



Modeling Distribution of High-Priority Exotic Plant Species on the Olympic Peninsula

National Park Service
North Coast and Cascades Network

Chad C. Jones
Charles B. Halpern
College of Forest Resources
Box 352100
University of Washington
Seattle, WA 98195-2100

FINAL REPORT

August 2007

Table of Contents

| | |
|-----------------------|----|
| Introduction..... | 1 |
| Methods | 4 |
| Results..... | 14 |
| Discussion..... | 20 |
| References Cited..... | 30 |

List of Figures

| | |
|--|----|
| Figure 1. The Olympic Peninsula | 4 |
| Figure 2. Suitability predictions for three statistical modeling techniques | 15 |
| Figure 3. Model accuracy measures for three statistical modeling techniques | 16 |
| Figure 4. Variability in model results among replicates | 17 |
| Figure 5. Habitat suitability maps for <i>Cirsium arvense</i> | 18 |
| Figure 6. Habitat suitability maps for <i>Geranium robertianum</i> | 19 |
| Figure 7. Habitat suitability maps for <i>Hedera helix</i> | 20 |
| Figure 8. Habitat suitability maps for <i>Ilex aquifolium</i> | 21 |
| Figure 9. Habitat suitability maps for <i>Rubus laciniatus</i> | 22 |
| Figure 10. Invasion-risk classification for <i>Geranium</i> , <i>Hedera</i> and <i>Ilex</i> | 23 |
| Figure 11. Invasion-risk classification in ONP for <i>Geranium</i> , <i>Hedera</i> and <i>Ilex</i> | 24 |
| Figure 12. Invasion-risk classification for <i>Cirsium</i> and <i>Rubus</i> | 25 |
| Figure 13. Invasion-risk classification in ONP for <i>Cirsium</i> and <i>Rubus</i> | 26 |

List of Tables

| | |
|---|----|
| Table 1. Summary of datasets used | 8 |
| Table 2. Number of locations where target species are present | 8 |
| Table 3. Habitat variables considered for modeling | 9 |
| Table 4. Climate variables used to create ‘native-range’ and ‘invaded-range’ models | 13 |
| Table 5. Accuracy for ‘native-range’ and ‘invaded-range’ models | 23 |
| Table 6. Percent of study area and ONP in different risk categories for <i>Geranium</i> , <i>Hedera</i> and <i>Ilex</i> | 25 |
| Table 7. Percent of study area and ONP in different risk categories for <i>Cirsium</i> and <i>Rubus</i> | 25 |

List of Appendices

| | |
|---|-----|
| Appendix A. Species Datasets..... | A-1 |
| Appendix B: Autocorrelation in Habitat Variables and Species Distributions..... | B-1 |
| Appendix C: Current Distributions of the Target Species on Olympic Peninsula..... | C-1 |
| Appendix D: Habitat Variable GIS Layers..... | D-1 |
| Appendix E: File Directory of GIS Layers..... | E-1 |

Executive Summary

- Invasive exotic plants can have important ecological impacts on native species and communities and are recognized as a management concern in national parks.
- Control of invasive species is easiest in the early stages of invasion, thus monitoring for these species is critical. If managers can predict where species are likely to occur, the limited resources available for monitoring can be focused in these areas.
- Habitat distribution models (HDMs) provide a potential method for predicting where invasive species are likely to occur by using information on habitat characteristics where species currently do and do not occur.
- Although HDMs are commonly used in conservation biology, there are several challenges in using them to predict the distributions of invasive species. First, these models assume that species are not spreading, which is clearly not the case for most invasives. Second, information about the distribution of invasive species is often concentrated in areas near roads or where the species is already common. This makes it difficult to make reliable predictions in other areas. Finally, there are many different techniques that can be used to develop these models. Logistic regression is most commonly used, but other techniques such as Genetic Algorithm for Rule-set Prediction (GARP) and Ecological Niche Factor Analysis (ENFA) may be more effective for modeling invasive species.
- Some of these challenges can be overcome or minimized by comparing models with others developed using data from other invaded areas or from a species' native range.
- In this project we modeled the potential distributions of five invasive species of concern in Olympic National Park: *Cirsium arvense* (Canada thistle), *Geranium robertianum* (herb Robert), *Hedera helix* (English ivy), *Ilex aquifolium* (English holly) and *Rubus laciniatus* (evergreen blackberry). We compared models for each species created using logistic regression, GARP and ENFA. For three of the species we compared these models with those based on the distributions of species in other invaded areas and in species' native ranges.
- Our goal was to predict the potential distributions of these five species on the Olympic Peninsula and to assess the potential of HDMs to aid in designing monitoring protocols for Olympic National Park (ONP). We addressed four questions: (1) How well do these models perform and how sensitive are they to clumped sample distributions and to the non-equilibrium nature of invasive species' distributions? (2) Do GARP and ENFA models perform better than logistic regression when applied to invasive species? (3) How do predictions from statistical modeling techniques compare with models based on data from the native range or other invaded areas? (4) Can the results of multiple modeling techniques be combined to better inform the development of monitoring protocols?
- We combined data on the current distributions (presence/absence) of the five species on the Olympic Peninsula from 13 data sources. These sources included exotic plant surveys, general vegetation surveys, vegetation sampling conducted for other purposes, herbarium specimens and personal observations of the species. In all, >4000 data points from across the Olympic Peninsula were included.

- From a larger set of potential habitat variables, we selected 12 as predictors of species' distributions. These included climate (e.g., number of frost days, annual precipitation), topography (e.g., slope, heat load), vegetation cover (conifers, total vegetation) and distance from the nearest wetland or stream.
- For each species and model type we created five replicate models using a portion of the available data. We then determined the accuracy of each model when used to predict data not used in developing the model. We also determined the variability of model results and the amount of suitable habitat predicted by each model.
- For *Geranium*, *Hedera* and *Ilex* we compared model results with models based on data from other invaded areas and from the species' native ranges.
- We combined results from all models to create maps for each species showing risk of invasion throughout the Olympic Peninsula.
- Model accuracy was good for *Geranium*, *Hedera* and *Ilex* but moderate to poor for *Rubus* and *Cirsium*.
- The amount and pattern of predicted suitable habitat varied considerably and was affected by clustering of the available data and by the continued spread of these species. In general, areas close to current infestations were predicted to be more suitable.
- Logistic regression was more accurate and less variable than GARP and ENFA for all species. There was little evidence that GARP and ENFA produced better models than did logistic regression.
- Although models had high accuracy for some species, they were clearly affected by dispersal patterns and clumping in the dataset. In addition, accuracy measures do not assess the ability of models to accurately predict future distributions, but only current distributions. Thus individual model results should be considered with care.
- Models based on data from other invaded areas and from species' native ranges predicted that much larger areas of the Olympic Peninsula were suitable. These models may more accurately represent the potential distributions of species.
- Combining models allowed us to develop maps indicating current, as well as long-term risks of invasion, while reducing the impact of error or bias in individual models.
- Maps of invasion risk can be used to develop monitoring protocols for Olympic National Park. Monitoring in turn can be used to refine and improve model predictions.

Introduction

Invasions of natural communities by exotic species are increasingly recognized as having major ecological and economic impacts (Vitousek et al. 1997, Levine et al. 2003) and as a major factor leading to the loss of biodiversity (Wilcove et al. 1998). Nonnative invaders directly compete with native species for resources and can alter disturbance regimes and nutrient cycles (Mack et al. 2000).

In theory, nature reserves such as national parks should be less susceptible to invasion by exotics than other areas both because of reduced opportunities for dispersal and generally lower levels of human-caused disturbance (Lonsdale 1999). Nevertheless, nature reserves throughout the world have been invaded by exotics (Usher 1988). In the USA, at least 1.1 million ha of land administered by the National Park Service is infested by exotic plants (Drees 2004).

Control of exotic species invasions is much easier in the early stages of invasion (National Invasive Species Council 2001) but early detection of invasions requires extensive monitoring. The National Park Service has recognized the importance of early detection of invaders and is developing plans for monitoring and early detection (Benjamin and Hiebert 2004). The North Coast and Cascade Network of the National Park Service is planning to monitor known populations and potential habitat locations for those invasive, exotic plant species considered to be most likely to damage natural ecosystems as part of the NPS Vital Signs Monitoring Program.

However, monitoring for invasive species is time consuming and costly (Rew et al. 2006). Because resources are limited it is necessary to prioritize and focus monitoring on areas of greatest concern (Benjamin and Hiebert 2004). One approach is to use habitat distribution models (HDMs) to determine where species are likely to invade and to focus monitoring on those areas. Habitat distribution models are widespread in conservation biology (Guisan and Zimmermann 2000, Scott et al. 2002) and are increasingly used to predict the potential distributions of invasive species (Rodríguez et al. 2007). If these models can accurately predict the potential distributions of invasive species, they can greatly benefit monitoring efforts.

There are several challenges, however, to applying HDMs to invasive species. First, these models assume that the species are in equilibrium with their environments (Guisan and Zimmermann 2000). This is clearly not the case for most invasive species, which are still spreading. Thus, models based on the current distribution of a species may under-predict the potential habitat for that species (Loo et al. 2007). Correlations between patterns of dispersal and patterns in habitat variables may also lead to biased results. Second, datasets on invasive species are often based on highly clumped sampling, focused along roads or other areas of high invasion. This clumped distribution can bias results toward areas of high sampling density. Third, and perhaps most importantly, we cannot determine how well the models are predicting the potential distributions of species that are spreading; we can only determine how well they predict current distributions. Thus, current model evaluation techniques are insufficient for evaluating model performance. Most studies that use HDMs for invasive species acknowledge these limitations, but it is unclear how important they are in shaping model predictions.

The most common HDM is logistic regression, which was used in 79% of studies reviewed by Manel et al. (2001). However, logistic regression requires a number of assumptions, leading to recent proliferation of modeling techniques that attempt to overcome these limitations (e.g., Guisan and Zimmermann 2000, Elith et al. 2006). In particular, modeling techniques that only require information on locations where species are present, have been suggested as better alternatives than logistic regression (e.g., Hirzel et al. 2001, Peterson 2003). With invasive species, absence may reflect either unsuitable habitat or suitable

habitat that has not yet been colonized, thus presence-only methods hold promise for modeling potential distributions.

Another approach that minimizes the effects of the aforementioned challenges is to develop a model based on a species' distribution in its native range, then apply the model to the invaded range (e.g., Peterson 2003, Thuiller et al. 2005). Models can also be based on distributions in another invaded area or on habitat requirements as reported in the literature (Jones and Reichard, in review). These approaches, however, have their own problems. In particular, constraints on species distributions may not be the same in the invaded range as in the native range leading to over- or under-predicting the potential distribution in the invaded range (Beerling et al. 1995, MacIsaac et al. 2000, Fitzpatrick et al. 2007, Loo et al. 2007).

Although each approach has its limitations, combining different approaches and modeling techniques may provide valuable insights into invasion patterns and the mechanisms controlling invasions (Fitzpatrick et al. 2007). Combining modeling approaches may also be useful for developing monitoring programs. Areas where models agree could be considered the highest priority for monitoring. This approach to developing monitoring protocol has not been reported in the literature.

In this project we modeled the potential distributions of five invasive species on the Olympic Peninsula, Washington: *Cirsium arvense* (Canada thistle), *Geranium robertianum* (herb Robert), *Hedera helix* (English ivy), *Ilex aquifolium* (English holly) and *Rubus laciniatus* (evergreen blackberry). These species are of particular concern in Olympic National Park. We developed models using three methods based on the current distributions of the species on the Peninsula. These methods were logistic regression and two presence-only techniques: Genetic Algorithm for Rule-set Prediction (GARP, Stockwell and Peters 1999) and Ecological Niche Factor Analysis (ENFA, Hirzel et al. 2002a). For three of the species, we compared these models to models developed based on information in the literature and models based on species' distributions in another invaded area. The overall goal of this project was to predict the potential distributions of these high-priority species on the Olympic Peninsula to inform development of monitoring protocol for Olympic National Park (ONP). To achieve this goal we addressed four questions: (1) How well do these models perform and how sensitive are they to clumped sample distributions and use with spreading species? (2) Do presence-only models perform better than logistic regression when applied to invasive species? (3) How do predictions from statistical modeling techniques compare with models based on data from the native range or other invaded areas? (4) Can multiple models be combined to better inform the development of monitoring protocols?

Methods

Study Area

The Olympic Peninsula is located in northwest Washington and covers an area

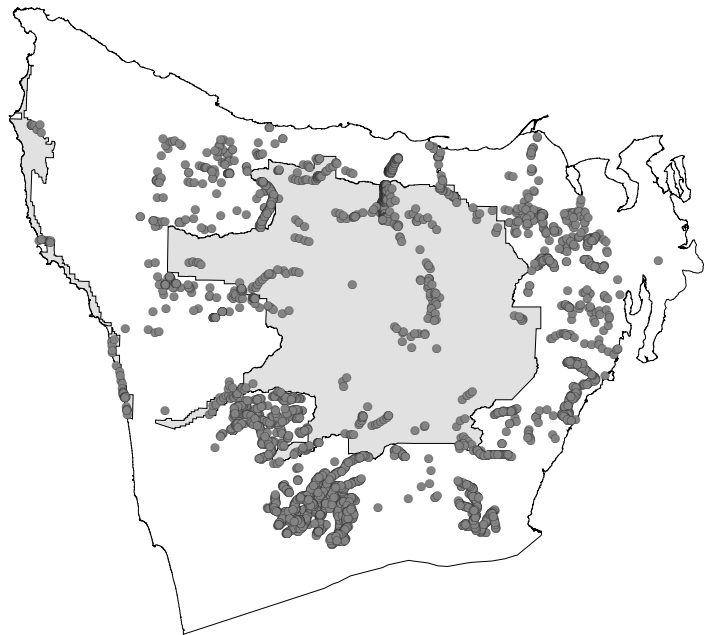


Figure 1. The Olympic Peninsula, Washington. Shading indicates the location of Olympic National Park. Filled circles indicate the locations of species data points.

of ~12,500 km² (Fig. 1). Olympic National Park covers almost 3700 km² in the center of the Peninsula and along the west coast. ONP is surrounded by Olympic National Forest and state and privately owned lands. It contains large roadless areas with undisturbed old-growth forest, whereas surrounding areas have a more extensive road network and many areas have been subjected to clearcut logging over the past century.

Elevation ranges from sea level to 2429 m, with the high elevations concentrated in the Olympic Mountains in the center of the Peninsula. The mountains create a strong gradient in annual precipitation with the driest areas (63 cm per year) in the northeastern part of the Peninsula and the wettest areas (709 cm per year) on the western slopes of the mountains. Precipitation during the growing season is low, with 6-12% of the total falling between July and September. Mean January temperatures range from 5° C at sea level to -9.5° C at high elevations. Mean July temperatures are mild, ranging from 11 to 18° C. Climate information is from DAYMET (www.daymet.org).

Study Species

Models were created for five species of concern to staff at ONP. They represent a range of growth forms and life histories and are at different stages of invasion on the Peninsula.

Cirsium arvense (Canada thistle)

Cirsium arvense is a perennial herb that can grow to 150 cm tall (Weber 2003). It can reproduce vegetatively from root fragments and from lateral roots that can spread up to 6 m per year (Moore 1975). Its seeds can be dispersed long distances by wind and can survive in the seed bank for up to 20 years, although most germinate in the first year (Donald 1994). *Cirsium* is most commonly found in open areas, such as grasslands, riparian areas and disturbed habitats, and does not colonize dense forest (Townsend and Groom 2006a). It does not grow in areas where the mean January temperature is less than -18° C and is limited in the southern USA by summer high temperatures (Moore 1975). It can tolerate a wide range of soil conditions but is uncommon in wet soils (Moore 1975).

Cirsium probably originated in southeastern Europe, but it is now common throughout Europe and has spread to northern Africa, South Africa, South America, New Zealand, Australia, and parts of Asia (Moore 1975, Weber 2003). It was first introduced in North America in the 1600s and was considered a problem weed in the eastern USA by 1795 (Morishita 1999). It first invaded the west coast of North America in the mid 1800s (Townsend and Groom 2006a). By 1936 it was considered to be “one of the most noxious and widespread weeds in cultivated ground on the Olympic Peninsula” (Jones 1936).

Cirsium can form very dense stands and crowd out native vegetation (Townsend and Groom 2006a). In addition, it has limited palatability and can therefore dominate grazed areas. There is some evidence that *Cirsium* is also allelopathic (Donald 1994).

Geranium robertianum (herb Robert)

Geranium robertianum is a biennial or short-lived perennial herb. It reproduces only by seed. Seeds are ejected from the fruit and can be dispersed by animals or water (Grime et al. 1988, Tofts 2004) and remain in the soil seed bank for up to five years (Bertin 2001). *Geranium* grows in a wide variety of habitats, from open, rocky areas to closed forest and can tolerate a wide range of climatic and soil conditions (Tofts 2004).

Geranium is native to Europe and has been introduced to many other areas of the world including eastern Asia, Chile, New Zealand, Australia and the USA (Tofts 2004). It is present but fairly rare in the

northeastern United States, where some consider it to be native (Fernald 1950, Seymour 1969). It is an aggressive invader in the Pacific Northwest, where it was first observed in 1906. *Geranium* was first recorded on the Olympic Peninsula near Heart O' the Hills in the mid 1970s and has gradually spread up the Hurricane Ridge road and to the Elwha Valley (Tisch 1992). Populations of *Geranium* in western Washington have spread rapidly in the past two decades (Barndt 2006).

Impacts of *Geranium* on native vegetation have not been well studied, but it can rapidly spread and cover 50-100% of the ground over large areas (Tisch 1992). It can outcompete native forest herbs because it grows rapidly and can photosynthesize over the winter (Tisch 1992).

Hedera helix (English ivy)

Hedera helix is an evergreen vine that grows in the open and in deep shade (Metcalf 2005). Individuals can live for up to 400 years (Rose 1996) and reproduce both sexually and asexually (Grime et al. 1988). Berries can be dispersed long distances by birds (Van Ruremonde and Kalkhoven 1991). *Hedera* is tolerant of freezing but does not reproduce when mean January temperatures are less than -2.5°C (Iverson 1944). It is drought tolerant (Laskurain et al. 2004) but only becomes invasive when there is sufficient moisture (Muys 2001).

The native range of *Hedera* extends from northern Europe to northern Africa and east to the Ukraine (Metcalf 2005). It has invaded Australia, Brazil, and New Zealand, as well as Hawaii, the Pacific Northwest, Mid-Atlantic, and southern states of the USA (Metcalf 2005). It has been present in the Pacific Northwest since at least 1892 (Murai 1999). It was not found in a botanical survey of the Olympic Peninsula in 1936 (Jones 1936) but was present by at least 1979 (Buckingham and Tisch 1979).

Hedera can form dense mats in the forest understory, reducing growth rates and density of understory herbs and tree seedlings (Thomas 1980, Dlugosch 2005). It also produces allelopathic compounds (Blood 2001) and can negatively affect native fauna by reducing availability of native forage species (Freshwater 1991). Although *Hedera* often covers trees, there is limited evidence of resulting tree mortality (Thomas 1980, Metcalf 2005), but it has been shown to reduce tree root growth more than other ground covers (Shoup and Whitcomb 1981).

Ilex aquifolium (English holly)

Ilex aquifolium is an evergreen tree or large shrub that can live for as many as 250 years (Peterken and Lloyd 1967). It reproduces mainly by seed. Berries can be dispersed long distances by birds (Peterken and Lloyd 1967) and survive for 4-5 years in the seed bank (Arrieta and Suarez 2004). *Ilex* tolerates deep shade but also grows in open areas (Peterken 1966). July temperatures must average at least 12°C and January temperatures at least -1°C for *Ilex* to survive and reproduce (Iverson 1944). It is not drought tolerant (Prentice and Helmisaari 1991).

The native range of *Ilex* includes western Europe and mountainous areas of northern Africa and southwest Asia (Peterken and Lloyd 1967). It is widely distributed as an ornamental and has become invasive in New Zealand, Australia and northwestern USA (Weber 2003). It has been grown in the Pacific Northwest as a landscape plant and for Christmas decorations since the late 1800s (Wieman 1961). It was not found in a botanical survey of the Olympic Peninsula in 1936 (Jones 1936) but was present by at least 1979 (Buckingham and Tisch 1979).

Little is known about the impacts of *Ilex* on native communities. Because it is evergreen, it casts deep shade year-round which may reduce plant regeneration beneath its canopy (Peterken 1966).

Rubus laciniatus (evergreen blackberry)

Rubus laciniatus is an evergreen vine that can grow to 3 m in height (Tirmenstein 1989). It sprouts vigorously from roots, stems and stem tips and can spread rapidly (Tirmenstein 1989). It also produces large numbers of seeds that can be dispersed long distances by animals and persist in the soil for many years (Murphy 2006). *Rubus* is most commonly found in disturbed areas along streambanks (Murphy 2006) and often colonizes forested areas after clearcutting, thinning or fire (Klinka et al. 1985). It often grows on barren unfertile soils but is not highly drought tolerant (Tirmenstein 1989).

Rubus is native to Europe and has invaded much of the northeastern and northwestern USA (Tirmenstein 1989). It was present on the Olympic Peninsula prior to 1900 (Jones 1936) and by 1906 was common in woodlands along the Washington coast (Piper 1906).

Rubus forms dense thickets and can shade out native vegetation. It can also provide food and shelter for a number of introduced animal species (Murphy 2006).

Species Data

To obtain information on current distributions of the five species, we searched for existing datasets that contained presence and absence data for these species on the Olympic Peninsula. For a data point indicating presence to be included it must have been collected no earlier than 1975 from within a plot of less than 2500 m². Plotless records of species presence (e.g. herbarium specimens or personal observations) could be included if their location was known within 50 m. Stricter criteria were used to verify absence: an absence point must have been collected after 1990, from within a plot of between 100 and 2500 m², and for which all species encountered were recorded. These criteria were used to reduce the chance that a species was present but not recorded.

Thirteen datasets contained points that met these criteria (Table 1, Appendix A). These were obtained from a variety of sources, including exotic species surveys, general vegetation inventories, localized vegetation sampling conducted for various other purposes, and locations where species were observed and recorded by one of the authors (C. Jones) or exotic plant management personnel. Almost all of the data points were collected since 1995. Together they represented a total of 4142 locations. Of these, 655 could not be used to verify absence. Because of this, the set of data points used for analysis varied slightly among species. Locations are distributed across the Peninsula although there is considerable clustering associated with the data sources (Fig. 1).

There is considerable autocorrelation in both the species data and environmental variables at short distances (Appendix B). To reduce this autocorrelation we set a minimum distance between plots at 100 m. Separately for each species we started by removing a random point from those with the greatest number of near neighbors (<100 m), then recalculated the number of near neighbors for the remaining points. This process was repeated until no points were within 100 m. To maintain as many presence points as possible we did this separately for presence and absence points.

The individual species have between 25 and 559 occurrences, with frequencies that range from 0.88 to 19.18% in the final datasets (Table 2). Distributions for individual species are shown in Appendix C.

Table 1. Summary of datasets used.

| Dataset | Contains plots in ONP? | Contains absence data? | # of sample points | Plot area (m ²) | Date sampled |
|---|------------------------|------------------------|--------------------|-----------------------------|--------------|
| Olympic NP Exotics Inventory ¹ | Y | Y | 256 | 100 | 2001 |
| Elwha Exotics Survey ¹ | Y | N | 580 | n/a | 2001 |
| Alpine Plant Community Survey ² | Y | Y | 13 | ~100 | 1990, 1999 |
| Exotic Plant Management Team ¹ | Y | N | 65 | n/a | 2002-2004 |
| LTEM Elwha plots ¹ | Y | Y | 61 | 400-625 | 1999-2000 |
| Collette DeFerrari dataset ³ | Y | Y | 95 | 50-100 | 1992 |
| Ann Lezberg dataset ⁴ | Y | Y | 100 | 100 | 1996-1997 |
| Olympic NF Exotics Survey ¹ | N | Y | 2881 | 400-2500 | 1995-2003 |
| Olympic Habitat Development Study ⁵ | N | Y | 60 | 100 | 2000-2004 |
| Olympic NF Ecology Plots ⁶ | N | Y | 21 | 250-1600 | 1980-2003 |
| Vegetation and Landform Database Development Study ¹ | Y | N | 1 | ~100 | 1997 |
| Olympic NP Herbarium Records ¹ | N | N | 5 | n/a | 1977-1983 |
| Personal Observations ¹ | Y | N | 4 | n/a | 2005 |

¹Unpublished data

²See appendix A1 and A2 in Houston et al. (1994)

³DeFerrari and Naiman (1994)

⁴Lezberg et al. (2001)

⁵See Carey and Harrington (2001) and Reutebuch et al. (2004)

⁶Henderson et. al (1989)

Table 2. Number of locations where target species are present (from datasets in Table 1).

| Species | Presences | Absences | Percent frequency |
|-----------------------------|-----------|----------|-------------------|
| <i>Cirsium arvense</i> | 549 | 2485 | 18.09 |
| <i>Geranium robertianum</i> | 219 | 2784 | 7.29 |
| <i>Hedera helix</i> | 25 | 2823 | 0.88 |
| <i>Ilex aquifolium</i> | 45 | 2809 | 1.58 |
| <i>Rubus laciniatus</i> | 559 | 2356 | 19.18 |

Habitat Variables

We created or adapted GIS layers for 23 climate, topographic, vegetation and soil variables (Table 3). Climate variables were available from DAYMET (www.daymet.org, Thornton et al. 1997). Vegetation variables were from the Interagency Vegetation Mapping Project (IVMP). Topographic variables were based on the USGS digital elevation model. Habitat layers are described in more detail in Appendix D. We eliminated four of these because they were not suitable for our study. The two soil variables (maximum depth to water table, maximum pH) were at too coarse a scale and therefore were not used. Tree quadratic mean diameter (QMD) was based on satellite imagery, but there were large gaps in these data, thus QMD was excluded. Finally, land cover — a categorical variable — did not fit well with the modeling structure because it would have required use of 13 dummy variables, thus it was also dropped.

Table 3. Habitat variables considered for modeling.

| Used in models | Considered but not used |
|--|--|
| Climate Variables¹ | Climate Variables¹ |
| Number of frost days | Mean temperature of the coldest month |
| Annual precipitation | Mean temperature of the warmest month |
| Frequency of precipitation | Growing degree days |
| Humidity ² | Temperature seasonality ⁷ |
| Solar radiation ² | Growing season precipitation ⁸ |
| Topographic Variables | Precipitation in driest month |
| Slope | Topographic Variable |
| Potential radiation ³ | Elevation |
| Heat load ³ | Soil Variables⁹ |
| Topographic moisture index ⁴ | Maximum depth to water table |
| Distance to nearest water⁵ | Maximum pH |
| Vegetation Variables⁶ | Vegetation and Land Cover Variables |
| Conifer cover | Quadratic mean diameter ⁶ |
| Total vegetation cover | Land cover ¹⁰ |

¹From DAYMET (Thornton et al. 1997)

²Based on precipitation and temperature (Thornton and Running 1999, Thornton et al. 2000)

³Based on latitude, slope and aspect (McCune and Keon 2002)

⁴Based on catchment area and slope (Moore et al. 1993)

⁵Log₁₀ transformed distance to the nearest stream or wetland

⁶From Interagency Vegetation Mapping Project

⁷Difference between mean temperature of coldest and warmest month

⁸July – September precipitation

⁹Maximum values for each map unit in STATSGO

(<http://www.ncgc.nrcs.usda.gov/branch/ssb/products/statsgo/>)

¹⁰Based on the National Land Cover Data of 1992

To assess multicollinearity, we calculated Spearman rank correlations among values of the remaining 19 variables at all locations for which we had species data. Two groups of variables were highly correlated ($r > 0.8$). The group of precipitation variables (annual precipitation, growing season precipitation, and precipitation during the driest month) were highly correlated ($r > 0.97$), thus only annual precipitation was used. Six variables related to temperature were strongly correlated (number of frost days, mean temperature of the warmest and coldest months, elevation, growing degree days and temperature seasonality). To reduce the number of variables, we used hierarchical partitioning, which measures the average model improvement that a variable provides when added to models with all possible combinations of the other variables (MacNally 2002). Number of frost days provided the greatest explanatory power and was thus included in the models. Removing correlated variables resulted in 12 habitat variables used in all models (Table 3).

Habitat layers were clipped to cover the same area (there were holes in the vegetation layers due to topographic shadow in the satellite imagery). All layers had 25 m resolution except for climate variables, which had 1 km resolution. Climate layers were converted to a 25 m grid size by assigning each grid cell the value of the corresponding 1 km grid cell.

Modeling Methods

We compared three different modeling techniques for each species: logistic regression, GARP, and ENFA. Logistic regression is the most common technique for habitat distribution models (Manel et al.

2001), whereas GARP and ENFA are more recent, presence-only techniques that may work better for invasive species that are still spreading (Hirzel et al. 2001, Peterson 2003).

Logistic Regression

Logistic regression is a form of generalized linear modeling (GLM). GLMs are an extension of linear regression that allow the response variable to have a distribution other than the normal distribution. The model consists of a linear predictor ($\mu = \alpha + \beta_1x_1 + \beta_2x_2 + \dots + \beta_ix_i + \varepsilon$) which is linked to the response, Y, with a “link function” (Nicholls 1989). For logistic regression (suitable for presence-absence data), the link function is $Y = \log [\mu/(1-\mu)]$. This produces a binomial rather than a normal distribution.

There is a large body of statistical theory supporting logistic regression with well-developed methods for determining the significance of parameters, as well as assessing goodness of fit. Logistic regression has been used widely for habitat distribution models in general (Manel et al. 2001) and for models of invasive species in particular (Buchan and Padilla 2000, McNab and Loftis 2002). However, it requires assumptions about the shape of species responses to each variable. Interactions among independent variables are not modeled unless they are explicitly included.

Because of the number of independent variables, we did not include quadratic or interaction terms in the models. For species we created models with all 12 independent variables using S-PLUS 6.2 (Insightful 2003). Because logistic regression is highly sensitive to the frequency of presence in a dataset, we down-weighted the importance of each absence so that the total weight of species absence points was the same as for species presence points. This procedure has been shown to improve model results (Maggini et al. 2006). We then removed variables in a stepwise fashion using the *step* function in S-plus. This procedure removed variables that reduced Akaike’s Information Criterion (AIC), a measure that combines model simplicity and goodness of fit (Crawley 2002). The resulting regression equations (reduced models) were then used to predict habitat suitability for each grid square in the study area. Suitability values ranged from 0 to 1 but were converted to a scale from 0 to 100 for comparability to the results of the other two methods.

Genetic Algorithm for Rule-set Prediction (GARP)

GARP is an artificial-intelligence based modeling system. The main structure of the model is a rule-based genetic algorithm (Stockwell and Peters 1999). The algorithm builds a set of rules for determining presence or absence, then modifies them slightly by mutation or crossing over to see if fit is improved. The rules include habitat envelopes, atomic rules and logistic regression and are different for each grid cell depending on which performs best. The model then predicts presence or absence for each grid cell. In areas where there is insufficient data the model does not give a prediction, but in the output these areas are not distinguishable from absences. Because rules are based on random mutations, each model run produces somewhat different results.

GARP is increasingly being used to model the potential distribution of invasive species (e.g., Underwood et al. 2004, Anderson et al. 2006). Because it can model interactions among variables, includes logistic regression and can accommodate data irregularities that occur with invasive species (Stockwell and Peters 1999), this technique may perform better than logistic regression.

For this study we used DesktopGARP 1.1.6. (<http://nhm.ku.edu/desktopgarp/index.html>). We used all rule types and all 12 of the habitat variables. For each species we created 100 GARP models, and then created an index for each grid cell based on the number of model runs that predicted presence in the cell. Thus each cell had a value between 0 and 100, with higher numbers indicating greater suitability.

Ecological Niche Factor Analysis

Ecological Niche Factor Analysis (ENFA) is a modeling technique akin to Principal Components Analysis that converts the environmental data (which may be correlated) into factors that are uncorrelated (Hirzel et al. 2002a). The first factor represents species marginality (i.e., the difference between the species' realized niche and the mean habitat conditions in the landscape). The remaining factors measure the specialization of the species (the breadth of the species' realized niche compared to the range of habitat conditions in the landscape). Each cell in the landscape is then given a score for each factor based on how far it is from the median of locations where the species is present. The scores are combined for all significant factors, and then standardized to create a habitat suitability index that ranges from 0 to 100. Like GARP, this technique does not require absence data. A test with virtual data showed that ENFA performed better than logistic regression for a species that is expanding its range (Hirzel et al. 2001). Therefore, this technique may work well for invasive species.

We created ENFA models for each species using Biomapper 3.0 (Hirzel et al. 2002b). We used the median distance measure and a broken-stick model to determine the number of factors to include (Hirzel et al. 2002a).

Model Building and Assessment

To allow for model validation, presences and absences were partitioned into five equal subsets for each species. Five models were created for each species using 80% of the data with a different set of 20% reserved to determine model accuracy in each case. We thus had five replicate models for each species and modeling technique.

We compared model types and species in three areas: (1) how accurate models were at predicting data not used in their creation, (2) amount of suitable habitat predicted, and (3) how variable the five replicate models were. We used several metrics to address each of these areas (see below). Differences in these measures among modeling techniques were compared using ANOVA followed by Bonferroni *post-hoc* comparisons.

1. Model accuracy

Traditionally, most measures of model accuracy are based on setting a threshold of suitability and converting model results to predictions of presence or absence in each grid cell based on the threshold (Fielding and Bell 1997). A 2 x 2 matrix is then created with numbers of true and false presences and absences. Many different metrics have been proposed to assess accuracy based on this matrix. We considered the following:

Sensitivity – proportion of presences in the assessment data that are correctly predicted by the model (a measure of how well the model captures where the species occurs).

Specificity – proportion of absences in the assessment data that are correctly predicted by the model (a measure of how well the model captures where the species does not occur).

Kappa – proportion of specific agreement, which includes the rate of both false positive and false negative errors (Fielding and Bell 1997). Kappa values <0.2 are poor, 0.2-0.6 are fair to moderate, and >0.6 are good (Landis and Koch 1977). Kappa is a measure of overall accuracy, but is sensitive to the frequency of presences in the dataset.

True Skill Statistic (TSS) – This measure is calculated as Sensitivity + Specificity – 1. It is equivalent to Kappa when frequency of presences and absences are equal, but is not sensitive to frequency (Allouche et al. 2006).

AUC – The measures above depend on the selected threshold and do not use the full amount of information available in the models. Therefore we also used Receiver Operating Characteristic (ROC) plots, which do not require a threshold, to assess accuracy. These plots are created by plotting Sensitivity against 1 – Specificity at all possible threshold levels (Fielding and Bell 1997). The area under this curve (AUC) is then used as a measure of model accuracy. Values of AUC generally range from 0.5 (equivalent to that due to chance), to 1 (perfect performance). Values >0.9 indicate good performance, 0.7-0.9 are moderate, and <0.7 are poor (Pearce and Ferrier 2000)

Because we were more interested in predictions of presence and because the frequency of most species was low, we used AUC, TSS and Sensitivity. We determined the threshold of suitability by calculating TSS at all possible threshold values and using the threshold that maximized TSS (e.g., Robertson et al. 2004). We used the maximum value of TSS and the value of Sensitivity at this threshold.

2. Amount of suitable habitat

Two metrics were used to evaluate amount of suitable habitat:

Percent suitable habitat – the percent of total area predicted to be suitable based on the threshold that maximized TSS. This measure is influenced by the threshold chosen.

Mean suitability – the mean of the suitability values in each grid cell. This measure is not affected by the threshold.

3. Model variability

Three metrics were used to examine model variability:

Variation in accuracy measures and amount of suitable habitat – for each model technique we calculated the standard deviation (SD) of each accuracy variable for the five model replicates.

Overlap – This is a measure of the overlap in suitable habitat predicted by each of the five model replicates. This is calculated as the percentage of grid cells that are suitable in any model that were suitable in all five models. This measure is affected by the variability in thresholds among the five replicates.

Mean SD – This is a measure of variability that is not affected by the choice of threshold. We calculated the standard deviation of suitability values for the five replicates for each grid cell. We then calculated the mean SD for the entire study area.

Comparison with Alternative Modeling Approaches

We compared model results with those from alternative modeling approaches for three of the species (*Geranium robertianum*, *Hedera helix* and *Ilex aquifolium*). The latter were based on information from other invaded areas or the species' native ranges and were used to assess how much model results are influenced by lack of equilibrium with the environment.

We used literature-based and climate envelope models for each of the three species (Jones and Reichard, in review). We developed literature-based models (hereafter ‘native-range’ models) using information on the climate tolerances of the species from the literature and climate and range maps from the species’ native ranges. Climate envelope models are created from extreme values (at locations where a species occurs) for climatic variables that are thought to shape a species’ distribution (Busby 1991). Typically, extreme values are removed to reduce the effects of outliers and to produce models of the species core habitat. Climate envelope models (hereafter ‘invaded-range’ models) were created for each of the three species using known location across the contiguous USA and removing the 5% most extreme values (Jones and Reichard, in review).

The ‘native-range’ and ‘invaded-range’ models created for the contiguous USA were applied to the Olympic Peninsula by overlaying them on the study area. Any 25 m grid cell that corresponded to suitable habitat in the ‘invaded-range’ or ‘native-range’ model was identified as suitable based on that model. However, the ‘native-range’ model for *Geranium* was problematic (Jones and Reichard, in review) and was not used. These models used only climate variables that were available for the entire contiguous USA (Table 4), thus they provide a coarse-scale representation of the potential range of the species on the Olympic Peninsula.

Summarizing Model Results for Monitoring

Each of the modeling methods has limitations, thus we developed an index that used information from all of the methods in developing recommendations for monitoring. For each of the five replicates (runs) of each statistical model, we set the threshold at the value that maximized TSS. For each grid cell we then calculated the number of model runs that predicted suitability (0-15). To simplify interpretation of this index we classified cells with values of 0-5 as “low risk of invasion,” values of 6-10 as “moderate risk of invasion,” and values of >10 as “high risk of invasion.” Thus, moderate risk areas must be considered suitable by models of at least two types and high risk areas by at least one model run of each type. For the three species for which ‘native-range’ and ‘invaded-range’ models were available, we also created a risk classification based on these models. We classified a cell as high risk for invasion if it was suitable in both models, as moderate risk if suitable in one model, and as low risk if suitable in neither model.

Role of Dispersal Corridors

Because current distributions may be affected by dispersal patterns, we examined species’ distributions with respect to potential dispersal corridors (roads, trails and streams). We used Chi-square tests to compare the proportions of presences that were near these features (within 25 m) with proportions of absences near the same features.

Table 4. Climate variables used to create ‘native-range’ and ‘invaded-range’ models.

| Variables | <i>Geranium</i> | | <i>Hedera</i> | | <i>Ilex</i> | |
|------------------------------|-----------------|-----------------------------|-----------------|-----------------------------|-----------------|-----------------|
| | ‘Invaded-range’ | ‘Native-range’ ¹ | ‘Invaded-range’ | ‘Native-range’ ¹ | ‘invaded-range’ | ‘invaded-range’ |
| Max July temp (° C) | >20.4, <30.6 | | | | | |
| Min January temp (° C) | >-13.1 | | | | | |
| Annual precip. (cm) | >70.5 | >70 | >55.1 | >71 | >54.2 | |
| Mean July temp (° C) | | >13 | >15.7 | | | |
| Mean January temp (° C) | | >-2.5 | >-2.5 | >-1 | >1.1 | |
| GDD ² (base 0° C) | | | | >2571, <6359 | >3169, <4499 | |

¹Values for *Geranium* are from Tofts (2004); for *Hedera* are from Iverson (1944), Muyt (2001), Hultén and Fries (1986) and Steinhauser (1970); and for *Ilex* are based on Iverson (1944) and Prentice and Helmisaari (1991).

²Growing degree-days

Results

General Modeling Results

Statistical models predicted that between 9.2% (GARP for *Geranium*) and 40.1% (LR for *Rubus*) of the habitat was suitable for these invasive species (Fig. 2c). Models for *Geranium*, *Hedera*, and *Ilex* were highly accurate based on values of AUC (Fig. 3a) and TSS (Fig. 3b). In contrast, models for *Rubus* and *Cirsium* had poor to moderate accuracy. Variation in both accuracy (error bars in Fig. 3) and predicted suitability (Fig. 4) among model replicates was generally high for *Hedera* and *Ilex*, although it differed among modeling techniques. This is because there were few positive occurrences for these species, so models were developed based on only a few data points. *Cirsium* also had fairly high variation in predicted suitability (Fig. 4) even though it had among the highest number of positive occurrences.

For *Cirsium*, predictions of suitable habitat varied considerably among modeling techniques (Fig. 5); moreover, models had poor accuracy making it difficult to summarize the distribution of suitable habitat. For *Geranium*, models predicted suitable habitat at low elevations mostly concentrated in the northeast portion of the Peninsula near the original point of invasion (Fig. 6). Suitability maps for *Hedera* and *Ilex* were very similar to each other, with suitability in coastal areas on all sides of the Peninsula (Figs. 7 and 8). The distribution of suitable habitat for *Rubus* was different from the other species; concentrated on the west side of the Peninsula where it is most common (Fig. 9).

Comparison of Modeling Techniques

GARP predicted the highest mean suitability value for each species, followed by logistic regression (LR) and ENFA (Fig. 2a). However, this led to a much higher threshold to maximize accuracy for GARP than for either of the other techniques (Fig. 2b). This resulted in smaller areas predicted to be suitable for *Geranium* and *Rubus* by GARP than by the other two methods (Fig. 2c).

Logistic regression was more accurate than GARP or ENFA for most species based on AUC and TSS (Fig. 3a,b). In general, AUC and TSS showed similar patterns. However, ENFA was more accurate than GARP for *Rubus* and *Cirsium* when measured by AUC, but was less accurate for *Cirsium* and *Hedera* when measured by TSS.

Logistic regression had the lowest variability and the highest overlap among model replicates for almost all species (Fig. 4). ENFA and GARP were more variable than LR, but there were no consistent differences between them.

For most species, suitability maps were generally similar among modeling techniques (Figs. 5-9). However there were large differences for *Cirsium* and *Rubus*. In general ENFA was more selective and GARP had the largest areas of high suitability. In areas where data points were sparse, GARP did not make adequate predictions (and treated those areas as unsuitable), leading to an area on the west coast of the Peninsula with low suitability for all species (compared to generally higher suitability using the other techniques).

‘Invaded-range’ and ‘Native-range’ Models

‘Invaded-range’ and ‘native-range’ models predicted that at least 50% of the Peninsula is suitable for invasion (Table 5), much greater than that predicted by the statistical models. For all three species (*Geranium*, *Hedera* and *Ilex*), models predicted that low elevation areas around the perimeter of the Peninsula are suitable, but models differ in how far into the mountains suitable habitat extends (Figs. 6-8).

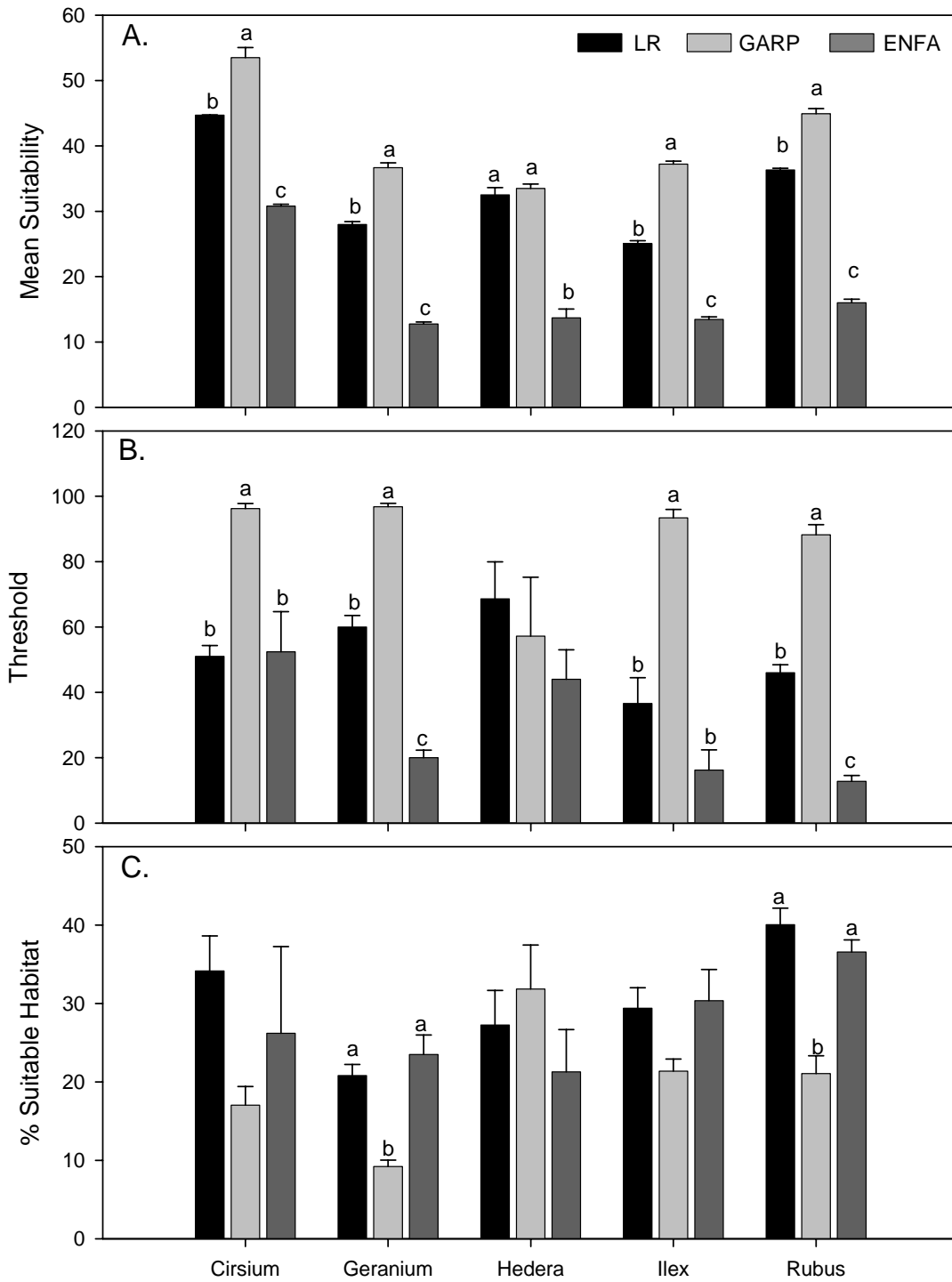


Figure 2. Suitability predictions for the three statistical modeling techniques. Mean suitability (A) is the average predicted suitability (on a scale of 0-100) for the entire study area. Threshold (B) is the cutoff for suitability for each model replicate that maximized model accuracy. Using this threshold we determined the percent of the study area (C) predicted to be suitable habitat. Error bars indicate 1 SE ($n=5$).

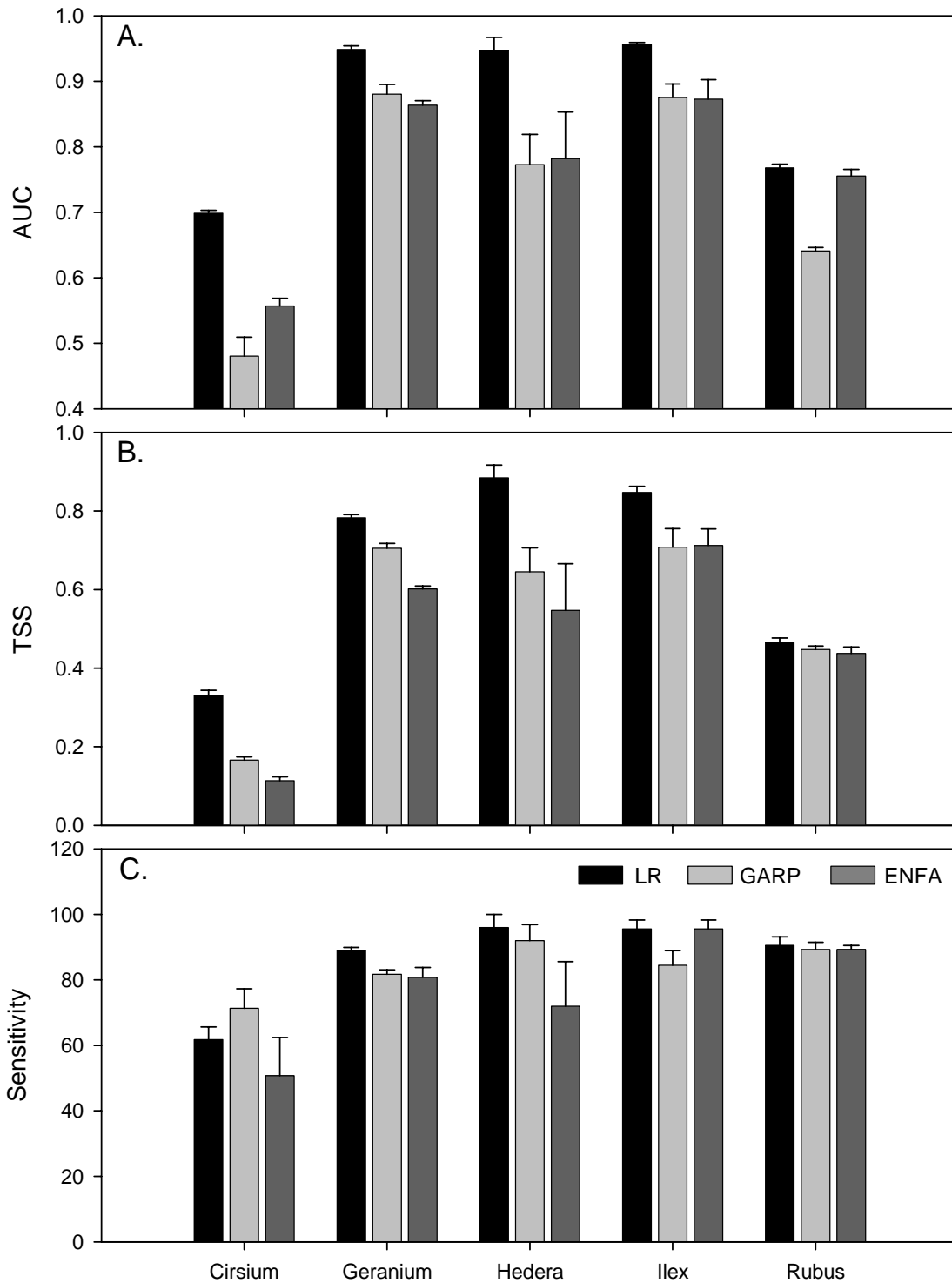


Figure 3. Model accuracy measures for the three statistical modeling techniques. AUC (A) is threshold independent: values >0.9 are good, 0.7-0.9 are moderate and <0.7 are poor (Pearce and Ferrier 2000). TSS (B) is threshold dependent: values >0.6 are good, 0.2-0.6 are moderate and <0.2 are poor (Landis and Koch 1977). Sensitivity (C) measures how accurately models predict occurrences. Error bars indicate 1 SE ($n=5$).

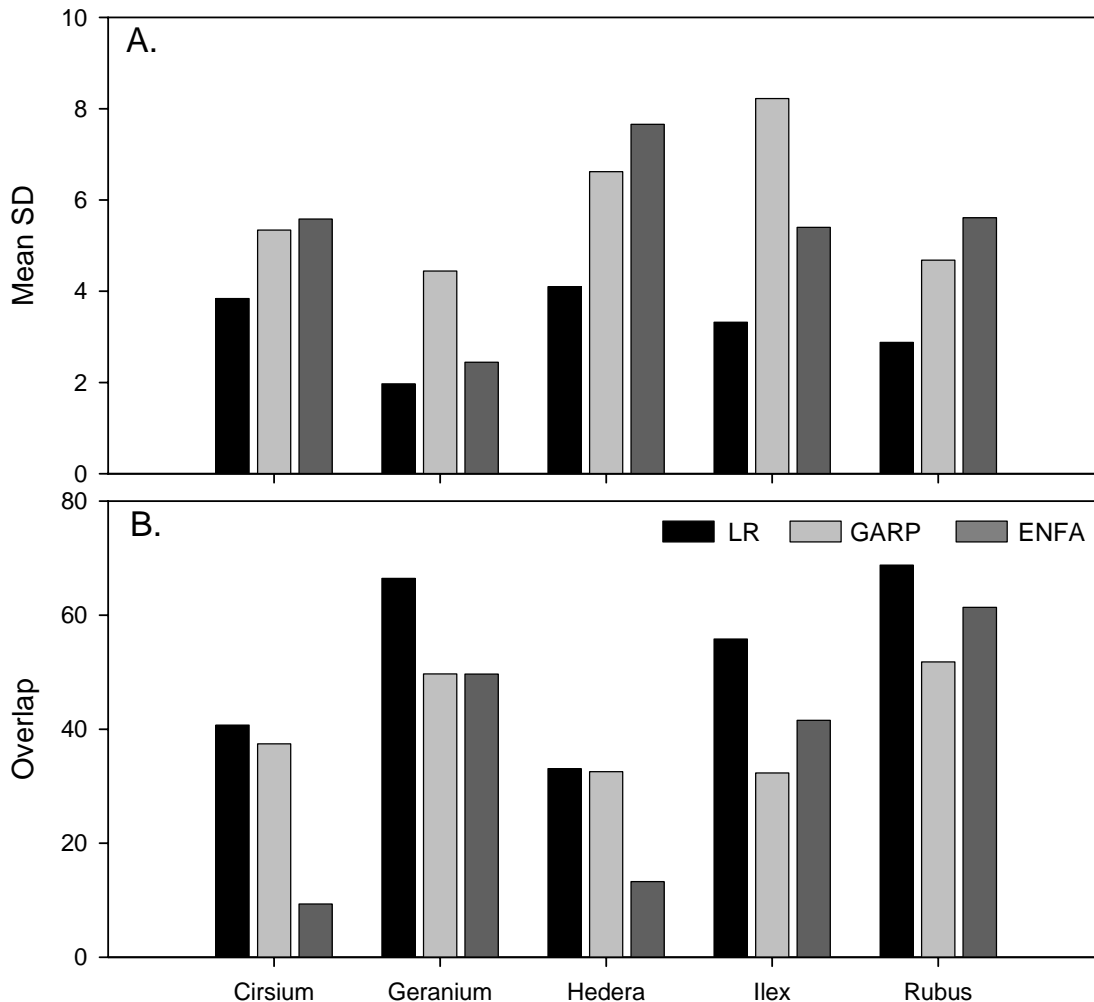


Figure 4. Variability in model results among replicates. Mean SD (A) is calculated as the standard deviation in suitability values of a grid cell averaged over all grid cells. Overlap (B) is the proportion of cells predicted to be suitable habitat by any of the replicates that are predicted to be suitable in all five replicates.

‘Native-range’ models predicted more suitable habitat than did ‘invaded-range’ models. Almost all of the locations where the species currently occur are predicted to be suitable by these models (Sensitivity >96%; Table 5). However, these models also predict suitability in many areas where the species do not currently occur, thus overall accuracy is low (TSS; Table 5). These results are consistent with the assumption that these models better reflect the long-term potential for species to spread to these habitats.

Using Multiple Models to Inform Monitoring

Because the three statistical modeling methods were influenced by the current distributions of species, and habitat close to current occurrences tended to be predicted as more suitable, the models generally reflected short-term risk of invasion. On the other hand, ‘invaded-range’ models were based on the extent to which the species have spread across the USA, and ‘native-range’ models reflected species’ distributions in their native ranges. Therefore, these models may more accurately portray the longer-term potential of the invasive species to spread throughout the Peninsula.

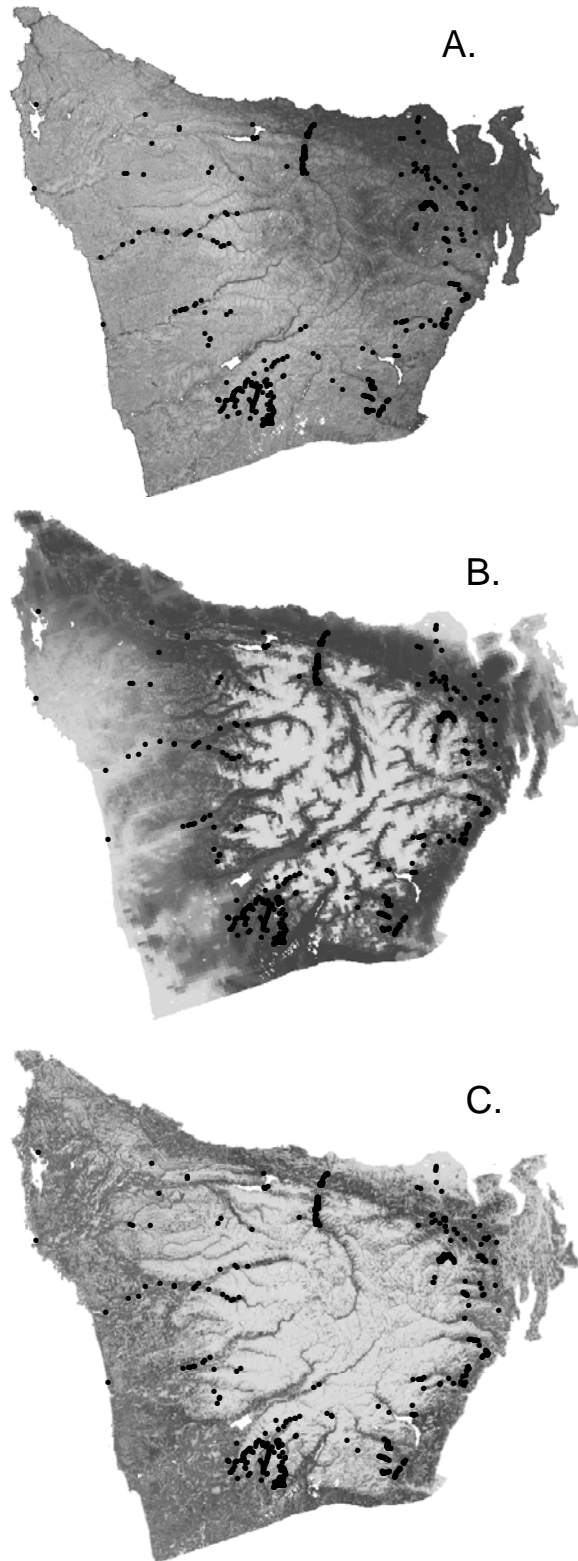


Figure 5. Habitat suitability maps for *Cirsium arvense* using logistic regression (A), GARP (B) and ENFA (C). Darker shading indicates greater suitability on a range of 0-100. Black circles indicate where *Cirsium* is present.

For *Geranium*, *Hedera*, and *Ilex* we combined the risk-classification from both types of models to create a classification that included both current and longer-term risk of invasion. If a grid cell was not suitable in any model it was given a minimal risk rating. Areas with low current risk were separated into those with moderate and high longer-term risk. However, areas that were considered moderate (index 6-10) or high (index 11-15) current risk were largely (>87%) classified as high long-term risk and thus were rated as moderate and high overall risk, respectively.

For the three species for which ‘native-range’ and/or ‘invaded-range’ models were available, combining models provided a straightforward risk-classification scheme that can inform monitoring on the Peninsula (Fig. 10) and in ONP in particular (Fig. 11). Between 13 and 30% of the Peninsula was classified as minimal risk for invasion while another 40-50% was classified as low current risk but greater (moderate or high) longer-term risk (Table 6). The majority of high risk areas were outside of ONP. Within ONP, moderate and high risk areas constituted <10% of the Park area (Table 6) and were concentrated in large river valleys (Fig. 11).

Combining statistical models for *Cirsium* and *Rubus* in the absence of ‘native-range’ or ‘invaded-range’ models was less informative. Nevertheless, combining models decreased the influence of errors in any one model or model type. Because suitability maps for the different modeling techniques were very different for these species, much of the Peninsula is classified as low to moderate risk, making these maps difficult to interpret (Table 7, Fig. 12). Risk classification for ONP, however, does show some intuitive patterns. High-risk habitat for *Cirsium* is concentrated in large river valleys, whereas for *Rubus* it includes both river valleys and coastal areas on the western side of the Peninsula (Fig. 13).

Role of Dispersal Corridors

Overall, 86% of sample plots were within 25 m of the nearest road or trail. This made it difficult to detect if species were more likely to occur along

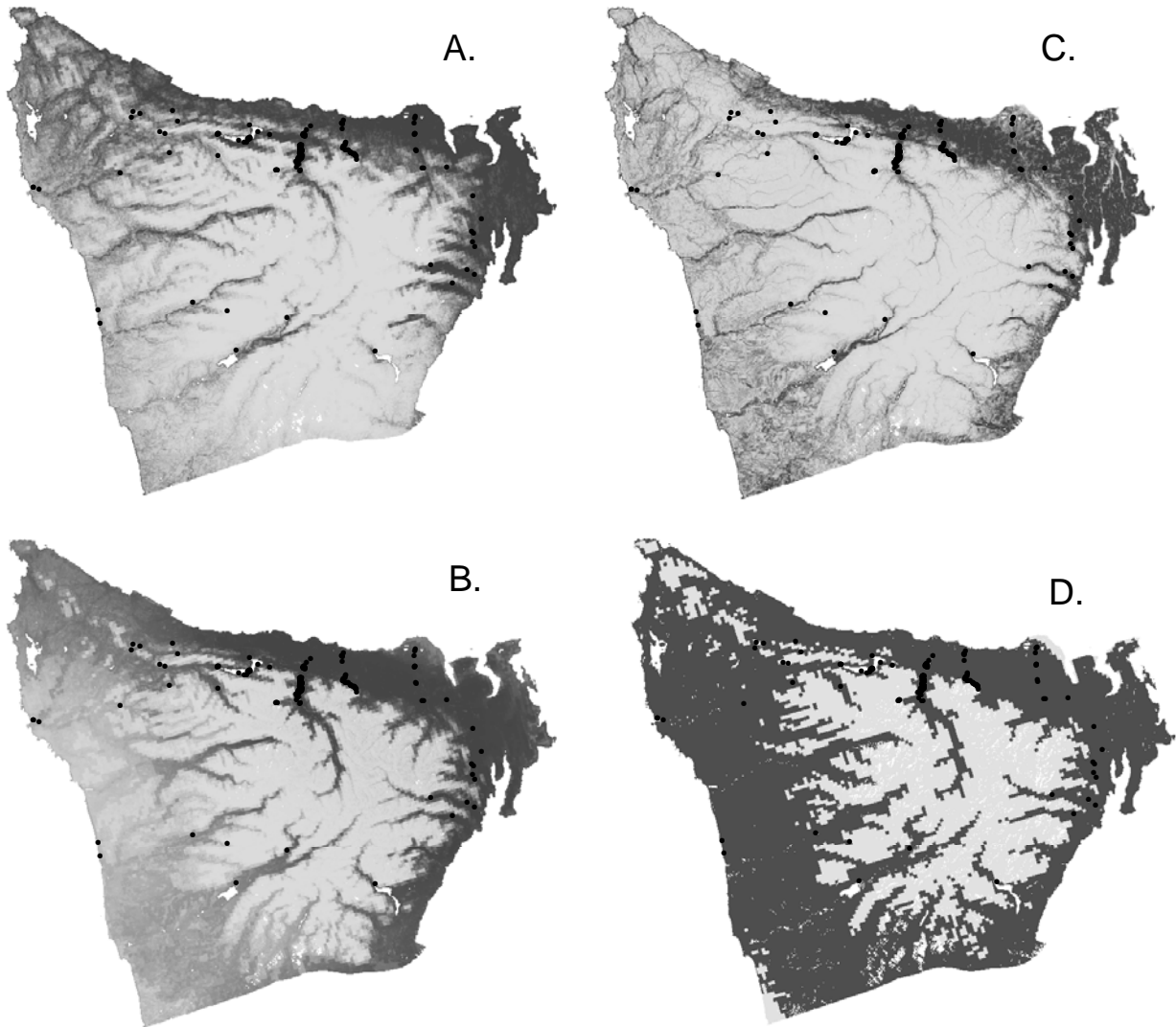


Figure 6. Habitat suitability maps for *Geranium robertianum* using logistic regression (A), GARP (B), ENFA (C) and 'invaded-range' models (D). For (A-C), darker shading indicates greater suitability on a range of 0-100. In (D), dark gray indicates suitable and light gray indicates unsuitable habitat. Black circles indicate where *Geranium* is present.

roads than expected by chance. However, four of the species were less common along roads than expected (*Geranium* 50% of occurrences near roads, $\chi^2 = 298$, $p < 0.001$; *Hedera* 68%, $\chi^2 = 7.9$, $p = 0.005$; *Ilex* 49%, $\chi^2 = 56.8$, $p < 0.001$; *Cirsium* 79%, $\chi^2 = 16$, $p < 0.001$) and only *Rubus* was more common along roads than expected ($\chi^2 = 10.2$, $p = 0.001$).

Only 21% of all plots were within 25 m of streams or wetlands. *Cirsium* (39% of occurrences, $\chi^2 = 51.6$, $p < 0.001$) and *Geranium* (46%, $\chi^2 = 74.2$, $p < 0.001$) were more common and *Ilex* (9%, $\chi^2 = 3.8$, $p = 0.05$) less common than expected near streams.

Although we did not quantitatively evaluate the importance of developed areas as sources of propagules, most populations of *Hedera* and *Ilex* in the park are near campgrounds, trailheads or maintenance facilities (C. Jones, personal observation).

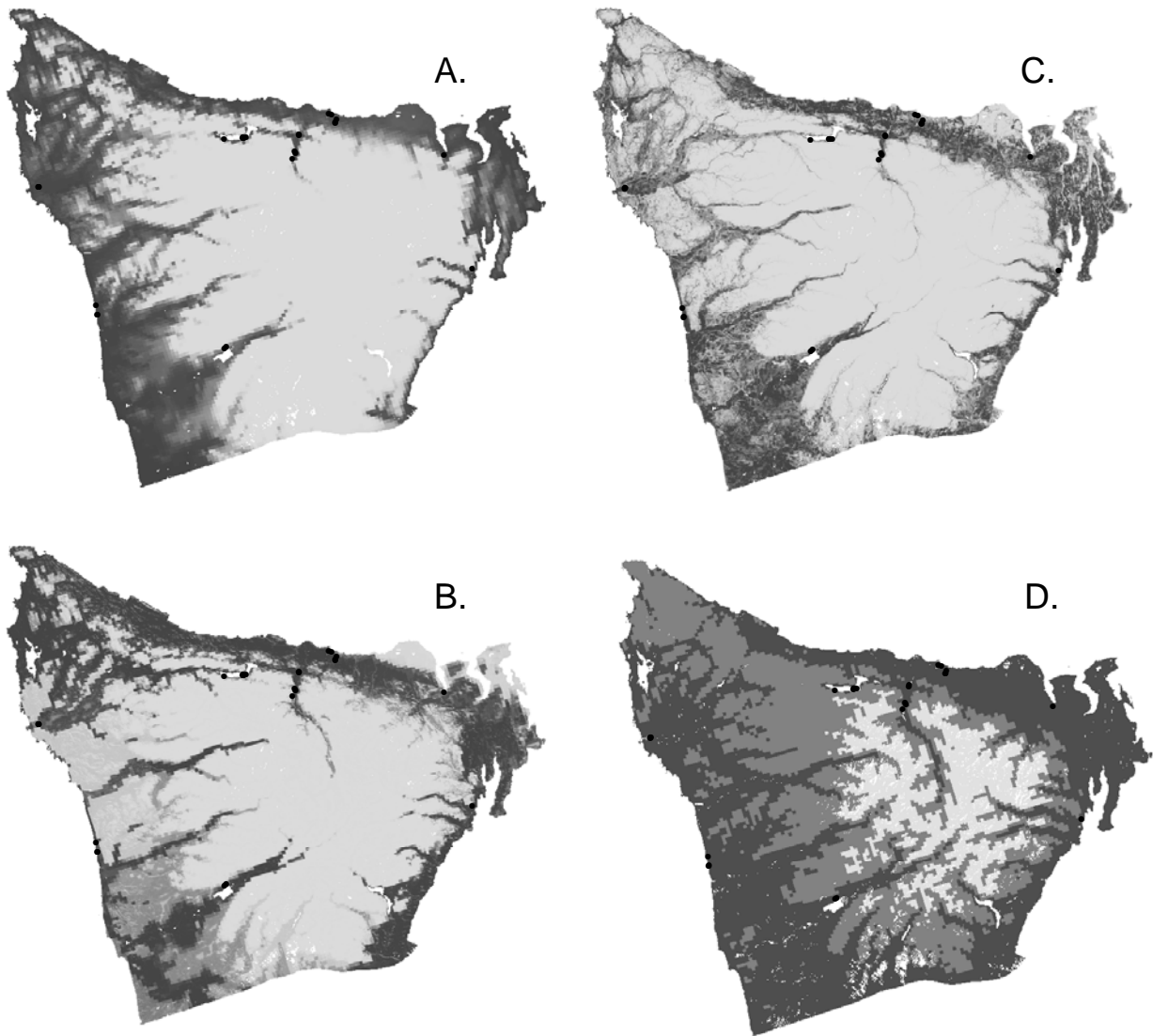


Figure 7. Habitat suitability maps for *Hedera helix* using logistic regression (A), GARP (B), ENFA (C) and alternative models (D). For (A-C), darker shading indicates greater suitability on a range of 0-100. In (D), dark gray indicates suitable habitat according to both 'native-range' and 'invaded-range' models, medium gray indicates suitable habitat according to the 'native-range' model only, and light gray indicates unsuitable habitat. Black circles indicate where *Hedera* is present.

Discussion

Habitat distribution models have the potential to be useful for modeling invasive species, but there are several challenges. Each type of model has assumptions that are often violated when used for invasive species. In addition, it is difficult to know how well models are performing. For these reasons, models must be developed and interpreted with care. Nevertheless, HDMs can be a useful tool to inform the development of monitoring plans for invasive species.

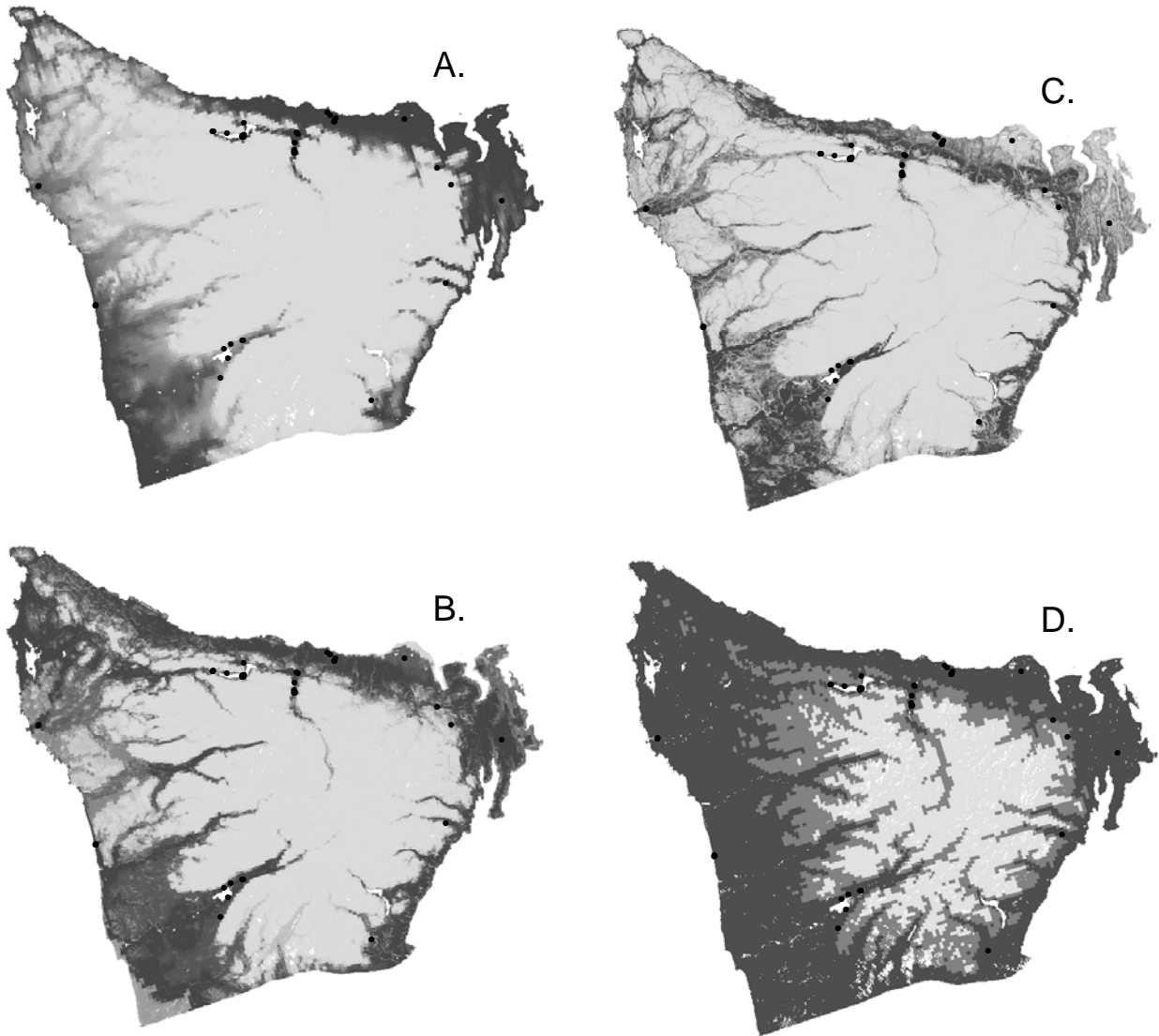


Figure 8. Habitat suitability maps for *Ilex aquifolium* using logistic regression (A), GARP (B), ENFA (C) and alternative models (D). For (A-C), darker shading indicates greater suitability on a range of 0-100. In (D), dark gray indicates suitable habitat according to both ‘native-range’ and ‘invaded-range’ models, medium gray indicates suitable habitat according to the ‘native-range’ model only, and light gray indicates unsuitable habitat. Black circles indicate where *Ilex* is present.

Model Performance

Based on measures of accuracy, models performed well for some species, but not for others. Some models, such as that for *Cirsium*, may not include habitat variables that are important in shaping species’ distributions. Alternatively, low accuracy may result from error in vegetation maps derived from satellite images: even though *Cirsium* generally does not occur in closed forest, 34% of its occurrences were in areas reported to have conifer cover >80%. Validation techniques can identify inconsistencies such as these and thus are important in model development.

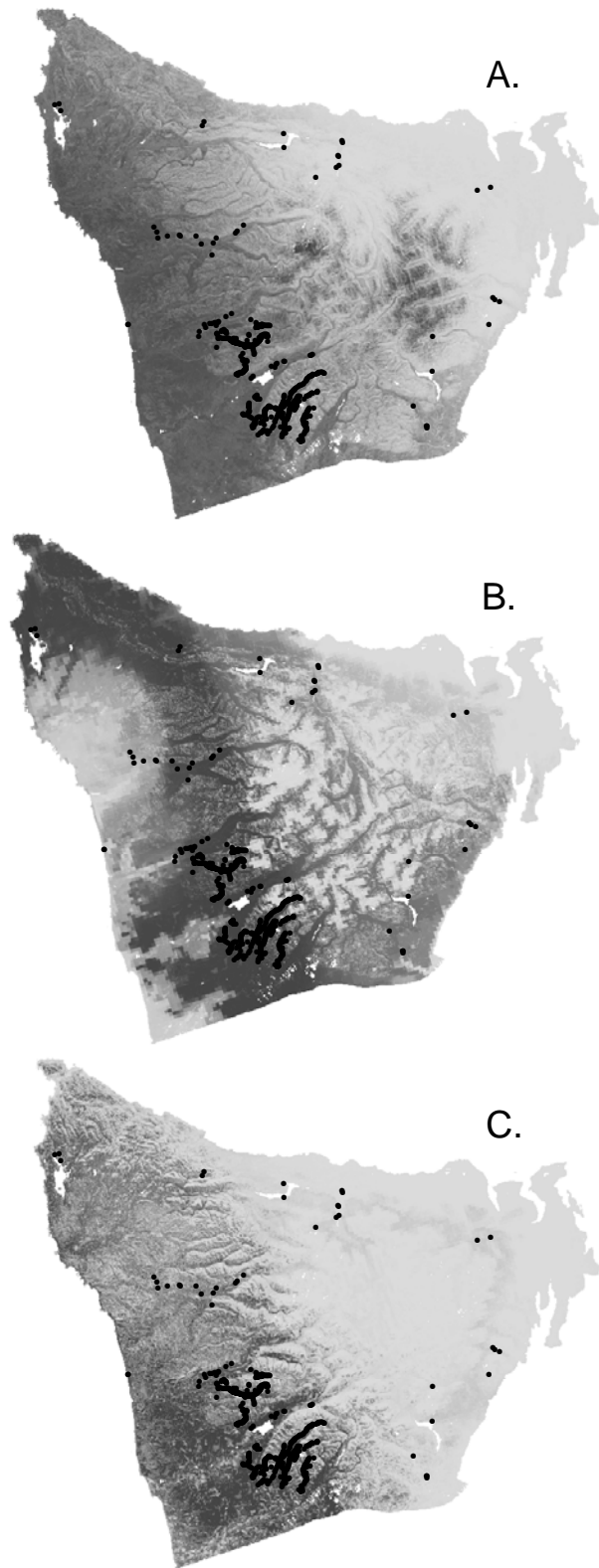


Figure 9. Habitat suitability maps for *Rubus laciniatus* using logistic regression (A), GARP (B) and ENFA (C). Darker shading indicates greater suitability on a range of 0-100. Black circles indicate where *Rubus* is present.

Various factors can lead to errors in prediction of potential habitat even when accuracy measures are high. First, when sample distributions are spatially clumped, model results can be influenced by areas with a high density of points. Model results for *Rubus* are illustrative. Because populations of *Rubus* are concentrated in the southwestern part of the Peninsula, model predictions of suitable habitat are also concentrated in this area. In contrast, areas where *Rubus* occurs in the northeastern part of the Peninsula are not predicted as suitable. Despite this error, model accuracy is high because there are relatively few occurrences in the northeast.

Second, because these invasive species are still spreading, models may be biased when dispersal patterns correspond with gradients in environment. For example, *Geranium* first invaded the Olympic Peninsula on the north coast near Port Angeles (Tisch 1992) and most occurrences are in this area. In the past 10 yr, however, it has started to spread around the perimeter of the Peninsula (C. Lucero, Clallam County Noxious Weed Control, personal communication). This gradient in dispersal shows a strong negative correlation with annual precipitation, which reaches a minimum in the area around Port Angeles. Thus models suggest that *Geranium* prefers areas with low precipitation when, in fact, other sources suggest that it prefers high moisture (Buckingham et al. 1995). As a result, much of the lowlands on the west side of the Peninsula are predicted as unsuitable habitat for *Geranium*. Separating the effects of environment from patterns of dispersal would be difficult without previous knowledge of the history of introduction and spread.

Finally, when numbers of occurrences are small, model results can be highly dependent on individual occurrences. Although model accuracy was high for *Hedera* and *Ilex*, there was considerable variation among model replicates. Thus, care should be taken when considering models based on few positive occurrences.

All of these issues underscore the challenge of using models to predict potential distributions. Measures of model accuracy only reflect how well models predict current distributions based on

Table 5. Accuracy and percent suitable habitat for ‘native-range’ and ‘invaded-range’ models

| | <i>Geranium</i> ‘Invaded-range’ | <i>Hedera</i> ‘Invaded-range’ | <i>Hedera</i> ‘Native-range’ | <i>Ilex</i> ‘Invaded-range’ | <i>Ilex</i> ‘Native-range’ |
|-------------|------------------------------------|----------------------------------|---------------------------------|--------------------------------|-------------------------------|
| Sensitivity | 0.98 | 0.96 | 1 | 1 | 1 |
| TSS | 0.19 | 0.42 | 0.01 | 0.49 | 0.09 |

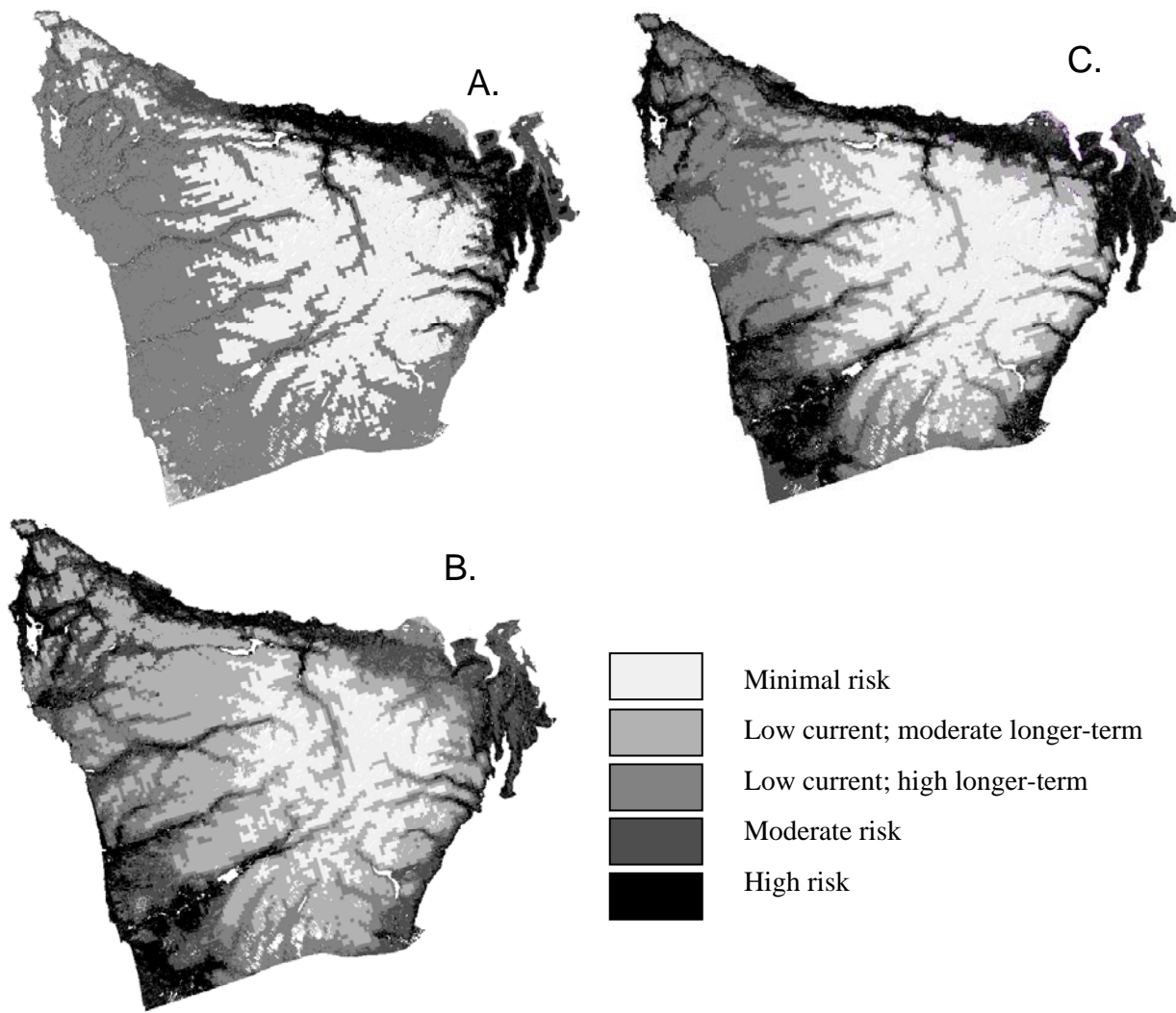


Figure 10. Invasion-risk classification for *Geranium robertianum* (A), *Hedera helix* (B), and *Ilex aquifolium* (C). Current risk is based on statistical modeling methods and longer-term risk is based on ‘native-range’ and ‘invaded-range’ models.

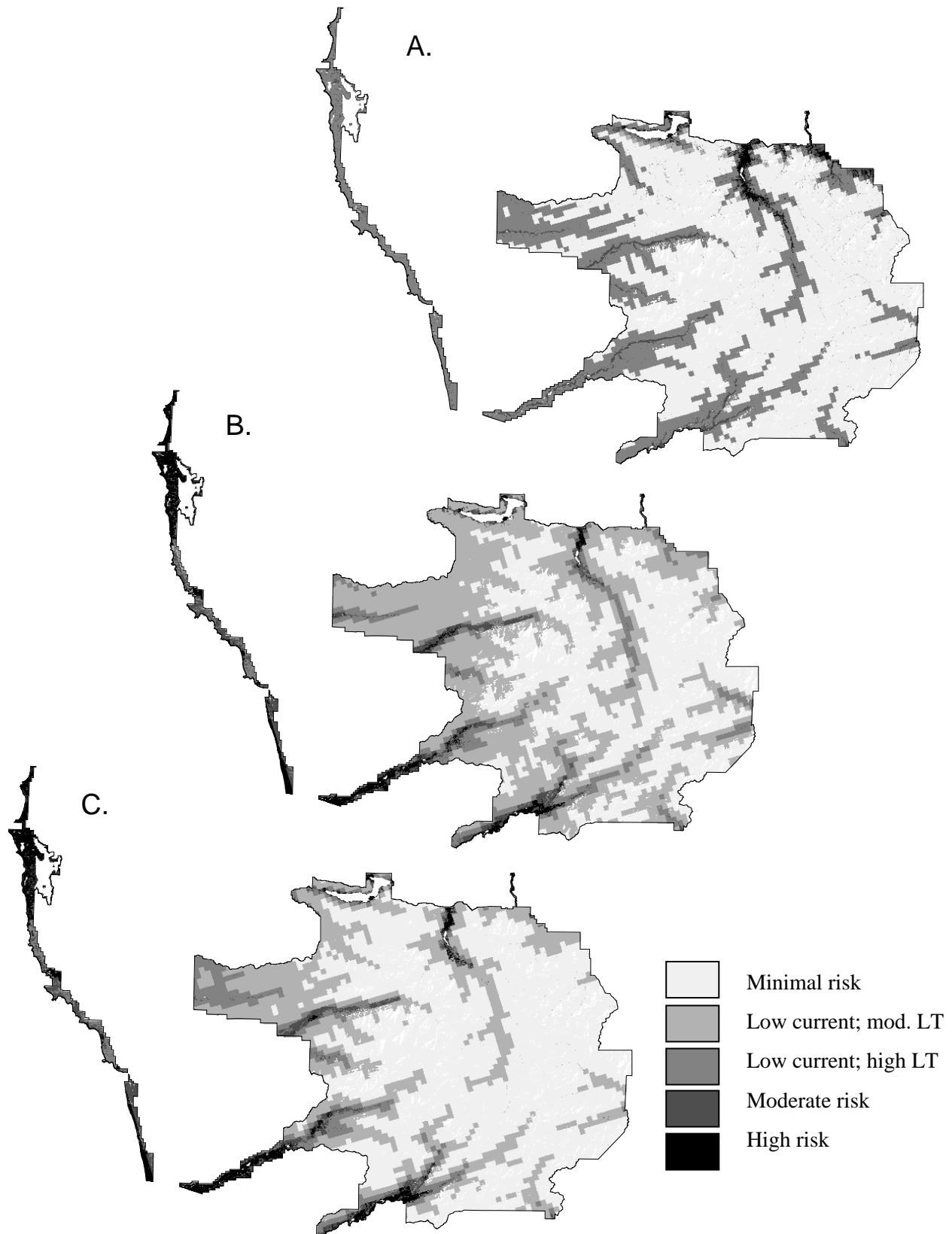


Figure 11. Invasion-risk classification for *Geranium robertianum* (A), *Hedera helix* (B), and *Ilex aquifolium* (C) in Olympic National Park. Current risk is based on statistical modeling methods and longer-term risk is based on ‘native-range’ and ‘invaded-range’ models.

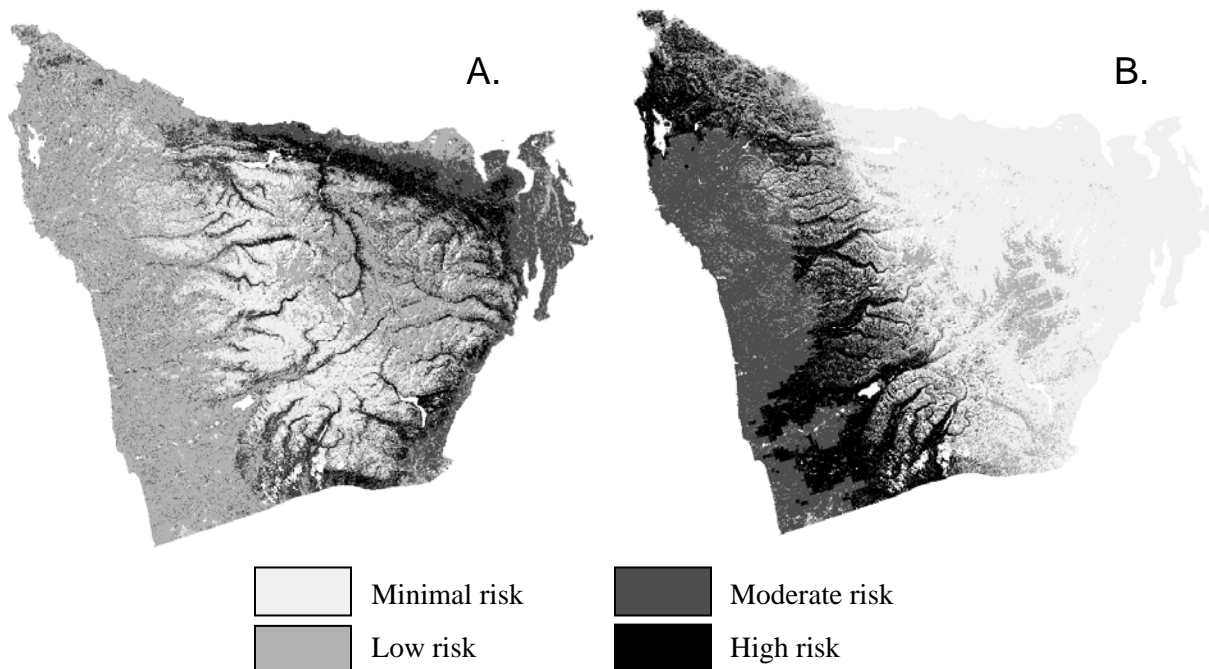
Table 6. Percent of study area (Penin.) and Olympic National Park (ONP) in different risk categories for *Geranium robertianum*, *Hedera helix* and *Ilex aquifolium*.

| Risk | <i>Geranium</i> | | <i>Hedera</i> | | <i>Ilex</i> | |
|-----------------------------------|------------------|------------------|---------------|------|-------------|------|
| | Penin. | ONP | Penin. | ONP | Penin. | ONP |
| Minimal | 29.4 | 63.9 | 13.3 | 39.2 | 23 | 58.5 |
| Low current, moderate longer-term | 1.8 ^a | 1.5 ^a | 33.7 | 43.5 | 21.1 | 26.2 |
| Low current, high longer-term | 49.6 | 29.8 | 17 | 9.0 | 22.9 | 7.8 |
| Moderate | 10.7 | 3.7 | 22.6 | 4.5 | 15.8 | 3.6 |
| High | 8.5 | 1.1 | 13.4 | 3.8 | 17.2 | 3.9 |

^aBecause there was no 'native-range' model for *Geranium* this value represents low current and low longer-term risk.

Table 7. Percent of study area (Penin.) and Olympic National Park (ONP) in different risk categories for *Cirsium arvense* and *Rubus laciniatus*.

| Risk (index score) | <i>Cirsium</i> | | <i>Rubus</i> | |
|--------------------|----------------|------|--------------|------|
| | Penin. | ONP | Penin. | ONP |
| Minimal (0) | 15.6 | 28.9 | 43.4 | 57.4 |
| Low (1-5) | 54.7 | 51.3 | 15.5 | 23.6 |
| Moderate (6-10) | 23.2 | 16.1 | 25.5 | 11.7 |
| High (11-15) | 6.5 | 3.7 | 15.6 | 7.3 |

**Figure 12.** Invasion-risk classification for *Cirsium arvense* (A) and *Rubus laciniatus* (B). Risk categories are based on numbers of models predicting suitability (0 = minimal, 1-5 = low, 6-10 = moderate, 11-15 = high).

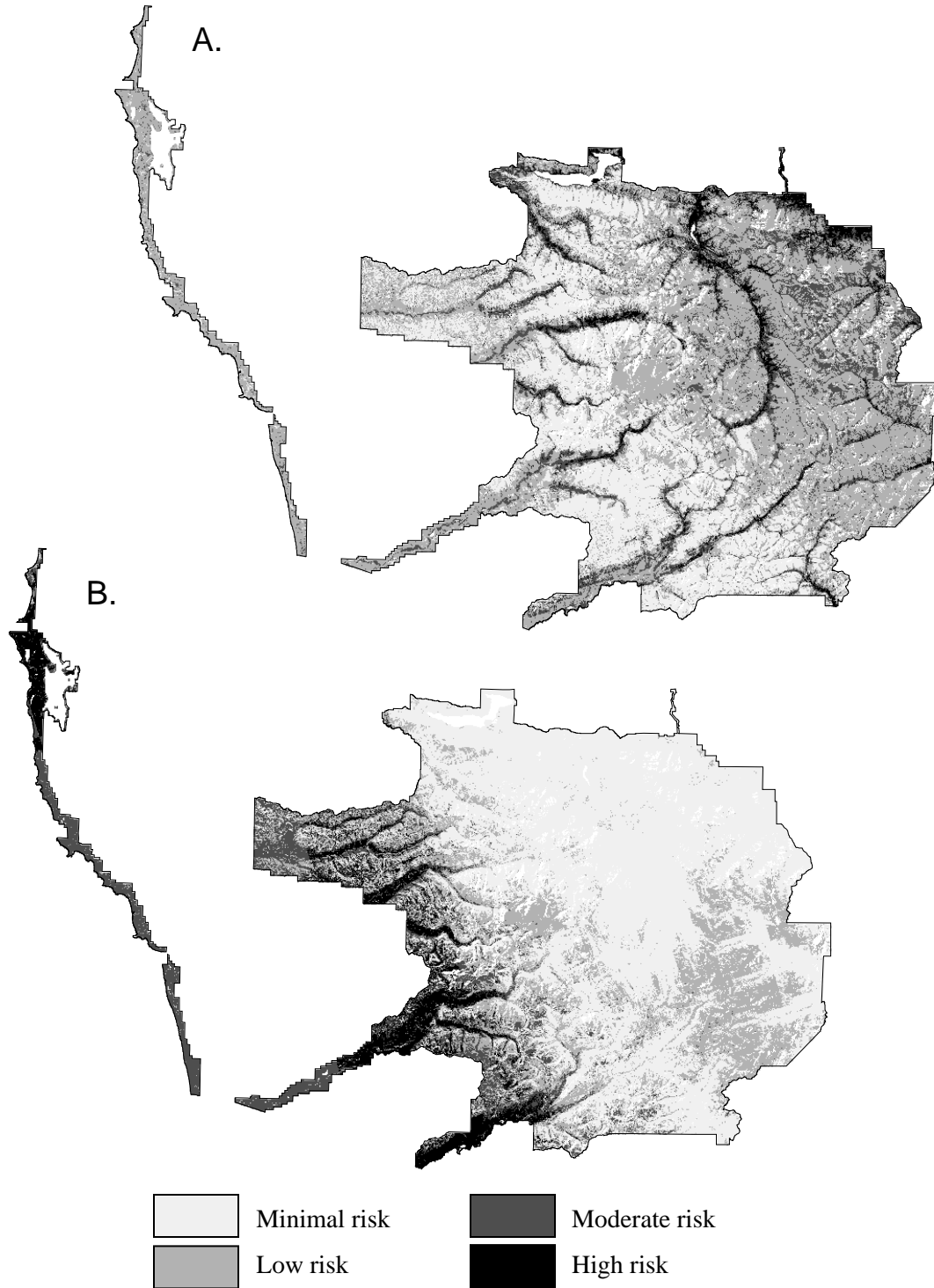


Figure 13. Invasion-risk classification for *Cirsium arvense* (A) and *Rubus laciniatus* (B) in Olympic National Park. Risk categories are based on numbers of models predicting suitability (0=minimal, 1-5 = low, 6-10 = moderate, 11-15 = high).

local validation data. Thus interpreting the predictions of models also requires judgment based on knowledge of the ecology of the species and the system of interest.

Logistic Regression vs. Presence-only Techniques

Logistic regression performed better than either of the two presence-only methods for almost every species in this study. Not only was accuracy higher, but there was less variability among replicates. This is surprising for two reasons. First, the model structures for both GARP and ENFA allow for interactions among habitat variables, whereas no interactions were included in the regression models. Second, logistic regression required inclusion of absence points, which, for these species, may reflect either unsuitable habitat or insufficient time for dispersal. Despite these limitations and assumptions, logistic regression outperformed presence-only models.

ENFA has been shown to work well for species whose ranges are continuing to spread when tested with simulated data (Hirzel et al. 2001). In this study, however, estimates of habitat suitability were quite low. Adjusting the threshold downward to maximize TSS increased area of suitable habitat, but even so, accuracy was only moderate. GARP, however, produced maps for *Geranium* and *Ilex* that were most similar to the ‘invaded-range’ and ‘native-range’ models, suggesting that it may be better at predicting future distributions. However, when higher thresholds for suitability were used to maximize current accuracy, the amount of predicted suitable habitat was very small, negating this advantage.

For most species, all of the modeling techniques predicted that habitat suitability would be low in the central western coastal region of the Peninsula. This result is probably an artifact of the sample-point distribution, which was very sparse in this area, rather than an effect of significant differences in habitat. However, this should be tested further. This effect was particularly pronounced in GARP models (Fig. 5-9b).

Models Based on Native vs. Invaded Range

Models based on species’ distributions across the contiguous USA (‘invaded-range’ models) predicted much larger areas of suitable habitat than did statistical models based on the data from the Olympic Peninsula. ‘Native-range’ models predicted even larger suitable areas. Models based on invaded vs. native ranges can differ for several reasons (Loo et al. 2007). Models based on the former may under-predict suitable habitat because the species has not yet spread through its potential range of habitat (MacIsaac et al. 2000). Alternatively, models based on the latter may under-predict suitable habitat (e.g., Beerling et al. 1995, Fitzpatrick et al. 2007) because the species is limited in its native range by competition, predation, or geographic or historical factors (Loo et al. 2007).

In this study, models based on current distributions appear to under-predict the potential habitats of all three of the species for which ‘invaded-range’ or ‘native-range’ models are available. Results suggest that all have spread through a greater proportion of their potential habitats in the contiguous USA, than in the Olympic Peninsula, which is not surprising given their relatively late arrival on the Peninsula. This suggests that on the Olympic Peninsula, these species are currently limited more by dispersal than environment and are likely to spread further as human or natural dispersal occurs. ‘Invaded-range’ models predict less suitable habitat than ‘native-range’ ones, suggesting that even across the contiguous USA, *Hedera* and *Ilex* do not occupy the full range of potential habitats.

There is, however, an important caveat to these comparisons. Both ‘invaded-range’ and ‘native-range’ models were developed from three coarse-scale climate variables and do not include any interactions. The statistical models from the Olympic Peninsula include many other habitat variables that may restrict

the distributions of these species within the climatically suitable range (e.g., vegetation cover, topography, distance to water). Thus, distributions predicted from ‘invaded-range’ and ‘native-range’ models should be interpreted as climatically suitable, but species should not be expected to occur everywhere within these ranges.

Combining Models to Inform Monitoring

It is clear that there are limitations with each kind of HDM and that model predictions should be treated more as hypotheses than as conclusions. However, these models can be useful in developing monitoring approaches to test these predictions. In addition, the combination of multiple modeling techniques shows great promise both for developing monitoring protocols and identifying areas for further research.

Combining different modeling techniques helps to reduce the influence of errors specific to individual methods. In particular, areas predicted to be suitable by all model types should be considered the most likely to be invaded. Conversely, areas that are consistently identified as unsuitable should be lower priority for monitoring. Because of the potential for model error, however, it is critical that some level of monitoring is implemented in these areas, particularly along roads or other potential dispersal corridors.

Using models based on both the current distribution of the species in the area of interest and other invaded areas or the species’ native range can further aid in monitoring. These models have different assumptions and allow different interpretation (i.e., current vs. longer-term risk of invasion). Our categorization of these models as predicting current and longer-term risk, respectively, is clearly a simplified approach. Nevertheless, for designing a monitoring approach, a simple characterization can be useful.

Implications for Monitoring in Olympic National Park

There are three important objectives that could be accomplished by an invasive species monitoring program in Olympic National Park. A primary objective would be to detect and control new populations in the early stages of invasion. However, a monitoring program could also be used to test and refine the predictions of these habitat models and could document temporal changes in species’ distributions. We discuss how results from our models could be used to develop such a monitoring program. We then outline a potential general approach to monitoring. Clearly, the details of this approach could be modified to address additional objectives or specific needs.

1. Intensity of monitoring: Risk-classification maps provide important information for developing monitoring approaches for *Geranium*, *Hedera*, and *Ilex*. Monitoring intensity could vary with the level of predicted risk.

a. Monitoring in areas of minimal risk: In combination, our models suggest that large areas of ONP (30% for *Cirsium*, 40% for *Hedera* and 60% for *Geranium*, *Ilex* and *Rubus*) are unlikely to be invaded. Monitoring can be conducted at low intensity in these areas, but some should be done for two reasons. First, it will help to test model predictions. If the invasive species establish in areas of minimal risk, there are problems with current models. Second, predictions for all of these models are based on current climate. As climate warms, species may spread to higher elevations than predicted, thus models will need to be modified (Guisan and Thuiller 2005)

b. Monitoring in areas of high risk: Monitoring should be most intensive in areas of high risk. In these areas there is the highest likelihood of new infestation and there is already considerable data for model development, thus efforts should be focused on control rather than model testing and refinement.

2. Monitoring *Cirsium* and *Rubus*: We have low confidence in the risk-classification maps for *Cirsium* and *Rubus*; these should be used with care. Both species are reported to occur mostly in disturbed areas (Murphy 2006, Townsend and Groom 2006), thus this may be a more productive way to focus monitoring. However, available GIS vegetation layers for the park will not be effective in identifying such habitats. Both the vegetation layers used in our models and those from Pacific Meridian Resources indicated that 35-60% of the occurrences of *Cirsium* and *Rubus* in ONP were in areas classified as having >70% canopy cover, which is unlikely for these species.

3. Dispersal sources and corridors: Our models suggest that all five species are in the early stages of invasion in Olympic National Park. Therefore, in the near future, distributions are likely to be controlled more by dispersal than by habitat variables. It may be useful to focus monitoring in areas where dispersal is likely. Roads and trails are likely to be important dispersal corridors (Parendes and Jones 2000, Trombulak and Frissell 2000, Gelbard and Belnap 2003, Watkins et al. 2003), but it was difficult to test for this because 86% of plots in our database were within 25 m of a road or trail. Only *Rubus* was more common near roads than would be expected by chance. Thus, it is clear that these species are not restricted to road corridors. Nevertheless, species may first colonize near roads and trails and then spread to surrounding areas.

Rivers and streams can also serve as invasion corridors. *Geranium* and *Cirsium* were more common within 25 m of streams than would be expected by chance. The pattern was particularly strong for *Geranium*. This may also indicate that moist riparian areas provide suitable habitat for these species. In either case, sampling along rivers and streams would be helpful for *Geranium* and *Cirsium*.

Hedera and *Ilex* are most common in ONP near campgrounds, trailheads or maintenance facilities (C. Jones, personal observation). These species appear to be spreading into forests from these locations, thus monitoring should be conducted for these species both in and around developed areas.

Potential monitoring program

A two-part monitoring program could be used to achieve the objectives of controlling the spread of these species, testing and refining models, and following temporal changes in distributions.

The first component would consist of a network of permanent plots measured on a regular basis. These plots would be established at low density in areas with minimal risk of invasion, moderate density where current risk is low but longer-term risk is high, and high density in all other areas (where longer-term risk is high). Actual plot densities could be adjusted based on resource availability.

Rectangular 100 m² plots would provide a balance between effort required to sample each plot and the likelihood of detecting the species. For each plot a complete list of non-native species should be recorded (and reproduced for all subsequent sampling). In addition, cover of overstory trees and understory vegetation should be estimated using a standard method for estimation.

These plots should be stratified by distance to roads, trails and streams. Because roads, trails and streams are likely dispersal corridors—and for ease of sampling—most plots should be within 100 m of one of these features. For example, plots could be divided into five equal categories (adjacent to roads, trails or streams, <25 m, 25-50 m, 50-100 m and >100 m).

These permanent plots would aid in detecting new infestations and over time will provide data for testing and refining model predictions. Because they can be resampled, they will provide information on changes in species' distributions.

The second component could supplement monitoring in areas of high risk. The major purpose would be to detect and control new infestations. This sampling would not be plot based, but would focus on species detection in larger defined areas (similar to the Elwha exotic plant survey). Surveys would focus on high-risk areas, along roads and trails and near development. If a species was found its location would be recorded and added to the database.

It could be possible to supplement these sampling approaches with information gathered by volunteers or park visitors. Frequent visitors to the park could be trained to identify key invasive species. A system could then be developed for reporting general locations to park staff who could then verify these reports. A similar approach has been used successfully by the Invasive Plant Atlas of New England (www.ipane.org).

Conclusions

There are a number of challenges to using HDMs to predict the potential distributions of invasive species. These challenges include errors in habitat or species data, clumping of sample distributions, species' lack of equilibrium with their environment, and assumptions specific to each modeling technique. Given these challenges, it is important to critically consider model results. Error and bias can be reduced by combining different modeling methods and developing models based on species data from the area of interest as well as from other invaded areas and the native ranges of species. Combining models can provide useful predictions on current and future risk and help to set priorities for developing efficient monitoring protocols. An iterative process of modeling and monitoring can then be used over time to test and refine predictions about species' distributions.

References Cited

- Allouche, O., A. Tsoar, and R. Kadmon. 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology* **43**:1223-1232.
- Anderson, R. P., A. T. Peterson, and S. L. Egbert. 2006. Vegetation-index models predict areas vulnerable to purple loosestrife (*Lythrum salicaria*) invasion in Kansas. *Southwestern Naturalist* **51**:471-480.
- Arrieta, S., and F. Suarez. 2004. Germination and seed bank depletion of holly (*Ilex aquifolium* L.) in four microhabitat types. *Seed Science Research* **14**:305-313.
- Barndt, J. K. 2006. Herb Robert (*Geranium robertianum*). Pages 56-57 in P. D. Boersma, S. H. Reichard, and A. N. Van Buren, editors. *Invasive Species in the Pacific Northwest*. University of Washington Press., Seattle and London.
- Berling, D. J., B. Huntley, and J. P. Bailey. 1995. Climate and the distribution of *Fallopia japonica* - use of an introduced species to test the predictive capacity of response surfaces. *Journal of Vegetation Science* **6**:269-282.
- Benjamin, P., and R. Hiebert. 2004. Assessing the invasive plant issue. *Park Science* **22**:27-31.
- Bertin, R. I. 2001. Life cycle, demography, and reproductive biology of herb Robert (*Geranium robertianum*). *Rhodora* **103**:96-116.

- Blood, K. 2001. Environmental weeds: a field guide for Southeast Australia. CH Jerran & Associates, Mount Waverly, Victoria, Australia.
- Buchan, L. A. J., and D. K. Padilla. 2000. Predicting the likelihood of Eurasian watermilfoil presence in lakes, a macrophyte monitoring tool. *Ecological Applications* **10**:1442-1455.
- Buckingham, N. M., E. G. Schreiner, T. N. Kaye, J. E. Burger and E. L. Tisch. 1995. Flora of the Olympic Peninsula. Northwest Interpretive Association: Washington Native Plant Society, Seattle, WA.
- Buckingham, N. M., and E. L. Tisch. 1979. Vascular plants of the Olympic Peninsula. NPS Cooperative Park Studies Unit Report B-79-2, Seattle, WA.
- Busby, J. R. 1991. BIOCLIM — a bioclimatic analysis and prediction system. Pages 64-68 *in* C. R. Margules and M. P. Austin, editors. Nature conservation: cost effective biological surveys and data analysis. CSIRO, Melbourne.
- Carey, A. B., and C. A. Harrington. 2001. Small mammals in young forests: implications for management for sustainability. *Forest Ecology and Management* **154**:289-309.
- Crawley, M. J. 2002. Statistical computing: an introduction to data analysis using S-plus. John Wiley & Sons, Chichester, UK.
- DeFerrari, C. M., and R. J. Naiman. 1994. A multi-scale assessment of the occurrence of exotic plants on the Olympic Peninsula, Washington. *Journal of Vegetation Science* **5**:247-258.
- Dlugosch, K. M. 2005. Understory community changes associated with English ivy invasions in Seattle's urban parks. *Northwest Science* **79**:52-59.
- Donald, W. W. 1994. The biology of Canada thistle (*Cirsium arvense*). *Reviews of Weed Science* **6**:77-101.
- Drees, L. 2004. A retrospective on NPS invasive species policy and management. *Park Science* **22**:21-26.
- Elith, J., C. H. Graham, R. P. Anderson, M. Dudik, et al. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* **29**:129-151.
- Fernald, M. L. 1950. Gray's manual of botany, 8th edition. Dioscorides Press, Portland, Oregon.
- Fielding, A. H., and J. F. Bell. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* **24**:38-49.
- Fitzpatrick, M. C., J. F. Weltzin, N. J. Sanders, and R. R. Dunn. 2007. The biogeography of prediction error: why does the introduced range of the fire ant over-predict its native range? *Global Ecology and Biogeography* **16**:24-33.
- Freshwater, V. 1991. Control of English ivy (*Hedera helix*) in Sherbrooke forest - a practical experience. *Plant Protection Quarterly* **6**:127.
- Gelbard, J. L., and J. Belnap. 2003. Roads as conduits for exotic plant invasions in a semiarid landscape. *Conservation Biology* **17**:420-432.

- Grime, J. P., J. G. Hodgson, and R. Hunt. 1988. Comparative plant ecology: a functional approach to common British species. Unwin Hyman, London.
- Guisan, A., and W. Thuiller. 2005. Predicting species distribution: offering more than simple habitat models. *Ecology Letters* **8**:993-1009.
- Guisan, A., and N. E. Zimmermann. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling* **135**:147-186.
- Henderson, J. A., D. H. Peter, R. D. Leshner, and D. C. Shaw. 1989. Forested plant associations of the Olympic National Forest. USDA Forest Service Region 6 Ecology Technical Paper 001-88.
- Hirzel, A. H., J. Hausser, D. Chessel, and N. Perrin. 2002a. Ecological-niche factor analysis: How to compute habitat-suitability maps without absence data? *Ecology* **83**:2027-2036.
- 2002b. Biomapper 3.0. Laboratory for Conservation Biology, University of Lausanne, Lausanne
- Hirzel, A. H., V. Helfer, and F. Metral. 2001. Assessing habitat-suitability models with a virtual species. *Ecological Modelling* **145**:111-121.
- Houston, D. B., E. G. Schreiner, and B. B. Moorhead 1994. Mountain goats in Olympic National Park: biology and management of an introduced species. Scientific Monograph NPS/NROLYM/NRSM-94/25.
2003. S-PLUS 6.2. Insightful Corporation, Seattle, WA
- Iverson, J. 1944. *Viscum*, *Hedera* and *Ilex* as climate indicators. *Geologiska Föreningens I Stockholm Förhandlingar* **66**:463-483.
- Jones, G. N. 1936. A botanical survey of the Olympic Peninsula, WA. University of Washington Press, Seattle, WA.
- Klinka, K., A. M. Scagel, and P. J. Courtin. 1985. Vegetation relationships among some seral ecosystems in southwestern British Columbia. *Canadian Journal of Forest Research* **15**:561-569.
- Laskurain, N. A., A. Escudero, J. M. Olano, and J. Loidi. 2004. Seedling dynamics of shrubs in a fully closed temperate forest: greater than expected. *Ecography* **27**:650-658.
- Levine, J. M., M. Vila, C. M. D'Antonio, J. S. Dukes, et al. 2003. Mechanisms underlying the impacts of exotic plant invasions. *Proceedings of the Royal Society of London Series B* **270**:775-781.
- Lezberg, A. L., C. B. Halpern, and J. A. Antos. 2001. Clonal development of *Maianthemum dilatatum* in forests of differing age and structure. *Canadian Journal of Botany* **79**:1028-1038.
- Lonsdale, W. M. 1999. Global patterns of plant invasions and the concept of invasibility. *Ecology* **80**:1522-1536.
- Loo, S. E., R. MacNally, and P. S. Lake. 2007. Forecasting New Zealand mudsnail invasion range: model comparison using native and invaded ranges. *Ecological Applications* **17**:181-189.
- MacIsaac, H. J., H. A. M. Ketelaars, I. A. Grigorovich, C. W. Ramcharan, et al. 2000. Modeling *Bythotrephes longimanus* invasions in the Great Lakes basin based on its European distribution. *Archiv fuer Hydrobiologie* **149**:1-21.

- Mack, R. N., D. Simberloff, W. M. Lonsdale, H. Evans, et al. 2000. Biotic invasions: causes, epidemiology, global consequences and control. *Ecological Applications* **10**:689-710.
- MacNally, R. 2002. Multiple regression and inference in ecology and conservation biology; further comments on identifying important predictor variables. *Biodiversity and Conservation* **11**:1397-1401.
- Maggini, R., A. Lehmann, N. E. Zimmermann, and A. Guisan. 2006. Improving generalized regression analysis for the spatial prediction of forest communities. *Journal of Biogeography* **33**:1729-1749.
- Manel, S., H. C. Williams, and S. J. Ormerod. 2001. Evaluating presence-absence models in ecology: the need to account for prevalence. *Journal of Applied Ecology* **38**:921-931.
- McCune, B., and D. Keon. 2002. Equations for potential annual direct incident radiation and heat load. *Journal of Vegetation Science* **13**:603-606.
- McNab, W. H., and D. L. Loftis. 2002. Probability of occurrence and habitat features for oriental bittersweet in an oak forest in the southern Appalachian mountains, USA. *Forest Ecology and Management* **155**:45-54.
- Metcalf, D. J. 2005. *Hedera helix* L. *Journal of Ecology* **93**:632-648.
- Moore, I. D., P. E. Gessler, G. A. Nielsen, and G. A. Peterson. 1993. Soil attribute prediction using terrain analysis. *Soil Science Society of America Journal* **57**:443-452.
- Moore, R. J. 1975. The biology of Canadian weeds, 13. *Cirsium arvense* (L.) Scop. *Canadian Journal of Plant Science* **55**:1033-1048.
- Morishita, D. W. 1999. Canada thistle. Pages 162-174 in R. L. Sheley and J. K. Petroff, editors. *Biology and management of noxious rangeland weeds*. Oregon State University Press, Corvallis, OR.
- Murai, M. 1999. Understanding the invasion of Pacific Northwest forests by English ivy (*Hedera* spp., Araliaceae). M.S. Thesis, University of Washington, Seattle, WA, USA.
- Murphy, M. 2006. Evergreen blackberry (*Rubus laciniatus*) and Himalayan blackberry (*Rubus armeniacus*). Pages 30-31 in P. D. Boersma, S. H. Reichard, and A. N. Van Buren, editors. *Invasive Species in the Pacific Northwest*. University of Washington Press., Seattle and London.
- Muyt, A. 2001. *Bush invaders of South-East Australia: a guide to the identification and control of environmental weeds found in South-East Australia*. RG & FJ Richardson, Meridith, Victoria, Australia.
- National Invasive Species Council. 2001. Meeting the invasive species challenge: national invasive species management plan. National Invasive Species Council, Washington DC.
- Nicholls, A. O. 1989. How to make biological surveys go further with generalized linear models. *Biological Conservation* **50**:51-75.
- Parendes, L. A., and J. A. Jones. 2000. Role of light availability and dispersal in exotic plant invasion along roads and streams in the H. J. Andrews Experimental Forest, Oregon. *Conservation Biology* **14**:64-75.
- Pearce, J., and S. Ferrier. 2000. Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling* **133**:225-245.

- Peterken, G. 1966. Mortality of holly (*Ilex aquifolium*) seedlings in relation to natural regeneration in the New Forest. *Journal of Ecology* **54**:259-269.
- Peterken, G. F., and P. S. Lloyd. 1967. Biological flora of the British Isles. *Ilex aquifolium* L. *Journal of Ecology* **55**:841-858.
- Peterson, A. T. 2003. Predicting the geography of species' invasions via ecological niche modeling. *Quarterly Review of Biology* **78**:419-433.
- Piper, C. V. 1906. Flora of the state of Washington. US Government Printing Office, Washington, D.C.
- Prentice, I. C., and H. Helmsaari. 1991. Silvics of North European trees - compilation, comparisons and implications for forest succession modeling. *Forest Ecology and Management* **42**:79-93.
- Reutebuch, S. E., C. A. Harrington, D. D. Marshall, and L. C. Brodie. 2004. Use of large-scale silvicultural studies to evaluate management options in Pacific Northwest forests of the United States. *Forest Snow and Landscape Research* **78**:191-208.
- Rew, L. J., B. D. Maxwell, F. L. Dougher, and R. Aspinall. 2006. Searching for a needle in a haystack: evaluating survey methods for non-indigenous plant species. *Biological Invasions* **8**:523-539.
- Robertson, M. P., M. H. Villet, and A. R. Palmer. 2004. A fuzzy classification technique for predicting species' distributions: applications using invasive alien plants and indigenous insects. *Diversity and Distributions* **10**:461-474.
- Rodríguez, J. P., L. Brotons, J. Bustamante, and J. Seoane. 2007. The application of predictive modelling of species distribution to biodiversity conservation. *Diversity and Distributions* **13**:243-251.
- Rose, P. Q. 1996. The gardener's guide to growing ivies. Timber Press, Portland, Oregon.
- Scott, J. M., P. J. Heglund, M. L. Morrison, J. B. Haufler, et al., editors. 2002. Predicting species occurrences: issues of accuracy and scale. Island Press, Washington, D.C.
- Seymour, F. C. 1969. The flora of New England; a manual for the identification of all vascular plants, including ferns and fern allies and flowering plants growing without cultivation in New England. C.E. Tuttle Co., Rutland, Vermont.
- Shoup, S., and C. E. Whitcomb. 1981. Interactions between trees and ground covers. *Journal of Arboriculture* **7**:186-187.
- Stockwell, D., and D. Peters. 1999. The GARP modelling system: problems and solutions to automated spatial prediction. *International Journal of Geographical Information Science* **13**:143-158.
- Thomas, L. K. J. 1980. The impact of three exotic plant species on a Potomac island. U.S. Department of the Interior, Washington D.C.
- Thornton, P. E., H. Hasenauer, and M. A. White. 2000. Simultaneous estimation of daily solar radiation and humidity from observed temperature and precipitation: an application over complex terrain in Austria. *Agricultural and Forest Meteorology* **104**:255-271.

- Thornton, P. E., and S. W. Running. 1999. An improved algorithm for estimating incident daily solar radiation from measurements of temperature, humidity and precipitation. *Agricultural and Forest Meteorology* **93**:211-228.
- Thornton, P. E., S. W. Running, and M. A. White. 1997. Generating surfaces of daily meteorological variables over large regions of complex terrain. *Journal of Hydrology* **190**:214-251.
- Thuiller, W., D. M. Richardson, P. Pysek, G. F. Midgley, et al. 2005. Niche-based modelling as a tool for predicting the risk of alien plant invasions at a global scale. *Global Change Biology* **11**:2234-2250.
- Tirmenstein, D. 1989. *Rubus laciniatus*. In: Fire Effects Information System. USDA Forest Service, Rocky Mountain Research Station, Fire Sciences Laboratory. Available online at: <http://www.fs.fed.us/database/feis/>. Accessed June 2 2007
- Tisch, E. 1992. Alien weed threatens Olympic National Park. *Voice of the Wild Olympics: Newsletter of the Olympic Park Associates* **1**:6.
- Tofts, R. J. 2004. *Geranium robertianum* L. *Journal of Ecology* **92**:537-555.
- Townsend, P. A., and M. J. Groom. 2006a. Bull thistle (*Cirsium vulgare*) and Canada thistle (*Cirsium arvense*). Pages 90-91 in A. N. Van Buren, editor. *Invasive Species in the Pacific Northwest*. University of Washington Press., Seattle and London.
- Townsend, P. A., and M. J. Groom. 2006b. Bull thistle (*Cirsium vulgare*) and Canada thistle (*Cirsium arvense*). Pages 90-91 in P. D. Boersma, S. H. Reichard, and A. N. Van Buren, editors. *Invasive Species in the Pacific Northwest*. University of Washington Press., Seattle and London.
- Trombulak, S. C., and C. A. Frissell. 2000. Review of ecological effects of roads on terrestrial and aquatic communities. *Conservation Biology* **14**:18-30.
- Underwood, E. C., R. Klinger, and P. E. Moore. 2004. Predicting patterns of non-native plant invasions in Yosemite National Park, California, USA. *Diversity and Distributions* **10**:447-459.
- Usher, M. B. 1988. Biological invasions of nature reserves - a search for generalizations. *Biological Conservation* **44**:119-135.
- Van Ruremonde, R., and J. T. R. Kalkhoven. 1991. Effects of woodlot isolation on the dispersion of plants with fleshy fruits. *Journal of Vegetation Science* **2**:377-384.
- Vitousek, P. M., C. M. D'Antonio, L. L. Loope, M. Rejmanek, et al. 1997. Introduced species: A significant component of human-caused global change. *New Zealand Journal of Ecology* **21**:1-16.
- Watkins, R. Z., J. Q. Chen, J. Pickens, and K. D. Brososke. 2003. Effects of forest roads on understory plants in a managed hardwood landscape. *Conservation Biology* **17**:411-419.
- Weber, E. 2003. *Invasive plant species of the world: a reference guide to environmental weeds*. CABI Publishing, Oxon, UK.
- Wieman, J. S. 1961. History of English holly (*Ilex aquifolium*) in Oregon and the Northwest. J.S. Wieman, Portland, Oregon.

Wilcove, D. S., D. Rothstein, J. Dubrow, A. Phillips, et al. 1998. Quantifying threats to imperiled species in the United States. *Bioscience* **48**:607-615.

Appendix A. Species Datasets

Olympic National Park Exotics Inventory

The purpose of this inventory was to determine the distribution of exotic species within Olympic National Park and to determine the habitats that are most susceptible to invasion. Sampling was conducted in four types of areas: in developed areas and along roads, trails and rivers. Developed areas or road/trail/river segments were randomly chosen from all those within the park and multiple plots were systematically sampled within each area or segment. All plots were 100 m² but of differing shape. For developed areas plots were 10 X 10 m, while for roads and trails plots were 1 X 100 m and located along the trail or road edge. For rivers the plot shape varied so as to include only riparian vegetation, but the plot size was always 100 m². In each plot the presence of all exotic species was recorded. For developed areas, 101 plots in 27 areas were surveyed. A total of 86 plots in 18 road segments, 31 plots in 10 trail segments, and 31 plots in 9 river segments were surveyed. There were a total of 249 plots surveyed. Data were collected in 2002.

Data and information about this dataset are held by the staff at Olympic National Park

Elwha River Restoration Area Exotics Survey

The purpose of this survey was to document the location and population size of exotic species within the Elwha River restoration area (River mile 5-16). The survey focused on riparian areas (all areas within 50 m from the river) and developed areas within 1.5 km of the river (including roads, power line corridors, trails, and administrative areas). Reservoir shorelines were surveyed by boat while other areas were surveyed by foot. The centers of each population of 26 target exotic species within the surveyed area were recorded. A total of 580 locations were recorded. Many of these contained populations of more than one exotic species so there were a total of 1981 populations recorded. Data were collected in 2001.

Because only the presence of species was recorded, for this project I will not any of these points to show species absence. Data and information about this dataset are held by the staff at Olympic National Park

Alpine and Subalpine Plant Community Survey Plots (Mountain Goat Study)

These data were collected to characterize and classify the vegetation in the summer range of the mountain goat in the Olympic Mountains. Cover for all species was estimated in plots ranging from 20m² to 100 m². General locations were determined based on the summer range of the mountain goats. Within each general area, plots were located and sized to characterize areas of uniform vegetation. Data were collected from 1981-1990. There were a total of 771 plots in 69 general areas. Some of the plots were focused on populations of the rare plant *Astragalus cottonii*. These plots were resampled in 1999.

For purposes of this project we only used plots that were 100m² and sampled since 1990. There were 13 that fit into this category. In one case there were two plots within 12 m of each other that together equaled 107 m². This pair of plots was included as a single data point.

Contact: Ed Schreiner, USGS BRD Olympic Field Station.

Olympic National Park Exotic Plant Management Team Data

These points are locations where infestations of the target exotic species have been located by the Olympic National Park Exotic Plant Management Team. There are 65 locations where *Geranium robertianum*, *Hedera helix*, and *Ilex aquifolium* were recorded in 2002-2004. None of these points indicate the absence of any species.

Contact: Dan Campbell, North Coast – Cascades Exotic Plant Management Team,
Dan_Campbell@nps.gov

LTEM Elwha Vegetation Survey

This project was carried out to determine if physical characteristics could be used to predict plant species associations within the Elwha Valley. Plots were randomly located within the valley and were classified into 500 m elevation bands and to lower, middle or upper reaches of the river. Plots were either 400 m² or 625 m² depending on the tree density (where trees were less dense the larger plot size was used). Percent cover of all understory species were recorded in 61 plots during 1999 and 2000 (two additional plots lack coordinates and were discarded for the current study).

Contact: Andrea Woodward, USGS-FRESC, andrea_woodward@usgs.gov

Collette DeFerrari Dataset

The purpose of this study was to examine patterns of exotic species across a landscape at multiple scales (DeFerrari and Naiman 1994). Plots were located in different vegetation patches in both upland and riparian areas in the Hoh and Dungeness watersheds. Data were collected in 1992.

All species (both native and exotic) were identified in plots that were either 50 m² (71 riparian plots) or 100 m² (138 upland plots). For this project we geo-referenced the plots based on comments on the general vicinity and hand drawn maps. After entering these plots into GIS we verified the locations based on elevation, slope, aspect and distance to the nearest road or trail. We included only plots that we could locate within 50 m or within 100 m if they were in an area with little variation in habitat variables. In addition, the riparian plots (because of their size) were only included if one of the target species were present. Based on these rules we included 29 riparian plots and 66 other plots.

DeFerrari, C.M. & R.J. Naiman. 1994 A multi-scale assessment of the occurrence of exotic plants on the Olympic Peninsula, Washington. *Journal of Vegetation Science* 5:247-258.

Ann Lezberg Dataset

The purpose of this study was to compare the development of understory vegetation and overstory structure among four age classes of low elevation forests on the Olympic Peninsula and to examine the relationship between understory vegetation and overstory characteristics.

Understory vegetation data were collected in 4-6 100 m² plots in each of 20 stands on the western Olympic Peninsula. There were a total of 100 plots. Plots were sampled in 1996 and 1997.

Lezberg, A. L., C. B. Halpern, and J. A. Antos. 2001. Clonal development of *Maianthemum dilatatum* in forests of differing age and structure. *Canadian Journal of Botany* 79:1028-1038.

Olympic National Forest Exotic Inventory

This project was designed by the Olympic National Forest to assess the distribution and abundance on exotic species in the forest, with an emphasis on roadside habitat. There were a total of 3193 plots sampled between 1995 and 2003. There are three major plot types: 1) 400 m² circular plots, 2) plots encompassing road segments that are ~12 m wide and 150-200 m long and 3) irregular polygons ranging in size from 200 m² to 40 ha. The presence or absence of all of the target species were recorded in each of these plots. For the current project we eliminated all plots greater than 2500 m², which left 2881 plots.

Contact: Joan Ziegltrum, Olympic National Forest, jziegltrum@fs.fed.us

Olympic Habitat Development Study

This dataset is from a large-scale silvicultural experiment conducted by the USDA Forest Service in Olympic National Forest. The goal of the project was to assess the effects of variable density thinning in mid-rotation forests on animals, understory and canopy structure (Reutebuch et al. 2004). Plots were assigned to treatment or control and 100 m² subplots were used to measure the understory. Eight sites were located throughout Olympic National Forest (Carey and Harrington 2001). Vegetation was sampled in each 100 m² subplot at two sites in 2002 (3 years after treatment) and two additional sites in 2004 (7 years after treatment). Pretreatment data were not available at the necessary scale. Because the GIS vegetation data for the current project is from before the treatment dates, we only included subplots from the control and areas of the treatment plots that were not thinned. This left 60 total subplots in two plots at each of the four sites.

Contact: Leslie Brodie, Olympia Forestry Sciences Laboratory, USDA Forest Service lbrodie@fs.fed.us

Carey, A.B.; Harrington, C.A. 2001. Small mammals in young forests: implications for management for sustainability. *Forest Ecology and Management*. 154: 289–309.

Reutebuch, S.E., Harrington, C.A., Marshall, D.D., & Brodie, L.C. 2004. Use of large-scale silvicultural studies to evaluate management options in Pacific Northwest forests of the United States. *Forest Snow and Landscape Research* 78 (1-2): 191-208

Olympic National Forest Ecology Plots

These are plots that were used to identify the forested plant associations of the Olympic National Forest (Henderson et al. 1989). Plots ranged in size from 250 m² to ~1600 m² and were sampled mostly between 1980 and 1986. Plots were spread throughout the National Forest. Some plots have been resampled as recently as 2003. We only included the 21 plots where one of the nine species was present.

Contact: Robin Leshner, Mt. Baker-Snoqualmie NF, rllesher@fs.fed.us

Henderson, J.A., D.H. Peter, R.D. Leshner, D.C. Shaw. 1989. Forested plant associations of the Olympic National Forest. USDA Forest Service Region 6 Ecology Technical Paper 001-88.

Vegetation and Landform Database Development Study (PMR)

The purpose of these plots was to assess the accuracy of vegetation classification maps based on satellite imagery for Olympic National Park. Understory vegetation was sampled in 1/40 acre plots (~101 m²).

Plots were systematically placed along 1.5 mile long transects. The transects were located to best capture the variation in the vegetation of the park as determined by remote sensing. All species with >5% cover and up to five rarer species were recorded. Because not all species were recorded, the absence of a species in the dataset does not indicate that it was not really there. Therefore only positive occurrences of the target species were included. Only one plot contained any of the target species. Data were collected in 1997.

Data and information about this dataset are held by the staff at Olympic National Park

ONP Herbarium Specimen Locations

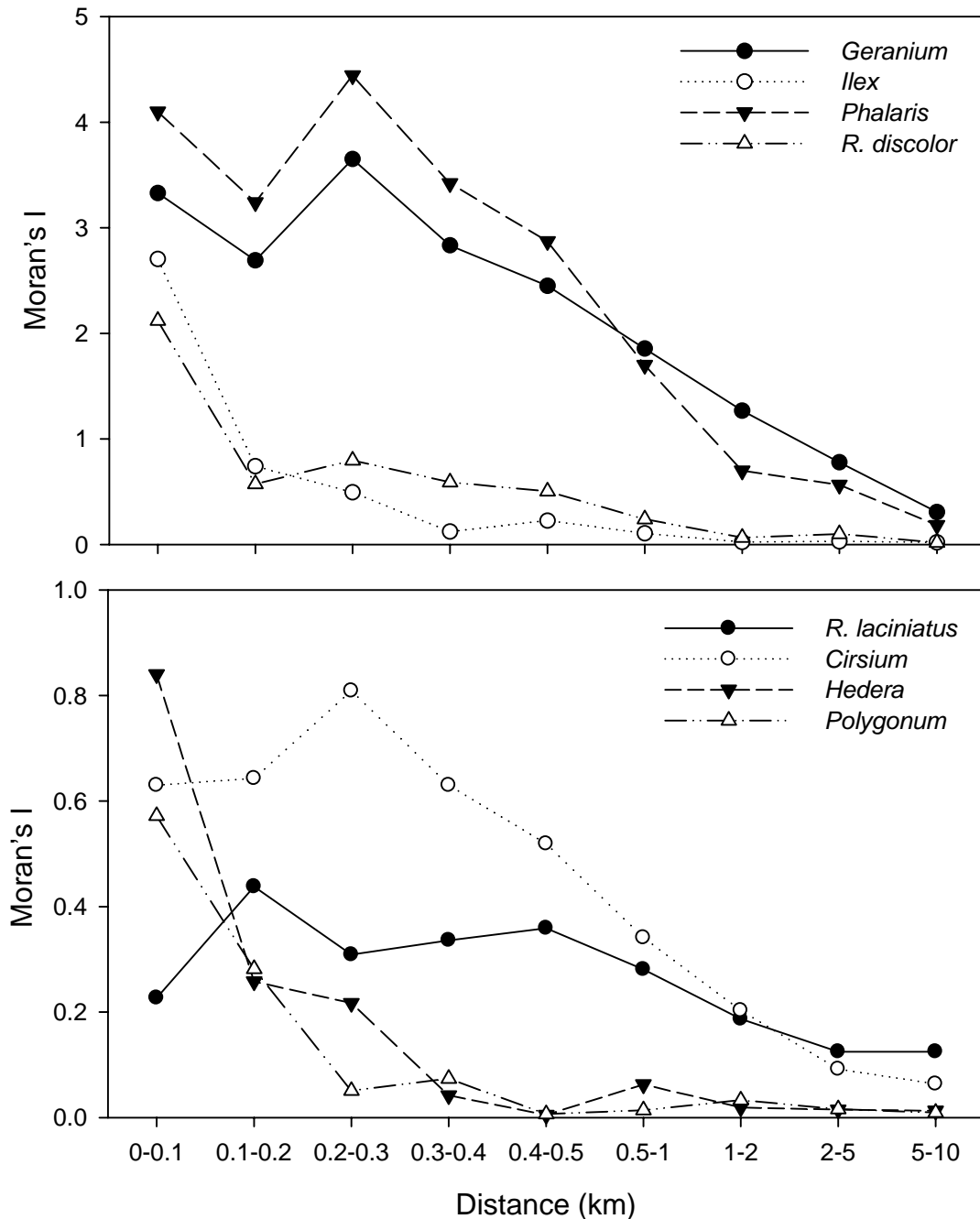
These are locations where specimens in the Olympic National Park Herbarium were collected. We geo-referenced them based on the locality information in the Herbarium database. Locations were only included if we could locate them within 50 m. A total of 5 records were included.

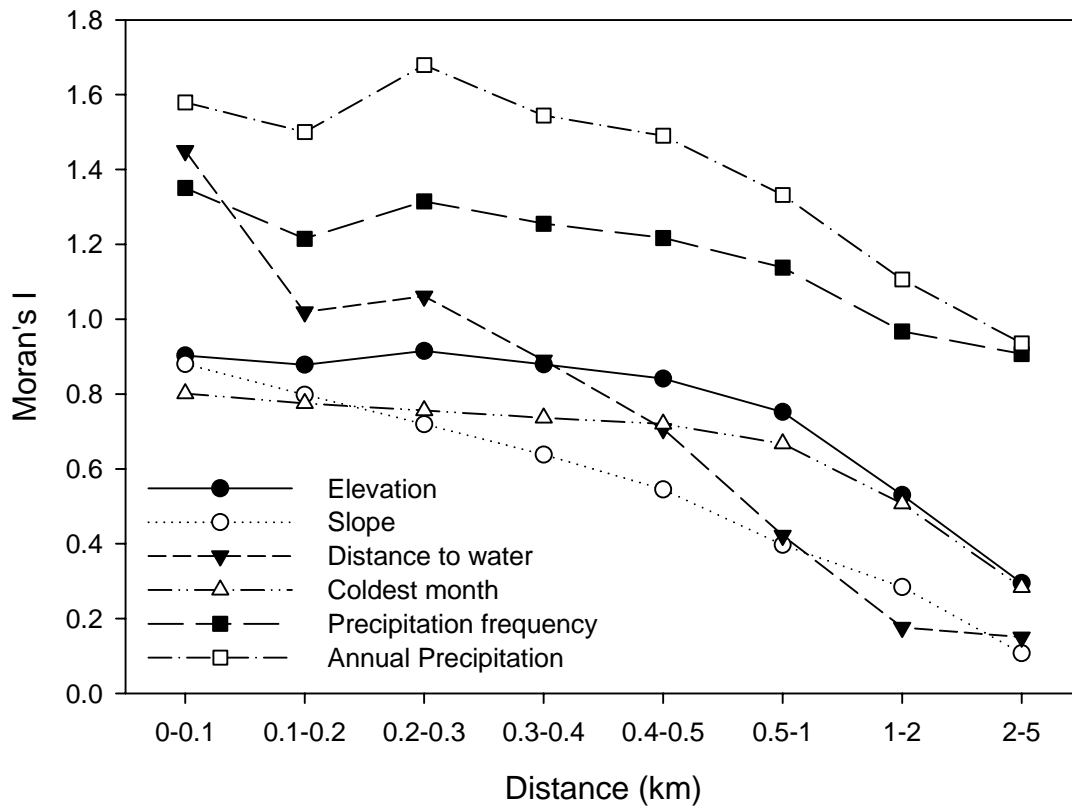
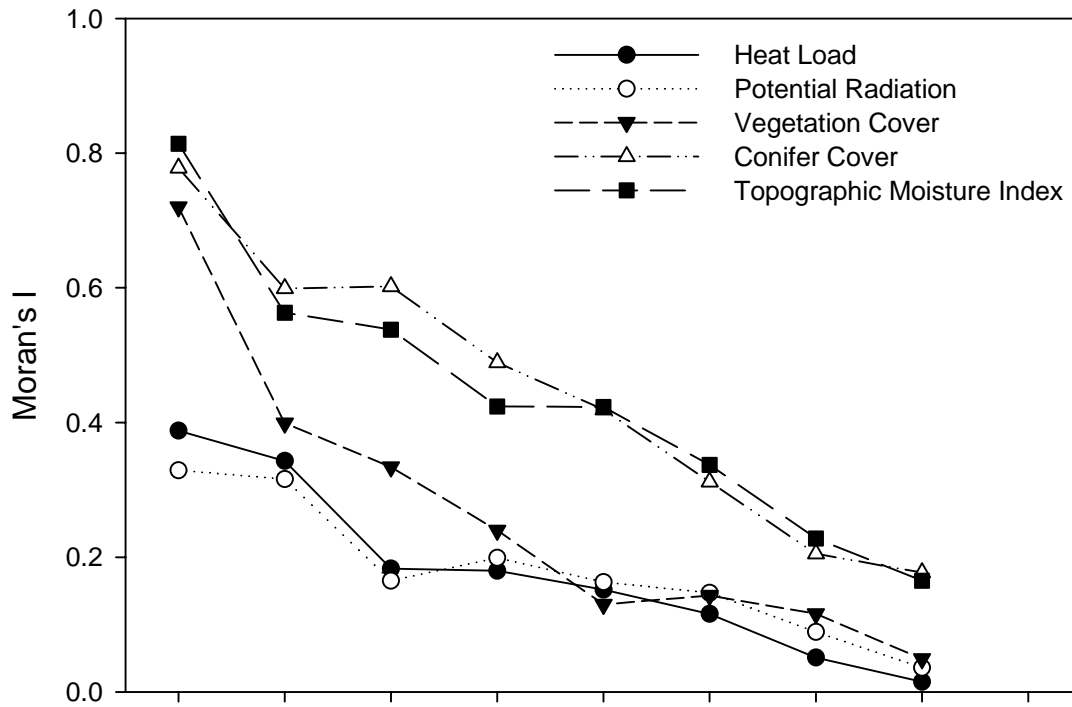
Personal Observations

These points are location on the Olympic Peninsula where we personally observed the species. Locations were recorded relative to roads, trails or other landmarks within 25 m. Four locations were included.

Appendix B: Autocorrelation in Habitat Variables and Species Distributions

We calculated Moran's I in nine distance classes (0-100, 100-200, 200-300, 300-400, 400-500, 500-1000, 1000-2000, 2000-5000 and 5000-10000 m) for each habitat variable and each species (Note that *Phalaris arundinacea*, *Polygonum cuspidatum* and *Rubus discolor* are included here even though they are not the subject of this report). Number of frost days, humidity and climate-based radiation were accidentally omitted while elevation and mean temperature of the coldest month were included. Moran's I usually ranges from 1 (strong positive autocorrelation) to -1 (strong negative correlation) although in some cases larger and smaller values are possible. Data points include all locations in the species dataset (n=4142)

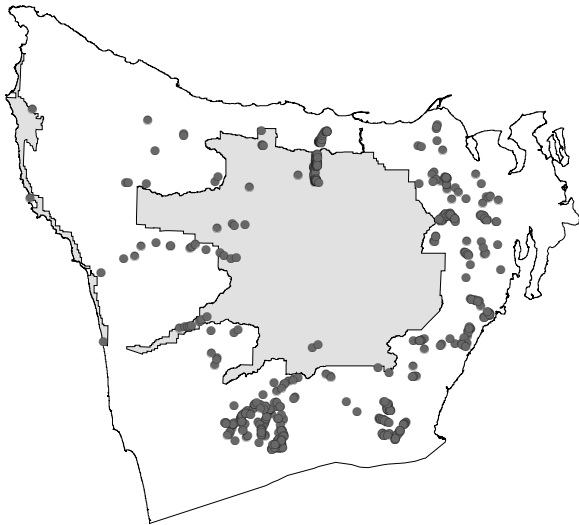




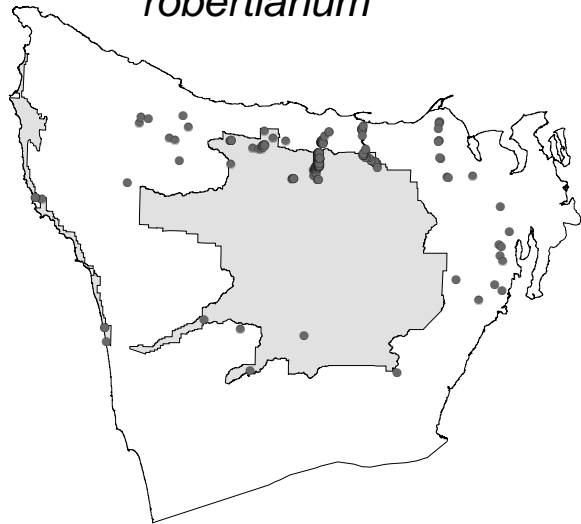
Appendix C. Current Distributions of the Five Target Species on the Olympic Peninsula

Dark gray circles indicate known locations. Shading indicates Olympic National Park.

Cirsium arvense



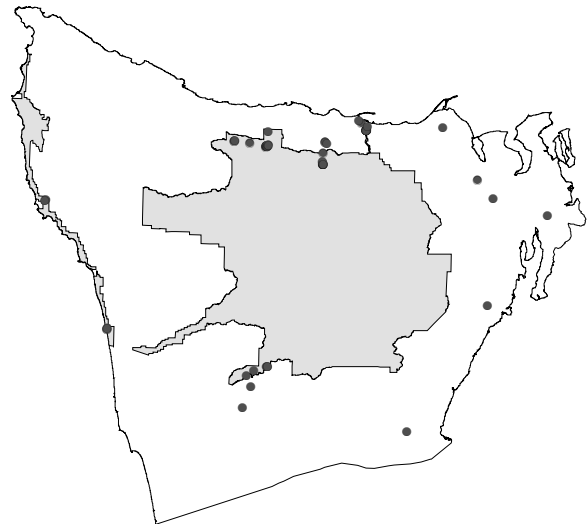
Geranium robertianum



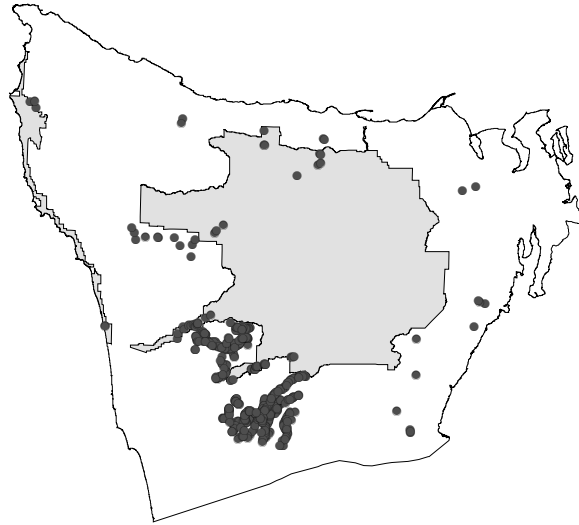
Hedera helix



Ilex aquifolium



Rubus laciniatus



Appendix D. Habitat Variable GIS Layers

We created or adapted GIS layers for the 24 variables that are listed below (*italics*). Of these 12 (**bold italics**) were used in modeling. Sources and methods for developing the layers are described below.

Climate Variables

All of the following climate layers were developed by the Numerical Terradynamic Simulation Group at the University of Montana using the DAYMET model (www.daymet.org, Thornton et al. 1997). The model is based on a digital elevation model and observations from meteorological stations over an 18 year period (1980-1997). The grids have a 1 km resolution grid for the contiguous USA. The original projection was Lambert Equal-Area Azimuthal (central meridian -100; central parallel 45).

For all climate layers, we clipped them to a rectangle encompassing the Olympic Peninsula, reset the cell size to 25 m and re-projected them to UTM Zone 10 NAD 1983. Changing the cell size to 25 m prior to projection reduced the distortion caused by projecting a grid.

Number of frost days

No change from the DAYMET layer. This is the number of days in the year with average minimum temperatures of 0° C or less.

Annual precipitation

No change from the DAYMET layer

Precipitation Frequency

No change from the DAYMET layer. This is the proportion of days with >0 precipitation.

Humidity

No change from the DAYMET layer. This layer is the average over the year of the daily partial pressure of water vapor near the surface. Humidity was modeled based on precipitation and temperature values (Thornton et al. 1999, Thornton et al. 2000).

Incident solar radiation

No change from the DAYMET layer. This layer is the annual average of the daily total shortwave radiation flux as estimated based on temperature, precipitation and humidity values (Thornton et al. 1999, Thornton et al. 2000).

Mean temperature of the coldest month

We compared the mean temperatures for individual months (Nov-Feb). For each grid cell the value of the coldest month. For the Olympic Peninsula the coldest month for all grid cells was December.

Mean temperature of the warmest month

We compared mean temperatures for June – September as explained above. In all cases for the Olympic Peninsula, August was the warmest month.

Growing degree days

No change from the DAYMET layer. This layer is the sum of the daily mean temperatures for all days in the year with mean temperatures above 0° C.

Temperature seasonality

This is the difference between the mean temperature of the warmest month and the mean temperature of the coldest month, calculated for each cell.

Summer precipitation

The sum of the monthly precipitation for July – September

Precipitation of the driest month

We compared monthly precipitation and assigned each cell the lowest value for any month. For the Olympic Peninsula the lowest values were split between July and August (about 2/3 August).

Topographic Variables

All of these variables are based on the USGS Digital Elevation Model (DEM), accessed via the University of Washington:

<http://duff.geology.washington.edu/data/raster/tenmeter/onebytwo10/index.html>

Original Projection: Zone 10 NAD 1927

Elevation

We used the original 10 m DEM with elevation in decimeters. We re-projected the DEM into NAD 1983 and rounded to the nearest meter. We then resampled to 25 m using bilinear interpolation.

Slope

We used ArcView 3.2 to create a slope grid based on the 10 m DEM.

Potential solar radiation

We calculated the slope, aspect and latitude (to the nearest 0.001 degree) for each 10 m cell. We then calculated the potential solar radiation using equations from McCune and Keon (2002). We then resampled to 25 m using bilinear interpolation.

Heat load

This is similar to potential solar radiation (and was developed using the same procedure) except that it reaches its maximum on southwest rather than south facing slopes.

Topographic moisture index

This is calculated for each grid cell as $\ln(A/\tan B)$ where A is the catchment area or the area from which water will flow to the selected grid cell, and B is the slope of the grid cell (Moore et al. 1993).

An ArcView extension for calculating this index was obtained from Tim Loesch, Minnesota DNR

www.dnr.state.min.us/mis/gis/tools/arcview/training.html

Soil Variables

Soil layers were based on data from STATSGO

<http://www.ncgc.nrcs.usda.gov/branch/ssb/products/statsgo/>

Original Projection: Albers Equal Area NAD 1927

Polygon vector data at 1:250,000 scale

We re-projected the layers into UTM Zone 10 NAD 1983 and clipped to the boundary of Olympic peninsula. The landscape in the STATGO layers is divided up into map units (which can include one or

more polygons) and for each map unit there are a number of soil components that make up the map unit. The soil components are not mapped but the proportion of the map unit that they represent is recorded.

For each soil component data on the minimum and maximum depth to water table and minimum and maximum pH of the top soil layer were available in data tables. For each map unit we calculated the mean of these four variables weighted by proportion of the map unit that each soil component represents. Through this procedure we calculated one value of min and max depth to water table and min and max pH for each map unit.

The min and max depth to water table are strongly correlated ($r = 0.98$) while the min and max pH are less strongly correlated ($r = 0.67$).

maximum depth to water table – We chose this measure because *Cirsium*, *Geranium*, *Hedera* and *Ilex* are limited in saturated soils. If maximum depth to water table is too small then the species would not be likely to occur within the polygon.

maximum pH – We chose this variable because *Geranium* and to some extent *Hedera* are limited by acidic soils. Presumably if max pH is too low then the species will not occur within the polygon.

However, for the Olympic Peninsula there are 37 map units. Because the spatial resolution is low and we averaged among soil components within map units these layers were not useful for this project.

Distance to Nearest Water

distance to water – This variable was created by combining two layers, one for wetlands and one for streams, and then calculating the base 10 logarithm of the distance to the nearest wetland or stream. This variable was log transformed because differences over the first 50 m are likely to be the most important.

We obtained the wetlands polygon layer obtained from the National Wetland Inventory via the ONRC Clearinghouse

http://www.onrc.washington.edu/clearinghouse/themes/hydro/hydro_theme.html

www.nwi.fws.gov

Original Projection: UTM Zone 10 NAD 1927

Wetland polygons were classified by hierarchical wetland type. We removed all upland polygons as well as all polygons without classification or outside the boundaries of the Olympic Peninsula. We then removed all polygons that indicated open-water marine and estuarine areas that are always flooded even at low tide (M1OWL and E10WL).

The streams layer was obtained from Olympic National Forest

<http://www.fs.fed.us/r6/data-library/gis/olympic/index.html>

Projection: UTM Zone 10 NAD 1927

This is a line layer with all of the stream segments on the Olympic Peninsula

Vegetation Variables

There were two available sources for vegetation data for the Olympic Peninsula:

- 1) Pacific Meridian Resources (PMR) Datasets for ONP

http://www.nps.gov/gis/park_gisdata/washington/olymp.htm

Original Projection: UTM Zone 10 NAD 1927

The canopy cover layer includes all tree species and is broken down into categories of 11-40%, 41-70%, and 71-100% as well as other types of cover (e.g. shrubs, meadow, water, barren, etc.). Remote Sensing Data are from 1992

Because this layer only contains data for ONP and a small buffer around it, we did not use it for this project.

2) Interagency Vegetation Mapping Project (IVMP) Datasets for the peninsula as a whole.

See the website at: <http://www.blm.gov/or/gis/index.php>

Original Projection: UTM Zone 10 NAD 1927

Four layers are available: vegetation cover, conifer cover, broadleaf cover (includes shrubs and herbs: essentially vegetation cover – conifer cover), quadratic mean diameter (estimate of average tree size)

They recommend only dividing these into 3 categories. About 2% of the cells are classified as unknown because of topographic shadow, smoke, clouds or other problems in the images. Remote Sensing Data are Landsat 5 TM images from 1996

conifer cover – We used the conifer cover layer from the IVMP dataset. We converted this layer into 6 classes: 0 = 0% conifer cover (or classified as non-forested), 1 = 1-20% conifer cover, 2 = 21-40% conifer cover, 3 = 41-60% conifer cover, 4 = 61-80% conifer cover and 5 = 81-100% conifer cover. There were some areas where conifer was not calculated (e.g., due to topographic shadow) and these areas were removed from the study. Accuracy for conifer cover with these intervals was only 64.6%

vegetation cover – For this we used the vegetation cover layer, which measures all vegetation (herbs, shrubs, deciduous and conifer trees). We divided the cover values into the same categories as used for the conifer cover layer. Accuracy is 87.9% for 20% intervals.

Quadratic Mean Diameter (QMD) – this layer has the modeled estimate of QMD for areas on the Olympic Peninsula with >30% conifer cover and >70% total vegetation cover. We divided the values for QMD (which range from 0-75 inches) into three categories: 0 (which includes areas with <70% vegetation, <30% conifer cover, and non-forested areas), 1-20, and 21-75 as recommended by the IVMP. Accuracy with only two categories (0-20 and >20) accuracy was 86.7%. Modeling in some areas (about 7.5% of total cells) resulted in obviously wrong estimates and were excluded from the layer by IVMP. We set these as well as the topographic shadow cells to null. Because large areas have no estimate for QMD, we did not use this layer in this project

Land Use

National Land Cover Data of 1992

<http://landcover.usgs.gov/natl/landcover.asp>

<http://edc.usgs.gov/products/landcover/nlcd.html>

This layer can be downloaded in either Geographic NAD 83 or Albers Conic Equal Area NAD 83. Cell size is 30 m. We projected the layer to UTM NAD 83 and resampled to 25 m cell size. Both of these steps introduce some error into the layer.

There are 21 land cover classes which we condensed into the following 13 (Definitions adapted from the NLCD web site). Because the statistical modeling techniques would have required the use of 12 dummy variables to include land use into the model we did not use this layer.

1. *Open Water* - all areas of open water, generally with less than 25% cover of vegetation/land cover.
2. *Perennial Ice/Snow* - all areas characterized by year-long surface cover of ice and/or snow.
3. *Developed* – Areas characterized by a high percentage (30 percent or greater) of constructed materials (e.g. asphalt, concrete, buildings, etc). Includes
4. *Barren* - Areas characterized by bare rock, gravel, sand, silt, clay, or other earthen material, with little or no "green" vegetation present regardless of its inherent ability to support life
5. *Deciduous Forest* - Areas dominated by trees where 75 percent or more of the tree species shed foliage simultaneously in response to seasonal change.
6. *Evergreen Forest* - Areas dominated by trees where 75 percent or more of the tree species maintain their leaves all year. Canopy is never without green foliage.
7. *Mixed Forest* - Areas dominated by trees where neither deciduous nor evergreen species represent more than 75 percent of the cover present.
8. *Shrubland* - Areas characterized by natural or semi-natural woody vegetation with aerial stems, generally less than 6 meters tall, with individuals or clumps not touching to interlocking. Both evergreen and deciduous species of true shrubs, young trees, and trees or shrubs that are small or stunted because of environmental conditions are included. Shrub canopy accounts for 25-100 percent of the cover.
9. *Non-Natural Woody* - Areas dominated by non-natural woody vegetation; non-natural woody vegetative canopy accounts for 25-100 percent of the cover (e.g. orchards, vineyards).
10. *Herbaceous Upland* - Upland areas characterized by natural or semi-natural herbaceous vegetation; herbaceous vegetation accounts for 75-100 percent of the cover.
11. *Planted/Cultivated* - Areas characterized by herbaceous vegetation that has been planted or is intensively managed for the production of food, feed, or fiber; or is maintained in developed settings for specific purposes. Herbaceous vegetation accounts for 75-100 percent of the cover. This category includes pastures, row crops, and parks.
12. *Woody Wetlands* - Areas where forest or shrubland vegetation accounts for 25-100 percent of the cover and the soil or substrate is periodically saturated with or covered with water.
13. *Emergent Herbaceous Wetlands* - Areas where perennial herbaceous vegetation accounts for 75-100 percent of the cover and the soil or substrate is periodically saturated with or covered with water.

Distance to developed area – we also created a layer measuring the distance to the nearest cell classified as developed. This was done by creating a polygon theme of all developed areas prior to projecting the land cover data. The polygon theme was then projected to UTM NAD 1983 and a distance grid with cell size of 25 m was calculated from the polygon theme

Trails and Roads (for examining the role of dispersal corridors)

Data for roads and trails came from three sources:

1) A layer showing the roads and a GPS verified layer showing the trails within the boundaries of Olympic National Park were obtained from park staff.

Projection: UTM Zone 10, NAD 83

2) Both trail and road coverages for Olympic National Forest were available from the USDA Forest Service web site:

<http://www.fs.fed.us/r6/data-library/gis/olympic/index.html>

Projection: UTM Zone 10 NAD 27

Roads theme has roads extending out of the forest into the park and surrounding areas.

This layer contains information on road name, current status, surface, # of lanes, etc.

The trail theme has all trails within the forest updated to 1996 plus some of the trails in the National Park

3) DNR layer – from the Washington Department of Natural Resources

<http://www3.wadnr.gov/dnrapp6/dataweb/dmmatrix.html>

Projection: Labert Conformal Conic

This layer includes roads, trails, railroads, and ferry crossings. The accuracy has not been assessed and there is likely some misclassification. In addition some roads may be closed. The focus of this layer is on state and private forest lands although transportation segments from all ownerships are included.

One issue that did arise is that for one area where the park expanded the DNR layer says there is a road and the park layer says it is a trail. Probably it was a road before the park was expanded. This means that there may be a number of roads in the DNR layer that are not used as roads anymore. This may explain why the DNR layer has more roads in ONF than the ONF layer (these roads may be closed)

Merging Layers – Roads

In ONP – there are a few minor differences between the DNR layer and the park layer. We used the park layer because there are missing segments in the DNR layer.

In ONF – Both the DNR layer and the ONF layer show roads that are not in the other layer. We calculated the minimum distance to a road in either layer in this area.

Elsewhere – The DNR layer is by far more complete outside of the Park and Forest. No roads show up in the NP and NF layers that do not show up in the DNR layer. For this area we only included the DNR layer

Merging Layers – Trails

In ONP – There are a few areas near the Sol Duc road where the DNR layer shows trails that are not in the ONP layer. There are also a few places where the trails do not match up well between the layers. We are more confident in the ONP layer so we only used it for this area.

In ONF – Both the DNR and ONF layers contain trails that the other does not. This is especially an issue in the northwest section of the forest. However, the DNR trails do not appear to be very precise and do not follow closely with the ONF trails in a lot of cases so we only used the ONF trails in this area.

Elsewhere – Only the DNR layer has trails outside of ONP and ONF

For roads and trails we calculated two different distance grids for the models:

1) Distance from nearest road or trail– This is the minimum distance from the nearest road or trail based on a combination of all three road layers and both trail layers. Within Olympic National Park the ONP layer was used. Within Olympic National Forest both the DNR and ONF layers were used, and everywhere else only the DNR layer was used.

Appendix E. File Names and Organization for GIS Layers

Folders are in **bold**, GIS layers in *italics*.

Olympic_Exotics_GIS **species_data**

All_species – a shapefile showing the locations of all species data points from the 13 datasets. The attribute table contains the following fields (in *italics*)

Plot Data

Plot_num – a unique identifier for each plot

X – X coordinate for plot location based on UTM 10 NAD 83

Y – Y coordinate for plot location based on UTM 10 NAD 83

Source – The original dataset from which the data point is taken (DEFERRARI = Collette DeFerrari dataset, ELWHAEXOTICS = Elwha Exotics Survey, ELWHALTEM = LTEM Elwha plots, EXOTICMNGMT = Exotic Plant Management Team, HDS = Olympic Habitat Development Study, LEZBERG = Ann Lezberg dataset, NPSEXOTICS = Olympic NP Exotics Inventory, ONFECOLOGY = Olympic NF Ecology Plots, ONFEXOTICS = Olympic NF Exotics Survey, ONPHERBARIUM = Olympic NP Herbarium Records, PERSOBS = Personal Observations, PMRRECON = Vegetation and Landform Database Development Study, SCHREINER = Alpine Plant Community Survey)

Date – Year of sampling

Size – plot size in m² (blank indicates no plot)

Info – additional information about the plot from the original dataset

Species Data – the following nine fields contain presence and absence data for each species (1=present, 0=absent, blank=no data).

Ciar – *Cirsium arvense*

Gero – *Geranium robertianum*

Hehe – *Hedera helix*

Ilaq – *Ilex aquifolium*

Phar – *Phalaris arundinacea*

Pocu – *Polygonum cuspidatum*

Posa – *Polygonum sachalinense*

Rudi – *Rubus discolor*

Rula – *Rubus laciniatus*

Habitat Variables – the last 24 fields contain the values of the potential habitat variables at each datapoint. The first 12 of these fields are the variables used in the final models

Ann_precip – annual precipitation

Con_cov – conifer cover in intervals (0, 1-20, 21-40, 41-60, 61-80, 81-100%)

Frost_days – number of days with temperatures below freezing

Heat_load – calculated based on slope, aspect and latitude

Humidity – based on temperature and precipitation

Log_dis_wat – base 10 logarithm of distance to nearest wetland or stream

Pot_rad – potential radiation calculated from slope, aspect and latitude

Pre_freq – proportion of days with measurable precipitation

Radiation – based on temperature and precipitation

Slope – in degrees

Tmi – topographic moisture index

Veg_cov – total vegetation cover in intervals (0, 1-20, 21-40, 41-60, 61-80, 81-100%)

Cold_mon – mean temperature of the coldest month
Dist_rd_tr – distance to the nearest road or trail
Driest_mo – precipitation for the driest month of the year
Elev – in meters
Gdd – growing degree days, sum of the mean temperature on all days above 0° C
Land_cover – 13 cover classes based on the National Land Cover Data of 1992
Max_ph – maximum pH in a soil map unit
Qmd – Quadratic Mean Diameter; average for conifer trees
Summer_pre – precipitation from July – September
Temp_range – temperature seasonality, the difference between *cold_month* and *warm_month*
Warm_mon – mean temperature of the warmest month
Wat_table – maximum depth to water table within a soil map unit

ciar_dist – a shapefile showing the locations of presence and absence points for *Cirsium* from the 13 datasets. All points are at least 100 m from the nearest neighbor. The attribute table contains the following fields (in italics)

Plot Data

Plot_num – a unique identifier for each plot
X – X coordinate for plot location based on UTM 10 NAD 83
Y – Y coordinate for plot location based on UTM 10 NAD 83
Source – The original dataset from which the data point is taken (DEFERRARI = Collette DeFerrari dataset, ELWHAEXOTICS = Elwha Exotics Survey, ELWHALTEM = LTEM Elwha plots, EXOTICMNGMT = Exotic Plant Management Team, HDS = Olympic Habitat Development Study, LEZBERG = Ann Lezberg dataset, NPSEXOTICS = Olympic NP Exotics Inventory, ONFECOLOGY = Olympic NF Ecology Plots, ONFEXOTICS = Olympic NF Exotics Survey, ONPHERBARIUM = Olympic NP Herbarium Records, PERSOBS = Personal Observations, PMRRECON = Vegetation and Landform Database Development Study, SCHREINER = Alpine Plant Community Survey)
Date – Year of sampling
Size – plot size in m² (blank indicates no plot)
Info – additional information about the plot from the original dataset
Ciar – *Cirsium arvense* presence or absence (1=present, 0=absent)
Group – Each presence and absence was randomly assigned to one of five equal groups for model validation

Habitat Variables – the next 12 fields contain the values of the habitat variables used in the final models

Ann_precip – annual precipitation
Con_cov – conifer cover in intervals (0, 1-20, 21-40, 41-60, 61-80, 81-100%)
Frost_days – number of days with temperatures below freezing
Heat_load – calculated based on slope, aspect and latitude
Humidity – based on temperature and precipitation
Log_dis_wat – base 10 logarithm of distance to nearest wetland or stream
Pot_rad – potential radiation calculated from slope, aspect and latitude
Pre_freq – proportion of days with measurable precipitation
Radiation – based on temperature and precipitation
Slope – in degrees
Tmi – topographic moisture index
Veg_cov – total vegetation cover in intervals (0, 1-20, 21-40, 41-60, 61-80, 81-100%)

Model results – the last fields contain habitat suitability values at the location of the data point for each model replicate

EnfaX (X ranges from 1-5) – habitat suitability for the X run of the ENFA model (suitability values range from 0-100)

LrX (X ranges from 1-5) – habitat suitability for the X run of the logistic regression model (suitability values range from 0-1)

GarpX (X ranges from 1-5) – habitat suitability for the X run of the GARP model (suitability values range from 0-100)

gero_dist – a shapefile showing the locations of presence and absence points for *Geranium* from the 13 datasets. All points are at least 100 m from the nearest neighbor. The attribute table contains the same fields as *ciar_dist* with two exceptions

Gero (replaces *Ciar*) – *Geranium robertianum* presence or absence (1=present, 0=absent)

Cli_env – predicted suitability based on the ‘invaded-range’ model (1=suitable, 0=unsuitable)

hehe_dist – a shapefile showing the locations of presence and absence points for *Hedera* from the 13 datasets. All points are at least 100 m from the nearest neighbor. The attribute table contains the same fields as *ciar_dist* with three exceptions

Hehe (replaces *Ciar*) – *Hedera helix* presence or absence (1=present, 0=absent)

Lit – predicted suitability from the ‘native-range’ model (1=suitable, 0=unsuitable)

Cli_env – predicted suitability based on the ‘invaded-range’ model (1=suitable, 0=unsuitable)

ilaq_dist – a shapefile showing the locations of presence and absence points for *Ilex* from the 13 datasets. All points are at least 100 m from the nearest neighbor. The attribute table contains the same fields as *ciar_dist* with three exceptions

Ilaq (replaces *Ciar*) – *Ilex aquifolium* presence or absence (1=present, 0=absent)

Lit – predicted suitability from the ‘native-range’ model (1=suitable, 0=unsuitable)

Cli_env – predicted suitability based on the ‘invaded-range’ model (1=suitable, 0=unsuitable)

rula_dist – a shapefile showing the locations of presence and absence points for *Rubus* from the 13 datasets. All points are at least 100 m from the nearest neighbor. The attribute table contains the same fields as *ciar_dist* with one exception

Rula (replaces *Ciar*) – *Rubus laciniatus* presence or absence (1=present, 0=absent)

potential_habitat_layers – these are the 24 potential habitat variables considered for use in models.

They are generally clipped to the Olympic Peninsula but not to the same extent

annual_precip – annual precipitation

cold_month – mean temperature of the coldest month

con_cov – conifer cover in intervals (0, 1-20, 21-40, 41-60, 61-80, 81-100%)

dist_rd_tr – distance to the nearest road or trail

driest_month – precipitation for the driest month of the year

elevation – in meters

frost_days – number of days with temperatures below freezing

gdd – growing degree days, sum of the mean temperature on all days above 0° C

heat_load – calculated based on slope, aspect and latitude

humidity – based on temperature and precipitation

land_cover – 13 cover classes based on the National Land Cover Data of 1992

log_dist_wat – base 10 logarithm of distance to nearest wetland or stream

max_ph – maximum pH in a soil map unit

pot_rad – potential radiation calculated from slope, aspect and latitude

precip_freq – proportion of days with measurable precipitation

qmd – Quadratic Mean Diameter; average for conifer trees
radiation – based on temperature and precipitation
slope – in degrees
summer_precip – precipitation from July – September
temp_range – temperature seasonality, the difference between *cold_month* and *warm_month*
tmi – topographic moisture index
veg_cov – total vegetation cover in intervals (0, 1-20, 21-40, 41-60, 61-80, 81-100%)
warm_month – mean temperature of the warmest month
water_table – maximum depth to water table within a soil map unit

final_habitat_layers – these are the 12 final habitat layers used in the model. They are all clipped to the same extent

annual_precip – annual precipitation
con_cov – conifer cover in intervals (0, 1-20, 21-40, 41-60, 61-80, 81-100%)
frost_days – number of days with temperatures below freezing
heat_load – calculated based on slope, aspect and latitude
humidity – based on temperature and precipitation
log_dist_wat – base 10 logarithm of distance to nearest wetland or stream
pot_rad – potential radiation calculated from slope, aspect and latitude
precip_freq – proportion of days with measurable precipitation
radiation – based on temperature and precipitation
slope – in degrees
tmi – topographic moisture index
veg_cov – total vegetation cover in intervals (0, 1-20, 21-40, 41-60, 61-80, 81-100%)

Models

Cirsium

ciar_enfa_x (x ranges from 1-5) – ENFA model replicates for *Cirsium*. Grid values range from 0 – 100
ciar_enfa_mn – Mean values from the five replicates. This is the layer used to produce the map.
ciar_garp_x (x ranges from 1-5) – GARP model replicates for *Cirsium*. Grid values range from 0 – 100
ciar_garp_mn – Mean values from the five replicates. This is the layer used to produce the map.
ciar_lr_x (x ranges from 1-5) – logistic regression model replicates for *Cirsium*. Grid values range from 0 – 100
ciar_lr_mn – Mean values from the five replicates. This is the layer used to produce the map.

Geranium

gero_enfa_x (x ranges from 1-5) – ENFA model replicates for *Geranium*. Grid values range from 0 – 100
gero_enfa_mn – Mean values from the five replicates. This is the layer used to produce the map.
gero_garp_x (x ranges from 1-5) – GARP model replicates for *Geranium*. Grid values range from 0 – 100
gero_garp_mn – Mean values from the five replicates. This is the layer used to produce the map.
gero_lr_x (x ranges from 1-5) – logistic regression model replicates for *Geranium*. Grid values range from 0 – 100
gero_lr_mn – Mean values from the five replicates. This is the layer used to produce the map.

Hedera

hehe_enfa_x (x ranges from 1-5) – ENFA model replicates for *Hedera*. Grid values range from 0 – 100

hehe_enfa_mn – Mean values from the five replicates. This is the layer used to produce the map.

hehe_garp_x (x ranges from 1-5) – GARP model replicates for *Hedera*. Grid values range from 0 – 100

hehe_garp_mn – Mean values from the five replicates. This is the layer used to produce the map.

hehe_lr_x (x ranges from 1-5) – logistic regression model replicates for *Hedera*. Grid values range from 0 – 100

hehe_lr_mn – Mean values from the five replicates. This is the layer used to produce the map.

Ilex

ilaq_enfa_x (x ranges from 1-5) – ENFA model replicates for *Ilex*. Grid values range from 0 – 100

ilaq_enfa_mn – Mean values from the five replicates. This is the layer used to produce the map.

ilaq_garp_x (x ranges from 1-5) – GARP model replicates for *Ilex*. Grid values range from 0 – 100

ilaq_garp_mn – Mean values from the five replicates. This is the layer used to produce the map.

ilaq_lr_x (x ranges from 1-5) – logistic regression model replicates for *Ilex*. Grid values range from 0 – 100

ilaq_lr_mn – Mean values from the five replicates. This is the layer used to produce the map.

Rubus

rula_enfa_x (x ranges from 1-5) – ENFA model replicates for *Rubus*. Grid values range from 0 – 100

rula_enfa_mn – Mean values from the five replicates. This is the layer used to produce the map.

rula_garp_x (x ranges from 1-5) – GARP model replicates for *Rubus*. Grid values range from 0 – 100

rula_garp_mn – Mean values from the five replicates. This is the layer used to produce the map.

rula_lr_x (x ranges from 1-5) – logistic regression model replicates for *Rubus*. Grid values range from 0 – 100

rula_lr_mn – Mean values from the five replicates. This is the layer used to produce the map.

risk_classification

ciar_index – Number of statistical model runs predicting suitability for *Cirsium* in each grid cell (0-15)

ciar_risk – Risk-classification for *Cirsium* for the entire peninsula (1=minimal risk, 2=low risk, 3=moderate risk, 4=high risk)

ciar_risk_np – Risk-classification for *Cirsium* for ONP (1=minimal risk, 2=low risk, 3=moderate risk, 4=high risk)

gero_index – Number of statistical model runs predicting suitability for *Geranium* in each grid cell (0-15)

gero_ce – ‘invaded-range’ model predictions for *Geranium* (1=suitable, 2=unsuitable)

- gero_risk* – Risk-classification for *Geranium* for the entire peninsula (1=minimal risk; 2=low current risk, moderate longer-term risk; 3=low current risk, high longer-term risk; 4=moderate risk, 5=high risk)
- gero_risk_np* – Risk-classification for *Geranium* for ONP (1=minimal risk; 2=low current risk, moderate longer-term risk; 3=low current risk, high longer-term risk; 4=moderate risk, 5=high risk)
- hehe_index* – Number of statistical model runs predicting suitability for *Hedera* in each grid cell (0-15)
- hehe_ce* – ‘invaded-range’ model predictions for *Hedera* (1=suitable, 2=unsuitable)
- hehe_lit* – ‘native-range’ model predictions for *Hedera* (1=suitable, 2=unsuitable)
- hehe_risk* – Risk-classification for *Hedera* for the entire peninsula (1=minimal risk; 2=low current risk, moderate longer-term risk; 3=low current risk, high longer-term risk; 4=moderate risk, 5=high risk)
- hehe_risk_np* – Risk-classification for *Hedera* for ONP (1=minimal risk; 2=low current risk, moderate longer-term risk; 3=low current risk, high longer-term risk; 4=moderate risk, 5=high risk)
- ilaq_index* – Number of statistical model runs predicting suitability for *Ilex* in each grid cell (0-15)
- ilaq_ce* – ‘invaded-range’ model predictions for *Ilex* (1=suitable, 2=unsuitable)
- ilaq_lit* – ‘native-range’ model predictions for *Ilex* (1=suitable, 2=unsuitable)
- ilaq_risk* – Risk-classification for *Ilex* for the entire peninsula (1=minimal risk; 2=low current risk, moderate longer-term risk; 3=low current risk, high longer-term risk; 4=moderate risk, 5=high risk)
- ilaq_risk_np* – Risk-classification for *Ilex* for ONP (1=minimal risk; 2=low current risk, moderate longer-term risk; 3=low current risk, high longer-term risk; 4=moderate risk, 5=high risk)
- rula_index* – Number of statistical model runs predicting suitability for *Rubus* in each grid cell (0-15)
- rula_risk* – Risk-classification for *Rubus* for the entire peninsula (1=minimal risk, 2=low risk, 3=moderate risk, 4=high risk)
- rula_risk_np* – Risk-classification for *Rubus* for ONP (1=minimal risk, 2=low risk, 3=moderate risk, 4=high risk)