

Bayesian Models for Pest Detection

Chris Malone

2022-04-09

Contents

1	Introduction	5
1.1	Introducing the research	5
1.1.1	Establish background area (biosecurity, monitoring) . . .	5
1.1.2	Establish specific problem (identify gap explicitly) . . .	5
1.1.3	State the research question	5
1.2	Outline of the thesis	5
2	Literature review	7
2.1	Inferring absence of cryptic pests	7
2.2	Standard approaches to inferring absence of a cryptic species . .	8
2.2.1	Limitations of the standard approach	9
2.3	[Introduce most recent work] Incorporating population growth and Bayesian posterior inference	11
2.3.1	Model specification	11
2.4	[Motivate my extension] Limitations of the basic Bayesian model	12
2.4.1	Summarise	14
3	A generalised model of pest species detection	15
3.1	Introduction	15
3.2	General outline	16
3.2.1	Spatial or non-spatial?	16
3.2.2	General outline	17
3.3	Prior distributions	18
3.4	Model specification	19
3.4.1	The growth model	19
3.4.2	The surveillance (detection) model	20
3.5	Computing the posterior distribution	20
3.5.1	Analogy with mixed models	21
3.5.2	Sampling algorithms when the number of unknowns is unknown	21
3.5.3	What is ABC?	22
3.6	A simple example	22
4	Case study: <i>Ceratatis capitata</i>	23

4.1	Example case: Medfly	23
4.1.1	Background info on Medfly	24
4.1.2	Medfly are economically important	24
4.1.3	Medfly are cryptic	24
4.1.4	Medfly belong to class of cases described in chapter 2	25
4.2	The data	27
4.3	Model specification	28
4.3.1	Population size	28
4.3.2	Fly locations	29
4.3.3	Detection model	32
4.4	Sampling	34
4.5	Results	34

Chapter 1

Introduction

1.1 Introducing the research

1.1.1 Establish background area (biosecurity, monitoring)

1.1.2 Establish specific problem (identify gap explicitly)

1.1.3 State the research question

1.2 Outline of the thesis

Chapter 2

Literature review

In this chapter, I review some literature relating to the core problem that I am concerned with in this thesis. I begin by discussing the general problem of inferring absence of a cryptic, incipient pest species population. In particular, I emphasise that the problem requires the marriage of separate mathematical modelling cultures; (a) mathematical/ecological modelling, and (b) statistical inference.

2.1 Inferring absence of cryptic pests

The primary problem this work is concerned with is to infer the absence of a cryptic pest species. A cryptic species is any species which is hard to detect. One example is snakes, which are small, immobile, camouflaged, and located in inaccessible habitats (Inferring the Absence of a Species: A Case Study of

Snakes, Kery 2002). A problem that environmental managers and governors face is to determine the minimal length of time required for. I will refer to this problem as the **problem of post-outbreak surveillance program design**, or, alternatively, as the problem of program design.

2.2 Standard approaches to inferring absence of a cryptic species

In this section, I outline the standard existing approach to inferring pest absence. The simplest approach to this problem is given by McArdle (1990). Let the *rarity* of the species $p \in [0, 1]$ be the probability that a species is detected in any given sampling unit. (Sampling units can be arranged spatially or temporally; e.g. a survey that involves checking w weeks at k locations would have wk sampling units.) Then, the number of surveys in which the species is detected is given by $X \sim \text{Binomial}(T, p)$. Accordingly, the probability of *not* detecting the species in T surveys is given by

$$\alpha = \Pr(X = 0) = 1 - (1 - p)^T.$$

The last formula allows us to compute any of the 3 quantities α , p , and t , assuming the other two are given.

Given this framework, the problem of planning program becomes the following. We decide *a priori* what the smallest “rarity” p worth detecting is. McArdle

2.2. STANDARD APPROACHES TO INFERRING ABSENCE OF A CRYPTIC SPECIES⁹

supposes that if a species is sufficiently difficult to detect (while, nonetheless, $p > 0$) then it cannot be considered a member of an ecological community, and therefore not worthy of being deemed “present”. Write the smallest rarity worth detecting as p_0 . Then, we choose a minimum detection probability α that we are willing to accept. For example, we might wish to have a chance of $\alpha > 0.95$ of detecting a species, given that it is present. Then, with the above formula, we can rearrange to get the smallest number of survey units T such that detection probability α is achieved.

Statisticians will recognise that the above is essentially power analysis for data modelled as identically and independently distributed Bernoulli trials. This analogy can be made more concrete. For any fixed rarity p , we can derive the probability of observing n or more negative surveys. This is the p-value. We can then reject the hypothesis that the rarity is greater than p_0 , the rarity worth detecting.

McArdle’s method is popular in the literature, most likely due to its conceptual simplicity.

1996. Using statistical probability to increase confidence of inferring species

2.2.1 Limitations of the standard approach

So far, I have discussed a general model, first proposed by McArdle (1990), for surveillance of a species in general. This model works well for a range of cases. However, for pest surveillance, this model has several limitations. Firstly,

for monitoring of cryptic pest populations the minimal permissible rarity p_0 may be arbitrarily small. For example, this may be the case if (a) a pest is extremely hard to detect, so that even moderate sized populations have low detection probability, and/or (b) pests have high invasive potential, so that even extremely small populations pose an invasive threat. However, when $p_0 = 0$, the probability of non-detection is 1 at each timestep.

The model extracts information from the

1. No probabilities
 - Probabilities help decision makers
2. No estimation for p .
 - For cryptic pest species, the model is even for dangerous levels of p .
3. No model of growth.
4. Not suitable for arbitrarily small p_0 .
 - There may be no $p_0 > 0$ that we would consider acceptable. Then, the goal is to reject $H_0 : p \leq p_0 = 0$. The problem is that the model is that non-detections have probability 1 under the null hypothesis that $p_0 \leq 0$. I.e. the probability of detecting a member of the species is at most 0 under this setting. In other words, $\alpha_n = 0$, for any number of surveys n . To get around this, we might choose p_0 to be arbitrarily small. However, such a choice would be arbitrary and not motivated by scientific theory or value-judgment considerations. Also, setting p_0 to be arbitrarily small

2.3. [INTRODUCE MOST RECENT WORK] INCORPORATING POPULATION GROWTH AND BAYESIAN PO

may force us to be overly conservative. The smaller that we set p_0 , the larger the number of surveys n required to reject $\mathbb{H}_0 : p \geq p_0$.

5. No prior information.

- This is related to the previous point. The above does not allow us to incorporate prior information about the rarity of the species. If we have reason to believe that the species is nearly extinct, then the above methodology may give overly conservative results. I.e. it might recommend surveying for longer than is required.

2.3 [Introduce most recent work] Incorporating population growth and Bayesian posterior inference

A Bayesian approach to the planning program problem is proposed by (Barnes et al. [2022]). Barnes extend the model in two basic ways. Firstly, Barnes et al. supplement the detection model with a model of the population dynamics for the species.

2.3.1 Model specification

As mentioned above, the Barnes model consists of (a) a growth model and (b) a detection model. The model for population growth gives us priors for the size of the population at each time and/or location where a survey is completed. In

particular, a Poisson-offspring model is assumed for the population size. The Poisson-offspring model is a type of branching process, initially developed by (???). A Poisson-offspring process is a process $X_{t=1}^T$ such that $X_0 \sim \text{Poisson}(\lambda)$, $\lambda \in [0, \infty)$, and $X_t | X_{t-1} \sim \text{Poisson}(X_{t-1}\delta)$. The typical situation that fits this process is the case where the number of individuals at time $t = 0$ is Poisson distributed with mean λ , and each individual born at time t has identically and independently Poisson distributed offspring with mean δ .

2.4 [Motivate my extension] Limitations of the basic Bayesian model

There exists a category of cases under which the assumptions of basic model above are not plausible. In this section, I describe this category of cases, and outline why these cases cause problems for the Barnes model. Three assumptions of the above model are as follows:

1. The growth of an incipient population can be described by a Poisson-offspring model.
2. The probability of detecting a randomly drawn member of the population p is known, or can be reasonably estimated from data.
3. The population grows exponentially at a constant rate. I.e. the growth rate of the population is not random and does not change over time.

When these assumptions are not met, the model may not be a realistic repre-

2.4. [MOTIVATE MY EXTENSION] LIMITATIONS OF THE BASIC BAYESIAN MODEL13

sensation of the processes being modelled. Firstly, the Poisson-offspring distribution is most natural when all detectable individuals at time t are offspring of the individuals that were detected at time $t - 1$. However, this will not be the case if members of the population continue to be detectable after giving birth.

It is not always possible to estimate a “rarity” parameter p from data. Recall that the rarity of a population is the probability of detecting a randomly drawn member of the population in a single survey. Thus, it depends fundamentally on the design and intensity of the survey employed. However, survey designs may vary significantly between regions, so that we cannot generalise estimates of p . For example, suppose an environmental manager has estimated p , under a standard, general surveillance program. However, they later suspect an outbreak has occurred, and therefore wish to implement an intensified surveillance program. A concrete example of this can be seen in the case of Mediterranean fruit fly, surveillance is performed using a spatial network of surveillance units (i.e. traps) which are inspected at regular intervals. When an outbreak has occurred, or is suspected, the manager may wish to deploy additional, *supplementary* survey units.

It may be objected that it should be possible to derive an estimate of the rarity p using the probability-given-distance function p_d . For example, we could simulate

In applied contexts, the probability of capture must be estimated from data. In the case of small insect species, such as Mediterranean fruit fly, probability of

capture is estimated with the aid of release-recapture experiments. Sterilised flies are released, in an area containing a standardised grid of surveillance units (traps). Then, after some time has passed, flies are counted. The resulting recapture data lends itself to two natural estimands. Firstly, there is the probability of a randomly selected fly being captured at all, given some trapping grid setup (e.g. a square grid with 400m spacing). This is essentially the “rarity” parameter discussed above. Secondly, there is the probability of a randomly selected fly being captured in a given trap, given the distance between the trap and the release location of the flies.

As mentioned above the third assumption is that the population grows exponentially at a constant rate. This assumption will be unreasonable in cases where population growth is seasonally dependent, or and/or when we are uncertain about the growth rate.

2.4.1 Summarise

In effect, in cases where the three assumptions listed above do not hold, it will not be ...

Chapter 3

A generalised model of pest species detection

3.1 Introduction

In the previous chapter, I reviewed recent work by Barnes et al. on inferring absence of a cryptic pest species. I concluded the chapter by discussing some of the model's restricting assumptions, and how they affect the reliability of the model. In the present chapter, I propose to develop a method that relaxes the assumptions listed above, and may work for a relatively broad class of cases. The general proposal is to combine agent based simulation (ABS) with approximate Bayesian computation (ABC) to infer probability of absence when the processes governing population dynamics are arbitrarily complex. I first start by outlining

the method, before giving a simple example of the method in action. I finish by discussing some theoretical and practical limitations of the method.

I propose a general methodology, rather than a specific mathematical model. In chapter 4, I focus on a specific implementation of the method, for a particular mathematical model.

3.2 General outline

3.2.1 Spatial or non-spatial?

The model may be spatial or non-spatial, depending on whether spatial information is available, whether the spatial location of the population is considered important for determining probability of capture, and how realistic the modeller wants to make the model. When the model is spatial, it is assumed that that we have a stochastic or deterministic function that relates the distance between an individual, and a surveillance unit, and that surveillance unit detecting that individual. In other words, we have some idea of how distance between detector and individual relates to the probability that the detector finds that individual.

For example, let $L_i \in \mathbb{R}^2$ be the location of individual i , and let $L_k \in \mathbb{R}^2$ be the location of survey unit or detector k . Then we might specify the probability of detector k finding individual i as a function of a linear transformation of the distance $d = \|L_k - L_i\|$, i.e.

$$p_d = \sigma(a + bd) = \frac{1}{1 + e^{-(a+bd)}}$$

for some scalar tuple $(a, b) \in \mathbb{R}^2$.

3.2.2 General outline

The basic model is Bayesian. The basic components of the model are the population size N_t at time points $t \in \{1, 2, \dots, T\}$, location of the incipient population L , and the number of individuals detected y_t , $t \in \{1, 2, \dots, T\}$. It is assumed that $\mathbf{y} = (y_1, \dots, y_T)$ is our data vector. In the typical case, we will assume that $\mathbf{y} = \mathbf{0}_T$, where $\mathbf{0}_T$ is the zero vector in \mathbb{R}^T . In other words, we consider the hypothetical case where we do not detect any species members at any time across T consecutive (discrete) time points.

Finally, we assume that there are a fixed number of spatially located survey units $K \in \mathbb{N}$, at locations L_k , each of which is a two dimensional random variable. These locations may be considered random, or fixed (known to the environmental manager). Thus, our parameter vector can be written as

$$\theta = (N_1, \dots, N_T, L, L_1, \dots, L_T).$$

For the inference problem we are concerned with in this work, we wish to perform inference only on the N_t terms; L and the L_t terms are considered “nuisance” parameters (Gelman, BDA3, p. 63).

I note a couple of changes one could make, briefly. One could also model the density of the population rather than the location and size. I do not consider that here. Such a model may be useful when we have data on densities rather

than population sizes.

I focus on the parameterisation described above, because it is intuitive for this problem, where we are concerned with a small, spatially delimited, incipient population.

Under this framework, the problem is to learn the distribution of $N_T \mid \mathbf{y} = \mathbf{0}_T$ for any $T \in \mathbb{N}$. This is the *posterior distribution* of N_T . More precisely, we want to compute the function of the distribution $\Pr(N_T = 0 \mid \mathbf{y} = \mathbf{0}_T)$ - in other words, the probability that the pest population has been eradicated, given that we have failed to detect the population at any point in time.

3.3 Prior distributions

Deriving the posterior distribution for N_T requires that we define a joint prior distribution over the joint parameter vector θ .

The priors are determined by a stochastic model of the population dynamics for the species. Such a model may or may not be “agent based”. For the purpose of this work, I take the term “agent based” to mean that the model takes into account differential individual properties of members of the species. (In this work, I consider differential locations). We model the population dynamics flexibly as a stochastic process. Simulation allows for the incorporation of an arbitrary degree of biological and ecological complexity into the model. Using an agent based model allows us to define a prior on each individual population member.

3.4 Model specification

Under the most general description, the basic model is composed of (a) a growth model and (b) a model of surveillance, or detection. The growth model describes our subjective beliefs about the size of the population initially, as well as how it is likely to change over time. The detection model describes our subjective beliefs about the size of the population.

3.4.1 The growth model

There are no a priori assumptions on the population dynamics for the growth model. For example, we might apply stochastic or deterministic models of logistic, exponential, or linear growth. The only thing that matters is that we specify a joint prior distribution over the population size across time points and locations.

For the present work, I focus on the case where there exists a single incipient population of unknown size.

A natural way to set a prior on the population size at each time point t is to set a prior on the population size at the initial time point, and then assume that the population sizes at other time points are given by some (deterministic or stochastic) function of the population size at $t - 1$, and the value of a covariate vector X_t , which includes variables relevant to population growth.

$$N_t = f(N_{t-1}, X_t)$$

3.4.2 The surveillance (detection) model

It is assumed that surveillances occur at regular time intervals $t \in \{1, \dots, T\}$.

3.5 Computing the posterior distribution

In this section, I discuss the problem of computing the posterior distribution, given a survey record. Above, I stated that the model could be defined flexibly. Without restrictions on the form of the growth and detection models, the posterior may be analytically intractable. In other words, we will not be able to write out the posterior density or mass as a function of the data and prior distributions. Such situations are common in the Bayesian framework, because of the tendency for the posterior density or mass to depend, implicitly or explicitly, on analytically intractable integrals.

In section 2, the model I outlined, given by Barnes et al., had a known analytic solution. In other words, the posterior probability of eradication could be computed as a relatively simple function of the number of negative surveys recorded (i.e., the data), and prior distributions on the population size at each time point and location.

So far, we have talked about situations when sampling is required for inference. Further problems arise when the model is *agent-based*. In other words, when we include uncertainty about individual-level features in the model. In this case, the detection probability is random, even when the location and population size is known. In other words, the probability of detecting at least one individual

is a function of the number of individuals, and also their individual (random) properties. This is a situation in which “the number of things you do not know is one of the things you do not know” (Richardson and Green, 1997).

3.5.1 Analogy with mixed models

3.5.2 Sampling algorithms when the number of unknowns is unknown

As mentioned above, standard MCMC algorithms for Bayesian inference will not work when the number of parameters is random. In this subsection, I discuss strategies for sampling from the posterior when this is the case. Firstly, there exist extensions to classical MCMC algorithms for the case where the number of parameters is random. Green (1995) outlines a method he calls “reversible jump MCMC”. This involves adding a step to the Metropolis Hastings algorithm, where . A second approach is to use approximate Bayesian computation (ABC). In this work, I focus on this method, as it has some nice properties for the issues we are concerned with here.

A sampler needs to move between points in different dimensional spaces.

3.5.3 What is ABC?

3.6 A simple example

In the following chapter, I give a detailed example of the method described above. However, for the purposes of explaining the method in simple terms, I will also give a simple example here.

Chapter 4

Case study: *Ceratatis capitata*

In the previous section, a general model for inferring absence/presence of a cryptic, incipient pest species population was introduced. In this chapter, I apply the model to a specific case study. The case study is for the outbreak of Mediterranean fruit fly (*Ceratatis Capitata*). Prior distributions are set and justified, and posterior inference is performed using approximate Bayesian computation.

4.1 Example case: Medfly

In chapter 2, I introduced a simple Bayesian model for inferring absence of a small population of a cryptic pest species, due to Barnes et al. (2021). I argued

that there existed a class of cases, where the assumptions of that simple model were not fully realistic. In chapter 3, I then introduced a method for relaxing the assumptions of this simple model. In the current section, I will introduce the reader to the problem of inferring absence of Medfly. I will then argue that the Medfly outbreak case belongs to the category of cases for which the Barnes model is not fully realistic.

4.1.1 Background info on Medfly

4.1.2 Medfly are economically important

Mediterranean fruit fly (*Ceratitidis Capitata*) or *medfly* is a fly species native to sub-Saharan Africa. It is considered to be of high economic importance, due to its potential for destruction of fruit production (Sciarette et al. 2018). Medfly has high invasive potential, as it can adapt to a relatively large range of climates and environments, and is known to have the capability to infest the fruits of over 300 species of plants (Ibid.).

4.1.3 Medfly are cryptic

I have mentioned above that Medfly are of high economic importance. However, they are also cryptic, insofar as they are very hard to detect at low levels. Monitoring for medfly is typically performed with the aid of lured traps (namely so-called Lynfield or Jackson traps). These traps are relatively ineffective for detecting medfly. For example, one study from the Adelaide metro area trapping grid found that only 0.02% of flies were recaptured from a release of 38.8 million

flies. Further, medfly are known to have low dispersals across space.

4.1.3.1 Medfly detection problems

Medfly have low dispersal. This means that flies may go undetected across generations. <https://onlinelibrary-wiley-com.virtual.anu.edu.au/doi/pdfdirect/10.1111/j.1570-7458.2006.00415.x>

4.1.4 Medfly belong to class of cases described in chapter

2

The reader may recall that the model introduced in chapter 3 was motivated by consideration of three restrictive assumptions, listed in chapter 2. These assumptions were that (a) population growth is a Poisson-offspring process, (b) detection probability does not depend on fly location, and (c) the growth rate is constant as a function of t .

I will briefly describe why these assumptions are difficult to justify in the case of medfly.

The first key assumption was that the pest population follows a Poisson-Offspring model. This is not a natural choice for the case of Medfly. Medfly increase rates are typically estimated in terms of a continuous exponential growth model. Secondly, the continuous exponential model is more common in ecology; therefore, it may be easier to understand to experts, from whom we may need to elicit priors. It may be objected that the Poisson-Offspring model has similarly interpretable parameters. The parameter is simply the number of

offspring that each adult gives birth to on average. However, this is only the case if the only trappable members of the population have been born in the current time interval. In the case of Medfly, however, this is not realistic. For example, adults give birth multiple times in their adult lifestage, and may be captured at any point. Secondly,

The second key assumption is that we can estimate the probability of capturing a randomly selected fly from data. This is difficult in the case of Medfly. For fruit flies, capture probability is typically estimated from data taken from release-recapture studies. In these studies, the researcher obtains a large collection of sterilised specimens, and releases them at a single point in space. Then, the

These experimental data can be useful when the trapping setup is similar to the setup we want to draw inference about. However, this will often not be the case. For example, studies vary in the number and type of traps used (SEE NOTE). Further, we may wish to infer eradication of pest populations in trapping systems that are highly unlike those in studies. For example, after an outbreak has occurred, and eradication measures have been stopped, it is common to set up supplementary trapping units to intensify monitoring and increase the likelihood of detecting flies, conditional on their presence in the area. (CITATION).

The third and final assumption was that the growth rate for the pest is roughly constant. This is not the case for Medfly, who reproduce in seasonal fruits. Reproduction rates are highly dependent on temperature and seasonal availability

of hosts. Therefore, growth varies systematically to a large degree across the year, in ways that are somewhat well understood. Most likely, the environmental manager should use information about the weather and time of year in setting priors on the growth rate for the species.

- Source for supplementary trapping: <https://nre.tas.gov.au/Documents/Review_of_IR_for_FruitFly.

- Problem 3: Growth rate is assumed constant.

In this section, I argued that the case of Medfly belongs to the class of cases, described in chapter 2, that do not fit the assumptions of the Barnes model. This justifies relaxing the assumptions of the model of Barnes et al., and using a more flexible model to infer pest absence, in the case of Medfly. I turn now to specifying a full model for this case.

4.2 The data

In this section, I do not use real data to estimate parameters. Instead, I model a hypothetical situation in which we observe $\mathbf{y} = \mathbf{0}_T$ (see above). The situation is as follows: We assume that at least one fly has been detected; eradication measures have since begun and then ceased; and we now proceed with intensified monitoring, while whatever population that may exist is free to grow relatively unhindered. By intensified monitoring, I mean that **supplementary** monitoring traps have been placed alongside the previously existing grid of **general** monitoring traps. More precisely, it is assumed that **general** exist year round in a 400×400 metre grid (DPIPWE, 2011, p. 50). The **supplementary** surveil-

lance system consists of a set of 16 traps in a circular area, centred at the site of the first fly detection.¹ The goal of the analysis is to infer the probability of eradication for the incipient population, given no flies detected at any point in this period.

I assume that traps are checked weekly.

4.3 Model specification

As is typical of Bayesian models, prior distributions must be specified over each of the parameters. I do not consider uninformative priors, as in practical cases, information will exist about the parameters, and should be used.

I break the model into the following three components:

1. Population size,
2. Fly locations, and
3. Detections.

4.3.1 Population size

I assume that our beliefs about the initial population size N_1 are described by a Poisson with random mean λ . I assume that λ follows an exponential distribution. This distribution for N_1 is chosen as it is a discrete distribution with right skew, and a relatively large amount of mass $f_{N_1}(x)$ at $x = 0$, corresponding to

¹It is typical to wait until at least 2 flies have been detected near each other for an outbreak to be declared. To illustrate the method in a simplified setting, I suppose that one fly detection is sufficient.

the situation where flies are already eradicated.

As for prior distributions on N_t , for $t \in \{1, \dots, T\}$, a growth model is used to structure the prior. Namely, an exponential growth model is assumed, so that $N_t = \text{round}[N_{t-1} \exp\{R_t\}]$, where R_t is the growth rate at time t . The exponential growth model is chosen for its ubiquity in ecological science in general, and in studies of fruit fly dynamics in particular. Rounding is introduced to give N_t discrete support. The growth parameter R_t is uncertain, and based on temperature.²

4.3.2 Fly locations

I first discuss the option of setting a uniform prior. Setting an uninformative prior is fairly straightforward for this problem. In particular, we might assume that, beyond a certain distance from the outbreak centre (say, 1km) any existing population of Medfly is distinct from the population of interest. Therefore, we might set the prior distribution for the population location to be uniform on the surface of a disk with 1km radius around the outbreak centre.

Despite the fact that an uninformative prior is relatively straightforward to set, it is most likely not advisable in specific applications. It will typically be the case that prior information is available to the decision maker. In particular, fruit flies are heavily dependent on the availability of suitable fruit trees for survival and reproduction. Therefore, someone with local area knowledge will

²Alternatively, we could leave the rounding step out, and interpret N_t as the expected number of flies at each step. I do not consider this possibility in any further depth.

be able to determine the most likely locations for an existing population. Also, the supplementary zone is not chosen arbitrarily. The choice of supplementary zone will typically reflect the beliefs of the decision maker about the location of the fly population.

When prior information exists, setting the prior distribution to be uninformative may cause us to underestimate the likelihood of observing captures in the supplementary surveillance zone. The overall effect will be to inflate $\Pr(N_T = 0 \mid \mathbf{y} = \mathbf{0}_T)$.

To update on detection location when the first fly is detected at a trap (say trap k) we can use a trick. The trick is to model the probability of the first detection being at trap k as the probability that a fly is detected at k in one period conditional on exactly one fly total being detected in that period. The benefit of this model is that it does not depend on how many weeks it took to get the first detection (which would require information about how long flies have been around before the first detection). See appendix for more details.

A mathematical trick can be used to derive a prior in some cases. Suppose we have K traps indexed by $k \in \{1, \dots, L\}$. Suppose also that we have a prior distribution over the population size N , given by $N \sim \text{Poisson}(\lambda)$, with $\lambda \sim \text{Exponential}(1/20)$. Here we assume no change in population size over time. Now, we suppose that each trap k is “competing” to catch the first trap each week. We suppose that the trap at the centre of the grid was the first to catch

a fly, and we want to use this information. Define the random variable

$$C_k = \begin{cases} 1 & \text{a fly is caught in trap } k \text{ before any other trap} \\ 0 & \text{otherwise.} \end{cases}$$

Under these assumptions, $L \mid C_k = 1$ is the distribution of L , given that a fly was caught in trap k before any other trap.

Whether or not we can analytically derive the posterior density depends on the probability of capture function $p(x)$. In the case we consider here, the function cannot be integrated, and so I resort to sampling. Under the above assumptions, the posterior resembles the convolution of a normal and a uniform distribution (see figure). See appendix for more details.

4.3.2.1 Dispersals

I assume that flies in the population are dispersed in space around the central location L . Let $D_{i,t}$ denote the location of fly i at time t , relative to the population centroid L . It is assumed that the population centroid does not change over time (i.e. L is independent of t , and everything else in the model). However, $D_{i,t}$ is independent of $D_{i',t'}$, for any $(i', t') \neq (i, t)$. Thus, our belief is essentially that flies are shuffled around at each time point, so that a fly's location at $t - 1$ tells us nothing about its location at t , except through the information both reveal about L . This assumption justifies not tracking individual flies across time – whether a fly lives across time periods, or instead dies and is replaced,

are equivalent scenarios under this model.

4.3.3 Detection model

Finally, I discuss the model for detecting individuals. Conditional on N_t , L , and

$D_{i,t}$.

Population size

Initial no. of flies:	$N_1 \mid \lambda \sim \text{Pois}(\lambda), \text{ where}$	
	$\lambda \sim \text{Exponential}(1/20)$	
Rate of increase:	$R_t \sim \text{Normal}(\mu_t, \sigma_t^2),$	$t \in \{2, \dots, T\}$
No. of flies:	$N_t := \text{round}\{N_{t-1} \exp(R_t)\}$	$t \in \{2, \dots, T\}$

Fly locations

Popn. loc.:	$L^{(U)} \sim \text{Uniform}^2(200, 600)$	
	$L^{(N)} \sim \text{Normal}^2(0, \sigma)$	
	$L := L^{(U)} + L^{(N)}$	
Fly dispersals:	$D_{i,t} \sim \text{Normal}(0, 20)$	$i \in \{1, \dots, N_t\},$
		$t \in \{1, \dots, T\}$
Fly locations:	$L_{i,t}^{\text{fly}} := L + D_{i,t}$	$i \in \{1, \dots, N_t\},$
		$t \in \{1, \dots, T\}$

Detection model

No. traps:	$K \in \mathbb{N}_+$	
Trap locations:	L_k^{trap}	$k \in \{1, \dots, K\}$
Dist. btw. fly i and trap k at time t :	$\delta_{i,k,t} := \ L_k^{\text{trap}} - L_{i,t}^{\text{fly}}\ $	$i \in \{1, \dots, N_t\},$
		$k \in \{1, \dots, K\},$
		$t \in \{1, \dots, T\}$
Individ. cap. prob.:	$p_{i,t} = 1 - \prod_{k=1}^K (1 - p(\delta_{i,k,t})),$	$i \in \{1, \dots, N_t\},$
		$k \in \{1, \dots, K\},$
		$t \in \{1, \dots, T\}$

4.4 Sampling

4.5 Results

Bibliography

B Barnes, M Parsa, F Giannini, and D Ramsey. Analytical bayesian models to quantify pest eradication success or species absence using zero-sighting records. *Theoretical population biology*, 144:70–80, 2022.