University of Massachusetts Amherst
ScholarWorks@UMass Amherst

Environmental Conservation Faculty Publication Series

Environmental Conservation

2017

Frequency of invasive plant occurrence is not a suitable proxy for abundance in the Northeast United States

Tyler Cross
University of Massachusetts Amherst

John T. Finn
University of Massachusetts - Amherst

Bethany A. Bradley
University of Massachusetts - Amherst

Follow this and additional works at: https://scholarworks.umass.edu/nrc_faculty_pubs

Part of the Environmental Monitoring Commons, and the Natural Resources and Conservation Commons

Frequency of invasive plant occurrence is not a suitable proxy for abundance in the Northeast United States

TYLER CROSS, JOHN T. FINN, AND BETHANY A. BRADLEY†

Department of Environmental Conservation, University of Massachusetts, Amherst, Massachusetts 01003 USA

Abstract. Measuring and predicting invasive plant abundance is critical for understanding impacts on ecosystems and economies. Although spatial abundance datasets remain rare, occurrence datasets are increasingly available across broad regional scales. We asked whether the frequency of these point occurrences can be used as a proxy for abundance of invasive plants. We compiled both occurrence and abundance data for 13 regionally important invasive plants in the northeast United States from herbarium records and several contributed distribution datasets. We integrated all available abundance information based on infested area, stem count, percent cover, or qualitative descriptions into abundance rankings ranging from 0 (absent) to 4 (highly abundant). Within equal-area grid cells of 800 m, we counted numbers of occurrence points and used ordinal regression to test whether higher densities of occurrence points increased the odds of a higher abundance ranking. We compiled a total of 86,854 occurrence points in 34,596 grid cells, of which 26,114 points (30%) within 11,976 cells (35%) had some form of abundance information. Eleven of the 13 species had a slight but significantly positive odds ratio; that is, more occurrence points of a species increased the odds that the species was abundant within the grid cell. However, the predictive ability of the models was poor ($\kappa < 0.2$) for the majority of species. Additionally, most grid cells contained only one or two occurrence points, making it impossible to infer abundance in all but a few locations. These results suggest that currently available occurrence datasets do not effectively represent abundance, which could explain why many distribution models based on occurrence data are poor predictors of abundance. Increased efforts to consistently collect and report invasive species abundance, ideally estimating both infested area and average cover, are strongly needed for regional-scale assessments of potential abundance and associated impact.

† **E-mail:** bbradley@eco.umass.edu

INTRODUCTION

Invasive species abundance is recognized as an important metric of potential impact on ecosystems (Parker et al. 1999, Daehler 2003, Stohlgren and Schnase 2006, Seabloom et al. 2013). Unfortunately, spatial data available for invasive species, such as museum/herbarium records and management records, are typically limited to occurrences only. Including all occurrences rather than just abundant infestations in species distribution models leads to vast overestimation of invasion risk (Bradley 2013), which is of limited use for guiding control efforts aimed at reducing ecological and economic impacts (Hulme 2006, McDonald et al. 2009). Hence, spatial data and associated spatial models of invasive species abundance at landscape and regional scales are strongly needed for

understanding potential impact and for coordinating monitoring and management.

Occurrence data alone are not typically effective for predicting abundance. Several studies have tested whether habitat suitability models based on presence-only or presence/absence data can effectively predict abundance, with generally poor results (Pearce and Ferrier 2001, Sakai et al. 2001, Jiménez-Valverde et al. 2009, VanDerWal et al. 2009). For invasive plants, presence-only models were effective for differentiating presence from absence, but could not predict increasing abundance (Pearce and Ferrier 2001), particularly when herbarium records were the source of data (Bradley 2016). However, spatial models trained with abundance data perform reasonably well for predicting invasive species abundance (Kulhanek et al. 2011, Bradley 2016). Thus, in order to effectively predict invasion risk associated with invasive species abundance, better spatial abundance data are needed, particularly at landscape to regional scales.

One approach for estimating abundance in the absence of explicit abundance data uses the frequency of individual occurrence as a proxy for local abundance through occupancy modeling. It is a well-accepted pattern in ecology that a species' abundance is positively correlated with the frequency of its occurrence across a region (He and Gaston 2000, Holt et al. 2002). Collection of occurrence data for occupancy modeling requires repeated observations to measure the frequency of species occurrence using consistent levels of search effort (Royle and Dorazio 2008). The resulting occurrence frequency and absence data are then used to model abundance. While contributed occurrence datasets do not meet these sampling criteria, it is possible that the increasingly widespread and repeated collections by research, monitoring, and management groups could provide sufficient spatial occurrence information to act as a proxy for local abundance.

In the United States, invasive plant occurrence data are available through herbarium records like those contained in the Global Biodiversity Information Facility (GBIF; www.gbif.org) as well as spatial data compilations like the Invasive Plant Atlas of New England (IPANE; Mehrhoff et al. 2003), iMAP Invasive Species (imapinvasives.org), or the Early Detection and Distribution MAPping System (EDDMAPS; Bargeron and Moorhead

2007). The latter databases contain data compiled from a range of sources, including both citizen scientists and conservation professionals. In some cases, invasive plant abundance data, either qualitative or quantitative, are included along with occurrence locations.

Botanical records like GBIF have long been accepted as an important source of occurrence data for use in species distribution or habitat modeling. For invasion ecology in particular, management and citizen science databases are also increasingly being used to model habitat suitability (Dickinson et al. 2010, 2012). While there is some concern that contributed datasets could contain information recorded by under-trained individuals (Crall et al. 2015), recent research has increasingly shown that data from citizen scientists are reliable (Danielsen et al. 2005, Fowler et al. 2013). Data are even more reliable when contributors are trained and/or when data are professionally verified, which is typical of invasive species datasets. Thus, citizen science and management records provide a robust dataset that increases numbers of occurrence records and broadens regional coverage (Fore et al. 2001, Delaney et al. 2008). However, to date, the majority of botanical and management records indicate occurrence alone. Information about abundance is often lacking and inconsistently reported.

Given the importance of abundance information for modeling invasion risk across landscapes and regions (Parker et al. 1999, Daehler 2003, Stohlgren and Schnase 2006, Seabloom et al. 2013), we aimed to test whether the spatial frequency of point occurrences can be used as an effective proxy for invasive plant abundance. Here, we compiled a comprehensive database of occurrence and abundance data for 13 problematic invasive plant species across the northeast United States. We hypothesized that the number of occurrences within equal-area grid cells will be positively related to invasive plant abundance. This analysis provides an empirical test of the relationship between the frequency of invasive plant occurrences and local plant abundance.

MATERIALS AND METHODS

Study species and area

We selected 13 non-native, invasive plants that are of concern to regional managers and have

negative environmental impacts in the northeast United States (Table 1). These species are some of the most common and widespread invaders within the study area, which included 13 states and the District of Columbia between Virginia and Maine. As the most commonly reported species, they provide a robust test of the relationship between frequency of occurrence and abundance.

Compilation of existing data sources

We compiled existing distribution and abundance data from four online databases that record geolocations of invasive species: the GBIF (www.gbif.org), the IPANE (Mehrhoff et al. 2003), the EDDMAPS (Bargeron 2016), and iMapInvasives (imapinvasive.org) for the states of Maine, New York, Pennsylvania, Vermont, and West Virginia (although Virginia also contributes to iMAP, they did not have records for our target species). Additionally, we collected and included invasive species occurrence and abundance information for selected species from two smaller databases compiled by researchers and managers in New England.

Ultimately, the included data ranged from botanical records collected by professional scientists, to citizen science efforts in which interested individuals collect and enter occurrences of invasive species into online repositories. All data included geographic location, with a subset also containing abundance information reported in a variety of formats from qualitative to quantitative. For databases containing polygons rather than points, we assumed that the polygon extents described the area of the invasive plant infestation. We removed duplicate points as well as points that fell outside of the study area and points where the spatial precision was coarser than one-thousandth of a decimal degree (equivalent to about 100 m within the study area). We also tested for points located at town or country centroids, which would likely have poor locational accuracy, but found none. We combined all available data in our analyses, but also calculated reporting differences between the datasets.

For all geographic locations also containing abundance information, we standardized abundance to a qualitative, ranked scale of 0–4 (Table 2) ranging from absent to highly abundant. Both qualitative and quantitative bins were arbitrary, but based upon commonly reported values within the management datasets for cover, stem count, and descriptions of invasion. These values

Table 1. List of target invasive plants.

Name	Growth habit	Impact example	Reference
Alliaria petiolata (Garlic Mustard)	Forb/herb	Disrupts mutualisms, reduces native diversity	Stinson et al. (2006)
Berberis thunbergii (Japanese barberry)	Shrub	Increases tick and Lyme disease prevalence	Williams et al. (2009)
Celastrus orbiculatus (Oriental Bittersweet)	Vine	Suppresses native plant growth	Fike and Niering (1999)
Cynanchum louiseae (Black Swallow-wort)	Vine, Forb/herb	Disrupts native communities, reduces native diversity	DiTommaso et al. (2005)
Elaeagnus umbellata (Autumn olive)	Shrub	Outcompetes native vegetation, increases nitrogen	Munger (2003)
Euonymus alatus (Burning bush)	Shrub	Outcompetes native vegetation	Mehrhoff et al. (2003)
Frangula alnus (Glossy Buckthorn)	Shrub, tree	Alters nutrient cycling	Stokdyk and Herrman (2014)
Lonicera morrowii (Morrow's honeysuckle)	Shrub	Shrub honeysuckles reduce growth of forest herbs	Miller and Gorchov (2004)
Lythrum salicaria (Purple Loosestrife)	Subshrub, forb/herb	Forms monocultures	Yakimowski et al. (2005)
Microstegium vimineum (Japanese Stilt-Grass)	Graminoid	Reduces native diversity	Morrison et al. (2007)
Polygonum cuspidatum (Japanese Knotweed)	Forb/herb, subshrub	Forms monocultures, alters nutrient cycling	Aguilera et al. (2010)
Rhamnus cathartica (Common buckthorn)	Shrub, tree	Alters nutrient cycling, eliminates leaf litter	Knight et al. (2007)
Rosa multiflora (Multiflora rose)	Subshrub, vine	Increases avian nest predation and nest failure	Borgmann and Rodewald (2004)

Table 2. Classification scheme used to combine quantitative and qualitative abundance estimates into abundance rankings.

Abundance ranking	Quantitative cover estimate (%)	Quantitative stem count	Quantitative range extents	Qualitative cover/stem estimate	Qualitative range extents
0	0	0	0	Absent	Absent/not present
1	≤1	1	≤1 m^2	Uncommon, trace, or single plant	Rare
2	1–5	2–99	1 m^2–1 acre	Few or scattered plants	Few or small patches
3	5–25	100–999	1–40 acres	Common or scattered dense plants	Moderate or several patches
4	>25	>999	40 + acres	Abundant or dense monoculture	Many small or several large patches

are consistent with previous rankings of relative invader abundance and importance (Rouget and Richardson 2003). Our decision to use >25% cover as the threshold for the highest bin was based on discussions with local invasive plant managers, who agreed that in the northeast >25% cover would be considered "abundant." Bins for quantitative range extent estimates were also based on commonly reported metrics of area (square meters for small areas, acres for larger areas). The break between ranks 3–4 (moderate vs. high abundance) of 40 acres for range extent was chosen to match the break in cover estimate of 25% of a grid cell (see below); a grid cell was approximately 160 acres.

Archived data

The dataset used in this analysis (not including iMAP invasives data, which are not public but can be requested from individual states) can be downloaded from: https://doi.org/10.7275/r55t3hfj.

Point count vs. abundance comparison

In order to calculate frequency of occurrences, we created an equal-area grid of 800 × 800 m cells encompassing the study area. This spatial resolution approximates 30 arc seconds, which is a typical gridded resolution for species distribution modeling at regional to continental scales. Polygon layers were transformed into point occurrences with one point per grid cell. We then summed the number of point occurrences within each grid cell. All spatial analyses were performed using ArcGIS 10.2 (ESRI, Redlands, California, USA).

In order to calculate abundance for each grid cell, we extracted the maximum abundance ranking associated with all points falling within each grid cell. We chose maximum abundance based on the assumption that it was a better indicator of potential invasion risk in a given cell than mean abundance. Additionally, mean abundance was problematic because ranks were based on unequally spaced bins (nonetheless, maximum and mean were highly correlated, $R = 0.85$). For polygon data, the area of the polygon overlapping each grid cell was calculated and grid cells were ranked according to the area category in Table 2.

In order to test whether frequency of point occurrences was related to abundance, we compared ranked abundance estimates at the 800-m grid cell resolution to the number of points falling within each grid cell. Grid cells only contained abundance information for a given species if one or more occurrences within the cell had associated abundance, or if polygon features identifying the extent of an invasion overlapped the grid cell. As a result, only grid cells containing both abundance and point occurrence information were included in the model.

Statistical analysis

We used ordinal regression analysis to test the hypothesis that frequency of occurrence points was positively related to abundance. Ordinal regression was appropriate in this case because the abundance classification bins were not equally spaced but increased with rank order. Ordinal regression tests for an overall relationship between occurrence frequency and ranked abundance based on proportional odds ratios. If the overall odds are not significantly different from 1 (i.e., 1:1), then there is no relationship. A significant odds ratio that does not overlap 1 can be interpreted as the odds that an increase of one occurrence will correspond to an increase of one

abundance rank. For example, overall odds of 2 gives 2:1 odds that one additional occurrence corresponds to a higher abundance rank. Ordinal regression was performed using the proportional odds logistic regression (polr) function in the MASS package in R (version 2.15.2, R Core Team 2014). We used the ordinal regression models to plot occurrence frequency vs. the probability of being in a given abundance rank (1–4) for each species. For visualization, we also created box-and-whisker plots for each study species showing abundance ranking vs occurrence frequency.

The occurrence frequency data were skewed toward single values; that is, most grid cells contained only a single occurrence point. To test whether the quantity of these single values biased the results, we repeated the analysis for the subset of grid cells with occurrence point frequency greater than one.

Finally, we randomly split each species dataset into 2/3 training and 1/3 testing and used the proportional odds logistic regression to predict abundance rank based on the occurrence frequencies in the testing dataset. We compared observed vs. predicted abundance ranks using 4 × 4 contingency tables and calculated Cohen's kappa (Cohen 1960) for each species. We interpreted kappa statistics using the benchmarks presented by Landis and Koch (1977) where 0 = poor, 0.2 = slight, 0.4 = fair, 0.6 = moderate, 0.8 = substantial, and 1.0 = perfect agreement between predicted and observed abundance ranks.

RESULTS

Occurrence frequency vs. abundance

Numbers of occurrences ranged from the smallest, *Cynanchum louiseae*, with 1088 points (26% with abundance information) spread across 526 cells to the largest, *Rosa multiflora*, with 12,355 points (35% with abundance information) spread across 5360 cells. Distribution and abundance data for these two example species are shown in Fig. 1. Maps of the remaining species are presented in Appendix S1. After removing duplicate and non-useable points, a total of 86,854 occurrence points remained within 34,596 grid cells across all 13 species (Table 3). Thirty percent of these points had associated abundance information based on quantitative or qualitative estimates of area or cover (or both). In most
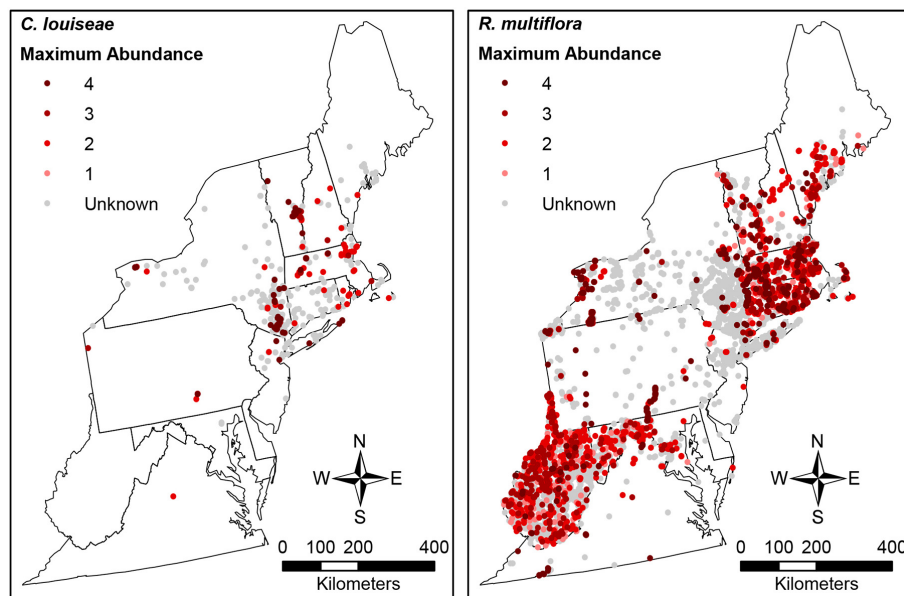

Fig. 1. Abundance maps for two example species *Cynanchum louiseae* and *Rosa multiflora* show that occurrence and abundance data are widespread across the northeast but data from mid-Atlantic states and Virginia were less well reported in the databases included in this analysis. Maximum abundance corresponds to rankings presented in Table 2.

Table 3. Total numbers of occurrence and abundance points compiled for each target species.

Species	Total points	Points w/abundance	% Points w/abundance	Total cells	Cells w/abundance	% Cells w/abundance
Alliaria petiolata	7974	2182	27.4	3739	1037	27.7
Berberis thunbergii	10,915	3042	27.9	3497	1316	37.6
Celastrus orbiculatus	8325	2748	33.0	2708	1276	47.1
Cynanchum louiseae	1088	279	25.6	526	148	28.1
Elaeagnus umbellata	5842	1727	29.6	2406	921	38.3
Euonymus alatus	3278	1301	39.7	1279	566	44.3
Frangula alnus	4108	1526	37.1	1141	599	52.5
Lonicera morrowii	2524	1046	41.4	1403	601	42.8
Lythrum salicaria	7031	1251	17.8	3534	701	19.8
Microstegium vimineum	8988	1599	17.8	2350	592	25.2
Polygonum cuspidatum	11,305	3063	27.1	5262	1574	29.9
Rhamnus cathartica	3121	1188	38.1	1391	486	34.9
Rosa multiflora	12,355	4295	34.8	5360	2159	40.3
Total	86,854	26,114	30.1	34,596	11,976	34.6

species studied, grid cells with and without abundance information contained similar frequencies of occurrence points. Across all species, the mean frequency of occurrences in grid cells with abundance information was 2.50 ± 0.04 (SE) points per grid cell vs. 2.65 ± 0.03 (SE) points per grid cell in grid cells without abundance information.

There was very little difference in frequency of point occurrences across the abundance rankings for any of the 13 species, with the possible exception of rank 4, which tended to have higher occurrence frequency (Fig. 2). However, even for rank 4, the median occurrence frequency values were typically either one or two points (note the log scale in Fig. 2). Indeed, of the 11,976 grid cells with some abundance information, the vast majority (9459, or 79%) contained only one or two occurrence points. This same pattern of low numbers of occurrences across all abundance ranks was still evident when we excluded all cells with a single occurrence point (Appendix S2).

Ordinal regression revealed overall odds significantly greater than one for 11 of the 13 study species (Table 4). That is, having one more occurrence point within a grid cell significantly increased the probability that the grid cell would have a higher abundance rank. However, the power of this predictive relationship was weak, with all but one of the significant odds ratios ranging from 1.13:1 to 1.31:1. In only one case, *Euonymus alatus*, could more occurrences be reasonably interpreted as higher abundance. When

we repeated these analyses including only grid cells with more than one occurrence point, we found a similar overall pattern (Appendix S2).

Probability plots based on the ordinal regression showed that there was little differentiation between abundance ranks 1–3 (Fig. 3). For the example species *Alliaria petiolata*, ordinal regression predicts that abundance rank 2 is most likely at low occurrence frequency, while abundance rank 4 is most likely once occurrence frequency becomes greater than ~5 occurrences per grid cell (Fig. 3). However, of the 229 *A. petiolata* grid cells with abundance rank 4, only 50 (22%) would have been predicted as high abundance based on this criterion. Probability plots for all 13 species are shown in Appendix S3.

Contingency tables of predicted vs. observed abundance rank showed that model predictions were generally poor (Appendix S4). Cohen's kappa statistics were below 0.1 (poor agreement) for all species except *C. louiseae* ($\kappa = 0.24$, slight agreement) and *E. alatus* ($\kappa = 0.16$, slight agreement). As expected from the probability plots (Appendix S3), predicted ranks tended to fall entirely into rank 2 and rank 4, creating a false-negative prediction for rank 1 and rank 3. An example contingency table for *A. petiolata* ($\kappa = 0.04$, poor agreement) is shown in Table 5.

Abundance and occurrence across datasets

The plurality of our occurrence points came from the five states contributing iMAP data, which accounted for 44,943 data points. The
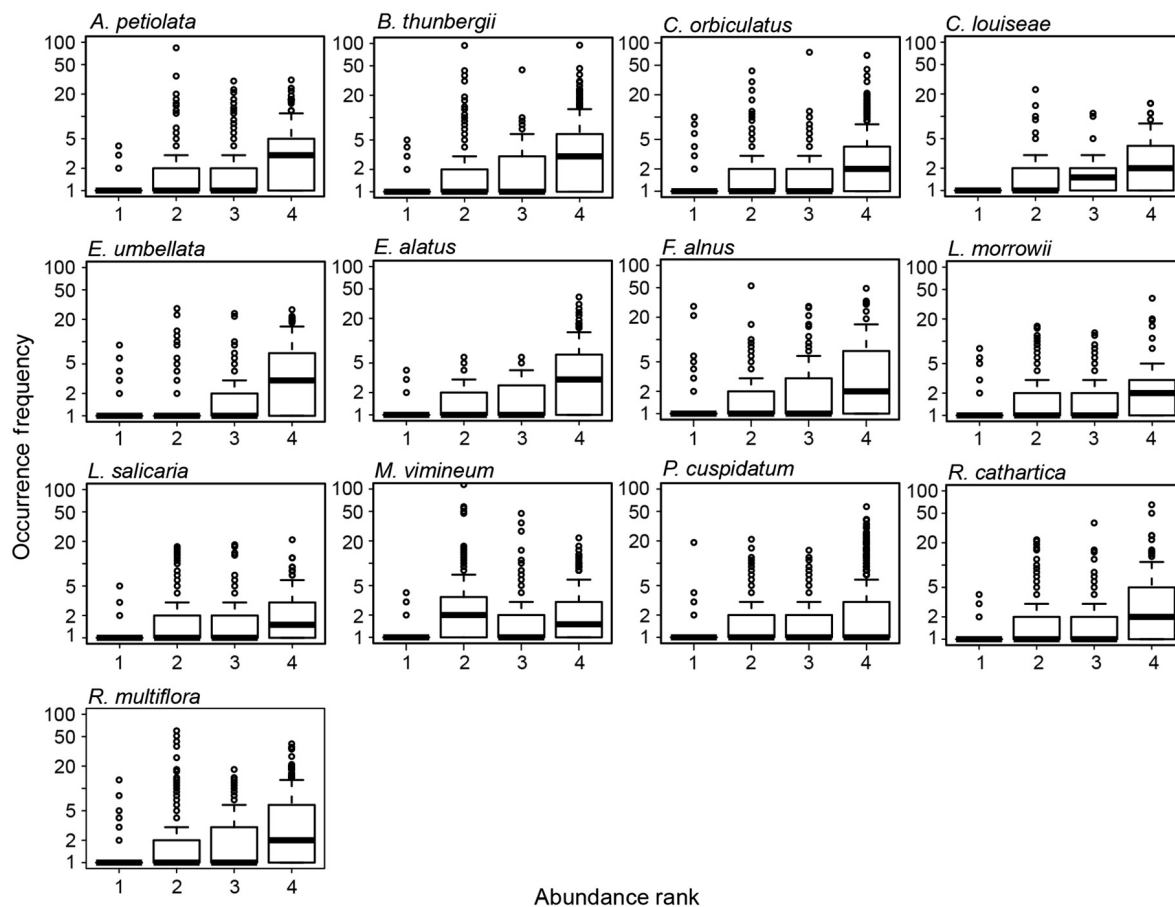
Fig. 2. Box plot of abundance ranking vs. the frequency of occurrences within grid cells. Most species show increasing numbers of occurrences as abundance increases. However, the median number of occurrences per grid cell typically only ranges from 1 to 2 between the lowest and highest abundance ranks.

second largest dataset was the EDDMAPS, with 26,821 data points (Table 6). Only EDDMAPS and GBIF targeted the full study region; the other datasets were regionally focused on single states or smaller portions of the northeast.

Among the databases, there was a range of availability of abundance information, from none in the GBIF herbarium records to 100% from IPANE, where contributors used the same methods to conduct quantitative surveys of vegetation (Mehrhoff et al. 2003). Records from iMAP and EDDMAPS, the two largest datasets, contained 11% and 42% abundance information, respectively (Table 6). This abundance information came in a variety of forms. Most common was percent cover; however, it was rare to find both percent cover and extent in the same record.

Thus, it is likely that some records of 100% cover, which suggests high abundance, could be describing a small, localized patch. Similarly, infested area was often reported, but typically without any cover information. The most directly interpretable quantitative metric was stem count, because thousands of individuals (particularly for woody species) can more readily be interpreted as "highly abundant" than either percent cover or area of infestation. Stem count was most often reported in the bins shown in Table 2.

Qualitative descriptions of cover and extent were also useful, although it was time-consuming to convert various unique comments into abundance ranks. It would be easier to compile qualitative descriptions if the data were reported in a consistent format such as the ones

Table 4. Overall odds are the odds that an increase in 1 occurrence per grid cell will correspond to an increase in the abundance ranking based on ordinal regression analysis.

Species	Overall odds	CI (95%)	Model fit (kappa)
Alliaria petiolata	1.16	1.11–1.21	0.04 (poor)
Berberis thunbergii	1.23	1.17–1.28	0.05 (poor)
Celastrus orbiculatus	1.19	1.14–1.24	0.00 (poor)
Cynanchum louiseae	1.08	0.98–1.21	0.24 (slight)
Elaeagnus umbellata	1.31	1.23–1.40	0.06 (poor)
Euonymus alatus	2.36	2.00–2.82	0.16 (slight)
Frangula alnus	1.14	1.09–1.19	0.03 (poor)
Lonicera morrowii	1.13	1.06–1.20	0.03 (poor)
Lythrum salicaria	1.15	1.09–1.22	0.01 (poor)
Microstegium vimineum	0.99	0.97–1.01	0.01 (poor)
Polygonum cuspidatum	1.22	1.16–1.29	0.00 (poor)
Rhamnus cathartica	1.14	1.08–1.20	0.07 (poor)
Rosa multiflora	1.25	1.21–1.30	0.08 (poor)

Note: Odds were significantly higher than 1:1 in 11 of 13 cases; however, although significant, the modeled relationship (based on Cohen's kappa) for all but *C. louiseae* and *E. alatus* was poor.

described in Table 2. Of lowest value (in fact, uninterpretable) were qualitative descriptions that did not clearly articulate whether they referred to individuals, cover, or infested area (e.g., "huge").
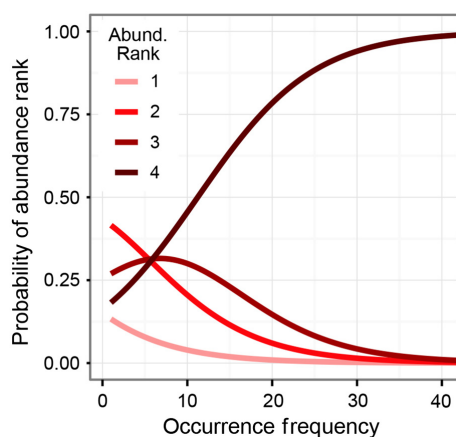

Fig. 3. Probability plot from ordinal regression model for *Alliaria petiolata* shows low differentiation between abundance ranks. At low occurrence frequency (<5 records per grid cell), the model predicts that abundance rank 2 is most likely. Above five records per grid cell, abundance rank 4 is most likely. Abundance ranks 1 and 3 are never the most likely.

Table 5. Contingency table for the example species *Alliaria petiolata* showing the number of observations in the testing dataset falling within each predicted rank.

Predicted	Observed Rank 1	Rank 2	Rank 3	Rank 4	Total
Rank 1	0	0	0	0	0
Rank 2	49	123	84	62	318
Rank 3	0	1	1	11	13
Rank 4	0	3	4	8	15
Total	49	127	89	81	346

Discussion

Ordinal regression analysis revealed weak but significant odds that an additional occurrence would correspond to a higher abundance rank in 11 of 13 species (Table 4). While this finding supports the hypothesis that the frequency of occurrence points in a grid cell relates to abundance, the model predictions overall were poor (Table 5; Appendix S4). Moreover, most grid cells contained only one or two occurrence points (Fig. 2), meaning that occurrence frequency as a metric of abundance could rarely be applied. Unfortunately, occurrence frequency is currently unlikely to provide a reliable means of distinguishing rare from abundant invasions (Fig. 2).

With sufficient, regular sampling, it has often been found that numbers of species occurrences are positively related to species abundance (He et al. 2002). Species with larger populations are more frequently observed, and thus, higher rates of observation indicate that more individuals are occupying the habitat (Royle and Dorazio 2008). However, effective occupancy modeling requires consistent, widespread sampling and resampling to measure occurrences (He et al. 2002). Although the invasive plant database we compiled contains substantial, widespread occurrences recorded by citizen scientists, invasive plant managers, and museum collectors, the occurrence data do not appear to be consistent enough to infer abundance. Regardless of the magnitude or extent of infestation, invasions are likely to be recorded only as a single point location. The majority of 800-m grid cells (63%) contained one occurrence point.

Several studies have modeled habitat suitability based on occurrence points and compared these suitability values to measured abundance (Pearce

Table 6. Total numbers of occurrence and abundance points compiled for each of the primary data sources.

Databases	Total points	Points w/abundance	% of points w/abundance	Total cells	Cells w/abundance	% of cells w/abundance
EDDMAPS	26,821	11,159	41.6	5692	2564	45
GBIF	5108	0	0	2685	0	0
iMapInvasives	44,943	4856	10.8	8733	1275	14.6
IPANE	6203	6203	100	1765	1765	100
WISP	3779	2959	78.3	668	256	38.3
All databases	86,854	25,177	29	19,543	5860	30

Notes: EDDMAPS, Early Detection and Distribution MAPping System; GBIF, Global Biodiversity Information Facility; IPANE, Invasive Plant Atlas of New England; WISP, Westfield Invasive Species Partnership. Cell sums are not the same as in Table 3 because here we are summing across databases rather than across species.

and Ferrier 2001, Sakai et al. 2001, Jiménez-Valverde et al. 2009, VanDerWal et al. 2009, Bradley 2016). In most cases, the relationship has been weak and models are only effectively able to differentiate between presence and absence. Our results, showing that the distribution of occurrences is poorly correlated with local abundance, suggest an underlying cause for the lack of modeled relationship. If frequency of occurrences only weakly reflects abundance, then it is not surprising that models based on these occurrences do a poor job of predicting abundance.

Previous research comparing abundance and occurrence frequency of invasive species has shown no relationship or even a negative relationship, with more occurrences in areas of low abundance (Marvin et al. 2009). Marvin et al. (2009) hypothesized that point data collected by invasive plant managers tended to focus on early detection and rapid response (EDRR), which targets small nascent infestations (Moody and Mack 1988). As a result, EDRR data collection efforts might, counterintuitively, tend to have more occurrences in areas of low abundance. Our results do not show evidence of this negative relationship at low abundance ranks; but there is no differentiation between abundance ranks 1–3 for most species (Fig. 3; Appendix S3). It is possible that the wide range of data sources we used includes the effect shown in Marvin et al. (2009) as well as clustered occurrences in areas of high abundance. For example, EDDMAPS has a strong focus on early detection, while IPANE conducted more systematic sampling. Additionally, our rankings merge together multiple quantitative and qualitative metrics of abundance, which are not necessarily directly comparable. The use of bins, while necessary to create sufficient data for sampling, adds

uncertainty to the analysis, and a more consistent ranking could lead to better predictive relationships. Asking scientists and managers to record and report more occurrence points in areas with abundant infestations would help to increase the strength of the relationship observed with existing data.

Data coverage and availability varied widely between states (Fig. 1; Appendix S1). States and agencies that archive to EDDMAPS make their data easily accessible for research and management applications. These types of free and open source datasets are increasingly important for supporting broad-scale invasion risk assessments. Herbarium records from GBIF are also easily accessible and used for a wide variety of applications in regional spatial modeling. For invasive plants, EDDMAPS spatial coverage was a marked improvement over GBIF. Within the study region, EDDMAPS contained nearly 15 times as many data points (Table 6). Additionally, 40% of EDDMAPS data included some form of qualitative or quantitative abundance estimate.

Spatial models predicting invasive species abundance have been shown to be more accurate when created using abundance data rather than simple occurrence information (Kulhanek et al. 2011, Bradley 2016). Habitat suitability models are useful tools for regional- and landscape-scale invasive species management; thus, there is a clear need for data collectors, professionals, and citizen scientists alike, to continue to collect and report abundance information. The contributed databases compiled here report a considerable amount of abundance information (Table 6). As yet, these abundance data are underutilized in modeling efforts. Landscape ecologists and biogeographers should consider including this

important source of data in predictive spatial risk assessments.

While existing abundance information is an excellent start, data collectors could consider modifications of collection methods to better inform spatial models. For example, abundance data were most often reported as a single metric, either by quantitatively or qualitatively describing cover, or by quantitatively or qualitatively describing extent. The combination of cover and extent information is much more informative for estimating the magnitude of an invasion, and we recommend that collectors report both pieces of information. Although quantitative measurements are preferable, we found that qualitative estimates using consistent rankings (e.g., Table 2) were very useful for interpreting relative abundance. We urge scientists, managers, and citizens collecting invasive plant occurrence data to include extent and cover information when archiving their data to online repositories.

Conclusions

Overall, we found significant positive relationships between occurrence frequency and abundance in 11 of 13 target invasive plants. However, the model predictions had generally poor agreement, making it unlikely that frequency of occurrence could be used as an effective proxy for abundance in risk assessments and management planning. Given the importance of abundance for understanding invasion risk, additional recording and reporting of abundance is needed, particularly data that include estimates of both cover and extent. A remarkably large proportion (30%) of the compiled occurrences already contained some abundance information, which could be a boon to regional-scale invasion risk assessments.

Acknowledgments

Literature Cited

Aguilera, A. G., P. Alpert, J. S. Dukes, and R. Harrington. 2010. Impacts of the invasive plant *Fallopia japonica* (Houtt.) on plant communities and ecosystem processes. Biological Invasions 12:1243–1252.

Bargeron, C. T., and D. J. Moorhead. 2007. EDDMapS —early detection and distribution mapping system for the southeast exotic pest plant council. Wildland weeds 10:4–8.

Borgmann, K. L., and A. D. Rodewald. 2004. Nest predation in an urbanizing landscape: the role of exotic shrubs. Ecological Applications 14:1757–1765.

Bradley, B. A. 2013. Distribution models of invasive plants over-estimate potential impact. Biological Invasions 15:1417–1429.

Bradley, B. A. 2016. Predicting abundance with presence-only models. Landscape Ecology 31:19–30.

Cohen, J. 1960. A coefficient of agreement for nominal scales. Educational and Psychological Measurement 20:37–46.

Crall, A. W., C. S. Jarnevich, N. E. Young, B. J. Panke, M. Renz, and T. J. Stohlgren. 2015. Citizen science contributes to our knowledge of invasive plant species distributions. Biological Invasions 17:2415–2427.

Daehler, C. 2003. Performance comparisons of co-occurring native and alien invasive plants: implications for conservation and restoration. Annual Review of Ecology, Evolution, and Systematics 34:183–211.

Danielsen, F., N. D. Burgess, and A. Balmford. 2005. Monitoring matters: examining the potential of locally-based approaches. Biodiversity and Conservation 14:2507–2542.

Delaney, D. G., C. D. Sperling, C. S. Adams, and B. Leung. 2008. Marine invasive species: validation of citizen science and implications for national monitoring networks. Biological Invasions 10:117–128.

Dickinson, J. L., J. Shirk, D. Bonter, R. Bonney, R. L. Crain, J. Martin, T. Phillips, and K. Purcell. 2012. The current state of citizen science as a tool for ecological research and public engagement. Frontiers in Ecology and the Environment 10:291–297.

Dickinson, J. L., B. Zuckerberg, and D. N. Bonter. 2010. Citizen science as an ecological research tool:

challenges and benefits. Annual Review of Ecology, Evolution, and Systematics 41:149–172.

DiTommaso, A., F. M. Lawlor, and S. J. Darbyshire. 2005. The biology of invasive alien plants in Canada. 2. *Cynanchum rossicum* (Kleopow) Borhidi [= *Vincetoxicum rossicum* (Kleopow) Barbar.] and *Cynanchum louiseae* (L.) Kartesz & Gandhi [= *Vincetoxicum nigrum* (L.) Moench]. Canadian Journal of Plant Science 85:243–263.

Fike, J., and W. A. Niering. 1999. Four decades of old field vegetation development and the role of *Celastrus orbiculatus* in the northeastern United States. Journal of Vegetation Science 10:483–492.

Fore, L. S., K. Paulsen, and K. O'Laughlin. 2001. Assessing the performance of volunteers in monitoring streams. Freshwater Biology 46:109–123.

Fowler, A., J. D. Whyatt, G. Davies, and R. Ellis. 2013. How reliable are citizen-derived scientific data? Assessing the quality of contrail observations made by the general public. Transactions in GIS 17: 488–506.

He, F., and K. J. Gaston. 2000. Estimating species abundance from occurrence. American Naturalist 156:553–559.

He, F., K. J. Gaston, and J. Wu. 2002. On species occupancy-abundance models. Ecoscience 9: 119–126.

Holt, A. R., K. J. Gaston, and F. He. 2002. Occupancy-abundance relationships and spatial distribution: a review. Basic and Applied Ecology 3:1–13.

Hulme, P. E. 2006. Beyond control: wider implications for the management of biological invasions. Journal of Applied Ecology 43:835–847.

Jiménez-Valverde, A., F. Diniz, E. B. de Azevedo, and P. A. V. Borges. 2009. Species distribution models do not account for abundance: the case of arthropods on Terceira Island. Annales Zoologici Fennici 46:451–464.

Knight, K. S., J. S. Kurylo, A. G. Endress, J. R. Stewart, and P. B. Reich. 2007. Ecology and ecosystem impacts of common buckthorn (*Rhamnus cathartica*): a review. Biological Invasions 9:925–937.

Kulhanek, S. A., B. Leung, and A. Ricciardi. 2011. Using ecological niche models to predict the abundance and impact of invasive species: application to the common carp. Ecological Applications 21: 203–213.

Landis, J. R., and G. G. Koch. 1977. The measurement of observer agreement for categorical data. Biometrics 159–174.

Marvin, D. C., B. A. Bradley, and D. S. Wilcove. 2009. A novel, web-based, ecosystem mapping tool using expert opinion. Natural Areas Journal 29:281–292.

McDonald, A., A. Riha, A. DiTommaso, and A. DeGaetano. 2009. Climate change and the geography of weed damage: Analysis of U.S. maize systems suggests the potential for significant range transformations. Agriculture, Ecosystems & Environment 130:131–140.

Mehrhoff, L. J., J. A. Silander Jr., S. A. Leicht, E. S. Mosher, and N. M. Tabak. 2003. IPANE: invasive plant atlas of New England. University of Connecticut, Storrs, Connecticut, USA.

Miller, K. E., and D. L. Gorchov. 2004. The invasive shrub, *Lonicera maackii*, reduces growth and fecundity of perennial forest herbs. Oecologia 139:359–375.

Moody, M. E., and R. N. Mack. 1988. Controlling the spread of plant invasions: the importance of nascent foci. Journal of Applied Ecology 25:1009–1021.

Morrison, J. A., H. A. Lubchansky, K. E. Mauck, K.-M. McCartney, and B. Dunn. 2007. Ecological comparison of two co-invasive species in eastern deciduous forests: *Alliaria petiolata* and *Microstegium vimineum*. Journal of the Torrey Botanical Society 134:1–17.

Munger, G. T. 2003. *Elaeagnus umbellata*. *in* Fire effects information system. U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fire Sciences Laboratory. http://www.fs.fed. us/database/feis/

Parker, I. M., et al. 1999. Impact: toward a framework for understanding the ecological effects of invaders. Biological Invasions 1:3–19.

Pearce, J., and S. Ferrier. 2001. The practical value of modelling relative abundance of species for regional conservation planning: a case study. Biological Conservation 98:33–43.

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

Rouget, M., and D. M. Richardson. 2003. Inferring process from pattern in plant invasions: a semimechanistic model incorporating propagule pressure and environmental factors. American Naturalist 162: 713–724.

Royle, A., and R. Dorazio. 2008. Hierarchical modeling and inference in ecology: the analysis of data from populations, metapopulations, and communities. Academic Press, Cambridge, Massachusetts, USA.

Sakai, A., F. Allendorf, J. Holt, D. Lodge, and J. Molofsky. 2001. The population biology of invasive species. Annual Review of Ecology and Systematics 32: 305–332.

Seabloom, E. W., et al. 2013. Predicting invasion in grassland ecosystems: Is exotic dominance the real embarrassment of richness? Global Change Biology 19:3677–3687.

Stinson, K. A., S. A. Campbell, J. R. Powell, B. E. Wolfe, R. M. Callaway, G. C. Thelen, S. G. Hallett, D. Prati,

and J. N. Klironomos. 2006. Invasive plant suppresses the growth of native tree seedlings by disrupting belowground mutualisms. PLoS Biology 4:e140.

Stohlgren, T., and J. Schnase. 2006. Risk analysis for biological hazards: what we need to know about invasive species. Risk Analysis 26:163–173.

Stokdyk, J. P., and K. S. Herrman. 2014. Short-term impacts of *Frangula alnus* litter on forest soil properties. Water, Air, and Soil Pollution 225: 1–9.

VanDerWal, J., L. P. Shoo, C. N. Johnson, and S. E. Williams. 2009. Abundance and the environmental niche: Environmental suitability estimated from niche models predicts the upper limit of local abundance. American Naturalist 174:282–291.

Williams, S. C., J. S. Ward, T. E. Worthley, and K. C. Stafford. 2009. Managing Japanese barberry (Ranunculales: Berberidaceae) infestations reduces blacklegged tick (Acari: Ixodidae) abundance and infection prevalence with *Borrelia burgdorferi* (Spirochaetales: Spirochaetaceae). Environmental Entomology 38:977–984.

Yakimowski, S. B., H. A. Hager, and C. G. Eckert. 2005. Limits and effects of invasion by the nonindigenous wetland plant *Lythrum salicaria* (purple loosestrife): a seed bank analysis. Biological Invasions 7:687–698.

Supporting Information