



# Optimal allocation of invasive species surveillance with the maximum expected coverage concept

Denys Yemshanov<sup>1\*</sup>, Robert G. Haight<sup>2</sup>, Frank H. Koch<sup>3</sup>, Bo Lu<sup>1</sup>, Robert Venette<sup>4</sup>, D. Barry Lyons<sup>1</sup>, Taylor Scarr<sup>5</sup> and Krista Ryall<sup>1</sup>

<sup>1</sup>Natural Resources Canada, Canadian Forest Service, Great Lakes Forestry Centre, 1219 Queen Street East, Sault Ste. Marie, ON P6A 2E5, Canada, <sup>2</sup>Northern Research Station, USDA Forest Service, 1992 Folwell Ave, St. Paul, MN 55108, USA, <sup>3</sup>Southern Research Station, Eastern Forest Environmental Threat Assessment Center, USDA Forest Service, 3041 Cornwallis Road, Research Triangle Park, NC 27709, USA, <sup>4</sup>Northern Research Station, USDA Forest Service, 1561 Lindig St., St. Paul, MN 55108, USA, <sup>5</sup>Ontario Ministry of Natural Resources, 70 Foster Drive, P6A 6V5 Sault Ste. Marie, ON, Canada

## ABSTRACT

**Aim** We address the problem of geographically allocating scarce survey resources to detect pests in their pathways of introduction given information about their likelihood of movement between origins and destinations. We introduce a model for selecting destination sites for survey that departs from the aim of reducing propagule pressure (PP) in pest destinations and instead aims to increase monitoring of pest origins. The model is a maximum expected coverage problem (MECP), which maximizes the expected number of origins that are covered by the survey system, where an origin is covered if at least one of its transmission pathways connects to a surveyed destination. For comparison, we present two models that aim to reduce PP in destination sites. One model maximizes the expected number of transmission pathways that are covered by survey locations and the other maximizes the expected number of survey locations that have one or more pest introductions.

**Location** United States, Canada.

**Methods** We demonstrate the models by analysing the human-mediated spread of the emerald ash borer (*Agrilus planipennis* Fairmaire), a major pest of ash trees in North America, by visitors to campgrounds in central Canada and the US Midwest. The models incorporate estimates of spread rates from a network of campers travelling from approximately 6500 invaded domains to 266 uninvaded campgrounds in three Canadian provinces (Ontario, Quebec and Manitoba) and three US states (Michigan, Minnesota and Wisconsin).

**Results** The MECP and PP-based model solutions agreed for large surveillance budgets but exhibited differences when the budgets were small. These results stem from differences between the coverage-based objective in MECP and the PP-based metrics in the PP models.

**Main conclusions** Our comparison of MECP and PP-based models reveals the trade-offs between objectives. Overall, the MECP is generic and can be adapted to survey species that are spread via other human-mediated vectors.

## Keywords

emerald ash borer, human-mediated spread, invasive species, maximum expected coverage problem, optimal survey allocation, pathways, propagule pressure.

\*Correspondence: Denys Yemshanov, Natural Resources Canada, Canadian Forest Service, Great Lakes Forestry Centre, 1219 Queen Street East, Sault Ste. Marie, ON P6A 2E5, Canada.  
E-mail: dyemshan@nrcan.gc.ca

## INTRODUCTION

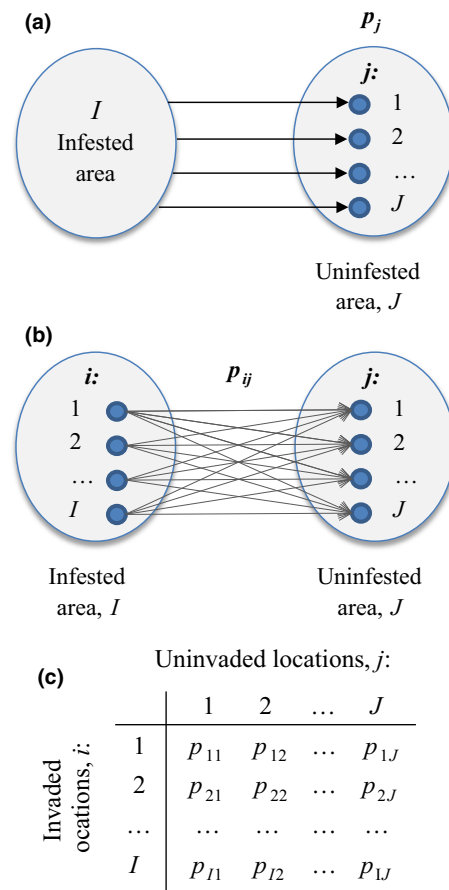
Human-assisted introductions of invasive alien species have resulted in significant economic damages world-wide (Meyerson & Reaser, 2003; Perrings *et al.*, 2005; Hulme

*et al.*, 2008; Aukema *et al.*, 2011). In North America, significant funding has been spent by federal, state and provincial agencies on large-scale pest surveillance programmes to prevent or mitigate these damages (NISC (National Invasive Species Council), 2007; Tobin, 2008).

A fundamental challenge for invasive species managers is deciding where to locate scarce survey resources in uninvaded areas. Many decision models design cost-effective surveillance programmes based on pest propagule pressure (PP) and establishment likelihood concepts. Propagule pressure is a measure of the expected number of individuals (e.g. the number of fertile adults of the species of interest) reaching an uninvaded location and is commonly expressed in terms of the rate, probability or likelihood of arrival (Johnston *et al.*, 2009; Simberloff, 2009). PP is location-specific and characterizes the *destinations* where invasive species may spread from their *origins* (Fig. 1a). Geographical variation in PP is a main determinant of invasive species' spread patterns (Simberloff, 2009). Various surveillance programmes based on PP (or similar establishment likelihood metrics) have been designed to detect and eradicate establishing pest populations before they become large and costly (Hauser & McCarthy, 2009; Cacho *et al.*, 2010; Epanchin-Niell *et al.*, 2012; Hester & Cacho, 2012; Horie *et al.*, 2013). Cost-effective surveillance programmes have also been designed to detect pests in their pathways of introduction with the aim of reducing PP (Surkov *et al.*, 2009; Springborn, 2014).

Here, we address the problem of geographically allocating survey resources to detect invasive pests in their pathways of introduction given information about the likelihood of pest movement between origins and destinations. We introduce a decision model for selecting *destination* sites for pest survey that departs from the aim of reducing PP and instead aims to increase monitoring of pest *origin* locations. Monitoring pest origins is important to enforce or assess pre-border phytosanitary or biosecurity quarantine measures in places where invasive pests are known to exist. Increasingly, biosecurity procedures are switching from border-centred to pathway-centred principles and aim to undertake additional pre-border mitigation measures to minimize the risk of invasive pest entries at the level of origin locations (Tanner, 1997; Hulme *et al.*, 2008; Maynard & Nowell, 2009; Bacon *et al.*, 2012). Monitoring pest origins is also important to assess the attributes of propagules that are transmitted from different sites. These attributes could affect the likelihood of pest establishment in uninvaded sites (Liebhold & Tobin, 2008). The model maximizes the expected number of pest origins that are covered by the survey system, where an origin is considered covered if it could transmit propagules to one or more surveyed sites. We formulate the model as a maximum expected coverage problem (MECP), which was originally designed to choose a fixed number of facility locations to maximize the expected number of demand nodes covered (Daskin, 1983) and has subsequently been used in biological conservation to select reserve sites for species protection (Haight *et al.*, 2000; Arthur *et al.*, 2002; Camm *et al.*, 2002).

The MECP concept is a new approach to the problem of invasive species surveillance. For comparison, we present two other survey models that aim to reduce PP in uninvaded destination locations. The first PP model is based on a



**Figure 1** Basic concept of invasive species' spread in a landscape: (a) – A schematic depiction of the location-specific propagule pressure metric; (b) – a schematic representation of invasive species' spread vectors from the invaded area  $I$  to uninvaded area  $J$ ; (c) –  $I \times J$  matrix with the species' spread rates from the invaded locations  $i$  to uninvaded locations  $j$ .

common knapsack allocation algorithm (Salkin & De Kluyver, 1975; Tulloch *et al.*, 2015) and maximizes the expected number of transmission pathways that are covered by survey locations, while the second PP model maximizes the expected number of survey locations that have one or more pest introductions. Overall, the MECP and PP models provide alternative ways of addressing the pest surveillance problem depending on decision-making goals.

We applied the MECP and PP models to plan the surveillance of the spread of the emerald ash borer (EAB), *Agrilus planipennis* Fairmaire (Coleoptera:Buprestidae), in central Canada and the US. Native to eastern Asia, EAB is a major threat to North American ash (*Fraxinus*) species, all of which are susceptible to attack. Since its initial introduction in Michigan, EAB has already caused significant damage in eastern USA and Canada (Poland & McCullough, 2006; Kovacs *et al.*, 2010). We solved our survey models using estimates of spread likelihoods from a network of campers travelling from approximately 6500 invaded locations in the area of EAB quarantine to 266 uninvaded campgrounds in three Canadian provinces (Ontario, Quebec and Manitoba) and three

US states (Michigan, Minnesota and Wisconsin). We then compared the MECP-based allocations of survey resources with the outcomes of the two models based on PP metrics.

**METHODS**

**Surveillance planning as a maximum expected coverage problem**

Consider a surveillance programme aimed at detecting propagules of an invasive species that is spreading in a heterogeneous landscape. The landscape consists of  $N$  locations of interest, where  $I$  origin locations are known to be invaded by the species and  $J$  destination locations are not invaded yet. The surveillance programme is intended, with a defined budget level  $C$ , to allocate survey resources across the subset of uninvaded locations  $J$ . Assume that the decision-maker can estimate the potential costs,  $c_j$ , of surveys at individual locations,  $j$ , ( $j \in J$ ). Due to budget constraints ( $C$ ), only a portion of uninvaded locations,  $M$ , can be covered by the survey, so  $M$  is a subset of  $J$ .

Any geographic location,  $i$ , in already-invaded areas (i.e. in subset  $I$ ) can serve as a potential origin of the species' spread to uninvaded locations  $j$  in subset  $J$ . Each pair of origin and destination locations,  $ij$ , can be characterized by the relative rate,  $p_{ij}$ , at which the species could spread from invaded location  $i$  to uninvaded location  $j$  along a corresponding vector,  $ij$  (Fig. 1b). Assume that the  $p_{ij}$  values are based on prior knowledge about the species' spread and reflect, in coarse terms, geographical variation in one or more factors deemed responsible for the species' spread from  $i$  to  $j$  (such as transportation, trade, recreation or movement of susceptible host organisms, such as nursery stock). For computational convenience, assume that the relative rate values,  $p_{ij}$ , derived from these factors are rescaled to fit to a 0–1 interval.

The anticipated rates of species spread from the invaded subset  $I$  to uninvaded subset  $J$  can be described by an  $I \times J$  matrix of the  $p_{ij}$  values where the rows,  $i = 1, \dots, I$ , denote the invaded locations and columns,  $j = 1, \dots, J$ , the uninvaded locations (Fig. 1b,c). Essentially, the  $I \times J$  matrix describes a bipartite network of spread vectors  $ij$  connecting each pair of  $i$  and  $j$  locations. We formulate the surveillance planning problem as the selection of particular destination locations  $j$  in subset  $J$  as survey sites to maximize the expected number of origin locations,  $i$ , ( $i \in I$ ), covered by the survey system, where a location is covered if it can transmit propagules to one or more surveyed sites, subject to budget constraint  $C$ . This is a general case of the MECP (Haight *et al.*, 2000; Polasky *et al.*, 2000; Arthur *et al.*, 2002); we adapt our MECP formulation from a biological conservation model described in Camm *et al.* (2002) and Arthur *et al.* (2004).

We start from the  $I \times J$  matrix of the  $p_{ij}$  values denoting the relative likelihood that each uninvaded destination location  $j$  will receive the invasive species from a given origin

location  $i$  (Fig. 1c) and  $0 \leq p_{ij} \leq 1$ . For our example, we focus only on human-mediated spread of invasive alien species and assume that the  $p_{ij}$  values primarily describe long-distance spread events beyond the range of the species' own biological spread capacity. Because we also assume that the geographical extent of our study is very large, and that the spatial resolution exceeds the species' dispersal distance by biological means, we have further made the simplifying assumption that the  $p_{ij}$  values are independent of the likelihood of arrival at adjacent locations within a single survey planning period (which we believe is a fair assumption when considering long-distance, human-mediated spread).

Let  $x_j$  be a binary decision variable that specifies, for all locations in subset  $J$ , whether a given location  $j$  is selected (i.e.  $x_j = 1$ ) or not selected (i.e.  $x_j = 0$ ) for survey. Then, the likelihood that the set of  $ij$  vectors, originating from a particular origin location  $i$ , by which the target species spreads to locations in uninvaded subset  $J$ , is *not* covered by the selected set of survey locations can be estimated as a product of the non-arrival rates over these locations:

$$\prod_{j \in J} (1 - p_{ij}x_j). \tag{1}$$

Equation 1 also implies that location  $i$  is *not covered* if none of its transmission pathways could reach the surveyed sites.

With the independence assumption, the likelihood that the pest entry vectors originating from location  $i$  are covered by at least one surveyed location can be written as:

$$1 - \prod_{j \in J} (1 - p_{ij}x_j). \tag{2}$$

The survey allocation problem can subsequently be formulated to select the set of survey locations that maximizes the expected number of source locations that are covered by the survey system, where a source location,  $i$ ,  $i = 1, \dots, I$  is considered covered if it can transmit propagules to one or more uninvaded survey locations,  $j$ ,  $j = 1, \dots, J$ , subject to a budget constraint:

$$\tau_{MECP} = \text{Max} \sum_{i=1}^I \left( 1 - \prod_{j=1}^J (1 - p_{ij}x_j) \right) \tag{3}$$

s. t.

$$\sum_{j=1}^J c_j x_j \leq C \tag{4}$$

$$x_j \in \{0, 1\} \quad \forall j \in J. \tag{5}$$

Note that the objective function in Eq. 3 only considers how a particular subset of source sites is covered by the surveyed locations, which in turn represent a subset of the uninvaded area. The MECP objective function is nonlinear but has a linearized approximation (Camm *et al.*, 2002; Arthur *et al.*, 2004). In this study, we applied a piecewise

linearized approximation of the MECP objective function from Camm *et al.* (2002) and Arthur *et al.* (2004) (see Appendix S1).

### Propagule pressure models

We compared the MECP model solutions with the solutions generated by two survey planning models based on the PP metric. Propagule pressure model 1 (PP1) follows the common definition of PP as a probabilistic measure of the expected number of individuals arriving at an uninvaded location from the already-invaded area (Simberloff, 2009) and depicts this pressure as the sum of the species' arrival rates to an uninvaded location  $j$  from all invaded (or presumably invaded) locations  $i$ :

$$\sum_{i=1}^I (p_{ij})_j \quad \forall j \in J. \quad (6)$$

The PP1 model allocates the surveys by selecting, from all possible destination locations, the particular subset of locations that maximizes the sum of the species' spread rates, subject to a budget constraint:

$$\tau_{PP1} = \text{Max} \sum_{j=1}^J \left( \left[ \sum_{i=1}^I (p_{ij}) \right]_j x_j \right) \quad (7)$$

$$\text{s.t.} \quad \sum_{j=1}^J c_j x_j \leq C \quad (8)$$

$$x_j \in \{0, 1\} \quad \forall j \in J. \quad (9)$$

Propagule pressure model 2 (PP2) uses a different PP metric that estimates the expected number of destination locations that may receive an invasive species from *one or more* origin locations. Let  $1 - p_{ij}$  be the likelihood that the species is *not* spread from source location  $i$  to uninvaded location  $j$ , and  $\prod_{i=1}^I (1 - p_{ij})$  is the likelihood that the species is *not* spread from one or more of the source locations  $i$  ( $i = 1, \dots, I$ ) to uninvaded location  $j$ . Then, the rate at which the species is spread from one or more source locations to the uninvaded location  $j$  can be estimated as:

$$\left[ 1 - \prod_{i=1}^I (1 - p_{ij}) \right]_j. \quad (10)$$

Using the survey selection binary variable  $x_j$  (defined above), we estimate the likelihood that the species is spread from one or more of the source locations to a surveyed location  $j$  as:

$$\left[ 1 - \prod_{i=1}^I (1 - p_{ij}) \right]_j x_j. \quad (11)$$

Note that when location  $j$  is not surveyed (i.e.  $x_j = 0$ ), the value of Eq. 11 is zero. The survey allocation problem is

formulated as maximizing the expected number of uninvaded destination locations to which a pest could spread and potentially be detected by surveys, subject to a budget constraint:

$$\tau_{PP2} = \text{Max} \sum_{j=1}^J \left( \left[ 1 - \prod_{i=1}^I (1 - p_{ij}) \right]_j x_j \right) \quad (12)$$

$$\text{s. t.} \quad \sum_{j=1}^J c_j x_j \leq C \quad (13)$$

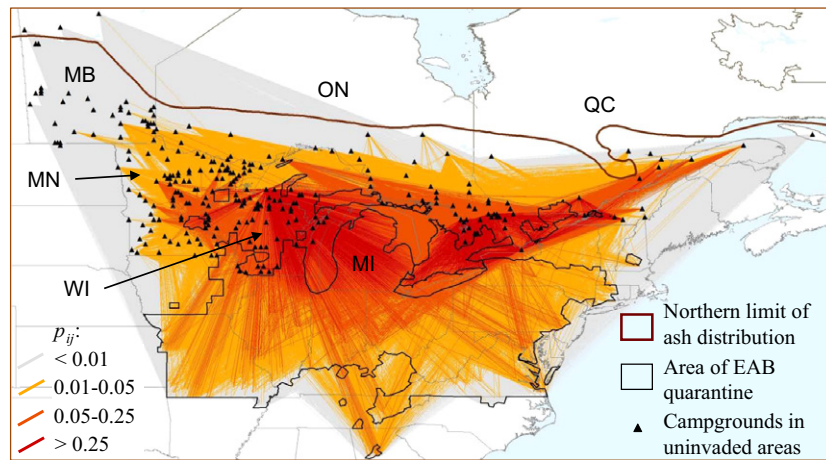
$$x_j \in \{0, 1\} \quad \forall j \in J. \quad (14)$$

In short, the PP1 model tries to capture as much of the total PP as possible with the selected survey locations and uses sums of species arrival rates from the already-invaded area  $I$  to an uninvaded location  $j$ ,  $\sum_{i=1}^I (p_{ij})$ . With the PP2 model, we are trying to capture as many of the likeliest invasion sites as possible with the selected survey sites and use the likelihood that the species moves from one or more locations  $i$  to location  $j$ ,  $1 - \prod_{i=1}^I (1 - p_{ij})$ . Notably, when only a small portion of the uninvaded area can be surveyed, the value in Eq. 10 for the selected survey sites may be close to 1, and the PP2 model could have multiple optimal solutions (ties). Each of these solutions, with similar  $\tau_{PP2}$  values, may have different  $\tau_{MECP}$  and  $\tau_{PP1}$  estimates. For each budget level, we determined whether multiple optimal PP2 solutions were present and reported the solutions with the highest  $\tau_{MECP}$  and  $\tau_{PP1}$  values. For this study, the MECP, PP1 and PP2 models used the same set of survey costs ( $c_j$ ), spread rates ( $p_{ij}$ ) and budget constraints ( $C$ ). The models were composed in the OpenSolver tool ([www.opensolver.org](http://www.opensolver.org)) and solved with the SCIP linear programming solver (<http://www.scip.zip.de>).

### Case study example

We applied the MECP and PP models to plan the surveillance of the spread of the EAB in eastern USA and Canada (Fig. 2). Long-distance EAB spread has been associated with human activities, primarily with commercial and passenger vehicles that could potentially move firewood or other infested materials (Haack *et al.*, 2002; Kovacs *et al.*, 2010; Koch *et al.*, 2012; Yemshanov *et al.*, 2012, 2014). There is also growing evidence that the species could hitchhike on vehicles (Buck & Marshall, 2008) hence suggesting that vehicles arriving at recreational facilities are a potential spread vector. Because the frequency of camper travel between locations has been recognized as a viable predictor of the human-mediated spread of wood-boring pests, including EAB (Haack *et al.*, 2010; Jacobi *et al.*, 2011; Koch *et al.*, 2012), we used a network of campgrounds as potential locations for EAB surveys. The geographical extent of our analysis was defined by the northern limit of the geographical distribution of ash species in central Canada and the





**Figure 2** Origin-destination vectors,  $ij$ , of potential emerald ash borer (EAB) spread with campers travelling from areas under EAB quarantine to campgrounds in uninvaded areas in three US states (Michigan, Minnesota and Wisconsin) and three Canadian provinces (Ontario, Quebec and Manitoba). The solid outline marks the area under EAB quarantine (as of February 2014) and the bold line depicts the northern limit of the ash distribution range (i.e. susceptible EAB host trees). US states: MI, Michigan; WI, Wisconsin; MN, Minnesota; Canadian provinces: ON, Ontario; QC, Quebec; MB, Manitoba.

US Midwest (Little, 1971), the current extent of EAB infestation and by the availability of data that document the movements of campers in Canada and USA. Such data are maintained by provincial ministries of natural resources (MNRs) in Canada, and state departments of natural resources (DNRs) in the USA. In addition, the US National Recreation Reservation Service (US NRRS) maintains an online reservation system at federal campground facilities nationwide (Koch *et al.*, 2014). The available information for Canada included provincial campground reservations for Quebec, Ontario and Manitoba, and in the USA, from Minnesota, Michigan and Wisconsin, as well as the reservations stored in the US NRRS database.

### Model-based assessment of potential EAB spread with campers

In our study, we considered the pathways of human-assisted EAB spread with campers to a network of campground facilities. We used a pathway-based model that simulated the long-distance spread of EAB via a network connecting invaded origin and uninvaded destination locations. The model was based on the frequencies of campers travelling to a network of state, provincial and federal campground facilities from areas currently under EAB quarantine. These frequencies were estimated from the campground reservation data, which provided visitor origin and destination campground locations.

We used the geographical locations of each visitor's origin and destination campgrounds, as well as the total annual number of visitors along unique 'source-destination' paths, to build a spread matrix  $I \times J$  where each element defined the relative rate of camper travel from a presumably invaded source location,  $i$ , to a destination location  $j$ . We tracked destination locations (set  $J$ ) in three Canadian provinces (Ontario, Quebec and Manitoba) and three US states (Michigan, Minnesota and Wisconsin) and origin locations

(set  $I$ ) from areas under EAB quarantine, which we treated as invaded sites (Fig. 2). We then used the matrix  $I \times J$  to simulate spread of EAB through the camper travel network from quarantined (and presumably invaded) areas to uninvaded campgrounds. Based on the stochastic simulations, each pair of invaded source – uninvaded destination locations (i.e. a spread vector  $ij$ ) was characterized by an EAB spread rate value ( $p_{ij}$ ). To keep the computing time reasonable, the spread matrix was assembled at a 15-km spatial resolution. Because our study was focused on assessing the long-distance human-mediated spread of EAB, the cell size was selected to exceed the natural spread capability of EAB (which has typical flight distance of mated females *c.* 3 km (Taylor *et al.*, 2010)). The multiple locations within a 15 km  $\times$  15 km map cell were aggregated into a single source (or destination) location. The final matrix resembled the conceptual diagram shown in Fig. 1(b), where each element depicted the rate of EAB movement,  $p_{ij}$  (Fig. 1c), from invaded location,  $i$ ,  $i \in I$  to campground destination  $j$  in the uninvaded area  $J$ . The size of the matrix was 6572 potentially invaded source locations (15 km  $\times$  15 km map cells) by 266 campgrounds in uninvaded areas (Fig. 2).

### Survey planning scenarios

We used the matrix of EAB spread rates to parameterize the MECP and PP-based survey models. The location-specific estimates of the annual survey costs included two components. The fixed cost portion included trapping supplies, safety and operational supplies, administrative support and salaries of survey staff (assuming deployment of the traps at survey locations and two checkups of EAB emergence in the middle and at the end of the season, plus the costs associated with identifying EAB from the collected material). The variable costs portion included the cost of

travel, vehicle lease and fuel expenses needed to access survey locations. We modelled the variable cost portion as linearly dependent on the distance from the survey location to the nearest urban area with likely availability of qualified personnel to conduct the survey. For Canadian locations, we used the summaries of costs of surveillance programmes provided by the Ontario Ministry of Natural Resources, and for US locations, we used similar data from the Minnesota Department of Natural Resources. The average programme costs for Canadian provinces were estimated at Cdn\$890 per year per surveyed location (in 2013 equivalent), with the corresponding portions of fixed, variable and administrative/supporting costs at 53%, 33% and 14%. The approximate costs for US locations were estimated at US\$696 (CAPS, 2013), and the ratio between fixed and variable costs was 78.5–21.5% (M. Abrahamson, pers. comm., Minnesota Dept. of Agriculture).

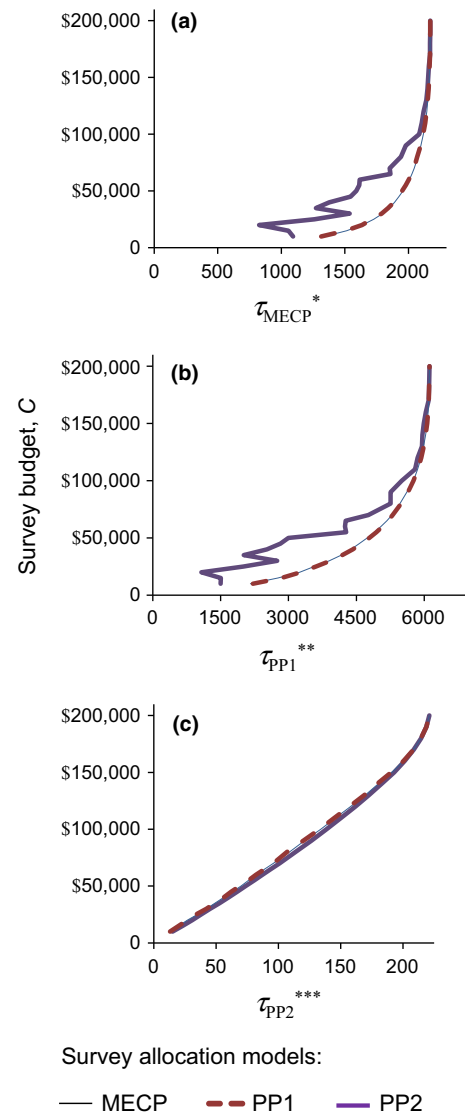
## RESULTS

### Total coverage and the survey budget

Figure 3 illustrates the performance of the MECP, PP1 and PP2 models in terms of the MECP and PP-based objective functions,  $\tau_{MECP}$ ,  $\tau_{PP1}$  and  $\tau_{PP2}$ . The slopes of the curves in Fig. 3 represent the marginal costs, showing the cost of each incremental increase in expected coverage for the MECP model (or in the ‘captured’ PP values for the PP1 and PP2 models). With respect to the  $\tau_{MECP}$  objective function (Fig. 3a), the coverage provided by the MECP model increases quickly as the budget level increases, already covering 75% of possible invaded origins when the budget reaches approximately \$25,000.

The behaviour of the survey models in terms of the  $\tau_{PP1}$  objective function (Fig. 3b) is similar, although in this case the PP1 model shows marginally better performance than the MECP model. The similarity is data driven: the PP values and the degree of connectivity between origin and destination sites are positively correlated. Differences between the cost curves would be more noticeable if the test scenario included multiple sources of isolated infestations that facilitated the spread of a pest to different portions of the uninvaded area, as we would expect to see with spread through a lattice or small-world network with a positive epidemiological threshold (Watts & Strogatz, 1998).

The PP2 model provides less coverage with respect to both the  $\tau_{MECP}$  and  $\tau_{PP1}$  objective functions when the survey budget is below \$100,000. The coverage deficit at low budget levels can be explained by the properties of the PP2 model, which attempts to maximize the capture of the spread likelihood from *one or more* invaded locations to surveyed destination locations in the uninvaded region. Because so many destination locations have multiple connections to invaded locations or are otherwise subject to very high spread rates,  $p_{ij}$ , a large proportion of these locations have species arrival likelihoods from one or more invaded locations,  $1 - \prod_{i=1}^I (1 - p_{ij})$ , that are close to 1. In turn,



**Fig. 3** Objective function value in the survey models as a function of the total survey budget ( $C$ ): (a) – expected coverage objective function values,  $\tau_{MECP}$ , for optimal allocation solutions generated with the maximum expected coverage problem (MECP), propagule pressure model 1 (PP1) and propagule pressure model 2 (PP2) models; (b) – propagule pressure PP1 objective function values,  $\tau_{PP1}$ ; (c) – propagule pressure PP2 objective function values,  $\tau_{PP2}$ ; X-axis – the objective function values: \* $\tau_{MECP}$  – expected number of source locations covered by the surveillance network (Eq. 3); \*\* $\tau_{PP1}$  – expected number of pathways covered by the surveillance network (Eq. 7); \*\*\* $\tau_{PP2}$  – expected number of destination locations covered by the surveillance network (Eq. 12); Y-axis – the survey budget constraint,  $C$ . Differences in the dimensions of the x-axes are due to the distinct formulations of the objective functions  $\tau_{PP1}$ ,  $\tau_{PP2}$  and  $\tau_{MECP}$  (Eqs. 3, 7 and 12).

the allocation of surveys among these high-risk locations is mostly driven by the location-specific survey costs,  $c_j$  (see Eq. 13). The abrupt changes in  $\tau_{PP1}$  and  $\tau_{MECP}$  values in small budget scenarios (Fig. 3a,b) are a result of discrete selections or omissions of individual survey sites.

### Geographical distribution of the potential sources of infestations covered by the survey system

Figure S1 in Appendix S2 provides a geographical depiction of how well the area of EAB quarantine is covered by each survey system. Each location (i.e. map cell) shows the likelihood that EAB will spread from this potentially invaded location  $i$  to one or more survey locations in the uninvaded area,  $(1 - \prod_{j=1}^J (1 - p_{ij})^{x_j})_i$ . The geographic coverage patterns under the MECP and PP1 models are similar: both emphasize large urban centres as well as areas with older infestations and high-density EAB populations (Fig. S1a,b, Appendix S2).

The coverage differences between the MECP and PP-based models follow consistent geographical patterns (Fig. 4). These differences can be explained by concentrated patterns of recreational travel in southern Ontario, where a relatively small number of prominent provincial parks (such as Sandbanks PP or Algonquin PP) receive very large numbers of visitors from large urban centres in the Greater Toronto Area. Alternatively, camper travel in south-western Minnesota and Iowa involves a larger number of low-profile campgrounds. These campgrounds may have lower estimated PP rates but are connected with a large number of locations in the EAB quarantine area, so their selection for surveys in the MECP model yields higher coverage levels (Fig. 4b). The PP2 model provides more even but lower coverage of the origin locations within the EAB quarantine area (Fig. 4c).

The provincial and state allocations of EAB surveys with the MECP, PP1 and PP2 models show some notable differences (Fig. 5). At small budget levels (\$25,000 and below), the PP1 model allocates high budget proportions to Ontario and Michigan, and the lowest proportion to Minnesota. In contrast, the PP2 model allocates lower budget proportions to Ontario and Michigan and higher proportions to Minnesota and Manitoba. At small budget levels, when there are far more locations with high PP rates than can be selected, the PP2 model chooses survey locations based mostly on their relative survey costs, so, when the budget level increases, the selection of additional sites often causes abrupt changes in state/provincial allocations of surveys. The survey models provided different resource allocation between campgrounds in the USA and Canada (Fig. 2, Appendix S2). The MECP model produced the most stable cross-border budget apportionment between the USA and Canada for total survey budgets of \$30,000 or more, whereas the PP1 model allocated a consistently higher portion of the budget to Canada than the MECP allocation, while the PP2 model allocated a lower portion to Canada.

## DISCUSSION

### General behaviour of the MECP and propagule pressure models

Our comparison of the MECP and propagule-pressure-based models revealed some noteworthy contrasts, but also some

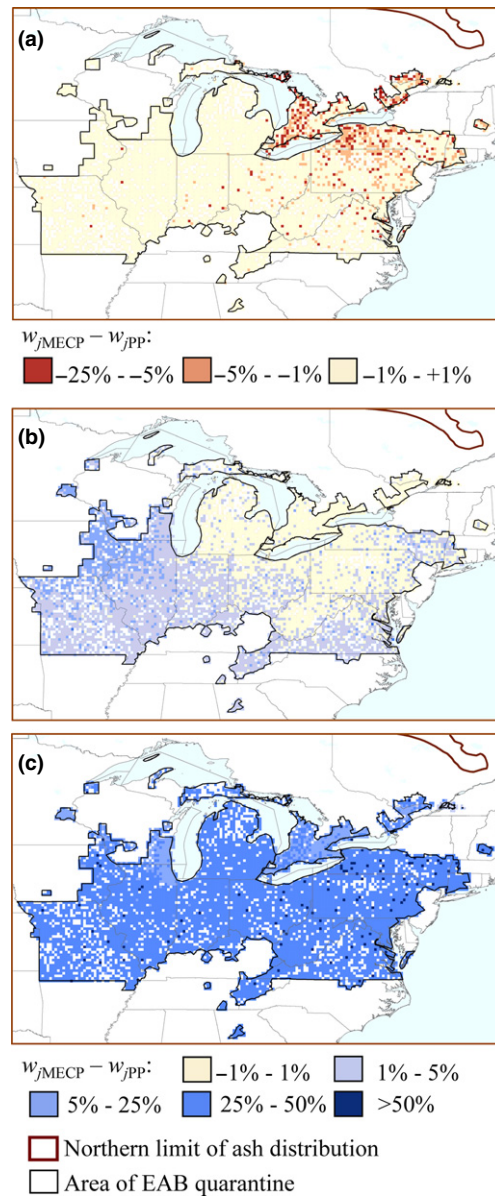


Fig. 4 Differences in coverage patterns,  $w_i = 1 - \prod_{j=1}^J (1 - p_{ij})^{x_j}$  between the optimal scenarios based on the maximum expected coverage problem (MECP) model, propagule pressure model 1 (PP1) and propagule pressure model 2 (PP2). Scenarios with a \$25,000 budget constraint are shown: (a) – negative differences between the  $w_j$  metrics in MECP and PP1 models (i.e. the areas with  $w_{jPP1} > w_{jMECP}$ ); (b) – positive differences between the  $w_j$  metrics in MECP and PP1 models (i.e. the areas with  $w_{jPP1} < w_{jMECP}$ ); (c) – differences between the  $w_j$  metrics in MECP and PP2 models. The  $w_j$  metric denotes the likelihood of an invasive species spreading from the invaded location  $i$  to one or more uninvaded locations  $j$  covered by the survey system in the uninvaded area  $J$ .

similarities. Fundamentally, differences in the behaviour of the models can be attributed to the distinct formulations of their objective functions. For example, the PP1 model attempts to capture the greatest possible proportion of the species spread

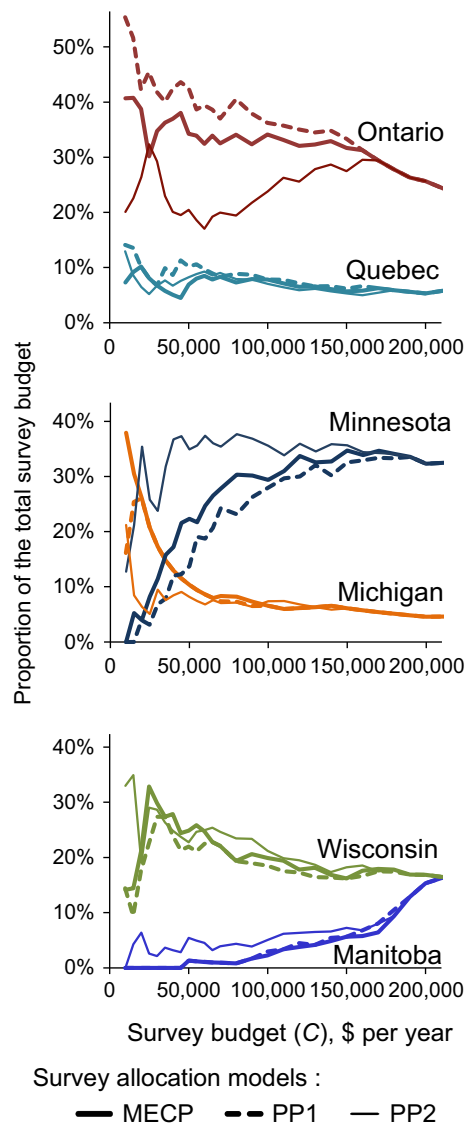


Fig. 5 State and provincial apportionments of the survey budget in the maximum expected coverage problem (MECP) model, propagule pressure model 1 (PP1) and propagule pressure model 2 (PP2).

that is expected to reach the destination locations and allocates surveys based on the trade-off between the survey costs and the PP values. Essentially, it behaves like the classical 'knapsack problem'. Alternatively, the objective function for the PP2 model prioritizes the destination locations that have high probability of *one or more* transmissions from invaded locations (which means that these locations may get infested without specifying the PP rates). At small budget levels, the locations with very high PP rates according to the PP1 model metric all had the probability of receiving one or more pest transmissions close to 1. For this group of high-risk locations, the selection of locations for survey was further stratified in the PP2 model by the location-specific survey costs,  $c_j$ , so the total number of sites surveyed for a given budget would be higher with the PP2 model than with the PP1 model.

Compared to the PP1 model, the MECP model appears to be more influenced by the degree of connectivity between the invaded and uninvaded sites. In the MECP model, destination locations that are highly connected, and therefore may receive pest transmissions from many source locations, are prioritized for survey. For example, between two candidate sites with the same PP levels ( $p_j$ ), the MECP model would select the site that can receive propagules from a larger number of invaded sites. This is consistent with the objective of the MECP model, which is to maximize the coverage of these origin locations.

The degree of similarity between the MECP and PP-based optimal solutions may also depend on the topological properties of the spread network. When the PP estimates and the degree of connectivity between the surveyed and invaded sites are positively correlated, the behaviour of the MECP and PP models will be close. Alternatively, when the connectivity and the PP have little or no correlation, or the species spreads from isolated (but invasive) sources to different parts of the uninvaded area, the solutions generated with the MECP and PP models will show some differences. In those situations, the MECP model will maximize the capacity of the survey network to detect the arrivals of the pest from as many invaded locations as possible, whereas the PP model will select the sites with the highest PP estimates, regardless of how these sites are connected to the invaded source locations.

In our study, the MECP and PP-based model solutions exhibited the most notable differences when the survey budgets were small. A small budget only permits establishment of a few survey sites in the uninvaded area, so the addition or deletion of even one survey site can cause abrupt changes of the objective function value as the budget gets smaller. At low budget levels, the PP1 model tends to select just a few locations with the most extreme PP estimates, whereas the MECP model may choose different locations with lower PP but numerous connections to the invaded sites. The PP2 model chooses the sites with the lowest survey costs because the values of Eq. 10 for the locations with the highest PP values are close or equal to 1, and this generally includes far more candidate locations than can be surveyed given a low budget.

### Potential applications for biosecurity and invasive species surveillance

While the MECP model has higher computational complexity than the PP-based models, the solution times are manageable even for large problems (see Appendix S1). For many applications, the preferred model type will depend on the survey objectives and the amount of available information about the spread of the pest of interest. The MECP and propagule-pressure-based models can be used to evaluate surveillance strategies in particular decision-making situations. Notably, the capacity of the MECP approach to maximize the coverage of invaded source locations helps to



capture geographical variation in the quality of propagules at different origin sites. If there is variability in the success rates of propagules coming from different source locations, then accounting for this variability should make surveillance efforts more cost-effective in the long run, which is especially critical when only a small portion of the uninvaded sites can be surveyed due to resource constraints. The MECP model is also useful in situations where surveys are intended to support decisions about restricting the arrival of an invader from particular origin locations, such as the decision to impose restrictions on imports of goods from foreign regions infested with a pest of interest. Because the MECP uses the matrix of individual spread pathways ( $p_{ij}$ ), it can be used to allocate surveys that maximize coverage of the subset of origin locations deemed high threat. This scenario could be implemented by adding a constraint (i.e. a threshold) to the model that specifies the minimum coverage level that needs to be maintained for the identified subset of high-threat origin locations.

The latter example highlights the potential value of the MECP model for pre-border biosecurity measures aimed at preventing the establishment of unwanted organisms. At the very least, successful detections at a selected set of survey locations could provide justification for thorough inspections of the corresponding origin locations, which may in turn uncover incipient (or potentially incipient) species populations. The approach can also be used to plan the surveillance of species that are spread via other human-mediated vectors, such as international trade or passenger transport. In cases where detailed proxy data are available to describe the human-assisted spread of an organism via individual shipments or movements between invaded and uninvaded locations (e.g. via shipping containers, vehicles or marine vessels), the MECP model can be used to allocate surveillance at the level of those individual shipments. For instance, the model could be parameterized from departmental data sources that track overseas shipments of cargoes and containerized goods to domestic inland destinations (such as trade manifest data collected by customs and border protection agencies), or based on the movement of commercial goods via ground transportation as captured by roadside commercial vehicle surveys.

The PP1 model appears to be more useful in knowledge-poor situations, such as when information about individual pathways of spread ( $ij$ ) is unavailable and a decision-maker only has access to destination-specific PP estimates ( $p_j$ ). However, if the PP values are very low and comparable with the amount of estimation error, or if a decision-maker believes that the invasion may have been started (or spread) by populations arriving from multiple isolated locations, the PP2 model would be more useful. The multiplicative metric in the PP2 model (Eq. 10) is coarser than the metric used in the PP1 model but is less affected by the number of locations from which the propagules might originate or by errors in the  $p_{ij}$  values (although it is still expected to be sensitive to the degree of connectivity between invaded and uninvaded

locations). Thus, the PP2 model could be useful for monitoring rare spread events involving very low arrival probabilities, or in cases where the  $p_j$  values are highly uncertain.

## REFERENCES

- Arthur, J.L., Haight, R.G., Montgomery, C.A. & Polasky, S. (2002) Analysis of the threshold and expected coverage approaches to the probabilistic reserve site selection problem. *Environmental Modeling and Assessment*, **7**, 81–89.
- Arthur, J.L., Camm, J.D., Haight, R.G., Montgomery, C.A. & Polasky, S. (2004) Weighing conservation objectives: maximum expected coverage versus endangered species protection. *Ecological Applications*, **14**, 1936–1945.
- Aukema, J.E., Leung, B., Kovacs, K., Chivers, C., Britton, K.O., Englin, J., Frankel, S.J., Haight, R.G., Holmes, T.P., Liebhold, A., McCullough, D.G. & Von Holle, B. (2011) Economic impacts of non-native forest insects in the continental United States. *PLoS ONE*, **6**, e24587.
- Bacon, S.J., Bacher, S. & Aebi, A. (2012) Gaps in border controls are related to quarantine alien insect invasions in Europe. *PLoS ONE*, **7**, e47689.
- Buck, J.H. & Marshall, J.M. (2008) Hitchhiking as a secondary dispersal pathway for adult emerald ash borer, *Agrilus planipennis*. *The Great Lakes Entomologist*, **41**, 197–198.
- Cacho, O.J., Spring, D., Hester, S. & MacNally, R. (2010) Allocating surveillance effort in the management of invasive species: a spatially-explicit model. *Environmental Modelling and Software*, **25**, 444–454.
- Camm, J.D., Norman, S.K., Polasky, S. & Solow, A.R. (2002) Nature reserve selection to maximize expected species covered. *Operations Research*, **50**, 946–955.
- CAPS (Cooperative Agricultural Pest Survey Program). (2013) CAPS – Annual Accomplishment Report 2012 (01/01/2012–21/31/2012): Survey. Available at: [http://www.ct.gov/caes/lib/caes/documents/caps/accomplishment\\_reports/caps\\_2012ct\\_survey\\_annual\\_accomplishment\\_report.pdf](http://www.ct.gov/caes/lib/caes/documents/caps/accomplishment_reports/caps_2012ct_survey_annual_accomplishment_report.pdf) (accessed 22 July 2014).
- Daskin, M.S. (1983) A maximum expected covering location model: formulation, properties and heuristic solution. *Transportation Science*, **17**, 48–68.
- Epanchin-Niell, R.S., Haight, R.G., Berec, L., Kean, J.M. & Liebhold, A.M. (2012) Optimal surveillance and eradication of invasive species in heterogeneous landscapes. *Ecology Letters*, **15**, 803–812.
- Haack, R.A., Jendek, E., Liu, H., Marchant, K.R., Petrice, T.R., Poland, T.M. & Ye, H. (2002) The emerald ash borer: a new exotic pest in North America. *Newsletter of the Michigan Entomological Society*, **47**, 1–5.
- Haack, R.A., Petrice, T.R. & Wiedenhoft, A.C. (2010) Incidence of bark- and wood-boring insects in firewood: a survey at Michigan's Mackinac Bridge. *Journal of Economic Entomology*, **103**, 1682–1692.
- Haight, R.G., ReVelle, C.S. & Snyder, S.A. (2000) An integer optimization approach to a probabilistic reserve site selection problem. *Operations Research*, **48**, 697–708.

- Hauser, C.E. & McCarthy, M.A. (2009) Streamlining 'search and destroy': cost-effective surveillance for invasive species management. *Ecology Letters*, **12**, 683–692.
- Hester, S.M. & Cacho, O.J. (2012) Optimization of search strategies in managing biological invasions: a simulation approach. *Human and Ecological Risk Assessment*, **18**, 181–199.
- Horie, T., Haight, R.G., Homans, F.R. & Venette, R.C. (2013) Optimal strategies for the surveillance and control of forest pathogens: a case study with oak wilt. *Ecological Economics*, **86**, 78–85.
- Hulme, P.E., Bacher, S., Kenis, M., Klotz, S., Kuhn, I., Minchin, D., Nentwig, W., Olenin, S., Panov, V., Pergl, J., Pysek, P., Roques, A., Sol, D., Solarz, W. & Vila, M. (2008) Grasping at the routes of biological invasions: a framework for integrating pathways into policy. *Journal of Applied Ecology*, **45**, 403–414.
- Jacobi, W.R., Goodrich, B.A. & Cleaver, C.M. (2011) Firewood transport by national and state park campers: a risk for native or exotic tree pest movement. *Arboriculture & Urban Forestry*, **37**, 126–138.
- Johnston, E.L., Piola, R.F. & Clark, G.F. (2009) The role of propagule pressure in invasion success. *Biological Invasions in Marine Ecosystems. Ecological Studies 104* (ed. by G. Rilov and J.A. Crooks), pp. 133–151. Springer-Verlag, Berlin, Heidelberg.
- Koch, F.H., Yemshanov, D., Magarey, R.D. & Smith, W.D. (2012) Dispersal of invasive forest insects via recreational firewood: a quantitative analysis. *Journal of Economic Entomology*, **105**, 438–450.
- Koch, F.H., Yemshanov, D., Haack, R.A. & Magarey, R.D. (2014) Using a network model to assess risk of forest pest spread via recreational travel. *PLoS ONE*, **9**, e102105.
- Kovacs, K.F., Haight, R.G., McCullough, D.G., Mercader, R.J., Siegert, N.W. & Liebhold, A.M. (2010) Cost of potential emerald ash borer damage in U.S. communities, 2009–2019. *Ecological Economics*, **69**, 569–578.
- Liebhold, A.M. & Tobin, P.C. (2008) Population ecology of insect invasions and their management. *Annual Review of Entomology*, **53**, 387–408.
- Little, E.L. Jr. (1971) *Atlas of United States Trees. Volume 1. Conifers and Important Hardwoods*. USDA For. Serv., Timb. Mgmt Res. Div, Washington, DC. Misc. Pub. No.1146.
- Maynard, G. & Nowell, D. (2009) Biosecurity and quarantine for preventing invasive species. *Invasive Species management: A Handbook of Technologies. Techniques in Ecology and Conservation Series* (ed. by M.N. Clout and P.A. Williams), pp.1–18. Oxford University Press, New York, NY.
- Meyerson, L.A. & Reaser, J.K. (2003) Bioinvasions, bioterrorism, and biosecurity. *Frontiers in Ecology and the Environment*, **1**, 307–314.
- NISC (National Invasive Species Council) (2007) *Fact Sheet: National Invasive Species Council Fiscal Year 2007 Interagency Invasive Species Performance Budget*. Available at: [http://www.invasivespecies.gov/global/org\\_collab\\_budget/organizational\\_budget\\_performance\\_based\\_budget.html](http://www.invasivespecies.gov/global/org_collab_budget/organizational_budget_performance_based_budget.html) (accessed 9 April 2015).
- Perrings, C., Dehnen-Schmutz, K., Touza, J. & Williamson, M. (2005) How to manage biological invasions under globalization. *Trends in Ecology and Evolution*, **20**, 212–215.
- Poland, T.M. & McCullough, D.G. (2006) Emerald ash borer: invasion of the urban forest and the threat to North America's ash resource. *Journal of Forestry*, **104**, 118–124.
- Polasky, S., Camm, J.D., Solow, A.R., Csuti, B., White, D. & Ding, R. (2000) Choosing reserve networks with incomplete species information. *Biological Conservation*, **94**, 1–10.
- Salkin, H.M. & De Kluyver, C.A. (1975) The knapsack problem: a survey. *Naval Research Logistics Quarterly*, **22**, 127–144.
- Simberloff, D. (2009) The role of Propagule pressure in biological invasions. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 81–102.
- Springborn, M.R. (2014) Risk aversion and adaptive management: insights from a multi-armed bandit model of invasive species risk. *Journal of Environmental Economics and Management*, **68**, 226–242.
- Surkov, I.V., Lansink, O., Alfons, G.J.M. & van der Werf, W. (2009) The optimal amount and allocation of sampling effort for plant health inspection. *European Review of Agricultural Economics*, **36**, 295–320.
- Tanner, C. (1997) Principles of Australian quarantine. *The Australian Journal of Agricultural and Resource Economics*, **41**, 541–558.
- Taylor, R.A.J., Bauer, L.S., Poland, T.M. & Windell, K.N. (2010) Flight performance of *Agrilus planipennis* (Coleoptera: Buprestidae) on a flight mill and in free flight. *Journal of Insect Behavior*, **23**, 128–148.
- Tobin, P.C. (2008) Cost analysis and biological ramifications for implementing the gypsy moth Slow the Spread Program. Gen. Tech. Rep. NRS-37. U.S. Department of Agriculture, Forest Service, Northern Research Station, Newtown Square, PA.
- Tulloch, A.I., Maloney, R.F., Joseph, L.N., Bennett, J.R., Di Fonzo, M.M., Probert, W.J., O'Connor, S.M., Densem, J.P. & Possingham, H.P. (2015) Effect of risk aversion on prioritizing conservation projects. *Conservation Biology*, **29**, 513–524.
- Watts, D.J. & Strogatz, S.H. (1998) Collective dynamics of 'small-world' networks. *Nature*, **393**, 440–442.
- Yemshanov, D., Koch, F.H., Lyons, D.B., Ducey, M. & Koehler, K. (2012) A dominance-based approach to map risks of ecological invasions in the presence of severe uncertainty. *Diversity and Distributions*, **18**, 33–46.
- Yemshanov, D., Koch, F.H., Lu, B., Lyons, D.B., Prestemon, J.P., Scarr, T. & Koehler, K. (2014) There is no silver bullet: the value of diversification in planning invasive species surveillance. *Ecological Economics*, **104**, 61–72.

## SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

**Appendix S1** Computational aspects.

**Appendix S2** Geographical distribution of the expected coverage values and the survey budgets in the MECP, PP1 and PP2 models.

## BIOSKETCH

**Denys Yemshanov** is a research scientist in the Canadian Forest Service at Natural Resources Canada. His specialization can be described as development of spatial models focused on various ecological and bioeconomic aspects of forest resources. Areas of his special interest include the development of risk modelling techniques and decision support tools for invasive alien species. His recent work is focused on modelling pathways of human-assisted

spread of invasive alien species with international trade, commercial transportation and recreational travel in North America.

Author contributions: DY: concept for the study, development of the pathway model, adaptation of the MECP model and wrote the first draft; RH: development of the MECP and the PP2 algorithms; FK and RV: development of the PP1 model; FK: linking the problem of human-assisted spread with recreation and biosecurity contexts and linking the pathway model with the campground reservation data; BL: programming case study applications and preparing the outputs; DY, FK, RH and RV: interpreting the performance of the MECP and PP models; DBL, KR and TS: expertise with EAB biology, surveillance and spread; DBL, RV, TS and KR: cost estimates of EAB surveys. All authors contributed to writing and editing of the manuscript.

---

Editor: Brian Leung