**Running head:**

**Title of paper:** Analysis of information from multiple herbarium calls into question existence of lag-phase in spread of *Frangula alnus* in North America

**Author:** Matthew E. Aiello-Lammens

**Affiliation:** Pace University, Department of Environmental Studies and Science

**Address:** 861 Bedford Road, Pleasantville, NY 10570

**Email of corresponding author:** matt.lammens@gmail.com; 631-327-2371

**Abstract**

Analysis of herbaria records allows for an examination of patterns of spatial spread of non-native plants in novel ranges, aiding in understanding the processes that govern non-native species invasions. Of particular importance has been the use of herbarium records to estimate extended lag-phase durations for non-native invasive species. I used herbaria records to investigate the rate of spread and pattern of establishment for the invasive plant *Frangula alnus* in northeastern and middle North America. I collected records spanning a temporal range from ca. 1880 - Present and a spatial range covering the entire invaded area in northeast North America. I addressed unequal sampling effort by comparing temporal and spatial patterns of *F. alnus* accessions to patterns in a group of ecologically similar native species. I found evidence for the potential of multiple initial introductions into North America, that were greatly separated geographically, ranging from southern Ontario to coastal New York and New Jersey. Trends in record collection in time and space show the rate of spread of *F. alnus* was initially slow, then increased rapidly during the early 20th century, and reached a relatively constant rate of spread in the later 20th century. Examining the spread of this species at the continental scale, there is little evidence that it experienced an extended lag phase between phases of establishment and rapid spatial spread, in contradiction to previous reports. Rather, it appears that *F. alnus* has steadily increased in area of occupancy since ca. 1920 to the present. These results raise the question of whether extended lag-phases may be a spatial-scale specific pattern. The analysis methods presented here provide one way to investigate this question further.

**Keywords:** species invasion; *Frangula alnus*; lag phase; herbarium; spatial spread

**Introduction (Currently ~1400 words)**

Trade and travel have contributed to the introduction of numerous species to places where they have no evolutionary or ecological history. Many introduced species perish, some establish self-sustaining populations, and a few grow to such high numbers that they negatively impact native species. The population dynamics of a non-native species transitioning from one established in a new region to one rapidly growing in abundance and occupied area are rarely directly observed. We must piece together this history by examining emerging patterns taken from snapshots in time. Thus, a full understanding of these dynamics remains a gap in our knowledge regarding species invasions. For most non-native invasive species, from the time of establishment to rapid spread, relatively little population growth in number and area, is observed. This is commonly referred to as the lag phase (Kowarik 1995, Crooks and Soulé 1999, Crooks 2005, Theoharides and Dukes 2007). Because the lag phase is early in the invasion process when population sizes are small, management actions taken during this time have the potential to be most effective at mitigating the negative impacts of a species invasion (Pluess et al. 2012).

Identifying a lag phase is not trivial. A newly established population in an environment suitable for growth and reproduction should increase exponentially in number as *N(t) = N(0) Rt*, where *t =* time (e.g., years), *N(t)* = the population size at time *t*, and *R* is the population growth rate. This assumes neither positive density dependence effects (i.e. Allee effects) or negative density dependent effects (i.e. intra-specific competition). A plot of population size through time yields an exponential growth curve. Crooks and Soulé (1999) describe “the shallow portion early in the growth curve when the population is growing relatively slowly in absolute number” as the inherent lag. However, because there is no mathematical definition of the inherent lag *per se*, it is difficult to apply this concept to identify a lag phase. In fact, a plot of the *log* of population size versus time shows a linear relationship, indicating no distinct transition point between the inherent lag phase and a population explosion phase.

In contract to an inherent lag, we can readily identify an extended lag phase. During an extended lag population growth is slower than its potential maximum. A plot of *log* of population size versus time is non-linear when *R* is accelerating. The acceleration can be smooth, resulting in a continuous curve, or abrupt, leading to a discontinuity in the curve. In either case, the period when the population is experiencing lower growth rates than maximum is the extended lag (Crooks and Soulé 1999).

Many invasive species show evidence of extended lags, but factors causing lags are not well understood; several ecological and evolutionary processes may be involved (Mack et al. 2000, Sakai et al. 2001, Pyšek and Hulme 2005, Gurevitch et al. 2011). The simplest explanation is that lags results from time constraints intrinsic to population growth and establishment, such as generation time and time to first reproduction (i.e., the time required for a population to achieve a stable age distribution). However, most lags are longer than can be explained by these processes. Our lack of understanding is a consequence of relatively little empirical work examining lags (Aikio et al. 2010a), though there has been much theoretical work (Crooks 2005). Historical biodiversity data from herbaria and museums can address this lack of empirical support, providing data to re-construct patterns of spatial spread of invasive species.

Retrospective spatial analyses can yield insights into the ecological processes involved in the spread of non-native invasive species. The specimen holdings of herbaria provide a rich data source for spatial analyses of plant species. In fact, analysis of the information from herbarium records have been used to estimate species rates of spread through time and space (e.g., (Delisle et al. 2003, Salo 2005, Miller et al. 2009, Crawford and Hoagland 2009, Newbold 2010, Aikio et al. 2010a, 2010b, Lavoie 2012). A major challenge in analyzing these data is that herbarium records may have been collected with unequal sampling effort in time and/ or space, resulting in biases. For example, widely observed periods of high and low herbaria specimen collection can influence interpretation of temporal trends (Prather et al. 2004). Spatial biases may result from herbarium specific emphases on regional specimen collection. Other biases may emerge because of issues of convenience. For example, in many holdings a larger number of specimens were collected near museums, herbaria, botanical gardens, and academic centers, as well as urban areas (Hijmans et al. 2000, Reddy and Dávalos 2003, Kadmon et al. 2004). These potential biases make it difficult to determine if observed trends in herbarium collections (e.g., increased number of records through time) are associated with changes in the population size or range size of the species of interest, or rather, representative of trends in specimen collection. It is important to distinguish these patterns to use these data for understanding ecological processes of species invasions.

Another challenge is that the data from herbaria are generally more conducive to studying spatial spread, rather than population increase. These processes are related, but we cannot assume that the rate of spatial spread is equivalent to the rate of population growth. This is problematic because whereas exponential growth is widely accepted as a null model for local population growth, there are many plausible null models for areal growth (i.e., spatial spread). A reaction-diffusion model is commonly applied, where spatial growth is a function of the radius of the population. The area of occupancy for a species spreading randomly on a landscape from a central point increases geometrically, with an exponent of 2. Thus, the square root of the area of occupancy through time for a spreading species should have a linear relationship with respect to time, analogous to linearity after log transform for exponential growth. While there are several null models, the square root transformation is appropriate in most situations (Crooks and Soulé 1999).

A small number of methods have been developed to address the problems associated with unequal sampling effort. Most compare trends in the distribution of records for a species of interest to other species with similar habitat requirements – i.e. associated species (Miller et al. 2009). For example, Catling and Porebski (1994) studying the spread of *Frangula alnus* in Ontario, compared patterns of collection records of *F. alnus* and native *Rhamnus alnifolia*, showing that botanists collected specimens well outside of urban areas during the time frame of the introduction and early spread of *F. alnus*. They thus concluded that *F. alnus* was primarily located near urban areas early in its invasion based on habitat, and not on unequal sampling effort. Their approach was a qualitative, primarily visual, assessment, but recent developments extend this idea to quantitative frameworks (e.g. (Delisle et al. 2003, Aikio et al. 2010a). These methods have been used to reconstruct patterns of range expansion for both non-native and native plants (e.g., Miller et al. 2009; Larkin 2011). At least one of these methods (i.e., Aikio et al. 2010a) was specifically developed and applied to identify the existence, and estimate the duration, of lag phases for invasive plants.

In this study, I reconstructed the spread of the non-native invasive species *Frangula alnus* throughout its novel range in northeastern North American. I compiled a dataset of historical presence observations of *F. alnus* based on herbarium records and the scientific literature. Adopting and modifying recently proposed methods, I calculated measures related to its rate of spatial expansion, while accounting for potential unequal sampling efforts. The approach I use here allowed me to combine data from multiple herbaria, while still accounting for potential biases. I hypothesized that I would find support for previous claims that *F. alnus* experienced an extended lag lasting from the time of its introduction (ca. 1860) to the early 1970s. Access to herbaria records and other historical biodiversity is increasing as more collections are digitized and made available to the public. Having techniques to combine information from desperate sources, such as those I present here, will help scientists use these data to investigate important standing questions regarding species range expansions.

**Materials and Methods (~2000 words)**

Study Species

Mention synonym Rhamnus frangula

Collection of presence records. I collected historical presence records for *F. alnus* throughout northeast North America by 1) searching on-line databases of several herbaria, 2) requesting record information from curators and collection managers of multiple herbaria, 3) searching the Global Biodiversity Information Facility (GBIF), and 4) searching the scientific literature using the keywords “Frangula alnus OR Rhammus frangula” in ISI Web of Science database. Curators or collection managers from Rutgers Chrysler Herbarium, Miami University of Ohio Herbarium, and Carnegie Mellon Herbarium, provided me with record information for all *F. alnus* or *R. frangula* in their collections. I searched the Global Biodiversity Information Facility (GBIF), the Canada Biodiversity Information Facility (CBIF), and the Invasive Plant Atlas of New England (IPANE) for records of *F. alnus,* and excluded records accessed in other searches, since many herbaria contribute data to these clearing houses. Similarly, I excluded any presences from the literature search corresponding with herbarium specimens. I used all presence records with information on the year and location (county level or finer) of observation in subsequent analyses.

Georeferencing presence records. Most records in the compiled dataset lacked latitude and longitude values noting where the specimen was collected. However, many had descriptive information, which I used to georeferenced the collection location. I primarily used GoogleEarth (Google Inc.) to georeference records (*sensu* Garcia-Milagros & Funk 2010), however some cases required additional searches on named locations, such as unique building names or geographic features. The descriptive location information allowed for varying degrees of precision in the final assigned latitude and longitude value. At a minimum, all records of specimens collected in the United States could be assigned to the county in which they were collected. For records with *only* county level information, I assigned the US Census Bureau designated latitude and longitude values. For records of specimens collected in Canada, I relied on the information included with the specimen to assign county equivalent values. For finer spatial resolution, some information could be used to determine the township the specimen was collected in, whereas others allowed for assigning precise locations. I assigned location uncertainty values following guidelines from the Biogeomancer Consortium (BioGeomancer Consortium 2006).

Number of records through time. Trends in the number of records added to an herbarium through time can be used to infer rates of spread for non-native invasive species and to determine if a species experienced an extended lag phase (*sensu* (Pyšek and Prach 1993, Aikio et al. 2010a). I calculated the cumulative number of *F. alnus* presence records per year and plotted the *log* of the cumulative number of records through time against years. I fit linear, quadratic, and cubic regression lines to these data, and used a likelihood ratio test to determine the best-fit model. I calculated an annual rate of growth for the cumulative number of presence records by dividing the cumulative number of records at year *t +* 1 by the cumulative number of records at year *t*. The mean rate of growth was calculated as the geometric mean of the annual growth rates. Additionally, I calculated 10-year moving window geometric mean growth rates, which minimizes the influence of extreme inter-annual fluctuations in growth rates. This yielded more consistent rates of growth for the earliest period of the invasion, when calculations are based on a relatively small number of records.

Area of occupancy through time. To examine range expansion of *F. alnus* throughout its novel range, I overlaid 5 x 5 arc minute grid (each cell approximately 57 km2 to 67 km2 in area) onto a map of all presence locations. The grid was constructed using the Quantum GIS software (QGIS Development Team 2012). I used the R statistical programming environment (R Core Team 2012) with additional functions from the “raster”, “rgdal”, and “dismo” packages (Hijmans 2012, Keitt et al. 2012, Hijmans et al. 2013) respectively) to carry out all other analyses. Each presence record was assigned membership to a grid cell based on its latitude and longitude value. If a grid cell contained at least one presence of *F. alnus*, I considered it occupied. I then calculated the total area occupied per decade and the cumulative area occupied from time of first introduction to the present. For the latter, I assumed once a grid cell was occupied, it remained so. These measures correspond to those used in other studies of invasive species spread (e.g., (Pyšek and Prach 1995, Weber 1998, Delisle et al. 2003). I calculated the rate of growth for the area of occupancy in the same way as I did for the number of records, substituting the cumulative number of records with the cumulative number of occupied grid cells. I then plotted the square root of the cumulative number of grid cells versus time (years) and tested for deviations from linearity.

Occupied counties through time. Many records could only be georeferenced to the county in which they occurred. Similar to the calculations ofarea of occupancy through time, I calculated the cumulative number of counties occupied by *F. alnus* through time. Again, I assumed that once *F. alnus* was found in a county, it was thereafter occupied. I calculated the growth rate for the cumulative number of counties occupied following the same method used for rate of growth of the number of records and area of occupancy through time.

Accounting for unequal sampling effort in time and space. A comparison of the trends in records of a species of interest to those of other species whose range and population size should be in equilibrium with their environment (e.g., native species) can aid in distinguishing patterns resulting from a species expansion versus unequal sampling effort (Delisle et al. 2003). Thus, I examined patterns in herbarium records for *Alnus incana* (Speckled Alder), *A. serrulata* (Smooth Alder), *Rhamnus alnifolia* (Alderleaf Buckthorn), *Salix peiolarisi* (Meadow Willow)*, Hamamelis virginica* (syn. *macrophylla*) (Witch Hazel), and *Fraxinus Americana* (White Ash). These are woody plants likely to be observed in ecological conditions where *F. alnus* is observed (personal observation, (Little et al. 1980, Catling and Porebski 1994, Houlahan and Findlay 2004). I searched GBIF for records of these species that were located within the area of interest and collected records reported in the University of Wisconsin, Ohio State University, University of Minnesota, Morton Arboretum Herbarium, Michigan State University, and Brooklyn Botanic Gardens herbaria databases. If geolocation information was missing, records were georeferenced to the county level. Grouping all of these records, I calculated the metrics for the number of records through time, the area of occupancy through time, and the number of counties occupied through time, as described above.

I compared the trends for *F. alnus* and the group of associated species for both area of occupancy through time and counties occupied through time using a modified method to that presented by Delisle et al. (2003). They compared trends in area of occupancy through time by dividing the cumulative number of occupied 10 x10 km grid cells for each non-native plant by the cumulative number of cells occupied by a group of five native species, yielding a proportion of non-native to native occupied cells for each year of their study period. A major assumption in the interpretation of this analysis is that spatial growth of native plants in the herbaria dataset does not represent the spread for these plants *per se*, but rather represents the increase in spatial coverage of herbaria records in general (i.e., an increase in the number of locations where samples are collected). Thus, if the proportion of occupied cells between non-native and native plants is increasing through time, this represents periods during which the spatial coverage of records for the non-native plant outpaces the background increase in spatial coverage. This can be interpreted as the result of the non-native plant rapidly spreading.

In this study, I was interested in examining spatial spread in a larger region than that studied in (Delisle et al. 2003), as well as in using records from multiple herbaria. Further, several presence records for the group of associated species were located in areas where *F. alnus* has not been observed, most likely due to the larger niche breadth of some of the associated species compared to *F. alnus,* (e.g., *Fraxinus americana*). Including these records in my analysis increased the area of occupancy measures for the group of associated species compared to the possible area of occupancy for *F. alnus*. Similarly, I collected *F. alnus* presence records from institutions I did not collect associated species records from (e.g., Miami University Herbarium and Rutgers Herbarium). The records from these institutions were largely geographically restricted (e.g., primarily located in the states of Ohio and New Jersey for the institutions mentioned) and these locations were not always represented well in the larger group of associated species dataset. Therefore, including these records had the effect of increasing the area of occupancy of *F. alnus* compared to the possible area of occupancy for the group of associated species. To account for both of these issues, I examined the ratio of cumulative area of occupancy of *F. alnus* to the cumulative area of occupancy of the associated group of species, limiting the records used to an area of coarse spatial overlap for both datasets. I defined the spatial overlap by creating a map of 30 x 30 arc minutes (i.e., 0.5° or approximately 50 x 50 km) and overlaying this on a map of all presence locations. As described in Area of occupancy through time, each record was assigned membership to one 30’ grid cell based on its latitude and longitude value. I then constructed restricted *F. alnus* and associated species presence datasets, in which only records that occurred in a 30’ grid cell occupied by at least one record from both datasets during the study period. Using these restricted datasets, I calculated the ratio of the increase in the cumulative area of occupancy of *F. alnus* to the group of associated species. In a separate analysis, I compared the cumulative number of counties occupied through time, while accounting for similar concerns regarding sampling regions in space that are unsuitable to *F. alnus*. I only included records from counties that were occupied at some time by both *F. alnus* and one of the associated plants. In this case, the ratio of the cumulative number of counties occupied at the end of the study period had to equal 1.0. The growth rates for the cumulative number of grid cells occupied and the cumulative number of counties occupied were compared between *F. alnus* and the group of associated species. To compare the growth rates I divided the annual growth rate of *F. alnus* records by the annual growth rate of the entire group of associated species.

Comparison of number of records through time. In addition to comparing trends in the areal increase of *F. alnus* and the associated species, I also compared trends in the cumulative number of records through time. For each year of the study period, I divided the cumulative number of *F. alnus* records by the cumulative number of associated species records. This correction method is similar to that proposed by (Delisle et al. 2003) and has been used in several studies to account for bias due to unequal sampling effort (e.g., Mihulka & Pyšek 2001; Aikio et al. 2010a; Larkin 2011). Additionally, to compare rates at which *F. alnus* and associated species presences were collected, I calculated the ratio of growth rates using both the annual and 10-year moving window average estimates. A ratio value less than 1.0 indicated that the rate at which *F. alnus* presences were being collected was lower than that of the associated species, while a ratio value of greater than 1.0 indicated that it was greater that that of the associated species.

**Results (~1040 words)**

I compiled a dataset of 749 *F. alnus* presence records and 5548 associated species presence records. These records show that *F. alnus* was collected across northeast and middle North America (Figure 2). The associated species collection locations were largely inclusive of where *F. alnus* was collected. The earliest dated presence of *F. alnus* was in 1879 in Hudson County, New Jersey (accessed from CHRB). The earliest dated presence of one of the associated species was for *Hamamelis virginiana* in 1836 in Richland County, Ohio (accessed from CM). Only 36 records of associated species pre-dated the first *F. alnus* record. A total of 14 *F. alnus* specimens were deposited in 4 separate herbaria prior to 1900; 12 were georeferenced to the metropolitan New York City region (9 in the state of New Jersey) and 2 were collected in southern Ontario (as reported in Catling & Porebski 1994). The number of herbarium specimens collected for both *F. alnus* and the group of associated species increased through time (Figure 3). The associated species showed a substantial increase in the number of records collected at the beginning of the 20th century, followed by another increase after 1950. There was a steady increase in the number of records collected per decade from the time of the first recorded presence of *F. alnus* in the United States to the end of the 20th century. For both groups, there was a dramatic decline in the number of records at the start of 21st century, which is a pattern also observed elsewhere (Prather et al. 2004).

The trend of the log cumulative number of records through time indicates a deviation from exponential growth for both *F. alnus* and the group of associated species (Fig. 4A). The best-fit regression models were a cubic polynomial regression for *F. alnus* (R2 = 0.989, df = 124, P << 0.05) and a quadratic polynomial regression for the associated species (R2 = 0.979, df = 174, P << 0.05 respectively). The cumulative number of records for the group of associated species increased steadily from the late 1800s to the present. The cumulative number of records for *F. alnus* increased rapidly from the late 1800s to approximately 1930. Ratios of the annual growth rates and ten year moving average (geometric mean) growth rates show that the rate of *F. alnus* collections was generally slower than that of the associated species prior to approximately 1920, but faster during most of the 20th century (Fig. 4B). The ratio of the cumulative increase in records of *F. alnus* to the associated species corroborates this pattern, indicating that the increase of *F. alnus* records was greater than the associated species from 1920 to the present (Fig. 4C).

The cumulative number of occupied 5 arc min grid cells for *F. alnus* and the associated species, constrained to overlap within 30 arc min grid cells, increased through time for both datasets (Fig. 5). Trends in the square root of cumulative occupied grid cells versus time supports a departure from a simple diffusion model of spatial spread in historical presence records (Fig. 5A). The best-fit regression models were a quadratic polynomial regression for *F. alnus* (R2 = 0.996, df = 129, P << 0.05) and a cubic polynomial regression for the group of associated species (R2 = 0.990, df = 140, P << 0.05), both indicating a departure from linearity. The rate of increase in presence records was low for both *F. alnus* and the associated species until 1890, after which the number of occupied cells rapidly increases for the group of associated species. In contrast, the number of occupied cells continued to increase slowly for *F. alnus* until approximately 1920, as is demonstrated by the calculated ratio between the two growth rates (Fig. 5B). Prior to 1915, with the exception of the earliest years of the *F. alnus* invasion, the rate at which new grid cells were considered occupied was lower than the corresponding rate for the associated species. From 1915 to the present, the rate at which grid cells are considered occupied by *F. alnus* was consistently higher than the corresponding rate of growth for the associated species. The ratio of the cumulative number of occupied grid cells of *F. alnus* to the associated species shows that after approximately 1910 the rate at which *F. alnus* increased in area of presence was greater than that of the group of associated species (Fig. 5C).

Trends in the cumulative number of counties occupied were similar to the results of the analysis of the cumulative number of occupied grid cells. The number of counties where herbarium samples were collected increased rapidly during the late 19th to early 20th century. In contrast, the number of counties where *F. alnus* was found increased slowly during the early part of the 20th century, but rapidly after 1940 (Fig. 6A). Assuming counties were selected for specimen collections randomly, the relationship between the square root of the cumulative number of occupied counties and time (years) should be linear for the associated species. However, this was not the case for either the associated species or *F. alnus* (Fig. 6A). A cubic polynomial regression model was the best-fit model for both the associated species (R2 = 0.967, df = 166, P < 0.001) and *F. alnus* (R2 = 0.997, df = 130, P < 0.001). The difference in the rate of growth of cumulative occupied counties between *F. alnus*­ and the associated species shows a similar pattern to that of the cumulative occupied grid cells (Fig. 6B). The rate at which counties are considered occupied is slower for *F. alnus* than the associated species early in the invasion history (prior to 1900) and faster during most of the 20th century. The ratio of occupied counties of *F. alnus* to those occupied by the associated species shows that early in the invasion history, *F. alnus* was less frequently collected in newly sampled counties, but that for most of the 20th century the number of counties occupied by *F. alnus* has increased more rapidly than the number of counties occupied by associated species (Fig. 6C). Additionally, *F. alnus* specimens were collected on average 48 years after at least one of the associated species had been collected in a county.

**Discussion (~2400 words)**

Early observations and likely region of first introduction. The exact introduction event for *F. alnus* is unknown, but based on its long history of use as a medicinal plant (United States Pharmacopeial Convention 1910) and evidence that it was planted as an ornamental shrub (Sherff 1912) Add reference to Sudworth and Fernow: (Sudworth, G. B. and B. E. Fernow.  1891.  *Trees of Washington, D.C.*, compliments of the Forestry Division. Geo. B. Sudworth, botanist ; B. E. Fernow, chief.  Washington, D.C.: Bell Lithographing Co.  [18] pp.; 2 folding plans; 14 x 22 cm.), it is likely it was purposely planted in many locations in North America. Based on the historic presence records, the location of first introduction was likely the metropolitan New York City area and areas of New Jersey along New York Harbor. Specimens for 12 of 14 records dating from before 1900 were collected in this region. Locations in southern Ontario account for the remaining 2 of 14 records dating before 1900, indicating that introduction into the region was likely early during the spread of *F. alnus*. New York Harbor was a large shipping and port area, raising the possibility that shipping played a role in an accidental introduction. Many non-native plants have been introduced to port areas by the unloading of solid ballast, with seeds mixed in with rocks and other materials used as ballast (Sorrie 2005, Barney 2006). However, this region historical had a high population density, thus there also may have been more people who purposely planted *F. alnus*.

Accounting for unequal sampling effort in historical biodiversity collections. The presence of unequal sampling effort in space or time is common in historical biodiversity collections (Reddy and Dávalos 2003, Graham et al. 2004, Kadmon et al. 2004, Boakes et al. 2010). Calculating the ratio of the cumulative number of records in space and time of non-native to native plants provides a way to account for such biases, making it possible to distinguish periods of relatively slow versus rapid spread (Delisle et al. 2003). The primary assumption in this correction method is that native species are in equilibrium with their environment prior to the collection of any records. That is, these plants have spread to all of the ecological conditions within the study range where they can survive and reproduce, and have a stable range distribution. There are some caveats to this assumption to consider, perhaps the most important being that the ecological conditions in northeast North America have not been stable over the last 130 years. There have been substantial changes in land use, resulting in changes to plant communities (Wright and Fridley 2010), as well as affecting plant invasion dynamics (Mosher et al. 2009). However, given that the species chosen for the group of associated species have similar ecological requirements as *F. alnus*, I expect that any such changes would affect trends in these species and *F. alnus* equally.

Given that the equilibrium assumption for the associated species is appropriate, then the rate of the cumulative number of records, grid cells, or counties occupied for these species is representative of the effort of specimen collectors, rather than the spread of these plants. As such, if the ratio of cumulative records of *F. alnus* to the associated species is increasing in time, this indicates a time when *F. alnus* is increasing more rapidly than background sampling, and thus experiencing positive growth rates. If the ratio is constant, *F. alnus* may be increasing, however indistinguishably from sampling effort. If the ratio is decreasing, *F. alnus* may still be increasing, but more slowly than the rate of sampling effort. Because the *cumulative* number of records was used in all three temporal trends I investigated, the absolute rate of change in samples cannot show a decline. Thus, time periods that have either decreasing or stable ratio values preceding periods of increasing ratio values may be considered lags.

Increases of *Frangula alnus* through time. The ratios between cumulative records through time, occupied grid cells through time, and occupied counties through time of *F. alnus* to the associated species each indicate that *F. alnus* has increased since ca. 1920 (Figs. 4C, 5C, and 6C). For each trend, the ratio of ten-year average growth rates was greater than 1.0 after 1920, indicating that the number of *F. alnus* samples increased more rapidly than those of the associated species (Figures 4B, 5B, and 6B). Prior to 1920, both the ratio values of cumulative number of records and growth rates fluctuated greatly. This is likely the result of the small number of records for both *F. alnus* and the associated species during this time. At this point, the addition of only a few records to either dataset would dramatically change the ratio between them. The accumulation of *F. alnus* presences early in its invasion history demonstrates this effect well. Generally *F. alnus* records increased more slowly than the associated species from 1879 to the 1920s, but there are anomalous years in which its growth rate was much larger than that of the associated species (Figures 4B, 5B, and 6B). For example, when the cumulative number of records for *F. alnus* increased from 4 to 10 from 1893 to 1894, the resulting growth rate was *R* = 2.5. Comparatively, during this time the number of records for the associated species increased from 171 to 191, resulting in a growth rate of *R* = 1.12. Thus, it is difficult to determine if *F. alnus* did in fact start its invasion with a very high growth rate, then immediately slowed, or rather this result is an artifact of calculating growth rates with small sample sizes. Based on the relative consistency of the total number of records collected in each decade, there is more support for the latter interpretation (Figure 3).

Trends of the square root of the cumulative number of grid cells (or counties) occupied by the associated species through time did not fit linear relationships (Figures 5A and 6A). Assuming these species were already present in most suitable environments (i.e., at equilibrium), this suggests that specimen collectors did simply start collecting at some central points, and spread out radially form there. Both trends can be described as initially concave up, followed by a steep linear trend, followed by a concave down curve, followed by another period of linearity. This suggests the rate of collections increased rapidly in the mid 19th century, was high throughout the second half of the 19th century, and then decelerated during the 20th century. The cubic regression fits do not necessarily support the presumption of an early concave up portion of the curve, but do support the interpretation of a rapid rate of increase early in the collection history, followed by a deceleration in collection rate. Regardless, it is clear that specimen collection rapidly spread across the landscape during the 19th century. Combined with the increase in the cumulative number of records during the earlier part of the study period (Figures 3 and 4A), these patterns suggest that this was a period of high sampling effort, coinciding with the beginning of an intense effort to collect specimens for herbaria (Prather et al. 2004).

The earliest presence records of *F. alnus* were collected during the period of high sampling effort near the end of the 19th century. However, it remained uncommon across the study region until at least the 1920s. Assuming that cumulative presence curves for the associated species represent the spatial spread of collection effort, and that a collector would collect *F. alnus* if it were present during a survey, it appears that the rate of spatial spread during the early 19th century of *F. alnus* was slow. This is most clearly exemplified by the trends in cumulative number of counties occupied through time. On average *F. alnus* was first observed 48 years after at least one of the associated species was observed in a county occupied by both by the end of the study period. By 1900, records for at least one of the associated species was collected in 42% (73 of 172) of counties, where as *F. alnus* was collected in 2% (4 of 172) of counties. This strongly suggests that *F. alnus* was not common at this time. However, the ratio of presence records shows a consistent rate of spread throughout the novel range during the following 100 years.

Evidence for an extended lag phase. In contrast to most previous studies, I examined the spread of an invasive species over a spatial area incorporating nearly all of the known naturalized range. Examining the relative rates of increase in occupied grid cells and occupied counties at this spatial scale, there is no compelling evidence for an extended lag phase persisting beyond the early 20th century. All three analyses suggest that *F. alnus* increased at a rate greater than sampling effort from at least 1920 on. Given that the rate of increase in record number and spatial presence for *F. alnus* was less than that of the associated species from 1880 to 1920, it is plausible that this period of time represents an extended lag phase. However, trends calculated for this period are based on a small number of *F. alnus* records, as is evident in the high variability in both the ratio of growth rates and the ratio of records or occupied area.

At least three previous studies examined aspects of the range expansion of *F. alnus* using herbarium records (Howell and Blackwell 1977, Catling and Porebski 1994, Larkin 2011). Each focused on a part of the invaded range. Howell & Blackwell (1977) examined its spread into, and throughout, Ohio. They found evidence that it likely entered Ohio in the 1920’s, being observed first in Lake County (northeastern Ohio). The authors speculate that the spread of *F. alnus* westward throughout Ohio was facilitated by the range-expansion of the non-native European Starling (*Sturnus vulgaris*). Many bird species are noted to eat the fruit of *F. alnus* and defecate undigested seeds, and *S. vulgaris* in particular have been known to eat these fruit (Howell & Blackwell 1977). The authors did not account for potential unequal sampling effort in herbarium records, but rather simply reported when and where *F. alnus* first appeared in Ohio. Catling & Porebski (1994) examined the historical spread of *F. alnus* in southern Ontario, Canada and found that it was first recorded in London, Ontario in 1898 and Ottawa in 1899. Their data suggested that *F. alnus* spread to other urban centers, but primarily remained confined to these areas until the 1970’s. From the 1970’s to the early 1990’s *F. alnus* spread into natural areas outside of urban areas at an increased rate compared to the previous 70 years. The authors suggest that this observation is indicative of a lag phase in the spread of *F. alnus*, though no quantitative analysis was carried out. As briefly mentioned in the Introduction, to address potential biases resulting from unequal sampling efforts, the authors visually compared the spatial distribution of the *F. alnus* records with that of native *Rhamnus alnifolia*, a species with similar habitat requirements. More recently, Larkin (2011) examined the lengths of lag phases for multiple (>200) species of non-native invasive species in the Wisconsin and the southern Lake Michigan region, applying the methods of (Aikio et al. 2010a). The earliest record of *F. alnus* was collected in 1908 in the southern Lake Michigan region, and based on quantitative analysis had a 31-year lag phase. Comparatively, it was found in southern and northern Wisconsin in 1927 and 1941, and had 36- and 15-year lag phases, respectively. The timings of these region-specific introductions suggest that *F. alnus* spread north from the southern Lake Michigan region into more northern parts of Wisconsin. The length of the lag phase calculated for southern Lake Michigan and southern Wisconsin generally agrees with my observations for the whole range. Further, the pattern of *F. alnus* spreading from urban areas into more rural regions corresponds with the results of recent genetic analyses (De Kort et al. 2016).

Caveats and potential sources of bias. I assumed the cumulative number of occupied grid cells, and counties, as representative of the area of extent. This may not be valid if *F. alnus* invades an area, but later goes locally extinct. However, this is an unlikely occurrence. With one notable exception (Cunard and Lee 2008), there are no documented examples of the local extinction of *F. alnus*, either by natural processes or management actions.This supports the notion that once a location is occupied, it remains so. Nevertheless, it is likely that some locations of historical *F. alnus* presences no longer have the species present because of changes in land use (e.g., development of once-woodland plots). This is the case for several of the records observed in the metropolitan New York region, the site of many of the earliest records. Disturbances, or removal of population from the landscape, due to anthropogenic influences may have a substantial effect on the spread of *F. alnus* throughout its novel region, but it can still be successful in very small, isolated, and disturbed plots within a urban or suburban land-use matrix (personal observation; (Del Tredici 2010). Given that the finest spatial resolution used in this study was 5 x 5 arc minutes (approximately 10 x 10 km), a complete local extinction of *F. alnus* is unlikely.

Most herbaria do not have electronic databases of their holdings that are easily accessed by the public (i.e., a web-base search interface) and many do not have a complete electronic database of records within the institution (Lavoie 2012). Therefore, I know there are several herbaria records absent from the compiled dataset. However, given the extent of the data compilation in space and across institutions, I am confident that the patterns and trends reported are generally representative of the spread of *F. alnus*. Future analyses that involve even greater effort in gathering herbarium specimens could increase the accuracy of my results, particularly providing a more accurate area of extent, but the trends presented here are very likely to hold up.

**Conclusions**

These results clearly suggest the presence of a rapid rate of spread of *F. alnus* throughout its invasion history. Calculating the ratio of the cumulative number of grid cells and counties occupied by *F. alnus* to those occupied by the group of associated species, I found that *F. alnus* expanded rapidly throughout its invaded range since the mid to late 1920s to the present. Patterns of spatial spread and estimates of lag phase duration likely vary between regions. However, a quantitative analysis from one region yielded a length for an extended lag phase consistent with what was observed throughout the entire invaded range (Larkin 2011). The patterns and processes of range expansion, particularly of lag phase dynamics, likely vary depending on scale and local ecological conditions (Theoharides and Dukes 2007). Future analyses that compare regional patterns to those for the whole range described here could further elucidate the role of spatial scale in processes of species invasions. Further, using these techniques to investigate the patterns of spread of other non-native species will help us to better understand the dynamics of extended lags.

**Table 1.** Historical presence record sources and counts. Abbreviations for “Accession Method” column: CBIF = Canadian Biodiversity Information Facility, GBIF = Global Biodiversity Information Facility, IPANE = Invasive Plant Atlas of New England, and vPlants = Virtual Herbarium of the Chicago Region.

|  |  |  |  |
| --- | --- | --- | --- |
| **Source** | **Herbarium Code** | **Record Count** | **Accession Method** |
| Harvard University Arnold Arboretum | A | 2 | IPANE |
| Acadia University Herbarium | ACAD | 8 | GBIF |
| Botanischer Garten und Botanisches Museum Berlin-Dahlem | B | 1 | GBIF |
| Brooklyn Botanical Garden | BKL | 35 | Institute website |
| Chicago Botanic Garden | CHIC | 3 | Institute website / vPlants |
| Rutgers University Chrysler Herbarium | CHRB | 10 | Provided by curator |
| Carnegie Museum of Natural History | CM | 68 | Provided by curator |
| University of Connecticut Torrey Herbarium | CONN | 85 | IPANE / GBIF |
| Field Museum of Natural History | F | 28 | Institute website / vPlants |
| Harvard University Herbaria | GH | 5 | IPANE |
| Royal Botanical Gardens | HAM | 7 | CBIF |
| University of Kansas | KANU | 3 | GBIF |
| Forest Products Laboratory | MAD | 1 | Institute website (U. of Wis.) |
| University of Massachusetts | MASS | 22 | IPANE |
| University of Minnesota | MIN | 48 | Institute website |
| Missouri Botanical Garden | MO | 2 | Institute website |
| Morton Arboretum | MOR | 43 | Institute website |
| Michigan State University | MSC | 25 | Institute website |
| Université de Montréal | MT | 22 | GBIF |
| Miami University | MU | 57 | Provided by curator |
| Yale University Connecticut Botanical Society | NCBS | 3 | IPANE |
| New England Botanical Club | NEBC | 63 | IPANE |
| New York Botanical Garden | NY | 5 | Institute website |
| Ohio State University | OS | 81 | Institute website |
| Queen’s University | QK | 2 | Institute website |
| University of Wyoming Rocky Mountain Herbarium | RM | 2 | Institute website |
| Smithsonian Institution | US | 3 | Institute website |
| University of Wisconsin – Green Bay | UWGB | 16 | Institute website (U. of Wis.) |
| University of Wisconsin – Stevens Point | UWSP | 6 | Institute website (U. of Wis.) |
| Naturhistorisches Museum Wien | W | 2 |  |
| University of Wisconsin – Madison | WIS | 85 | Institute website (U. of Wis.) |
| Yale University Peabody Museum of Natural History | YU | 4 | IPANE / GBIF |
| Literature Search |  | 5 | ISI Web of Science |

Figure 1. Population size versus time relationship for an exponentially growing population. Black dots represent constant population growth rate *R* = 1.2. Grey dots represent an increasing growth rate for the first 20 time steps from *R* = 1.0 to *R* = 1.2, then constant *R* = 1.2 from time points 20 to 40. (A) Population size versus time. Inset plot is Population size versus time for *R* = 1.2 for the first 20 time steps. (B) Log (Population size) versus time. For constant *R*, note the non-linear relationship in (A) versus the linear relationship in (B). This relationship is non-linear in both (A) and (B) for a population with an increasing growth rate (grey dots).



Figure 2. Geographic locations of collected records for *F. alnus* and a group of associated species. Red points represent records from the compiled historical presence records for *F. alnus* and green points represent records from the compiled historical presence records for the group of associated species. Some locations were assigned geographic locations based on the latitude and longitude values of counties as defined by the US Census Bureau. Red outlined box delineates the study region.



Figure 3. Total number of records collected in each decade for *F. alnus* (black bars) and the combined group of associated species (grey bars).



Figure 4. (A) Log cumulative number of records through time. Linear (solid) and cubic polynomial (dot-dash) regression predictions are plotted over the cumulative increase curves. (B) Ratio of growth rates of cumulative number of records of *F. alnus* versus associated species calculated annually (black circles) and by10 year moving window average (geometric mean) (red dots). Note that some extreme data points are not shown (those >1.2 or <0.8, but contribute to the moving window average values. (C) Ratio of the cumulative number of records of *F. alnus* to those of associated species.

**B**

**A**



**C**

Figure 5. (A) Square root of the cumulative number of grid cells through time. Shown here are the linear and polynomial regression lines for a models using year as a predictor variable and the square root of the cumulative number of grid cells occupied as the response variable. Linear regression predictions are shown for both *F. alnus* and the group of associated species. The best-fit polynomial regression fit is shown for each set (quadratic for *F. alnus* and cubic for the group of associated species) (B) Ratio of growth rates of cumulative occupied grid cells calculated annually (black points) and by 10 year moving window average (geometric mean) (red points). Outlier data not shown (>1.2 or <0.8), but do contribute to the moving window average values. (C) Ratio of square root of the 5 arc min grid cells occupied by *F. alnus* and associated species at a given time step. Occupied 5 arc min grid cells were constrained to be within 30 arc min grid cells occupied by both *F. alnus* and associated species at least once during the study period.

**B**

**A**

**C**

****

Figure 6. (A) Square root of the cumulative number of counties occupied through time. Shown here are the linear and polynomial regression lines for a models using year as a predictor variable and the square root of the cumulative number of counties occupied as the response variable. Linear regression predictions are shown for both *F. alnus* and the group of associated species (solid red and black lines respectively). The best-fit polynomial regression fit is shown for each set (cubic for both *F. alnus* and the group of associated species) (B) Ratio of the rate of growth for cumulative occupied counties calculated annually (black points) and by 10 year moving window average (geometric mean) (red points). Outlier data not shown (those >1.2 or <0.8, but contribute to the moving window average values. (C) Square root of the ratio of the cumulative number of counties occupied by *F. alnus* to those occupied by the group of associated species.

**C**

**B**

**A**

****

**References**

Aikio, S., R. P. Duncan, and P. E. Hulme. 2010a. Lag-phases in alien plant invasions: separating the facts from the artefacts. Oikos 119:370–378.

Aikio, S., R. P. Duncan, and P. E. Hulme. 2010b. Herbarium records identify the role of long-distance spread in the spatial distribution of alien plants in New Zealand. Journal of Biogeography 37:1740–1751.

Barney, J. N. 2006. North American history of two invasive plant species: phytogeographic distribution, dispersal vectors, and multiple introductions. Biological Invasions 8:703–717.

BioGeomancer Consortium. 2006. Guide to Best Practices for Georeferencing. Page (A. D. Chapman and J. Wieczorek, Eds.). Global Biodiversity Information Facility, Copenhagen.

Boakes, E. H., P. J. K. McGowan, R. A. Fuller, D. Chang-qing, N. E. Clark, K. O’Connor, and G. M. Mace. 2010. Distorted views of biodiversity: spatial and temporal bias in species occurrence data. PLoS Biology 8:e1000385.

Catling, P. M., and Z. S. Porebski. 1994. The history of invasion and current status of glossy buckthorn, *Rhamnus frangula*, in southern Ontario. Canadian field-naturalist 108:305–310.

Crawford, P. H. C., and B. W. Hoagland. 2009. Can herbarium records be used to map alien species invasion and native species expansion over the past 100 years? Journal of Biogeography 36:651–661.

Crooks, J. 2005. Lag times and exotic species: the ecology and management of biological invasions in slow-motion. Ecoscience 12:316–329.

Crooks, J. A., and M. E. Soulé. 1999. Lag times in population explosions of invasive species: causes and implications. Pages 103–125*in* O. T. Sandlund, P. J. Schei, and A. Viken, editors.Invasive species and biodiversity management. Kluwer Academic Dordrecht, The Netherlands.

Cunard, C., and T. D. Lee. 2008. Is patience a virtue? Succession, light, and the death of invasive glossy buckthorn (*Frangula alnus*). Biological Invasions 11:577–586.

Delisle, F., C. Lavoie, M. Jean, and D. Lachance. 2003. Reconstructing the spread of invasive plants: taking into account biases associated with herbarium specimens. Journal of Biogeography 30:1033–1042.

Garcia-Milagros, E., and V. A. Funk. 2010. Improving the use of information from museum specimens: using Google Earth© to georeference Guiana Shield specimens in the US National Herbarium. Frontiers of Biogeography 2:71–77.

Graham, C. H., S. Ferrier, F. Huettman, C. Moritz, and A. T. Peterson. 2004. New developments in museum-based informatics and applications in biodiversity analysis. Trends in Ecology & Evolution 19:497–503.

Gurevitch, J., G. A. Fox, G. M. Wardle, Inderjit, and D. Taub. 2011. Emergent insights from the synthesis of conceptual frameworks for biological invasions. Ecology Letters 14:407–418.

Hijmans, R. J. 2012. Introduction to the’raster’package (version 2.0-08). R.

Hijmans, R. J., K. A. Garrett, Z. Huaman, D. P. Zhang, M. Schreuder, and M. Bonierbale. 2000. Assessing the geographic representativeness of genebank collections: the case of Bolivian wild potatoes. Conservation Biology 14:1755–1765.

Hijmans, R. J., S. J. Phillips, J. R. Leathwick, and J. Elith. 2013. dismo: Species distribution modeling. R.

Houlahan, J. E., and С. S. Findlay. 2004. Effect of invasive plant species on temperate wetland paint diversity. Conservation Biology 18:1132–1138.

Howell, J. A., and W. H. J. Blackwell. 1977. The history of *Rhamnus frangula* (glossy buckthorn) in the Ohio flora. Castanea 42:111–115.

Kadmon, R., O. Farber, and A. Danin. 2004. Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models. Ecological Applications 14:401–413.

Keitt, T. H., R. Bivand, and B. Rowlingson. 2012. rgdal. R.

De Kort, H., J. Mergeay, H. Jacquemyn, and O. Honnay. 2016. Transatlantic invasion routes and adaptive potential in North American populations of the invasive glossy buckthorn, Frangula alnus. Annals of Botany:mcw157.

Kowarik, I. 1995. Time lags in biological invasions with regard to success and failure of alien species. Page *in* P. Pyšek, K. Prach, M. Rejmánek, and M. Wade, editors. Plant invasions: General aspects and special problems. SPB Adademic Publishing, Amsterdam.

Larkin, D. J. 2011. Lengths and correlates of lag phases in upper-Midwest plant invasions. Biological Invasions 14:827–838.

Lavoie, C. 2012. Biological collections in an ever changing world: herbaria as tools for biogeographical and environmental studies. Perspectives in Plant Ecology, Evolution and Systematics 15:68–76.

Little, E. L., S. Bullaty, and A. Lomeo. 1980. National Audubon Society Field Guide to North American Trees: Eastern Region. Alfred A. Knopf.

Mack, R. N., D. Simberloff, W. Mark Lonsdale, H. Evans, M. Clout, and F. A. Bazzaz. 2000. Biotic invasions: causes, epidemiology, global consequences, and control. Ecological Applications 10:689–710.

Mihulka, S., and P. Pyšek. 2001. Invasion history of Oenothera congeners in Europe: a comparative study of spreading rates in the last 200 years. Journal of Biogeography 28:597–609.

Miller, R. J., A. D. Carroll, T. P. Wilson, and J. Shaw. 2009. Spatiotemporal analysis of three common wetland invasive plant species using herbarium specimens and geographic information systems. Castanea 74:133–145.

Mosher, E. S., J. A. Silander Jr, and A. M. Latimer. 2009. The role of land-use history in major invasions by woody plant species in the northeastern North American landscape. Biological Invasions 11:2317–2328.

Newbold, T. 2010. Applications and limitations of museum data for conservation and ecology, with particular attention to species distribution models. Progress in Physical Geography 34:3–22.

Pluess, T., R. Cannon, V. Jarošík, J. Pergl, P. Pyšek, and S. Bacher. 2012. When are eradication campaigns successful? A test of common assumptions. Biological Invasions 14:1365–1378.

Prather, L. A., O. Alvarez-Fuentes, M. H. Hayfield, and C. J. Ferguson. 2004. The decline of plant collecting in the United States: a threat to the infrastructure of biodiversity studies. Systematic Botany 29:15–28.

Pyšek, P., and P. E. Hulme. 2005. Spatio-temporal dynamics of plant invasions: linking pattern to process. Ecoscience 12:302–315.

Pyšek, P., and K. Prach. 1993. Plant invasions and the role of riparian habitats: a comparison of four species alien to central Europe. Journal of Biogeography 20:413–420.

Pyšek, P., and K. Prach. 1995. Invasion dynamics of *Impatiens glandulifera* - a century of spreading reconstructed. Biological Conservation 74:41–48.

QGIS Development Team. 2012. QGIS Geographic Information System.

R Core Team. 2012. R: A language and environment for statistical computing. Vienna, Austria.

Reddy, S., and L. Dávalos. 2003. Geographical sampling bias and its implications for conservation priorities in Africa. Journal of Biogeography 30:1719–1727.

Sakai, A. K., F. W. Allendorf, J. S. Holt, D. M. Lodge, J. Molofsky, K. A. With, S. Baughman, R. J. Cabin, J. E. Cohen, N. C. Ellstrand, D. E. McCauley, P. O’Neil, I. M. Parker, J. N. Thompson, and S. G. Weller. 2001. The population biology of invasive species. Annual Review of Ecology and Systematics 32:305–332.

Salo, L. F. 2005. Red brome (*Bromus rubens* subsp. *madritensis*) in North America: possible modes for early introductions, subsequent spread. Biological Invasions 7:165–180.

Sherff, E. E. 1912. Range extenstions of *Rhamnus frangula* and *Sporobolus asperifolius*. Rhodora 14:227–229.

Sorrie, B. A. 2005. Alien vascular plants in Massachusetts. Rhodora 107:284–329.

Theoharides, K. A., and J. S. Dukes. 2007. Plant invasion across space and time: factors affecting nonindigenous species success during four stages of invasion. New Phytologist 176:256–273.

Del Tredici, P. 2010. Wild Urban Plants of the Northeast: A Field Guide. Comstock Publishing Associates, Ithica.

United States Pharmacopeial Convention. 1910. The Pharmacopeia of the United States of America. Ninth Dece. P. Blakiston’s Son & Company, Philadelphia.

Weber, E. 1998. The dynamics of plant invasions: a case study of three exotic goldenrod species (*Solidago* L.) in Europe. Journal of Biogeography 25:147–154.

Wright, J. P., and J. D. Fridley. 2010. Biogeographic synthesis of secondary succession rates in eastern North America. Journal of Biogeography 37:1584–1596.