (corresponding to the relatively constant β_t). However, this suggests that R_0 is still above one, and therefore, the epidemic will continue to propagate in these areas and potentially beyond.

3.3 | Forward prediction

Predictions of the spread of OPM are needed to inform control strategies. To test the validity of the SIR model with the inferred parameters from Section 3.1 and thus how well the model can capture future expansions in OPM, we can calculate a one-year prediction for a known data point. We remove the last data point, $R(2020)$, and re-infer the parameters for the new shortened observed time series. We then use these parameters to run the model forward (10×10^3) simulations, matching the number of iterations in the MCMC and obtain an estimate for $R(2020)$. The median predictions with upper and lower quartiles for 1000 runs are shown in Figure 6(a,c) for Bushy and Richmond, respectively. In both cases, the predictive interval captures the observed data. Realizations from 100 forward runs are shown in

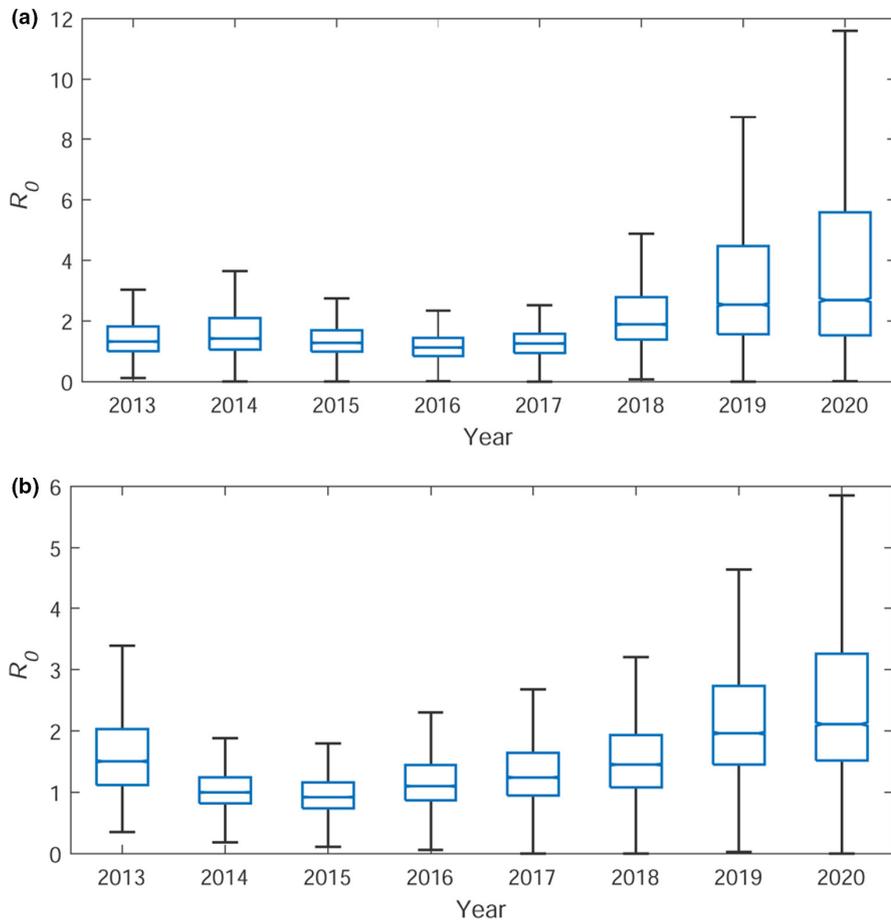


FIGURE 5 Posterior distributions of $R_0(t) = \beta_t N / \gamma$ for (a) Bushy and (b) Richmond Park. The central line indicates the median, with the bottom and top edges of the box showing the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers, which are not shown here

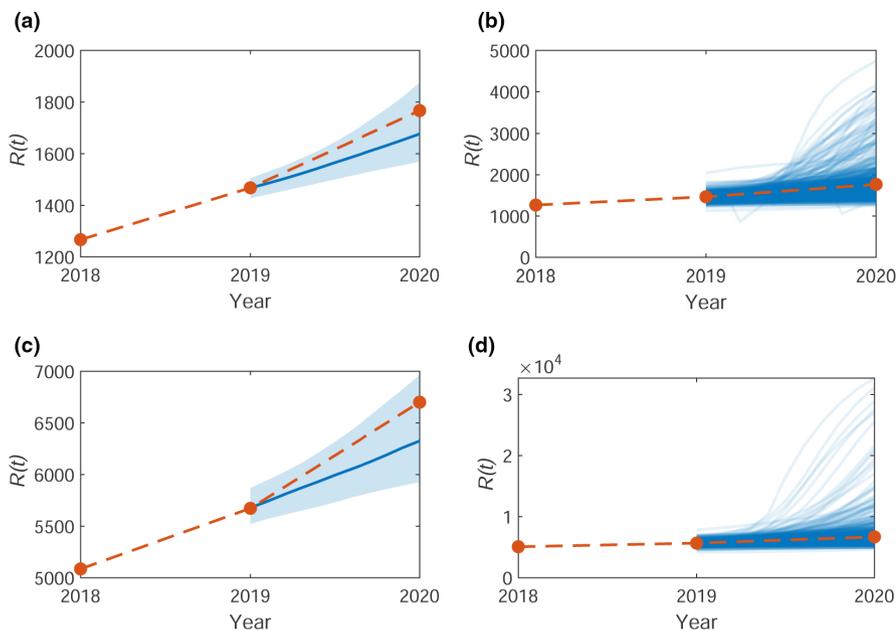


FIGURE 6 Model predictions for the total number of trees with removed nests up to 2020, $R(2020)$, with median (blue line) for (a, b) Bushy and (c, d) Richmond. In (a) and (c), the shaded area shows the 50% credible region. In (b) and (d), 100 simulations are shown from the forward model. The orange line shows the observed data

Figure 6(b,d) to show that results are mostly concentrated around the observed data, with some outliers over-estimating $R(t)$. One-step predictions for the whole time series are shown in Appendix S1, Figure S3. This suggests that the available data are sufficient to inform the model to make predictions over these time scales.

Similarly, we can produce predictions for the number of infested locations in 2021, $R(2021)$. The median predictions with upper and

lower quartiles are shown in Figure 7(a,b) for Bushy and Richmond, respectively. This corresponds to an average (median) of 350 new infested locations (lower-upper quartile estimate range 150–800) in Bushy Park and 1100 (700–2000) in Richmond Park. Since submission of this manuscript, the data for 2021 were recorded, with 167 new infested locations in Bushy and 523 in Richmond Park. This is lower than our mean predicted estimates, which could be due to

increased efficacy of the control methods with time. Longer-term predictions up to the year 2025 are shown in Figure 7(c,d) for Bushy and Richmond, respectively. By this time, $R(t)$ is beginning to saturate due to a depletion of available trees that have not yet previously been infested, that is, the infestation has spread through the whole susceptible population.

4 | DISCUSSION

Recent modeling work has suggested that the surroundings of the current OPM infestation area in the UK are highly climatically suitable and, therefore, at very high risk from future infestations (Godefroid et al., 2020). Since the government strategy for the containment of OPM relies on targeted control at the boundary of the current infested area, it is crucial to understand and be able to predict the future spread to optimize both the cost and efficacy of these control programs (Contingency Plan, 2021).

We have shown the applicability of a SIR compartmental model with a time-varying infestation rate to describe the OPM epidemic in the UK between the years 2013 and 2020. Such models have previously been used to describe the spread of tree diseases (Rodriguez-Quinones & Gordillo, 2019) and invasive species (Ferrari et al., 2014; Wildemeersch et al., 2019). The statistical methodology used is a powerful tool for inferring the parameters of such models from real data and is transferable to other epidemiological and ecological datasets. Previously, similar statistical methodology has been used to describe the spread of infectious diseases (e.g., measles (Cauchemez & Ferguson, 2008) and Ebola (Fintzi et al., 2020)) and the spatial expansion of non-native plants (Cook et al., 2007), but has not yet been applied to the study of invasive insects.

Our results show, along with previous analysis (Suprunenko et al., 2021), that the spread of OPM is continuing at a stable rate despite

the current intervention methods. Correspondingly, we show that the basic reproduction number R_0 has been above one since 2013. To see a reduction in the OPM population density and to protect the surrounding areas, a reduction of R_0 to below one would need to be seen. Although the basic reproduction number R_0 is typically used in the modeling of infectious diseases (Dietz, 1993; Ma, 2020), here it gives an analogous measure for the expected number of infested trees caused by a single currently infested tree through its infested lifetime.

Driven by the nature of the data collected, we chose to make the assumption that the trees with removed nests best represented the removed category in the SIR model. Although not explicitly formulated in this model, we expect that after nest removal and spraying with a biological insecticide, these trees will not be susceptible to future infestation on the short time scales considered here. This limitation of the model could be explored further through assuming this data instead represented the infested category (with the caveat that these trees would not actually be infective to others at the times the data were collected) or by extending the model to an SIRS formulation, which would allow for re-infestation after a period of immunity.

For simplicity and to be better described by a SIR model, we counted each tree that had undergone nest removal as one removed tree, regardless of how many nests were recorded as being removed from it. However, the defoliation effects and risks to human health from OPM are closely related to nest density (i.e., the numbers of nests per tree) (Jactel et al., 2011). In future work, nest density could be taken into account through a nest density-dependent infestation rate.

A challenge of modeling OPM and other tree pests and diseases is the lack of a complete inventory oak trees in the UK, representing the susceptible population in our SIR model. This has been previously noted and highlighted as a priority for future data collection by other modeling studies (Cowley et al., 2015). It is of particular importance for future spatial models of OPM, which require an estimate of the distribution of oak trees in the areas of interest.

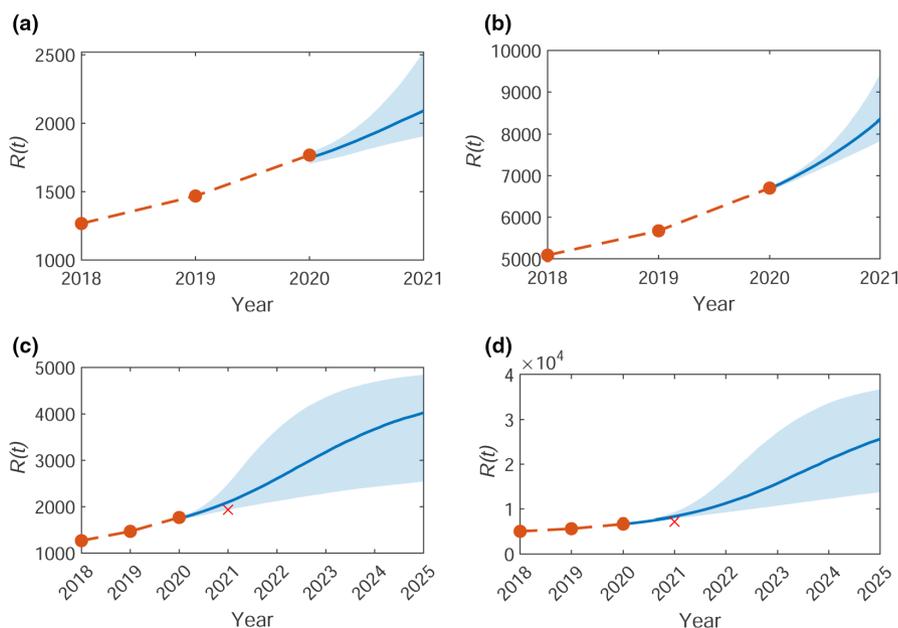


FIGURE 7 Model predictions for the median (solid blue line) total number of trees with removed nests up to 2021, $R(2021)$, for (a) Bushy and (b) Richmond park, and up to 2025, $R(2025)$, for (c) Bushy and (d) Richmond park. The shaded area shows the 50% credible region. The orange line shows the observed data up to 2020 used to parameterize the model, and the red cross shows the most recent data for 2021

It is also worth noting that many areas infested with OPM have been undergoing control measures (Contingency Plan, 2021) and so any inferred infestation rates represent the dynamics under these controls, rather than the inherent parameters of the uncontrolled pest spread. In Richmond and Bushy Parks, control measures include the yearly nest removal and limited spraying with a biological insecticide. It would be interesting to conduct a similar analysis on a contained area that had undergone no (or different) control measures to assess the differences in the infestation rates and thus assess the efficacy of the controls. The effect of confounding factors such as the weather, difference in landscapes, and the presence of other pests and parasitoids should also be investigated.

The results from this work can inform the development of future mathematical models for the spread of OPM. These models can be used to identify at-risk regions (Cowley et al., 2015) and predict the distribution of OPM on a national scale. The development of these models will require further targeted data collection to obtain complete oak tree inventories, as well as data on the population numbers and locations of OPM (or indeed any other invasive insect or pathogen).

ACKNOWLEDGMENTS

We thank Gillian Jonusas from The Royal Parks for sharing the OPM data for Bushy and Richmond parks. LW, AG, NP, and AB acknowledge support of the EPSRC New Horizons Grant EP/V048511/1.

AUTHOR CONTRIBUTIONS

Laura E. Wadkin involved in conceptualization (equal); formal analysis (equal); investigation (lead); methodology (equal); project administration (lead); software (equal); visualization (lead); and writing—original draft (lead). **Julia Branson** and **Andrew Hoppit** involved in writing—review and editing (supporting). **Nicholas G. Parker** and **Andrew W. Baggaley** involved in conceptualization (equal); funding acquisition (equal); methodology (equal); supervision (equal); and writing—review and editing (equal). **Andrew Golightly** involved in conceptualization (equal); formal analysis (equal); funding acquisition (equal); methodology (equal); software (equal); supervision (equal); and writing—review and editing (equal).

DATA AVAILABILITY STATEMENT

OPM data: The data used in this paper were collected, processed, and kindly shared by The Royal Parks charity. This data can be provided from The Royal Parks upon reasonable request. Statistical modeling software: all inference schemes and modeling software with example time series are available on the Newcastle University data repository at 10.25405/data.ncl.19292216.

ORCID

Laura E. Wadkin  <https://orcid.org/0000-0001-7355-2023>

Julia Branson  <https://orcid.org/0000-0002-7511-1026>

Nicholas G. Parker  <https://orcid.org/0000-0002-7238-8492>

Andrew Golightly  <https://orcid.org/0000-0001-6730-1279>

Andrew W. Baggaley  <https://orcid.org/0000-0002-6813-2443>

REFERENCES

- Andersson, H. K., & Britton, T. (2000). *Stochastic epidemic models and their statistical analysis*. Lecture Notes in Statistics. Springer-Verlag.
- Aukema, J. E., Leung, B., Kovacs, K., Chivers, C., Britton, K. O., Englin, J., Frankel, S. J., Haight, R. G., Holmes, T. P., Liebhold, A. M., McCullough, D. G., & Von Holle, B. (2011). Economic impacts of non-native forest insects in the continental United States. *PLoS One*, 6(9), e24587. <https://doi.org/10.1371/journal.pone.0024587>
- Bartlett, M. (1949). Some evolutionary stochastic processes. *Journal of the Royal Statistical Society: Series B (Methodological)*, 11, 211–229. <https://doi.org/10.1111/j.2517-6161.1949.tb00031.x>
- Boyd, I. L., Freer-Smith, P. H., Gilligan, C. A., & Godfray, H. C. J. (2013). The consequence of tree pests and diseases for ecosystem services. *Science*, 342(6160), 1235773. <https://doi.org/10.1126/science.1235773>
- Bradshaw, C. J. A., Leroy, B., Bellard, C., Roiz, D., Albert, C., Fournier, A., Barbet-Massin, M., Salles, J., Simard, F., & Courchamp, F. (2016). Massive yet grossly underestimated global costs of invasive insects. *Nature Communications*, 7(1), 1–8. <https://doi.org/10.1038/ncomms12986>
- Cauchemez, S., & Ferguson, N. M. (2008). Likelihood-based estimation of continuous-time epidemic models from time-series data: Application to measles transmission in London. *Journal of the Royal Society, Interface*, 5(25), 885–897. <https://doi.org/10.1098/rsif.2007.1292>
- Contingency Plan (2021). *Oak Processionary Moth (Thaumetopoea processionea) Contingency Plan*. [Online; accessed 28-October-2021].
- Cook, A., Marion, G., Butler, A., & Gibson, G. (2007). Bayesian inference for the spatio-temporal invasion of alien species. *Bulletin of Mathematical Biology*, 69(6), 2005–2025. <https://doi.org/10.1007/s11538-007-9202-4>
- Cowley, D. J., Johnson, O., & Pocock, M. J. O. (2015). Using electric network theory to model the spread of oak processionary moth, *Thaumetopoea processionea*, in urban woodland patches. *Landscape Ecology*, 30(5), 905–918. <https://doi.org/10.1007/s10980-015-0168-6>
- Dietz, K. (1993). The estimation of the basic reproduction number for infectious diseases. *Statistical Methods in Medical Research*, 2(1), 23–41. <https://doi.org/10.1177/096228029300200103>
- Dureau, J., Kalogeropoulos, K., & Baguelin, M. (2013). Capturing the time-varying drivers of an epidemic using stochastic dynamical systems. *Biostatistics*, 14(3), 541–555. <https://doi.org/10.1093/biostatistics/kxs052>
- Ellison, A. M. (2004). Bayesian inference in ecology. *Ecology Letters*, 7(6), 509–520. <https://doi.org/10.1111/j.1461-0248.2004.00603.x>
- Fearnhead, P., Giagos, V., & Sherlock, C. (2014). Inference for reaction networks using the Linear Noise Approximation. *Biometrics*, 70, 457–466. <https://doi.org/10.1111/biom.12152>
- Ferrari, J. R., Preisser, E. L., & Fitzpatrick, M. C. (2014). Modeling the spread of invasive species using dynamic network models. *Biological Invasions*, 16(4), 949–960. <https://doi.org/10.1007/s10530-013-0552-6>
- Fintzi, J., Wakefield, J., & Minin, V. N. (2020). A linear noise approximation for stochastic epidemic models fit to partially observed incidence counts. *Biometrics*, 1–12. <https://doi.org/10.1111/biom.13538>
- Freer-Smith, P. H., & Webber, J. F. (2017). Tree pests and diseases: The threat to biodiversity and the delivery of ecosystem services. *Biodiversity and Conservation*, 26(13), 3167–3181. <https://doi.org/10.1007/s10531-015-1019-0>
- Fuchs, C. (2013). *Inference for diffusion processes: With applications in life sciences*. Springer Science & Business Media.
- Gertsev, V. I., & Gertseva, V. V. (2004). Classification of mathematical models in ecology. *Ecological Modelling*, 178(3–4), 329–334. <https://doi.org/10.1016/j.ecolmodel.2004.03.009>

- Godefroid, M., Meurisse, N., Groenen, F., Kerdelhué, C., & Rossi, J.-P. (2020). Current and future distribution of the invasive oak processionary moth. *Biological Invasions*, 22(2), 523–534. <https://doi.org/10.1007/s10530-019-02108-4>
- Golightly, A., Henderson, D. A., & Sherlock, C. (2015). Delayed acceptance particle MCMC for exact inference in stochastic kinetic models. *Statistics and Computing*, 25, 1039–1055. <https://doi.org/10.1007/s11222-014-9469-x>
- Gottschling, S., & Meyer, S. (2006). An epidemic airborne disease caused by the oak processionary caterpillar. *Pediatric Dermatology*, 23(1), 64–66. <https://doi.org/10.1111/j.1525-1470.2006.00173.x>
- Heesterbeek, J. A. P., & Dietz, K. (1996). The concept of R0 in epidemic theory. *Statistica Neerlandica*, 50(1), 89–110.
- Ho, L. S. T., Xu, J., Crawford, F. W., Minin, V. N., & Suchard, M. A. (2018). Birth/birth-death processes and their computable transition probabilities with biological applications. *Journal of Mathematical Biology*, 76(4), 911–944. <https://doi.org/10.1007/s00285-017-1160-3>
- Jactel, H., Birgersson, G., Andersson, S., & Schlyter, F. (2011). Non-host volatiles mediate associational resistance to the pine processionary moth. *Oecologia*, 166(3), 703–711. <https://doi.org/10.1007/s00442-011-1918-z>
- Kenis, M., Auger-Rozenberg, M., Roques, A., Timms, L., Pere, C., Cock, M. J. W., Settele, J., Augustin, S., & Lopez-Vaamonde, C. (2009). Ecological effects of invasive alien insects. *Biological Invasions*, 11(1), 21–45. <https://doi.org/10.1007/s10530-008-9318-y>
- Kermack, W. O., & McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and*, 115(772), 700–721. <https://doi.org/10.1098/rspa.1927.0118>
- Komorowski, M., Finkenstadt, B., Harper, C., & Rand, D. (2009). Bayesian inference of biochemical kinetic parameters using the linear noise approximation. *BMC Bioinformatics*, 10(1), 343. <https://doi.org/10.1186/1471-2105-10-343>
- Kurtz, T. G. (1972). The relationship between stochastic and deterministic models for chemical reactions. *The Journal of Chemical Physics*, 57, 2976–2978. <https://doi.org/10.1063/1.1678692>
- Kypraios, T., Neal, P., & Prangle, D. (2017). A tutorial introduction to Bayesian inference for stochastic epidemic models using Approximate Bayesian Computation. *Mathematical Biosciences*, 287, 42–53. <https://doi.org/10.1016/j.mbs.2016.07.001>
- Lenzner, B., Latombe, G., Capinha, C., Bellard, C., Courchamp, F., Diagne, C., Dullinger, S., Golivets, M., Irl, S. D. H., Kühn, I., Leung, B., Liu, C., Moser, D., Roura-Pascual, N., Seebens, H., Turbelin, A., Weigelt, P., & Essl, F. (2020). What will the future bring for biological invasions on islands? An expert-based assessment. *Ecology and Evolution*, 8, 280. <https://doi.org/10.3389/fevo.2020.00280>
- Ma, J. (2020). Estimating epidemic exponential growth rate and basic reproduction number. *Infectious Disease Modelling*, 5, 129–141. <https://doi.org/10.1016/j.idm.2019.12.009>
- Maier, H., Spiegel, W., Kinaciyan, T., & Hönigsmann, H. (2004). Caterpillar dermatitis in two siblings due to the larvae of *Thaumetopoea processionea* L., the oak processionary caterpillar. *Dermatology*, 208(1), 70–73. <https://doi.org/10.1159/000075051>
- Maier, H., Spiegel, W., Kinaciyan, T., Krehan, H., Cabaj, A., Schopf, A., & Hönigsmann, H. (2003). The oak processionary caterpillar as the cause of an epidemic airborne disease: Survey and analysis. *British Journal of Dermatology*, 149(5), 990–997. <https://doi.org/10.1111/j.1365-2133.2003.05673.x>
- Manchester, S. J., & Bullock, J. M. (2000). The impacts of non-native species on UK biodiversity and the effectiveness of control. *Journal of Applied Ecology*, 37(5), 845–864. <https://doi.org/10.1046/j.1365-2664.2000.00538.x>
- McKinley, T. J., Ross, J. V., Deardon, R., & Cook, A. R. (2014). Simulation-based Bayesian inference for epidemic models. *Computational Statistics & Data Analysis*, 71, 434–447. <https://doi.org/10.1016/j.csda.2012.12.012>
- Mindlin, M. J., le Polain de Waroux, O., Case, S., & Walsh, B. (2012). The arrival of oak processionary moth, a novel cause of itchy dermatitis, in the UK: Experience, lessons and recommendations. *Public Health*, 126(9), 778–781. <https://doi.org/10.1016/j.puhe.2012.06.007>
- Oksendal, B. (1995). *Stochastic differential equations: An introduction with applications* (6th ed.). Springer-Verlag.
- Orozco-Fuentes, S., Griffiths, G., Holmes, M. J., Ettelaie, R., Smith, J., Baggaley, A. W., & Parker, N. G. (2019). Early warning signals in plant disease outbreaks. *Ecological Modelling*, 393, 12–19. <https://doi.org/10.1016/j.ecolmodel.2018.11.003>
- Parry, M., Gibson, G. J., Parnell, S., Gottwald, T. R., Irey, M. S., Gast, T. C., & Gilligan, C. A. (2014). Bayesian inference for an emerging arboreal epidemic in the presence of control. *Proceedings of the National Academy of Sciences of the United States of America*, 111(17), 6258–6262. <https://doi.org/10.1073/pnas.1310997111>
- Plant Health (England) (Amendment) (no. 3) Order (2018). [Online; accessed 06-December-2021].
- Rahlenbeck, S., & Utikal, J. (2015). The oak processionary moth: A new health hazard? *British Journal of General Practice*, 65(637), 435–436. <https://doi.org/10.3399/bjgp15X686341>
- Rodriguez-Quinones, L., & Gordillo, L. F. (2019). Low-density barriers for controlling plant crowd diseases: How far and fast can pathogens spread? *Nat. Natural Resource Modeling*, 32(4), e12222. <https://doi.org/10.1111/nrm.12222>
- Roy, B. A., Alexander, H. M., Davidson, J., Campbell, F. T., Burdon, J. J., Sniezko, R., & Brasier, C. (2014). Increasing forest loss worldwide from invasive pests requires new trade regulations. *Frontiers in Ecology and the Environment*, 12(8), 457–465. <https://doi.org/10.1890/130240>
- Scholten, T. P. H. (2021). *Predicting the oak processionary moth population using statistical modeling*. Master's thesis. Eindhoven University of Technology.
- Suprunenko, Y. F., Castle, M. D., Webb, C. R., Branson, J., Hoppit, A., & Gilligan, C. A. (2021). Estimating expansion of the range of oak processionary moth (*Thaumetopoea processionea*) in the UK from 2006 to 2019. *Agricultural and Forest Entomology*, 24, 53–62. <https://doi.org/10.1111/afe.12468>
- Tomlinson, I., Potter, C., & Bayliss, H. (2015). Managing tree pests and diseases in urban settings: The case of oak processionary moth in London, 2006–2012. *Urban Forestry and Urban Greening*, 14(2), 286–292. <https://doi.org/10.1016/j.ufug.2015.02.009>
- van Kampen, N. G. (2001). *Stochastic processes in physics and chemistry*. North-Holland.
- Wang, X., & Song, X. (2008). Mathematical models for the control of a pest population by infected pest. *Computers & Mathematics with Applications*, 56(1), 266–278. <https://doi.org/10.1016/j.camwa.2007.12.015>
- Wildemeersch, M., Franklin, O., Seidl, R., Rogelj, J., Moorthy, I., & Thurner, S. (2019). Modelling the multi-scaled nature of pest outbreaks. *Ecological Modelling*, 409, 108745. <https://doi.org/10.1016/j.ecolmodel.2019.108745>

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Wadkin, L. E., Branson, J., Hoppit, A., Parker, N. G., Golightly, A., & Baggaley, A. W. (2022). Inference for epidemic models with time-varying infection rates: Tracking the dynamics of oak processionary moth in the UK. *Ecology and Evolution*, 12, e8871. <https://doi.org/10.1002/ece3.8871>