Ecological Applications, 31(8), 2021, e02449

Optimal surveillance against bioinvasions: a sample average approximation method applied to an agent-based spread model

Hoa-Thi-Minh Nguyen [1], Pham Van Ha,[1] and Tom Kompas [2,3]

[1]*Crawford School of Public Policy, Australian National University, Crawford Building (132), Lennox Crossing, Canberra, Australian Capital Territory 2601 Australia*
[2]*Centre of Excellence for Biosecurity Risk Analysis, School of Biosciences and School of Ecosystem and Forest Sciences, University of Melbourne, Melbourne, Victoria 3010 Australia*

Citation: Nguyen, H. T. M., P. V. Ha, and T. Kompas. 2021. Optimal surveillance against bioinvasions: a sample average approximation method applied to an agent-based spread model. Ecological Applications 31(8):e02449. 10.1002/eap.2449

Abstract. Trade-offs exist between the point of early detection and the future cost of controlling any invasive species. Finding optimal levels of early detection, with post-border active surveillance, where time, space and randomness are explicitly considered, is computationally challenging. We use a stochastic programming model to find the optimal level of surveillance and predict damages, easing the computational challenge by combining a sample average approximation (SAA) approach and parallel processing techniques. The model is applied to the case of Asian Papaya Fruit Fly (PFF), a highly destructive pest, in Queensland, Australia. To capture the non-linearity in PFF spread, we use an agent-based model (ABM), which is calibrated to a highly detailed land-use raster map (50 m × 50 m) and weather-related data, validated against a historical outbreak. The combination of SAA and ABM sets our work apart from the existing literature. Indeed, despite its increasing popularity as a powerful analytical tool, given its granularity and capability to model the system of interest adequately, the complexity of ABM limits its application in optimizing frameworks due to considerable uncertainty about solution quality. In this light, the use of SAA ensures quality in the optimal solution (with a measured optimality gap) while still being able to handle large-scale decision-making problems. With this combination, our application suggests that the optimal (economic) trap grid size for PFF in Queensland is ~0.7 km, much smaller than the currently implemented level of 5 km. Although the current policy implies a much lower surveillance cost per year, compared with the $2.08 million under our optimal policy, the expected total cost of an outbreak is $23.92 million, much higher than the optimal policy of roughly $7.74 million.

Key words: *agent-based model; early detection; optimal surveillance; optimization; papaya fruit flies* (Bactrocera papayae); *sample average approximation; spatial-dynamic process; stochastic programming.*

Introduction

Alien invasive species (AIS) cause enormous environmental and economic damage (Wilcove et al. 1998, Pimentel et al. 2000), due, in part, to the delay in detection and control. It is impractical to prevent every post-border incursion that occurs behind the border and biosecurity measures generally struggle to keep up with the increasing risks of bioinvasions in an increasingly connected world. Therefore, post-border surveillance has received considerable attention (Epanchin-Niell and Hastings 2010), with recent literature highlighting the importance of combining early detection with the eradication of AIS to enhance biosecurity efforts, a practice that used to be applied to only highly economically damaging AIS (see Liebhold et al. 2016, for a review).

Manuscript received 28 April 2020; revised 20 February 2021; accepted 22 March 2021; final version received 26 August 2021. Corresponding Editor: William L. Kendall.
[3] Corresponding Author. E-mail: tom.kompas@unimelb.edu.au

Finding the optimal level of post-border surveillance is difficult, for at least two reasons. The first is how to model the AIS spread in the most realistic manner. Dangerous AIS often spread quickly and randomly, over time and space, at a rate that can be highly dependent on their characteristics, such as age (Van den Bosch et al. 1992), the landscape with which they interact, the interactions among themselves, and the prevailing weather conditions (Kot et al. 1996, Shigesada and Kawasaki 1997, Keeling et al. 2001). Furthermore, while most of the spread is local, some spread can be over a long distance, via jumps, and therefore substantially alter an outbreak size. This feature is typical in biology but hard to calibrate in modeling (Hastings et al. 2005, Meentemeyer et al. 2011). The second reason is the computational challenge in solving large-dimensioned problems. Overlaying an optimization routine on a realistic spread model, in which millions of cells are considered, can make the problem intractable. This aspect alone explains why quite a few studies have greatly

simplified their spread models, assuming, for example, constant spread rates (Meentemeyer et al. 2012).

In this study, we constructed a large-dimensional model, where time, space and randomness are explicitly considered, to find the optimal level of surveillance and predict damages. We circumvented the computational challenge using a combination of the sample average approximation (SAA) approach and parallel processing techniques. The model was applied to the case of Asian Papaya Fruit Fly (PFF), a highly destructive pest, in Queensland, Australia. To capture the non-linearity in PFF spread, we use an agent-based model (ABM), which is calibrated to a highly detailed land-use raster map (50 m × 50 m) and weather-related data, validated against a historical outbreak.

The combination of an SAA approach and an ABM sets our work apart from the existing literature. ABMs have become a powerful analytical tool in various fields, especially in ecology, given its granularity and capability to model the system of interest adequately (DeAngelis and Grimm 2014). However, the complexity of an ABM limits its application in optimizing frameworks. Indeed, there is considerable uncertainty about solution quality in studies in which an ABM is used in optimization problems (Barbati et al. 2012). In this light, the use of the SAA approach ensures quality in the optimal solution (where the optimality gap can be measured) while still being able to handle large-scale decision-making problems.

The paper is structured as follows. One section briefly reviews the related literature. The following section describes a discrete-time stochastic spatial-dynamic model to find optimal surveillance against biological invasions and the approach to solving it. We apply the model to the case of PFF in Queensland in the next section, in which the empirical dispersal model is an ABM. The last section provides some discussion, especially on the model's applicability to other contexts/systems as well as any caveats and limitations that are worth exploring in future research.

Approaches to Modeling Biosecurity

The economics of biosecurity has evolved rapidly alongside the increasing risk of bioinvasions, with a burgeoning literature since the early 2000s (see Olson [2006], Lovell et al. [2006] and Perrings et al. [2000] for reviews). Although biosecurity is managed as a continuum from pre-border to post-border, its activities are classified into three types, namely prevention (pre-border and at the border), post-border surveillance and post-border control and eradication, each of which are often studied separately. At first, most studies focused on prevention, designing trade policies in a way to minimize or prevent the introduction of AIS (e.g., Horan et al. 2002, Costello and McAusland 2003, McAusland and Costello 2004). Another strand analyzed strategies that are optimal for controlling established AIS incursions s (e.g., Olson and Roy 2002, Taylor and Hastings 2004, Epanchin-Niell and Hastings 2010, Epanchin-Niell and Wilen 2012). Over the last decade, however, with the recognition that it is impossible to fully prevent every entry at the border (including pests that arrive on wind or other environmental pathways), post-border surveillance efforts for early detection have attracted considerable attention (e.g., Mehta et al. 2007, Hauser and McCarthy 2009, Barrett et al. 2010, Jarrad et al. 2011).

When biosecurity measures are studied together, the economic trade-off among them is typically the focus. Existing literature tells us that control and prevention are generally substitutes for each other while control and surveillance are complementary (Polasky 2010). Specific bioeconomic modeling results also suggest that investment in prevention may likely yield higher returns than for control (e.g., Leung et al. 2002, 2005, Finnoff et al. 2007). However, this depends very much on context, the set of AIS incursions and the control measures (Kompas et al. 2019). The trade-off between control and surveillance is generally not clear. As a result, the existing literature conventionally finds an optimal surveillance and search effort by minimizing the sum of upfront surveillance cost and the probabilistic damage of an outbreak with needed control measures (e.g., Mehta et al. 2007, Epanchin-Niell et al. 2012, 2014, Horan et al. 2018).

Regardless of the focus, there has been increasing demand for bioeconomic models that can capture the typical features of AIS spread (Wilen 2007, Albers et al. 2010, Meentemeyer et al. 2012). This added demand has resulted in the rapid growth of ABM applications over the past two decades in all fields, including ecology (DeAngelis and Grimm 2014). Nonetheless, due to their complexity, ABMs have been used mainly in simulations. Despite being insightful, simulations can reveal only the relative efficiency of one, or at best a limited number of, policy choices, without determining an optimal outcome. In addition, full and explicit consideration of time, space and randomness in AIS spread in an optimization routine is generally limited by computational complexity. This aspect explains why the literature that applies an ABM for optimization problems is scant, and the quality of their solutions is yet to be assured (Barbati et al. 2012).

Against this background, we propose a bioeconomic model that uses an ABM to capture the realistic spread of a dangerous AIS and an optimization routine to determine an optimal level of post-border surveillance against it. Our work contributes directly to the growing literature that finds optimal surveillance and eradication of invasive species in heterogeneous landscapes. Specifically, similar to Epanchin-Niell et al. (2012) and different from most studies in the existing literature (e.g., Hauser and McCarthy 2009, Horie et al. 2013, Bonneau et al. 2019), we designed an optimal long-term surveillance program. As in related work by Epanchin-Niell et al. (2012), our model allows stochastic invasion, establishment and detection, unknown infected cells before

their detection, and increasing detectability with increasing age of infection. The distinct feature of our work lies in the use of an ABM spread model to represent the non-linear and random spread over time and space of an AIS. This paper indeed presents the first agent-based optimization surveillance model in the literature, to the best of our knowledge.

A Discrete-Time Stochastic Spatial-Dynamic Surveillance Model of Bioinvasions

In this section, we develop and solve a model to find the optimal level of surveillance to detect a bioinvasion early, considering the cost of surveillance itself and its potential benefit in reducing the economic damage of an outbreak. The spread model is stochastic, temporal and spatial, allowing for the growth and spread of a species over both discrete-time and space, with jumps, all in a random manner. It is worth noting that the description of the model here is general to accommodate many kinds of discrete-time stochastic spatial-dynamic dispersal processes. When being applied, the nature of stochasticity, the scales for time and space are context specific, as illustrated in the section on Empirical Application and Results. Our model is most applicable to situations in which the bioinvasion is probabilistic, and also potentially damaging to a sector or the economy enough to require (potential) eradication within a finite period.

Random dispersal model

To begin, consider a land area divided into q small raster cells that is either habitable or not by an AIS. The cell is considered to be small enough that the within-cell AIS population growth can be ignored. We denote x_{it} as the infestation state of a habitable cell i at time t where x_{it} can take on one of the two values: $x_{it} = 1$ means the cell is infested, while $x_{it} = 0$ means the cell is susceptible. Vector $\mathbf{X_t}$ represents infestation states of all habitable cells at time t. At $t = 0$, there is only one random cell infested. Moving forward in time, the infestation state of each cell i at time t where $t > 0$ depends on four factors in the t and/or previous A periods, during which an AIS can survive and search for a new host to colonize. The first factor is the cell's infestation status. The second is the realization of the random dispersal ξ in terms of range, direction and quantity. The third factor is the realization of the probability γ of an infested cell being detected without the aid of any active surveillance measures (or so-called "naturally detected," typically by a farmer or someone in the community). The fourth factor is the intensity of an active surveillance measure (g) to detect a bioinvasion early. Of these four factors, the last two factors affect the cell status in the sense that, if it is detected, it will be eradicated right away.

To this end, the random dispersal of an AIS over time and heterogenous space can be expressed using the following transition equation:

$$\begin{aligned}\mathbf{X_t} &= f(\mathbf{X_{t-A}}, \ldots, \mathbf{X_{t-1}}; \boldsymbol{\xi}_{\mathbf{t-A}}, \ldots, \boldsymbol{\xi}_{\mathbf{t-1}}; \boldsymbol{\gamma}_{\mathbf{t-A}}, \ldots, \boldsymbol{\gamma}_{\mathbf{t-1}}; X_0; g) \\ &= f(\boldsymbol{\Xi_t}; X_0; g) \end{aligned} \tag{1}$$

where $\boldsymbol{\Xi_t}$ is a matrix combining all information on the infestation states of all habitable cells and the realizations of random events; and X_0 is the initial condition, that is, the initial cell from which an outbreak starts.

Economic costs and optimization set-up

An AIS outbreak incurs several economic costs. For example, it damages production in the infested area, and is usually expensive to remove, both directly (eradication costs) and indirectly (management costs). Sales revenues also generally suffer due to trade sanctions during the outbreak and market closure periods. In many cases, products in non-infested (but nearby areas) can be sold to the market only after meeting some specific conditions such as being sprayed with particular chemicals, which results in treatment costs in "suspension zones."

Without active intervention, an invasion will eventually be detected when the outbreak becomes large enough. Implementing active surveillance will help to detect the bioinvasion early, therefore shortening the time and reducing the size of an outbreak. However, although surveillance cost is relatively known, upfront and ongoing, outbreak costs are contingent on an incursion probability λ, and the realization of various random factors as discussed earlier. A relevant policy question is whether it is worthwhile to implement active surveillance activities and, if so, at what level or intensity (i.e., how much to spend), so that the total cost of an outbreak, along with total damages and the cost of detecting it early, is the smallest. In this light, our surveillance optimization problem is formulated as follows:

$$\begin{aligned}\min_{g \geq 0} \text{TC} = & \underbrace{gs}_{\text{surveillance}} + \lambda \left[\sum_{t=t_{\text{detect}}}^{T_{\text{obrk}}} \left(\underbrace{\mathbf{D}'_{\mathbf{t}}(e + \mathbf{d})}_{\text{eradication + damages}} + \underbrace{\mathbf{Z}'_{\mathbf{t}} z}_{\text{suspension}} \right) \right. \\ & \left. + \underbrace{(\mathbf{D}'_{T_{\text{obrk}}} + \mathbf{Z}'_{T_{\text{obrk}}}) m}_{\text{management cost}} + \underbrace{(T_{\text{obrk}} + T_{\text{mkt}}) r}_{\text{revenueloss}} \right] \\ = & S(g) + \lambda [C(\Xi, g)] \equiv f(\Xi, g) \end{aligned} \tag{2}$$

subjectto

$$\begin{cases} \text{Transitionequation}: & \mathbf{X_t} = f(\boldsymbol{\Xi_t}; X_0; g) \text{ detailed in}() \\ \text{Initialcondition}: & \sum x_{i0} = 1 \\ \text{End} - \text{pointcondition}: & \sum x_{iT_{\text{obrk}}} = 0 \text{ where } 0 \leq T_{\text{obrk}} < \infty \end{cases}$$

where S is surveillance cost, depending on surveillance level or intensity g and surveillance marginal cost s. $\mathbf{D_t}$

and $\mathbf{Z_t}$ are $X \times 1$ vectors indicating whether at time t, inhabitable cells are in the eradication and suspension zones, respectively. The former encircles the infested cell on the radius of $r_{\text{eradication}}$ while the latter surrounds the former on the radius of $r_{\text{suspension}}$. Similarly, e and z are marginal (and average) eradication suspension costs, respectively, while $\mathbf{d}$ is a vector $X_t \times 1$ of cell-specific damage costs, all one-off during the entire outbreak. For simplicity, we assume that production damage is negligible before the invasion is detected at t_{detect}, and therefore, damages, together with eradication and suspension costs, are incurred only after detection. Furthermore, once infested, cells remain so until they are removed. A management cost to cover the coordination of response activities, if any, is applied to the entire outbreak at the unit cost of m per cell. That means it depends linearly on the size of the quarantine area that combines both eradication and suspension zones. Finally, revenue losses are proportional to the "business-as-usual" revenues at the rate r, and dependent on the length of outbreak and market closure periods (i.e., $T_{\text{obrk}} + T_{\text{mkt}}$).

The problem in Eq. 2 can be viewed as a two-stage optimization problem without recourse e (Shapiro 2003). The decision on surveillance intensity g, usually the size of the trapping grid or the distance between traps, has to be made regardless of incursion and before a possible outbreak unfolds. However, this decision has significant bearing on the realization of the random dispersal and detection and, therefore, the cost of an incursion $C(\Xi, g)$. The objective function, therefore, boils down to finding g, the correct point of early detection, so that the combination of the upfront deterministic cost of surveillance and the probabilisitic cost of an incursion, denoted as $f(\Xi, g)$, is the smallest.

Finally, it is worth noting that a discount rate is not needed in our model as we focus on equilibrium analysis. This approach has been used previously in the literature on optimal surveillance (e.g., Epanchin-Niell et al. 2012, Kompas et al. 2017).

Solution method, algorithm and parallel processing techniques

The model in (2) has many scenarios due to the realization of various random factors. Therefore, a closed-form solution is unlikely to be attainable. We use instead the SAA approach (Shapiro 2003) to obtain the optimal solution on average. Accordingly, the objective function in the model (2) is re-written in the form:

$$\min_{g \geq 0} TC = \underbrace{S(g)}_{\text{Surveillance cost}} + \underbrace{\lambda \mathbb{E}[C(\Xi, g)]}_{\text{Expected damages}} \equiv \underbrace{\mathbb{E}[f(\Xi, g)]}_{\text{Total expected cost of an outbreak}} \tag{3}$$

where $\mathbb{E}$ is the expectations operator. The SAA method minimizes the expected objective function in (3) approximately using a sample average estimate derived from a finite set of samples from the underlying probability space. The resulting sampled problem scenarios are then solved by deterministic optimization techniques. As SAA uses an exterior sampling method, in which the random matrix Ξ is realized outside the optimization routine, it separates sampling procedures and optimization techniques, therefore making it easy to implement and more efficient to solve (Shapiro 2003). More importantly, each sampled scenario has both treated (i.e., with policy interventions) and untreated (i.e., without any policy interventions) outcomes, allowing a precise measurement of the policy impact.

Underpinned by the Law of Large Numbers and the Central Limit Theorem, the SAA estimator has desired statistical properties. They include unbiasedness, consistency and asymptotic normality under some regularity conditions (see Shapiro [2003] for detailed proofs). Put differently, the SAA estimate of the optimal solution converges to the true solution value of the problem with probability 1 as the sample size $N \to \infty$ (Shapiro 2003). Finally, statistical inference for SAA is well developed, therefore allowing us to calculate error bounds for obtained solutions, validate them, and set-up stopping rules in our algorithm (see Kleywegt et al. 2002, Norkin et al. 1998, Mak et al. 1999, Verweij et al. 2003 for detailed proofs).

Regarding implementation, the SAA method involves a three-stage procedure. This procedure is repeated until convergence toward the true objective function value TC^* is achieved, as described in Algorithm 3.3 (Box 1).

Detailed spatial heterogeneity coupled with the need to keep track of AIS results is an incredibly large-dimensional problem. In this case, serial computing is not as efficient as needed. We, therefore, use a parallel processing algorithm to solve the problem. That is, several processes in several computers (cores) generate simultaneously many fractions of sample size N, N' and N'', and calculate lower and upper bound estimates across processes, which are then sent to a master process. Similarly, for the optimization process that is based on a direct search for the optimal point, we apply the same technique to simultaneously find the optimal solution candidate, and estimate the optimality gap.

Empirical Application and Results

Our surveillance model is applied to the case of PFF in Queensland (QLD) in Australia. The random dispersal model used in this application is an ABM, which is arguably the best model to present a stochastic spatial-dynamic dispersal process of invasive species. Having said that, it is worth noting that the model described in the section can be applied to other spread model types, and not necessarily just for an ABM.

This section starts with a description of the ABM model's architect. We then explain the choice of parameter values for both the dispersal process and economic costs. Given the large dimension involved in our analysis,

Box 1
Algorithm 1 The SAA procedure

Input: M independently and identically distributed (iid) samples of N iid training scenarios, a sample N' of iid testing scenarios, and a sample N'' of iid validating scenarios. Samples are independent from each other. N' and $N''$are much larger than N.

Stage 1:

- For each sample of N training scenarios in the set of M samples, estimate the minimum of average objective function value $\overline{TC}_{\mathrm{N}}^{\mathrm{m}} = \min_{\mathrm{g}} \frac{1}{\mathrm{N}} \sum_{\mathrm{n}=1}^{\mathrm{N}} \mathrm{TC}\left(\Xi_{\mathrm{n}}^{\mathrm{m}},\ \mathrm{g}\right)$, and the candidate optimal policy solution $\hat{\mathrm{g}}^{\mathrm{m}}$, where $m\varepsilon[1,M]$.
- The lower bound of the optimal value TC^{*} is the average of the estimated minimum function value of M samples, or $\overline{TC}_{\mathrm{N,M}} = \frac{1}{\mathrm{M}} \sum_{\mathrm{m}=1}^{\mathrm{M}} \overline{TC}_{\mathrm{N}}^{\mathrm{m}}$

Stage 2: Use the testing sample N' to find the best optimal solution $\hat{\mathrm{g}}^{*}$ among all $\hat{\mathrm{g}}^{\mathrm{m}}$, or

$$\hat{\mathrm{g}}^{*} \in \arg\min\left\{\frac{1}{\mathrm{N}'} \sum_{\mathrm{n}'=1}^{\mathrm{N}'} \mathrm{TC}(\Xi_{\mathrm{n}'},\ \hat{\mathrm{g}}) : \hat{\mathrm{g}} \in \left\{\hat{\mathrm{g}}^{1},\ \hat{\mathrm{g}}^{2},\ \ldots,\ \hat{\mathrm{g}}^{\mathrm{M}}\right\}\right\}$$

Stage 3:

- Use the validating sample N'' to find the upper bound of the optimal value TC^{*}, or

$$\widehat{TC}_{\mathrm{N}''}(\hat{\mathrm{g}}^{*}) = \frac{1}{\mathrm{N}''} \sum_{\mathrm{n}''=1}^{\mathrm{N}''} \mathrm{TC}(\hat{\mathrm{g}}^{*},\ \Xi_{\mathrm{n}''})$$

- Estimate the "optimality gap" or estimation error gap $(\hat{\mathrm{g}}^{*}) = \widehat{TC}_{\mathrm{N}''}(\hat{\mathrm{g}}^{*}) - \overline{TC}_{\mathrm{N,M}}$ to check whether the convergence is achieved.

Stop or increase the sample N and repeat the process until the desired optimality gap is achieved.

we describe our computing process, alongside the software packages used. Finally we present results and check their sensitivity subject to some key parameter values.

The research site, QLD, is under a constant threat of PFF as it is located near South-East Asia and Pacific Island Countries where PFF is native and widespread. With monsoonal winds in the wet season, PFF can "travel" through the Torres Strait Islands (TSI), just north of mainland Australia, to QLD, along environmental and human-assisted pathways. To date, an ongoing and strong surveillance and trapping program in the TSI has largely prevented the permanent establishment of PFF in the TSI and reduced the risk of flies moving south to mainland Australia, via QLD. Furthermore, local traps in QLD can potentially detect PFF early should it escape the quarantine zone in the TSI (DAFF 2005). However, the threat to QLD and Australia from PFF remains high with estimates of the damage of a country-wide spread of PFF of Australian Dollars (AUD) $3.3 billion (Hafi et al. 2013).

An agent-based PFF spread model

The outbreak in our model starts from an invasion of PFF migrating via the TSI, the most likely event. A PFF outbreak begins when fruit flies settle in a random cell within the incursion area. Once settled, they will gradually expand in a southerly direction given wind and weather patterns.

The architect of our agent-based PFF spread model has three main components: weather conditions, the environment and agents (Fig. 1). Weather conditions include data on seasonal features such as wind, temperature, soil moisture, and host suitability that affect PFF dispersal (Bateman 1967, Dominiak et al. 2003, Yonow et al. 2004). The environment spans 1.85 million km^2 with approximately 1.4 billion 50 m × 50 m raster cells in QLD. Its detailed land-use raster map, from ABARES (2015), provides us with information on six broad categories of land use, with up to as many as 60 different smaller land-use purposes in each category. Based on this map and the fact that PFF infests only horticultural crops, we classified the research area into non-habitable and habitable raster cells. Finally, agents in our model are propagules of PFF, each of which can consist of two (one male and one female) or more fruit flies to enhance their chance of survival and successful reproduction in a new host. The choice of propagules over individual

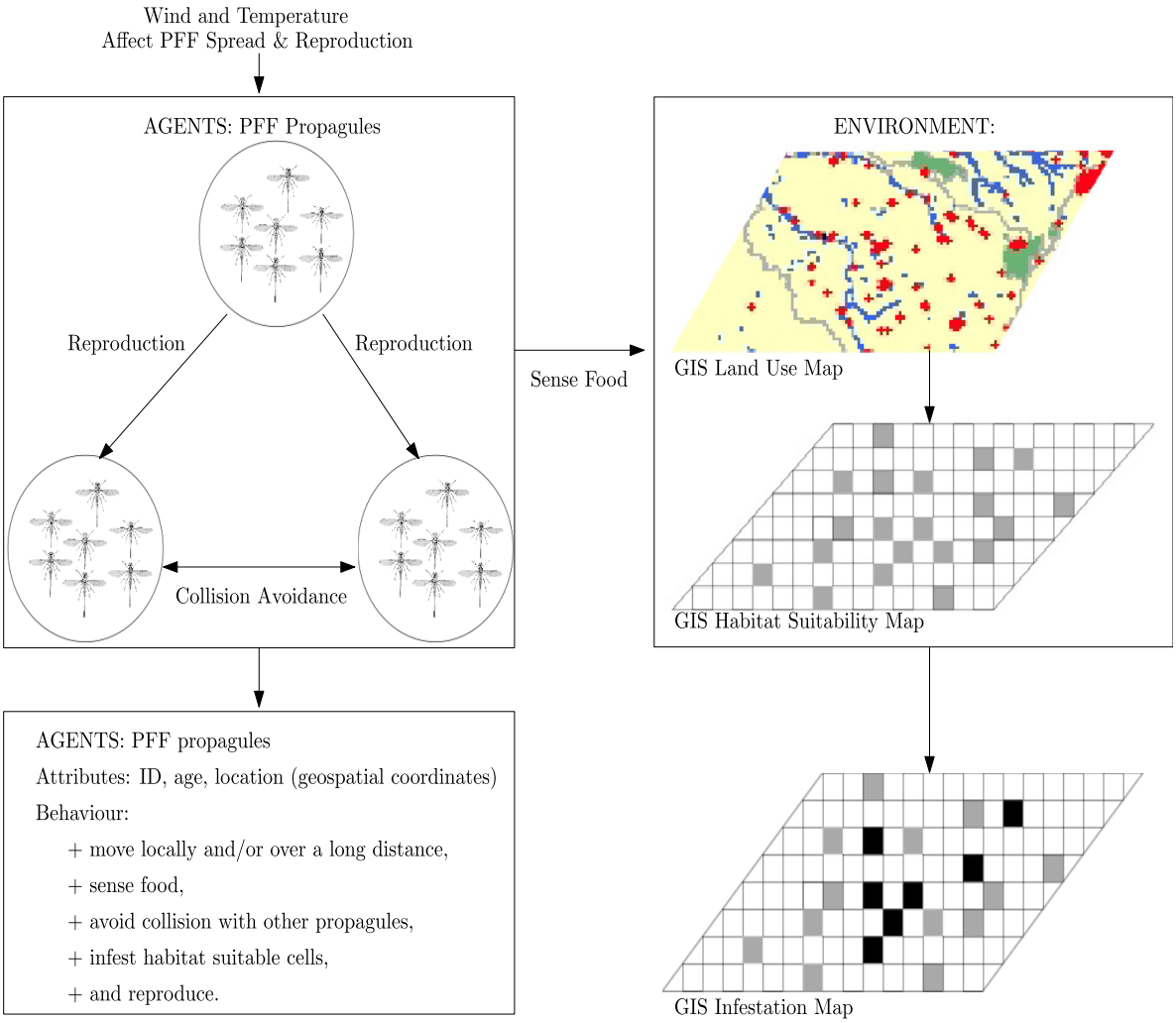
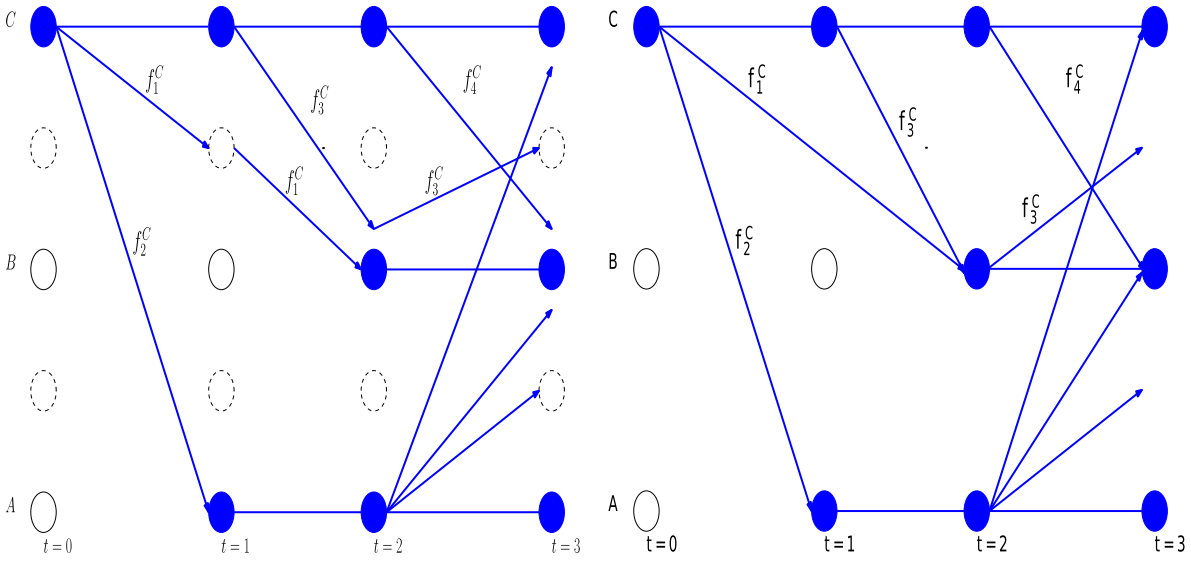


FIG. 1. Schematic of the Papaya Fruit Fly (PFF) spread model. GIS = Geographic Information System.

fruit flies as the model agent reduces unnecessary model complexity while remaining useful to represent the realistic dispersal of PFF. In particular, agents travel together and have their own specific age and geospatial location at each model time step (which is 1 week, to be in line with their life cycle).

There are two stages in the life of each agent. The early development stage lasts for four weeks when the agent develops from egg to larva and then pupa. During this stage, it stays "latent" at the host (Yonow et al. 2004). In the adult stage the agent can potentially reproduce, and more importantly migrate to other hosts (Yonow et al. 2004).

For an outbreak as a whole, PFF dispersal is characterized by three important factors, namely the range, the direction and the quantity of released agents (Adeva et al. 2012). Regarding the range, some agents can make a jump over a long distance of r_{jump} in the first week of their adult stage when they are the strongest and most active (Adeva et al. 2012, Dominiak 2012). We denote p_{jump} as the probability of an agent making such a jump. The agents that do not make a long jump but instead move locally within a range of r_{local}/week. Depending on whether an agent moves locally or over a long distance, the actual distance it makes is a random event following uniform distributions unif(0,r_{local}) or unif(0,r_{jump}), respectively. In terms of the direction, an agent making a long jump does so in any direction. After the long jump, it travels in a similar way as the ones that stay local. The movement direction of locally traveling agents depends on the proximity of a nearby host or food that they can sense (Adeva et al. 2012). We denote the probability of an agent finding a nearby host as β, a function of the distance between the two. When β is equal to zero, or all hosts are too far away to be detected, the agent will move randomly in any direction. Finally, the number of agents released from each host per week, π, depends on seasonal features such as wind, temperature, soil moisture, and host suitability (Bateman 1967, Dominiak et al. 2003, Yonow et al. 2004).



(a) With non-habitable nodes (b) Without non-habitable nodes

FIG. 2. Papaya Fruit Fly random network dispersal model. Solid circles represent hosts that are habitable for PFF while the broken ones are not. Starting from host C, which is random, agents such as f_1^C, f_2^C, f_3^C, etc. are released to find new hosts. From these new hosts, more agents are released. For example, agent f_2^C occupies host A in $t = 1$ and starts releasing new agents from $t = 2$. Finally, agents avoid cells that have already occupied by other agents. For example, agent f_3^C arrives at host B at $t = 2$, but leaves immediately as it has been occupied by agent f_1^C.

The dispersal process in our model can be represented in an extended network (Fig. 2). In panel (a), solid circles represent hosts or cells that are habitable for PFF while the broken ones are not. Starting from host C, agents are released to find new hosts, expanding their colony. It is worth noting that many flies can be born in host C, but not all of them will leave the host. To model the colony expansion, we only focus on those who leave as propagules and can form a new host for tractability. The departing or leaving agents can find a new host within one period of time (e.g., agent f_2^C), or they can land on a non-habitable cell and have to keep searching until they find a new host (e.g., agent f_1^C). If a host has already been occupied, an arriving agent has to leave immediately and continue their search for a new host in the next period to avoid "collision" (e.g., agent f_3^C). The reason is that eggs are inserted directly into the host fruit and once larvae starts to feed, an unidentified change occurs in the fruit, which generally causes females to avoid it (Waterhouse and Sands 2001). Our dispersal model can be seen more easily in panel (b), where all non-habitable nodes and "unsuccessful" connections (i.e., connections to/from non-habitable cells) are hidden, and therefore, a node at a particular time step t can contact multiple nodes at the following time steps within the length of a PFF adult stage.

Parameter values of the PFF dispersal process

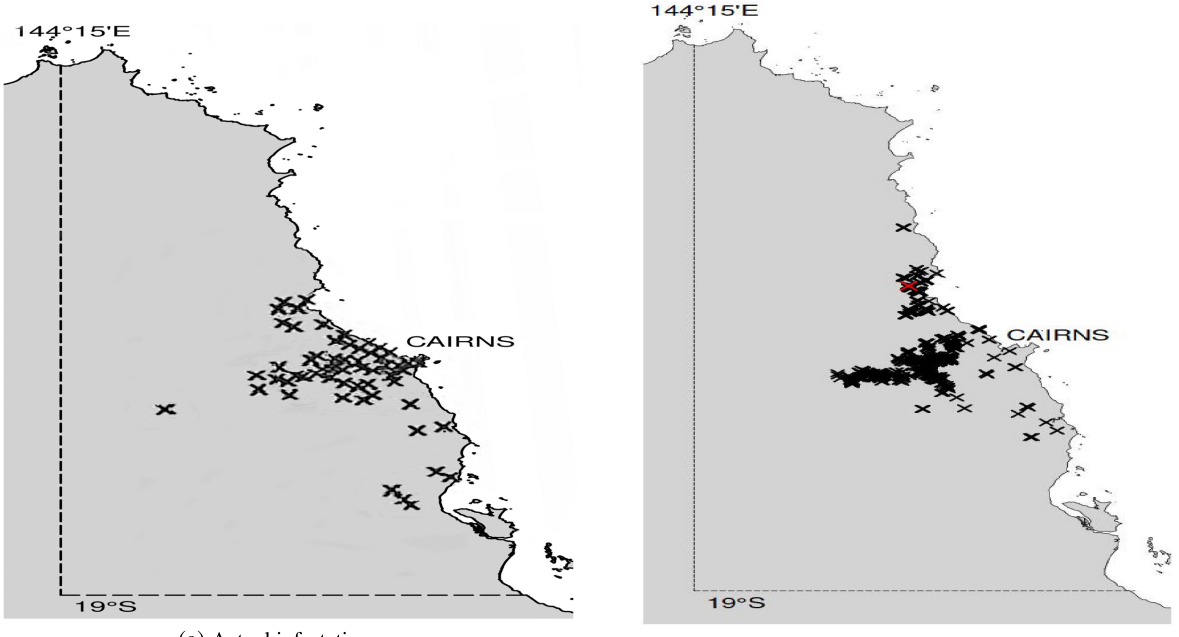
Parameter values for the random dispersal model are largely drawn from the literature (Table 1). In particular, the outbreak arrival rate λ is one in every five years (Kompas and Che 2009) while the lifespan of an agent is $A = 10$ weeks, based on Bateman (1967); Yonow et al. (2004); Adeva et al. (2012). Regarding the range of dispersal, the probability of an agent making a long jump p_{jump} is 0.3 based on the dispersal distribution in Adeva et al. (2012) while the maximum range of local and distance travel are $r_{\text{local}} = 1.4$ km/week and $r_{\text{jump}} = 94$ km in the first week, respectively, based on Adeva et al. (2012); Dominiak (2012). For the direction of dispersal, we adopt the probability β of an agent finding a nearby host from Adeva et al. (2012, p. 101). Accordingly, an agent will detect a habitable host within 0.1 km with certainty. But the further away a habitable host is, the less certain an agent can detect it. When habitable hosts are located beyond 3 km far away, an agent cannot detect it, thereby having to keep moving randomly in the direction until it finds a habitable host, or die due to the lack of food or reaching the end of its life (details in Appendix S1).

It is vital to calibrate correctly the quantity π of propagules released from each infested cell as this parameter largely determines the extent of PFF spread, therefore the size of a PFF outbreak. Therefore, its calibration deserves special attention and is made based on two sources of information. First, we use historical data on the spread of the first PFF outbreak in north Queensland in 1995 (Fay et al. 1997, p. 260b). It is widely believed that PFF were present for 12–15 months before a massive eradication campaign commenced (Cantrell et al. 2002). In this light, we let PFF disperse freely (undetected) for 16 months in our simulations using our

TABLE 1. Model parameterization.

Parameters	Description	Unit	Value
Parameters of random dispersal			
λ	Outbreak incursion probability[(a)]	Per year	0.2
A	Life span of a PFF propagule[(b)]	Week	10
p_{jump}	Probability of an agent making a long jump[(c)]		0.3
r_{jump}	Maximum range of distance travel[(d)]	km/1st week	94
r_{local}	Maximum range of local travel[(c)]	km/week	1.4
β	Probability of a PFF to find a nearby host[(c)]		Appendix S1
π	No. propagules released from an infested cell[(e)]	No./week	2
γ	Probability of an infested cell detected naturally[(f)]		=1 if six/more months; 0 otherwise
Economic costs			
$r_{eradication}$	Radius of the eradication zone[(d1)]	km	1.5
$r_{suspension}$	Radius of the suspension ring[(d2)]	km	13.5
e	Eradication cost[(a)]	\$/km^2	539
d	Damage cost	\$/cell	Appendix S3
z	Suspension cost[(i)]	\$/ton	143
m	Management cost[(g)]	\$/cell	114
r	Weekly trade-related revenue loss[(g)]	\$ mil/year	25
T^{mkt}	Market closure period[(h)]	Month	8.5
s	Marginal cost of surveillance		Appendix S4

Notes: All parameter values are in Australian Dollars (AUD) in the years of their respective sources. They are converted into 2015 AUD in our computation to generate model results. [(a)]Kompas and Che (2009); [(b)]Bateman (1967), Yonow et al. (2004) & Adeva et al. (2012); [(c)]Authors' assumption based on Adeva et al. (2012); [(d1)]Dominiak (2007); [(d2)]Dominiak (2012); The South Australian Government Gazette (2020); [(e)]Authors' calibration based on historical data of the first outbreak in 1995 (Fay et al. 1997, p. 260b) and seasonal patterns & Atlas of Living Australia (2015) described in Appendix S2; [(f)]Authors' assumption; [(g)]Authors' calculation based on Cantrell et al. (2002); [(h)]Underwood (2007); [(i)]Authors' calculation from (Hafi et al. 2013).



(a) Actual infestations (b) Simulated infestations

FIG. 3. (a, b) Papaya Fruit Fly outbreak: actual vs. a simulated medium-sized outbreak. The actual infestations are from the outbreak that occurred in north Queensland in November 1995 (Fay et al. 1997, p. 260b). The simulated infestations are generated by our model.

random dispersal model, and compare the simulated infestations with the historical snapshot data. Panel (b) in Fig. 3 shows infested raster cells in a medium-sized outbreak among our 100 simulation runs. With two propagules/week, on average, released from an infested cell, there is a visual correspondence between our simulated spread in panel (b) and the actual spread in panel (a). The second source of information is the monthly occurrence records of fruit flies (*Bactrocera* (*Bactrocera*) *tryoni*) in Queensland during 1950–2015 (Atlas of Living Australia 2015). As PFF and *Bactrocera tryoni* belong to the same genus, *Bactrocera*, they share some common biological characteristics including seasonal patterns of incursion and migration. For this reason, we can use information on *Bactrocera tryoni* as a proxy to estimate seasonal patterns of PFF. These estimated seasonal patterns are then incorporated into our calibration of the number of propagules released per week (details in Appendix S2).

Finally, the probability γ of an infested cell being detected naturally is 1 if the cell is infested for more than six months, and zero otherwise. The reason is two-fold. First, PFF is hard to detect. Its eggs are inserted directly into the host fruit well before it ripens, and the rapidly growing tissues quickly cover any marks made by the fruit fly, making it difficult for all but the trained eye to see where eggs have been laid (Cantrell et al. 2002). As enlisting "trained eyes" has never been used as a measure to detect PFF early in practice, we consider it exogenous. As PFF makes infested fruit look ripe earlier, this will get noticed by growers, therefore being eventually detected. Second, it would generally take horticultural crops in Queensland about six months to ripen.

Economic costs

Parameter values for costs are also largely drawn from the literature and Australian biosecurity regulations (Table 1). Specifically, the radius to define the eradication zone $r_{\text{eradication}}$ is 1.5 km (Dominiak 2007) with an eradication cost e of \$539/km^2, largely to cover labor and chemicals (Kompas and Che 2009). Also, in this zone, the production damage $\mathbf{d}$ is cell-specific and accounts for 45% of the cell production value (Kompas and Che 2009) which is estimated using the data on land use (ABARES 2015) and production value (ABS 2011) (details in Appendix S3). Conversely, the radius to define the suspension ring $r_{\text{suspension}}$, which encircles the eradication zone, is 13.5 km (Dominiak 2012, The South Australian Government Gazette 2020) with a suspension cost z of \$143 per ton, based on Hafi et al. (2013). Management cost m is \$114/cell, while revenue losses r are \$25 million/year, both estimated based on the actual cost of the outbreak in 1995 in Cantrell et al. (2002). The market closure T^{mkt} of 8.5 months was applied to all scenarios following Underwood (2007), while the outbreak time T_{obrk} depended on how each simulated scenario is realized. Nonetheless, if the realized outbreak time was longer than 15 months, we adopted the eradication strategy in 1995 (Cantrell et al. 2002) to change the radius of the suspension zone to 80 km to ensure that the outbreak will be terminated. Finally, adopted from Florec et al. (2010), the ongoing surveillance cost depends on the surveillance intensity g, and the marginal cost s that covers the cost for labor, equipment and materials, as well as the time required for workers to check traps and travel between them (details in Appendix S4).

Computing process

We used 12 processes over three quad core CPU computers with Hyper-Threading to generate numerical results. Such a parallel processing method helps increase the possible simulation numbers in our computional system by 12-fold, compared with a similar uni-processing process. Algorithm 3.3 is executed until the optimality gaps are stabilized at <1% (Table 1). To achieve this optimality gap, we increase the training sample size N gradually to 672 while keeping the number of training samples M constant at 50. It is worth noting that the convergence to the true solution under the LLN depends on the sample size N, not the number of samples M. To this end, the number of simulations in the first stage grows to 33,600. In addition, the sample sizes N' and N'' used to find the candidate optimal solution and check its quality in the second and third stages remain, remaining constant at 33,600.

Numerical results and ploting in this paper are obtained using C, and R (R Core Team 2014). For R, in particular, we use the following packages: *maptools* 0.8-36 (Bivand and Lewin-Koh 2015), *raster* 2.3-40 (Hijmans 2015), *rasterVis* 0.31 (Perpiñán and Hijmans 2014), *rgdal* 0.9-2 (Bivand et al. 2015), *sp* 1.1.0 (Pebesma and Bivand 2005, Bivand et al. 2013), *plotrix* 3.5.11 (Lemon 2006), and *ggplot2* 1.0.1 (Wickham 2009).

Results

Numerical results of one sample of N training scenarios in Stage 1 are illustrated in Fig. 4. Other samples have similar patterns and, therefore, are not presented for brevity. The vertical axis shows the total expected cost and its components, which include surveillance cost and expected damages. In addition, the horizontal axis shows surveillance intensity that is the PFF trap grid size or the distance between traps. The smaller is the grid size, the more traps will be laid and the more intense will be the surveillance activity. As can be seen, there is a clear trade-off between spending on early detection and the total expected cost of an outbreak. When the surveillance grid is much smaller than optimal (and therefore a larger number of traps), the total expected cost is largely driven by surveillance cost. In contrast, when the grid is more extensive than optimal, the expected damages dominate the total expected cost. Both cases are, more importantly, not justified on economic grounds.

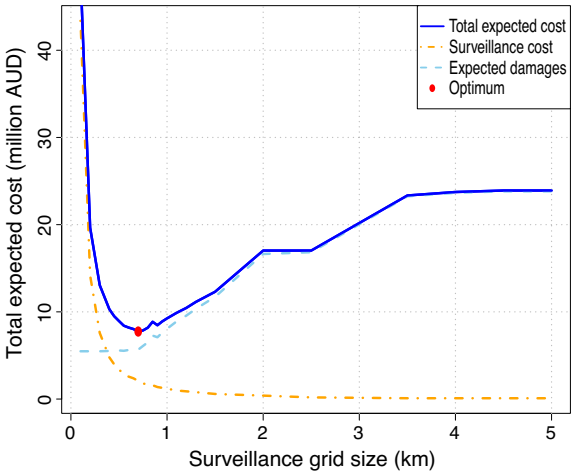


FIG. 4. Total expected cost of an outbreak and its components including surveillance and expected damage against surveillance intensity.

Results of all sampled scenarios are summarized in Table 2. The optimal trap grid size, $\hat{g}^*$, is 0.7 km, which is equivalent to roughly 6,782 traps to be laid. With this level of surveillance intensity, the minimized total expected outbreak cost is estimated to range approximately from \$7.73 million (the lower bound) to \$7.75 million (the upper bound). The quality of our estimation or the convergence of our estimates to the true optimal value can be seen in the optimality gap (i.e., the difference between the two bounds) which is <1% of the lower bound value. Furthermore, the gap is not statistically significant from zero at the 5% level.

We next compare our optimal surveillance level with the currently implemented policy in Queensland, which is apparently effective given that no sizable outbreaks have been detected since 1995. Fig. 5 shows that the traps currently laid are less dense, at the grid size of roughly 5 km, compared with those at the optimal (economic) level of 0.7 km suggested by our results. Although the current policy implies a much lower surveillance cost per year, compared with the \$2.08 million under our optimal policy, the expected total outbreak cost under the current scenario is \$23.92 million, much higher than the optimal cost of roughly \$7.74 million.

Finally, our results support but differ from the existing literature in a similar context. Specifically, the aggregate and deterministic model of surveillance against PFF by Kompas and Che (2009) for the whole of Australia suggests a 60% increase in surveillance expenses against PFF, or a trap grid size of ~3 km. This difference highlights the need for considering spatial heterogeneity and randomness in optimization models of surveillance against AIS.

Sensitivity analysis

This subsection checks whether our results are sensitive to key parameter values. It is worth noting that our model, unlike deterministic ones, fully incorporates randomness. Therefore, changes in parameter values will alter the entire data generating processes for simulations. New solution estimates, therefore, reflect the changes in parameter values, not the randomness.

Here we only focus on five key parameters that are hard to estimate precisely but are instrumental in determining model results. They are the incursion probability λ, the number of agents released from each infested cell π, the production loss rate to calculate cell-specific damages $\mathbf{d}$, the management cost m and the market closure period T^{mkt}. Other parameters have a relatively minor impact, are determined by regulatory requirements or are well established in the literature, and therefore are not discussed here. Detailed results are available upon request.

The optimal surveillance policy and outbreak costs for different incursion probabilities are presented in

TABLE 2. Estimated Papaya Fruit Fly total expected outbreak costs and optimality gaps.

	N (the number of training scenarios)								
Optimality indicators	48	144	240	336	432	528	576	624	672
The optimal trap grid size (km)	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
The lower bound of the true optimal value, \$AU million (A)	7.656 (0.024)	7.714 (0.015)	7.730 (0.011)	7.743 (0.009)	7.735 (0.007)	7.745 (0.007)	7.755 (0.009)	7.744 (0.009)	7.736 (0.006)
The upper bound of the true optimal value, \$AU million (B)	7.749 (0.007)	7.749 (0.007)	7.739 (0.006)	7.742 (0.006)	7.749 (0.006)	7.744 (0.006)	7.746 (0.006)	7.747 (0.007)	7.742 (0.006)
Optimality gap, \$AU million (C = B−A)	0.093 (0.031)	0.035 (0.021)	0.009 (0.017)	−0.002 (0.016)	0.013 (0.014)	−0.001 (0.013)	−0.009 (0.015)	0.003 (0.015)	0.006 (0.012)
Gap as percentage of the lower bound D = (C/A) × 100%	1.216	0.449	0.123	−0.019	0.174	−0.014	−0.112	0.045	0.078

Notes: $M = 50$; $N' = N'' = 33{,}600$; Number of processes = 12. Values in AUD million (2015). Standard errors in parentheses. All estimates are significant at the 1% level except for the ones of the optimality gap which, as expected, are statistically insignificant at the 5% level.

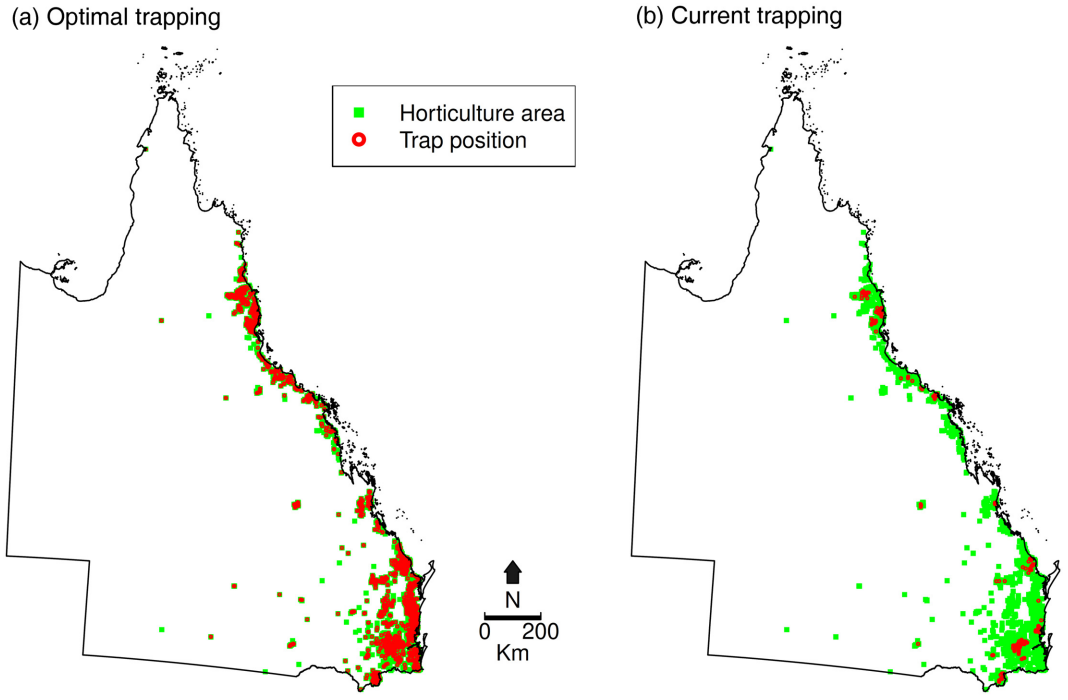


FIG. 5. Papaya Fruit Fly traps in Queensland: optimal vs. current. Panel (a) presents the optimal trap density generated by our model while panel (b) illustrates the current trap density.

Table 3. Clearly, if the incursion probability is zero, there is no need for surveillance. But the higher the probability is, the higher the total expected cost would be. Indeed, if the probability is doubled (i.e., $\lambda = 0.4$), the expected total outbreak cost is estimated to increase by 80%, from ~$7.4 million to $13.4 million. In contrast, if the probability is halved (i.e. $\lambda = 0.1$), the expected total outbreak cost is estimated to decrease by 35%, from ~$7.4 million to $3.8 million. Despite this big variation in the optimal cost, there is little sensitivity in the optimal surveillance policy. In particular, the optimal surveillance level is 0.9 km if $\lambda = 0.05$ and reduces slightly to 0.75 km if $\lambda = 0.1$ For the most likely range of the incursion probability ($\lambda\varepsilon[0.15,0.5]$), the optimal surveillance level remains highly stable at 0.7 km. This lack of sensitivity is because the benefit of early detection depends on how much benefit is gained when an early detection system is put in place or how fast the outbreak grows undetected. This, in turn, depends on the biological characteristics of the species and the environment. The incursion probability will not change that; rather, it relates more to the trade-off between the cost of the surveillance system and the benefit of early detection.

Fig. 6a reveals some sensitivity in the optimal surveillance policy outcome when π, the number of agents released from each infested cell, varies. More agents being released makes them easier to detect but also extends the outbreak size, therefore its cost. As these effects move in two opposite directions, the result is a small reduction in the optimal level of surveillance.

Changes in the market closure period T^{mkt} do alter the expected outbreak cost (Fig. 6b), but in a relatively linear way. That is, the longer the market closure period, the larger would be the expected outbreak cost. But the optimal surveillance level remains unchanged. Finally, there is little sensitivity in the production loss factor (Fig. 6c) as well as management cost (Fig. 6d).

DISCUSSION

AIS causes an extensive economic loss and is the second most ranked threat to biodiversity. Reducing its damage has become one of the United Nations Millennium Development Goals (Gurevitch and Padilla 2004, Butchart et al. 2010). Spending on pre-border and border measures alone is often not economically effective as complete prevention is impossible and prevention measures cannot keep up with the increasing risks of a bioinvasion due to globalization and new or enhanced environmental pathways. In this context, added priority is needed to identify optimal surveillance intensity that takes into account both ongoing surveillance costs, expected damages and management costs.

Contribution to the literature and policy implications

Contributing to the growing body of literature on surveillance, we propose and solve a bioeconomic model to find the optimal surveillance for the early detection of

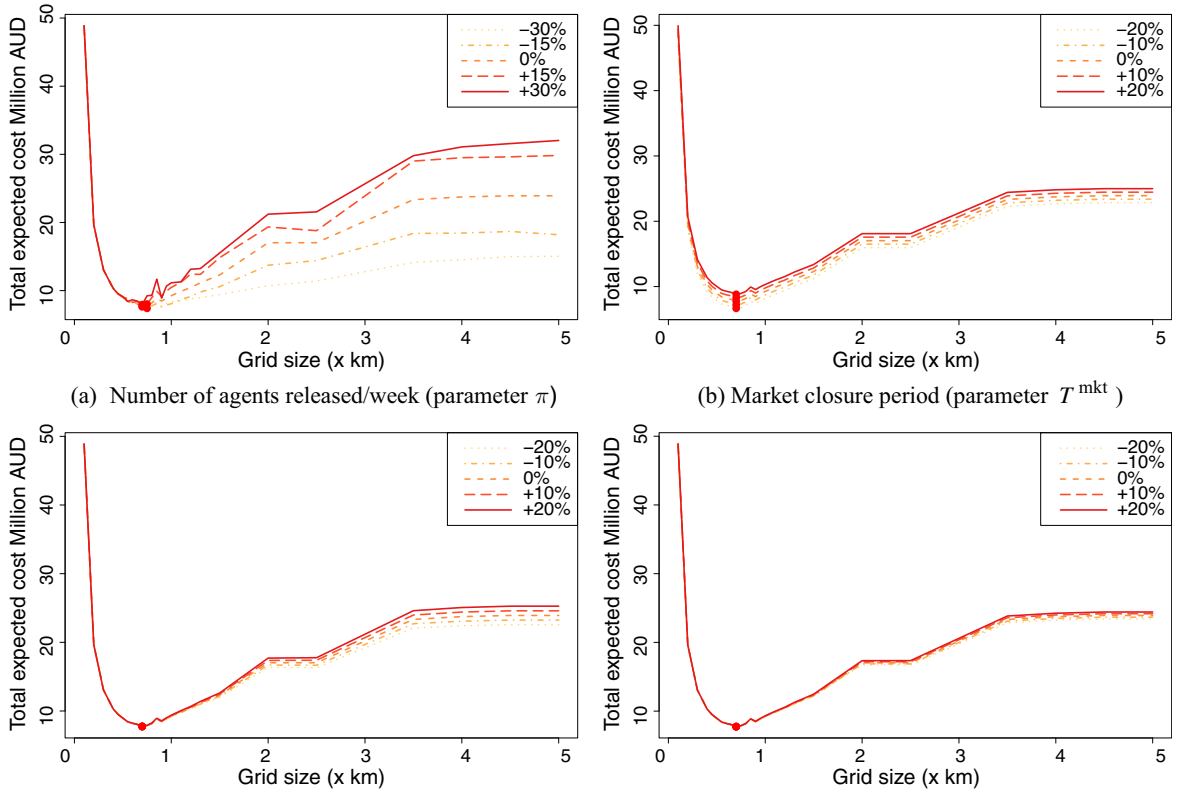
TABLE 3. Sensitivity analysis of the optimal results when the outbreak incursion probability or arrival varies.

	N (the number of training scenarios)								
Optimality indicators	48	144	240	336	432	528	576	624	672
Outbreak incursion probability λ = 0.05									
Optimal trap grid size (km)	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9
Lower bound of the true optimal value, $AU million (A)	3.057 (0.022)	3.094 (0.015)	3.129 (0.008)	3.146 (0.010)	3.143 (0.008)	3.156 (0.008)	3.146 (0.007)	3.147 (0.006)	3.149 (0.007)
Upper bound of the true optimal value, $AU million (B)	3.152 (0.007)	3.155 (0.007)	3.160 (0.007)	3.167 (0.007)	3.169 (0.007)	3.168 (0.007)	3.156 (0.007)	3.160 (0.007)	3.162 (0.007)
Optimality gap (C = B−A)	0.095 (0.029)	0.061 (0.022)	0.031 (0.015)	0.021 (0.017)	0.026 (0.015)	0.012 (0.014)	0.009 (0.014)	0.013 (0.013)	0.013 (0.014)
Gap as percentage of the lower bound D = (C/A) × 100%	3.099	1.965	0.994	0.666	0.841	0.377	0.300	0.408	0.424
Outbreak incursion probability λ = 0.10									
Optimal trap grid size (km)	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75
Lower bound of the true optimal value, $AU million (A)	4.738 (0.025)	4.794 (0.015)	4.813 (0.011)	4.826 (0.011)	4.821 (0.009)	4.839 (0.009)	4.833 (0.008)	4.834 (0.007)	4.831 (0.008)
Upper bound of the true optimal value, $AU million (B)	4.832 (0.007)	4.827 (0.007)	4.829 (0.007)	4.839 (0.007)	4.843 (0.007)	4.841 (0.008)	4.831 (0.007)	4.834 (0.007)	4.835 (0.007)
Optimality gap (C = B−A)	0.094 (0.033)	0.034 (0.022)	0.016 (0.018)	0.014 (0.019)	0.021 (0.016)	0.003 (0.016)	−0.002 (0.015)	0.000 (0.015)	0.004 (0.016)
Gap as percentage of the lower bound D = (C/A) × 100%	1.985	0.701	0.326	0.283	0.444	0.059	−0.046	0.001	0.076
Outbreak incursion probability λ = 0.15									
Optimal trap grid size (km)	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
Lower bound of the true optimal value, $AU million (A)	6.214 (0.023)	6.280 (0.014)	6.295 (0.010)	6.307 (0.009)	6.303 (0.008)	6.318 (0.007)	6.321 (0.008)	6.317 (0.006)	6.316 (0.005)
Upper bound of the true optimal value, $AU million (B)	6.332 (0.005)	6.332 (0.005)	6.324 (0.005)	6.326 (0.005)	6.332 (0.005)	6.328 (0.005)	6.330 (0.005)	6.330 (0.005)	6.327 (0.005)
Optimality gap (C = B−A)	0.117 (0.028)	0.051 (0.018)	0.029 (0.015)	0.019 (0.014)	0.029 (0.012)	0.010 (0.012)	0.009 (0.013)	0.013 (0.011)	0.011 (0.010)
Gap as percentage of the lower bound D = (C/A) × 100%	1.888	0.815	0.464	0.306	0.453	0.158	0.145	0.212	0.173
Outbreak incursion probability λ = 0.40									
Optimal trap grid size (km)	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
Lower bound of the true optimal value, $AU million (A)	13.311 (0.036)	13.379 (0.025)	13.399 (0.019)	13.421 (0.016)	13.405 (0.014)	13.420 (0.012)	13.430 (0.018)	13.410 (0.017)	13.393 (0.012)
Upper bound of the true optimal value, $AU million (B)	13.418 (0.013)	13.417 (0.013)	13.398 (0.012)	13.403 (0.013)	13.417 (0.013)	13.408 (0.013)	13.412 (0.013)	13.414 (0.013)	13.405 (0.012)
Optimality gap (C = B−A)	0.107 (0.049)	0.038 (0.038)	−0.001 (0.031)	−0.018 (0.029)	0.012 (0.026)	−0.012 (0.025)	−0.017 (0.030)	0.004 (0.030)	0.012 (0.024)
Gap as percentage of the lower bound D = (C/A) × 100%	0.806	0.282	−0.006	−0.133	0.091	−0.088	−0.129	0.031	0.091
Outbreak incursion probability λ = 0.50									
Optimal trap grid size (km)	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
Lower bound of the true optimal value, $AU million (A)	16.124 (0.042)	16.204 (0.030)	16.225 (0.023)	16.258 (0.020)	16.236 (0.017)	16.254 (0.016)	16.267 (0.022)	16.243 (0.021)	16.221 (0.015)
Upper bound of the true optimal value, $AU million (B)	16.253 (0.016)	16.251 (0.016)	16.227 (0.015)	16.234 (0.016)	16.251 (0.016)	16.240 (0.016)	16.246 (0.016)	16.248 (0.016)	16.236 (0.015)
Optimality gap (C = B−A)	0.128 (0.059)	0.047 (0.046)	0.002 (0.038)	−0.024 (0.036)	0.015 (0.033)	−0.015 (0.031)	−0.022 (0.038)	0.005 (0.037)	0.015 (0.030)
Gap as percentage of the lower bound D = (C/A) × 100%	0.794	0.290	0.012	−0.149	0.094	−0.091	−0.133	0.032	0.094

Notes: $M = 50$; $N' = N'' = 33{,}600$; Number of processes = 12. Values in AUD million (2015). Standard errors in parentheses. Setting the optimal gap aside, all estimates are significant at the 1% level.

an AIS. Specifically, we extend the current literature by explicitly modeling spatial attributes and randomness in full detail. The numerical solution of a large-dimensioned problem is made possible due to a novel combination of the SAA approach and parallel processing techniques. Our study is the first, to the best of our knowledge, that uses this solution method in environmental and biosecurity economics.

We apply our model to the case of a potential entry of PFF in Queensland. For this AIS, time, space and randomness, all play vital roles in defining its damage and, therefore, should be fully considered. The lack of



(a) Number of agents released/week (parameter π)

(b) Market closure period (parameter T^{mkt})

(c) Production loss factor

(d) Management cost (pamater m)

FIG. 6. (a–d) Sensitivity analysis of some key parameter values.

complete prevention, coupled with massive damage caused by PFF, makes early detection and eradication indispensable to reduce its potential impact, a shared context of many bioinvasions.

Our PFF spread model is agent based with parameters calibrated using a detailed land-use spatial map, historical outbreak snapshot data and seasonal patterns of fruit fly spread. Results from our optimal surveillance model suggest that the current surveillance against PFF in Queensland, although effective, is below what would be economically optimal. These results not only have a direct bearing on Australia, a key agricultural crop exporting country, but also raise questions on whether surveillance measures are appropriate in many parts of the world where fruit flies are a significant threat to horticultural crops. For example, the potential damage caused by the Mediterranean fruit fly (*Ceratitis capitata*) in the United States is estimated to be roughly $15 billion alone, if left uncontrolled (Hagen et al. 1981), while the annual damage from fruit flies is worth millions of dollars in Africa (National Research Council 1992).

Model applicability, limitations and future research

Because of the challenge in optimizing a problem that involves uncertainty and spatial dynamics, our research has at least three limitations that are worth addressing in future research.

First, to enhance tractability, our model applies only to bioinvasions that have to be eradicated swiftly so that the time horizon is finite. A missing feature and an important future improvement in this work would be to allow for different response strategies such as limited or spatial containment. A model framework that considers such a feature has been proposed but only for a relatively small-dimensional problem, and either in a deterministic setting (Epanchin-Niell and Wilen 2012) or one with a limited range of stochasticity (Chalak et al. 2017). Also, to enhance tractability, we only consider the cases in which cells, once infested, remain so until being detected and removed. We therefore exclude cases in which infested cells "wink on and off" on their own. With more computational power, future research could incorporate more heterogeneity in the policy response as well as a wider class of bioinvasions to aid decision-making, while retaining the fundamentally stochastic and spatial-dynamic nature of the problem.

The second limitation concerns a few of the parameter values in the empirical models in general. An example is the detection value that is set as a binary in our application. With further information, potentially estimated from monitoring data, this can be improved.

Finally, future research should consider enlisting the "trained eye," as noted, to spot potentially damaged fruit, however difficult, as an option to speed up early detection of PFF and therefore partially mitigate the damage. At this moment, lack of information on the costs and benefits of this measure compels us to exclude this case, and not enlist it as a policy option.

Acknowledgments

We thank the Subject Matter Editor and two reviewers for their helpful comments and advice, which have greatly improved the paper. Funding from the Australian Department of Agriculture and Water Resources and the Centre of Excellence for Biosecurity Risk Analysis at the University of Melbourne (Projects 1405C and 1608A) is also gratefully acknowledged.

Literature Cited

ABARES. 2015. Catchment scale land use of Australia - Update March 2015. Australian Bureau of Agricultural and Resource Economics and Sciences, Canberra, Australia. http://data.daff.gov.au/anrdl/metadata_files/pb_luausg9abll20150415_11a.xml

ABS. 2011. Agricultural commodities, Australia, 2010-2011. Australian Bureau of Statistics. http://www.abs.gov.au. Accessed 21 January 2014.

Adeva, J. G., J. Botha, and M. Reynolds. 2012. A simulation modelling approach to forecast establishment and spread of *Bactrocera* fruit flies. Ecological Modelling 227:93–108.

Albers, H. J., C. Fischer, and J. N. Sanchirico. 2010. Invasive species management in a spatially heterogeneous world: effects of uniform policies. Resource and Energy Economics 32:483–499.

Atlas of Living Australia. 2015. *Bactrocera* (*Bactrocera*) *tryoni*. http://biocache.ala.org.au/occurrences/search?q=Bactrocera+(Bactrocera)+tryoni. Accessed 25 June 2015.

Barbati, M., G. Bruno, and A. Genovese. 2012. Applications of agent-based models for optimization problems: a literature review. Expert Systems with Applications 39:6020–6028.

Barrett, S., P. Whittle, K. Mengersen, and R. Stoklosa. 2010. Biosecurity threats: the design of surveillance systems, based on power and risk. Environmental and Ecological Statistics 17:503–519.

Bateman, M. 1967. Adaptations to temperature in geographic races of the Queensland fruit fly *Dacus* (*Strumenta*) *tryoni*. Australian Journal Zoology 15:1141–1161. https://www.publish.csiro.au/zo/zo9671141

Bivand, R., T. Keitt, and B. Rowlingson. 2015. rgdal: bindings for the geospatial data abstraction library. R package version 0.9-2. https://cran.r-project.org/web/packages/rgdal/rgdal.pdf

Bivand, R., and N. Lewin-Koh. 2015. maptools: tools for reading and handling spatial objects. R package version 0.8-36. https://cran.r-project.org/web/packages/maptools/maptools.pdf

Bivand, R. S., E. Pebesma, and V. Gomez-Rubio. 2013. Applied spatial data analysis with R. Second edition. Springer, New York, New York, USA.

Bonneau, M., J. Martin, N. Peyrard, L. Rodgers, C. M. Romagosa, and F. A. Johnson. 2019. Optimal spatial allocation of control effort to manage invasives in the face of imperfect detection and misclassification. Ecological Modelling 392:108–116.

Butchart, S. H., et al. 2010. Global biodiversity: indicators of recent declines. Science 328:1164–1168.

Cantrell, B., B. Chadwick, and A. Cahill. 2002. Fruit fly fighters: eradication of the Papaya Fruit Fly. CSIRO PUBLISHING, Clayton, Victoria, Australia.

Chalak, M., M. Polyakov, and D. J. Pannell. 2017. Economics of controlling invasive species: a stochastic optimization model for a spatial-dynamic process. American Journal of Agricultural Economics 99:123–139.

Costello, C., and C. McAusland. 2003. Protectionism, trade, and measures of damage from exotic species introductions. American Journal of Agricultural Economics 85:964–975.

DAFF. 2005. Papaya fruit fly. https://www.business.qld.gov.au/industry/agriculture/land-management/health-pests-weeds-diseases/weeds-and-diseases/surveillance-programs/papaya-fruit-fly. Accessed 26 June 2015.

DeAngelis, D. L., and V. Grimm. 2014. Individual-based models in ecology after four decades. F1000Prime Reports 6:39.

Dominiak, B. 2007. Fruit fly eradication in the fruit fly exclusion zone: questions and answers. NSW DPI, Prime fact 552. https://www.dpi.nsw.gov.au/biosecurity/insect-pests/qff

Dominiak, B. C. 2012. Review of dispersal, survival, and establishment of *Bactrocera tryoni* (Diptera: Tephritidae) for quarantine purposes. Annals of the Entomological Society of America 105:434–446.

Dominiak, B. C., A. E. Westcott, and I. M. Barchia. 2003. Release of sterile Queensland fruit fly, *Bactrocera tryoni* (Froggatt) (Diptera: Tephritidae), at Sydney, Australia. Australian Journal of Experimental Agriculture 43:519–528.

Epanchin-Niell, R. S., E. G. Brockerhoff, J. M. Kean, and J. A. Turner. 2014. Designing cost-efficient surveillance for early detection and control of multiple biological invaders. Ecological Applications 24:1258–1274.

Epanchin-Niell, R. S., R. G. Haight, L. Berec, J. M. Kean, and A. M. Liebhold. 2012. Optimal surveillance and eradication of invasive species in heterogeneous landscapes. Ecology Letters 15:803–812.

Epanchin-Niell, R. S., and A. Hastings. 2010. Controlling established invaders: integrating economics and spread dynamics to determine optimal management. Ecology Letters 13:528–541.

Epanchin-Niell, R. S., and J. E. Wilen. 2012. Optimal spatial control of biological invasions. Journal of Environmental Economics and Management 63:260–270.

Fay, H., R. Drew, and A. Lloyd. 1997. The eradication program for papaya fruit fly (*Bactrocera papayae* Drew and Hancock) in North Queensland. Pages 259–261 *in* A. J. Allwood and R. A. I. Drew, Editors. Management of Fruit Flies in the Pacific. A Regional Symposium, Nadi, Fiji 28–31 October 1996. No. 76. 267pp in ACIAR Proceedings. ACIAR Proceedings. Australian Centre for International Agricultural Research, Canberra, Australia.

Finnoff, D., J. F. Shogren, B. Leung, and D. Lodge. 2007. Take a risk: Preferring prevention over control of biological invaders. Ecological Economics 62, 216–222, special Section: Ecological-economic modelling for designing and evaluating biodiversity conservation policies.

Florec, V., R. Sadler, and B. White. 2010. Surveillance in fruit flies free areas: an economic analysis. Working Papers 100882. University of Western Australia, School of Agricultural and Resource Economics, Western Australia, Australia.

Gurevitch, J., and D. K. Padilla. 2004. Are invasive species a major cause of extinctions? Trends in Ecology and Evolution 19:470–474.

Hafi, A., T. Arthur, M. Symes, and N. Millist. 2013. Benefit-cost analysis of the long term containment strategy for exotic fruit flies in the Torres Strait. https://portal.biosecurityportal.org.

au/nffsac/Documents/BCA%20Containment%20strategy_140401_v1%200%200.pdf

Hagen, K., W. Allen, and R. Tassan. 1981. Mediterranean fruit fly: the worst may be yet to come. California Agriculture 35:5–7.

Hastings, A., et al. 2005. The spatial spread of invasions: new developments in theory and evidence. Ecology Letters 8:91–101.

Hauser, C. E., and M. A. McCarthy. 2009. Streamlining "search and destroy": cost-effective surveillance for invasive species management. Ecology Letters 12:683–692.

Hijmans, R. J. 2015. raster: geographic data analysis and modeling. https://cran.r-project.org/web/packages/raster/raster.pdf

Horan, R. D., D. Finnoff, K. Berry, C. Reeling, and J. F. Shogren. 2018. Managing wildlife faced with pathogen risks involving multi-stable outcomes. Environmental and Resource Economics 70:713–730.

Horan, R. D., C. Perrings, F. Lupi, and E. H. Bulte. 2002. Biological pollution prevention strategies under ignorance: the case of invasive species. American Journal of Agricultural Economics 84:1303–1310.

Horie, T., R. G. Haight, F. R. Homans, and R. C. Venette. 2013. Optimal strategies for the surveillance and control of forest pathogens: a case study with oak wilt. Ecological Economics 86:78–85.

Jarrad, F. C., S. Barrett, J. Murray, R. Stoklosa, P. Whittle, and K. Mengersen. 2011. Ecological aspects of biosecurity surveillance design for the detection of multiple invasive animal species. Biological Invasions 13:803–818.

Keeling, M. J., M. E. J. Woolhouse, D. J. Shaw, L. Matthews, M. Chase-Topping, D. T. Haydon, S. J. Cornell, J. Kappey, J. Wilesmith, and B. T. Grenfell. 2001. Dynamics of the 2001 UK foot and mouth epidemic: stochastic dispersal in a heterogeneous landscape. Science 294:813–817.

Kleywegt, A. J., A. Shapiro, and T. Homem-de Mello. 2002. The sample average approximation method for stochastic discrete optimization. SIAM Journal on Optimization 12:479–502.

Kompas, T., and N. Che. 2009. A practical optimal surveillance measure: the case of Papaya Fruit Fly in Australia. Australian Centre for Biosecurity and Environmental Economics, ANU; Centre of Excellence for Biosecurity Risk Analysis, University of Melbourne, Canberra, Australia.

Kompas, T., L. Chu, P. V. Ha, and D. Spring. 2019. Budgeting and portfolio allocation for biosecurity measures. Australian Journal of Agricultural and Resource Economics 63:412–438.

Kompas, T., P. V. Ha, and H. T. M. Nguyen. 2021. Data file at OSF, SAA PFF Surveillance. https://osf.io/95j8s/

Kompas, T., P. V. Ha, H. T. M. Nguyen, I. East, S. Roche, and G. Garner. 2017. Optimal surveillance against foot-and-mouth disease: the case of bulk milk testing in Australia. Australian Journal of Agricultural and Resource Economics 61:515–538.

Kot, M., M. A. Lewis, and P. van den Driessche. 1996. Dispersal data and the spread of invading organisms. Ecology 77:2027–2042.

Lemon, J. 2006. Plotrix: a package in the red light district of R. R-News 6:8–12.

Leung, B., D. Finnoff, J. F. Shogren, and D. Lodge. 2005. Managing invasive species: rules of thumb for rapid assessment. Ecological Economics 55:24–36.

Leung, B., D. M. Lodge, D. Finnoff, J. F. Shogren, M. A. Lewis, and G. Lamberti. 2002. An ounce of prevention or a pound of cure: bioeconomic risk analysis of invasive species. Proceedings of the Royal Society B: Biological Sciences 269:2407.

Liebhold, A. M., et al. 2016. Eradication of invading insect populations: from concepts to applications. Annual Review of Entomology 61:335–352.

Lovell, S. J., S. F. Stone, and L. Fernandez. 2006. The economic impacts of aquatic invasive species: a review of the literature. Agricultural and Resource Economics Review 35:195–208.

Mak, W.-K., D. P. Morton, and R. K. Wood. 1999. Monte Carlo bounding techniques for determining solution quality in stochastic programs. Operations Research Letters 24:47–56.

McAusland, C., and C. Costello. 2004. Avoiding invasives: trade-related policies for controlling unintentional exotic species introductions. Journal of Environmental Economics and Management 48:954–977.

Meentemeyer, R. K., N. J. Cunniffe, A. R. Cook, J. A. Filipe, R. D. Hunter, D. M. Rizzo, and C. A. Gilligan. 2011. Epidemiological modeling of invasion in heterogeneous landscapes: spread of sudden oak death in California (1990–2030). Ecosphere 2:1–24.

Meentemeyer, R. K., S. E. Haas, and T. Václavk. 2012. Landscape epidemiology of emerging infectious diseases in natural and human-altered ecosystems. Annual Review of Phytopathology 50:379–402.

Mehta, S. V., R. G. Haight, F. R. Homans, S. Polasky, and R. C. Venette. 2007. Optimal detection and control strategies for invasive species management. Ecological Economics 61:237–245.

National Research Council. 1992. Restoration of aquatic ecosystems: science, technology, and public policy. National Academy Press, Washington D.C., USA. https://www.nap.edu/catalog/1807/restoration-of-aquatic-ecosystems-science-technology-and-public-policy

Norkin, V. I., G. C. Pflug, and A. Ruszczyński. 1998. A branch and bound method for stochastic global optimization. Mathematical Programming 83:425–450.

Olson, L. J. 2006. The economics of terrestrial invasive species: a review of the literature. Agricultural and Resource Economics Review 35:178.

Olson, L. J., and S. Roy. 2002. The economics of controlling a stochastic biological invasion. American Journal of Agricultural Economics 84:1311–1316.

Pebesma, E., and R. Bivand. 2005. Classes and methods for spatial data in R. R News 5:9–13.

Perpiñán, O., and R. Hijmans. 2014. rasterVis. R package version 0.32. https://oscarperpinan.github.io/rastervis/

Perrings, C., S. Dalmazzone, and M. H. Williamson. 2000. The economics of biological invasions. Edward Elgar Publishing, Northampton, UK.

Pimentel, D., L. Lach, R. Zuniga, and D. Morrison. 2000. Environmental and economic costs of nonindigenous species in the United States. BioScience 50:53–65.

Polasky, S. 2010. A model of prevention, detection, and control for invasive species. Pages 100–107 *in* C. Perrings, H. Mooney, and M. Williamson, editors. Bioinvasions and globalization. Oxford University Press, Oxford, UK.

R Core Team. 2014. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Shapiro, A. 2003. Monte carlo sampling methods. Pages 353–425 *in* A. Ruszczyński and A. Shapiro, editors. Handbooks in operations research and management science. Volume 10. Elsevier Science, Amsterdam, The Netherlands.

Shigesada, N., and K. Kawasaki. 1997. Biological invasions: theory and practice. Oxford University Press, Oxford, UK.

Taylor, C. M., and A. Hastings. 2004. Finding optimal control strategies for invasive species: a density-structured model for

Spartina alterniflora. Journal of Applied Ecology 41:1049–1057.

The South Australian Government Gazette. 2020. State Government Instruments. Plant health ACT 2009. Fruit Fly Suspension Area. https://governmentgazette.sa.gov.au/sites/default/files/public/documents/gazette/2020/July/2020_057.pdf

Underwood, R. 2007. Fruit fly: Likely impact of an incursion of fruit fly in the bay of Plenty, Hawkes bay or Nelson. Fruition Horticulture (BOP) Ltd, Tauranga, New Zealand.

Van den Bosch, F., R. Hengeveld, and J. Metz. 1992. Analysing the velocity of animal range expansion. Journal of Biogeography 19:135–150.

Verweij, B., S. Ahmed, A. Kleywegt, G. Nemhauser, and A. Shapiro. 2003. The sample average approximation method applied to stochastic routing problems: a computational study. Computational Optimization and Applications 24:289–333.

Waterhouse, D. F., and D. P. A. Sands. 2001. Classical biological control of arthropods in Australia. CSIRO Entomology, Canberra, Australian Capital Territory, Australia.

Wickham, H. 2009. ggplot2: elegant graphics for data analysis. Springer, New York, New York, USA.

Wilcove, D. S., D. Rothstein, J. Dubow, A. Phillips, and E. Losos. 1998. Quantifying threats to imperiled species in the United States. BioScience 48:607–615.

Wilen, J. E. 2007. Economics of spatial-dynamic processes. American Journal of Agricultural Economics 89:1134–1144.

Yonow, T., M. Zalucki, R. Sutherst, B. Dominiak, G. Maywald, D. Maelzer, and D. Kriticos. 2004. Modelling the population dynamics of the Queensland fruit fly, *Bactrocera* (*Dacus*) *tryoni*: a cohort-based approach incorporating the effects of weather. Ecological Modelling 173:9–30.

Supporting Information

Additional supporting information may be found online at: http://onlinelibrary.wiley.com/doi/10.1002/eap.2449/full

Open Research

Data available from OSF (Kompas et al. 2021): https://osf.io/95j8s/.