Journal of Applied Ecology

Journal of Applied Ecology 2017, **54,** 225–233 doi: 10.1111/1365-2664.12754

Predicting costs of alien species surveillance across varying transportation networks

Laura Blackburn[1]*, Rebecca Epanchin-Niell[2], Alexandra Thompson[2] and Andrew Liebhold[1]

[1]*USDA Forest Service, Northern Research Station, 180 Canfield Street, Morgantown, WV 26505, USA; and*
[2]*Resources for the Future, 1616 P Street NW, Washington, DC 20036, USA*

Summary

1. Efforts to detect and eradicate invading populations before they establish are a critical component of national biosecurity programmes. An essential element for maximizing the efficiency of these efforts is the balancing of expenditures on surveillance (e.g. trapping) versus treatment (e.g. eradication). Identifying the optimal allocation of resources towards surveillance requires an underlying model of how costs and the probability of detection fluctuate with survey intensity across various landscapes. Here, we develop a model, widely applicable across biological systems, for predicating costs associated with varying surveillance intensities across diverse road networks.

2. We assumed that surveillance is conducted across a set of point locations. Resources needed to conduct surveillance include the fixed costs associated with surveying a point (e.g. cost of materials or labour time spent at the survey point) and variable costs that correspond to the expense of the time and distance travelled between points. We estimated travel time and distance between points as functions of surveillance intensity and road network characteristics using data from simulated least cost driving routes connecting points located on real-world road networks. Time and distance estimates were then combined with cost data from an actual gypsy moth *Lymantria dispar* surveillance programme in the state of Washington to predict per trap costs of surveillance across varying road network densities and surveillance intensities.

3. Per point driving time, driving distance and total costs all decline with increasing survey point density and increasing road density. Surveillance intensity (planned point spacing) explains ˜94% of the average time driven per point and 97% of the distance driven per point – thus representing the primary explanatory variable. Incorporating road density and dead end road density explains relatively little additional variance in the model, although they improve goodness of fit.

4. *Synthesis and applications*. This work predicts costs associated with surveillance of invasive species populations. We find that the cost per survey point diminishes with increasing survey point density and also depends on road network characteristics. When combined with maps for the relative risk of alien species establishment across landscapes and measures of surveillance efficacy dependent on effort, these cost predictions can increase efficiency of surveillance and eradication efforts for the gypsy moth and other invasive species.

Key-words: biosecurity, cost-effective monitoring, gypsy moth, invasive alien species, network analyst, non-native species, pest detection and eradication, point sampling, survey costs, vehicle routing

Introduction

Given the serious environmental consequences and economic impacts of biological invasions (Perrings *et al.* 2002; Simberloff *et al.* 2013), there is a need to limit future invasions. A primary strategy for excluding species is preventing their initial arrival via import quarantines, inspection and quarantine treatments (Hulme 2009). However, efforts to limit species entry by most pathways can never be 100% effective; given trends of increasing

*Correspondence author. E-mail: laurablackburn@fs.fed.us

globalization, new species will continue to arrive and establish. Consequently, exclusion via detection and eradication of new populations must play an important role in biosecurity portfolios (Holden, Nyrop & Ellner 2016; Liebhold *et al.* 2016).

A crucial precursor to eradication is the identification of the presence and spatial location of the target population. Early detection of populations is critical, such that they have not grown to a size that renders eradication either impractical or too expensive (Sakai *et al.* 2001; Rejmánek & Pitcairn 2002; Liebhold & Tobin 2008; Tobin *et al.* 2014). Effective early detection necessitates that some sort of surveillance system exists prior to the arrival of the target species. Surveillance may be carried out via visual survey, traps or other types of detection efforts, such as public vigilance, that are distributed across large regions where target species arrival may be anticipated. Often, this is accomplished using a grid or other network of survey points (Berec *et al.* 2015). While such active surveillance programmes are frequently used to detect high-risk species, invading populations of other species may be detected through passive surveillance such as through reporting by the public or industry groups.

Active surveillance for invasive species over large geographic areas is inherently costly. Practitioners must choose how much to invest in surveillance and how to allocate limited surveillance resources across the landscape to minimize the probability or costs of invasion establishment. There is a general need to minimize surveillance costs while still providing information needed to effectively target eradication efforts (Liebhold *et al.* 2016). Several authors have identified that there is an inherent balance between the costs of surveillance and eradication (e.g. Mehta *et al.* 2007; Bogich, Liebhold & Shea 2008; Hauser & McCarthy 2009; Cacho *et al.* 2010; Epanchin-Niell *et al.* 2014). Managers can invest heavily in surveillance, which results in locating invading populations when they are small, thereby reducing eradication costs. Or they may opt to spend less money on surveillance, which results in finding populations when they have grown to a larger average size and are more expensive and less feasible to eradicate. Optimization of an integrated surveillance/eradication programme thus aims to identify the balance between expenditures on surveillance and eradication that minimizes total costs (e.g. Mehta *et al.* 2007; Bogich, Liebhold & Shea 2008; Hauser & McCarthy 2009; Cacho *et al.* 2010; Epanchin-Niell *et al.* 2012).

While bioeconomic models exist for optimization of surveillance/eradication programmes, they depend upon information about how both surveillance and eradication costs vary across landscapes. Modelling surveillance cost is complex because it depends not only upon the number of survey points, but also the costs associated with travel among survey points. Mayo, Straka & Leonard (2003), who conducted the only empirical evaluation of surveillance costs that we are aware of, estimated costs as a function of road density and topography using programme cost data from the gypsy moth Slow-the-Spread program. Their study did not examine the effects of survey intensity on surveillance costs, likely because the density of survey points across regions was relatively homogeneous in this programme. Past surveillance optimization models have made a variety of assumptions about surveillance costs. Some have assumed a fixed cost per survey point or per unit time spent at a site (Bogich, Liebhold & Shea 2008; Hauser & McCarthy 2009; Epanchin-Niell *et al.* 2012), while others have assumed that cost per unit of surveillance effort increases with increasing effort (e.g. Mehta *et al.* 2007; Holden, Nyrop & Ellner 2016). Another study assumed decreasing costs per survey point with increasing surveillance density based on expert opinion (Epanchin-Niell *et al.* 2014). Hence, surveillance cost functions have been quite variable and often had little empirical justification.

In most alien species surveillance programmes, travel between survey points is accomplished via driving on road networks. This involves selecting a driving route to visit each point. Transportation costs of visiting a set of locations can be reduced by optimally sequencing stops to minimize total time or distance travelled. This scheduling challenge is known as the travelling salesperson problem (TSP). Implementing vehicle route optimization (i.e. TSP) is a part of effective decision-making in many fields ranging from solid municipal waste (e.g. Ghose, Dikshit & Sharma 2006; Tavares *et al.* 2009) to delivery and home service fleets (e.g. Weigel & Cao 1999) and bus routes (e.g. Thangiah & Nygard 1992).

Here we modelled surveillance as a TSP to identify travel routes that minimized total time to visit all points in a surveillance grid. We identified optimal travel routes, and associated time and distance per survey point, across a diverse sample of real-world road networks and for surveillance grids of varying point density. From these data, we estimated the travel time and distance to visit a single survey point, as a function of surveillance intensity and road network characteristics. We then illustrate how these results can be used to estimate a surveillance cost function using surveillance for invading gypsy moth *Lymantria dispar* populations in the western USA as a model system.

Materials and methods

STUDY AREAS

Our analysis utilized road network data from 30 sites in three states in the USA: Washington, Iowa and Missouri (Fig. 1). These locations were chosen in order to obtain a range of high to low road densities and varying road topologies, from highly irregular to highly gridded, reflecting both human population density and the topography of the landscape. Each site represents a 12×12 km block (144 km^2); 15 sites are in Washington, eight sites are in Missouri and seven sites are in Iowa. Site locations were selected such that they exhibited visually homogenous road networks within the 12×12 km blocks.

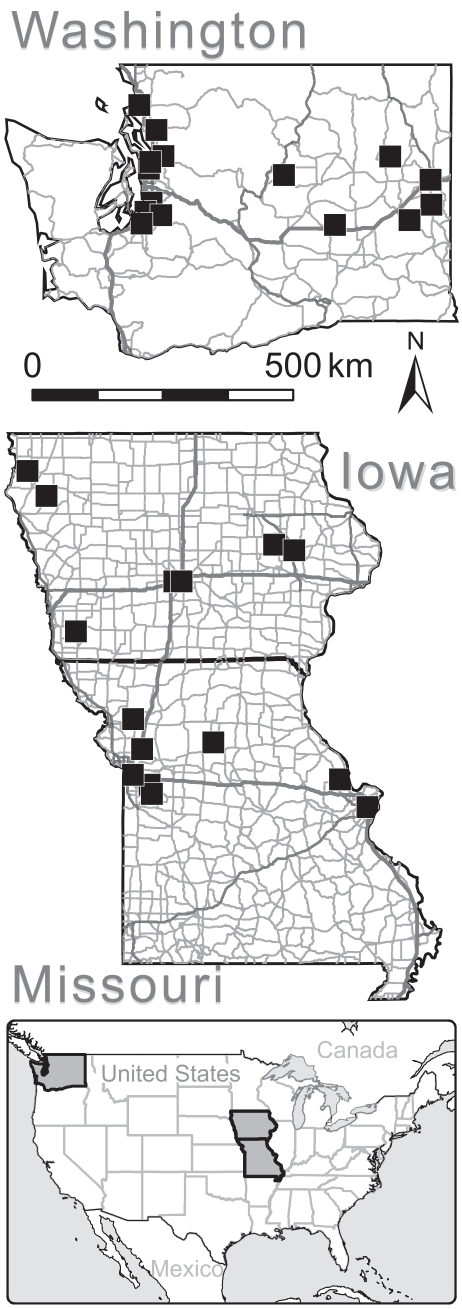

Fig. 1. Study area. The map at the bottom shows the three states used in this analysis, highlighted in grey. On the state maps, the black squares represent the study sites, the thick lines represent interstate highways, and the thin lines represent United States and state highways.

ROUTING SIMULATIONS

Within each of the 30 study blocks, survey point placement was simulated for a series of point densities (0·03, 0·06, 0·11, 0·25, 1 and 4 points km^{-2}), where survey points were located on square grids with intervals of 6, 4, 3, 2, 1 and 0·5 km, respectively, as described below. The ARCGIS 10.1 Network Analyst extension (ESRI, Redlands, CA, USA) was used to identify optimal surveyor driving routes connecting all survey points. Requirements for solving such a vehicle routing problem (VRP) include a network data set (network on which travel occurs), at least one depot (the starting and ending point), at least one order (the destination or survey location) and a layer to record the travel route. The network data set used in this analysis was ESRI's U.S. and Canada Detailed Streets (streets.sdc) from the 2008 Data and Maps DVD set that accompanies the ARCGIS software. Each of the 12 × 12 km study blocks represents the bounds or spatial extent of the driving analysis. Depots were created at the northeast corner of each block and snapped to the nearest road in the road network. PYTHON (Python Software Foundation. Python Language Reference, version 2.7; available at: http://www.python.org) scripting was used to create orders or 'stops' in each study block at 6-, 4-, 3-, 2-, 1- and 0·5-km intervals, such that there are surveillance intensity scenarios ranging from 4 to 576 points per block (see Appendix S1, Supporting Information). Each combination of study block and stop interval represents a unique point feature class (n = 180). Stops were snapped to the nearest and the most major road in the road network within a search radius equal to half of the stop interval. If no road was found within the search radius, that stop was omitted. The route layer identified the starting and ending depot location, but otherwise was a placeholder for the computed route in our analyses.

Developing optimal routes is a key component of the ARCGIS Network Analyst extension that allows users to account for network characteristics such as turn restrictions, speed limits and one-way streets. In this study, the TSP was implemented to determine the most efficient route to visit each order (stop or survey point) and the sequence in which the orders are visited (Fig. 2). To solve a TSP, Network Analyst applies the Dijkstra's algorithm which uses an origin-destination cost matrix between all of the stops to be sequenced and a tabu search-based algorithm, a type of metaheuristic algorithm, to find the best sequence of stops (Dijkstra 1959; ESRI 2014). Attribute parameter values were updated such that road speeds of <32 km h^{-1} were reassigned as 32 km h^{-1}. Python scripting allowed for all routes to be solved programmatically, outputting the order count, total travel time, total travel distance, block ID and stop interval for each optimized route (see Appendix S2).

MEASURING ROAD TOPOLOGY

Both road density and road topology, measures of the transportation network connectivity, affect the efficiency with which drivers transit across a given road network (Litman 2011; Levinson 2013). Road length and the count of dead end roads, both converted to density, were computed for each block in order to provide a basic characterization of road topology. We created python scripts to calculate the length of the roads in each block and the number of dead end roads per block (see Appendices S3 and S4). Total road length and dead end road count for each block were divided by block area to compute road density and dead end road density.

MODELLING DRIVING TIME AND DISTANCE

To predict time and distance travelled between points as functions of planned survey point densities, road network characteristics and their interactions, we utilized multiple linear regression. Because each of the 30 unique study blocks had six observations of varying point density, standard errors of estimated coefficients were likely to be correlated within each block. To ensure robust inference, we

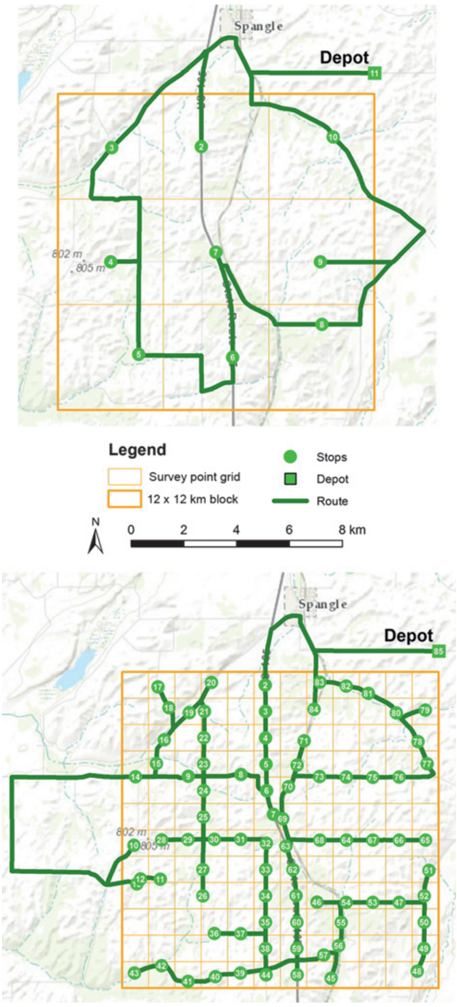

Fig. 2. Route optimization for two different point densities, 0·06 points km^{-2} (top) and 1 point km^{-2} (bottom).

used cluster-robust standard errors, clustered by block. This specification of error terms is robust to heteroscedasticity and allows for correlation of error terms within clusters (Cameron & Miller 2015). In contrast to specifying a random or fixed effects model, using cluster-robust standard errors does not require specific assumptions about the structure of error correlation. All analyses were done using STATA 13.0 (StataCorp. 2013. *Stata Statistical Software: Release 13*. College Station, TX, USA).

Table 1 lists dependent and independent variables, their transformations and summary statistics. Given the substantial range of potential model specifications, we explored relationships between variables both quantitatively and graphically in STATA. Based on this, in our models time, distance and point density (PD) were natural log transformed to achieve linear relationships between point density and the dependent variables and to reduce heteroscedasticity. We did not transform road density (RD), but included its quadratic term, as well as its interaction with the natural log of point density ($\ln PD$) to allow for interdependencies in the effects of surveillance intensity and road density. We also considered the density of dead end roads in a block (DER) and the count of dead end roads per road length (DER_RD) as measures of road network topology.

We evaluated models for travel distance and travel time in a tiered approach, first considering models that only included planned point density ($\ln PD$), then models with both planned point density and road density variables and finally models that also included measures of dead end roads. Models were evaluated and selected using a combination of adjusted R^2 values, variable significance and Akaike and Bayesian information criteria. We present the best models from each tier of models for comparison and because prediction from each tier requires different intensities of data input. Fifteen models were considered at the highest tier, representing each potential combination of road density and dead end road variables, in addition to the natural log of point density.

CALCULATING A SURVEILLANCE COST FUNCTION

We assumed that the total cost per survey point equals the sum of fixed costs associated with a sampled location (e.g. material cost, cost of time spent surveying a location) and the per point variable costs of surveillance, which depend on the time and distance to travel between survey points. Thus, the cost C per survey point can be represented as

$$C = F + V_d \hat{d} + V_t \hat{t}, \qquad \text{eqn 1}$$

where, F is the fixed cost associated with each survey point, V_d and V_t are the costs per unit distance and time travelled,

Table 1. List of dependent and independent variables from each block (n = 180)

Variable	Definition	Units	Mean	SD	Min	Max
Dependent variables						
T	Time cost per point	min	4·26	3·32	0·60	15·78
$\ln T$	Natural log (T)		1·13	0·84	−0·51	2·76
D	Distance cost per point	km	3·37	2·61	0·45	11·63
$\ln D$	Natural log (D)		0·88	0·88	−0·80	2·45
Independent variables						
pPD	Planned point density	traps km^{-2}	0·91	1·43	0·03	4·00
$\ln pPD$	Natural log (pPD)		−1·43	1·69	−3·58	1·39
RD	Road density	km km^{-2}	3·22	2·80	0·57	9·88
RD^2	RD squared		18·15	24·83	0·33	97·64
DER	Dead end road density by block	no. km^{-2}	2·55	3·07	0·02	10·12
DER_RD	Dead end roads per road length ($DER\ RD^{-1}$)	no. km^{-1}	0·57	0·49	0·00	1·47
$\ln pPD \times RD$	Interaction between $\ln pPD$ & RD		−4·59	8·22	−35·41	13·70

respectively, and $\hat{d}$ and $\hat{t}$ are the predicted distance and time between survey points. Parameters F, V_d and V_t may be surveillance programme-specific and estimated based on programme characteristics. Variables $\hat{d}$ and $\hat{t}$ may depend on surveillance intensity and road network characteristics and are the focus of the driving time and distance modelling described above.

GYPSY MOTH APPLICATION

The availability of extensive information about the costs and effectiveness of surveillance systems for gypsy moth makes this species an excellent model system for estimating costs and optimizing surveillance activities (Mayo, Straka & Leonard 2003; Bogich, Liebhold & Shea 2008; Epanchin-Niell *et al.* 2012; Berec *et al.* 2015). Native to temperate Eurasia, the gypsy moth initially established in N. America near Boston, Massachusetts in the late 1860s. Since then, the species has slowly expanded its range, which currently includes most of the north-eastern USA and eastern Canada. Occasionally, gypsy moth life stages are accidentally transported outside of the infested area and may found isolated colonies in distant uninfested portions of the USA such as the Pacific states. Every year, about 100 000 pheromone-baited traps are placed in the uninfested region for surveillance purposes. These traps may detect newly invaded populations and trigger programmes to eradicate them. This large-scale surveillance/eradication programme has been in place for half a century and has been successful at excluding the species from establishing in the western states.

The models we developed for predicting driving time and distance were applied to estimating gypsy moth surveillance cost functions using cost parameter estimates (F, V_d and V_t) derived from information provided by the Washington State Department of Agriculture Pest Program (http://agr.wa.gov/plantsinsects/pest-program). Trappers are paid a wage of \$19·74 h^{-1} (including benefits). Vehicles cost \$261 month^{-1}, which is equivalent to an hourly usage cost of \$1·63. Additional vehicle usage costs are charged at \$0·22 km^{-1} driven. Time spent hanging, checking and retrieving a trap is approximately 20 min, costing \$7·13 per trap in wages and vehicle costs. Trap materials, including the trap, lure and killing agent, cost \$0·75 per trap. Four field supervisors, earning \$4153 month^{-1} including benefits, spend about 75% of their time conducting quality assurance and quality control of trap operations during the trapping season, about four and a half summer months. Using the 2014 number of deployed traps (19 000), supervisors cost \$2·96 per trap. Together, the overhead cost plus the vehicle and labour costs associated with time spent at a trap result in a fixed cost F of \$10·83 per trap. Finally, variable costs of getting to trap locations four times a year were calculated based on travel time and distance predictions developed in our models. Total variable costs equal the estimated travel time $\hat{d}$ per trap multiplied by the cost per h V_t (\$21·38) plus the estimated travel distance $\hat{d}$ multiplied by cost per km V_d, (\$0·22).

$$C = 10{\cdot}83 + 0{\cdot}22\hat{d} + 21{\cdot}38\hat{t}. \qquad \text{eqn 2}$$

Results

ROUTING SIMULATIONS

Simulated routing successfully connected survey points at most planned point densities. However, especially at highest planned point densities in blocks with low road densities, there were some points that were excluded from the survey grid because no road fell within the search radius around the point (Fig. S1).

MODELLING DRIVING TIME AND DISTANCE

Time and distance per survey point consistently decreased with planned point density (Fig. 3). For a given point density, both travel time and distance per point tended to decrease with increasing road density, though the relationship was weak for some point densities.

The natural log of planned survey point density (ln*PD*) was the primary explanatory variable for both time and distance (explaining ~94% and 97% of average time and distance per survey point in a block, respectively). Results are shown in Tables 2 and 3 and Fig. S2 (Models T1 and D1). Coefficients are negative and statistically significant in both time and distance models under all models evaluated. We build upon this simple model to consider the effects of other road network characteristics.

In the second tier of models, including road density variables and an interaction between road density and natural log of planned survey point density (ln*PD* × *RD*) explained an additional 2·4% of variance for the best-fitting time model and <1% for the distance model. For time, the negative coefficient for RD^2 indicates decreasing time per survey point with increasing road

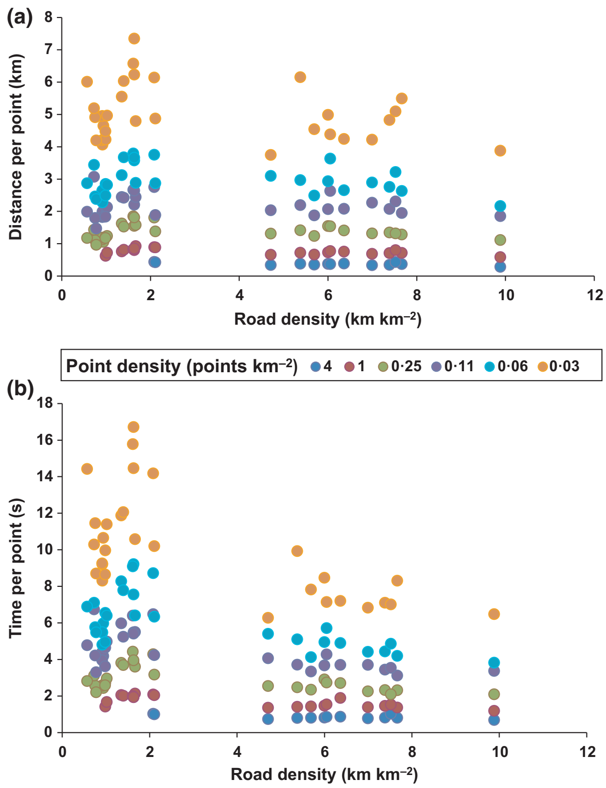

Fig. 3. Graphs of distance (a) and time (b) as a function of road density with colour-coding of planned point density.

Table 2. Estimated coefficients (SE) for select models predicting time per survey point (min point^{-1})

	T1	T2	T3
ln*PD*	−0·482***	−0·507***	−0·509***
	(0·01)	(0·01)	(0·01)
RD^2		−0·004***	−0·001*
		(0·00)	(0·00)
ln*PD* × *RD*		0·008***	0·008***
		(0·00)	(0·00)
DER_RD			0·379***
			(0·08)
DER			−0·064***
			(0·01)
Constant	0·443***	0·513***	0·417***
	(0·03)	(0·03)	(0·04)
R^2	0·937	0·961	0·971
Adjusted R^2	0·936	0·961	0·97
AIC	−45·238	−129·74	−176·753
BIC	−38·852	−116·968	−157·595
RMSE	0·212	0·167	0·146
N	180	180	180

*$P < 0.05$, ***$P < 0.001$.

density (Model T2; Table 2; Fig. S2). The positive coefficient on the interaction term indicates that the effect of road density is reduced at higher planned survey point densities. For distance, the relative magnitudes and signs of road density coefficients indicate a concave quadratic relationship between road density and the natural log of distance, such that travel distance per survey point is highest at intermediate road densities, when controlling for planned survey point density (Model D2; Table 3; Fig. S2).

The above quadratic relationship and observation that driving costs and time were generally highly variable in low-medium density road networks (Fig. 3) supported that road topology may explain some of the variation not explained by survey point density or road density alone. We considered that the density of dead end roads might be a useful characterization of road topology that likely influences driving times and distances. Indeed, we found that the density of dead end roads varies with road density, with the greatest density of dead end roads found at intermediate road densities and the lowest density of dead end roads found in the lowest road densities (Fig. 4).

Including information on dead end roads improved model fit for both the time and distance models according to AIC/BIC/adjusted R^2, but explained <1% additional variance. The best-fitting time model included both dead end road density (*DER*) and dead end road count per length of road (*DER_RD*), as well as the same road density variables as in the best-fitting model without dead end road variables. However, the absolute magnitude and significance of the road density variable declined (Model T3; Table 2). The best-fitting distance model also included both dead end road variables, but excluded road density variables (Model D3; Table 3). When road density variables were included in a model with dead end road variables, they were insignificant and led to a poorer fit according to AIC and BIC indicators (Model D3·2; Table 3).

The effects of dead end road variables in both time and distance models are comparable, with dead end roads per km of road (*DER_RD*) having significant positive effects and dead end road density (*DER*) having significant negative effects (Models T3 & D3; Tables 2 and 3). The latter (*DER*) actually represents the interaction between the former (*DER_RD*) and *RD*, such that increasing road density mediates the positive effect of dead end roads per length of road on time and distance. Practically, these findings suggest that a concentration of dead end roads in an area increases travel time and distance between survey points,

Table 3. Estimated coefficients (SE) for select models predicting distance per survey point (km point^{-1})

	D1	D2	D3	D3·2
ln*PD*	−0·511***	−0·511***	−0·511***	−0·511***
	(0·00)	(0·01)	(0·01)	(0·01)
RD		0·083**		0·034
		(0·03)		(0·05)
RD^2		−0·010***		−0·004
		(0·00)		(0·01)
DER_RD			0·327***	0·284***
			(0·06)	(0·07)
DER			−0·041***	−0·038*
			(0·01)	(0·02)
Constant	0·146***	0·066	0·065*	0·045
	(0·02)	(0·04)	(0·03)	(0·06)
R^2	0·967	0·972	0·976	0·976
Adjusted R^2	0·967	0·971	0·976	0·976
AIC	−148·804	−172·184	−200·262	−198·75
BIC	−142·419	−159·412	−187·491	−179·592
RMSE	0·159	0·148	0·137	0·137
N	180	180	180	180

*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

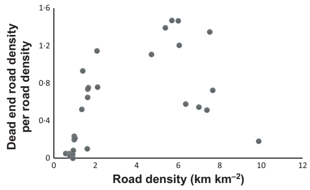

Fig. 4. Density of dead end roads relative to block road density.

but as travel route options improve with increasing road density, the adverse effects of dead end roads are reduced.

The dependent variables in our models (time and distance) were natural log transformed. Prediction of time and distance by simply exponentiating the predicted log values would underestimate their values (Wooldridge 2006). We therefore predict time and distance by multiplying the exponentiated prediction by $\exp(\mathrm{RMSE}^2/2)$, where RMSE is the root mean squared error (Wooldridge 2006). Focusing on the models that account for planned survey point density and road density, travel time (T) and distance (D) per survey point can be predicted as follows:

$$T = 1{\cdot}01\exp(0{\cdot}513 - 0{\cdot}507\ln(PD) - 0{\cdot}00388RD^2 + 0{\cdot}00779(\ln PD \times RD)), \qquad \text{eqn 3}$$

$$D = 1{\cdot}01\exp(0{\cdot}066 - 0{\cdot}511\ln(PD) + 0{\cdot}0823RD - 0{\cdot}0102RD^2)), \qquad \text{eqn 4}$$

where RD is road density (km km^{-2}) and PD is survey point density (points km^{-2}). Surfaces for these models highlight the dominant role of planned point density on both time and distance between survey points (Fig. 5). When focused on models that only include planned survey point density, which explained ~94% and 97% of average time and distance per survey point, respectively, travel time (T) and distance (D) per survey point can be predicted as follows:

$$T = 1{\cdot}02\exp(0{\cdot}443 - 0{\cdot}482\ln(PD)), \qquad \text{eqn 5}$$

$$D = 1{\cdot}01\exp(0{\cdot}146 - 0{\cdot}511\ln(PD)). \qquad \text{eqn 6}$$

GYPSY MOTH APPLICATION

By applying estimated fixed and variable trapping costs for gypsy moth surveys to predicted travel time and distance provided by regression model outputs, we estimated per trap gypsy moth surveillance costs for a range of trap and road densities and generated contours of estimated costs (Fig. 6). Although planned surveillance intensity was the primary driver of travel time and distance under our assumptions, results also demonstrate a quantifiable relationship between travel costs and road network variables. Compared to surveillance intensity, though, these variables have low overall explanatory power.

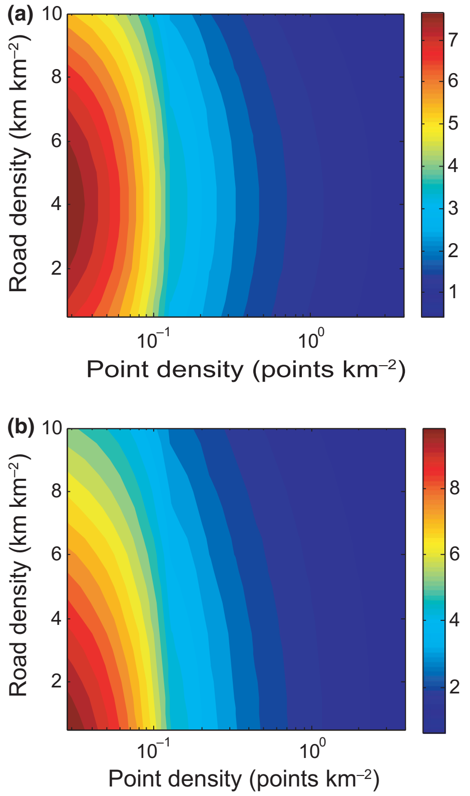

Fig. 5. Contour graphs of modelled driving (a) distance (km point^{-1}) and (b) time (min point^{-1}) as functions of planned point density and road density.

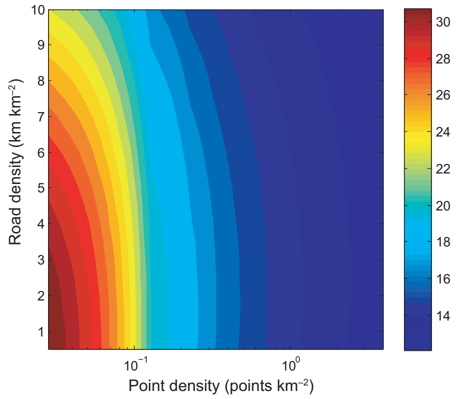

Fig. 6. Contour graph showing modelled total costs per survey point (\$ point^{-1}) as a function of planned point density and road density.

Discussion

The analyses presented here, illustrated using gypsy moth detection programmes in the western USA, are potentially of broad applicability to other invasive species surveillance programmes. In most such programmes, surveillance is conducted across a network of points distributed across a region, and individuals use motor vehicles to access these points. For efforts to optimize the allocation of resources to surveillance across landscapes, the same time and distance functions (eqns 3 and 4) should be widely applicable for estimation of surveillance costs. More broadly, these models may be applicable to other types of ecological sampling. Other applications where point sampling is used for estimation of spatial patterns include species range characterization, detection of rare species and other types of spatial monitoring.

Our modelling results show that planned survey point density is the primary driver of costs, accounting for 94% and 97% of average travel time and distance per point in a block, respectively. Thus, planned survey point density can be used to predict costs (eqns 5 and 6). However, when network data are available and projects are large, including road network characteristics in survey cost models can improve cost predictions and thus improve the cost-efficiency of surveillance plans. While the inclusion of both road density and dead end road density provided the best model fit, the additional explained variance is relatively small, and thus may not be generally of practical relevance. The primary effect of road density is on the number of survey points that can be placed, and thus the total size and cost of a trapping programme rather than per trap costs.

Optimizing surveillance routes by using a TSP algorithm can minimize the operational costs of routing activities such as in surveillance programmes. For example, incorporating route optimization into municipal waste collection greatly reduced route length (Zamorano *et al.* 2009). However, due to the computational complexity of a TSP, routing identified by a GIS, as used here, will likely not provide a true optimal solution to a route with more than 10 stops (Curtin *et al.* 2014). While network analysis software can provide a conservative estimate of time and distance to travel a route, it may not provide the most optimal route when solving a TSP.

One surprising result here was that road density had a relatively small impact on driving time, distance and total costs. The reason for this could be due to the heterogeneity of road speeds, since the routing algorithm used here solves routes by optimizing for faster and higher-level roads. Localities with high road densities tend to occur in areas of high residential densities (Hawbaker *et al.* 2005) and thus are more likely to have low speed limits. Ultimately, it is perhaps convenient that road density has a negligible effect on surveillance costs as it suggests that total surveillance costs can be realistically estimated based solely on the survey point density.

When such a cost model is applied in bioeconomic models for optimization of surveillance intensity, the lack of a strong effect of road density simplifies the problem. However, it should be noted that while road density may not have a strong effect on surveillance costs, it quite likely may be strongly associated with the probability of an alien species colonizing an area. Specifically, there is a tendency for alien species, including the gypsy moth, to arrive in densely populated areas (Colunga-Garcia *et al.* 2010). Such an association may still drive optimal surveillance activities to be concentrated in urban environments even though trapping costs may not be significantly lower there (Epanchin-Niell *et al.* 2014).

Low road density can hinder the ability to navigate to a location. In our simulations, we used a 'fuzzy grid' to plan survey locations. That is, survey points were placed at locations along roads that were closest to the planned grid point centres and within radii of one-half the spacing between grid points. At high survey point densities and low road densities, a substantial fraction of points were omitted because no roads fell within radii (Fig. S2). When omitted survey points comprise a large fraction of all survey locations, this would have obvious adverse impacts on the probability of detecting newly established populations. The magnitude of the radius used around each point would likely affect the fraction of cells with missing survey locations.

The analysis reported contains some obvious simplifications. For example, confining the survey area to a rectangular grid and assignment of the driving depot to the north-east corner of every grid is likely different from real field scenarios though these differences are not likely to substantially affect the conclusions drawn from simulations here. Also, our findings are based on characteristics of US road networks. Travel times could increase in regions with poor road quality or substantially lower typical speed limits. Perhaps a more important character of real surveillance problems is that individual survey personnel may navigate routes that travel through multiple grids of varying point density and may not always follow the cost-minimizing route. However, the relationships among point density, road density, driving time and driving distance are not likely to substantially change, though total trapping costs might be higher than predicted. Future work could evaluate the efficiency of trappers' travel routes and potential gains from improved route planning, such as by using vehicle route optimization software.

In future studies, incorporating multiple transportation modes into route mapping and location analytics will help to further characterize costs associated with surveillance. For example, in our analysis, we assumed that survey points could only be accessed via driving. However, in reality, it may be possible to move between survey points via some combination of driving and walking. In such a case, the problem would be slightly more complex since the costs associated with walking time and distance are likely quite different from those associated with driving. Nevertheless, in many surveillance programmes, this may be more realistic. The ability to walk to survey points would also potentially eliminate the need to drop survey points when no roads are available.

This model provides a means for estimating the cost of broad-scale surveillance and can be applied to a wide range of biological monitoring programmes. In the case of invasive species, when combined with a species landscape establishment model, these predictions will optimize efficiency of detection and eradication efforts. Operational programmes will benefit by incorporating route optimization through a reduced route length, which lends to a more efficient surveillance programme. Ultimately, this model contributes to more cost-effective large-scale surveillance programmes and eliminating invading populations before they establish.

Acknowledgements

We thank James Marra, Washington Department of Agriculture, who provided background information and technical data. We also thank Richard Church, University of California, Santa Barbara for technical advice. This work was made possible, in part, by Agreements from the United States Department of Agriculture's Animal and Plant Health Inspection Service (APHIS) and US Forest Service (USFS) Eastern Forest Environmental Threat Assessment Center. It may not necessarily express APHIS' or USFS' views.

Data accessibility

PYTHON scripts: uploaded as online supporting information, Appendices S1–S4.

References

Berec, L., Kean, J.M., Epanchin-Niell, R., Liebhold, A.M. & Haight, R.G. (2015) Designing efficient surveys: spatial arrangement of sample points for detection of invasive species. *Biological Invasions*, **17**, 445–459.

Bogich, T.L., Liebhold, A.M. & Shea, K. (2008) To sample or eradicate? A cost minimization model for monitoring and managing an invasive species. *Journal of Applied Ecology*, **45**, 1134–1142.

Cacho, O.J., Spring, D., Hester, S. & Mac Nally, R. (2010) Allocating surveillance effort in the management of invasive species: a spatially-explicit model. *Environmental Modelling and Software*, **25**, 444–454.

Cameron, A.C. & Miller, D.L. (2015) A practitioner's guide to cluster-robust inference. *Journal of Human Resources*, **50**, 317–372.

Colunga-Garcia, M., Magarey, R.A., Haack, R.A., Gage, S.H. & Qi, J. (2010) Enhancing early detection of exotic pests in agricultural and forest ecosystems using an urban-gradient framework. *Ecological Applications*, **20**, 303–310.

Curtin, K.M., Voicu, G., Rice, M.T. & Stefanidis, A. (2014) A comparative analysis of traveling salesman solutions from geographic information systems. *Transactions in GIS*, **18**, 286–301.

Dijkstra, E.W. (1959) A note on two problems in connexion with graphs. *Numerische Mathematik*, **1**, 269–271.

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.

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

ESRI (2014) *ArcGIS Help 10.1*. ESRI, Redlands, CA, USA.

Ghose, M., Dikshit, A.K. & Sharma, S. (2006) A GIS based transportation model for solid waste disposal – a case study on Asansol municipality. *Waste Management*, **26**, 1287–1293.

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

Hawbaker, T.J., Radeloff, V.C., Hammer, R.B. & Clayton, M.K. (2005) Road density and landscape pattern in relation to housing density, and ownership, land cover, and soils. *Landscape Ecology*, **20**, 609–625.

Holden, M.H., Nyrop, J.P. & Ellner, S.P. (2016) The economic benefit of time-varying surveillance effort for invasive species management. *Journal of Applied Ecology*, **53**, 712–721.

Hulme, P.E. (2009) Trade, transport and trouble: managing invasive species pathways in an era of globalization. *Journal of Applied Ecology*, **46**, 10–18.

Levinson, D.M. (2013) Access across America. (No. CTS 13-20) p. 38. University of Minnesota, Twin Cities, MN, USA.

Liebhold, A.M. & Tobin, P.C. (2008) Population ecology of insect invasions and their management. *Annual Review of Entomology*, **53**, 387–408.

Liebhold, A.M., Berec, L., Brockerhoff, E.G., Epanchin-Niell, R.S., Hastings, A., Herms, D.A. *et al.* (2016) Eradication of invading insect populations: from concepts to applications. *Annual Review of Entomology*, **61**, 335–352.

Litman, T. (2011) *Evaluating Accessibility for Transportation Planning*. p. 49. Victoria Transport Policy Institute, Victoria, BC, Canada.

Mayo, J.H., Straka, T.J. & Leonard, D.S. (2003) The cost of slowing the spread of the gypsy moth (Lepidoptera: Lymantriidae). *Journal of Economic Entomology*, **96**, 1448–1454.

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

Perrings, C., Williamson, M., Barbier, E.B., Delfino, D., Dalmazzone, S., Shogren, J., Simmons, P. & Watkinson, A. (2002) Biological invasion risks and the public good: an economic perspective. *Conservation Ecology*, **6**, 1.

Rejmánek, M. & Pitcairn, M. (2002) When is eradication of exotic pest plants a realistic goal. *Turning the Tide: The Eradication of Invasive Species* (eds C. Veitch & M. Clout), pp. viii + 414. IUCN SSC Invasive Species Specialists Group, IUCN, Gland, Switzerland and Cambridge, UK.

Sakai, A.K., Allendorf, F.W., Holt, J.S., Lodge, D.M., Molofsky, J., With, K.A. *et al.* (2001) The population biology of invasive species. *Annual Review of Ecology and Systematics*, **32**, 305–332.

Simberloff, D., Martin, J.L., Genovesi, P., Maris, V., Wardle, D.A., Aronson, J. *et al.* (2013) Impacts of biological invasions: what's what and the way forward. *Trends in Ecology and Evolution*, **28**, 58–66.

Tavares, G., Zsigraiova, Z., Semiao, V. & Carvalho, M.G. (2009) Optimisation of MSW collection routes for minimum fuel consumption using 3D GIS modelling. *Waste Management*, **29**, 1176–1185.

Thangiah, S.R. & Nygard, K.E. (1992) School bus routing using genetic algorithms. *Proceedings of the SPIE Conference on Applications of Artificial Intelligence X: Knowledge Based Systems,* **1707**, pp. 387–398. Orlando, FL, USA.

Tobin, P.C., Kean, J.M., Suckling, D.M., McCullough, D.G., Herms, D.A. & Stringer, L.D. (2014) Determinants of successful arthropod eradication programs. *Biological Invasions*, **16**, 401–414.

Weigel, D. & Cao, B. (1999) Applying GIS and OR techniques to solve Sears technician-dispatching and home delivery problems. *Interfaces*, **29**, 112–130.

Wooldridge, J.M. (2006) *Introductory Econometrics: A Modern Approach*, third edn. Thompson South-Western, Mason, OH, USA.

Zamorano, M., Molero, E., Grindlay, A., Rodríguez, M., Hurtado, A. & Calvo, F. (2009) A planning scenario for the application of geographical information systems in municipal waste collection: a case of Churriana de la Vega (Granada, Spain). *Resources, Conservation and Recycling*, **54**, 123–133.

Received 17 March 2016; accepted 12 July 2016
Handling Editor: Jacqueline Beggs

Supporting Information

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

Fig. S1. Relationship between road density and the proportion of 'omitted' points.

Fig. S2. Observed versus fitted points in select models.

Appendix S1. Python script used to create point locations or 'stops' for each study block at various intervals.

Appendix S2. Python script used to programmatically solve all routes and determine the order of stops.

Appendix S3. Python script used to calculate the length of roads in each study block.

Appendix S4. Python script used to calculate the number of dead end roads per block.