Received Date : 11-Aug-2015

Revised Date : 09-Dec-2016

Accepted Date : 22-Dec-2016

Article type : Research Article

Handling editor: Dr. Ryan Chisholm

Selecting simultaneous actions of different durations to optimally manage an ecological network

Optimising simultaneous actions

Martin Péron[1,2,*], Cassie C. Jansen[2,3], Chrystal Mantyka-Pringle[2,4], Sam Nicol[2], Nancy A. Schellhorn[2], Kai Helge Becker[5], Iadine Chadès[2,6]

[1] Queensland University of Technology, Mathematical school, Brisbane, QLD 4000, Australia

[2] Commonwealth Scientific and Industrial Research Organisation, Dutton Park, QLD 4102, Australia

[3] Queensland Health, Metro North Public Health Unit, Windsor, QLD 4030, Australia

[4] The University of Saskatchewan, Global Institute for Water Security, School of Environment and Sustainability, Saskatoon, SK S7N 5B3, Canada.

[5] University of Strathclyde, Department of Management Science, Glasgow G1 1XQ, United Kingdom.

[6] ARC Centre of Excellence for Environmental Decisions, University of Queensland, Brisbane QLD 4072

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/2041-210X.12744
This article is protected by copyright. All rights reserved.

Accepted Article

CJ: cassie.jansen@health.qld.gov.au; CMP: c.mantyka-pringle@usask.ca

SN: sam.nicol@csiro.au; NAS: nancy.schellhorn@csiro.au; IC: iadine.chades@csiro.au; KHB: kai.becker@strath.ac.uk

Author for correspondence:

*Martin Péron:

Phone: +61 (4) 2148 8778;

Email: martin.peron@csiro.au

Accepted Article

Abstract

1. Species management requires decision-making under uncertainty. Given a management objective and limited budget, managers need to decide what to do, and where and when to do it. A schedule of management actions that achieves the best performance is an optimal policy. A popular optimisation technique used to find optimal policies in ecology and conservation is stochastic dynamic programming (SDP). Most SDP approaches can only accommodate actions of equal durations. However, in many situations, actions take time to implement or cannot change rapidly. Calculating the optimal policy of such problems is computationally demanding and becomes intractable for large problems. Here, we address the problem of implementing several actions of different durations simultaneously.
2. We demonstrate analytically that synchronising actions and their durations provide upper and lower bounds of the optimal performance. These bounds provide a simple way to evaluate the performance of any policy, including rules of thumb. We apply this approach to the management of a dynamic ecological network of *Aedes albopictus*, an invasive mosquito

This article is protected by copyright. All rights reserved.

that vectors human diseases. The objective is to prevent mosquitoes from colonising mainland Australia from the nearby Torres Straits Islands where managers must decide between management actions that differ in duration and effectiveness.

3. We were unable to compute an optimal policy for more than eight islands out of 17, but obtained upper and lower bounds for up to 13 islands. These bounds are within 16% of an optimal policy. We used the bounds to recommend managing highly populated islands as a priority.
4. Our approach calculates upper and lower bounds for the optimal policy by solving simpler problems that are guaranteed to perform better and worse than the optimal policy, respectively. By providing bounds on the optimal solution, the performance of policies can be evaluated even if the optimal policy cannot be calculated. Our general approach can be replicated for problems where simultaneous actions of different durations need to be implemented.

Keywords: optimal management, simultaneous actions, Markov decision processes, susceptible-infested-susceptible, SIS, performance bounds, invasive species, threatened species, mosquito, *Aedes albopictus*.

Introduction

Managing dynamic ecological systems is often constrained by limited resources, leading managers to use mathematical methods to make cost-effective decisions (Duke, Dundas & Messer 2013). Given a specified management objective, sequential decisions can be optimised with an algorithm called stochastic dynamic programming (SDP, Marescot *et al.* (2013)). When computational resources are

This article is protected by copyright. All rights reserved.

Accepted Article

sufficient relative to the complexity of the problem, SDP returns an optimal *policy,* i.e. action to implement in each state of the system, and the *performance* (or *value)* of this policy (Puterman 1994). In behavioural ecology, SDP is used to assess if species optimise their reproductive fitness over time (Houston *et al.* 1988; Venner *et al.* 2006). In applied ecology, SDP has become an essential decision-making tool when information is missing, with applications in prioritizing global conservation effort (Wilson *et al.* 2006), weed control (Firn *et al.* 2008), disease management (Chadès *et al.* 2011), species migration (Nicol *et al.* 2015), fire regime management (McCarthy, Possingham & Gill 2001) and adaptive management (Walters & Hilborn 1978; Hauser & Possingham 2008). To achieve a management objective faster, several actions can be implemented simultaneously. In particular, for spatial problems, simultaneous actions in different locations must be optimised, for example in forestry (Forsell *et al.* 2011) or invasive or threatened species management (Monterrubio, Rioja-Paradela & Carrillo-Reyes 2015; Mantyka-Pringle *et al.* 2016). Additionally, there are many examples in the ecological literature of actions of different durations (Phelan, Norris & Mason 1996; Pelizza *et al.* 2010).

To date, little research has focused on simultaneous actions (Boutilier & Brafman 1997). In artificial intelligence, simultaneous actions have become important when several decision problems merge (Singh & Cohn 1998), or when actions have random durations (Rohanimanesh & Mahadevan 2002). Accommodating simultaneous actions of different durations is challenging because they terminate at different timesteps (Barto & Mahadevan 2003) and thus, computing an exact SDP to find an optimal policy requires high memory demands and computation time. A workaround is to use approximate methods. Approximate algorithms focus on maximising the value of policies but, in practice, policies that cannot be explained in ecological terms will not be applied by managers (Walters 1986). Identifying sensible rules of thumb, i.e. simplified versions of more complex policies, is often preferred (Chadès *et al.* 2008; Grechi *et al.* 2014). However, this simplification causes a loss of value that is often unknown to managers (Pichancourt *et al.* 2012).

This article is protected by copyright. All rights reserved.

Here, we introduce two approximate models that provide upper and lower bounds on the optimal performance at an advantageous computational cost and will allow decision-makers to find well performing rules of thumb. Obtaining an *upper bound* and *lower bound* of the unknown optimal performance is useful, as calculating the error in the performance of a rule of thumb relative to the upper bound, e.g. 10%, guarantees that this rule of thumb is within 10% of the optimal performance. We apply our approach to the management of invasive mosquito *Aedes albopictus* in the Torres Strait Islands, Australia. This approach can be replicated for large problems when simultaneous actions have different durations to evaluate and increase the reliability of rules of thumb.

Materials and Methods

Markov decision problems and stochastic dynamic programming

Markov decision processes (MDP) are mathematical frameworks for modelling sequential decision problems where the outcome is partly stochastic and partly controlled by a decision-maker. A MDP is defined by five components $<S, A, P, r, C>$ (Puterman 1994) : (i) a state space S, (ii) an action space A, (iii) a transition function P for each action, (iv) immediate rewards r and (v) a performance criterion C.

The decision-maker aims to direct the process towards rewarding states, motivated by a performance criterion. From a given state s, the decision-maker selects an action a and receives a reward $r(s, a)$. At the next timestep, the system transitions to a subsequent state s' with probability $P(s'|s, a)$. The performance criterion C specifies the objective (e.g. maximise or minimise a sum of expected future rewards), the time horizon (finite or infinite), the initial state s_0 and whether there is a discount rate (γ). A *policy* π describes which decisions are made in each state, i.e. $\pi: S \rightarrow A$.

This article is protected by copyright. All rights reserved.

Solving a MDP means finding a policy that optimises the performance criterion (*optimal policy*). Stochastic dynamic programming (SDP) denotes a collection of solution methods to solve MDPs, such as policy iteration and value iteration (see Marescot *et al.* (2013) and Appendix S1 for an overview). SDP is an efficient algorithm since it runs in polynomial time, but may not be tractable when the state or action spaces are very large, thus requiring alternative approaches (Nicol & Chadès 2011).

Extension of MDP for simultaneous decisions of different durations

A limitation of MDPs is that all actions must occur for the same duration. Herein we provide a method to overcome this limitation. Specifically, we address decision problems where an action a can be decomposed into N sub-actions $a_1, a_2, \dots, a_N$ at each timestep, with $a_i \in A_i$. As an example of the distinction between actions and sub-actions, consider a management strategy for a network of N connected sites. An action is comprised of N sub-actions applied to the individual sites. Each sub-action may have a different duration $d(a_i)$ and must be implemented for its full duration. The transition function and rewards may depend on the sub-actions $a_1, a_2, \dots, a_N$ currently implemented.

We propose a MDP model that solves this decision problem optimally, called the *exact model*. To fit the MDP framework, we need to respect the Markov property, which requires that subsequent states can be predicted using only the current state and action. To ensure that all actions are implemented until completion, we augment each state $s \in S$ with information about which sub-actions are currently implemented $(a_1, a_2, \dots, a_N)$ and the number of timesteps until each finishes (noted $t_1, t_2, \dots, t_N, with\ t_i \in \mathbb{N}$). Formally, each state of the exact model becomes $(s, a_1, t_1, a_2, t_2, \dots, a_N, t_N)$. The new state space is denoted S_{exact}. The set of possible actions $A(s, a_1, t_1, \dots, a_N, t_N)$ that can be implemented depends on the current MDP state: if a_i is not finished $(t_i > 0)$, then a_i must continue; if a_i has just terminated $(t_i = 0)$, all sub-actions are possible.

This article is protected by copyright. All rights reserved.

The transition function P_{exact} should not only contain the transition function P on states of S but also update the elements (a_i, t_i), i.e. initialise $(a_i, d(a_i) - 1)$ when a_i begins, and then subtract 1 from t_i at each timestep until a_i is completed. The rewards r are the same in the original MDP and the augmented exact model.

We can apply SDP to this exact model $< S_{exact}, A, P_{exact}, r >$ to find the optimal policy and its performance, noted V^*. However, the state space S_{exact} is exponential in N: $|S_{exact}| = |S| \prod_{i=1}^{N} \left(1 + \sum_{a_i \in A_i} (d(a_i) - 1)\right)$ (Appendix S2). SDP is likely to be intractable for all but trivial values of N (see Results). This motivates us to introduce two approximate models, the *lower bound model* and the *upper bound model*.

These two models are obtained by synchronising sub-actions, which forces sub-actions to finish simultaneously. As a consequence, the performance of the lower and upper bound models will be lower and higher than the performance of the exact model. With all actions finishing simultaneously, the number of states can be reduced dramatically and larger problems can thus be addressed.

Lower bound model

The *lower bound model* is obtained by modifying the exact model in two steps.

First, we add a *synchronisation constraint*, which forces all sub-actions to be implemented as many times as necessary to end simultaneously (Fig. 2). To do so, we forbid changing any sub-action while at least one sub-action is in progress. The resulting MDP leads to a performance equal or lower than the exact model (Appendix S3). Intuitively, since the lower bound model is less flexible than the exact model, fewer policies are possible and performance decreases.

Second, we reformulate the state space obtained by synchronisation to remove unnecessary states. The states where at least one sub-action is in progress are unnecessary to obtain the optimal policy. Given an action a, the sub-actions $a_1, a_2, \dots a_N$ will finish simultaneously after the least common multiple (LCM) of the durations $d(a_1), d(a_2), \dots d(a_N)$:

This article is protected by copyright. All rights reserved.

$$LCM\{d(a_i), 1 \leq i \leq N\} = \min\{d \in \mathbb{N}: 1 \leq i \leq N, d = k_i d(a_i)\ for\ some\ k_i \in \mathbb{N}\}$$

We note this duration $LCM(a)$. We remove states with a duration shorter than $LCM(a)$ and assign the sub-actions of a with duration $LCM(a)$. The new state space is $S \subset S_{exact}$.

The new action space is the same as in the exact model except that it is defined on the subset $S \subsetneq S_{exact}$.

Because our aim is to compare the performance of the exact and lower bound models, we define the transition function and rewards such that a policy calculated with the lower bound model will have the same performance as when evaluated using the exact model. Since the new transitions last several timesteps during which the action does not change, a transition over d timesteps is made of d times the same transition: for every action a, P_{lower} is the function (matrix) P raised to the power of the duration of a:

$$P_{lower}(.|.,a) = P(.|.,a)^{LCM(a)} \qquad (eqn\ 1)$$

The rewards $r_{lower}(s,a)$ should account for both the immediate reward $r(s,a)$ and the expected rewards of the states that are not computed in the lower bound model (Fig. 2B):

$$r_{lower}(s,a) = r(s,a) + \sum_{i=1}^{LCM(a)-1} \gamma^i \sum_{s' \in S} P(s'|s,a)^i r(s',a) \qquad (eqn\ 2)$$

The resulting model $< S, A, P_{lower}, r_{lower} >$ is a semi-MDP (Bradtke & Duff 1994), an extension of MDPs. Only one action can be implemented at a time, however different actions can have different durations. Semi-MDPs can be solved efficiently with SDP. Reformulating the model by removing unnecessary states does not affect performance, so the optimal performance in the lower bound model V^*_{lower} is a lower bound of the optimal performance in the exact model V^*:

$$V^*_{lower}(s) \leq V^*(s),\ \ \forall s \in S \qquad (eqn\ 3)$$

The number of states in this model is greatly reduced, from $|S_{exact}| = |S| \prod_{i=1}^{N}\left(1 + \sum_{a_i \in A_i}(d(a_i) - 1)\right)$ to $|S|$. This model can be solved for problems of larger sizes than the exact problem.

This article is protected by copyright. All rights reserved.

Upper bound model

Like the lower bound model, the *upper bound model* is built in two steps from the exact model.

First, we allow management actions to be interrupted before completion in order to start different management actions (Fig. 2C). We reduce the duration of all actions to a unique duration (noted GCD for convenience), which equals the greatest common divisor of the durations of all sub-actions:

$GCD\{d(a): a \in \bigcup_{i=1}^{N} A_i\} = \max\{d \in \mathbb{N}: a \in \bigcup_{i=1}^{N} A_i, d(a) = dk_a \; for\ some\ k_a \in \mathbb{N}\}$

Note that GCD does not depend on implemented sub-actions, but rather the set of all sub-actions available. GCD must evenly divide all durations $d(a)$ to ensure that a management action a applied repeatedly in the upper bound model can last $d(a)$ timesteps—its duration in the exact model. Any policy in the exact model can also be implemented in the upper bound model. Since the upper bound model is more flexible than the exact model due to shorter actions (technically, a *relaxation*), the resulting MDP leads to an equal or higher performance than the exact problem formulation (Appendix S3).

Second, we reformulate the state space to remove unnecessary states. As per the lower bound model, the states between times $t = 0$ and $t = GCD$ require no decisions and can be removed.

The state space and the action space of the upper bound model are the same as in the lower bound model (S and A). For each action, the new transition function P_{upper} is P raised to the power of the duration GCD. The new rewards $r_{upper}(s,a)$ must take into account both the immediate reward $r(s,a)$ and the expected rewards in the removed states, which are no longer computed:

$$r_{upper}(s,a) = r(s,a) + \sum_{i=1}^{GCD-1} \gamma^i \sum_{s' \in S} P(s'|s,a)^i r(s',a), \forall s \in S, \forall a \in A(s) \quad (eqn\ 4)$$

The resulting MDP $< S, A, P_{upper}, r_{upper} >$ is a semi-MDP whose optimal performance, V^*_{upper}, is an upper bound of the exact performance V^*:

$$V^*(s) \leq V^*_{upper}(s), \quad for\ all\ s \in S \quad (eqn\ 5)$$

This article is protected by copyright. All rights reserved.

Accepted Article

In conclusion, we have constructed two MDP models which provide lower and upper bounds of the exact performance:

$$V^*_{lower}(s) \leq V^*(s) \leq V^*_{upper}(s), \quad for\ all\ s \in S \quad (eqn\ 6)$$

These approximate models require $|S|$ states to be solved, many fewer than the exact model. Importantly, unlike the policy of the lower bound model, the policy of the upper bound model cannot be implemented (as it violates the duration constraints of the actions). However, it provides a valuable upper bound against which to compare viable sub-optimal policies.

We provide the MATLAB code solving our case study at dx.doi.org/10.6084/m9.figshare.4557565. It uses the MDPSolve package (https://sites.google.com/site/mdpsolve/). The necessary input parameters are provided in Appendix S4.

Case study: Managing *Aedes albopictus* in the Torres Strait Islands

Aedes albopictus is a highly invasive species and a vector of several arboviruses, including chikungunya and dengue viruses (Bonizzoni *et al.* 2013). *Aedes albopictus* was first detected in the Torres Strait Islands in 2005 (Ritchie *et al.* 2006), where it persists today. These islands are potential sources of dispersal between Indonesia, Papua New Guinea (PNG) and mainland Australia (Beebe *et al.* 2013) via numerous human-mediated pathways including local boats, airplanes and ferries (Fig. 1). Herein, for simplicity, we consider Indonesia and PNG as a single potential source of *Aedes albopictus* referred to as PNG.

If *Aedes albopictus* were to establish on mainland Australia, its invasion is expected to be widespread and persistent (Hill, Axford & Hoffmann 2014), and extremely challenging to control (Beebe *et al.* 2013). Further, Australia's main population centres would likely become receptive to dengue transmission (Russell *et al.* 2005) and subject to significant biting nuisance (Beebe *et al.* 2013).

This article is protected by copyright. All rights reserved.

Since the detection of *Aedes albopictus* in the Torres Strait, several management actions have been implemented. These include community education, insecticide applications to harbourage areas (e.g. vegetation which provides resting habitat for adult mosquitoes) and domestic housing, and chemical treatment or disposal of container larval habitats (e.g. plant pots, sagging tarps, etc.). We distinguish two levels of such management actions: *light* and *strong*, the latter being costlier but more effective. Since budget is limited, not all islands can be managed simultaneously. At each timestep (six months), decision-makers must decide which of the 17 inhabited islands should be managed to protect mainland Australia. Since we assume that mainland Australia cannot be successfully managed if infested, our objective is to maximise the mean time until *Aedes albopictus* invades mainland Australia.

States and transition function of the SIS model

We model the mosquitoes' dispersal over time using a Susceptible-Infected-Susceptible (SIS) network (Pastor-Satorras & Vespignani 2001). The SIS model allows the locations (islands hereafter) to be either *infested* with, or *susceptible* to an invasive species. In our case study, each state s represents the *infestation status* of all Torres Strait Islands and mainland Australia. Because we assume that mainland Australia cannot be managed, we define 'mainland Australia infested' as an absorbing state (i.e. sink), noted σ. All other states are of the form $(s_1, s_2, \dots, s_N)$, with s_i the status of island i. Formally, $S = \{\sigma\} \cup \{(s_1, s_2, \dots, s_N); s_i \in \{infested, susceptible\}, 1 \leq i \leq N\}$.

In an SIS model, the status of each location may change at each timestep in two ways. First, infested locations can become susceptible when the species goes locally extinct: in our case study, the extinction probability equals the effectiveness of the action currently implemented (see next paragraph). Second, links between locations represent risks of reinfestation of susceptible locations from infested ones. Here, we defined the (re)infestation probability as follows: mainland Australia remains susceptible in the next timestep (i.e. $s_{t+1} \neq \sigma$) with probability $\prod_{i \in \{1,2,..,N\}} \left(1 - p_M^i\right)$ where p_M^i is the infestation probability from an infested island i to mainland Australia. Conversely, the

This article is protected by copyright. All rights reserved.

probability of the mainland becoming infested (i.e. $s_{t+1} = \sigma$) is $1 - \prod_{i \in \{1,2,..,N\}}(1 - p_M^i)$. The (re)infestation probability of all islands from neighbouring islands follows suit.

Rewards and actions of the SIS model

The reward r reflects our objective to prevent the infestation of mainland Australia. We set a reward of 0.5 when the mainland is not infested ($s \neq \sigma$), and 0 if it is infested ($s = \sigma$). With a discount of $\gamma = 1$, we receive 0.5 every timestep (six months) until mainland Australia becomes infested. The *performance* of any policy (i.e. expected cumulative reward obtained) equals the mean time until infestation. Although $\gamma = 1$, the mean time to infestation is finite because we assume that PNG is an infinite source of mosquitoes.

Two management actions are possible (*light* and *strong*), but the budget allows implementation of only one light management and one strong management across all islands, or three light managements, at each timestep. A sub-action a_i can take values in $A_i = \{no\ action, light, strong\}$. The maximum budget is accounted for by reducing the set of possible actions $A(s) \subseteq \prod_{i=1}^{N} A_i$ (unaffordable actions and their related states are not computed). The effectiveness $p(a_i)$ of sub-action a_i is defined as the probability of eradicating the mosquito over one timestep and depends on the characteristics of island i (Appendix S5-6). Finally, the durations $d(a_i)$ vary: *no action* lasts one timestep (six months), and *light and strong management* last six timesteps each.

Parameters

Data was collected at an expert elicitation workshop to estimate the effectiveness of actions based on characteristics of the Torres Strait Islands (Martin *et al.* (2012), Appendix S5-6). Experts in invasive species, vector biology and ecology, mosquito control, public health management and biosecurity estimated the effectiveness of all three management actions. Estimates accounted for island characteristics including size, vegetation refuge and accessibility (terrain), which influence both the operational feasibility of actions and the habitat suitability for *Aedes albopictus*. We used a

This article is protected by copyright. All rights reserved.

Accepted Article

Bayesian network to calculate the effectiveness of actions depending on these characteristics (Appendix S6, Clark (2005)).

We were unable to collect data on the probability of transmission of *Aedes albopictus* between islands. In the absence of data, the probability of transmission between islands was derived using Cauchy dispersion kernels (Pitt 2008). Experts agreed that the transmission between two given islands likely depended on the number of inhabitants and the distance between islands; larger populations and proximal islands have higher transmission probabilities. The transmission probability p_i^j from an island i to j depends on the island populations, pop_i and pop_j, and the distance between the islands d_{ij}:

$$p_i^j = D \times \frac{pop_i \times pop_j}{1 + \left(\frac{d_{ij}}{\beta}\right)^2} \quad (\text{Cauchy} - eqn\ 7)$$

where D is a constant influencing the speed of transmissions through the network, and β is the shape parameter. We calibrated two sets of parameters arbitrarily, namely low and high transmission probabilities, leading to slow and fast infestations of mainland Australia, respectively. The range of mean times to infestation captures the time to infestation estimated by experts (Appendix S7-8).

Computational experiments

We compared the optimal performances of our three models on our case study. Recall that the performance of a policy equals the mean time until infestation of mainland Australia. It was not necessary to run simulations for these proposed MDP models because both the optimal policy and its performance are direct outputs of SDP. Solving the exact 17-island network problem was computationally intractable (runs out of the 1000GB memory), so we gradually evaluated the performance of our proposed models on networks including an increasing number of islands

This article is protected by copyright. All rights reserved.

(remaining islands do not affect the system). We first added the islands with the highest probability of directly infesting the mainland (rule 'highest transmission', see Appendix S5). We tested the robustness of our approach on two dispersal scenarios, with low and high mosquito transmission probabilities (see Parameters).

We also evaluated several simple rules of thumb, which consist of managing the highest-ranked infested islands according to the following rankings: (i) largest populations; (ii) closest to the mainland; (iii) easiest to manage (islands where actions have the highest probability of success); and (iv) highest transmission probability toward the mainland. We calculated the performance of (v) continuously implementing strong managements on all islands ('all managed', i.e. unlimited budget) and (vi) no actions. We ran 10,000 simulations to assess the performance of these rules of thumb and recorded the mean times of infestation and 90% confidence intervals.

Results

Low transmission probabilities

For all models, infestation of mainland Australia happens sooner as more islands are included in the analysis (as expected; Fig. 3). A steep decrease in the mean time until infestation occurs when considering up to five islands (20-30 years/island), followed by a gradual decrease (approximately one year/island) until all islands are included in the analysis. This is because we incrementally included the islands with the rule 'highest transmission'.

As expected, the performance of the exact model is between the upper bound model and the lower bound model and all rules of thumb. The exact model runs out of memory for more than eight islands while the lower and upper bound models are tractable until 13 islands. This difference is attributable to a higher number of states in the exact model, with a ratio up to $\prod_{i=1}^{N}(1+$

This article is protected by copyright. All rights reserved.

$\sum_{a_i \in A_i}(d(a_i) - 1)) = 11^N$ (the ratio is lower in practice since states related to unaffordable actions are disregarded - see Material and Methods).

All model performances are equal when only one island (Thursday Island) is included because there are no simultaneous management actions. For more than one island, all performances remain similar, in particular those of the lower bound and the exact models. We have chosen to display the performance of the best rule of thumb only, 'highest transmission', which performs equally to our lower bound. Other rules of thumb ('largest population', 'closest to mainland' and 'easiest to manage') perform worse than 'highest transmission' (Appendix S9). The performances of 'no actions' and 'all managed' illustrate the worst and best possible outcomes, respectively, but these bounds are less informative about the optimal performance, because they are much wider than the lower and upper bound models.

To assess the quality of the upper bound, we calculated the relative errors of all models compared to the upper bound (Table 1). The relative error of the exact model remains less than 14% and shows that this upper bound remains close to the exact performances when islands are added. The relative error of the lower bound remains less than 16% for all numbers of islands considered, guaranteeing that the lower bound equals at least 84% of the exact performance in our case study. Note that the lower bound is a very close approximation of the exact policy (see Discussion). The relative error of 'highest transmission' is similar to that of the lower bound and remains less than 17%.

High transmission probabilities

When assuming high transmission probabilities, the mean time until infestation of mainland Australia under our best rule of thumb is less (13 years for 17 islands) than that calculated using low transmissions probabilities (50 years for 17 islands; Fig. 4). The differences between all models are smaller than with the low transmission probabilities. The rule of thumb 'highest transmission' performs consistently well, while others (not shown) underperformed.

This article is protected by copyright. All rights reserved.

Accepted Article

The lower and upper bounds are closer together than with the low transmission probabilities. This is confirmed by the relative error compared to the upper bound (Table 2). The relative error of the exact model (<7%) shows that this upper bound remains close to the exact performances regardless of the number of islands considered. The relative error of the lower bound model reaches 9% (6% when the exact model is no longer tractable). The relative error of 'highest transmission' is less than 9%.

Optimal policies - low and high transmissions probabilities

When running simulations of the policies recommended by the exact, lower bound and upper bound models, some islands appear more important than others. It is therefore possible to identify an order (the *prioritisation ranking*) in which islands should be managed until eradication. Further, this prioritisation ranking is very similar for all three policies (when tractable) and for both high and low transmission probabilities.

When considering four islands (Appendix S10), all policies prioritise Thursday and Horn Islands (in this order) before Mulgrave and Banks Islands, i.e. this prioritisation ranking matches the 'highest transmission' ranking exactly. These two rankings are not the same when more islands are included (upper bound model for 11 islands; Fig. 5 & Appendix S5), because other factors than 'highest transmission' also affect the optimal policies. One such factor is the effectiveness of management: ineffective management actions on Banks cause it to be ranked 8th on the prioritisation ranking against 4th on the 'highest transmission' ranking. Another factor is the proximity of islands: Jervis Island is 5th on the prioritisation ranking against 9th on the 'highest transmission' ranking. A possible interpretation is that Jervis, when compared to Yam or Coconut for example, is close to critical islands such as Thursday, Horn and Mulgrave.

Accepted Article

This article is protected by copyright. All rights reserved.

Discussion and concluding remarks

We developed a new approach to assist decision-makers when actions are simultaneous and of different durations. This approach modifies time constraints to reduce the model size by several orders of magnitude to obtain bounds of the unknown exact performance. We applied this to the spatial management of an invasive mosquito, *Aedes albopictus*, modelled as a SIS network. The bounds provide a narrow range guaranteed to contain the performance of the exact optimal policy, for problems too large to compute the exact solution. This research impacts metapopulations and network management problems in biosecurity, health and ecology when the budget allows the implementation of simultaneous actions.

Our two approximate models share a number of advantages when compared to rules of thumb. First, they account for the consequences of actions on future events, which is necessary to select the best immediate action. The sensitivity analysis on low and high transmission probabilities shows that the lower bound model is less likely to underperform than rules of thumb, which are not guaranteed to perform well (Abel 2003). Second, our models can be evaluated exactly with SDP rather than using simulations. Third, the policies generated by our models can be used to derive efficient rules of thumb.

The performances of the lower and upper bound models are sensitive to the least common multiple (LCM) and greatest common divisor (GCD) of the duration of management actions. In our case study, the lower bound likely performs well because the LCM is exactly the duration of the management actions. We have run the tool with various durations to evaluate the sensitivity of bound models to the GCD and the LCM (Appendix S11). When these durations share many divisors, the LCM and GCD are close, which leads to small relative errors between bounds. By contrast, when durations do not share many divisors, the relative errors between bounds increase.

Although the lower and upper bound models can be solved at a reduced computational cost, in our case study the memory size and computation times required still grow exponentially with the

This article is protected by copyright. All rights reserved.

Accepted Article

number of islands considered (Appendix S12). Here, the number of states is $2^N + 1$ because we used a *flat* representation of states, i.e. each possible combination of island states is accounted for. To optimise the management of SIS networks, Chadès *et al.* (2011) used *factored* MDPs to take advantage of the network structure, i.e. the independence of conditional probabilities. In our model, all islands are connected (complete network) and using factored MDP provides no advantage. As data becomes available, it is likely that small transmission probabilities could be ignored to create a network structure that could be exploited by factored MDPs (Hoey *et al.* 1999; Forsell *et al.* 2011). We increased the number of islands managed incrementally, ignoring the influence of other islands. An alternative would be to aggregate the remaining islands as one island. However, this is not a trivial task as it requires aggregating a large number of states (Li, Walsh & Littman 2006). How to do so in the best way possible will be the aim of future research.

Management implications

All models target Thursday, Horn and Mulgrave Islands as management priorities in this order, because these islands are highly populated and close to mainland Australia and, hence, have the highest probability of transmission to mainland. Knowing that these islands are close to each other (favouring transmissions) and that Horn Island is the 'transport hub' of the Torres Strait adds further credence to their high prioritisation. The prioritisation of these three islands is insensitive to the number of islands included (1-13) and to the transmission probabilities (low/high), showing the robustness of this policy. However, the mean time until infestation greatly depends on the dataset: it ranges from 13 to 50 years when calculated using low (Fig. 3) and high transmission probabilities (Fig. 4), respectively. Obtaining more precise estimates of the transmission probabilities will produce a narrower time range estimate. Higher budgets allocated to management can also postpone infestation, more sensitively when transmission probabilities are low (40 years with no budget/80 years with unlimited budget) than high (10/15 years). A comprehensive sensitivity analysis would help the decision maker set the most suitable budget.

This article is protected by copyright. All rights reserved.

Accepted Article

Additional factors may influence our management recommendations. For example, *Aedes albopictus* is difficult to detect and decision-makers cannot be certain that an island is susceptible (Hawley 1988). It is possible to provide management recommendations that accounts for imperfect detection using partially observable MDPs (Chadès et al. 2008). However, these models do not yet account for actions of different durations and are even more difficult to solve than MDPs. Other unknown factors may influence management recommendations such as species interactions, increased migration flow and effects of climate change. Value of information studies could help decision-makers determine whether these unknown factors warrant adapting management recommendations (Canessa *et al.* 2015).

In our case study, practitioners keep managing *Aedes albopictus* for a fixed period of time on the targeted island despite the mosquitoes being undetected. This constraint was motivated by the imperfect detectability of *Aedes albopictus* (Hawley 1988), which may occur in other applications. For example, Chadès *et al.* (2008) show that managing a threatened species with imperfect detectability can be optimal, even when we do not observe the species. This typically happens when the species is still deemed very likely to be present. Similarly, Regan, Chadès and Possingham (2011) recommend managing invasive plants up to four years since the last detection to ensure eradication. Another motivation for having prolonged management in absence of sighting is to decrease the suitability of mosquito habitat. For instance, managing soil organically for several years reduces the susceptibility of a species of maize to an insect pest significantly (Phelan, Norris & Mason 1996). Furthermore, to control a weed of rice, McIntyre, Mitchell and Ladiges (1989) recommend combining management actions of various durations and starting times, such as delaying flooding and establishing a sward of pasture during the coolest seasons. Our general approach could help optimise the spatial management of these problems. Other reasons for long actions may include operational constraints, such as fixed-length contracts for workers implementing actions.

Accepted Article

This article is protected by copyright. All rights reserved.

Accepted Article

Contributions

The project was devised by all authors. The optimisation models were developed and implemented by MP. These optimisation models include the exact model, the lower and upper bounds models. MP and IC wrote the manuscript; all authors substantially edited the manuscript. Data was collected by IC, CJ, CMP and NAS. Bayesian network analysis was conducted by MP, CJ, CMP and IC.

Acknowledgements

This research is supported by an Industry Doctoral Training Centre scholarship (MP) and CSIRO Julius Career Awards (IC, NAS). We acknowledge the critical contributions of time and expertise provided by the *Aedes albopictus* Technical Advisory Group, and other experts who participated in the expert elicitation workshop. We would also like to thank Andrew Higgins for valuable feedback on this manuscript. Computational resources and services used in this work were provided by the HPC and Research Support Group, Queensland University of Technology, Brisbane, Australia.

Data Accessibility

All data relevant to this study, i.e. the effectiveness of actions and the probabilities of transmission of mosquitoes, is available at dx.doi.org/10.6084/m9.figshare.4557562. The MDP transition matrices are generated using the MATLAB code, which is available at dx.doi.org/10.6084/m9.figshare.4557565.

This article is protected by copyright. All rights reserved.

References

Abel, C.F. (2003) Heuristics and problem solving. *New Directions for Teaching and Learning,* **2003,** 53–58.

Barto, A. & Mahadevan, S. (2003) Recent advances in hierarchical reinforcement learning. *Discrete Event Dynamic Systems,* **13,** 341–379.

Beebe, N.W., Ambrose, L., Hill, L.A., Davis, J.B., Hapgood, G., Cooper, R.D., Russell, R.C., Ritchie, S.A., Reimer, L.J., Lobo, N.F., Syafruddin, D. & van den Hurk, A.F. (2013) Tracing the tiger: population genetics provides valuable insights into the *Aedes (Stegomyia) albopictus* invasion of the Australasian Region. *PLoS Negl Trop Dis,* **7,** e2361.

Bonizzoni, M., Gasperi, G., Chen, X. & James, A.A. (2013) The invasive mosquito species *Aedes albopictus*: Current knowledge and future perspectives. *Trends in Parasitology,* **29,** 460-468.

Boutilier, C. & Brafman, R. (1997) Planning with concurrent interacting actions. *Conference of the Association for the Advancement of Artificial Intelligence*.

Bradtke, S.J. & Duff, M.O. (1994) Reinforcement learning methods for continuous-time Markov decision problems. *Advances in Neural Information Processing Systems,* **7,** 393-400.

Canessa, S., Guillera-Arroita, G., Lahoz-Monfort, J.J., Southwell, D.M., Armstrong, D.P., Chadès, I., Lacy, R.C. & Converse, S.J. (2015) When do we need more data? A primer on calculating the value of information for applied ecologists. *Methods in Ecology and Evolution,* **6,** 1219-1228.

Chadès, I., Martin, T.G., Nicol, S., Burgman, M.A., Possingham, H.P. & Buckley, Y.M. (2011) General rules for managing and surveying networks of pests, diseases, and endangered species. *Proceedings of the National Academy of Sciences of the United States of America,* **108,** 8323-8328.

Chadès, I., McDonald-Madden, E., McCarthy, M.A., Wintle, B., Linkie, M. & Possingham, H.P. (2008) When to stop managing or surveying cryptic threatened species. *Proceedings of the National Academy of Sciences of the United States of America,* **105,** 13936.

Clark, J.S. (2005) Why environmental scientists are becoming Bayesians. *Ecology Letters,* **8,** 2-14.

Duke, J.M., Dundas, S.J. & Messer, K.D. (2013) Cost-effective conservation planning: Lessons from economics. *Journal of Environmental Management,* **125,** 126-133.

Firn, J., Rout, T., Possingham, H. & Buckley, Y.M. (2008) Managing beyond the invader: manipulating disturbance of natives simplifies control efforts. *Journal of Applied Ecology,* **45,** 1143-1151.

Forsell, N., Wikström, P., Garcia, F., Sabbadin, R., Blennow, K. & Eriksson, L.O. (2011) Management of the risk of wind damage in forestry: a graph-based Markov decision process approach. *Annals of Operations Research,* **190,** 57-74.

Grechi, I., Chadès, I., Buckley, Y., Friedel, M., Grice, A.C., Possingham, H.P., van Klinken, R.D. & Martin, T.G. (2014) A decision framework for management of conflicting production and biodiversity goals for a commercially valuable invasive species. *Agricultural Systems,* **125,** 1-11.

Hauser, C.E. & Possingham, H.P. (2008) Experimental or precautionary? Adaptive management over a range of time horizons. *Journal of Applied Ecology,* **45,** 72-81.

Hawley, W. (1988) The biology of *Aedes albopictus*. *Journal of the American Mosquito Control Association. Supplement,* **1,** 1-39.

Hill, M.P., Axford, J.K. & Hoffmann, A.A. (2014) Predicting the spread of *Aedes albopictus* in Australia under current and future climates: Multiple approaches and datasets to incorporate potential evolutionary divergence. *Austral Ecology,* **39,** 469-478.

Hoey, J., St-Aubin, R., Hu, A. & Boutilier, C. (1999) SPUDD: Stochastic planning using decision diagrams. *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*, pp. 279-288. Morgan Kaufmann, Stockholm, Sweden.

Houston, A., Clark, C., McNamara, J. & Mangel, M. (1988) Dynamic models in behavioural and evolutionary ecology. *Nature,* **332,** 29-34.

Accepted Article

This article is protected by copyright. All rights reserved.

Li, L., Walsh, T.J. & Littman, M.L. (2006) Towards a unified theory of state abstraction for MDPs. *9th International Symposium on Artificial Intelligence and Mathematics*.

Mantyka-Pringle, C.S., Martin, T.G., Moffatt, D.B., Udy, J., Olley, J., Saxton, N., Sheldon, F., Bunn, S.E. & Rhodes, J.R. (2016) Prioritizing management actions for the conservation of freshwater biodiversity under changing climate and land-cover. *Biological Conservation,* **197,** 80-89.

Marescot, L., Chapron, G., Chadès, I., Fackler, P., Duchamp, C., Marboutin, E. & Gimenez, O. (2013) Complex decisions made simple: a primer on stochastic dynamic programming. *Methods in Ecology and Evolution,* **4,** 872-884.

Martin, T.G., Burgman, M.A., Fidler, F., Kuhnert, P.M., Low-Choy, S., McBride, M. & Mengersen, K. (2012) Eliciting expert knowledge in conservation science. *Conservation Biology,* **26,** 29-38.

McCarthy, M.A., Possingham, H.P. & Gill, A.M. (2001) Using stochastic dynamic programming to determine optimal fire management for *Banksia ornata*. *Journal of Applied Ecology,* **38,** 585-592.

McIntyre, S., Mitchell, D. & Ladiges, P. (1989) Seedling mortality and submergence in Diplachne fusca: a semi-aquatic weed of rice fields. *Journal of Applied Ecology***,** 537-549.

Monterrubio, C.L., Rioja-Paradela, T. & Carrillo-Reyes, A. (2015) State of knowledge and conservation of endangered and critically endangered lagomorphs worldwide. *Therya,* **6,** 11-30.

Nicol, S. & Chadès, I. (2011) Beyond stochastic dynamic programming: a heuristic sampling method for optimizing conservation decisions in very large state spaces. *Methods in Ecology and Evolution,* **2,** 221-228.

Nicol, S., Fuller, R.A., Iwamura, T. & Chadès, I. (2015) Adapting environmental management to uncertain but inevitable change. *Proceedings of the Royal Society of London B: Biological Sciences,* **282,** 20142984.

Pastor-Satorras, R. & Vespignani, A. (2001) Epidemic spreading in scale-free networks. *Physical review letters,* **86,** 3200.

Pelizza, S.A., Scorsetti, A.C., Bisaro, V., Lastra, C.C.L. & Garcıa, J.J. (2010) Individual and combined effects of Bacillus thuringiensis var. israelensis, temephos and Leptolegnia chapmanii on the larval mortality of Aedes aegypti.

Phelan, P., Norris, K. & Mason, J. (1996) Soil-management history and host preference by Ostrinia nubilalis: evidence for plant mineral balance mediating insect–plant interactions. *Environmental Entomology,* **25,** 1329-1336.

Pichancourt, J.B., Chadès, I., Firn, J., van Klinken, R.D. & Martin, T.G. (2012) Simple rules to contain an invasive species with a complex life cycle and high dispersal capacity. *Journal of Applied Ecology,* **49,** 52-62.

Pitt, J.P.W. (2008) Modelling the spread of invasive species across heterogeneous landscapes. Doctor of Philosophy, Lincoln University.

Puterman, M.L. (1994) *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, Inc. New York, NY, USA.

Regan, T.J., Chadès, I. & Possingham, H.P. (2011) Optimally managing under imperfect detection: a method for plant invasions. *Journal of Applied Ecology,* **48,** 76-85.

Ritchie, S.A., Moore, P., Carruthers, M., Williams, C., Montgomery, B., Foley, P., Ahboo, S., Van Den Hurk, A.F., Lindsay, M.D. & Cooper, B. (2006) Discovery of a widespread infestation of *Aedes albopictus* in the Torres Strait, Australia. *Journal of the American Mosquito Control Association,* **22,** 358-365.

Rohanimanesh, K. & Mahadevan, S. (2002) Learning to take concurrent actions. *Advances in Neural Information Processing Systems*, pp. 1619–1626.

Russell, R.C., Williams, C.R., Sutherst, R.W. & Ritchie, S.A. (2005) *Aedes (Stegomyia) albopictus* -a dengue threat for southern Australia? *Communicable diseases intelligence,* **29,** 296-298.

Singh, S. & Cohn, D. (1998) How to dynamically merge Markov decision processes. *Advances in Neural Information Processing Systems*, pp. 1057–1063. MIT Press.

Accepted Article

This article is protected by copyright. All rights reserved.

Venner, S., Chadès, I., Bel-Venner, M.-C., Pasquet, A., Charpillet, F. & Leborgne, R. (2006) Dynamic optimization over infinite-time horizon: Web-building strategy in an orb-weaving spider as a case study. *Journal of theoretical biology,* **241,** 725-733.

Walters, C.J. (1986) *Adaptive management of renewable resources*. Macmillan, New York, USA.

Walters, C.J. & Hilborn, R. (1978) Ecological Optimization and Adaptive Management. *Annual Review of Ecology and Systematics,* **9,** 157-188.

Wilson, K.A., McBride, M.F., Bode, M. & Possingham, H.P. (2006) Prioritizing global conservation efforts. *Nature,* **440,** 337-340.

Online supporting information (two files):

A. The file "Supporting Information.docx" (also available at dx.doi.org/10.6084/m9.figshare.4557562) contains 12 appendices:
    1) Description of Markov decision processes.
    2) Calculation of the number of states of the exact model.
    3) Proofs of upper and lower bounds.
    4) Description of the inputs parameters required for the program.
    5) Table of the effectiveness of different management actions on all Torres Strait Islands.
    6) Belief Bayesian network providing the effectiveness of actions depending on four islands characteristics.
    7) Parameters used in the Cauchy formula for the low and high transmissions.
    8) Human population size and distances between islands.
    9) Mean time until infestation of mainland Australia for the three models and six rules of thumb.
    10) Prioritisation ranking on four islands for low and high transmission probabilities.
    11) Relative errors of model performances compared to the upper bound with different sub-action durations.
    12) Computational times of the exact, lower bound and upper bound models for low transmission probabilities.

B. The file "Simultaneous actions.rar" (also available at dx.doi.org/10.6084/m9.figshare.4557565) contains the MATLAB code used for computational experiments.

Accepted Article

This article is protected by copyright. All rights reserved.

#islands included	2	3	4	5	6	7	8	9	10	11	12	13
Exact	3.91	9.42	11.5	12.2	13.0	13.1	**13.6**	**intractable**				
Lower bound	8.29	14.2	15.1	15.3	15.8	15.7	**15.9**	**15.9**	**15.9**	15.2	14.9	14.8
Highest transmission	2.82	10.3	14.6	14.7	13.9	**16.9**	16.6	16.7	16.8	15.7	15.2	15.7

Table 1: Relative errors (%) of model performances compared to the upper bound with low transmissions probabilities for an increasing number of islands. When the exact performance is unknown, the relative error of any model to the upper bound specifies the guaranteed percentage difference between the model performance and the optimal performance. For example, a relative error of 10% guarantees that this model is within 10% of the optimal performance. The highest relative errors for each model are shown in bold. Intractability occurred due to memory limits.

#islands included	2	3	4	5	6	7	8	9	10	11	12	13
Exact	3.22	**6.34**	5.32	4.85	4.96	4.68	4.44	**intractable**				
Lower bound	7.31	**8.74**	6.78	5.98	6	5.6	5.24	4.78	4.66	5.42	4.2	4.08
Highest transmission	3.14	**8.7**	7.78	4.93	6.53	6.65	6.28	4.93	5.61	3.21	5.14	6.37

Table 2: Relative errors (%) of model performances compared to the upper bound with high transmissions probabilities for an increasing number of islands.

Accepted Article

This article is protected by copyright. All rights reserved.

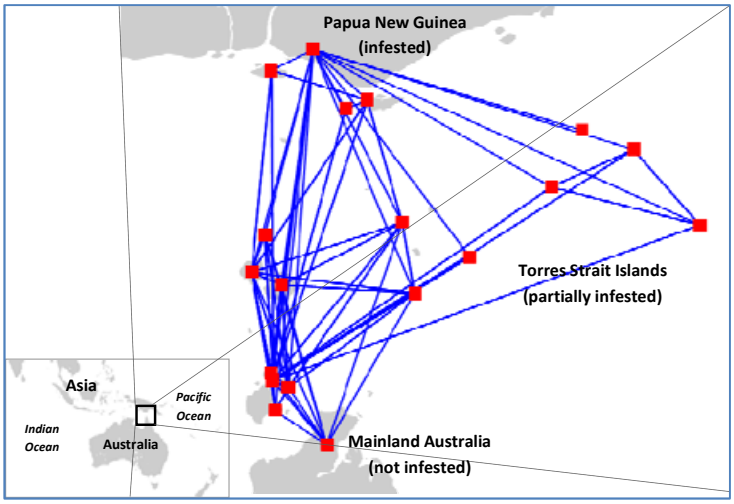


Fig. 1: Map of the Torres Strait showing the nodes of the model (PNG, populated Torres Strait Islands and mainland Australia) as red squares. Blue lines illustrate possible invasion pathways of *Aedes albopictus* between nodes via human-mediated transport including local boats, airplanes or ferries. Pathways with a small transmission probability are not shown for clarity.

Accepted Article

This article is protected by copyright. All rights reserved.

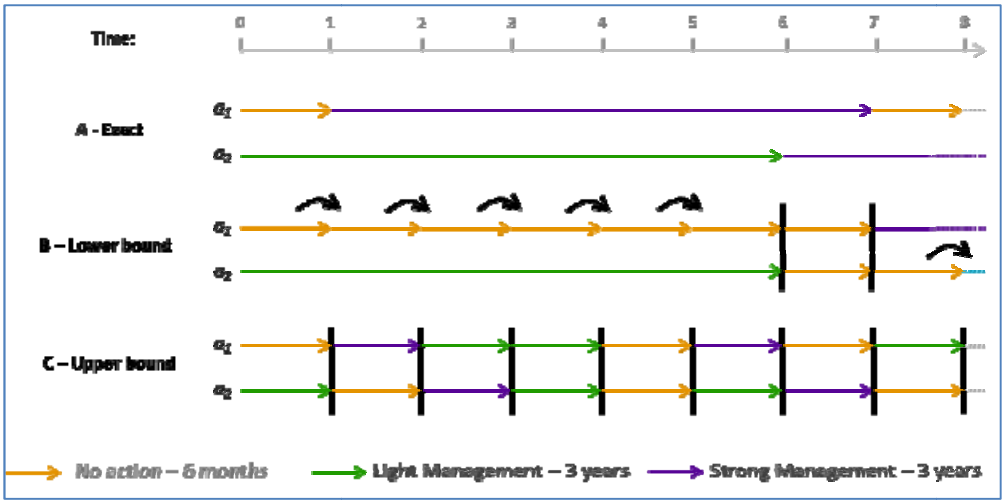


Fig. 2: Schematic of exact (A), lower bound (B) and upper bound (C) policies over eight timesteps. We use the sub-actions of the case study as an example, with $N = 2$ islands and $A_1 = A_2 = \{no\ action, light, strong\}$ of durations one, six and six timesteps respectively (i.e. six months, and three years). Management actions can be asynchronous in the exact model (A). In the lower bound model (B), the curved arrows illustrate the supplementary constraint forcing the management action to continue as the lowest common multiplier of one and six is six. In the upper bound model (C), all actions are interrupted after $GCD(1,6,6) = 1$ timestep. For models B and C, actions do not change between vertical bars allowing a decrease in the number of states.

Accepted Article

This article is protected by copyright. All rights reserved.

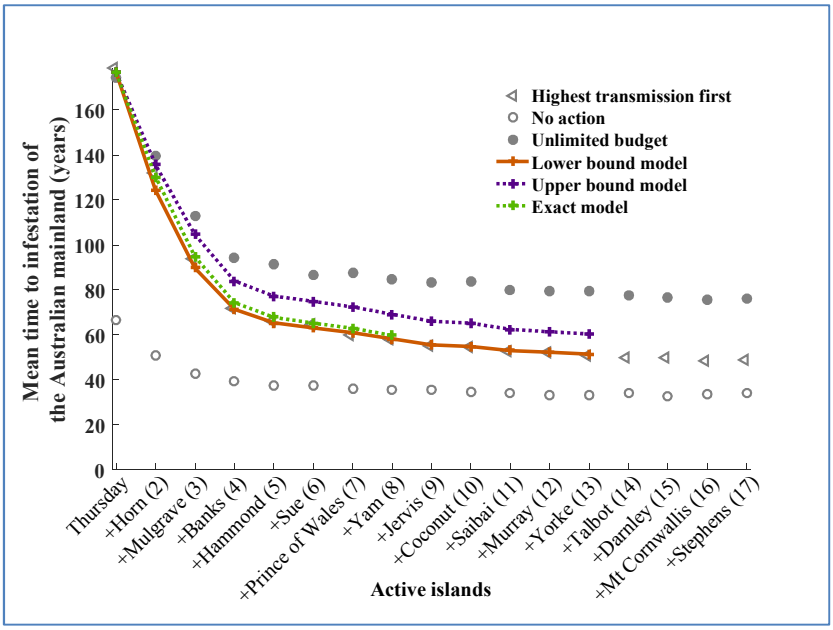


Fig. 3: Mean time to infestation of mainland Australia for the exact, lower bound and upper bound models in the case of low transmission probabilities. Only the best rule of thumb, 'highest transmission', is shown. Islands are progressively included in the analysis until the model is not tractable. The exact model is tractable for up to eight islands, the lower and upper bound model 13 islands. All rules of thumb are tractable up to 17 islands. The 90% confidence intervals are smaller in size than the symbols displayed in the graph and not displayed for clarity.

Accepted Article

This article is protected by copyright. All rights reserved.

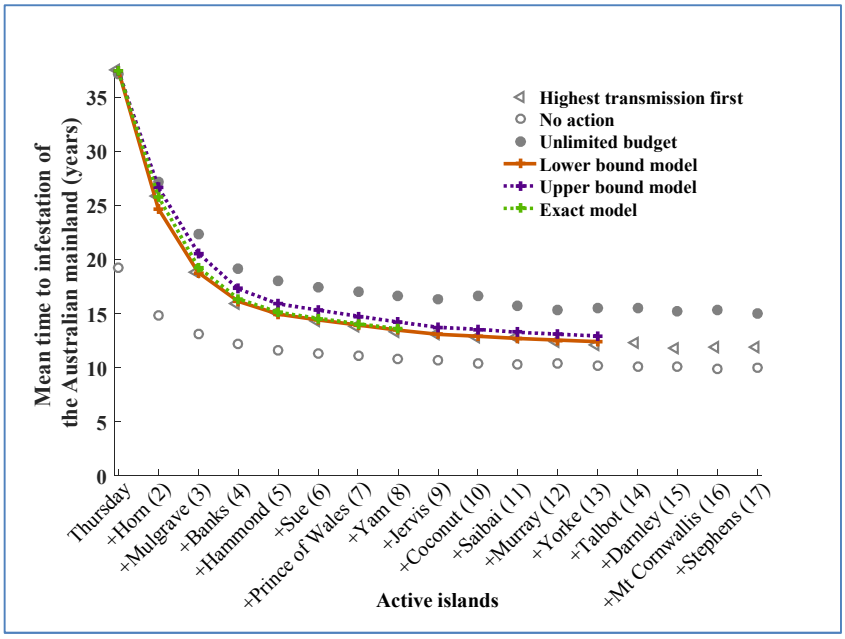


Fig. 4: Mean time to infestation of the Australian mainland for each model with high transmission probabilities, with one to 17 islands included.

Accepted Article

This article is protected by copyright. All rights reserved.

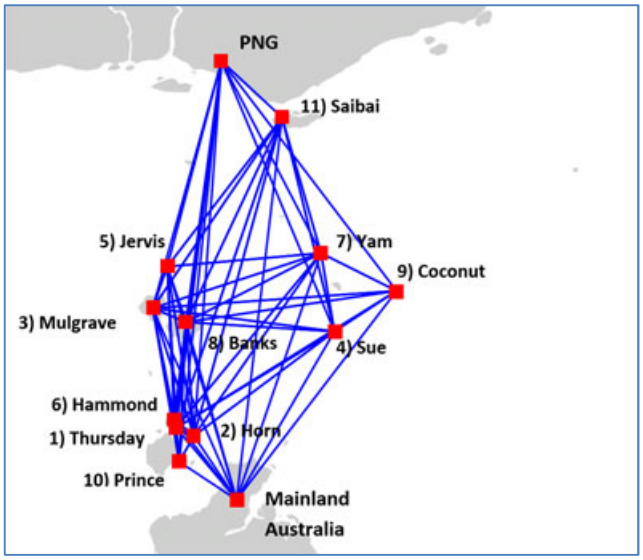


Fig. 5: Prioritisation ranking of 11 islands using the upper bound policy. The rankings that emerge from the exact, lower bound and upper bound policies are the same when tractable. At each timestep, only the two infested islands with highest ranking are managed, due to limited budget.

Accepted Article

This article is protected by copyright. All rights reserved.

Appendix S1:

A Markov decision problem (MDP) is a mathematical framework used to model a sequential decision problem. The system dynamics are partly random and partly under the control of a decision maker (Bellman 1957). When modelling an optimisation problem as a MDP, we assume that the Markov property holds, i.e. the process history has no impact on future dynamics (Puterman 1994).

A MDP is defined by five components: (i) a state space S, (ii) an action space A, (iii) a transition probability matrix P for each action, (iv) immediate rewards r for each state and action and (v) a performance criterion (Puterman 1994). The set of possible action can depend on the current state $s \in S$ and is noted $A(s)$. Solving a MDP means finding a best policy π (action to take in each state, i.e. $\pi: S \rightarrow A$) to maximise (or minimise) the sum of expected future rewards. The performance criterion provides details about the objective (maximisation or minimisation), the time horizon (finite or infinite), the initial state s_0 and the presence of a discount factor (γ). We focus on the maximisation of the discounted sum in infinite time horizon:

$$E^{\pi}\left[\sum_{t=0}^{\infty} \gamma^t r(s_t, a_t) \,|s_0\right] \qquad (eqn\ 1)$$

Stochastic dynamic programming (SDP) denotes a collection of solution methods to solve MDPs, such as policy iteration and value iteration. The peculiarity of policy iteration is that it puts the emphasis on the policy instead of the value: starting from any initial policy π_0, the best current policy π is evaluated (step 1) and improved (step 2) repeatedly until it is optimal (π^*).

1) The evaluation consists of calculating a value (or performance) $V_{\pi}(s)$ for each state $s \in S$. This value corresponds to the sum of future rewards one can expect, starting from the state s when implementing the policy π. Formally,

$$V_{\pi}(s) = E^{\pi}\left[\sum_{t=0}^{\infty} \gamma^t r(s_t, a_t) \,|s_0 = s\right]$$

This can be calculated iteratively (backwards induction) or through a *matrix inversion*, since the value function satisfies the following equation

$$V_\pi(s) \; = r(s,\pi(s)) + \gamma \sum_{s' \in S} P(s' \mid s,\pi(s)) \; \times V_\pi(s') \qquad \forall\, s \in \; S$$

$$V_\pi \; = r_\pi + \gamma P_\pi \, V_\pi$$

with r_π and P_π the reward and transition matrix associated with the policy π. Noting I the identity matrix, this implies

$$V_\pi \; = (I - \gamma P_\pi)^{-1} r_\pi \qquad (eqn\ 2)$$

Note that $I - \gamma P_\pi$ is always invertible when $\gamma < 1$ because P_π is a transition matrix. That we deal with undiscounted sums ($\gamma = 1$) in this manuscript is not an issue because of the absorbing state is reachable from any state and has reward zero.

2) Once the value V_π of the policy π has been evaluated, we can improve this policy by applying Bellman's equation on all states:

$$\pi(s) \; = \underset{a \in \; A(s)}{\operatorname{argmax}} \left[r(s,a) + \gamma \sum_{s' \in S} P(s' \mid s,a) \; \times V_\pi(s') \right] \qquad \forall\, s \in \; S \qquad (eqn\ 3)$$

When equations 2 and 3 are computed several times, π converges to the optimal policy π^*. The outputs of policy iteration (and other SDP techniques) are the optimal policy π^* and the optimal value V_{π^*}. In this manuscript, the value is of high importance because it equals the expected time until the mainland becomes infested starting from a given state.

Appendix S2:

We show how the number of states of the exact model can be calculated. First, note that $t_i = 0$ indicates that the sub-action a_i has just terminated, and thus does not restrict the choice of future

sub-actions; therefore, a_i need not be stored in the state, and can be replaced by 'null'. For each $1 \leq i \leq N$, (a_i, t_i) belongs to $A_i^+ = \{'null', 0\} \cup \{(a_i, t_i): a_i \in A_i, 1 \leq t_i \leq d(a_i) - 1\}$. Then, the state space of the exact model is: $S_{exact} = \{(s, a_1, t_1, a_2, t_2, \dots, a_N, t_N): s \in S, (a_i, t_i) \in A_i^+, 1 \leq i \leq N\} = S \times \prod_{i=1}^{N} A_i^+$.

For each $1 \leq i \leq N$:

$$|A_i^+| = |\{'null', 0\} \cup \{(a_i, t_i): a_i \in A_i, 1 \leq t_i \leq d(a_i) - 1\}|$$

$$= 1 + |\{(a_i, t_i): a_i \in A_i, 1 \leq t_i \leq d(a_i) - 1\}|$$

$$= 1 + \sum_{a_i \in A_i} |\{(a_i, t_i): 1 \leq t_i \leq d(a_i) - 1\}|$$

$$= 1 + \sum_{a_i \in A_i} (d(a_i) - 1)$$

Finally,

$$|S_{exact}| = |S| \prod_{i=1}^{N} |A_i^+| = |S| \prod_{i=1}^{N} \left(1 + \sum_{a_i \in A_i} (d(a_i) - 1) \right)$$

The number of states is exponential in the number of sub-actions N. Also, the exponentiation base grows with the durations of actions.

In our case study, the set of possible actions on each island is $\mathcal{A} = \{no\ action, light\ management, strong\ management\}$ of durations one, six and six timesteps, respectively (i.e. six months, and three years). The number of states, that also accounts for the absorbing state σ ('mainland Australia infested') equals:

$$|S_{exact}| = |\{\sigma\}| + |S| \prod_{i=1}^{N} \big(1 + (d(no\ action) - 1) + (d(light) - 1) + (d(strong) - 1)\big)$$

$$= 1 + 11^N |S|$$

With $|S| = 2^N$ because each of the N islands is either *infested* or *susceptible*.

Appendix S3:

This appendix includes three proofs. We first prove that reducing the action set reduces the performance in a maximisation problem (a). Based on this, we then prove that the lower bound model has a lower performance than the exact model (b). Then, we show that the upper bound model has a higher performance than the exact model (c).

a) Let V and V' denote the performances of any MDPs $< S, A, P, r >$ and $< S, A', P, r >$ and Π and Π' be their sets of policies, respectively. The following holds:

$$[\forall\, s \in S, A'(s) \subseteq A(s)] \;\Rightarrow\; \Pi' \subseteq \Pi$$

It follows, for every state s:

$$V'(s) = \max_{\pi \in \Pi'} E^{\pi}\left[\sum_{t=0}^{\infty} \gamma^t r(s_t, \pi(s_t)) | s_0 = s\right] \leq \max_{\pi \in \Pi} E^{\pi}\left[\sum_{t=0}^{\infty} \gamma^t r(s_t, \pi(s_t)) | s_0 = s\right] = V(s)$$

b) We prove that the lower bound model has a lower performance than the exact model $< S_{exact}, A, P_{exact}, r >$. We consider the definition of the lower bound in the first step (see Materials and Methods), i.e. after addition of the *synchronisation constraint*. In this definition, the lower bound model has the same state space, transition function and rewards as the exact model.

Let A' denote the new action space in the lower bound model. Let $s_{exact} \in S_{exact}$ and $S_{progress}$ denote the states where at least one action is in progress, i.e. $S_{progress} = \{(s, a_1, t_1, a_2, t_2, \dots, a_N, t_N) \in S_{exact} : t_i > 0 \; for\; some\; 1 \leq i \leq N$. If $s_{exact} \in S_{progress}$, the set of possible action $A'(s_{exact})$ only contains the action applied at the previous timestep (in order to extend the action). The set of actions allowed $A'(s)$ is a subset of the action space in the exact model:

$$A'(s_{exact}) \subseteq A(s_{exact}), \forall\, s_{exact} \in S_{progress}$$

For the other states, the set of possible actions is left unchanged:

$$A'(s_{exact}) = A(s_{exact}) \ \forall \ s_{exact} \in S_{exact} \backslash S_{progress}$$

So, $A'(s_{exact}) \subseteq A(s_{exact})$ for all $s_{exact} \in S_{exact}$. This implies (a) that the lower bound model has a lower performance than the exact model.

The second step in designing the lower bound model (see Materials and Methods) is a reformulation that relies on the observation that states where only one action is possible can be removed without increasing or decreasing the performance. However, the transition function and rewards have to be modified to account for the states removed.

c) We prove that the upper bound model has a higher performance than the exact model $< S_{exact}, A, P_{exact}, r >$. We define the set $S_{stop} \subseteq S$, made of states separated by GCD timesteps. Formally,

$$S_{stop} = \{s_{exact} \in S_{exact}, \forall \ 1 \leq i \leq N, GCD \ \mid s_i^T\}$$

All actions are stopped after GCD timesteps. Equivalently, we can modify the action space by setting

$$A''(s_{exact}) = \prod_{i=1}^{N} A_i \, , \forall \ s_{exact} \in \ S_{stop}$$

i.e. all actions are possible in states S_{stop}. We have:

$$A(s_{exact}) \subseteq A''(s_{exact}), \forall \ s_{exact} \in S_{stop}$$

For the other states, the set of possible actions is left unchanged:

$$A(s_{exact}) = A''(s_{exact}) \ \forall \ s_{exact} \in S_{exact} \backslash S_{stop}$$

So, $A(s_{exact}) \subseteq A''(s_{exact})$ for all $s \in S_{exact}$. This implies (a) that the upper bound model has a better performance than the exact model. As in the lower bound model, the reformulation in the second step (see Materials and Methods) does not alter the performance of the upper bound.

Appendix S4:

The inputs parameters required for the program are:

- The number of islands N;
- Two $1 \times |A_1|$ arrays describing the durations and costs of each sub-action;
- The budget received per timestep;
- The effectiveness of each action on each island;
- The colonisation probability between each pair of island (including Papua New Guinea and mainland Australia);
- The discount factor γ. The time horizon is infinite.

Appendix S5:

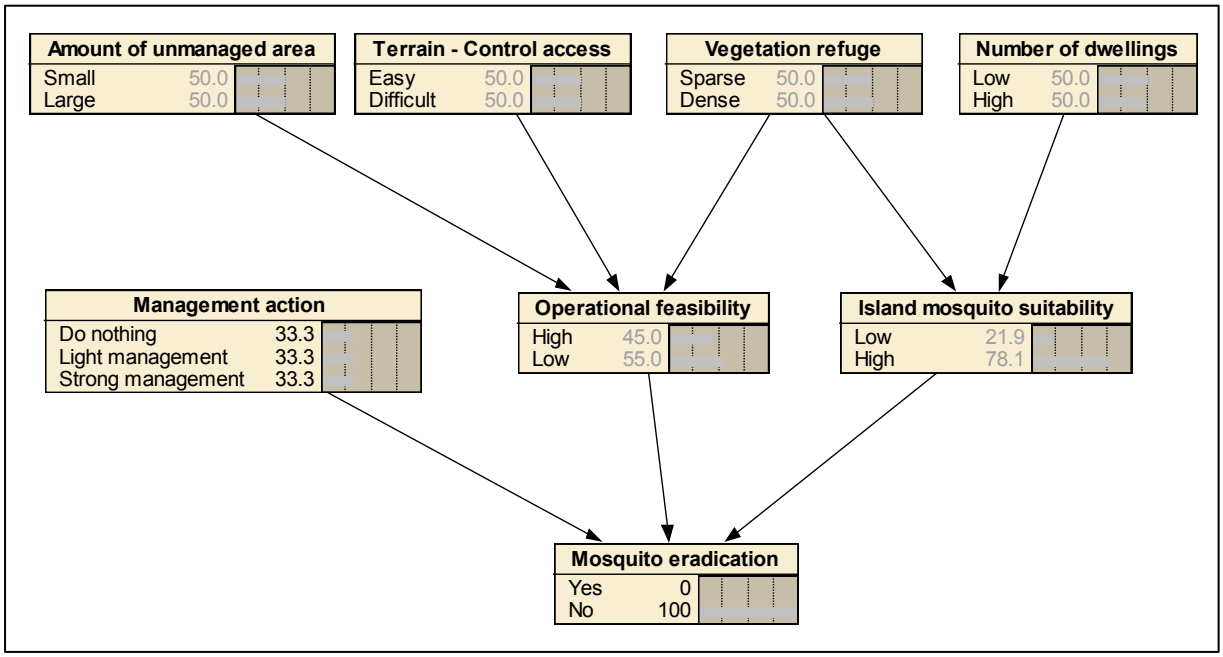
Effectiveness of different management actions on all Torres Strait Islands. The effectiveness of an action is defined as the probability of eradicating the tiger mosquito over one timestep and depends on key characteristics such as vegetation type, number of human dwellings (human population) and terrain type (as a measure of accessibility; Appendix S6). We collected data estimated by experts on management effectiveness for each management action and each island at an expert elicitation workshop held in 2013. Islands are ranked from the highest to lowest probability of infesting the mainland in one timestep (second last column), assuming low transmission probabilities. In our computational experiments, islands are added following this ranking, which is also the rule of thumb 'highest transmission first'. The last column shows the prioritisation ranking that emerges from the upper bound policy for 11 islands.

Management action	**No action**	**Light Management**	**Strong Management**	**Probability of transmission to mainland**	**Management Prioritisation ranking**
Cost	0	x	2x		
Budget (every timestep)	3x				
Duration	six months (one timestep)	three years (six timesteps)	three years (six timesteps)		
Island					
Thursday	0.020379	0.112205	0.173365	0.019841	1
Horn	0.033678	0.114894	0.169567	0.0053089	2
Mulgrave	0.033678	0.114894	0.169567	0.0030053	3
Banks	0.020379	0.036144	0.073427	0.0020876	8
Hammond	0.033678	0.061252	0.10015	0.0015947	6
Sue	0.033678	0.138956	0.203725	0.00098455	4

Prince of Wales	0.033678	0.05848	0.096742	0.00093403	10
Yam	0.020379	0.112205	0.173365	0.00073332	7
Jervis	0.020379	0.112205	0.173365	0.00063267	5
Coconut	0.033678	0.138956	0.203725	0.00038382	9
Saibai	0.020379	0.036144	0.073427	0.00038163	11
Murray	0.026951	0.046704	0.08442	0.00034181	
Yorke	0.033678	0.138956	0.203725	0.00034083	
Talbot	0.020379	0.036144	0.073427	0.00026109	
Darnley	0.028933	0.050074	0.087939	0.00023087	
Mt Cornwallis	0.033678	0.061252	0.10015	0.00017667	
Stephens	0.028933	0.050074	0.087939	0.00005999	

Appendix S6:

Belief Bayesian network providing the effectiveness of actions depending on four islands characteristics. The 11 participants comprised experts in invasive species, vector biology and ecology, mosquito control, public health management and biosecurity. Experts provided anonymous estimates of the actions effectiveness, i.e. probability of eradicating the mosquito over one timestep (six months), for each island and each management action: *no action* (one timestep), *light management* (six timesteps) and *strong management* (six timesteps). The amount of unmanaged area, accessibility/terrain, vegetation refuge and number of dwellings affect the infestation probability. Experts first estimated the operational feasibility and the suitability to mosquitoes for different combinations of these four characteristics (top arrows) and second estimated the probability of mosquito eradication for different combinations of operational feasibility, mosquito suitability and management action (bottom arrows). The effectiveness of any action on any island can then be obtained, provided the four island characteristics are known. Note that the estimates of the actions effectiveness were presented individually and then discussed as a group; subsequently, experts could revise their estimates (Martin *et al.* 2012) before an average was calculated.



Appendix S7:

Parameters used in the Cauchy formula for the low and high transmissions. The Cauchy formula is:

$$p_i^j = C \times \frac{pop_i \times pop_j}{1 + \left(\frac{d_{ij}}{\beta}\right)^2} \quad (\text{Cauchy} - eqn\ 13)$$

Configuration	**Low transmissions**	**High transmissions**
Constant	C = 5×10^{-8}	C = 10^{-7}
Shape parameter **(distances are in km)**	β = 50	β = 50

153 **Appendix S8:**

154 Human population size (bottom row) and distances (shortest coast to coast distance in kilometres)

155 between islands:

ISLAND NAME	AUSTRALIAN MAINLAND	THURSDAY	HORN	MULGRAVE	BANKS	HAMMOND	SUE	PRINCE OF WALES	YAM
AUSTRALIAN MAINLAND	0	27	16	66	53	29	61	16	90
THURSDAY	27	0	2	43	35	1	76	1	94
HORN	16	2	0	45	34	4	69	2	89
MULGRAVE	66	43	45	0	2	38	69	46	66
BANKS	53	35	34	2	0	30	52	38	55
HAMMOND	29	1	4	38	30	0	74	2	91
SUE	61	76	69	69	52	74	0	79	33
PRINCE OF WALES	16	1	2	46	38	2	79	0	98
YAM	90	94	89	66	55	91	33	98	0
JERVIS	86	66	67	10	17	61	73	69	62
COCONUT	91	108	101	95	80	105	30	111	34
SAIBAI	140	134	132	85	85	130	87	138	52
MURRAY	181	211	202	203	188	209	136	212	138
YORKE	140	157	150	138	124	154	80	160	69
TALBOT	157	141	142	85	91	136	118	144	86
DARNLEY	179	199	192	180	168	197	122	203	112
MT CORNWALLIS	138	130	128	79	80	125	91	133	57
STEPHENS	171	186	179	161	150	183	110	190	94
PAPUA NEW GUINEA	151	144	142	93	94	139	100	147	65
POPULATION		2548	586	818	439	212	247	103	313

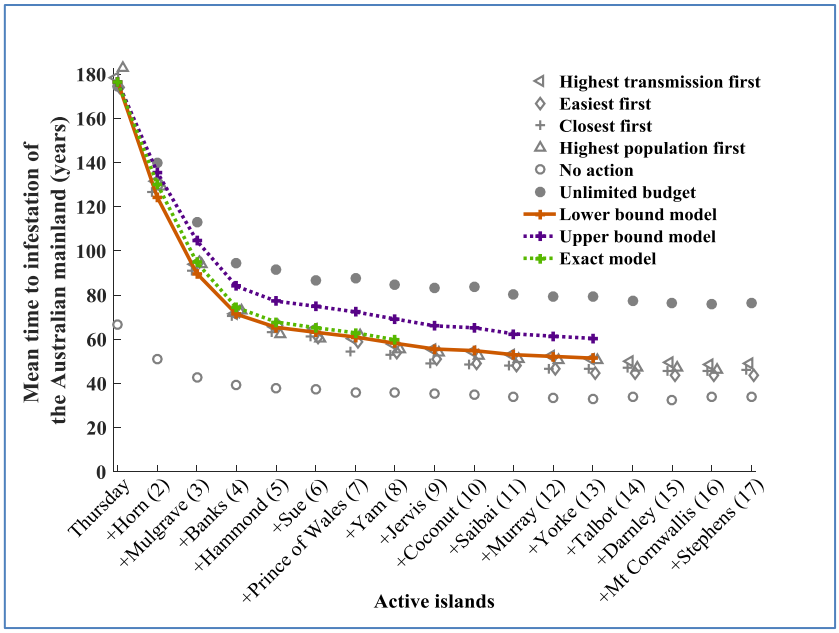
156

ISLAND NAME	JERVIS	COCONUT	SAIBAI	MURRAY	YORKE	TALBOT	DARNLEY	MT CORNWALLIS	STEPHENS	PAPUA NEW GUINEA
AUSTRALIAN MAINLAND	86	91	140	181	140	157	179	138	171	151
THURSDAY	66	108	134	211	157	141	199	130	186	144
HORN	67	101	132	202	150	142	192	128	179	142
MULGRAVE	10	95	85	203	138	85	180	79	161	93
BANKS	17	80	85	188	124	91	168	80	150	94
HAMMOND	61	105	130	209	154	136	197	125	183	139
SUE	73	30	87	136	80	118	122	91	110	100
PRINCE OF WALES	69	111	138	212	160	144	203	133	190	147
YAM	62	34	52	138	69	86	112	57	94	65
JERVIS	0	95	74	202	133	72	175	67	155	81
COCONUT	95	0	78	106	48	120	90	89	79	92
SAIBAI	74	78	0	148	77	36	107	5	83	5

MURRAY	202	106	148	0	70	205	46	172	70	125
YORKE	133	48	77	70	0	133	41	100	30	80
TALBOT	72	120	36	205	133	0	164	30	140	7
DARNLEY	175	90	107	46	41	164	0	134	24	76
MT CORNWALLIS	67	89	5	172	100	30	134	0	109	11
STEPHENS	155	79	83	70	30	140	24	109	0	58
PAPUA NEW GUINEA	81	92	5	125	80	7	76	11	58	0
POPULATION	251	166	337	484	300	284	320	153	76	

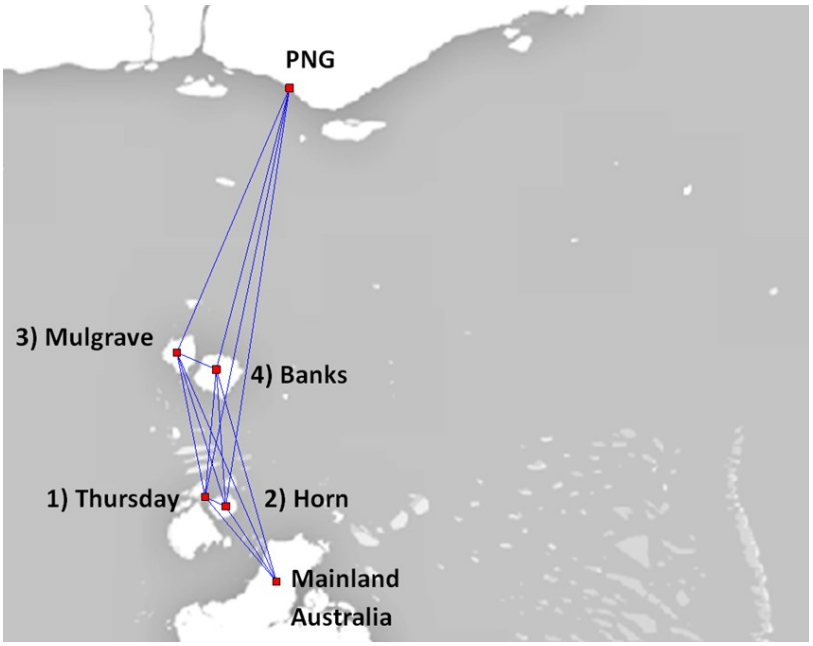
Appendix S9:

Mean time until infestation of mainland Australia for the three models, and six rules of thumb when transmissions are low. The best rule of thumb is 'highest transmission first', followed by 'highest population first', 'closest first' and 'easiest first'.



Appendix S10:

Prioritisation ranking on four islands for low and high transmission probabilities. The rankings that emerge from the exact, lower bound and upper bound models are the same. At each timestep, only the two infested islands with highest ranking are managed, due to limited budget.



Appendix S11:

Relative errors (%) of model performances compared to the upper bound with different sub-action durations. Recall that in our case study, durations are 1, 6 and 6 for a relative error up to 16%. With durations 3, 6 and 6, the LCM and GCD are close: GCD(3,6,6) = 3 and LCM(3,6) = 6, which leads to relative errors less than 6%. With durations 2, 5 and 7, we have GCD(2,5,7) = 1 and LCM(2,5,7)=70. The maximum relative error between the bounds increases and remains under 20%. For durations (3,6,6) and (2,5,7), the bound models are tractable until 12 islands (compared to 13 islands for durations (1,6,6)) because the transition matrices for durations (1,6,6) are sparser than with durations (3,6,6) and (2,5,7). For durations (3,6,6) and (2,5,7), the exact model is intractable above six islands (compared to eight islands for durations (1,6,6)) because long durations mean a high number of states in the exact model.

Durations (no action, light and strong manage-ment)	**Transmis-sion probabi-lities**	**#islands included**	2	3	4	5	6	7	8	9	10	11	12
3, 6, 6	**Low**	**Exact**	1.6	4.9	5.4	5.5	**5.9**	**intractable**					
		LB	1.6	4.9	5.4	5.5	5.9	5.8	**6**	**6**	**6**	5.5	5.3
	High	**Exact**	1.2	**2.4**	1.6	1.3	1.2	**intractable**					
		LB	1.2	**2.4**	1.6	1.3	1.2	1.1	1	0.9	0.8	0.8	0.7
2, 5, 7	**Low**	**Exact**	4.2	10.6	12.6	13.2	**13.8**	**intractable**					
		LB	4.6	18.2	18.6	18.6	**19.4**	19.2	**19.4**	19.2	19.2	18	17.6
	High	**Exact**	3.2	**6.6**	5.5	5	5.1	**intractable**					
		LB	3.2	**8.6**	6.6	5.8	5.9	5.5	5.1	4.7	4.6	4.2	4.1

Appendix S12:

Computational times of the exact, lower bound and upper bound models for low transmission probabilities. The computational times for the exact model are several orders of magnitudes larger than those of the bound models. The computational times have been obtained on a dual 3.46GHz Intel Xeon X5690, which could not solve the largest instances (eight islands for the exact model and 13 islands for the bound models).