

The Importance of Geographic and Biological Variables in Predicting the Naturalization of Non-Native Woody Plants in the Upper Midwest¹

Mark P. Widrlechner², Emily J. Kapler³, Philip M. Dixon⁴ and Janette R. Thompson⁵

*Departments of Ecology, Evolution and Organismal Biology and Horticulture
Iowa State University, Ames, IA 50011*

Abstract

The selection, introduction, and cultivation of non-native woody plants beyond their native ranges can have great benefits, but also unintended consequences. Among these consequences is the tendency for some species to naturalize and become invasive pests in new environments to which they were introduced. In lieu of lengthy and costly field trials, risk-assessment models can be used to predict the likelihood of naturalization. We compared the relative performance of five established risk-assessment models on species datasets from two previously untested areas: southern Minnesota and northern Missouri. Model classification rates ranged from 64.2 to 90.5%, biologically significant errors ranged from 4.4 to 9.3%, and horticulturally limiting errors ranged from 6.6 to 30.4%. For the random forest model, we investigated the importance of variables used to predict naturalization by examining datasets for five distinct study areas across the Upper Midwest. Geographic-risk ratios were the most important predictors of species' tendency to naturalize. Other factors, such as quick maturity, record of invading elsewhere, and production of fleshy, bird-dispersed fruit were also important in the random forest models. Although some models tested need additional refinement, the random forest models maintain robustness and provide additional information on plant-specific characteristics that contribute to naturalization.

Index words: invasive plants, risk-assessment models, Minnesota, Missouri, random forest, variable importance.

Significance to the Nursery Industry

The introduction of new woody landscape plants can generate economic benefits for consumers and the nursery industry. Some new introductions escape from cultivation, naturalize, and invade existing plant communities. This is a concern to many stakeholders, including gardeners, members of the nursery industry, and land managers. By studying past introductions, researchers can uncover patterns in life-history characteristics and native geographic ranges that allow prediction of naturalization and invasion. We seek to develop methods to safely introduce beneficial non-native plants while protecting native ecosystems and agricultural lands from invasion by non-native plants. To this end, we developed models to predict naturalization of non-native woody landscape plants cultivated in southern Minnesota and northern Missouri. We then combined these findings with results of similar analyses from Iowa and two study areas in the Chicago region to investigate which life-history and geographic characteristics are the strongest

predictors across the Upper Midwest. In southern Minnesota and northern Missouri, the random forest method generated the best combination of species classification (as naturalizing or not naturalizing) and low error rates. When we ranked the importance of plant characteristics in random forest model predictions across the five datasets, by far the most powerful and consistent predictor of naturalization was a plant's geographic-risk ratio (G-value), a statistic based on the mean proportion of naturalizing species across an introduced plant's full native range. The next strongest predictive characteristic was whether a plant displayed rapid reproductive maturity, a trait that can be desirable for marketability, in cases when floral or fruiting display attract buyers. This suggests a need to carefully balance some desirable production characteristics with risks of naturalization and invasion.

Introduction

Over millennia, the native ranges of woody plant species have naturally expanded and contracted. In recent decades, human actions have greatly accelerated their expansion on a global basis (36). While many woody plant introductions have benefits that far outweigh their costs (42), some human-assisted plant introductions have resulted in naturalization (the ability of a plant to propagate and sustain a population outside of cultivation) and invasion (the ability of a naturalized plant to aggressively colonize and displace native plants) into new habitats, causing harm to native ecosystems and species (2, 6, 11). A recent inventory indicates that more than 700 species of woody plants (425 trees and 303 shrubs) are considered invasive in some part of the world (35). While this is only a small proportion of species that have been deliberately introduced for horticultural purposes, it is a serious and growing problem worldwide (37).

Within the scientific community, there has long been interest in understanding the characteristics that lead to species invasiveness (8) and this interest has rapidly expanded in recent years (e.g. 7, 9, 14, 15, 20). There is also growing interest in the specific impacts of these invasions on natural

¹Received for publication February 14, 2013; in revised form March 22, 2013. Journal paper of the Iowa Agriculture and Home Economics Experiment Station, Ames, IA, and supported by Hatch Act, McIntire-Stennis, and State of Iowa funds. We acknowledge additional financial support from USDA-ARS through the Floral and Nursery Crops Research Initiative. We also thank Anita Cholewa, Robin Kennedy, Matt O'Hearn, and Welby Smith for herbarium assistance, and Jeffery Iles and two anonymous peer reviewers for their useful critiques of our manuscript. Mention of commercial brand names does not constitute an endorsement of any product by the U.S. Department of Agriculture or cooperating agencies.

²Affiliate Associate Professor. Departments of Ecology, Evolution and Organismal Biology and Horticulture, Iowa State University, Ames, IA. isumw@iastate.edu.

³Master of Science. Department of Natural Resource Ecology and Management, Iowa State University. ekapler@gmail.com.

⁴University Professor. Department of Statistics, Iowa State University. pdixon@iastate.edu.

⁵Professor. Department of Natural Resource Ecology and Management, Iowa State University. jrrt@iastate.edu.

ecosystems' structure and function (12, 31, 47). It is clear that naturalization and invasion depend on both the life-history characteristics of the introduced plant and how well it is pre-adapted to a new habitat area (36).

Considerable study has been devoted to understanding how woody plants successfully naturalize or invade once arriving in a new habitat area. Researchers have proposed general schemes that include life-history characteristics, such as population fitness, generation time and level of fecundity, rate of population expansion, and individuals' competitiveness, as critical elements affecting the likelihood of becoming invasive (34, 36). Additionally, the native range of imported species, or the degree of the environmental match between the population source and the location of introduction, are important considerations (32, 38), especially at local and regional scales (49, 51, 52).

By using information gathered about introduced species, a growing body of research has specifically focused on developing predictive risk-assessment models to evaluate (ideally, before their release into the market and the landscape) the probability that a new woody plant will naturalize, invade, and cause harm (16, 20, 27, 33). As part of a larger strategy, risk assessments can proactively balance the societal and economic benefits of plant introductions against potential harms and avoid costly efforts to contain or eradicate species after escape (13, 28).

To these ends, we have developed and tested a set of predictive models to assess the likelihood of naturalization (a necessary prerequisite of invasion) for non-native woody plants in the Upper Midwest. The models use various statistical approaches to make their predictions based on both life-history characteristics (e.g., quick maturity, quick vegetative spread, invasive outside North America, and fleshy, bird-dispersed fruit) and a native-range criterion (G-value, described fully in Widrlechner et al. (52) and reiterated herein) (14, 52, 53). Our past work includes testing decision trees, CART models, combination models, and random forest techniques developed at continental, regional, and local scales on a series of introduced woody plant population datasets for Iowa (14, 52) and the Chicago area (53). Risk-assessment models generally assign one of three possible outcomes: 'accept' for plants not likely to naturalize, 'reject' for plants that are likely to naturalize, and 'further analysis' for plants when model results are unclear. Accuracy of risk-assessment models is assessed based on analysis of known naturalizing and non-naturalizing species in two ways. First, a high classification rate indicates the model is successfully placing species in 'accept' or 'reject' outcomes. Successful classification does not necessarily indicate correct classification, so the second assessment evaluates the degree to which the classifications are correct. Errors may either be described as 'horticulturally limiting' (rejecting a species not known to naturalize) or 'biologically significant' (accepting a species known to naturalize). In our previous work, classification rates have ranged between 62.0% (for a CART model applied to an Iowa dataset) and 93.1% (for decision tree and matrix models also applied to an Iowa dataset) (14); 'horticulturally limiting' errors have ranged from 3.7 to 38.5%, and 'biologically significant' errors have ranged from 1.8 to 18.5% (14, 52, 53).

This paper represents another step in the ongoing work to assess classification and error rates of risk-assessment models in the Upper Midwest by testing the relative performance of

previously developed models on two new woody plant datasets: one for southern Minnesota and another for northern Missouri. Second, we explore the relative importance of the various life-history and native-range criteria included in random forests generated from each dataset.

Materials and Methods

Our earlier analyses included plant datasets for Iowa and the Chicago area. For the research reported in this paper, we first established the scope of two new study areas for Minnesota and Missouri. Both of these study areas and corresponding datasets were originally compiled as part of an ongoing effort to validate the risk-assessment models developed for Iowa (52), and were therefore tailored to roughly resemble the range of variation in climatic conditions found in Iowa. We evaluated moisture balance, January mean temperatures, and natural geologic and plant community subdivisions to determine the counties to be included in the Minnesota and Missouri study areas (Fig. 1). Study areas in both states were required to have a positive moisture balance of $\leq 40 \text{ I}_m$ (48). In addition to this moisture requirement, the study area in Minnesota was defined by January mean air temperatures of $\geq -15\text{C}$ (59F) (43), excluding the Laurentian Mixed Forest plant community area (24). The study area for Missouri was defined by a positive moisture balance and January mean air temperatures of $\leq 0\text{C}$ (32F), excluding the Ozark and Mississippi Lowlands regions (25).

In order to include non-native woody plant taxa (naturalizing or otherwise) in our lists for both of these newly established study areas, they needed to be commonly and historically cultivated in the respective state for significant periods of time (at least 30 years). We began by consulting lists of woody plants drawn from previous studies (52, 53), pre-1980s nursery and gardening catalogs, and books on woody plants cultivated in these areas (19, 39, 40, 41). These lists were reviewed by regional experts (for southern Minnesota: Neil Anderson, Jeff Gilman, Gary Johnson, Laurie Robinson, Harold Pellett, and Mike Zins; for northern Mis-

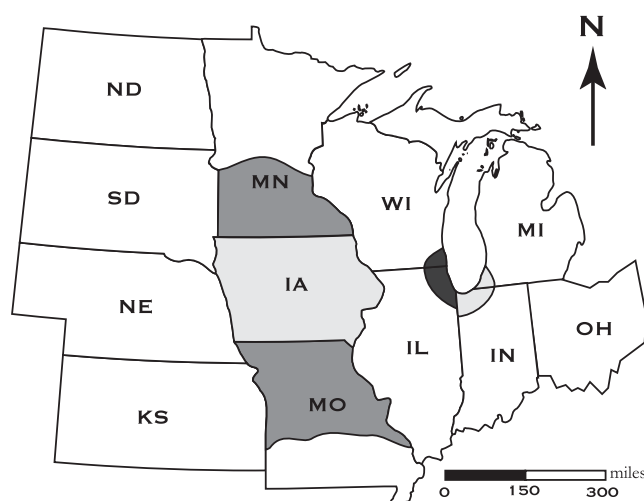


Fig. 1. Map of study areas for five risk-assessment datasets in the Upper Midwest. Study areas for their respective states are shaded. Chicago A is the more darkly shaded area in Illinois and Wisconsin and Chicago B is the less darkly shaded area in Indiana and Michigan.

souri: Alan Branhagen, June Hutson, Chris Starbuck, and Jan Vinyard) to generate a final list of species. Naturalization status was determined by consulting herbarium records at the University of Minnesota — Saint Paul, George Yatskievych's Flora of Missouri Database, and the University of Missouri — Columbia. We required a minimum of two herbarium vouchers that documented naturalization from different locations within each study area in order to designate a non-native woody plant taxon as naturalized in that study area. Taxa with ambiguous naturalization status were removed from the study. The final list for southern Minnesota included 23 naturalizing and 71 non-naturalizing species ($n = 94$), and the final list for northern Missouri included 39 naturalizing and 87 non-naturalizing species ($n = 126$) (Table 1).

Life-history characteristics were populated with data from previous studies (14, 52, 53) and additional reference sources (4, 30, 40, 41, 45). Because non-native taxa can behave differently across geography, our life-history characteristics were reviewed by horticultural experts to ensure local accuracy (for southern Minnesota: Jeff Gilman, Harold Pellett, Laurie Robinson, Nancy Rose, and Mike Zins; for northern Missouri: Alan Branhagen, Boyce Tankersley, Chris Starbuck, and Guy Sternberg). Geographic-risk ratios (G-values) were calculated independently for the southern Minnesota and northern Missouri datasets by using native range data for each species. Native range data were primarily obtained from the USDA-ARS Germplasm Resources Information Network database (46) and previous studies (14, 52, 53), with supplementation from published floras (5, 17, 44). Native range data were organized into 360 geographic subdivisions for the southern Minnesota study area and 390 geographic subdivisions for northern Missouri. G-values were calculated as the proportion of species native to a geographic subdivision that have naturalized in the study area (southern Minnesota or northern Missouri), as described by Widrlechner et al. (52). Life-history characteristics and G-values were compiled into spreadsheets for analysis and may be accessed at <http://www.nrem.iastate.edu/research/jan-t/index.php>.

After reviewing the life-history and geographic characteristics for both datasets, each dataset was independently assessed by the five risk-assessment models described in Kapler et al. (14). Four of these five models are discussed in detail in Widrlechner et al. (52) and include a 'continental decision tree' developed by Reichard and Hamilton (33), plus three models developed specifically for Iowa (52): 1) a 'modified decision tree' that adds ten steps to the continental decision tree, 2) a 'decision tree/matrix model' that focuses on reevaluating 'further analysis' species produced by the continental decision tree, and 3) a classification and regression tree model ('CART model') developed specifically for Iowa. In addition, fitted random forest models, which are an extension of CART modeling techniques, were generated with the randomForest package (21) in R software (29) for the southern Minnesota and northern Missouri datasets following Kapler et al. (14). Random forest models provide a predicted probability of naturalization from zero (will not naturalize) to one (will naturalize). We set cutoff rules within these probabilities for each dataset in order to create 'accept', 'reject', and 'further analysis' outputs as follows:

For southern Minnesota:

- If the predicted probability is < 0.205 , then classify as 'accept';

- If the predicted probability is ≥ 0.370 , then classify as 'reject'; and
- If the predicted probability is between 0.205 and 0.370, then classify as 'further analysis'.

For northern Missouri:

- If the predicted probability is < 0.278 , then classify as 'accept';
- If the predicted probability is ≥ 0.490 , then classify as 'reject'; and
- If the predicted probability is between 0.278 and 0.490, then classify as 'further analysis'.

These cutoff values were set to maximize classification rates while keeping horticulturally limiting errors and biologically significant errors at, or below, values that are considerable acceptable by a broad cross-section of stakeholders (13).

The power and accuracy of each model were assessed in three ways. First, the 'classification rate' or proportion

Table 1. Study area and naturalization status of species used in the southern Minnesota and northern Missouri data sets.

Species	Naturalized in southern Minnesota	Naturalized in northern Missouri
<i>Abies concolor</i> (Gordon & Glend.) Lindl. ex Hildebr.	N	N
<i>Acer griseum</i> (Franch.) Pax		N
<i>Acer palmatum</i> Thunb.		N
<i>Acer platanoides</i> L.	Y	N
<i>Acer tataricum</i> L. subsp. <i>ginnala</i> (Maxim.) Wesm.	Y	Y
<i>Aesculus hippocastanum</i> L.	N	
<i>Ailanthus altissima</i> (Mill.) Swingle		Y
<i>Albizia julibrissin</i> Durazz.		Y
<i>Alnus glutinosa</i> (L.) Gaertn.	N	
<i>Amelanchier canadensis</i> (L.) Medik.	N	N
<i>Aronia arbutifolia</i> (L.) Pers.	N	N
<i>Berberis koreana</i> Palib.	N	
<i>Berberis thunbergii</i> DC.	Y	Y
<i>Berberis vulgaris</i> L.		Y
<i>Betula lenta</i> L.	N	
<i>Betula pendula</i> Roth.	N	
<i>Broussonetia papyrifera</i> (L.) Vent.		Y
<i>Buddleja davidii</i> Franch.		Y
<i>Buxus microphylla</i> Siebold & Zucc.		N
<i>Callicarpa dichotoma</i> (Lour.) K. Koch		N
<i>Caragana arborescens</i> Lam.	Y	
<i>Caragana frutex</i> (L.) K. Koch	N	
<i>Caragana pygmaea</i> DC.	N	
<i>Carpinus betulus</i> L.		N
<i>Castanea mollissima</i> Blume		N
<i>Catalpa speciosa</i> (Warder) Engelm.	Y	
<i>Celastrus orbiculatus</i> Thunb.		Y
<i>Chionanthus virginicus</i> L.		N
<i>Cladrastis kentukea</i> (Dum. Cours.) Rudd		N
<i>Clematis terniflora</i> DC.		Y
<i>Clethra alnifolia</i> L.	N	N
<i>Cornus alba</i> L.	N	N
<i>Cornus kousa</i> Hance		N
<i>Cornus mas</i> L.		N
<i>Cornus sanguinea</i> L.		N
<i>Cornus sericea</i> L. subsp. <i>sericea</i>		N
<i>Corylus avellana</i> L.		N
<i>Cotinus coggygria</i> Scop.	N	N

Continued...

Table 1. Continued.

Species	Naturalized in southern Minnesota	Naturalized in northern Missouri
<i>Cotinus obovatus</i> Raf.	N	
<i>Cotoneaster acutifolius</i> Turcz.	N	N
<i>Cotoneaster adpressus</i> Bois		N
<i>Cotoneaster apiculatus</i> Rehder & E.H. Wilson		N
<i>Cotoneaster divaricatus</i> Rehder & E.H. Wilson	N	N
<i>Cotoneaster horizontalis</i> Dence.		N
<i>Cotoneaster integerrimus</i> Medik.	N	
<i>Cotoneaster lucidus</i> Schldt.	N	
<i>Crataegus crus-galli</i> L.	N	
<i>Deutzia gracilis</i> Siebold & Zucc.		N
<i>Diervilla lonicera</i> Mill.		N
<i>Diervilla sessilifolia</i> Buckley		N
<i>Elaeagnus angustifolia</i> L.	Y	Y
<i>Elaeagnus umbellata</i> Thunb.		Y
<i>Euonymus alatus</i> (Thunb.) Siebold		Y
<i>Euonymus europaeus</i> L.	N	
<i>Euonymus fortunei</i> (Turcz.) Hand.-Mazz.	N	Y
<i>Euonymus nanus</i> M. Bieb.	N	
<i>Fagus sylvatica</i> L.		N
<i>Frangula alnus</i> Mill.	Y	N
<i>Hibiscus syriacus</i> L.		Y
<i>Hippophae rhamnoides</i> L.	N	
<i>Hydrangea arborescens</i> L.	N	
<i>Hydrangea macrophylla</i> (Thunb.) Ser.		N
<i>Hydrangea paniculata</i> Siebold.	N	N
<i>Hydrangea quercifolia</i> W. Bartram		N
<i>Ilex crenata</i> Thunb.		N
<i>Juniperus chinensis</i> L.	N	N
<i>Juniperus horizontalis</i> Moench		N
<i>Juniperus sabina</i> L.	N	N
<i>Juniperus scopulorum</i> Sarg.	N	N
<i>Juniperus squamata</i> Buch.-Ham. ex D. Don	N	N
<i>Kerria japonica</i> (L.) DC.		N
<i>Koeleruteria paniculata</i> Laxm.		Y
<i>Kolkwitzia amabilis</i> Graebn.		N
<i>Lespedeza thunbergii</i> (DC.) Nakai		Y
<i>Ligustrum obtusifolium</i> Siebold & Zucc.		Y
<i>Ligustrum vulgare</i> L.	N	Y
<i>Lonicera japonica</i> Thunb. ex Murray		Y
<i>Lonicera maackii</i> (Rupr.) Maxim.	N	Y
<i>Lonicera morrowii</i> A. Gray	Y	
<i>Lonicera sempervirens</i> L.	N	
<i>Lonicera tatarica</i> L.	Y	
<i>Lonicera xylosteum</i> L.	Y	
<i>Lycium barbarum</i> L.	Y	Y
<i>Maackia amurensis</i> Rupr. & Maxim.	N	
<i>Maclura pomifera</i> (Raf.) C.K. Schneid.		Y
<i>Magnolia grandiflora</i> L.		N
<i>Magnolia stellata</i> (Siebold & Zucc.) Maxim.	N	N
<i>Malus pumila</i> Mill.	Y	
<i>Metasequoia glyptostroboides</i> Hu & W.C. Cheng		N
<i>Microbiota decussata</i> Kom.	N	
<i>Morus alba</i> L.	Y	Y
<i>Parthenocissus triscuspidata</i> (Sieb. & Zucc.) Planach.	N	
<i>Phellodendron amurense</i> Rupr.	N	
<i>Philadelphus coronarius</i> L.	N	
<i>Picea abies</i> (L.) Karst.	N	N
<i>Picea glauca</i> (Moench) Voss		N
<i>Picea pungens</i> Engelm.	N	N
<i>Pinus cembra</i> L.	N	
<i>Pinus mugo</i> Turra	N	N
<i>Pinus nigra</i> Arnold.	N	N
<i>Pinus ponderosa</i> P. Lawson & C. Lawson	N	
<i>Platycladus orientalis</i> (L.) Franco		N
<i>Populus alba</i> L.	Y	Y
<i>Prunus cerasifera</i> Ehrh.	N	N
<i>Prunus glandulosa</i> Thunb.		N

Continued ...

Table 1. Continued.

Species	Naturalized in southern Minnesota	Naturalized in northern Missouri
<i>Prunus maackii</i> Rupr.		N
<i>Prunus mahaleb</i> L.		Y
<i>Prunus mandshurica</i> (Maxim.) Koehne	N	
<i>Prunus padus</i> L.	N	
<i>Prunus pumila</i> L. var. <i>besseyi</i> (L.H. Bailey) Gleason		N
<i>Prunus sargentii</i> Rehder		N
<i>Prunus serrulata</i> Lindl.		N
<i>Prunus subhirtella</i> Miq.		N
<i>Prunus triloba</i> Lindl.		N
<i>Pseudotsuga menziesii</i> (Mirb.) Franco	N	
<i>Pyracantha angustifolia</i> (Franch.) C.K. Schneid.		N
<i>Pyracantha coccinea</i> M. Roem.		N
<i>Pyrus ussuriensis</i> Maxim.	N	
<i>Quercus acutissima</i> Carruthers		Y
<i>Rhamnus cathartica</i> L.	Y	Y
<i>Rhamnus davurica</i> Pall.		Y
<i>Rhododendron molle</i> (Blume) G. Don	N	
<i>Rhodotypos scandens</i> (Thunb.) Makino		N
<i>Rhus typhina</i> L.		N
<i>Ribes alpinum</i> L.		N
<i>Robinia pseudoacacia</i> L.	Y	
<i>Rosa gallica</i> L.		Y
<i>Rosa luciae</i> Franch. & Rochebr. ex Crép.		Y
<i>Rosa multiflora</i> Thunb.	Y	Y
<i>Rosa multiflora</i> L.		Y
<i>Rosa rugosa</i> Thunb.		N
<i>Salix alba</i> L.	N	Y
<i>Salix caprea</i> L.	N	N
<i>Salix fragilis</i> L.	Y	
<i>Salix integra</i> Thunb.		N
<i>Salix matsudana</i> Koidz.		N
<i>Salix pentandra</i> L.	Y	
<i>Salix purpurea</i> L.	N	N
<i>Sambucus nigra</i> L.		N
<i>Sorbaria sorbifolia</i> (L.) A. Braun	N	
<i>Sorbus aucuparia</i> L.	Y	
<i>Spiraea albiflora</i> (Miq.) Zabel		N
<i>Spiraea betulifolia</i> Pall.	N	N
<i>Spiraea fritschiana</i> C.K. Schneid.	N	
<i>Spiraea japonica</i> L.f.	N	N
<i>Spiraea nipponica</i> Maxim.	N	N
<i>Spiraea prunifolia</i> Siebold & Zucc.		Y
<i>Spiraea thunbergii</i> Siebold ex Blume	N	N
<i>Spiraea trilobata</i> L.	N	
<i>Styphnolobium japonicum</i> (L.) Schott		N
<i>Symphoricarpos orbiculatus</i> Moench	Y	
<i>Syringa meyeri</i> C.K. Schneid.	N	
<i>Syringa pubescens</i> Turcz.	N	N
<i>Syringa reticulata</i> (Blume) H. Hara		N
<i>Syringa villosa</i> Vahl		N
<i>Syringa vulgaris</i> L.		N
<i>Tamarix ramosissima</i> Ledeb.		Y
<i>Taxus cuspidata</i> Siebold & Zucc.	N	N
<i>Thuja occidentalis</i> L.		N
<i>Tilia cordata</i> Mill.	N	N
<i>Tilia ×europaea</i> L.		N
<i>Tsuga canadensis</i> (L.) Carrière		N
<i>Ulmus parvifolia</i> Jacq.		N
<i>Ulmus pumila</i> L.	Y	Y
<i>Vaccinium corymbosum</i> L.		N
<i>Viburnum carlesii</i> Hemsl.		N
<i>Viburnum cassinoides</i> L.		N
<i>Viburnum dentatum</i> L.	N	
<i>Viburnum opulus</i> L. var. <i>opulus</i>	Y	Y
<i>Viburnum prunifolium</i> L.	N	
<i>Weigela florida</i> (Bunge) A. DC.	N	N
<i>Wisteria frutescens</i> (L.) Poir.		Y
<i>Wisteria sinensis</i> (Sims) Sweet		Y
<i>Zelkova serrata</i> (Thunb.) Makino		N

Table 2. Summary of classification and error rates for five risk-assessment models by data set.

Model	Classification rate (%)	Biologically significant error rate (%)	Horticulturally limiting error rate (%)
Continental decision tree			
Southern Minnesota	72.3	5.9	30.9
Northern Missouri	72.2	4.4	28.6
Modified decision tree			
Southern Minnesota	90.4	9.4	23.5
Northern Missouri	89.7	5.3	25.7
Decision tree/matrix model			
Southern Minnesota	88.3	7.2	26.5
Northern Missouri	87.3	4.6	29.1
CART model			
Southern Minnesota	63.8	10.0	6.7
Northern Missouri	81.0	7.8	11.8
Random forest model			
Southern Minnesota	80.0	5.9	11.8
Northern Missouri	81.8	7.8	9.7

of species that a model successfully classifies ('accept' or 'reject') was examined as a measure of a model's power. We also assessed two types of errors expressed as the proportion of misclassifications to the total number of classified species: 1) 'horticulturally limiting errors' or non-naturalizing species that the models rejected as naturalizing and, 2) 'biologically significant errors' or naturalizing species that the models accepted as non-naturalizing. Both of these metrics are described in detail by Widrlechner et al. (52, 53). To determine the relative contributions of each individual life-history and geographical characteristic to the predictive strength of the random forest models, variable importance plots were generated in R software as part of the modeling process for each dataset. The quality of a node in one tree in the random forest can be measured by the change in the Gini index (10) when that node is added to the tree:

$$\text{Gini} = 1 - \sum_{j=1}^2 p_j^2$$

where p_j is the probability that a species classified into that node naturalizes ($j = 1$) or does not ($j = 2$). A node where all species are correctly classified has $\text{Gini} = 0$. The importance of a variable is calculated by considering all splits based on that variable and calculating the total drop in the Gini index for those splits (10). Because the sum of variable importance values for all traits depends on the overall proportion of naturalizing species in the study and the number of nodes in the tree, we calculated the relative importance of each variable by expressing the importance as the proportion of the total importance for all variables.

Results and Discussion

Model performance. Fitting the models to predict naturalization status in southern Minnesota and northern Missouri generated similar results to those from Iowa (14, 52) and two regions in the Chicago area (53). The extended Reichard and Hamilton models (the modified decision tree and the decision tree/matrix model) have higher classification rates (between 87.3 and 90.4%) than does the original Reichard and Hamilton continental model (about 72% for both areas; Table 2). All three models have similar horticultural (23.5 to 30.9%) and biological (4.4 to 9.4%) error rates (Table 2). The CART model has a much lower classification rate for the southern Minnesota study area (63.8%), a somewhat

lower classification rate for the northern Missouri study area (81.0%) and slightly elevated biological error rates (10.0% in Minnesota and 7.8% in Missouri) compared to the original and extended Reichard and Hamilton models. To its credit, the CART model produced much lower horticulturally limiting error rates in both states (6.7% in southern Minnesota and 11.8% in northern Missouri) compared to the original and extended Reichard and Hamilton models. Relative to Reichard and Hamilton, the modified decision tree, and the decision tree/matrix model, the random forest model had lower classification rates (80.0% for southern Minnesota and 81.8% for northern Missouri), similar biological error rates (5.9% in southern Minnesota and 7.8% in northern Missouri), and much lower horticulturally limiting error rates (11.8% in southern Minnesota and 9.7% in northern Missouri) (Table 2).

Misclassified species. As we have previously noted (14, 52), it is instructive to examine misclassified species more carefully. The random forest model gives unanticipated results for several species in the two study areas (Table 3). Two naturalizing species were misclassified as 'accept' in both study areas: *Berberis thunbergii* and *Viburnum opulus*. *Berberis thunbergii* is endemic to Japan, and in contrast to many woody plants endemic to Japan (50), this shrub is well adapted to seasonal moisture deficits similar to those that characterize the Upper Midwest. We have also observed it to be increasing in woodlands in Iowa and Illinois that are heavily browsed by deer. Notably, the Japanese barberry was also misclassified by risk-assessment models for both Iowa and the Chicago region (52,53).

Old World populations of *V. opulus* have also been naturalizing throughout the Midwest, although typically in relatively low abundance. It is possible that the introduced Old World populations are hybridizing with local populations of *V. opulus* var. *americanum*, at least in the northern part of our region, which may be increasing their adaptation to local conditions.

The naturalization of *Wisteria frutescens* in northern Missouri may be related to a very different phenomenon. Compared to other regions of the world, the southeastern United States is a low-risk region for the introduction of naturalizing woody plants into the Midwest (51). But the native range of this southeastern vine extends north and

Table 3. Species producing biologically significant and horticulturally limiting errors in local random forests for the southern Minnesota and northern Missouri datasets.

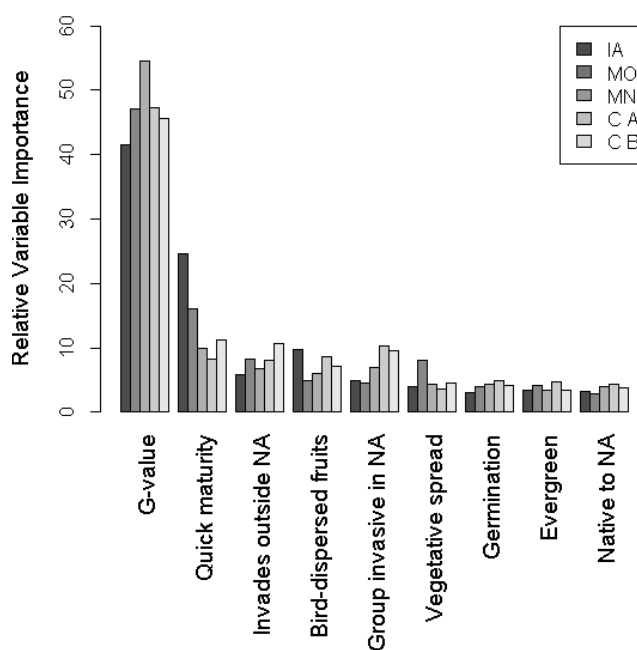
Study area	Species	Naturalized?	Predicted probability of naturalizing (%)
Southern Minnesota	<i>Berberis thunbergii</i> DC.	Y	8.4
	<i>Viburnum opulus</i> L. var. <i>opulus</i>	Y	5.4
	<i>Alnus glutinosa</i> (L.) Gaertn.	N	69.0
	<i>Euonymus europaeus</i> L.	N	81.5
	<i>Ribes alpinum</i> L.	N	59.0
Northern Missouri	<i>Berberis thunbergii</i> DC.	Y	6.1
	<i>Viburnum opulus</i> L. var. <i>opulus</i>	Y	1.3
	<i>Spiraea prunifolia</i> Siebold & Zucc.	Y	9.0
	<i>Wisteria frutescens</i> (L.) Poir.	Y	9.1
	<i>Castanea mollissima</i> Blume	N	69.0
	<i>Cotoneaster divaricatus</i> Rehder & E.H. Wilson	N	91.6
	<i>Prunus cerasifera</i> Ehrh.	N	73.8
	<i>Rosa rugosa</i> Thunb.	N	76.1

west into the forests of southeastern Missouri and southern Illinois (19, 26) very near the boundary of our Missouri study area. It is possible that a combination of its cultivation as an ornamental vine and natural dispersal events is leading to its range expansion.

When examining species that are predicted to naturalize but have not, one must also keep in mind the possibility that these species may still be in a lag phase (18) prior to naturalization or they may be undocumented naturalizers. Long-lived trees, such as *Castanea mollissima*, often have lengthy lag phases. *Euonymus europaeus* may be an example of the latter, an undocumented naturalizer. In this case, the European spindle-tree is quite difficult to distinguish from its native congener, *E. atropurpureus*, except when these plants are in flower, which only occurs for a short period in the spring greatly reducing the likelihood of detection.

There are other potential explanations for the lack of naturalization among species predicted to naturalize. For example, *Alnus glutinosa* is known to naturalize in northern Illinois (53) and has become so invasive along waterways in Du Page County, Illinois, that extensive efforts have been directed toward large-scale removal (Joseph Suchecki, personal communication). In the Old World, *A. glutinosa* has an extensive native range, covering an extremely broad latitudinal band. However, provenance testing across multiple locations in the Midwest has shown that each Old World population has a very narrow band of latitudinal adaptation (3, 23). Thus, it is likely that *A. glutinosa* will not naturalize in southern Minnesota unless populations that are well adapted to that latitude and climate are widely cultivated there. The case of *Ribes alpinum* is somewhat different. It is very well adapted in southern Minnesota and is widely planted as a low-growing, dense shrub in residential and commercial landscapes. For this dioecious species, however, fruit production is only rarely observed, which may be due to a preference for the cultivation of single clones in blocks, and for selection of staminate clones. Without fruit production, it is unlikely to escape from cultivation. And, finally, in the case of the species with the highest predicted chance of naturalization (Table 3), *Cotoneaster divaricatus*, we suspect that the discrepancy may be the result of a gradual decline in its cultivation in northern Missouri. As noted by Lockwood et al. (22), a critical degree of propagule pressure is often needed before naturalization or invasion is observed.

Variable importance. We determined the relative importance of life-history and geographical characteristics in the random forest models for each study area (one Iowa and two Chicago datasets, along with the datasets for Minnesota and Missouri presented in this paper) to look for any patterns in the results (Fig. 2). Using simulation, Archer and Kimes (1) show that the random forest variable importance plot successfully identifies the correct variables to include in a classification model. The two variables most closely associated with native range, ‘Native to North America’ and ‘G-value’, display strikingly different results. ‘Native to North America’ is the least important variable in model construction, whereas ‘G-value’ is, by far, the most important determinant in all five study areas. The contribution of ‘G-value’ to each of the five random forest models, as reflected in standardized relative importance, ranges between 41.5 and 54.5% (Fig. 2). At least

**Fig. 2. Variable importance in the random forest model for five datasets representing the Upper Midwest.**

within this region, the role that geographic pre-adaptation plays in allowing cultivated woody plants to naturalize cannot be overemphasized. Previous research has demonstrated that much of this pre-adaptation is related to climatic analogs (a close match between climatic characteristics of a plant's native range and those of the location where it is introduced). This is true both within our region (49, 51) and more broadly (as reviewed by Richardson and Thullier, 38).

Of the biological characteristics included in the random forest model, 'Quick maturity' stands out as being of particular importance, especially in Missouri and Iowa, where its relative importance is 2.0 to 2.5 times greater than any other biological trait. This underscores the importance of a relatively brief establishment phase leading to consistent reproduction. In other words, minimum generation time, especially for woody plants, leads to greater propagule pressure in a shorter period of time, which has widely been documented as a likely precursor of naturalization and invasion (36).

'Fleshy, bird-dispersed fruits' and 'Quick vegetative spread' are of somewhat lesser importance. Of these two, 'Fleshy, bird-dispersed fruits' is most important in Iowa. This characteristic was added to risk-assessment models developed solely from the Iowa dataset (51) in order to improve upon the classification and error rates of the continental decision tree, so this result is not surprising. We hypothesize that seed dispersal by birds may be most important in landscapes resembling those of Iowa, where forest areas are highly fragmented. Bird dispersal can facilitate colonization from managed landscapes to new sites even across inhospitable areas that may exist between them. Of the five study areas, 'Vegetative spread' is most important in northern Missouri, and it may prove to be more important in densely wooded plant communities in milder climates, whenever seed propagation is limiting.

Taken together, the two characteristics related to invasion history for each species, 'Group invasive in North America' and 'Invades outside North America', are comparable to 'Quick maturity' in contributing to our random forest models. These two traits play key roles in Reichard and Hamilton's continental model (33) for North America, wherein 'Invades outside North America' is the first criterion, and 'Group invasive in North America' is used at three subsequent nodes in the decision tree.

In our study areas, the remaining characteristics contribute relatively little to predicting the probability of naturalization in the random forest models. Information on seed dormancy and on leaf persistence of broad-leaved evergreens under a wide range of environmental conditions is not always readily available. Our results suggest that special efforts may not be needed to identify these two character states before proceeding to model development. In contrast, for most taxa, it is relatively easy to determine whether a plant is native to North America (required to calculate G-values), but that fact *per se* also contributes little to random forest models.

Concluding thoughts. Classification and error rates for the five risk-assessment models applied to species datasets from southern Minnesota and northern Missouri resembled those reported in earlier work for other areas within the Upper Midwest (Iowa, and two areas near Chicago; 14, 52, 53). Across all of these datasets, the random forest models have consistently produced acceptable classification rates, as well

as reasonable biological and horticultural error rates (results reported here, as well as in Kapler et al., 14). In addition, the ability to discern which variables in the model are making important contributions to predicting species' naturalization, as is true for the random forest models, is helpful in understanding the relative performance of different models. In our ongoing work, we are developing regional-scale models and comparing their performance to those of the local models that have been generated for each of these datasets. It is important to note that the utility of any floristic risk assessment model is only as good as the data available to construct it. We encourage regular floristic surveys and voucher specimen collection, particularly documenting non-native species, to provide better data on floristic change.

Literature Cited

1. Archer, K.J. and R.V. Kimes. 2008. Empirical characterization of random forest variable importance measures. *Computational Statistics and Data Analysis* 52:2249–2260.
2. Brooks, M., C. D'Antonio, D. Richardson, J. Grace, J. Keeley, J. DiTomaso, R. Hobbs, M. Pellant, and D. Pyke. 2004. Effects of invasive alien plants on fire regimes. *BioScience* 54:677–688.
3. DeWald, L.E. and K.C. Steiner. 1986. Phenology, height increment, and cold tolerance of *Alnus glutinosa* populations in a common environment. *Silvae Genetica* 35:205–211.
4. Dirr, M.A. 1998. *Manual of Woody Landscape Plants*, 5th Ed. Stipes, Champaign, IL.
5. eFloras. 2010. Flora of China online volumes. Accessed 2010. http://www.efloras.org/flora_page.aspx?flora_id=2.
6. Ehrenfeld, J. 2003. Effects of exotic plant invasions on soil nutrient cycling processes. *Ecosystems* 6:503–523.
7. Eldridge, D.J., F.T. Maestre, S. Maltez-Mouro, and M.A. Bowker. 2012. A global database of shrub encroachment effects on ecosystem structure and functioning. *Ecology* 93:2499.
8. Elton, C.S. 1958. *The Ecology of Invasions by Animals and Plants*. Methuen, London, UK.
9. Grotkopp, E., J. Erskine-Ogden, and M. Rejmánek. 2010. Assessing potential invasiveness of woody horticultural plant species using seedling growth rate traits. *J. Appl. Ecol.* 47:1320–1328.
10. Hastie, T., R. Tibshirani, and J. Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd Ed. Springer, New York, NY.
11. Hejda, M., P. Pyšek, and V. Jarošík. 2009. Impact of invasive plants on the species richness, diversity, and composition of invaded communities. *J. Ecol.* 97:393–403.
12. Holland-Clift, S., D. O'Dowd, and R. Nally. 2011. Impacts of invasive willow (*Salix rubens*) on riparian bird assemblages in south-eastern Australia. *Austral Ecol.* 36:511–520.
13. Kapler, E.J., J.R. Thompson, and M.P. Widrechner. 2012. Assessing stakeholder perspectives on invasive plants to inform risk analysis. *Invasive Plant Science and Management* 5:194–208.
14. Kapler, E.J., M.P. Widrechner, P.M. Dixon, and J.R. Thompson. 2012. Performance of five models to predict the naturalization of non-native woody plants in Iowa. *J. Environ. Hort.* 30:35–41.
15. Keller, R., D. Kocov, and S. Džeroski. 2011. Trait-based risk assessment for invasive species: High performance across diverse taxonomic groups, geographic ranges and machine learning/statistical tools. *Divers. Distrib.* 17:451–461.
16. Keller, R., D. Lodge, and D. Finnoff. 2007. Risk assessment for invasive species produces net bioeconomic benefits. *Proc. Natl. Acad. Sci. (USA)* 104:203–207.
17. Komarov, V.L. et al. (Eds.). 1934–1964. *Flora of the USSR*, 20 vols. Academy of Sciences of the USSR, Moscow and Leningrad.
18. Kowarik, I. 1995. Time lags in biological invasions with regard to the success and failure of alien species. p. 15–38. *In: Plant Invasions: General*

Aspects and Special Problems. P. Pyšek, K. Prach, M. Rejmánek, and M. Wade (Eds.). SPB Academic Publishing, Amsterdam.

19. Kurz, D. 1997. Shrubs and Woody Vines of Missouri. Missouri Department of Conservation, Jefferson City, MO.

20. Leung, B., N. Roura-Pascual, S. Bacher, J. Heikkilä, L. Brotons, M.A. Burgman, K. Dehnen-Schmutz, F. Essl, P.E. Hulme, D.M. Richardson, D. Sol, and M. Vilà. 2012. TEASIng apart alien species risk assessments: A framework for best practices. *Ecol. Lett.* 12:1475–1493.

21. Liaw, A. and M. Wiener. 2002. Classification and regression by random Forest. *R News* 2(3):18–22.

22. Lockwood, J.L., P. Cassey, and T. Blackburn. 2005. The role of propagule pressure in explaining species invasions. *Trends in Ecology & Evolution* 20:223–228.

23. Maynard, C.A. and R.B. Hall. 1980. Early results of a range-wide provenance trial of *Alnus glutinosa* (L.) Gaertn. *Proc. NE For. Improv. Conf.* 27:184–201.

24. Minnesota Department of Natural Resources. 2012. Laurentian Mixed Forest Province. Accessed 2012. <http://www.dnr.state.mn.us/ecs/212/index.html>.

25. Missouri Department of Conservation. 2012. The Natural Divisions of Missouri. Accessed 2012. <http://extra.mdc.mo.gov/nathis/birds/birdatlas/natdiv.htm>.

26. Mohlenbrock, R.H. and D.M. Ladd. 1978. Distribution of Illinois Vascular Plants. Southern Illinois Univ. Press, Carbondale, IL.

27. Ou, L., C. Lu, and D. O'Toole. 2008. A risk-assessment system for alien plant bio-invasion in Xiamen, China. *J. Environ. Sci.* 20:989–997.

28. Pimentel, D. 2011. Biological Invasions: Economic and Environmental Costs of Alien Plant, Animal and Microbe Species. 2nd Ed. CRC Press, New York, NY.

29. R Development Core Team. 2011. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Accessed 2012. <http://R-project.org>.

30. Randall, R. 2003. Rod Randall's big weed list. Accessed 2010. <http://www.invasive.org/gist/biglist.html>.

31. Rascher, K.G., C. Hellmann, C. Máguas, and C. Werner. 2012. Community scale ¹⁵N isoscapes: Tracing the spatial impact of an exotic N₂-fixing invader. *Ecol. Lett.* 15:484–491.

32. Reichard, S. 2001. The search for patterns that enable the prediction of invasion. p. 10–19. *In*: R. Groves, F. Panetta, and J. Virtue (Eds.). *Weed Risk Assessment*. CSIRO Publishing, Collingwood, Australia.

33. Reichard, S.H. and C.W. Hamilton. 1997. Predicting invasions of woody plants introduced into North America. *Conserv. Biol.* 11:193–203.

34. Rejmánek, M. 2011. Invasiveness. p. 379–385. *In*: D. Simberloff and M. Rejmánek (Eds.). *Encyclopedia of Biological Invasions*. University of California Press, Berkeley, CA.

35. Rejmánek, M. and D.M. Richardson. 2013. Trees and shrubs as invasive alien species — 2013 update of the global database. *Divers. Distrib.* Accessed 2013. DOI:10.1111/ddi.12075.

36. Rejmánek, M., D. Richardson, and P. Pyšek. 2013. Plant invasions and invisibility of plant communities. p. 387–424. *In*: E. van der Maarel

and J. Franklin (Eds.). *Vegetation Ecology*. John Wiley and Sons, New York, NY.

37. Richardson, D. and M. Rejmánek. 2011. Trees and shrubs as invasive alien species — a global review. *Divers. Distrib.* 17:788–809.

38. Richardson, D. and W. Thullier. 2007. Home away from home: Objective mapping of high-risk source areas for plant introductions. *Divers. Distrib.* 13:299–312.

39. Rocca, J.P., D. Starkey, and E. Heflin. 1979. Trees for Urban Missouri. Missouri Department of Conservation, Jefferson City, MO.

40. Rose, N., D. Selinger, and J. Whitman. 2000. Growing Shrubs and Trees in Cold Climates, 1st edition. NTC/Contemporary Publishing Group Inc., Chicago, IL.

41. Smith, W.R. 2008. Trees and Shrubs of Minnesota. University of Minnesota Press, Minneapolis, MN.

42. Smith, C., W. Lonsdale, and J. Fortune. 1999. When to ignore advice: Invasion predictions and decision theory. *Biol. Invasions* 1:89–96.

43. Steinhauser, F. (Ed.). 1979. Climatic atlas of North and Central America. Vol. 1. Maps of mean temperature. WMO, UNESCO, Geneva, Switzerland. 28 maps.

44. Tutin, T.G. et al. (Eds.). 1964–1994. *Flora Europaea*, 5 vols. Cambridge University Press, Cambridge, UK.

45. U.S. Department of Agriculture, Forest Service. 2008. The Woody Plant Seed Manual. Accessed 2010. http://www.nsl.fs.fed.us/nsl_wpsm.html.

46. U.S. Department of Agriculture, Agricultural Research Service. 2010. Germplasm Resources Information Network database. Accessed 2011. <http://www.ars-grin.gov/npgs/>.

47. Vilà, M., J. Espinar, M. Hejda, P. Hulme, V. Jarošík, J. Maron, J. Pergl, U. Schaffner, Y. Sun, and P. Pyšek. 2011. Ecological impacts of invasive alien plants: A meta-analysis of their effects on species, communities, and ecosystems. *Ecol. Lett.* 14:702–708.

48. Widrlechner, M.P. 1999. A zone map for mean annual moisture balance in the North Central United States. *Landscape Plant News* 10(2):10–14.

49. Widrlechner, M.P. 2001. The role of environmental analogs in identifying potentially invasive woody plants in Iowa. *J. Iowa Acad. Sci.* 108:158–165.

50. Widrlechner, M.P., J.B. Hebel, D.E. Herman, J.K. Iles, G.J. Kling, A.P. Ovrom, J.C. Pair, E.T. Paparozzi, S.R. Poppe, N. Rose, R.E. Schutzki, C. Tubesing, and D.K. Wildung. 1998. Performance of landscape plants from northern Japan in the north central United States. *J. Environ. Hort.* 16:27–32.

51. Widrlechner, M.P. and J. Iles. 2002. A geographic assessment of the risk of naturalization of non-native woody plants in Iowa. *J. Environ. Hort.* 20:47–56.

52. Widrlechner, M.P., J.R. Thompson, J.K. Iles, and P.M. Dixon. 2004. Models for predicting the risk of naturalization of non-native woody plants in Iowa. *J. Environ. Hort.* 22:23–31.

53. Widrlechner, M.P., J.R. Thompson, E.J. Kapler, K. Kordecki, P.M. Dixon, and G. Gates. 2009. A test of four models to predict the risk of naturalization of non-native woody plants in the Chicago Region. *J. Environ. Hort.* 27:241–250.