View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Harvard University - DASH

DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD

Observer Bias and the Detection of Low-Density Populations

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters.

Citation	Fitzpatrick, Matthew C., Evan L. Preisser, Aaron M. Ellison, and Joseph S. Elkinton. Forthcoming. Observer bias and the detection of low-density populations. Ecological Applications 19.
Published Version	http://esapubs.org/esapubs/journals/applications.htm
Accessed	February 17, 2015 9:31:06 PM EST
Citable Link	http://nrs.harvard.edu/urn-3:HUL.InstRepos:3123278
Terms of Use	This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

(Article begins on next page)

Running head: Observer bias and surveying invasive species

Observer bias and the detection of low-density populations

Matthew C. Fitzpatrick[1,2*], Evan L. Preisser[1], Aaron M. Ellison[2], and Joseph S. Elkinton[3]

[1]Department of Biological Sciences, University of Rhode Island, Kingston, RI 02881

[2]Harvard University, Harvard Forest, Petersham, MA 01366

[3]Department of Plant, Soil, and Insect Sciences, University of Massachusetts at Amherst, Amherst, Massachusetts

*Author to whom correspondence should be addressed:

Matt Fitzpatrick, Harvard Forest,

324 North Main Street, Petersham, MA 01366

e-mail: mfitzpat@fas.harvard.edu; phone: 978 724-3302; fax: 978 724-3595

Manuscript type: *Communication*

Abstract

Monitoring programs increasingly are used to document the spread of invasive species in the hope of detecting and eradicating low-density infestations before they become established. However, interobserver variation in the detection and correct identification of low-density populations of invasive species remains largely unexplored. In this study, we compare the abilities of volunteer and experienced individuals to detect low-density populations of an actively spreading invasive species and we explore how interobserver variation can bias estimates of the proportion of sites infested derived from occupancy models that allow for both false negative *and* false positive (misclassification) errors. We found that experienced individuals detected small infestations at sites where volunteers failed to find infestations. However, occupancy models erroneously suggested that experienced observers had a higher probability of falsely detecting the species as present than did volunteers. This unexpected finding is an artifact of the modeling framework and results from a failure of volunteers to detect low-density infestations rather than from false positive errors by experienced observers. Our findings reveal a potential issue with site occupancy models that can arise when volunteer and experienced observers are used together in surveys.

Keywords: Citizen science, hemlock woolly adelgid, invasive species, monitoring, occurrence probability, site occupancy models, survey, volunteer

Introduction

The growing threat posed by invasive species has focused increased attention on the importance of documenting the distribution and spread of introduced organisms. Monitoring programs aimed at detecting low-density 'founder' populations can play a critical role in slowing or even stopping the spread of harmful invasives by identifying recently-established populations that can be targeted for control and/or eradication (Lodge et al. 2006). Even partially successful programs of this sort can lower densities sufficiently for Allee effects and stochastic events to substantially increase the probability of subsequent population collapse (Liebhold and Tobin 2008). These efforts have proven remarkably successful against actively-dispersing species like the gypsy moth, *Lymantria dispar* L., that respond to pheromones or other cues (e.g., the gypsy moth ‘Slow the Spread’ program; Sharov et al. 2002). Low-density populations of species that disperse passively by means of wind, water, or phoresy, however, often prove far more difficult to locate. Without the ability to attract the organisms to a trapping location, researchers face the often-daunting task of repeatedly searching potential habitats for low-density populations of the invading species.

The challenges of successfully completing the labor-intensive surveys necessary to document the spread of invasive species have been met in part by volunteer-based or ‘citizen science’ monitoring programs (e.g., CitSci.org). Such programs rely on concerned individuals, from schoolchildren to retirees, as cost-effective early warning and continual monitoring systems that provide the primary data for large-scale scientific studies and management responses. There are now more than 200 citizen-science programs operating in North America and their popularity is growing worldwide (Cohn 2008).

Although the educational and scientific benefits of volunteer-based invasive species monitoring programs are clear, the reliability of data collected by novice individuals has sometimes been questioned (Cohn 2008, Delaney et al. 2008). These concerns stem mostly from a lack of studies comparing the quality of volunteer- versus professionally-collected data rather than from studies demonstrating that volunteers collect unreliable data. In the context of monitoring low-density populations of invasive species, the main concern is that novice observers may have a lower probability of detecting the species when it present and/or a higher probability of misidentification (i.e., falsely observing the species as present when it is in fact absent) than do experienced individuals. If true, then differences in the ability of observers to detect and correctly identify low-density populations of invasive species may represent an important, but largely undocumented source of sampling variation and bias in invasive species monitoring programs.

The detectability of species and observer bias both have important implications for documenting current distributions of invasive species and for developing reliable estimates of changes in these distributions. Site occupancy modeling (MacKenzie et al. 2006) has emerged in recent years as a means of estimating the proportion of sites truly occupied by a species given that organisms are often detected imperfectly, i.e., the probability of detecting the species is often less than one. If the probability of detecting a species is <1, as is certainly the case for low-density populations of actively spreading invasive species, then some individuals will go undetected and the actual number of occupied sites will be greater than the number of sites at which the species was actually detected. The initial model developed for estimating site occupancy rates (MacKenzie et al. 2002) considered only the possibility of 'false negatives', cases in which the species is present at a location but goes undetected. Royle and Link (2006) extended the MacKenzie et al. (2002) model to include the possibility of 'false positives', situations in which observers misidentify the target species and report it as present when the species is in fact absent. If misidentifications are common in a survey, then the true number of sites occupied could be less than the number of sites at which the species was observed. Even low false positive rates have been shown to induce extreme bias in estimates of the proportion of occupied sites (Royle and Link 2006), but the impacts of observer bias on estimates of the proportion of sites infested by invasive species remains poorly explored.

In this study, we first compare the abilities of inexperienced volunteers and experienced observers to detect low-density populations of an actively spreading forest pest, the hemlock woolly adelgid. We then use these data to explore the general question of how interobserver variation can bias estimates of the proportion of sites infested derived from occupancy models. We hypothesized that relative to experienced observers, novice individuals should be less likely to detect low-density populations and would be more prone to misidentification of the study species. To explore these hypotheses, we use maximum likelihood methods to select among occupancy models that consider differences in the ability of observers to both detect and correctly identify the hemlock woolly adelgid. We parameterize these models using data from a 420-tree survey conducted by nine volunteers and three experienced individuals. Our results support the notion that volunteers and experienced observers differ in their ability to detect low-density populations and that such differences in observer ability can bias estimates of the proportion of sites occupied. However, this bias manifests itself in unexpected ways.

Materials and Methods

Study species

The hemlock woolly adelgid, *Adelges tsugae* Annand ('HWA'; Hemiptera: Adelgidae) is an actively-spreading invasive pest of eastern hemlock (*Tsuga canadensis* (L.) Carr.) and Carolina hemlock (*Tsuga caroliniana* Englemann) in the eastern United States (McClure and Cheah 1999). HWA is a minuscule (<1-mm long adult), flightless insect that in the US is both obligately parthenogenetic and exclusively passively dispersed (McClure 1990). The parthenogenetic nature of HWA means that even a single colonizing individual can start a new infestation, producing an initially low-density population that only can be detected by costly and time-consuming surveys (Evans and Gregoire 2007). Further, Costa and Onken (2007) list several objects common on hemlock foliage that might be confused with HWA by observers with varying skill levels. These include spider ovisacs, pine sap from adjacent conifers, froth from spittle bugs, and wool from white pine aphids blown from neighboring trees.

Study area

We sampled hemlock trees in the 487-ha Cadwell Memorial Forest in Pelham, Massachusetts (N42.37°, W72.42°), an experimental forest managed by the University of Massachusetts at Amherst. Cadwell Forest is located in the central hardwood region of southern New England and includes discrete stands of eastern hemlock. Before 2007, no HWA infestations had been detected at Cadwell Forest and the local hemlock trees appeared uniformly healthy (J. Elkinton, unpublished data). In the late winter of 2008, however, ad hoc surveys revealed low levels of HWA infestations on several trees. Hemlock stands in this forest thus provide an ideal venue to compare the ability of volunteer and experienced observers to detect early low-density HWA invasions.

Sampling design

Hemlock often grows in nearly monospecific stands that are patchily distributed across the landscape (Ellison et al. 2005). We selected five hemlock stands (~ 1×10^4 m^2 each) for sampling that were primarily (>50%) comprised of hemlock trees ≤10 m in height such that a portion of each tree could be sampled from the ground. All stands were bordered by hardwood forests, allowing the natural boundaries of each stand to be readily identified. Within each stand, all hemlock trees ≥0.5 m in height were numbered using aluminum tags and marked with flagging tape to improve visibility. We marked a total of 420 hemlock trees in the five stands (mean number of trees per stand = 80, range = 31 to 146).

Twelve observers participated in the sampling effort: three experienced individuals who perform field research on HWA and nine volunteers who had no prior experience sampling for HWA. Prior to the sampling, the volunteers were trained for fifteen minutes on the sampling methodology (see below) and on identifying HWA infestations, including objects that could be confused with HWA. Each person was then assigned to one of four groups (n=3 persons per group). Two of the groups entirely were comprised of volunteers (hereafter referred to as 'volunteer-only'). The remaining two groups contained one experienced and two volunteer individuals and two experienced and one volunteer individual (hereafter referred to as 'volunteer/experienced'). Each group was provided a numbered list of trees to sample that could be located in the field by the corresponding numbered tag on each tree. To control for possible heterogeneity in infestation and detection rates between stands, each group was randomly assigned trees to sample in multiple stands.

Our sampling design followed the protocol described by MacKenzie et al. (2006) for a single-species, single-season occupancy model, with individual hemlock trees regarded as sites. Occupancy modeling requires that sites must be visited by at least two independent observers, with each observer recording the presence/absence of the target species at each site. In this study, three observers from the same group visited each tree independently. Observers searched all accessible branches for evidence of white woolly masses characteristic of the HWA sistens generation. Each search continued until either HWA was detected or a two-minute sampling period had expired. To ensure that sampling was independent, no two observers sampled a tree at the same time and observers were instructed not to communicate the infestation status of trees to the other observers in their group. Sampling occurred on April 26th, 2008, when the white woolly masses produced by HWA are at their largest and most visible; this time period is generally considered the optimal sampling period for HWA (Costa and Onken 2007). The sessile nature of the HWA sistens generation precludes any changes in infestation status during our study.

To examine whether there were differences between volunteers and experienced individuals in terms of the density of infestations detected by each type of observer, two experienced individuals involved in the original survey returned the following week to all trees where HWA was detected. All accessible branches thoroughly were searched and the number of white wooly masses observed on the tree was counted. This second, more thorough survey provided an estimate of the number of detectable individuals on the tree. We used a paired t-test on log-transformed HWA abundance to compare the mean abundance of HWA infestations that were detected by any of the nine volunteers to the mean abundance of HWA infestations that were detected by only the three experienced individuals and but not by any of the nine volunteers.

Occupancy modeling

We examined how differences in detection abilities between observers influence estimates of the proportion of infested hemlock trees. The occupancy model framework proposed by Royle and Link (2006) allows the estimation of three parameters: ψ, the proportion of sites occupied (in our case, the proportion of infested hemlock trees), and two classification probabilities. These probabilities are (A) p_{11}, the 'detection probability', the probability of detecting the species, given that the species is actually present at the site; and (B) p_{10}, the 'misclassification probability', the probability of falsely detecting the species at an unoccupied site. Given our randomized sampling design, the number of trees sampled by each observer (minimum n=85, Tables 1 & 2), and the sessile nature of HWA, heterogeneity in detection and misclassification probabilities should result almost entirely from interobserver variation.

We considered four models that make different assumptions regarding p_{11} and p_{10}. The simplest model was the standard framework proposed by MacKenzie et al. (2002) that assumes false positives are not possible ($p_{10} = 0$) and that detection probabilities are constant across observers, or "ψ; $p_{11}(\cdot)$; $p_{10}(0)$". The second model again assumes that false positives were not possible, but allows observers to differ in their probability of detecting HWA: "ψ; $p_{11}(t)$; $p_{10}(0)$". The final two models both incorporate the possibility of misclassification ($p_{10} > 0$, Royle and Link 2006), with the simpler of the two assuming that observers do not differ in their probability of detecting or misclassifying HWA: "ψ; $p_{11}(\cdot)$; $p_{10}(\cdot)$". The more complex of these two models assumes that observers can differ in their probability of detecting and misclassifying HWA: "ψ; $p_{11}(t)$; $p_{10}(t)$". Maximum-likelihood estimates of the model parameters can be obtained by maximizing numerically

$$L(p_{11}, p_{10}, \psi \mid y) \propto \prod_{i=1}^{n} \left\{ \left[p_{11}^{y_i} (1 - p_{11})^{T - y_i} \right] \psi + \left[p_{10}^{y_i} (1 - p_{10})^{T - y_i} \right] (1 - \psi) \right\},$$

where n is the number of sites (trees), T is the number of samples (observers), and $\boldsymbol{y} = \{ y_{i=1}^{n} \}$ with y_i representing the site-specific number of detections. See Royle and Link (2006) for details. We used the small sample size form of Akaike's Information Criterion (AIC_c) to determine the model best supported by the data (Burnham and Anderson 2002). Statistical analyses were performed in R 2.7.2 (R Development Core Team 2006) using code modified from Royle and Link (2006) and in Microsoft Excel using Excel spreadsheets developed by Donovan and Hines (2007). Sample data, R code, and Excel spreadsheets are provided in the Supplement to this paper.

Results

The two volunteer-only groups detected HWA infestations on a smaller proportion of trees than did the two volunteer/experienced groups. One of the volunteer-only groups detected HWA on 14 of 86 sampled trees (naïve infestation rate = 0.163), and the other on 33 of 95 trees (naïve infestation rate = 0.347). In contrast, the two volunteer/experienced groups detected HWA on 57 of 125 trees (naïve infestation rate = 0.456) and on 69 of 114 trees (naïve infestation rate = 0.605). Of the two volunteer/experienced groups, the group with the fewest volunteers realized the highest overall naïve infestation rate (0.605). When two experienced observers returned to the 173 trees to estimate the abundance of detected HWA infestations, HWA was found on 164 trees. Experienced individuals detected smaller HWA infestations than volunteers (paired t-test, $p = 0.017$).

The form of the best-supported model differed between volunteer-only groups and volunteer/experienced groups. For volunteer-only groups, model comparison by ΔAICc and normalized Akaike model selection weights (Burnham and Anderson 2002) revealed that models where the probability of misidentifying HWA was zero ($p_{10} = 0$) were best supported by the data (Table 1). However, the best-supported model for volunteer-only groups differed in their assumptions regarding whether observers differed in their probability of detecting HWA infestations. The best-supported model for one of the volunteer-only groups assumed that observers differed in their detection probabilities, $\psi p_{11}(\mathrm{t})p_{10}(0)$, while the data for the other volunteer-only group most strongly supported the model $\psi p_{11}(\cdot)p_{10}(0)$, which did not make this assumption. In contrast, the form of the best-supported model was the same for both volunteer/experienced groups (Table 2). For such groups, strongest support was for model $\psi p_{11}(\mathrm{t})p_{10}(\mathrm{t})$, where misclassification probabilities were greater than zero and both detection and misclassification probabilities differed between observers. There was little support for models where experienced and volunteer observers were assumed to have equal probabilities of detecting HWA infestations.

When compared to volunteers in their group, experienced observers had a higher probability of detecting HWA infestations (Table 2). Unexpectedly, this was also true of the probability of misclassifying other organisms as HWA, with experienced observers having a higher probability of misclassifying HWA infestations than volunteers. This finding is an artifact of the models, the origin of which we discuss below. When comparing across groups, estimates of detection probabilities from the best-supported models ranged from 0.28-0.94, with the highest value obtained by an experienced observer and the lowest by a volunteer (Tables 1, 2). Detection probabilities for experienced observers were always greater than 0.75 and had a smaller range than those of volunteers (0.19 versus 0.44).

Estimates of the proportion of trees infested from the best-supported models ranged from 0.12-0.41. For volunteer-only groups, the estimated infestation rate was higher than the naïve infestation rate (Table 1). In contrast, the estimated infestation rate was considerably lower than the naïve infestation rate for groups containing an experienced observer (Table 2).

Discussion

The reliability of data collected from field surveys is directly related to sampling variation and bias in the methods used to gather the data and interobserver variation is one such source of bias. Our findings suggest that observer experience can be an important source of sampling variation and bias in the detection of low-density populations. However, when such surveys are used in an occupancy modeling framework that allows for misidentification, interobserver bias can be manifested in an unexpected manner.

We found that experienced observers differed from volunteers in their ability to detect low-density infestations. Relative to volunteers, experienced observers (1) detected infestations at a greater proportion of trees, (2) had a higher probability of detecting infestations, and (3) detected smaller infestations. Although we were not surprised by these findings, we were surprised by the apparent result that experienced observers were *more* likely to misclassify HWA than volunteers. Although the possibility that experienced individuals are more likely to misidentify HWA cannot be discounted, Costa and Onken (2007) note that once detected, HWA are nearly unmistakable to a well-trained individual. An alternative explanation is suggested by a closer inspection of the detection histories (Table 3). For the team with one experienced observer and two volunteers, the two volunteer observers detected HWA on only 1 of 125 trees when the experienced observer did not. In contrast, the experienced individual detected HWA 23 times when the two volunteers did not. However, when the infested trees were resurveyed by two experienced observers to estimate the abundance of HWA, this additional survey detected infestations on 19 of these 23 trees. The detection histories for the group with two experienced individuals reveal a similar pattern.

Taken together, our results (A) suggest a failure by volunteers to detect low-density infestations rather than misidentification by experienced observers and (B) reveal an issue regarding the absence of statistical weighting in the model. In essence, the misclassification model assumes that there are two types of sites and the probability of detection is lower at one type of site than the other. The differences in detection probabilities between these two sites can arise either through misclassification (Royle and Link 2006) or through heterogeneity in detection. In this study, heterogeneity in detection associated with variation in abundance of HWA and differences in the ability of observers to detect low-density populations, rather than misclassification, is the factor most likely to be driving differences in detection between sites. In other words, the two types of sites in our study are those with relatively dense infestations that were detected by both volunteers and experienced observers and those with relatively low density infestations that were detected only by experienced individuals. However, as formulated, our models give equal weight to the quality of any individual's observations. Therefore, when a low-density infestation is detected by one experienced observer, but missed by the remaining two volunteers, statistical support tips in favor of misclassification. This issue became apparent only when surveys completed by experienced observers were paired with those made by volunteers. Thus our findings caution against the use of observers of differing levels of experience in the same survey and suggest the need to include in models that allow for false positive errors survey-specific covariates that account for biases in detection probabilities introduced by differences in observers (e.g., Bailey et al. 2004).

Our findings also speak to how strongly misidentifications can bias estimates of the proportion of sites occupied (Royle and Link 2006). In the most extreme case, the modeled proportion of infested trees was nearly 4 times lower (0.12 versus 0.58, naïve infestation =

0.456), when misclassification probabilities were assumed to be greater than zero versus when they were assumed to be zero. Again, the modeled rate of 0.12 when misclassification probabilities were assumed to be greater than zero appears to be primarily a function of the model's spurious interpretation of valid detections made by experienced observers as instances of misclassification.

What do our results say about the adequacy of data on the distribution of low-density populations collected by volunteers? We suggest that the answer to this question depends on the ultimate use of the data and on the system under study. For example, recent studies have demonstrated that volunteers can provide accurate data on the presence of invading species (Boudreau and Yan 2004; Delaney et al. 2008). These studies, both involving aquatic invasive species, dealt with either a relatively large and easy-to-detect organism (Delaney et al. 2008) or used volunteers to collect samples that were later verified by professionals (Boudreau and Yan 2004). In contrast, HWA, though easy to identify to the trained eye, can be extremely difficult to detect when occurring at low densities (Evans and Gregoire 2007); our results suggest field experience can improve the ability to detect such infestations. Thus, we argue our findings speak more to issues regarding the importance of properly training volunteers and to the challenges of monitoring low-density or difficult-to-detect organisms (e.g., Milberg et al. 2008), rather than to the reliability of volunteer-based monitoring programs per se. For example, Lotz and Allen (2007) found that there was no difference in error rates between professional scientists and volunteers who had received the same training and who had little difference in actual field anuran-call-survey experience (see also Shirose et al. 1997; Genet and Sargent 2003). Further, multiple studies have demonstrated that observer bias generally decreases as observers become more experienced (Sauer et al. 1994; McLaren and Cadman 1999; Delaney et al. 2008). Taken together, our results underscore the importance of adequate training for volunteers taking part in monitoring programs and the need to document and account for interobserver variation in analytical estimates of site occupancy rates (Lotz and Allen 2007; Pierce and Gutzwiller 2007). Future work in this area should consider the role of survey-specific covariates that account for interobserver variation in detection probabilities.

ACKNOWLEDGEMENTS

We thank the volunteers and J. Backer, D. Cox, L. Ingwell, A. Paradis, A. Roehrig, and J. Turner for field assistance. Advice and comments from Andy Royle and Adam Porter resulted in substantial improvements to this manuscript. This research was funded by NSF DEB-0715504 to EP.

LITERATURE CITED

Bailey, LL, et al. (2004) "Estimating site occupancy and species detection probability parameters for terrestrial salamanders." Ecological Applications **14**(3): 692-702.

Boudreau, SA and ND Yan (2004) "Auditing the accuracy of a volunteer-based surveillance program for an aquatic invader bythotrephes." Environmental Monitoring and Assessment **91**(1): 17-26.

Burnham, KP and DR Anderson (2002) Model selection and multimodel inference: A practical information-theoretic approach, Springer.

Cohn, JP (2008) "Citizen science: Can volunteers do real research?" BioScience **58**(3): 192-197.

Costa, S and B Onken (2007) Standardizing sampling for detection and monitoring of hemlock woolly adelgid in eastern hemlock forests. Forest Health Technology Enterprise Team. Morgantown, WV, U.S. Department of Agriculture, Forest Service**:** 11.

Delaney, DG, et al. (2008) "Marine invasive species: Validation of citizen science and implications for national monitoring networks." Biological Invasions **10**(1): 117-128.

Donovan, T and J Hines (2007) Exercises in occupancy modeling and estimation.

Ellison, AM, et al. (2005) "Loss of foundation species: Consequences for the structure and dynamics of forested ecosystems." Frontiers in Ecology and the Environment **3**(9): 479-486.

Evans, AM and TG Gregoire (2007) "The tree crown distribution of hemlock woolly adelgid, adelges tsugae (hem., adelgidae) from randomized branch sampling." Journal of Applied Entomology **131**(1): 26-33.

Genet, KS and LG Sargent (2003) "Evaluation of methods and data quality from a volunteer-based amphibian call survey." Wildlife Society Bulletin **31**(3): 703-714.

Liebhold, A and P Tobin (2008) "Population ecology of insect invasions and their management." Annual Review of Entomology **53**: 387-408.

Lodge, DM, et al. (2006) "Biological invasions: Recommendations for us policy and management." Ecological Applications **16**(6): 2035-2054.

Lotz, A and CR Allen (2007) "Observer bias in anuran call surveys." Journal of Wildlife Management **71**(2): 675-679.

MacKenzie, DI, et al. (2002) "Estimating site occupancy rates when detection probabilities are less than one." Ecology **83**(8): 2248-2255.

MacKenzie, DI, et al. (2006) Occupancy estimation and modeling: Inferring patterns and dynamics of species occurrence, Academic Press.

McClure, M (1990) "Role of wind, birds, deer, and humans in the dispersal of hemlock woolly adelgid (homoptera: Adelgidae)." Environmental Entomology **19**(1): 36-43.

McClure, MS and C Cheah (1999) "Reshaping the ecology of invading populations of hemlock woolly adelgid, adelges tsugae (homoptera: Adelgidae), in eastern north america." Biological Invasions **1**(2): 247-254.

McLaren, MA and MD Cadman (1999) "Can novice volunteers provide credible data for bird surveys requiring song identification?" Journal of Field Ornithology **70**(4): 481-490.

Milberg, P, et al. (2008) "Observer bias and random variation in vegetation monitoring data." Journal of Vegetation Science **19**(5).

Pierce, BA and KJ Gutzwiller (2007) "Interobserver variation in frog call surveys." Journal of Herpetology **41**(3): 424-429.

R-Development-Core-Team (2006) R: A language and environment for statistical computing, reference index version 2.4.1. Vienna, Austria. , R Foundation for Statistical Computing.

Royle, JA and WA Link (2006) "Generalized site occupancy models allowing for false positive and false negative errors." Ecology **87**(4): 835-841.

Sauer, JR, et al. (1994) "Observer differences in the north america breeding bird survey." The Auk **111**(1): 50-62.

Sharov, AA, et al. (2002) "" Slow the spread": A national program to contain the gypsy moth." Journal of Forestry **100**(5): 30-35.

Shirose, LJ, et al. (1997) "Validation tests of an amphibian call count survey technique in ontario, canada." HERPETOLOGICA **53**: 312-320.

Table 1 – Comparison of models and parameter estimates for detection of HWA for groups comprised entirely of volunteers. K, number of parameters in model; ΔAICc, small sample size form of Akaike's information criterion (AICc) for each model, minus the AICc of the model with minimum AICc; *w*, normalized model selection weights; p_{11}, the probability of detecting the species given that the species is actually present at the site; p_{10}, the probability of falsely detecting the species at an unoccupied site. For model notation, symbols within parentheses indicate whether probabilities are assumed to be constant (·) or different (t) across surveys.

	ΔAICc	*w*	K	ψ-hat	$p_{11,1}$	$p_{11,2}$	$p_{11,3}$	$p_{10,1}$	$p_{10,2}$	$p_{10,3}$
n = 95, $\psi_{,\text{naïve}}$ = 0.347										
ψ; p_{11}(t); p_{10}=0	0	0.85	4	0.41	0.28	0.43	0.61	0	0	0
ψ; p_{11}(·); p_{10}=0	4.47	0.09	2	0.43	0.43	0.43	0.43	0	0	0
ψ; p_{11}(t); p_{10}(t)	6.57	0.03	7	0.39	0.27	0.46	0.65	0.02	0.00	0.00
ψ; p_{11}(·); p_{10}(·)	6.61	0.03	3	0.43	0.43	0.43	0.43	0.00	0.00	0.00
n = 86, $\psi_{,\text{naïve}}$ = 0.163										
ψ; p_{11}(·); p_{10}=0	0	0.62	2	0.17	0.72	0.72	0.72	0	0	0
ψ; p_{11}(·); p_{10}(·)	1.68	0.27	3	0.15	0.78	0.78	0.78	0.01	0.01	0.01
ψ; p_{11}(t); p_{10}=0	3.44	0.11	4	0.17	0.63	0.77	0.77	0	0	0
ψ; p_{11}(t); p_{10}(t)	9.34	0.01	7	0.14	0.67	0.86	0.86	0.01	0.01	0.01

Table 2 – Comparison of models and parameter estimates for detection of HWA for groups comprised of both volunteers (V) and experienced observers (E, parameter estimates for experienced individuals are italicized). K, number of parameters in model; ΔAICc, small sample size form of Akaike's information criterion (AICc) for each model, minus the AICc of the model with minimum AICc; *w*, normalized model selection weights; p_{11}, the probability of detecting the species given that the species is actually present at the site; p_{10}, the probability of falsely detecting the species at an unoccupied site. For model notation, symbols within parentheses indicate whether probabilities are assumed to be constant (·) or different (t) across surveys.

	ΔAICc	*w*	K	ψ-hat	$p_{11,1}$	$p_{11,2}$	$p_{11,3}$	$p_{10,1}$	$p_{10,2}$	$p_{10,3}$
n = 114, $\psi_{,\text{naïve}}$ = 0.605					E	E	V	E	E	V
ψ; p_{11}(t); p_{10}(t)	0	0.65	7	0.26	*0.78*	*0.75*	0.34	*0.07*	*0.44*	0.01
ψ; p_{11}(t); p_{10}=0	1.2	0.35	4	0.72	*0.36*	*0.72*	0.13	*0*	*0*	0
ψ; p_{11}(·); p_{10}=0	56.71	0.00	2	0.84	*0.35*	*0.35*	0.35	*0*	*0*	0
ψ; p_{11}(·); p_{10}(·)	56.82	0.00	3	0.10	*0.75*	*0.75*	0.75	*0.24*	*0.24*	0.24
n = 125, $\psi_{,\text{naïve}}$ = 0.456					E	V	V	E	V	V
ψ; p_{11}(t); p_{10}(t)	0	0.92	7	0.12	*0.94*	0.72	0.79	*0.25*	0.15	0.04
ψ; p_{11}(t); p_{10}=0	5.06	0.08	4	0.58	*0.57*	0.37	0.22	*0*	0	0
ψ; p_{11}(·); p_{10}(·)	12.75	0.00	3	0.1	*0.84*	0.84	0.84	0.15	0.15	0.15
ψ; p_{11}(·); p_{10}=0	19.18	0.00	2	0.61	*0.37*	0.37	0.37	*0*	0	0

Table 3 – Detection histories of HWA populations by group. Histories indicate whether HWA was determined to be present (1) or absent (0) for each of the three surveys. For groups with experienced observers surveys are ordered such that reading from left to right moves from experienced (E) to volunteer (V) observers (e.g., 100 for the group with one experienced observer and two volunteers indicates an instance when the experienced observer detected HWA but the two volunteers did not).

Detection history	EEV	EVV	VVV	VVV
111	6	8	2	6
110	14	6	2	1
011	2	1	9	3
101	2	4	4	1
100	7	23	3	1
010	37	12	4	1
001	0	3	9	1
000	45	68	62	72