Modelling the Spread of European Buckthorn in the Region of Waterloo

by
Sarah Endicott

A thesis
presented to the University of Waterloo
in fulfilment of the
thesis requirement for the degree of
Master of Environmental Studies
in
Planning

Waterloo, Ontario, Canada, 2015

Authors Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

European buckthorn (*Rhamnus cathartica* L.*)* is a common invasive species in the Region of Waterloo. Management of buckthorn is expensive and time consuming and as a result it has continued to spread in the Region of Waterloo despite being on the Ontario noxious weed list. Control of buckthorn will require long term management to prevent reinvasion of managed areas. Therefore this project seeks to provide land managers with a model to predict the likely spread of buckthorn over time and space so that the cost of management can be predicted and different management strategies can be compared.

A generalized linear mixed-effects model was used to create a habitat suitability model and a spread model for the probability of buckthorn presence for cells in a grid. The habitat suitability model predicts the presence of buckthorn based on environmental characteristics while the spread model predicts the likelihood of buckthorn invasion based on the suitability of habitat and the presence of buckthorn in the surrounding area. The spread model indicates that the invasion of buckthorn is influenced by the suitability of habitat and the presence of buckthorn in neighbouring cells. The success of the spread model suggests that this approach can be used to create a spatiotemporally explicit model with limited sampling effort. To explore the utility of the spread model for conservation management purposes, a simulation model was created that is based on the spread model. Simulations were performed to test the spread of buckthorn in a sample forest patch and to test different management strategies. The simulations showed that buckthorn can be controlled within a patch with a limited amount of effort.

Acknowledgements

I would like to thank my supervisor Dr. Michael Drescher and my committee member Dr. Alexander Brenning for their guidance and support throughout the entire process of creating my thesis. I would also like to thank Dr. Stephen Murphy for his contributions as the external reader. It was greatly appreciated and I enjoyed their valuable guidance. I would like to thank the volunteers who assisted me in data collection for all of their help; this research would not have been possible without it. Finally, I would like to express my appreciation to my family, and friends for support and assistance along the way.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

This project aims to improve understanding of how the invasive shrub European buckthorn (*Rhamnus cathartica* L.*)* spreads in the Region of Waterloo. The research question is what factors influence the likelihood of an area being invaded by buckthorn over one year and can the process of spread be predicted by a spatio-temporally explicit model. The objectives are to create a model of buckthorn spread that will allow land managers to make informed decisions on what management strategies will be most effective, use simulations to demonstrate how this model could be used, and discuss the policy context for management of invasive species. This thesis begins by introducing the study species and reviewing strategies for modeling invasive species spread. Next, the chosen modeling approach is used to create a habitat suitability model showing the relationship between environmental factors and the presence of buckthorn. Next, a spread model is created using habitat suitability and the presence of nearby buckthorns to predict the likelihood of an area being invaded. The use of the buckthorn spread model to inform land management decisions is demonstrated by simulating the impact of different management strategies on the population of buckthorn and the effort required for management. In addition, legislative and policy strategies are discussed for managing buckthorn on a landscape scale.

Invasive species are species that have spread outside of their native range and cause significant damages to the ecosystems they invade (Government of Canada, 2004). These invasive species cause damage to both the natural environment and the economy. Economic damage caused by invasive species is due to losses to the agricultural, fisheries and forestry industries and requires money to be spent on control efforts. For example a study by Colautti *et al.* (2006) estimated the cost of 16 invasive species for which data was available in Canada to be between $13.3 and $34.5 billion per year. Invasive species can cause many changes to the ecosystems they invade including changes to ecosystem services, ecosystem functioning and extirpation or extinction of native species. The International Union for Conservation of Nature (IUCN) names invasive species as the second largest threat to biodiversity, after habitat loss (IUCN, 2011). In Canada 24% of Species at Risk are threatened by invasive species (Government of Canada, 2004). However, because of considerable uncertainties in the valuation of ecosystem services, it is difficult to provide good cost estimates for invasive species impact beyond agricultural effects.

European buckthorn has negative impacts on natural ecosystems and agriculture. The impacts on natural ecosystems are discussed in the next chapter. European buckthorn has negative impacts on agriculture primarily because it is a host for two crop pests: soybean aphid and oat crown rust. Soybean aphid is considered one of the most significant threats to soybean production in North America (Ragsdale *et al*., 2007). In the 2003 soybean aphid outbreak in Iowa it has been estimated that more than a quarter of a billion dollars was lost due to lost yield and spraying costs (Pilcher *et al*., 2005). A study by Bahlai *et al.*, (2010) found that the ratio of buckthorn density to field area was the most important factor for predicting the early-season density of soybean aphids. Buckthorn is considered to be an important source for oat crown rust spores in temperate areas (Heimpel *et al.*, 2010; United States Department of Agriculture, 2008). Chong *et al.*, 2008 found that fields near buckthorn populations had high severities of buckthorn invasion. According to the USDA oat crown rust is the most damaging disease for oat and during moderate to severe epidemics it can cause between 10 and 40 % yield reduction. More research would be needed to show the impact of removing buckthorn on either oat crown rust or soy bean aphids.

Many invasive species, including European buckthorn, have become so well established in Canada that it would be impractical to attempt to eradicate them entirely (Government of Canada, 2004; Knight *et al.*,2007). Therefore, they are managed by attempting to prevent their spread and mitigate their impacts. One important precondition for managing the impacts of invasive species is to understand how they are likely to spread to in the future (Hastings *et al.*, 2005). If the spread dynamics can be understood then land managers can make informed decisions on when and how to control invasive species. If the spread dynamincs of buckthorn could be well understood it might be possible to decide whether it is worthwhile to implement costly control measures, for example in valuable natural areas when there could be contiued reinvasion from the surrounding landscape. This information would allow land managers to make strong cases for the implementation of buckthorn control measures.

Chapter 2: Literature Review

Study Species

Rhamnus cathartica L. (European buckthorn) (Figure 1) is a dioecious shrub or small tree that is native to Europe, which was introduced to North America as an ornamental species in the early 1800s (Knight *et al.*, 2007). Buckthorn has several features that have made it a successful invader of habitat in North America. Buckthorn can grow in a wide range of habitats in South Central and Eastern Canada and the North Central and Eastern United States (Kurylo *et al.*, 2007). Within this range it occupies a wide variety of sites including open areas, forest edges, forest interiors and drier areas of wetlands (Knight *et al.*, 2007; Kurylo *et al.*, 2007). Buckthorn fruits are small drupes born by female trees; they are roughly 5-7 mm and turn from green to black with ripening (Knight *et al.*, 2007). These fruits contain 2-4 seeds, which have high germination rates ranging from 85% to 100%, which contributes to buckthorn's ability to spread quickly (Archibold *et al.*, 1997; Knight *et al.*, 2007). Buckthorn seeds in the soil can remain viable for up to 5 years (Ontario Invasive Plant Council, 2012).

Figure 1: Common buckthorn leaves
Illustration by Andrea Kingsley
http://www.invadingspecies.com/invaders/plants-terrestrial/common-buckthorn/

Germination is inhibited by the litter layer, herbaceous plant cover, desiccation, water logging and very dense shade (<1% of full light in midsummer) (Knight *et al.*, 2007; Kurylo *et al.*, 2007). Optimal germination rates occur in somewhat shaded, bare, moist soils (Knight *et al.*, 2007). Buckthorn seedlings are especially adept at capitalizing on gaps in the canopy because they can grow in shadier conditions and are more responsive to light when it becomes available than 11 other European shrub species (Grubb *et al.*, 1996). In addition, buckthorn's leaves emerge earlier in spring and senesce later in the fall than those of native species, which allows it to photosynthesize while the canopy is leafless (Archibold *et al.*, 1997; Knight *et al.*, 2007). The characteristics described above help buckthorn to outcompete native species, especially in disturbed areas with bare soil and on the edges of forests (Knight *et al.*, 2007).

Buckthorn seeds either fall under the parent plant or are dispersed farther afield by animals or water (Knight *et al.*, 2007 and Becker *et al.*, 2012). Many bird species have been observed eating buckthorn fruit, and buckthorn stands that are associated with fences and perch trees indicate that the seeds survive the bird's digestive system and are released with their droppings (Knight *et al.*, 2007). The correlation of places where birds perch with buckthorn shrub density indicates that birds are a vector for dispersing buckthorn seeds (Knight *et al.*, 2007). Mice have also been seen to remove buckthorn seeds from under trees, helping the seeds to be dispersed farther (Knight *et al.*, 2007). Animal mediated dispersal contributes to the ability of buckthorn to invade new areas because animals can carry buckthorn seeds long distances to new patches of habitat (Knight *et al.*, 2007).

The ecological impacts of buckthorn make its increasing abundance and swift dispersal a concern for land managers. One of the clearest impacts of buckthorn is that it alters soil chemistry by increasing nitrogen, pH, total carbon, and water content (Knight *et al.*, 2007). These soil chemistry changes have a negative impact on arthropod communities, which could cause problems for birds and mammals that depend on certain arthropods for food (Knight *et al.*, 2007). Evidence also shows that light levels and herbaceous plant cover are lower beneath buckthorn compared to other shrubs, which could impact the growth of native plants (Knight *et al.*, 2007). Controlled experiments with individual buckthorn shrubs did not show a negative impact of buckthorn on native understory plants (Kollman and Grubb 1999). However these results may not be generalizable to buckthorn in a stand of many trees. The experiment used individual shrubs but resource competition and allelopathy are density dependant and may only be apparent under larger stands of buckthorn (Knight *et al.*, 2007). Mascaro and Schnitzer (2007) found that buckthorn was capable of becoming the dominant species at a variety of forested sites in Wisconsin. The result of buckthorn's competitive success indicates that it can supress the growth of native hardwood species, which has potential impacts on forestry and maple syrup industries (Derickx and Antunes, 2013).

In addition to the negative effects caused by buckthorn on native plants, buckthorn has a mutualistic interaction with invasive European earthworms. This is due to the fact that these earthworms prefer buckthorn leaf litter, which is high in nitrogen, and because buckthorn grows well in bare soils created by earthworms (Heimpel *et al.*, 2010). Therefore, if the spread of buckthorn increases, invasive earthworms are likely to spread with it, causing additional negative impacts (Heimpel *et al.*, 2010).

Because buckthorn is a host for crop pests such as the soy bean aphid and oat crown rust, it is considered a noxious weed under Ontario's Weed Control Act (Archibold *et al.*, 1997; Heimpel *et al.*, 2010 and Becker *et al.*, 2012; Government of Ontario, 2012). Because of its ability to disperse widely and dominate habitat, buckthorn is also classified as a category 1 invasive species, meaning that it is a high priority for control (Urban Forest Associates Inc., 2002).

Distribution Modelling

Predicting the distribution of invasive species is important because it allows land managers to allocate their control efforts more effectively (Hastings *et al.*, 2005; Smolik *et al.*, 2010). Predicting the distribution is a complex problem because the spread of invasive species over a landscape is dependent on many factors including the suitability of the habitat, the plant's mode of dispersal, temporal changes in habitat and climate, and the growth of the source population (Leung *et al.* 2012; Hastings *et al.*, 2005). In this thesis the distribution of a species is defined as the specific locations within the study region where the species is present. The spread is the change in the distribution of the species over time. There are two main types of methods for predicting the distribution of invasive species: predicting the suitability of habitat or predicting the spread of the species from a known source. These can also be combined in a hybrid spatially explicit model that considers the suitability of habitat as a component of the dynamics of the species' spread.

Habitat suitability models are created using statistical models, which relate the distribution of species to characteristics of the environment. This allows the probability of a species' occurrence at an unsurveyed site to be estimated (Pearce and Ferrier, 2000). These are often called species distribution models (SDM) and they are used to model the area that could potentially be covered by an invasive species but they do not attempt to model spread. There are many methods that have been used to create SDMs, including logistic regression, generalized linear model (GLM), generalized additive model (GAM), climate envelope (e.g., BIOCLIM), classification and regression tree (CART), neural network (NN), and genetic algorithm (e.g., GARP) (Jeschke and Strayner, 2008). One limitation of SDMs is that they assume that the climate and physical environment are the only factors that affect the distribution of species (Jeschke and Strayner, 2008). In fact, constraints on the dispersal of a species can also have important impacts on their distribution (Jeschke and Strayner, 2008). Dispersal constraints are particularly important for invasive species at local scales because although the climate may be suitable, invasive species may not have had the opportunity yet to disperse to all suitable sites in the local area (Václavík and Meentemeyer, 2009). Despite these limitations SDMs have been used to effectively predict the distribution of many different species including invasive species (Jeschke and Strayner, 2008).

A review of SDMs conducted by Jeshke and Strayner (2008) found that there was no consistently superior method for creating SDMs. For the current analysis GLM was chosen because it incorporates both presence and absence data and is easy to interpret (Hastings *et al.*, 2005). Václavík and Meentemeyer (2009) found that presence and absence data produced better results than presence only data. Because data for this project was collected through structured sampling it was possible to obtain reliable absences, therefore GLM was chosen over other methods that use presence only data. Logistic regression is a specific type of GLM which uses a logit link function and is appropriate for binary, presence/absence data (Hosmer and Lemeshow, 2000).

As mentioned above, one of the problems with using GLM to model the habitat suitability of an invasive species is that the invasive may not have had an opportunity to invade all habitat types that are suitable for it (Jeschke and Strayner, 2008; Smolik *et al.*, 2010). Since habitat suitability is determined based on the types of habitat that are occupied, habitat that has not yet been invaded despite being suitable, may be incorrectly labelled as unsuitable (Smolik *et al.*, 2010). However, this problem is not significant if the species is widely dispersed within the study region and can be assumed to have reached all suitable habitat available (Smolik *et al.*, 2010). My research has found that European buckthorn is very widely distributed in Waterloo Region and therefore this is not expected to be a large problem.

Another issue that arises when using logistic regression to model invasive species distributions is the spatial autocorrelation that is common for this kind of data. Spatial autocorrelation occurs when the distance between points is not independent of the similarity between the points (Dormann *et al.*, 2007). This can be accounted for by using a generalized linear mixed effects model (GLMM), which can incorporate random effects as well as complex spatial correlation structures (Thiele and Markussen, 2012). Random effects are explanatory variables that influence only the variance of the response variable as opposed to fixed effects which influence only the mean (Crawley, 2007; Zuur *et al.* 2009). Random effects can be used to account for grouping of the data by estimating the distribution of the mean of the response variable at each level of the grouping variable (Crawley, 2007; Thiele and Markussen, 2012). This has the advantage of using fewer degrees of freedom than the alternative of estimating a fixed regression coefficient for each level (Crawley, 2007; Thiele and Markussen, 2012). Therefore a GLMM was used to model the suitability of habitat for buckthorn in the Region of Waterloo.

Habitat suitability alone cannot produce spatio-temporally explicit predictions of how buckthorn will spread. Therefore habitat suitability will be combined with a model of buckthorn spread over time. One of the first models of invasive spread was the simple reaction diffusion model, which was developed by Skellam (1951) and since has been applied to many invasive species (Hastings *et al.*, 2005). Frappier *et al* (2003) used this model to predict the linear velocity (V) of spread of buckthorn based on the diffusion coefficient (D) and the intrinsic rate of increase of the population (r):

$$V = 2(rD)^{\frac{1}{2}}$$

Frappier *et al.* (2003) found that the simple reaction diffusion model was able to predict the spread of buckthorn at the scale of a single forest. However, the reaction diffusion model does not account for environmental factors that impact reproduction, environmental heterogeneity, or long-distance dispersal of seeds to new sites, which are important for predicting spread at the landscape level (Hastings *et al.*, 2005; Higgins *et al.*, 1996).

Another method that is often used to predict the spread of invasive species is the discrete time integro-difference model. Integro-difference equations represent the probability of finding an individual at location x at time t based on the sum of the probabilities for all locations y that an individual was produced at y and dispersed to x (Hastings *et al.*, 2005 and Kot *et al.*, 1996). The equation for this is shown below:

$$p_{t+1}(x) = \int p_t(y) f[p_t(y)] k(x,y) dy$$

The dispersal kernel $k(x,y)$ is the probability density function for propagules dispersing from location y to x, and is usually simplified to depend only on the relative distance of dispersal, i.e. $k(x-y)$ (Hastings *et al.*, 2005; Kot *et al.*, 1996). The $f[p_t(y)]$ portion of the equation represents the population growth function and requires estimates of reproduction parameters (Hastings *et al.*, 2005; Kot *et al.*, 1996). The dispersal kernel is estimated by fitting the model to observed data (Hastings *et al.*, 2005). If the dispersal kernel has a Gaussian distribution then the integro-difference model is equivalent to the reaction diffusion model (Hastings *et al.*, 2005; Kot *et al.*, 1996). The integro-difference model allows for non-random dispersal kernels but it still does not often include environmental heterogeneity and does not account for long-distance dispersal (Hastings *et al.*, 2005; Kot *et al.*, 1996).

Another approach that can be used to model spread utilizes grid-based models such as interacting particle systems or cellular automata (Smolik *et al.,* 2010; Fitzgerald *et al.,* 2012). Interacting particle systems operate on a grid of cells and predict the likelihood of each cell transitioning to a different state based on the state of neighbouring cells (Durrett and Levin, 1994). Interacting particle systems are a generalization of cellular automata, which allow for the inclusion of environmental heterogeneity and stochasticity in the transition rules (Smolik *et al.,* 2010). In the case of modeling the spread of an invasive species, the probability of a cell transitioning from uninvaded to invaded is calculated based on the state of surrounding cells and other factors such as suitability of habitat or presence of dispersal vectors or barriers (Smolik *et al.,* 2010; Fitzgerald *et al.,* 2012).

Grid-based models are commonly used in hybrid models where one model is used to predict the suitability of the habitat in each grid cell while another model is used to predict the likelihood of a propagule arriving at that cell (Smolik *et al.,* 2010; Fitzgerald *et al.,* 2012; Fennell *et al.,* 2012; Catterall *et al.,* 2012; Cook *et al.,* 2007). Hybrid models have been developed using various methods to model spread. Fennel *et al.* (2012) uses a mechanistic individual based model using experimental knowledge to parameterize the model, which predicts the spread from the initial introduction point. Catterall *et al.* (2012) and Cook *et al.* (2007) use a Bayesian model to parameterize a grid-based model, which uses distribution data from multiple time points and suitability factors to model the spread of an invasive plant. Similarly, Smolik *et al.* (2010) used logistic regression to model habitat suitability and combined it with a grid-based model parameterized using numerical optimization based on a time series of distribution data for an invasive plant. Percolation theory has also been used to describe the susceptibility of landscapes to invasion (With, 2004). However, the application of this approach for buckthorn is limited by a lack of knowledge about the dispersal distances of buckthorn seeds because percolation models require knowledge of the maximum dispersal distance of seeds.

A problem with applying many of the above methods for modeling the spread dynamics of buckthorn is that they use a dispersal kernel to model the relationship between the spread of the invasive species and the initial distribution. Unfortunately, fitting the dispersal kernel requires a large, spatially continuous dataset for the invasive species in the study region. Many studies deal with this problem by using coarse resolution data that enables them to indicate the presence/absence of species over a large space, though this approach does not allow for local scale predictions of spread. Other studies use data collected over many years or with very large surveying efforts. However, since a high intensity sampling effort was not feasible in the scope of this project, we used a different method to model spread at a local scale. Another approach would be to use mechanistic models but these would also be difficult to apply because data does not exist in the literature on the dispersal distance of buckthorn seeds by birds or rodents. Any estimate of buckthorn's dispersal kernel would therefore involve extensive fieldwork beyond the scope of this study or an arbitrary guess of the average dispersal distance. Currently the literature contains evidence that buckthorn is eaten by many birds and that dispersal is linked to perch locations but the origin of dispersed seeds has not been determined (McCay and McCay, 2009, Archibold *et al.*, 1997)

We used a GLMM with the probability of a cell transitioning from not containing buckthorn (absent) to containing buckthorn (present) at a particular time step as the response variable. The predictor variables were habitat suitability, the number of immediate neighbours containing buckthorn and the number of second order neighbours containing buckthorn. In this model, the coefficients of the neighbour variables and of the habitat variable accounted for short-distance dispersal, while the intercept accounted for the probability that a cell could be invaded by buckthorn if none of the neighbouring cells contained buckthorn, i.e., long-distance dispersal. This method assumes that there is a widely dispersed population of buckthorn throughout the landscape that could be a source of seeds for this long-distance dispersal.

Another type of model that could have been used is a Generalized Estimating Equation (GEE) that explicitly specifies an association structure for the observations. GLMMs do not assume an autocorrelation structure between observations in the same cluster and therefore do not account for temporal autocorrelation in longitudinal data (Zuur *et al.*, 2009; Zhang *et al.,* 2011). Despite these possible issues GLMMs have been used in other studies (Fitzgerald *et al.,* 2012) because they are easily interpreted and explained to land mangers that might be using these models. In the current study, temporal and spatial autocorrelations were tested for in the model residuals to check the extent to which they are a problem in this application.

By avoiding the need for a continuous data set this method allows the characterization of buckthorn spread over the variety of different habitats in the region without an unreasonably intense sampling effort. A similar approach was used by Fitzgerald *et al.* (2012) who modeled the invasion of ants using a grid-based model parameterized with a GLMM where the number of neighbouring cells containing ants was used as an explanatory variable. However the ants could only spread 100 m a year, which was the size of the grid cells, thus eliminating the possibility of dispersal from second-order neighbours or long-distance dispersal (Fitzgerald *et al.*, 2012). Using a GLMM to model the spread of buckthorn, including long distance dispersal, based on the presence of buckthorn in first-order and second-order neighbouring cells and habitat suitability represents a novel approach to modeling spread with limited data requirements

Management of Invasive Species

There are two areas of research for the management of invasive species, the best methods for killing or removing the species and the most effective strategies for implementing these methods. Methods of buckthorn removal have been well researched and best management practices are in place. The Ontario Invasive Plant Council (2012) recommends pulling, burning, grazing, repeated mowing, flooding or cutting/girdling followed by herbicide application. Delanoy and Archibold (2007) recommend chemical girdling as the most cost effective method of control for large populations. The appropriate control method will depend on the size and density of the invasion and the resources of land managers.

Management strategies are more difficult to evaluate because they can be concerned with a range of goals and will differ depending on the scale being considered. For established invasive species the options for management range from eradication, to containment, to abandonment (Epanchin-Niell and Hastings, 2010).

Decisions for the best management scenario for buckthorn will depend on a range of factors including the cost of management, and positive and negative effects of invasion, and cultural values. The negative impacts of buckthorn are discussed above but the costs associated with these impacts are difficult to quantify. Given buckthorn's wide distribution and the density of invasion at some sites it needs to be questioned whether any management is worthwhile. In some situations it may be best to treat an invaded site as a novel ecosystem. A novel ecosystem is one where the ecosystem has been altered by human actions leading to a new self-sustaining ecosystem with different properties and functions from the unaltered system (Morse *et al.,* 2014; Hobbs, Higgs, & Harris, 2009). This system would require major management interventions to be returned to a more natural state. Therefore managers must weigh the value of the novel ecosystem against the value of a more natural ecosystem and the cost of restoration.

For example, a site that is heavily invaded by buckthorn will still provide many ecosystem services such as pollution filtration, erosion control, habitat for some species, shade cultural enjoyment. In the case of buckthorn these services may be better provided by a more natural forest but they are still present in a buckthorn dominated system. If it is determined that the value of a restored forest would not be high enough to justify the costs of restoration, a site can be managed as a novel ecosystem where the goal of management is to preserve the ecosystem functions rather than restore the site to a more natural state (Bhagwat *et al.,* 2012).

This decision can be made on a site by site basis based on the extent of invasion and the presence or proximity of highly valued natural features. The City of Kitchener, for example, has decided to manage some areas as novel ecosystems while restoring others. Their decision is based on the level of buckthorn invasion, ecological value, and public use (Joshua Shea, Personal Communication). If eradication is not feasible on a landscape scale it may be desirable for individual properties with high quality habitat or where the invasion is not locally established. At all scales making informed decisions on the best management strategy requires knowledge of the cost of control, cost of damages, and spread dynamics (Epanchin-Niell and Hastings, 2010). Through the model created in this research managers will be able to predict the costs of long-term buckthorn management in areas where restoration has been completed.

It will be important for conservation planners to consider the long-term cost of managing buckthorn at highly valued sites over many years against the shorter term but substantial costs of attempting buckthorn control on a landscape scale. If private landowners can be mobilized to reduce the amount of buckthorn in the landscape it could greatly reduce the long-term costs of protecting natural areas by reducing the propagule pressure coming from outside natural areas. Models of spread can help to provide information on the feasibility of control by providing estimates of the likelihood of reinvasion and potential extent of damages. Land managers can implement invasive species control on individual properties but landscape level control will require cooperation between multiple landowners, which can be encouraged by government policy and legislation (Yu and Leung, 2006).

Policy and Legislation

Once the likely future spread of an invasive species has been determined and areas that are at risk have been identified, it is important to have policies and legislation in place that allows actions to be taken to prevent damages to sensitive areas. The only legislation that applies directly to invasive plants in Ontario is the *Weed Control Act* (1990). In addition, there is the *Invasive Species Act* (Bill 37, 2015) that is currently being reviewed by the provincial legislature. The *Weed Control Act* requires all landowners to remove any noxious weed found on their land if there is a danger of it spreading to land used for agricultural or horticultural purposes. This legislation is enforced by inspectors who may require the removal of noxious weeds and if they are not removed the inspector may have them removed at the expense of the owner. A noxious weed is any plant that is listed in the regulations. In addition, the *Weed Control Act* gives municipalities the authority to create a by-law designating a plant as a local weed, which will be treated as a noxious weed in that area under the Act. European buckthorn is listed in the regulations as a noxious weed because it is a host for oat crown rust and soy bean aphid. However, since the *Weed Control Act* only applies to noxious weeds, which could spread to agricultural or horticultural lands, sensitive natural areas may not be protected by this legislation.

The *Invasive Species Act* will operate similarly to the *Weed Control Act* in that it will give inspectors the ability to require that actions be taken to prevent the spread of invasive species that pose a significant risk to the natural environment. This is an improvement to the current legislation because it applies directly to invasive species and unlike the *Weed Control Act* it can apply to species that do not threaten agriculture. It also gives the Minister the power to temporarily designate a species as an invasive species without the time required to add it to the regulations. The *Invasive Species Act* will give the Ontario government the ability to react quickly to new invasions and the power to control the spread of existing invasive species. However it will only be applicable to those species considered a significant threat and exceptions may be granted. The *Invasive Species Act* is unlikely to influence the control of European buckthorn since the powers created by the Act already apply to buckthorn through its status as a noxious weed.

Municipal governments contribute to invasive species control through public education programs and policies for management of municipally owned lands. For example, the City of Waterloo's Environmental Land Acquisition and Maintenance Policy identifies the removal of invasive species as one of the management priorities for environmentally sensitive lands protected by the City. Public education programs such as fact sheets and public notices are used to try to encourage private landowners to control invasive species on their land, but landowners are not required to control these species unless they are determined to be at risk of spreading to agricultural areas.

Improvements to this policy framework that would motivate invasive species control on private lands would help to prevent the spread of invasive species to sensitive natural areas. One improvement that will be achieved once the *Invasive Species Act* becomes law, is updating legislation so that similar rules to those that apply to noxious weeds would also apply to invasive plants that could spread to natural areas as well as agricultural areas. Another strategy would be an incentives or cost sharing program for voluntary removal of invasive plants. One example of this type of strategy has been used in British Columbia, where the Regional District of East Kootenay implemented a strategy of combining enforcement of legislation with incentives such as a cost rebate for removal of invaders and an herbicide rebate (Craig, 2012). A similar environmental stewardship incentives program that already exists in Ontario is the Managed Forest Tax Incentive Program, which grants a lower municipal tax rate to landowners who agree to conserve and actively manage their forests (Government of Ontario, 2012). Reimbursing the costs of invasive species removal would encourage the management of invasive species on private lands.

Chapter 3: Modelling

Methods

Buckthorn Distribution

The Region of Waterloo is situated in the Grand River watershed in Southwestern Ontario. It is made up of three cities and four townships, the cities are highly urbanized while the townships are predominantly devoted to agriculture. Land in the Region of Waterloo is 65% devoted to agriculture with 75% of that land being used for the growth of crops, including corn, soy beans, canola, wheat and hay (Region of Waterloo, 2011). The remaining natural cover is primarily woodlots on land zoned as agricultural and protected areas. Surficial geological substrates of the region mainly consist of glacial till, sand, and gravel deposits (Ontario Geological Survey, 1998).

The Region of Waterloo is too large for a complete inventory of all buckthorn in the area. Therefore random sampling was used to obtain an unbiased sample of the land cover of the Region. Random sampling of a number of smaller sampling areas was used instead of a complete sampling of one larger, continuous sampling area to attempt to represent the variety of habitat types in the area. ArcGIS 10.2 (ESRI, 2014) Geographic Information System (GIS) was used with data on the land cover of the Region of Waterloo to define the sampling areas for this study. The Region of Waterloo contains many environments where buckthorn is unable to grow, such as cropped agricultural fields, roads, open water, and quarries. Consequently, the Southern Ontario Land Resource Information System (SOLRIS) Land Use Data (Ontario Ministry of Natural Resources, 2008) was used to identify and exclude these environments from the area that was sampled. The SOLRIS Land Use Data classifies the land cover of the Region of Waterloo into 15 classes (Table 1).

Table 1: Land cover classes in the Region of Waterloo (Ontario Ministry of Natural Resources, 2008)

Land Cover Type	Land Cover Class	Description
Built	Transportation	Roads and highways
Built	Built-up area: impervious	Built areas
Built	Built-up area: pervious	Parks in cities
Built	Extraction	Pits and quarries
Built	Undifferentiated	Primarily agricultural fields but also some meadows
Semi-natural	Hedge rows	Linear strips of forest between fields
Semi-natural	Plantations- tree/ cultivated	Trees of uniform height planted in rows
Natural	Forest: coniferous/ deciduous/ mixed	Forested areas
Natural	Swamp	Forested wetland
Natural	Bog	Wetland
Natural	Fen	Wetland
Natural	Marsh	Wetland
Natural	Open water	Lakes and Rivers

Although the bog, fen and marsh land cover classes are present in the Region of Waterloo, they are not common, and were not captured by any of the randomly selected sampling areas. Therefore bog, fen and marsh are not considered in the rest of this analysis.

First, all natural land cover classes (i.e., forest, hedge row, plantation and swamp) were combined into one new land cover class called natural vegetation. All built land cover classes (i.e., built up im-/pervious and transportation) plus land cover classes open water, undifferentiated and extraction were excluded from the sampling region because they are unlikely to contain buckthorn. Next a 400 m buffer was added around all natural vegetation contiguous with built up areas and included in the sampling region. These buffers were used to add some areas with built up impervious and built up pervious land cover classes back to the sampling region because buckthorn commonly grows in urban backyards and around fences and is likely to be found near urban natural areas (Personal Observation). Finally, any undifferentiated areas added back to the sampling region by the buffer were removed.

Next, 15 points were randomly located within the sampling region and a circle of 200 m radius was drawn around each point to create 15 sampling areas. Each sampling area was overlaid with a grid of 25 m by 25 m to create 224 cells in each sampling area (Figure 2). Since many parts of the sampling region were less than 400 m across, the creation of sampling areas created some grid cells, which overlaid unsuitable land cover. These cells were not sampled and were excluded from the analysis. The 15 sampling areas were determined to be a large enough sample because the proportion of cells containing buckthorn in the first eight sampling areas was approximately 0.25. Based on this proportion and aiming for a standard error of 0.02 the following sample size calculation was carried out where $\hat{p}_{LC}$ is the proportion of cells with land cover type LC that contain buckthorn, n is the number of cells to sample and SE is the standard error:

$$SE(\hat{p}_{LC}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n_{LC}}}$$

$$n_{LC} = \frac{0.25(1-0.25)}{0.02^2} \times 5\ Landcover\ classes$$

$$n = 2344\ cells\ sampled \div 185\ cells\ sampled\ per\ sampling\ area$$

$$= 12.7\ sampling\ areas$$

Therefore, to achieve an adequate sample size at least 13 sampling areas had to be sampled and, erring on the side of caution, 15 sampling areas were selected. Ultimately, 2400 cells were sampled in the 15 sampling areas. This number of cells is smaller than the total number of cells in the 15 sampling areas (= 3,360 cells), because approximately 900 cells were inaccessible or overlaid land cover classes that were excluded from analysis.

Sampling

With the help of volunteer field assistants, it was attempted to visit each cell in the sampling areas that was included in the analysis to determine the presence or absence of buckthorn. Cells were located with the help of a Juno Trimble SB GPS, which was loaded with a map including the sampling grid, and the complete area of each cell was walked using the GPS icon indicating the position on the map. Using an increment borer, a core sample was taken from the stem of the largest buckthorn that was identified by sight, its diameter at breast height measured and a description of the land cover of the cell recorded. If buckthorn was present but the largest individual was less than ~2.5 cm in basal diameter, it was not possible to take a core but the presence of buckthorn was noted. Some cells were inaccessible so these cells were removed from the dataset.

Processing Cores

The core samples were glued to wooden mounts and then sanded to a polished finish. At first the sanding was done by hand only but later a random orbital sander was used for rough sanding, which was then finished by hand. Growth rings were counted using a dissecting microscope and buckthorn age was assumed to be identical to the number of growth rings. In cases when the core missed the pith or the centre of the tree was rotten, the number of missing rings was estimated and added to the counted number of rings. To estimate the number of missing rings, the radius of the oldest ring was measured by matching it to a clear sheet with concentric rings of known radius. Then the growth rate of the three oldest rings was used to interpolate the number of missing rings (Ranius, Nikalsson, and Berg, 2009).

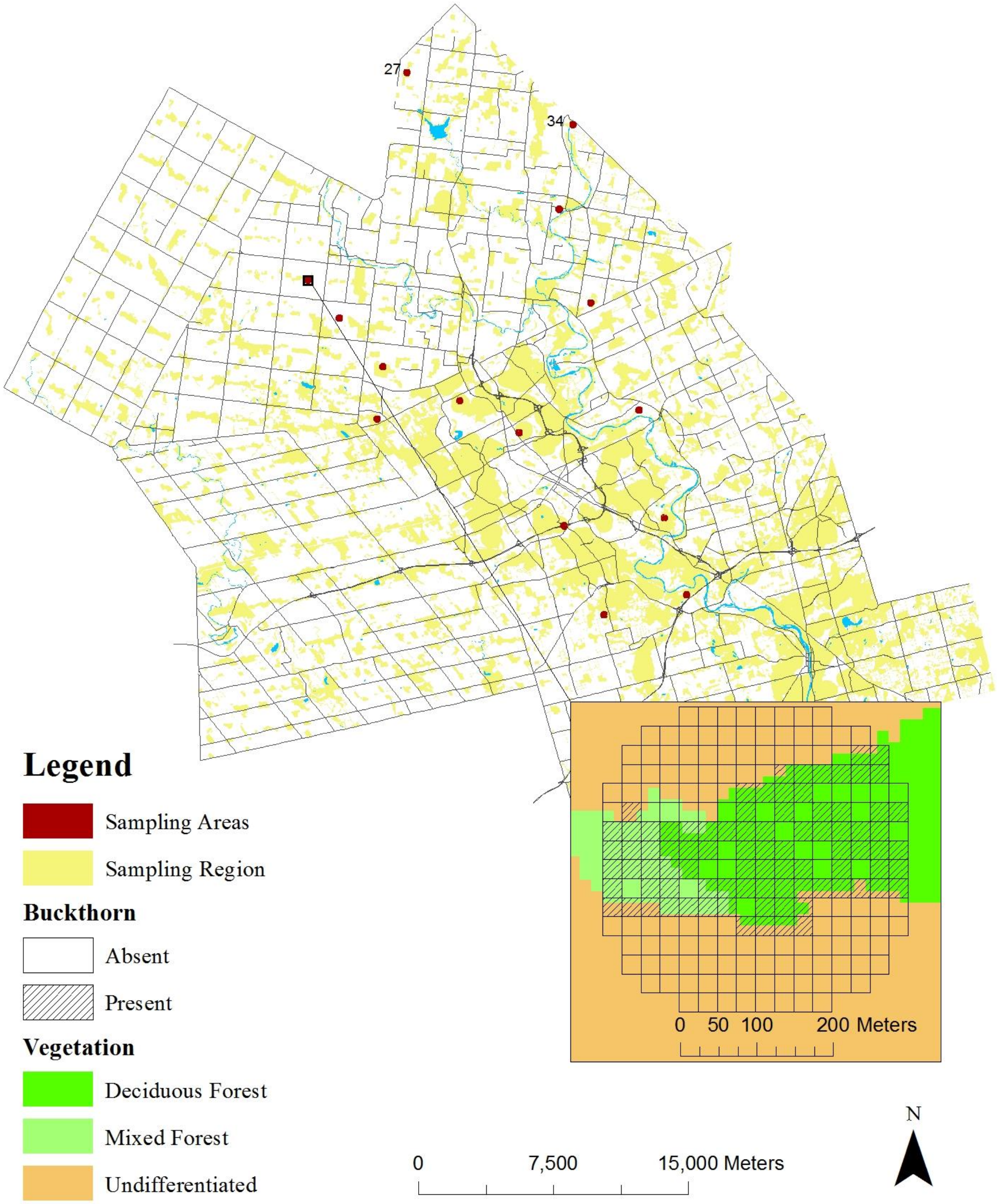

Figure 2: Map of the Region of Waterloo showing the sampling region within which sampling areas were randomly located. The inset shows the grid of a sampling area over a map of the land cover

Growth rings of 590 tree cores were counted. In order to determine the intra-observer error in the buckthorn age, 113 cores from 10 randomly chosen mounts were recounted. The buckthorn ages from the first and second counts were compared by a paired t-test of the log of buckthorn age. The buckthorn ages were not significantly different $(t(112) = -1.4899, p >$

0.05) in the two repetitions of counting and therefore the buckthorn ages can be considered reliable

Habitat Suitability

The habitat suitability model was created using generalized linear models and generalized linear mixed models. All the analysis was done using ArcGIS 10.2 (ESRI, 2014) and R: A language and environment for statistical computing, version 3.1.2 (R Core Team, 2014). Various R packages were used in the analysis; references are given the first time each package is used. (R scripts are provided in Appendix A)

Initially, six habitat characteristics were considered for the habitat suitability model including land cover class (Table 2). Early inspection of the land cover class data did not indicate any significant difference between different types of forests. Therefore, to decrease the number of levels in the land cover class variable, the deciduous forest, coniferous forest, mixed forest, and hedgerow land cover classes were combined into a new land cover class called 'all forest', while the land cover classes undifferentiated, extraction, transportation and open water were combined into a new land cover class called 'other'.

The land cover data has a resolution of 10 m while the sampling area grid cells are 25 m, to adjust for these mismatched resolutions the area of each land cover class within each cell was calculated. To convert each cell back to a single land cover class, all cells were reclassified according to the land cover class area within each cell using the following rules: (1) If a cell contained either all forest or swamp, it was assigned as the one with the greatest area of these two. (2) If a cell contained neither all forest nor swamp but it contained built-up pervious or built-up impervious, it was assigned as the one with the greatest area of these two. (3) If a cell contained neither all forest, swamp, built-up pervious nor built-up impervious, it was assigned as other. Cells with the land cover class other were removed from the data set because they are assumed to have zero probability of containing buckthorn. In a few cases, cells that did contain buckthorn were classified as other but nevertheless were removed from the data set. This situation occurred when the land cover had either recently changed or was not differentiable from agricultural fields. The approach to exclude these cases from analysis was deemed appropriate because otherwise the calculations would have resulted in a small likelihood of buckthorn growing in an agricultural field, which is unreasonable because buckthorn establishment is limited by cropping.

Distance to the edge of the forest patch was included as a habitat variable as a proxy for characteristics that often vary between the edge and the interior of forests. For example, light levels are usually higher near the edge of a forest and there is evidence for differences in nutrient availability and species composition (Vallet *et al.,* 2010)

Following Hosmer and Lemeshow (2000), first a univariable logistic regression of each variable on the presence of buckthorn was completed. For each regression a likelihood ratio test and Wald test were performed to determine whether there was a significant difference between the univariable regression and a null model. All variables that showed a significant $(p < 0.25)$ relationship with the presence of buckthorn were combined in a maximal model. We use such a high value of p so that variables that may be important in combination with the other variables are included (Hosmer and Lemeshow, 2000). The model was then simplified by backwards selection from the maximal model until the minimal acceptable model was determined. A maximal model is the model that includes all potential variables and their interactions even if

Table 2: List of habitat characteristics considered in habitat suitability model and a description of how a value for each characteristic was assigned to each cell

Habitat Characteristic	Description	Interpretation for cells	Source
Land Cover (LC)	15 m by 15 m raster of land cover classes	Calculated the area of each land cover type in cell. Classified it as the land cover type most likely to contain buckthorn	Southern Ontario Land Resource Information System (SOLRIS) Land Use Data
Compound Topographic Index (CTI)	Steady state wetness index. Function of both the slope and the upstream contributing area per unit width orthogonal to the flow direction.	Mean CTI within each cell	Geomorphometry and Gradient Metrics toolbox by Jeffrey Evans and DMTI Digital Elevation Model
Nearest Property Line (NearPropLi)	Distance from each grid cell to the nearest property line	Distance from each cell to the nearest property line	Regional Municipality of Waterloo Property Parcels
Distance to Patch Edge (PatchDist)	Patches were made by dissolving All Forest with Swamp	Distance from each cell to the edge of the nearest patch	SOLRIS Land Use Data
Permeability (Permea)	Permeability of the surficial material	Classified as permeability of whichever class had the largest area in the cell	Surficial Geology of Southern Ontario (MRD128-REV) from the Ontario Geological Survey
Material	Surficial Material	Classified as material of whichever class had the largest area in the cell	Surficial Geology of Southern Ontario (MRD128-REV) from the Ontario Geological Survey

their coefficients are not significant (Crawley, 2007). The minimal acceptable model is the simplest model that is not significantly worse than the maximal model (Crawley, 2007). Next, generalized linear mixed effects models (GLMM) were fit to account for the effect of the hierarchical sampling structure and to include a spatial autocorrelation structure. The glmer function from the lme4 package (Bates *et al.*, 2014) was used to create a model with the sampling area ID number as a random effect. The glmer function estimates the model parameters using maximum likelihood, which allows the model to be simplified by comparing nested models with likelihood ratio tests (Zuur *et al.*, 2009). The glmmPQL function from the MASS package (Venables and Ripley, 2002) was then used to incorporate a spatial autocorrelation structure as well as the random effect. glmmPQL uses Penalized Quasi-Likelihood to estimate model parameters and therefore cannot be used for model simplification by likelihood ratio tests. Therefore the glmmPQL model was run with the same formula as the glmer model.

The Akaike Information Criterion (AIC) is given for all models and is used to compare non-nested models (Zuur *et al.*, 2009). AIC gives a penalized log-likelihood which penalizes the addition of more parameters. A lower AIC indicates a more parsimonious model relative to a model for the same data with higher AIC. AIC was reported here because it is commonly used in the literature and is recommended for comparisons of non-nested models (Thiele and Markussen, 2012; Zuur *et al.*, 2009; Fitzgerald and Gordon, 2012)

To test the ability of the model to discriminate between cells containing buckthorn and unoccupied cells, leave-one-out cross-validation was performed. This was done by running the habitat suitability model with each sampling area removed in turn and then testing the habitat suitability model performance on the sampling area that was left out using the sperrorest package (Brenning, 2012). Discrimination is measured by the area under the receiver operating characteristic curve (AUROC). AUROC is equal to the proportion of times that for any pair of cells where buckthorn is present in one cell and absent from the other, the model predicts a higher probability of buckthorn presence in the occupied cell (Hosmer and Lemeshow, 2000). AUROC ranges from zero to 1 where one is perfect discrimination and 0.5 is no discrimination. As a rule of thumb AUROC > 0.7 is considered acceptable (Hosmer and Lemeshow, 2000).

Spread Model

In order to represent the change in the distribution of buckthorn over time, a hybrid GLMM was used to model the likelihood of a cell becoming invaded by buckthorn based on both the suitability of habitat and the likelihood of a seed being dispersed to that cell. The likelihood of a seed being dispersed to a cell was modelled based on the presence of buckthorn in a neighbouring cell. To model the spread of buckthorn, a time-series was created for each cell based on the age of buckthorn in the cell. The time-series values were 1 for the year in which the cell transitioned from buckthorn absent to buckthorn present and 0 for all years in which buckthorn remained absent. The years after buckthorn became present were not included in the model because it was assumed that once buckthorn was present in a cell, it would remain so. As well, for each cell the number of first and second order neighbour cells containing buckthorn was also calculated for each step of the time series (Figure). The time series start ten years before the present because the age is not known for trees that were too small to be cored (i.e., smaller than 2.5 cm basal diameter). The smallest tree cored was five years old, therefore limiting the model to ten years before the present ensures that only trees with known age are included. Ten years was chosen in order to err on the side of caution because although some five year old trees were large enough to be cored it is possible that a tree of greater than five years may have been too small to core.

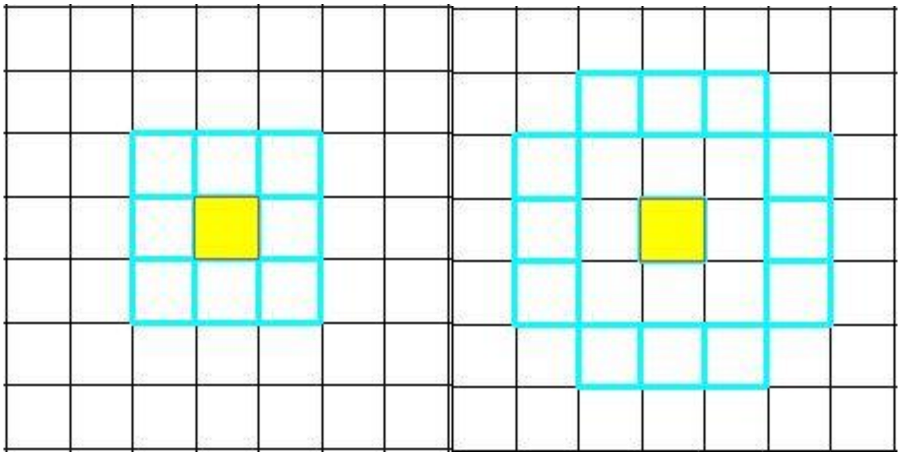

Figure 3: Left: cells outlined in blue are considered first order neighbours to the yellow cell. Right: the cells outlined in blue are considered second order neighbours to the yellow cell

The GLMM for the spread model used the event of the transition from buckthorn absent to buckthorn present as the response variable, the centred habitat suitability, year, and number of first and second order neighbour cells containing buckthorn as the explanatory variables, and the sampling area ID as the random effect. The habitat suitability was centred by subtracting the mean habitat suitability from each cell's habitat suitability. This was done so that the intercept in the model would reflect the likelihood of a cell transitioning from buckthorn absent to present when a cell had average habitat suitability. The variable year is counted as "years before present" and reflects the fact that in the past less buckthorn was present in the Region and therefore the likelihood of any cell being invaded was lower in the past. The relationship between the number of neighbour cells with buckthorn present and the response variable was examined by plotting a generalized additive model (GAM) using the gam function from the mgcv package (Wood, 2011).

The discrimination of the model was again determined with leave-one-out cross-validation. The fit of the model was validated using plots of the individual variables and the response versus the cumulative residuals using the cumres function from the gof package (Holst, 2014; Thiele and Markussen, 2012). The cumres function does not accept GLMM models so a

GLM was fit, with the sampling area ID as a fixed effect, for the validation step. The cumulative residual method of model validation involves comparing the cumulative sum of residuals observed in the model to simulated cumulative residuals, which reflect the expected random variation within the data (Lin, Wei, and Ying, 2002). If the observed residuals are within the range of the simulated residuals then any fluctuations in the residuals can be attributed to random variation and the null hypothesis that the model is correctly specified cannot be rejected. The cumres function also produces the results of two goodness-of-fit tests, the Kolmogorov-Smirnov and Cramer-von-Mises tests, which give a quantitative evaluation of whether the observed cumulative residuals are significantly different from what would be expected due to random variation (Lin, Wei, and Ying, 2002).

Autocorrelation in the spread model could be present both as spatial autocorrelation and temporal autocorrelation because observations of a cell that are closer together in time or space may be more similar than those that are farther away in time or space. The residual spatial autocorrelation for the spread model was tested by plotting a semivariogram for the data from years 10 to 20 over distances from 0 m to 300 m. Temporal autocorrelation was tested by plotting the semivariogram based on the distance in time between data points for the same cell. The semivariance at each time lag was then averaged over all the cells in five randomly selected sampling areas. A subset of the dataset was used for both the spatial and temporal autocorrelation in order to reduce the computational requirements for the semivariogram. The function variog from the package geoR (Diggle and Ribeiro, 2007) was used to plot the semivariogram.

Results

Buckthorn Distribution

Buckthorn was present in all 15 sampling areas but there were great differences between sampling areas in the number of cells containing buckthorn (Table 3). The average age of all cored trees was 13.7 years and the oldest tree recorded was 56 years old. The number of cells occupied by buckthorn has increased exponentially from 1959 to the present except for years one to five because accurate ages were not possible to obtain for these years (Figure 4).

Table 3: Summary of the number of cells containing buckthorn and the average age of the oldest buckthorn within each cell for the 15 sampling areas

Sampling area ID	Mean age	Standard deviation	Number of cells containing buckthorn
1	13.40	10.71	35
15	23.33	10.96	168
18	12.31	9.66	65
21	10.70	7.77	61
23	9.68	6.91	25
24	12.21	10.11	48
27	11.23	10.85	13
30	10.81	10.20	86
32	8.75	8.48	8
34	9.38	8.09	50
36	4.09	6.37	46
37	8.28	7.51	57
40	10.42	10.68	43
41	16.00	5.66	2
43	15.95	9.71	111
Total:	**13.69**	**11.02**	**818**

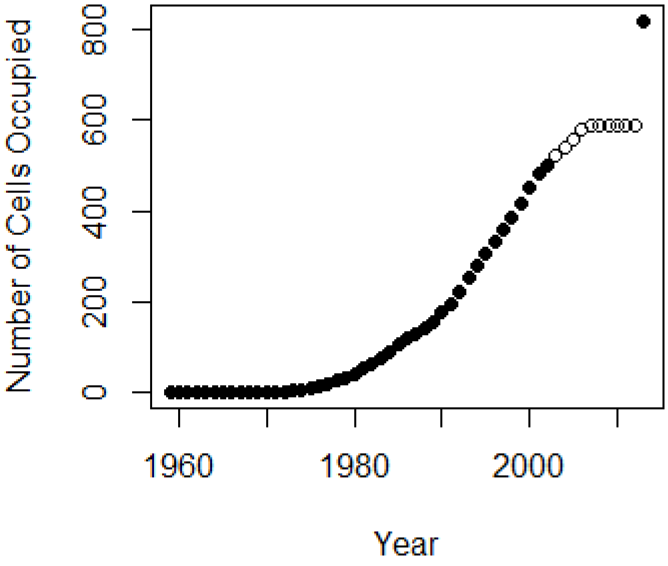

Figure 4: Number of cells occupied by buckthorn per year from 1959 to 2013. Solid circles represent years in which age data was available, open circles represent years which are ignored because age data was not available because the trees were too small to core (ie <2.5 cm in basal diameter). The number of occupied cells was calculated based on the age of the oldest buckthorn in the cell and assuming that this was the first buckthorn to occupy the cell.

Habitat Suitability

The results of the univariable regressions indicate that all of the variables had a significant ($p < 0.25$) relationship with the presence of buckthorn, except for the variable 'distance to the nearest property line' (NearPropLi) and the compound topographic index (CTI) (Table 4). Therefore NearPropLi and CTI were excluded from the rest of the analysis.

Table 4: Results of univariable logistic regression for each variable to test for a significant relationship with the presence of buckthorn. The likelihood ratio test compares the model to a null model with just the intercept while the Wald test tests the hypothesis that the coefficient is equal to zero. For explanation of variable names see Table 2.

Variable		Likelihood Ratio Test P Value	Wald Test P Value	GLM Coefficient	GLM Std. Error
NearPropLi		0.780	0.779	2.746E-04	0.001
PatchDist		<0.001	<0.001	-0.005	0.001
CTI		0.400	0.401	-0.026	0.030
LC		<0.001	<0.001		
LC Levels:	(Intercept)			-0.375	0.066
	LCBuiltImp			-1.474	0.142
	LCBuiltPerv			-0.800	0.125

	LCSwamp			-0.660	0.130
Permea		<0.001	<0.001		
Permea Levels:	(Intercept)			-1.629	0.081
	PermeaLow			0.783	0.124
	PermeaMediumLow			1.636	0.115
	PermeaVariable			0.694	0.199
Material		<0.001	<0.001		
Material Levels:	(Intercept)			-0.380	0.060
	MaterialGravel			-0.785	0.143
	MaterialOrganicDeposition			-1.229	0.777
	MaterialSand			-1.303	0.109
	MaterialSilt			-12.186	324.744

The formula for the maximal model was:

$$Buckthorn\ Presence \sim Material\ + Permea\ + log10(PatchDist\ + 1) + LC\ + LC\!:\!Permea$$

This model failed because of collinearity between material and permeability. Since it was clear from the univariable regression that the silt level of material did not fit well, material was removed. The interaction term was removed at this point because although a likelihood ratio test showed that the model with the interaction was better than the model without the interaction, some of the coefficients for interaction terms had excessive standard errors (-12.20 ± 240). From this point on, the model was simplified by removing each term in turn and then performing a likelihood ratio test to compare the simplified model to the more complex model. If the simplified model was not significantly worse, the simplified model was kept (Table 5).

Table 5: Results of model simplification by backward selection. Each variable was removed in turn and the model without the variable was tested against the model containing the variable. A likelihood ratio test was used for this comparison. If the model with fewer variables was not significantly different the variable was removed from the model. For explanation of variable names see Table 2, intBT represents the presence of buckthorn.

Model	Formula	Test	Log Likelihood	Likelihood Ratio Test P Value
Maximal model	intBT ~ Material + Permea + log10(PatchDist + 1):InPatch +	None	N/A	N/A
Model 1	intBT ~ Permea + log10(PatchDist + 1):InPatch + LC	None	-1239.736	N/A
Model 2	intBT ~ log10(PatchDist + 1):InPatch + LC	1 vs 2	-1333.139	<0.001
Model 3	intBT ~ Permea + LC	1 vs 3	-1272.005	<0.001
Model 4	intBT ~ Permea + log10(PatchDist + 1):InPatch	1 vs 4	-1326.290	<0.001

Based on the results of the model simplification the minimal acceptable model was found to be Model 1. The likelihood ratio test between Model 1 and the null model showed that Model 1 was significantly different from the null model ($p\ =<\ 0.001$). Therefore the null hypothesis that the models are equivalent was rejected and it was concluded that at least one of the coefficients in Model 1 is significant. The Wald test results in the summary of Model 1 (Table 6) show that all the coefficients are significant ($\alpha = 0.05$).

Table 6: Summary of the result of the minimum acceptable model. The P value is based on a Wald test comparing the coefficient estimate to zero. For explanation of variable names see Table 2, intBT represents the presence of buckthorn.

Model 1: Formula: intBT ~ Permea + log10(PatchDist + 1) + LC

Variable	**Estimate**	**Std. Error**	**z value**	**Pr(>|z|)**	
(Intercept)	-0.49258	0.11857	-4.154	<0.001	***
PermeaLow	0.52717	0.13199	3.994	<0.001	***
PermeaMediumLow	1.59911	0.12096	13.22	<0.001	***
PermeaVariable	0.46651	0.21471	2.173	0.0298	*
LCBuiltImp	-1.86964	0.16025	-11.667	<0.001	***
LCBuiltPerv	-1.30197	0.15115	-8.614	<0.001	***
LCSwamp	-0.64997	0.14188	-4.581	<0.001	***
log10(PatchDist + 1):InPatch	-0.61277	0.07773	-7.884	<0.001	***

Next the output from Model 1 was tested for spatial autocorrelation by plotting the spline correlogram of the Pearson residuals from Model 1 (Zuur *et al.*, 2009). The results show that spatial autocorrelation is present at both the within and between sampling area levels (Figure 5A,B).

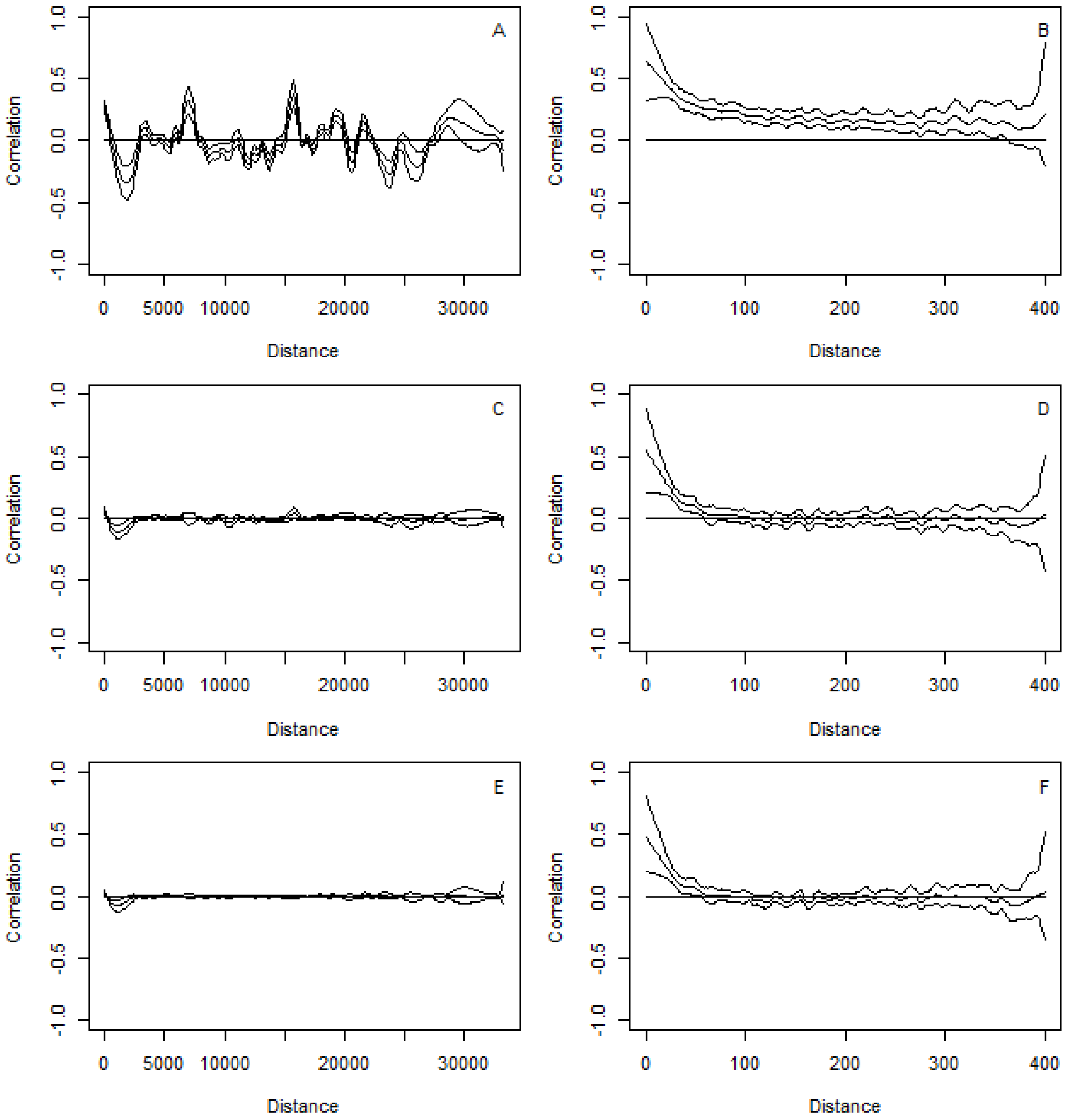

Figure 5: Spline correlograms with 95 % pointwise bootstrap confidence intervals showing the spatial autocorrelation in the Pearson residuals of the fitted values for: A) the GLM model at all distances, B) the GLM model up to 400 m, C) the mixed effects model fit with glmer1 at all distances, D) the mixed effects model fit with glmer1 up to 400 m, E) the mixed effects model fit with glmmPQL at all distances, and F) the mixed effects model fit with glmmPQL up to 400 m.

To deal with the between-sampling area spatial autocorrelation, generalized linear mixed effects model with sampling area as a random effect were run using the glmer function from the package lme4. Model simplification was carried out by backwards selection similar to the GLM model. The results of the model simplification indicate that the minimal acceptable model is glmer1 with permeability, distance to patch edge, and land cover as explanatory variables (Table 7).

Table 7: Results of model simplification by backward selection for the GLMM with sampling area ID as a random effect. Each variable was removed in turn and the model without the variable was tested against the model containing the variable. A likelihood ratio test was used for this comparison. If the model with fewer variables was not significantly different the variable was removed from the model. For explanation of variable names see Table 2, intBT represents the presence of buckthorn.

Model	Formula	Test	Log Likelihood	Likelihood Ratio Test P Value
glmer1	intBT ~ Permea + log10(PatchDist + 1):InPatch + LC	None	-951.6425	N/A
glmer2	intBT ~ log10(PatchDist + 1):InPatch + LC	1 vs 2	-961.0722	<0.001
glmer3	intBT ~ Permea + LC	1 vs 3	-959.9048	<0.001
glmer4	intBT ~ Permea + log10(PatchDist + 1):InPatch	1 vs 4	-1028.1198	<0.001

The correlograms for the glmer1 model (Figure 5C,D) show that the between sampling area spatial autocorrelation was accounted for by adding the random effect. To test for a spatial autocorrelation structure within sampling areas, we used the glmmPQL function to include a spherical correlation structure. This model had a range of 25 m and a nugget of 0.95 and produced effectively the same coefficients as the glmer1 model (results not shown). The correlogram of the Pearson residuals from the glmmPQL model (Figure 5E,F) is also very similar to the glmer1 model correlogram. An exponential correlation structure was tested with similar results. These results indicate that including the within sampling areas correlation structure did not improve the model. Therefore the glmer1 model was used for the rest of the analysis (Table 8).

Table 8: Summary of the result of the glmer1 model. The p value is based on a Wald test comparing the coefficient estimate to zero. For explanation of variable names see Table 2, intBT represents the presence of buckthorn

glmer1: Formula: intBT ~ LC + log10(PatchDist + 1):InPatch + Permea + (1 | ID_1)					
Random effects:					
Groups Name	**Variance**	**Std.Dev.**			
ID_1(Intercept)	3.503	1.872			
Number of obs: 2400, groups: 15					
Fixed effects:					
	Estimate	**Std. Error**	**z value**	**Pr(>|z|)**	
(Intercept)	0.181	0.517	0.35	0.726	
LCBuiltImp	-2.923	0.258	-11.323	<0.001	***
LCBuiltPerv	-2.262	0.261	-8.684	<0.001	***
LCSwamp	-0.525	0.278	-1.892	0.058	.
PermeaLow	-0.073	0.329	-0.223	0.823	
PermeaMediumLow	0.681	0.233	2.92	0.004	**
PermeaVariable	-0.753	0.262	-2.876	0.004	**
log10(PatchDist+1):InPatch	-0.502	0.123	-4.091	<0.001	***

The result of the leave-one-out cross-validation showed the AUROC for the model was 0.63 which is considered poor discrimination (Hosmer and Lemeshow, 2000).

Spread Model

The GAM showed that there was a non-linear relationship between the number of neighbour cells in which buckthorn was present and the invasion probability. The largest change in invasion probability occurred going from zero to one neighbour cells in which buckthorn was present with a less pronounced impact of additional neighbours (Figure 6). Therefore two binary variables of the presence or absence of buckthorn in any first order and second order neighbours were added to the model.

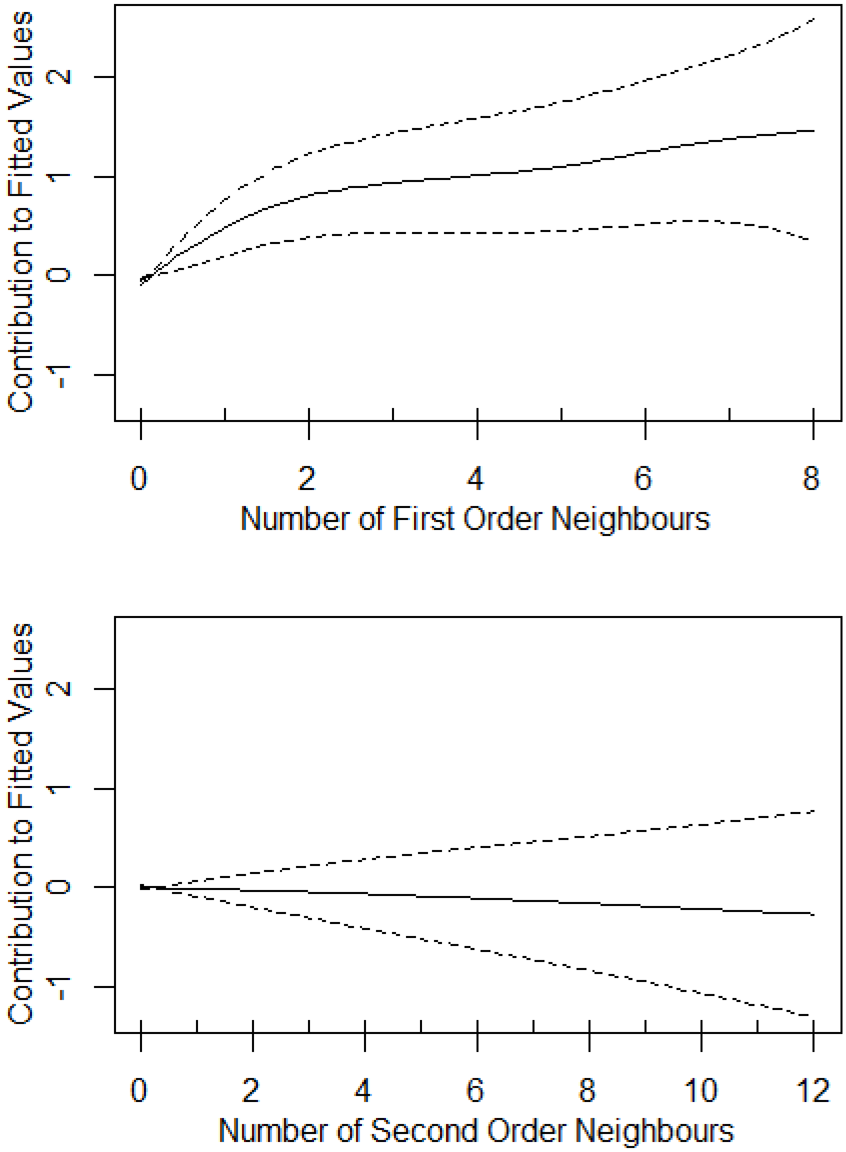

Figure 6: Results of generalized additive model (GAM) for the number of first and second order neighbours that are invaded by buckthorn, showing the nonlinear relationship between the number of invaded first order neighbours and the likelihood of a cell becoming invaded by buckthorn.

The results from the spread model showed that the presence of at least one first order neighbour cell containing buckthorn significantly increased the probability of a cell transitioning from buckthorn absent to buckthorn present (Table 9). The exact number of first order neighbour cells with buckthorn present had a much smaller though still significant effect. Neither the presence of second order neighbour cells with buckthorn present nor their number had a significant effect. Habitat suitability and year also had significant effects on the probability of a cell transitioning from buckthorn absent to buckthorn present (Figure 7).

Table 9: Summary of the result of the spread model. The p value is based on a Wald test comparing the coefficient estimate to zero. For explanation of variable names see Table 2.

Formula: trans~(neigh1 > 0) + neigh1 + (neigh2 > 0) + neigh2 + hsc + year + (1|ID_1)					
Random effects:					
Groups Name	**Variance**	**Std.Dev.**			
ID_1 (Intercept)	1.345	1.16			
Number of obs: 54361, groups: ID_1, 15					
Fixed effects:					
	Estimate	**Std. Error**	**z value**	**Pr(>|z|)**	
(Intercept)	-4.169	0.453	-9.212	<0.001	***
(neigh1 > 0)	0.748	0.241	3.103	0.002	**
neigh1	0.134	0.063	2.122	0.034	*
(neigh2 > 0)	0.341	0.231	1.477	0.140	
neigh2	-0.017	0.045	-0.378	0.706	
hsc	3.299	0.595	5.542	<0.001	***
year	-0.104	0.013	-8.005	<0.001	***

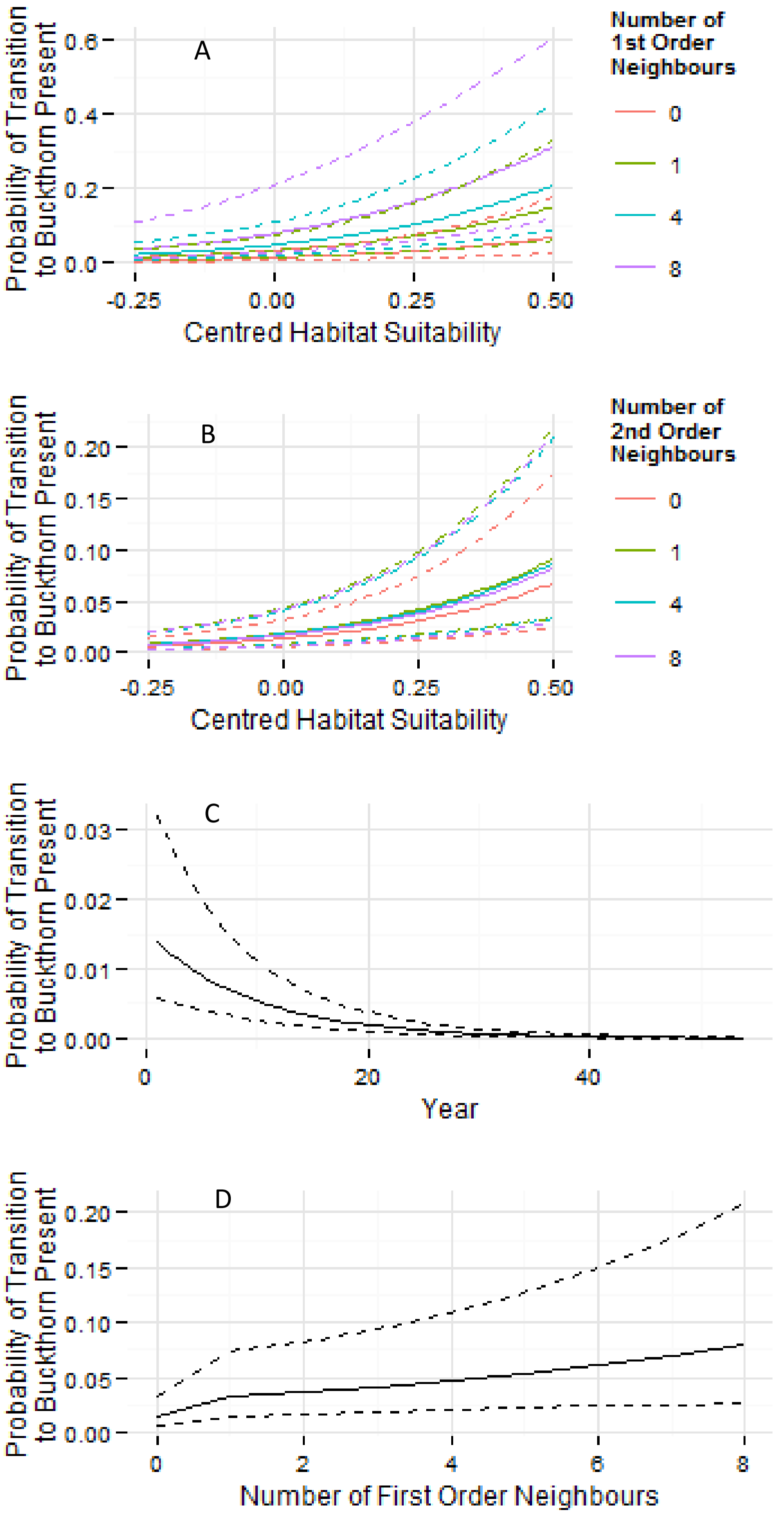

Figure 7: A) The relationship between the probability of a cell transitioning from buckthorn absent to present and the centred habitat suitability with four different numbers of invaded first order neighbours, when year = 1; B) The relationship between the probability of a cell transitioning from buckthorn absent to present and the centred habitat suitability with four different numbers of invaded second order neighbours, when year = 1; C) The relationship between the probability of a cell transitioning from buckthorn absent to present and year when there are zero invaded first or second order neighbours, and centred habitat suitability=0; D) The relationship between the probability of a cell transitioning from buckthorn absent to present and the number of invaded first order neighbours, when year=1 and centred habitat suitability=0

The semivariogram for the spatial autocorrelation in the spread model shows that there is weak (nugget-to-sill ratio of approximately 0.045:0.06) residual autocorrelation in the model. (Figure 8).

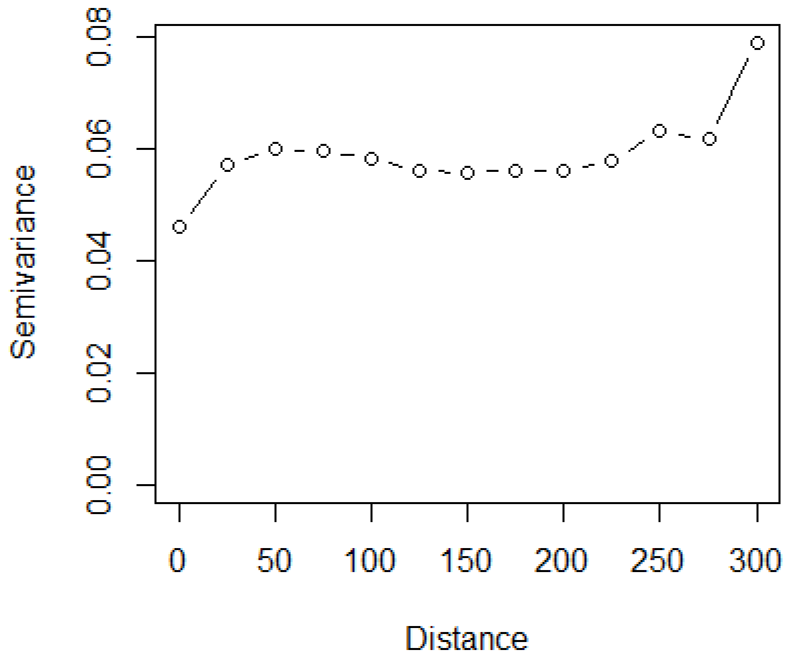

Figure 8: Semivariance of spread model residuals for a subset of the data containing years 10 to 20. The distances reflect the range of distances for cells within the same sampling area.

The semivariogram for the temporal autocorrelation in the model shows that there is nonstationarity in the model residuals with respect to time (Figure 9).

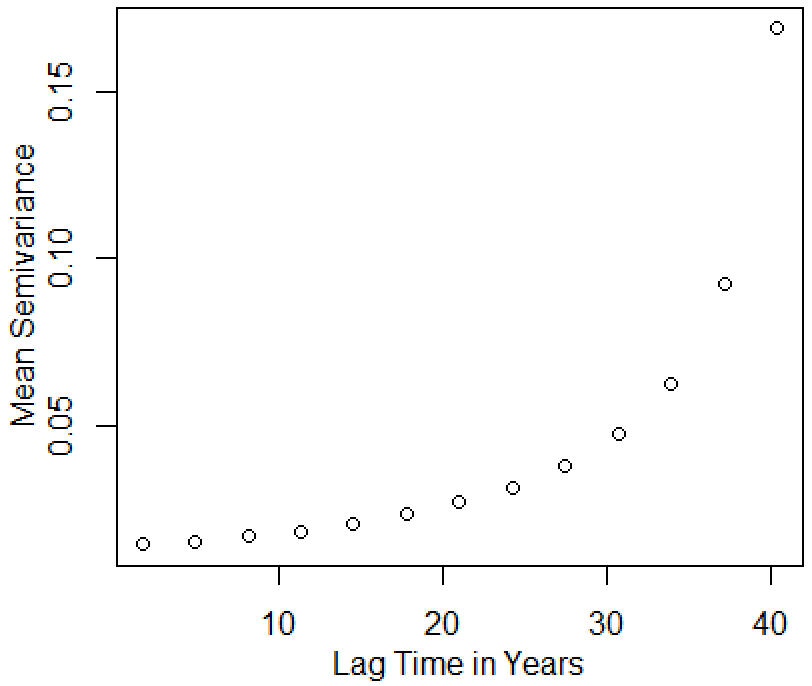

Figure 9: Temporal autocorrelation of the residuals from the spread model for five randomly selected sampling areas at time lags ranging from 1 to 40 years. The semivariance was calculated for each cell over the range of time lags and then the mean semivariance taken for each time lag.

The cumulative residual plots show that the observed cumulative residuals are in the expected range (Figure 10). The model shows good discrimination with an AUROC of 0.88.

Discussion

The hybrid GLMM spread model had an AUROC of 0.88, suggesting good discrimination, while the habitat suitability model had an AUROC of 0.63, suggesting poor discrimination. Nevertheless, these two AUROC values should not be compared directly because they are based on different data sets and because the arbitrary length of the transition time series used to estimate the spread model can affect the AUROC value.

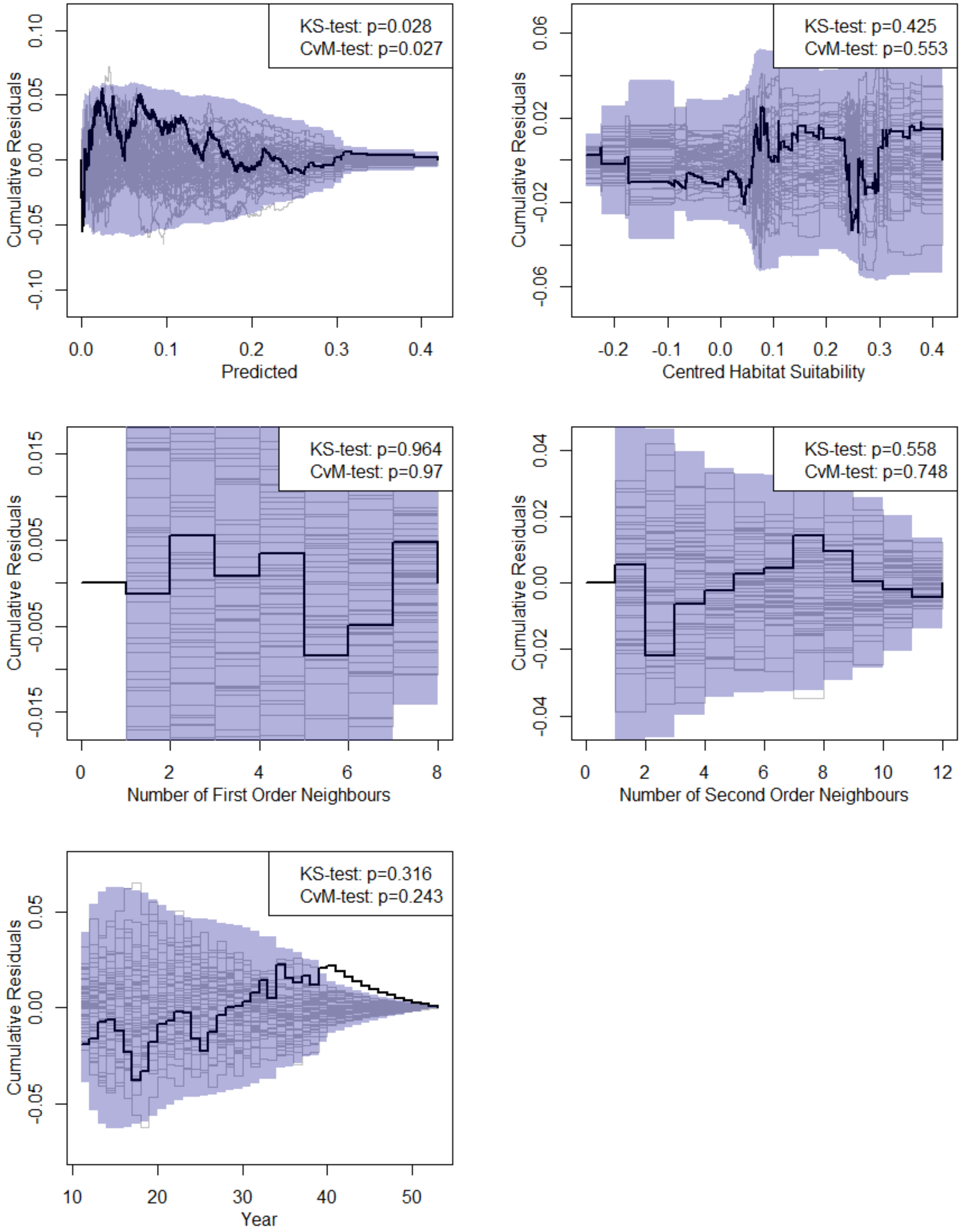

Figure 10: The black lines represent the cumulative residuals plotted against each variable and the predicted values of the spread model. The grey lines represent simulated cumulative residuals based on random variation expected if the model is correctly specified. The KS-test and the CvM-test are the Kolmogorov-Smirnov and Cramer-von-Mises tests which test the hypothesis that the variation in the observed cumulative residuals is significantly different from the expected random variation.

The GLMM habitat suitability model that included random effects was a significant improvement over the GLM habitat suitability model. This improvement reflects the fact that GLMM is the appropriate modeling framework when the data are grouped by variables that are not directly of interest for the study, in this case the sampling areas (Thiele and Markussen, 2012). The model was greatly improved, as indicated by the AIC changing from 2495 to 1921, when the hierarchical sampling structure was accounted for by including sampling area as a random effect. Including the random effect also accounted for the between sampling area spatial autocorrelation in the data (Figure 5). The habitat suitability model predicted that cells near the edge of a forest with medium-low permeability are most likely to contain buckthorn (Figure 11). This is consistent with expectations based from the literature because buckthorn is known to grow well near the edges of forests (Knight *et al.*, 2007). It is important to note that standard errors for the habitat suitability model (Table 8) are fairly high relative to the absolute value of the coefficient estimates and therefore the 95% confidence intervals for the probability of buckthorn presence for cells with different characteristics overlap (Figure 11).

It is harder to interpret the results for the permeability because buckthorn is known to grow on a wide range of soil moistures (Kurylo *et al.*, 2007). The low suitability for cells with variable permeability is probably influenced by the fact that low permeability is only present in three sampling areas, one of which is in an urban area so buckthorn presence is reduced by human management. Cells with medium-low permeability had the highest habitat suitability which may have been influenced by the fact that the three sampling areas with the highest proportion of cells occupied by buckthorn have medium-low permeability. From the literature there is no evidence that permeability of the surficial material is an important determinant of buckthorn presence so it is possible that there is a different process occurring that happens to correlate with permeability. It seems likely that permeability could be reflecting a difference related to the location of the cell because the predominate permeability changes from medium-low to high to low as you move from east to west across the Region of Waterloo. The high suitability of medium-low permeability cells could be related to the historical spread of buckthorn, since 40 of the 44 buckthorn samples older than 30 years were taken from two of the sampling areas with medium-low permeability.

One source of errors in the model are the missing interaction terms that had to be removed because of insufficient sampling for some combinations of permeability and land cover. Although several of the interaction terms did not have significant coefficients, some did and including them may have improved the model's performance. This problem could have been avoided by using a stratified sampling approach. However, then the population wide results, such as the total number of cells invaded by buckthorn, would have been less representative of the Region in general because the data would not have been a random sample.

Applying the model to the whole region showed that suitable habitat is widespread throughout the Region of Waterloo (Figure 12).

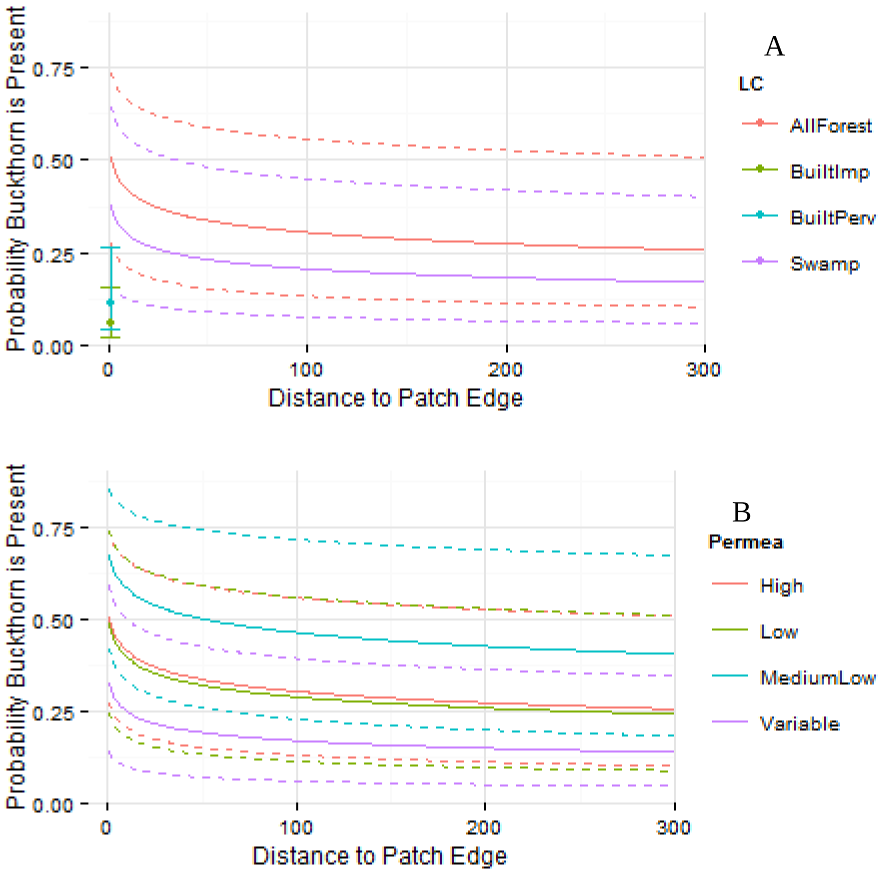

Figure 11: Dashed lines represent confidence intervals based on the fixed effects. A) Probability of a cell containing buckthorn for four levels of permeability when land cover class is 'AllForest'. B) Probability of a cell containing buckthorn for four classes of land cover. See Table 1 for explanation of variable names.

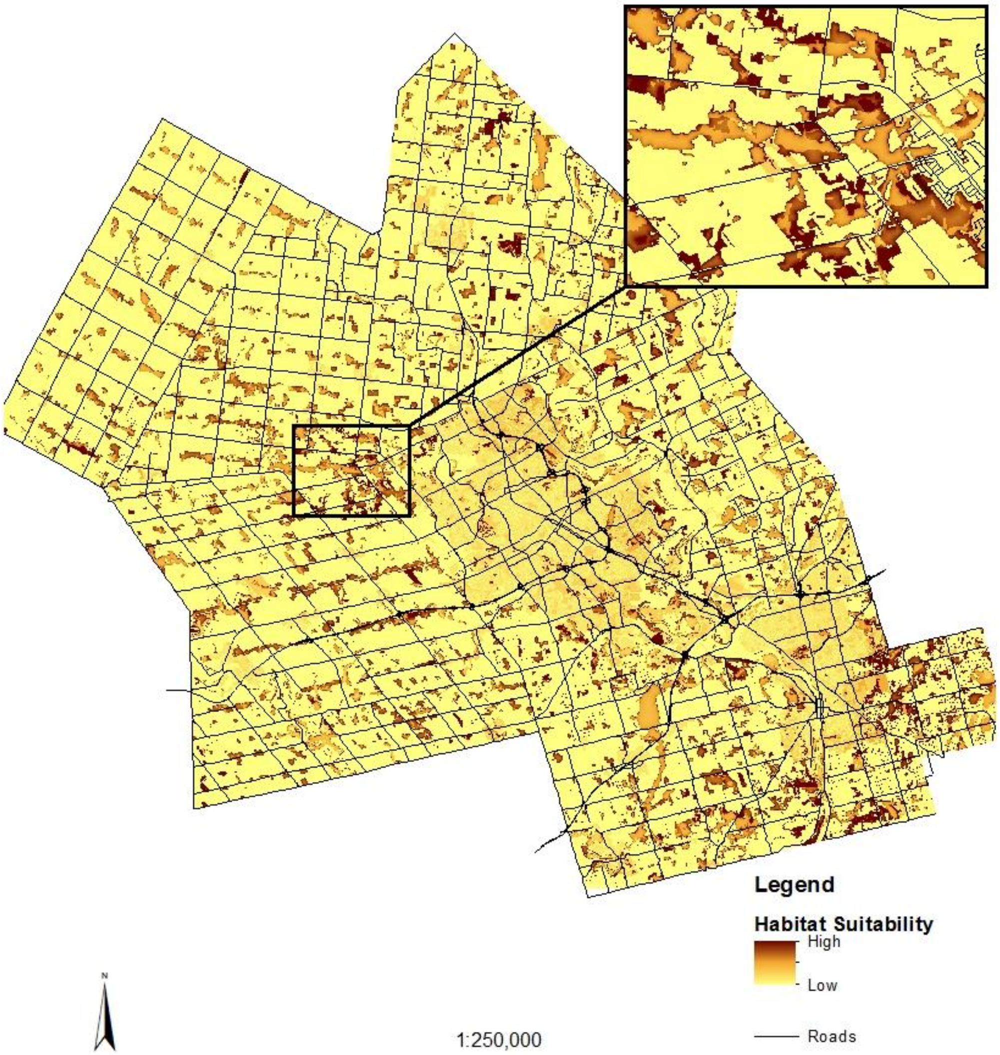

Figure 12: Map of the Region of Waterloo showing the result of applying the habitat suitability model to the habitat characteristics of the whole region. Yellow areas were assigned a habitat suitability of zero because they were not included in the model.

The AUROC value of 0.63 for the habitat suitability model is considered poor discrimination based on the rule of thumb that 0.7 is acceptable discrimination (Hosmer and Lemeshow, 2000). There are several possible explanations for the poor performance of the habitat suitability model. First, buckthorn is a generalist species known to grow on a range of habitats from dry to moist and open fields to forest interiors (Kurylo *et al.*, 2007). The wide range of habitats tolerated by buckthorn may mean that the suitability of habitat is not a major determinant of the distribution of buckthorn. Secondly, buckthorn is an invasive species and is still spreading to suitable habitats, which violates the assumption of species distribution models that the species distribution is at equilibrium across the landscape (Jeschke and Strayner, 2008; Smolik *et al.*, 2010). However, buckthorn was found to be present in all the sampling areas, which indicates that it has had some opportunity to spread to all sampling areas. Nevertheless, the amount of propagule pressure may not be the same in all sampling areas thus leading to variation in the saturation of suitable sites. This issue was dealt with in the spread model by including the effect of dispersal through the number of neighbouring cells containing buckthorn.

Another factor that could have caused errors in the habitat suitability model is the classification of land cover types. The SOLRIS data from the Ontario Ministry of Natural Resources (2008) is based on satellite imagery in order to cover a large area. However, the data does not take into account all aspects of the landscape that are potentially relevant for the spread and establishment of buckthorn. For example, a patch that is labeled by the SOLRIS data as swamp may have standing water year round or may be inundated only at certain times of year. This would impact the ability of buckthorn to grow (Kurylo *et al.*, 2007) but would not necessarily be reflected in the land cover class. An additional source of error in the model is the spatial accuracy of the data. The land cover data as well as the GPS that was used for localizing grid cells during fieldwork both have an accuracy of +/- 10 m. Therefore cells at the intersection of two land cover classes may be classified incorrectly and a buckthorn at the border between two cells may have been assigned to the wrong cell.

An additional factor that may impact the ability of buckthorn to invade an area is its history of disturbance. Field observations showed that although buckthorn was found in every forest, more mature forests appeared to have fewer buckthorns. This might indicate that mature forests have some ability to resist invasion of buckthorn (McCay *et al.*, 2009). Some possible reasons mature forests might be able to resist buckthorn invasion may be lower light levels, greater competition for nutrients or lower propagule pressure because birds are more likely to perch at forest edges than in the forest interior (McCay and McCay, 2009; Yates, Levia, & Williams, 2004). However, sites that were more recently disturbed would have been vulnerable to buckthorn invasion during the early stages of succession allowing it to become established. Although it is encouraging that mature forest may be able to resist buckthorn invasion, it remains troubling that buckthorn was present in all forests, independent of level of maturity, emphasizing the ability of this species to take advantage of any disturbance that does occur. Unfortunately, disturbance data was not available for the whole landscape and therefore this factor could not be included in the model.

Modeling the distribution of buckthorn through a spread model was far more effective than with the habitat suitability model. The results show that although habitat suitability is an important factor for predicting an invasion, including the impact of local dispersal through the presence of buckthorn in first and second order neighbour cells improves the discrimination of the model. This is similar to the results of Fitzgerald *et al.* (2012) who found that including the effect of neighbours improved their predictions of an invasion of Argentine ants. Although spatial autocorrelation is present within the spread model residuals it is very weak, so it is not expected to affect the interpretation of the model. The nonstationarity in the temporal semivariogram may be due to the year variable not capturing the full impact of time in the model. However the semivariance only increases for long time lags for which there is less data. The cumulative residuals do not show a serious misspecification in the model so the model results are still considered valid.

The strongest effect on the probability of a cell transitioning to buckthorn present was the centred habitat suitability. However, the significance of this relationship should be interpreted cautiously because the errors in the habitat suitability model are propagated into the spread model. Therefore there is greater uncertainty in the spread model than is apparent from the standard errors (Table 9).

The presence of temporal autocorrelation in the spread model residuals suggests that it would be useful to compare the results of the GLMM spread model with a model using a GEE approach instead. A GEE model would account for the autocorrelation between observations of the same cell at different times (Zuur *et al.*, 2009). This was not attempted here because of the difficulty in implementing the appropriate association structure for GEE with binary data, which is not currently available in the R GEE packages (Zuur *et al.*, 2009).

The significant relationship between the probability of invasion and year reflects the fact that buckthorn has been continuing to spread across the Region of Waterloo in the past 50 years and therefore closer to the present, more seed sources are available throughout the landscape and the likelihood of invasion is higher. This is also reflected by the fact that the total number of cells in which buckthorn is present has increased exponentially over the last 50 years. This suggests that buckthorn is still actively spreading within the Region of Waterloo and has not reached an equilibrium. Since the invasion probability is continuing to grow, it is important to take action to control buckthorn as soon as possible because it will only get more expensive and difficult to control buckthorn as it becomes more prevalent.

All the limitations and sources of error for the habitat suitability model also affect the spread model. Another limitation of the spread model is the fact that only the presence of buckthorn in first and second order neighbours were used in the model as opposed to the distance to a buckthorn source in the whole population. This has the advantage of decreasing the burden of sampling, since the entire area did not need to be surveyed. However, it removes the possibility of modelling differences in the likelihood of invasion based on the distance from the cell to the nearest buckthorn propagule source. In addition, the model did not consider the effect of buckthorn age on the number of seeds produced (Knight *et al.*, 2007). Doing so might have improved the model's performance but would have added to the complexity of the model and therefore was not attempted. Finally, there may have been errors in the calculation of buckthorn age based on the number of tree rings due to difficulty in deciphering rings and estimates used when the pith of the tree was missed. However, the test of intra-observer error shows that the ages were reasonably reliable, so even if they were somewhat inaccurate the relative ages of buckthorn in different cells should still be meaningful.

In the spread model the intercept plus the coefficient for the year variable can be interpreted as the base level probability that a cell might be invaded by buckthorn over one year if it has average habitat suitability and has no first or second order neighbours with buckthorn, and if the effect of the year is assumed to be the same as in year one of the model. This gives land managers the ability to predict the likelihood of buckthorn invading an area that does not have any buckthorn in the immediate vicinity but where buckthorn is known to be present in the more distant surroundings. This is important for land managers who are working to remove buckthorn from a property because it can be used to model the time it would take for buckthorn to reinvade an area if all the local buckthorn was initially removed. This will allow land managers to show the need for continued long-term management after an initial removal of buckthorn. An example of how simulations can be used with this model to compare the effectiveness of different management strategies is described in the next chapter.

Chapter 4: Applications

To demonstrate how the model of buckthorn spread could be used, it was applied to an example area of habitat in the Region of Waterloo and the future spread of buckthorn was simulated under different management strategies. The management strategies are compared to each other based on the criteria effectiveness and effort and recommendations are made for useful buckthorn control strategies in the Region of Waterloo. For the purposes of this analysis effectiveness is defined as the amount of buckthorn remaining at the site and effort is the amount of work required to control buckthorn.

Comparing the success of different management strategies is important because removal of buckthorn is expensive and the application of sub-optimal approaches is a waste of resources. Chemical girdling by spraying the basal bark with herbicide has been established as the most effective way to reduce the extent and spread of mature buckthorn stands (Delanoy and Archibold, 2007). However, it is less clear following which management strategy (i.e., frequency and extent) is most effective. The cost of chemical girdling treatment is estimated at \$500-\$3,000 /ha depending on the density of buckthorn at the site (Ron Wu-Winters, Personal Communication). After initial removal of mature buckthorn, continued control is needed to remove seedlings arising from the seed bank or from outside seed sources but this does not require chemical treatment since smaller stems can be removed mechanically.

In the context of the Region of Waterloo, where eradication of buckthorn from the entire area is not a feasible objective, ongoing management will be a key step for limiting buckthorn in natural areas because reinvasion from surrounding properties will be a constant possibility. The spread model described above has the advantage of producing a likelihood of buckthorn invasion even for areas that are not directly adjacent to properties that contain buckthorn. The spread model can therefore be used to model the likelihood of reinvasion of a managed natural area even when the location of buckthorn in the surroundings is not known. Although in many cases it might be possible to obtain funding for initial removal of invasive species, it can be more difficult to secure funding for long term management (Ron Wu-Winters, Personal Communication). Having a model that predicts how much management (i.e., frequency and extent) will be required to control buckthorn in the long term, will allow land managers to provide evidence for the need for funding of long-term management. For the most effective control of buckthorn at local sites, it might also be necessary to create management strategies to control buckthorn on a landscape scale. These strategies should include legislative or policy approaches that encourage the management of buckthorn on a variety of property types, such as private and public lands.

Methods

A typical forested patch in the Region of Waterloo was arbitrarily chosen for the simulation. The simulated area is 145,575 m^2, has low permeability, and a mixture of forest and swamp land cover types surrounded by agricultural fields (Figure 13). A 25 m by 25 m grid was superimposed on the simulated area resulting in 1,133 cells. Then, land cover type, distance to forest patch edge, and permeability were used to calculate the habitat suitability of each cell in the simulated area based on the coefficients from the habitat suitability model. The probability of invasion for each cell was then calculated at each time step based on the habitat suitability and the state (i.e., invaded by buckthorn or not) of neighbouring cells using the previously fitted parameters of the buckthorn spread model. Using a random number generator, it was then determined whether the cell was invaded by buckthorn. Once invaded, cells remained invaded unless they were subject to removal of buckthorn.

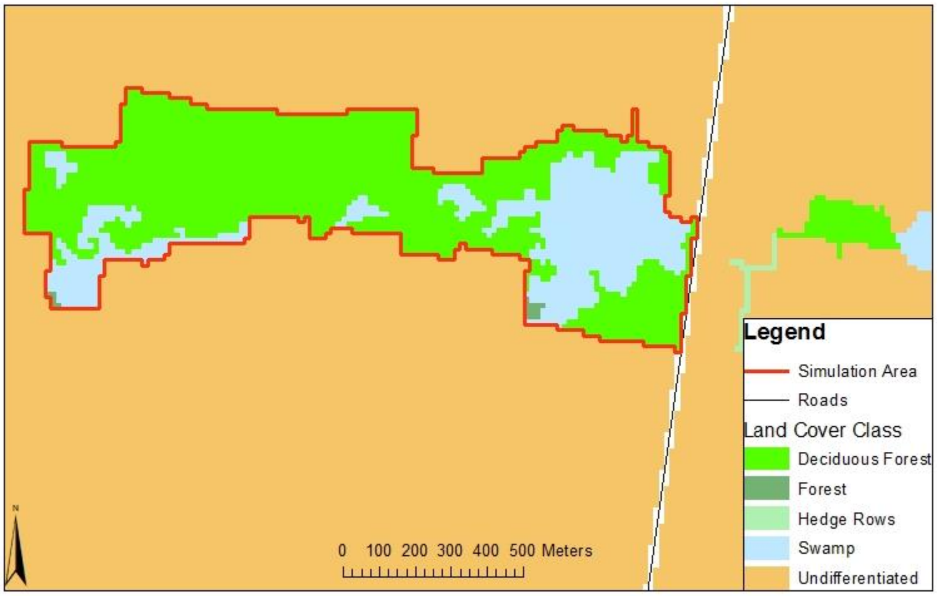

Figure 13: Map of the land cover classes for the example area used in the simulation

Two different management scenarios were simulated and compared. In the first scenario buckthorn removal was carried out over the whole patch at a range of frequencies from every year to every ten years. In the second scenario the simulated area was managed incrementally, the area was broken into subsections and buckthorn was removed from one subsection each year. The simulation was run with different numbers of subsections ranging from two to ten. A third parameter included in the simulation was the effect of different levels of removal success. Removal success is defined as the probability that all buckthorn is successfully removed from a cell to which management is applied. Including removal success allows land managers to know what level of removal success they need to reach in order to get the expected results. Management was included in the simulation by comparing a random number to the removal success for each cell to be managed at that time step. If the random number was less than the removal success parameter the cell was changed from buckthorn present to buckthorn absent.

The R package simecol (Petzoldt and Rinke, 2007) was used to create a simulation model predicting how buckthorn would spread throughout the simulated area. The simulation was run for 50 years with one year time steps and the total number of occupied cells was recorded in each year. The simulation was run over 50 years to show the long-term success of buckthorn control and so that the equilibrium level of buckthorn presence could be determined under each management scenario. The simulation was repeated 1,000 times and the average number of occupied cells calculated for each year. For the scenario where buckthorn was removed at different frequencies a periodic graph was produced which was smoothed by plotting the average of the number of cells with buckthorn present over the years between removals. For example, if removal occurred every five years the sum of the cells with buckthorn present for years one to five was divided by five and this average was used in the graph. This smoothing was done to improve the ease of comparison between different frequencies of buckthorn removal. This was repeated for all levels of the simulation model parameters to reflect the outcomes of different management scenarios. The number of cells occupied by buckthorn, once an apparent equilibrium was reached between reinvasion and buckthorn removal, was used to reflect the effectiveness of the management strategy.

To show the amount of effort needed for each management scenario, a value of one was assigned to each cell where buckthorn was removed and 0.5 for cells that were surveyed but buckthorn was not removed. This reflects the scenario where all mature buckthorns were removed from the patch before the beginning of the simulation and therefore buckthorn removal will only require hand pulling and will not require a large amount of effort relative to surveying. The formula for effort (E) is shown below where R is the number of cells where buckthorn is removed and M is the number of cells that are managed in each time step.

$$E = R + 0.5M$$

The total effort for 50 years of management is the sum of the effort in each year. The amount of effort required for each management scenario was compared by graphing total effort for different levels of each management variable. The different levels of removal success were modelled with management frequency set at every four years. The different levels of management frequency and number of subsections were modelled with removal success set to 0.8. The total effort for 50 years of management and the effectiveness were compared to determine the preferred management scenario.

Results

The number of cells in the simulation area that are occupied by buckthorn varies greatly with the management strategy that is applied. All management strategies result in the number of cells with buckthorn present increasing from the initial condition where zero cells are occupied to an equilibrium once buckthorn removal balances reinvasion. The unsmoothed results (not shown) for management frequency ranging from every year to every ten years, with removal success set at 0.8, show that the number of cells occupied by buckthorn increases quickly each year after buckthorn is removed when there is no management and then drops when buckthorn is removed. When management is carried out every nine or ten years the rate of increase in number of occupied cells starts to slow as the area becomes saturated. For the smoothed graphs the equilibrium number of occupied cells increases with decreasing frequency of management (Figure 14A). After 50 years, the largest difference in the number of cells occupied by buckthorn is between management frequencies every 4 years and every 5 years (Figure 14A). The average number of cells occupied when the area is managed every 4 years is 157.

When the removal success parameter is varied with management frequency held at every 4 years the results show a wide range (Figure 14B). When removal success is zero, i.e. no buckthorn is removed, all 1,133 cells in the simulated area become occupied within 20 years. When removal success is one, i.e. every buckthorn is removed when a cell is managed, the equilibrium average number of occupied cells is 29. There is a large decrease in occupation between removal success of 0.7 where the average number of occupied cells is 607 and removal success of 0.8 where the average number of occupied cells is 156.

When the number of subdivisions is varied with removal success held at 0.8 the resulting graph is similar to when management frequency is varied (Figure 14C). The main difference is that the fluctuations in the number of occupied cells are far less pronounced which is why it was not necessary to smooth the graph. Although all the cells are never occupied, when there are ten subdivisions the equilibrium is reached when approximately 900 cells are occupied. When there are two subdivisions the equilibrium is reached when approximately 20 cells are occupied. The largest increase in occupation occurs between 4 and 5 subdivisions where the numbers of occupied cells are approximately 191 and 510 respectively.

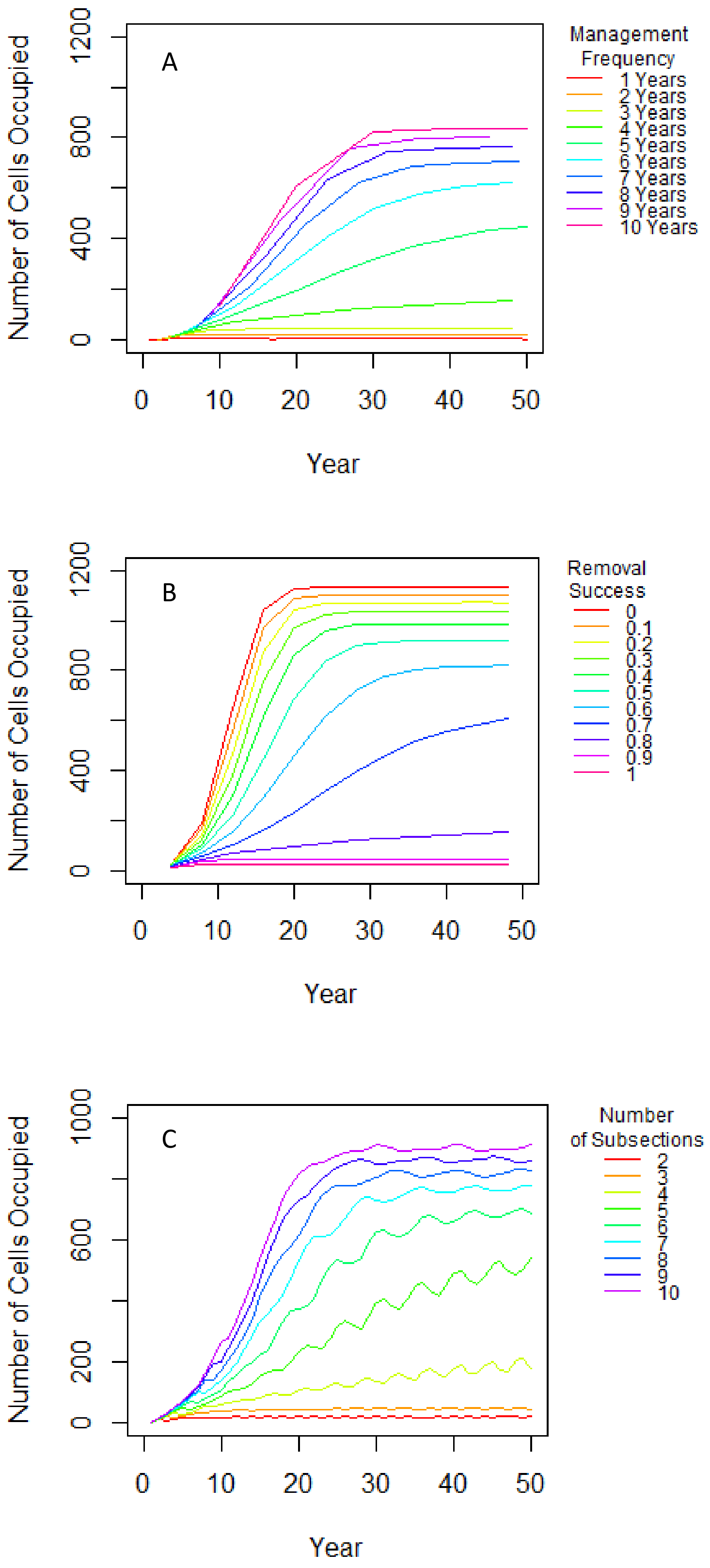

Figure 14: Comparison of the number of cells occupied by buckthorn under different management scenarios after 50 years of buckthorn management in a simulation area in the Region of Waterloo. The simulation area has 1,133 cells that could contain buckthorn. A and B were smoothed by taking the average number of occupied cells over the number of years between buckthorn removals. The 95% confidence intervals are not shown because the widest confidence interval is only ± 4 cells. A) Number of cells occupied over 50 years with management frequency varying from every year to every ten years and removal success set to 0.8. B) Number of cells occupied over 50 years with removal success varying from 0 to 1 and management frequency set to every four years. C) Number of cells occupied over 50 years with number of subsections varying from two to ten, removal success set to 0.8, and assuming one subsection is managed every year.

Total effort for 50 years of buckthorn control was calculated for management frequency ranging from every year to every ten years when removal success was set at 0.8. The plot of effort against management frequency shows that effort is highest for management every year. Effort declines quickly as frequency of management decreases with a local minimum of 8,602 when management occurs every four years. After a slight increase, effort then approaches a minimum of 6,690 at ten years (Figure 15A).

Total effort for 50 years was calculated for simulations where removal success varied from 0.1 to 1 and management frequency was set at every four years. The plot of effort against removal success shows a maximum of 11,985 at 0.6 and a minimum of 7,785 at 1 (Figure 15B).

Total effort for 50 years was calculated for simulations where the number of subsections was varied from two to ten and removal success was set at 0.8. The plot of effort against number of subsections shows that effort decreases with increasing numbers of subsections (Figure 15C). The steepest decline in effort is found between four and six subsections where effort decreases from 6,900 to 6,395 respectively.

To compare the effectiveness and effort required for each management scenario the number of occupied cells was plotted against the total effort for 50 years of management (Figure 16). The strategies that are most effective for the least effort are in the bottom left corner of this graph.

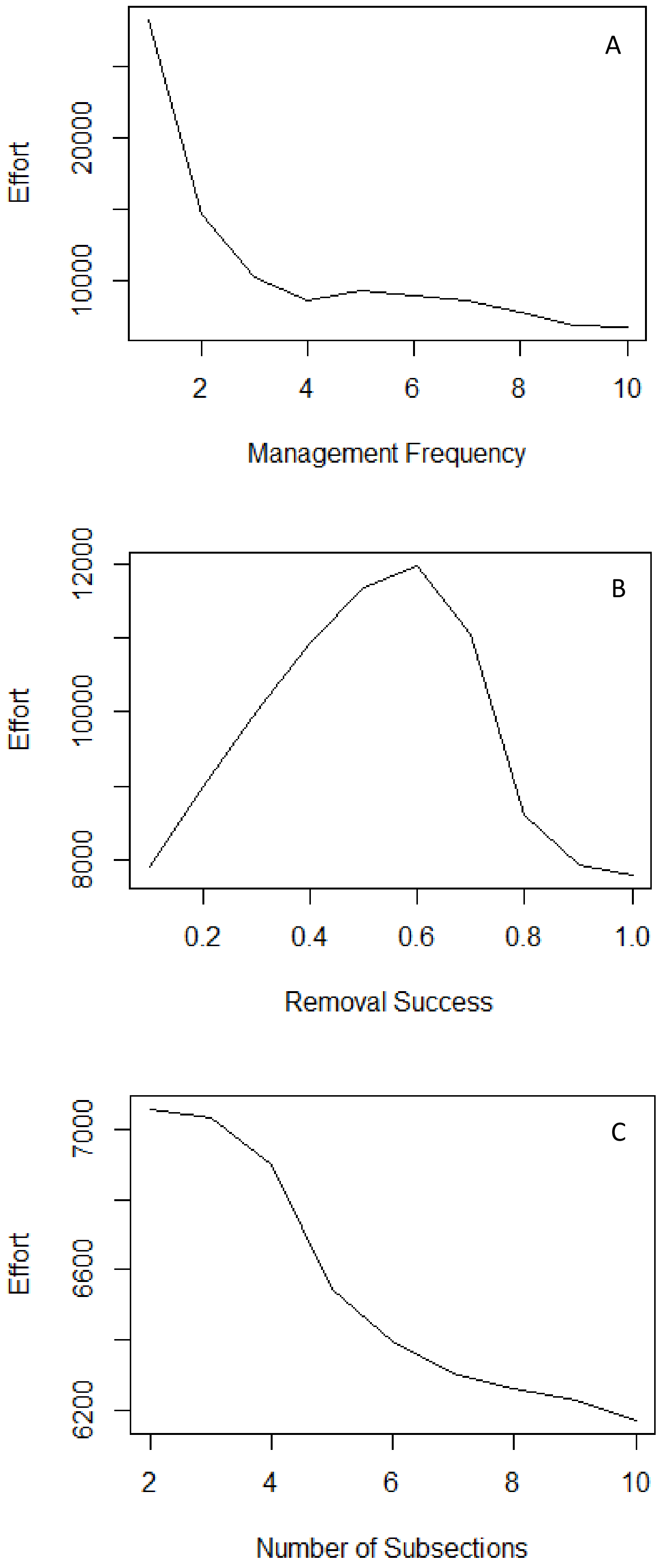

Figure 15: The total effort for management of buckthorn over 50 years in a simulation patch in the Region of Waterloo. The simulation area has 1,133 cells that could contain buckthorn and effort was calculated as 0.5 times the number of cells surveyed plus the number of cells where buckthorn was removed for each year that management occurred. A) The total effort for 50 years when management frequency was varied from every year to every ten years and removal success was set to 0.8. B) The total effort for 50 years when removal success was varied from 0.1 to 1 and management frequency was set to every four years. C) The total effort for 50 years when the number of subsections was varied from two to ten and removal success was set to 0.8, and assuming one subsection is managed every year.

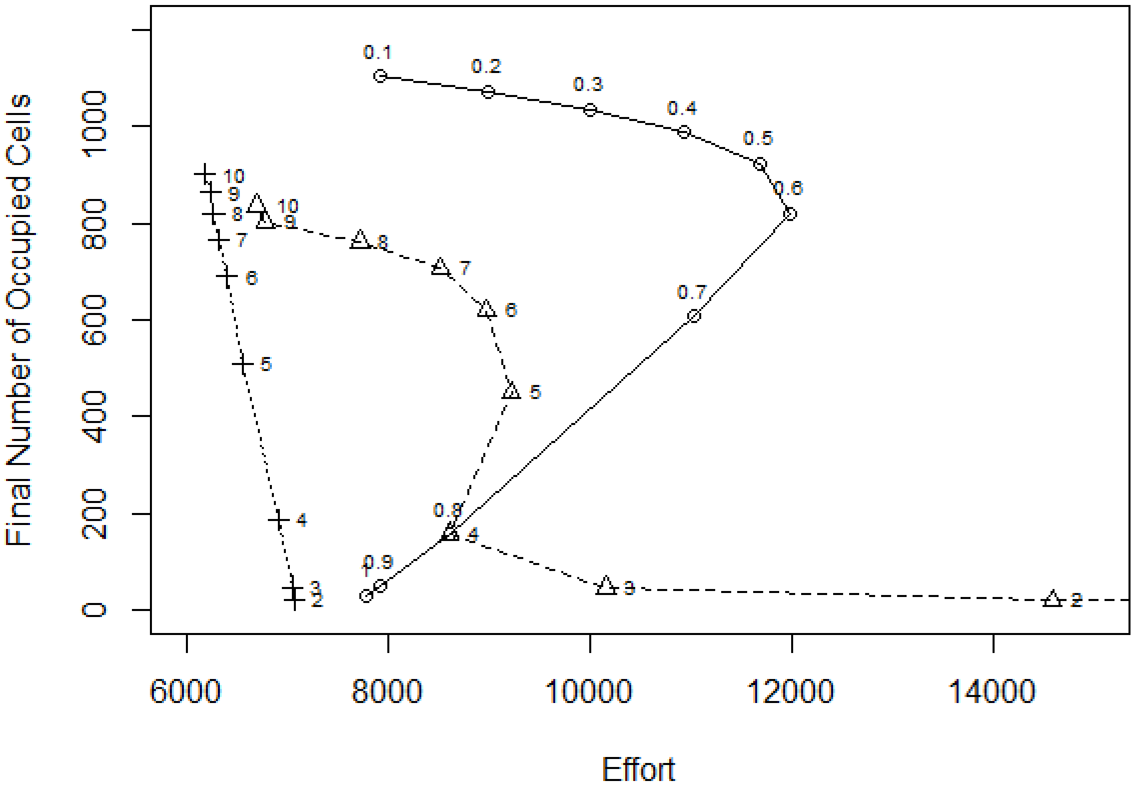

Figure 16: Comparison of total effort and final number of occupied cells after 50 years of management for different management scenarios. The "+" symbol with short dashes symbolizes the results when the model is run with varying numbers of subsections indicated by the number and removal success set to 0.8. The "Δ"symbol with long dashes symbolizes the results when the model is run with varying management frequency and removal success set to 0.8; the frequency is indicated by the number. The "○"symbol with solid line symbolizes the results when the model is run with varying removal success and a frequency of every four years; the level of removal success is indicated by the number.

Discussion

The results of the simulations provide estimates of the relative effort required for different management strategies. The appropriate management strategy will be dependent on the resources and goals of the land managers in each situation, but the process outlined above shows how this problem can be approached. For the simulated area, the recommended management strategy is to break the area into four subsections and remove buckthorn from one subsection each year, i.e., cycling management through all subsections over four years. This is consistent with Delanoy and Archibold (2007) who suggest that treatment every 5 years is an effective for controlling buckthorn populations. The recommended strategy requires 6,550 units of effort and would maintain the buckthorn population at approximately 191 occupied cells assuming that the likelihood of buckthorn being successfully removed from a cell that was managed is 0.8 (Figure 16). This is significantly less effort than if the whole patch was managed every four years which would give 157 cells occupied by buckthorn but would require 8,600 units of effort. This difference is caused by the fact that in the incremental management strategy some part of the area is managed each year so there will be fewer neighbouring cells containing buckthorn, which will slow down the progress of the invasion. Also because a smaller area is managed each year, less effort is required for surveying.

There were two main sources of uncertainty in the simulations. The uncertainty created by the stochastic nature of the spread process was accounted for by averaging the results over 1,000 simulations and is indicated by the confidence intervals (Figure 14). The uncertainty propagated from the spread model was not explicitly included in the results but it is important to note that the predictions include all the sources of error and uncertainty discussed in the previous chapter.

The simulation model was run with the removal success parameter set at a variety of levels, but in reality removal success will not be directly controlled by land managers. However, the results show that aiming for removal success of 0.8, i.e. buckthorn is successfully removed from 80% of cells that are managed, is a desirable goal as there is a large difference in effort between removal success of 0.7 and 0.8. Removal success of greater than 0.8 also leads to lower effort, but the gains are diminishing. Removal success of 0.8 is expected to be an achievable goal based on the experience of Delanoy and Archibold (2007) who found a miss rate of 21% for buckthorn stumps that were not sprayed with herbicide. Land managers should assess the success of their removal efforts to investigate whether the 0.8 removal success goal is being reached. Assessing removal success will be especially important if land managers are deciding between using volunteer or professional labour because if volunteers can achieve high enough removal success it may not be necessary to hire expensive professionals.

The results of the simulation model when removal success is set to zero show what would happen if buckthorn was initially removed from an area and no follow-up monitoring or removal was carried out. By 20 years after the initial removal every cell in the area would contain buckthorn. This demonstrates the importance of long-term buckthorn management, because if long-term management is not carried out, the financial investment for the initial removal will be wasted. Showing the importance of long-term buckthorn management is necessary because it is currently more difficult for land managers to get funding for long-term management then initial removal (Ron Wu-Winters, Personal Communication).

This analysis only considers the effort required for surveying and removal of buckthorn but other costs related to the overhead required to assemble a crew and visit the site every year might make less frequent management of the whole area more efficient. On the other hand, management that occurs every year on a smaller area might provide continuity of the buckthorn control program. This could lead to better management and less training costs as workers benefit from last years’ experience. Also, if volunteers are used for buckthorn monitoring and removal, managing a portion of the site every year could help to keep volunteers engaged from year to year.

In this analysis it was assumed that the simulated area had been cleared of buckthorn before the simulation began. However, if the simulation model were to be run without this assumption it would be important to consider that the effort required to remove buckthorn will change over time. Initial removal of mature buckthorn stands will be far more intensive than long term monitoring and removal of seedlings as they appear. Also, if buckthorn management was very infrequent, trees would become large enough to require chemical treatment instead of only mechanical uprooting, which would increase the effort required for removal. Such changes in effort for different management strategies could be incorporated into the simulation model by linking the effort required to manage a cell to the time since it was invaded. Time since invasion would reflect the greater size of individual trees and the higher density of buckthorn that are likely to be present in cells that were invaded a longer time ago (McCay *et al.*, 2009).

In addition, the simulation model does not account for the presence of seeds in the seed bank, which would be present following the removal of mature buckthorn (Delanoy and Archibold, 2007). As a result, the simulation model probably underestimates the likelihood of a cell becoming invaded in the first five years after the mature buckthorn was removed because buckthorn seeds remain viable for up to five years (Ontario Invasive Plant Council, 2012). It was not possible to include the presence of buckthorn seeds in this analysis because the current distribution of buckthorn in the simulation area is not known. However, this could be modelled by assuming that seedlings would grow out of the seed bank in all cells where mature buckthorn was removed. The likelihood of seedlings growing form the seed bank would have to be determined experimentally.

Another assumption of the spread model is that the influence of an invaded cell on its neighbour cells is independent of the length of time since the cell was invaded. However, this does not reflect the fact that buckthorn trees in a cell that was invaded longer ago will produce more seeds and therefore might increase the likelihood of a neighbouring cell being invaded. Including the time since buckthorn invasion of cells may have improved the spread model, and therefore the simulation, but would have added to its complexity and was therefore not attempted at this time.

A different possible application of this spread model would be to simulate spread at a landscape scale. This would identify patches that are currently known not to be impacted by buckthorn but that are in danger of future invasion. In this way the model could be used to prioritize management of patches that can still be protected from the ecological damage of a buckthorn invasion.

Many additional management actions could be taken to affect buckthorn control, which were not included in the simulation model (Delanoy and Archibold, 2007). For example, best management practices for buckthorn recommend planting native tree species to compete with buckthorn for light and nutrients in the disturbed area (Ontario Invasive Plant Council, 2012). In addition, buckthorn is less common in areas with fewer canopy gaps, so planting could help to make the site more resistant to invasion in the future (McCay *et al.,* 2009). This management action would decrease the likelihood of a cell changing from buckthorn absent to buckthorn present and, though it would be cost intensive, would allow for less frequent management and would improve the overall biodiversity of the site (Ontario Invasive Plant Council, 2012). This management action could be included in the simulation model as an additional factor that would decrease the habitat suitability for buckthorn of cells where planting occurred.

Other strategies could target the population of buckthorn in the surrounding landscape, which would decrease the likelihood of cells being invaded due to long distance dispersal. The spread model shows the impact of the abundance of buckthorn in the landscape through the year variable. As the year moves closer to the present, the likelihood of a cell becoming invaded by buckthorn increases which is a reflection of the increasing abundance of buckthorn in the landscape. The management strategy of reducing the level of buckthorn in the landscape could be included in the simulation model by changing the level of the year variable within the spread model. For example, if the presence of buckthorn in the surrounding landscape could be reduced by 25%, this would be approximately equal to the level of buckthorn in the landscape ten years ago. Then the spread model would be run with the year variable set to ten and the likelihood of a cell with no buckthorn in neighbouring cells being invaded would have decreased, reflecting the lower propagule pressure from the surrounding landscape. Reducing the presence of buckthorn in the landscape would reduce the amount of effort that would be needed to keep buckthorn out of important natural areas but would require the cooperation of a large number of different stakeholders and landowners (Epanchin-Niell and Wilen, 2015). Landscape level management would require the use of government policies and legislation to promote the management of buckthorn on private property (Yu and Leung, 2006).

Landscape level management methods are generally either incentives to encourage management or regulations that require removal or ban proliferation of invasive species. For example, the *Weed Control Act* (1990) gives municipalities the power to order a landowner to remove noxious weeds from their property. This is a provincial law enforced by municipalities that only applies to listed species and only if they threaten agricultural enterprises and it is rarely enforced as relating to invasive species (Credit Valley Conservation, 2009). Although it would be initially effective to enforce the *Weed Control Act*, its enforcement might create an adversarial relationship between landowners and organizations working to prevent the spread of invasive species. Enforcement of the Act would also require municipalities to clean up invasive species on their own properties, which would be extremely expensive. In addition to the expense for municipalities to enforce the *Weed Control Act* it would also be unfair to require a landowner to shoulder the expense of removing buckthorn since they did not cause its arrival on their properties. For these reasons, application of the *Weed Control Act* or the anticipated *Invasive Species Act* (Bill 37, 2015) is unlikely and would be unfair for species that are very widespread.

An alternative to legal enforcement of invasive species removal is the roll out of voluntary programs. These programs can be used by landowners to apply for financial assistance or consulting services that could help them manage invasive species on their property. Some existing provincial programs include invasive species removal as activities that could qualify landowners for financial incentives, such as the Managed Forests Tax Incentive Program (Government of Ontario, 2012), the Conservation Lands Tax Incentive Program (Government of Ontario, 2014) and the Species at Risk Farm Incentive Program (Ontario Soil and Crop Improvement Association, 2014). With more local focus, some conservation authorities and municipalities will provide funding or in kind services to help private landowners manage invasive species (Credit Valley Conservation, 2009). While voluntary removal programs are fairer to landowners, they would create substantial costs for the governments providing these programs. In addition, landscape scale management will only be effective if all landowners implement controls, which creates a situation where one weak link can lead to diminished returns from control for all landowners (Fenichel, Richards, & Shanafelt, 2014; Epanchin-Niell and Wilen, 2015). Therefore cooperation among landowners should be encouraged through the creation of groups that link landowners in the same region who share common weed problems. Municipalities can play a facilitating role in creating and guiding such groups.

Although buckthorn has been recognized as a noxious weed in Ontario since 1990 there has been a continued increase in its presence in Waterloo Region. Therefore, the current strategies employed to limit the spread of buckthorn do not appear to be working. This may be due to ineffective strategies or a lack of available funding to properly implement the strategies. This topic is too broad to explore here but the current project shows that buckthorn presence is increasing rapidly and is likely to continue to do so. Clearly, a revised approach is needed to prevent the negative effects of buckthorn increasing in the future.

The simulation model shows that buckthorn can be controlled on a discrete area of land with relatively infrequent management that is applied consistently over the long-term. This provides land managers with evidence of the feasibility and necessity of long-term management. However, if the goal is management of buckthorn at a landscape scale then the growth of the buckthorn population over the past 50 years shows that new policies and or regulations will need to be implemented to achieve control. Options for encouraging wider control of buckthorn include enforcement of existing regulations and promotion of related incentive programs or the introduction of new targeted policies and programs to help private landowners control buckthorn on their land.

Chapter 5: Conclusion

This project aimed to determine what factors influence the likelihood of an area being invaded by European buckthorn and whether the process of spread can be predicted by a spatio-temporally explicit model. It was determined that buckthorn spread is influenced by the suitability of habitat and the distribution of buckthorn in the immediate surroundings. Habitat suitability was affected by the distance to the edge of a forest patch, the permeability of the surficial material, and the land cover class. Though habitat suitability alone was a poor predictor of buckthorn presence, the spread model, which combined the habitat suitability with the distribution of buckthorn in the immediate surroundings, was successful at predicting the likelihood of an area becoming invaded by buckthorn. A potential use of this model was demonstrated by simulating the spread of buckthorn in an uninvaded patch and the impact of different management strategies. The simulation showed that long term management is required to control buckthorn. The most effective strategy in the sample patch was to break the patch into four subsections and remove buckthorn from one subsection each year, assuming 80% removal success.

Contributions

This project has contributed to research on the modeling of invasive species by providing a method to predict the spread of an invasive plant with limited sampling effort. The use of a GLMM to model spread based on habitat suitability and the presence of buckthorn in first-order and second-order neighbouring cells is a novel approach which provides information on the likelihood of long distance dispersal to any cell in the landscape. One of the important results of the model is that it enables land managers to predict the likelihood of reinvasion of an area where invasive species management has occurred. This is done without requiring extensive surveying to determine the distribution of the invasive species over the whole landscape, which allows limited resources to be targeted at the control of the invasive species directly.

This project has contributed to the field of environmental planning and management by creating a tool for land managers to predict and plan for the spread of buckthorn. This is important because funding for long term management can be hard to obtain, partially caused by difficulty to predict management effectiveness and cost. The findings from this research will allow land managers to provide a scientific argument for the need for long-term management as well as evidence for the most effective management regime. In addition, predicting the cost of long-term management will allow land managers to decide which areas to target for remediation. If long-term funding cannot be secured it would be a waste of resources to attempt removing any buckthorn populations.

The results also show the extent of buckthorn invasion in the Region of Waterloo. Buckthorn was present in all fifteen sampling areas visited. In addition, the examination of tree cores demonstrated that the population of buckthorn has been growing exponentially over the last 50 years. This suggests that buckthorn is a growing problem in the Region for which new initiatives will be required to prevent continued spread because the current management approach seems to have been ineffective at achieving this goal.

This research also included a discussion of current policies and legislation that relate to buckthorn management. The continued spread of buckthorn, despite being on the noxious weed list and being considered a high priority invader, shows that the strategies employed to date have been ineffective at managing buckthorn. Although the new *Invasive Species Act* that is currently being considered by the legislature would give governments the tools to regulate invasive species, this is unlikely to have a large impact on the regulation of buckthorn since similar regulations already apply to buckthorn as a noxious weed under the *Weed Act*. Due to the large amount of buckthorn present on both public and private lands collaborative policies will be needed to help land managers fund and implement buckthorn management.

Research Opportunities

The current research has made apparent several opportunities for future research to expand our understanding of buckthorn invasion in Ontario. A direct extension of this project would be to return to the same sites in five years to analyse the expansion of the buckthorn population. This would allow predictions of the model to be tested against the actual spread of buckthorn and facilitate further improvements to the model. Another interesting extension would be to improve the model by including a measure of the historical disturbance at each site. This could be done by examining historical air photos of each site to investigate if there has been any change in the land cover over the past 50 years. Additionally, landowners at each site could be interviewed to obtain an understanding of past and current land management practices that might be linked to buckthorn spread. For example, woodlots with different levels of wood harvesting and other disturbances might be related to the presence or absence of buckthorn as they open up interior forest areas and enable buckthorn establishment. In addition, this would provide information on any past attempts to manage buckthorn populations at the site, which could help to explain the remaining unexplained variation in buckthorn presence/absence.

Another way this project could be extended would be to perform more in depth analyses of the characteristics of forest patches where buckthorn is present, and especially of those areas where buckthorn presence is very limited. Characteristics that might be of interest would be disturbance history, light availability, soil moisture and soil composition. This would help to discern what factors give forest patches the ability to resist further invasion, once buckthorn has established itself. Better understanding of the factors that help prevent invasion would help land managers make management decisions to protect currently unaffected sites from invasion or prevent reinvasion after remediation of a site. In addition, the abundance and age distribution of buckthorn at each site could be determined and included in the model to estimate demographic effects on buckthorn invasion at different sites. This would give land managers more information about which sites are in danger of becoming severely impacted by buckthorn.

Another topic of interest would be a more extensive analysis of strategies used to combat buckthorn and other invasive species, both at the patch and landscape scale. Strategies that could be compared might include the involvement of citizen scientists for monitoring invasions, use of volunteer labour to remove invasive species from public property, positive incentives compared to negative incentives for private land management, and education of private land managers. Such an analysis would give land managers and policy makers information about which strategies have been effective and whether new initiatives in the Region of Waterloo could be successful.

A better understanding of all cost factors related to buckthorn invasion would help to justify the kind of large scale actions that would be needed to control buckthorn on a landscape scale. For example, how much does the presence of buckthorn affect the abundance of soybean aphids and how is this reflected in the yield lost and pesticide spraying costs? Or, what are the environmental costs to the whole ecosystem of aphid spraying due to the presence of buckthorn? The literature does not appear to contain any studies on the savings that could be achieved by removing buckthorn from an area. Answering these questions might help to convince politicians and the agricultural sector of the importance of combating the spread of buckthorn and help governments and landowners justify buckthorn control programs.

Another area of interest in the management of all invasive species, including buckthorn, is the potential impact by climate change. Kurylo *et al.* (2007) hypothesize that despite rising temperatures the Canadian Shield will remain a barrier to invasion due to the acidic soils derived from the granite parent material. However current distribution maps of buckthorn show that it is present in Parry Sound, Ontario, as well as the Superior National Forest, Minnesota, which are both on the southern edge of the Canadian Shield (EDDMaps Ontario, 2015). To ascertain whether the Canadian Shield will be a significant barrier to buckthorn spread on a regional scale, it would be important to investigate whether these leading buckthorn populations are growing on soils typical of the Canadian Shield or if they are located in small patches of more basic soil. The perceived limit of buckthorn's range by the Canadian Shield could also be related to other factors such as the presence of herbivores, density of shade, or a lack of dispersal agents. For example, in Europe, buckthorn is associated with many more insect and fungal species than in North America (Gassmann and Tosevski, 2014). It would be useful to determine what factors limit the range of buckthorn in its native European habitat to discover whether the same factors will exist in North America.

Final Conclusion

European buckthorn is a growing concern for environmental planners in the Region of Waterloo, this research provides a tool for them to predict the invasion of an area by buckthorn allowing them to make informed decisions on the cost of long term management. The simulation modelling results show that buckthorn presence can be kept to a low level if management is carried out consistently at long intervals. These findings will allow planners to make a case for management of environmentally sensitive areas despite the continued presence of buckthorn in the surrounding landscape.

References

Bill 37, *An Act Respecting Invasive Species*, 1st Sess, 41th Leg, 2015.

Archibold, O. W., Brooks, D., & Delanoy, L. (1997). An investigation of the invasive shrub European buckthorn, Rhamnus cathartica L., near Saskatoon, Saskatchewan. Canadian Field-Naturalist, 111(4), 617-621.

Bahlai, C., Sikkema, S., Hallett, R. H., Newman, J., & Schaafsma, A. (2010). Modeling distribution and abundance of soybean aphid in soybean fields using measurements from the surrounding landscape. Environmental Entomology, 39(1), 50-56.

Bates D, Maechler M, Bolker B and Walker S (2014). “lme4: Linear mixed-effects models using Eigen and S4”. R package version 1.1-7, http://CRAN.R-project.org/package=lme4.

Becker, R. H., Zmijewski, K. A., & Crail, T. (2013). Seeing the forest for the invasives: mapping buckthorn in the Oak Openings. Biological invasions, 15(2), 315-326.

Bhagwat, S. A., Breman, E., Thekaekara, T., Thornton, T. F., & Willis, K. J. (2012). A battle lost? report on two centuries of invasion and management of *Lantana camara* L. in Australia, India and South Africa. Plos One, 7(3), e32407.

Brenning, A. 2012. Spatial cross-validation and bootstrap for the assessment of prediction rules in remote sensing: the R package 'sperrorest'. IEEE International Symposium on Geoscience and Remote Sensing IGARSS, in press.

Catterall, S., Cook, A. R., Marion, G., Butler, A., & Hulme, P. E. (2012). Accounting for uncertainty in colonisation times: A novel approach to modelling the spatio-temporal dynamics of alien invasions using distribution data. Ecography, 35(10), 901-911.

Chong, J., Gruenke, J., Dueck, R., Mayert, W., & Woods, S. (2008). Virulence of oat crown rust [Puccinia coronata f. sp avenae] in Canada during 2002-2006. Canadian Journal of Plant Pathology-Revue Canadienne De Phytopathologie, 30(1), 115-123.

City of Waterloo (1999). Environmental lands acquisition and maintenance policy. Retrieved April 7, 2013, from http://www.waterloo.ca/en/contentresources/resources/government/environmental_land_acquisition_and_maintenance_policy.pdf

Colautti, R. I., Bailey, S. A., van Overdijk, C. D., Amundsen, K., & MacIsaac, H. J. (2006). Characterised and projected costs of nonindigenous species in Canada. Biological Invasions, 8(1), 45-59.

Cook, A., Marion, G., Butler, A., & Gibson, G. (2007). Bayesian inference for the spatio-temporal invasion of alien species. Bulletin of Mathematical Biology, 69(6), 2005-2025.

Crawley, M. J. (2007). The R book. Chichester, England ; Hoboken, N.J.: Chichester, England ; Hoboken, N.J. : Wiley.

Craig, J. (2012). Invasive Plant Management in the CSRD Region: Backgrounder for discussing an invasive species association/committee. Silverwing Ecological Consulting.

Credit Valley Conservation. (2009). An Invasive Species Strategy. Retrieved March 11, 2015 from http://www.creditvalleyca.ca/wp-content/uploads/2011/02/InvasiveSpeciesStrategy-draft.pdf

Delanoy, L., & Archibold, O. W. (2007). Efficacy of control measures for European buckthorn (*Rhamnus cathartica* L.) in Saskatchewan. Environmental management, 40(4), 709-718.

Derickx, L. M., & Antunes, P. M. (2013). A guide to the identification and control of exotic invasive species in Ontario's hardwood forests. Invasive Species Research Institute.

Diggle, P.J. & Ribeiro Jr, P.J. Model Based Geostatistics Springer, New York, 2007

DMTI Digital Elevation Model [computer file]. Markham, Ontario: DMTI Spatial Inc. , [2011].

Dormann, C. F., McPherson, J. M., Araújo, M. B., Bivand, R., Bolliger, J., Carl, G., et al. (2007). Methods to account for spatial autocorrelation in the analysis of species distributional data: A review. Ecography, 30(5), 609-628.

Durrett, R., & Levin, S. A. (1994). Stochastic spatial models - a users guide to ecological applications. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences, 343(1305), 329-350.

EDDMapS Ontario. 2015. Early Detection & Distribution Mapping System for Ontario. The University of Georgia - Center for Invasive Species and Ecosystem Health. Available online at http://www.eddmaps.org/ontario/; last accessed March 21, 2015.

Epanchin-Niell, R. S., & Hastings, A. (2010). Controlling established invaders: Integrating economics and spread dynamics to determine optimal management. Ecology Letters, 13(4), 528-541.

Epanchin-Niell, R. S., & Wilen, J. E. (2015). Individual and cooperative management of invasive species in human-mediated landscapes. American Journal of Agricultural Economics, 97(1), 180-198.

ESRI (Environmental Systems Resource Institute). 2014. ArcGIS 10.2. ESRI, Redlands, California.

Evans JS, Oakleaf J, Cushman SA, Theobald D (2011) An ArcGIS Toolbox for Surface Gradient and Geomorphometric Modeling, version 1.01. Available: http://evansmurphy.wix.com/evansspatial. Accessed: 2014 Apr

Fenichel, E. P., Richards, T. J., & Shanafelt, D. W. (2014). The Control of Invasive Species on Private Property with Neighbor-to-Neighbor Spillovers. Environmental & Resource Economics, 59(2), 231-255.

Fennell, M., Murphy, J. E., Armstrong, C., Gallagher, T., & Osborne, B. (2012). Plant spread simulator: a model for simulating large-scale directed dispersal processes across heterogeneous environments. Ecological modelling, 230, 1-10.

Fitzgerald, K., & Gordon, D. M.. (2012). Effects of Vegetation Cover, Presence of a Native Ant Species, and Human Disturbance on Colonization by Argentine Ants. Conservation Biology, 26(3), 525-538.

Fitzgerald, K., Heller, N., & Gordon, D. M. (2012). Modeling the spread of the Argentine ant into natural areas: Habitat suitability and spread from neighboring sites Elsevier.

Frappier, B., Lee, T. D., Olson, K. F., & Eckert, R. T. (2003). Small-scale invasion pattern, spread rate, and lag-phase behavior of *Rhamnus frangula* L. Forest Ecology and Management, 186(1-3), 1-6.

Gassmann, A., & Tosevski, I. (2014). Biological control of Rhamnus cathartica: is it feasible? A review of work done in 2002-2012. Journal of Applied Entomology, 138(1-2), 1-13.

Grubb, P. J., Lee, W. G., Kollmann, J., & Wilson, J. B. (1996). Interaction of irradiance and soil nutrient supply on growth of seedlings of ten European tall-shrub species and *fagus sylvatica*. Journal of Ecology, , 827-840.

Government of Canada. (2004). An Invasive Alien Species Strategy for Canada. Retrieved March 25, 2013, from http://www.ec.gc.ca/Publications/26E24C67-2299-4E7A-8014-9FB6B80695C5%5Ciassc-sneee_eng.pdf

Government of Ontario. (2012). Common Buckthorn Fact Sheet. Retrieved March 18, 2013, from www.invadingspecies.com/download/publications/fact.sheets/Buckthorn_fact_sheet_web.pdf

Government of Ontario. (2010). Conservation land tax incentive program: Community conservation lands guide Retrieved from http://www.web2.mnr.gov.on.ca/cltip/CLTIP_Community_Conservation_Lands_Guide.pdf

Government of Ontario. (2012). Ontario managed forest tax incentive program (MFTIP) guide Retrieved from http://www.ontario.ca/document/managed-forest-tax-incentive-program-guide

Hastings, A., Cuddington, K., Davies, K., Dugaw, C., Elmendorf, S., Freestone, A., et al. (2005). The spatial spread of invasions: New developments in theory and evidence. Ecology Letters, 8(1), 91-101.

Heimpel, G. E., Frelich, L. E., Landis, D. A., Hopper, K. R., Hoelmer, K. A., Sezen, Z., et al. (2010). European buckthorn and Asian soybean aphid as components of an extensive invasional meltdown in North America. Biological Invasions, 12(9), 2913-2931.

Higgins, S. I., Richardson, D. M., & Cowling, R. M. (1996). Modeling invasive plant spread: The role of plant-environment interactions and model structure. Ecology, 77(7), 2043-2054.

Hobbs, R. J., Higgs, E., & Harris, J. A. (2009). Novel ecosystems: implications for conservation and restoration. Trends in ecology & evolution, 24(11), 599-605.

Holst, K. K. , (2014). gof: Model-diagnostics based on cumulative residuals. R package version 0.9.1.

Hosmer Jr, D. W., & Lemeshow, S. (2000). Applied logistic regression John Wiley & Sons.

International Union for Conservation of Nature. (2011). Invasive Species. Retrieved April 9, 2013, from http://www.iucn.org/about/union/secretariat/offices/iucnmed/iucn_med_programme/species/invasive_species/

Jeschke, J. M., & Strayer, D. L. (2008). Usefulness of bioclimatic models for studying climate change and invasive species. Year in Ecology and Conservation Biology 2008, 1134, 1-24.

Knight, K. S., Kurylo, J. S., Endress, A. G., Stewart, J. R., & Reich, P. B. (2007). Ecology and ecosystem impacts of common buckthorn (*Rhamnus cathartica*): A review. Biological Invasions, 9(8), 925-937.

Kollmann, J., & Grubb, P. J. (1999). Recruitment of fleshy-fruited species under different shrub species: control by under-canopy environment. Ecological Research, 14(1), 9-21.

Kot, M., Lewis, M. A., & vandenDriessche, P. (1996). Dispersal data and the spread of invading organisms. Ecology, 77(7), 2027-2042.

Kurylo, J. S., Knight, K. S., Stewart, J. R., & Endress, A. G. (2007). *Rhamnus cathartica*: Native and naturalized distribution and habitat preferences. Journal of the Torrey Botanical Society, 134(3), 420-430.

Leung, B., Roura-Pascual, N., Bacher, S., Heikkilä, J., Brotons, L., Burgman, M. A., et al. (2012). TEASIng apart alien species risk assessments: A framework for best practices. Ecology Letters, 15(12), 1475-1493.

Lin, D. Y., Wei, L. J., & Ying, Z. (2002). Model-checking techniques based on cumulative residuals. Biometrics, 58(1), 1-12.

Mascaro, J., & Schnitzer, S. A. (2007). *Rhamnus cathartica* L.(common buckthorn) as an ecosystem dominant in southern Wisconsin forests. Northeastern Naturalist, 14(3), 387-402.

McCay, T. S., & McCay, D. H. (2009). Processes regulating the invasion of European buckthorn (*Rhamnus cathartica*) in three habitats of the northeastern United States. Biological Invasions, 11(8), 1835-1844.

McCay, T. S., McCay, D. H., Caragiulo, A. V., & Mandel, T. L. (2009). Demography and distribution of the invasive *Rhamnus cathartica* in habitats of a fragmented landscape. The Journal of the Torrey Botanical Society, 136(1), 110-121.

Morse, N. B., Pellissier, P. A., Cianciola, E. N., Brereton, R. L., Sullivan, M. M., Shonka, N. K., ... & McDowell, W. H. (2014). Novel ecosystems in the Anthropocene: a revision of the novel ecosystem concept for pragmatic applications. Ecol. Soc, 19(2).

Ontario Geological Survey 2010. Surficial geology of Southern Ontario; Ontario Geological Survey, Miscellaneous Release--Data 128-REV ISBN 978-1-4435-2483-4 [DVD] ISBN 978-1-4435-2482-7 [zip file]

Ontario Geological Survey and Planning and Engineering Initiatives Limited 1998. Aggregate resources inventory of the Regional Municipality of Waterloo, townships of North Dumfries, Wellesley, Wilmot, and Woolwich and the cities of Cambridge, Kitchener, and Waterloo; Ontario Geological Survey, Aggregate Resources Inventory Paper 161, 64p Retrieved February 12, 2015 from http://www.geologyontario.mndmf.gov.on.ca/mndmfiles/pub/data/imaging/ARIP161/ARIP161.pdf

Ontario Invasive Plant Council. (2012). Invasive Common (European) Buckthorn: Best Management Practices in Ontario.

Ontario Ministry of Natural Resources. (2009). Biodiversity: Programs and Tools . Retrieved April 7, 2013, from http://www.mnr.gov.on.ca/en/Business/Biodiversity/2ColumnSubPage/STEL02_166973.html

Ontario Soil and Crop Improvement Association, Government of Canada, Government of Ontario, & Ontario Farm Environmental Coalition. (2014). SARFIP: Species at risk farm incentive program 2014 Retrieved from http://www.ontariosoilcrop.org/docs/sarfip_booklet_2014_web.pdf

Pearce, J., & Ferrier, S. (2000). Evaluating the predictive performance of habitat models developed using logistic regression. Ecological Modelling, 133(3), 225-245.

Petzoldt, T. and K. Rinke (2007). simecol: An Object-Oriented Framework for Ecological Modeling in R.
Journal of Statistical Software, 22(9), 1--31. http://www.jstatsoft.org/v22/i09/.

Pilcher, C., Rice, M. E., & Vagts, T. (2005). Economic impact of soybean aphid. Integrated Crop Management, 494(13)

Ragsdale, D. W., Landis, D. A., Brodeur, J., Heimpel, G. E., & Desneux, N. (2011). Ecology and management of the soybean aphid in North America. Annual Review of Entomology, 56, 375-399.

Ranius, T., Niklasson, M., & Berg, N. (2009). A comparison of methods for estimating the age of hollow oaks. Ecoscience, 16(2), 167-174.

Regional Municipality of Waterloo Property Parcels [computer file]. Toronto, Ontario: Teranet Incorporated, [2012].

Region of Waterloo. (2011) Census Bulletin: Agriculture Retrieved February 12, 2015 from http://www.regionofwaterloo.ca/en/doingBusiness/resources/2011_Census_Bulletin_2_Agriculture.PDF

Skellam, J. G. (1951). Random dispersal in theoretical populations. Biometrika, 38(1-2), 196-218.

Smolik, M. G., Dullinger, S., Essl, F., Kleinbauer, I., Leitner, M., Peterseil, J., et al. (2010). Integrating species distribution models and interacting particle systems to predict the spread of an invasive alien plant. Journal of Biogeography, 37(3), 411-422.

Southern Ontario Land Resource Information System (SOLRIS) Land Use Data. Toronto, Ontario: The Ontario Ministry of Natural Resources, 2008.

Thiele, J., & Markussen, B. (2012). Potential of GLMM in modelling invasive spread. CAB Reviews: Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources, 7(16)

United States Department of Agriculture. (2008). Oat crown rust. Retrieved 04/22, 2015, from http://www.ars.usda.gov/Main/docs.htm?docid=9919

Urban Forest Associates Inc. (2002). Invasive Exotic Species Ranking for Southern Ontario. Retrieved March 20, 2013, from http://serontario.org/pdfs/exotics.pdf

Václavík, T., & Meentemeyer, R. K. (2009). Invasive species distribution modeling (iSDM): Are absence data and dispersal constraints needed to predict actual distributions? Ecological Modelling, 220(23), 3248-3258.

Vallet, J., Beaujouan, V., Pithon, J., Roze, F., & Daniel, H. (2010). The effects of urban or rural landscape context and distance from the edge on native woodland plant communities. Biodiversity and Conservation, 19(12), 3375-3392. doi:10.1007/s10531-010-9901-2

Venables, W. N. & Ripley, B. D. (2002) Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0

Weed Control Act. R.S.O. 1990, Chapter W.5

With, K. A. (2004). Assessing the risk of invasive spread in fragmented landscapes. Risk Analysis, 24(4), 803-815.

Wood, S.N. (2011) Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. Journal of the Royal Statistical Society (B) 73(1):3-36

Yates, E. D., Levia, D. F., & Williams, C. L. (2004). Recruitment of three non-native invasive plants into a fragmented forest in southern Illinois. Forest Ecology and Management, 190(2-3), 119-130.

Yu, R., & Leung, P. (2006). Optimal pest management: A reproductive pollutant perspective. International Journal of Pest Management, 52(3), 155-166.

Zhang, H., Xia, Y., Chen, R., Gunzler, D., Tang, W., & Tu, X. (2011). Modeling longitudinal binomial responses: implications from two dueling paradigms. Journal of Applied Statistics, 38(11), 2373-2390. doi:10.1080/02664763.2010.550038

Zuur, A., Ieno, E. N., Walker, N., Saveliev, A. A., & Smith, G. M. (2009). Mixed effects models and extensions in ecology with R Springer.

Appendix A: R Scripts

Beware some variable names are repeated with different values. I recommended to removing all previous values after each section. New sections are denoted by a line of #.

```
##Model simplification with glm ##
data3<-read.table("DataWithoutOther.txt")
attach(data3)
MaxMod<-glm(intBT ~ Material  + Permea + log10(PatchDist + 1):InPatch + LC, family=binomial(logit))
summary(MaxMod)

Mod1<-glm(intBT~Permea+log10(PatchDist + 1):InPatch+LC, family=binomial(logit))
summary(Mod1)

Mod2<-glm(intBT~log10(PatchDist + 1):InPatch+LC, family=binomial(logit))
summary(Mod2)
lrt12<-lrtest(Mod1,Mod2)

Mod3<-glm(intBT~Permea+LC, family=binomial(logit))
summary(Mod3)
lrt13<-lrtest(Mod1,Mod3)

Mod4<-glm(intBT~Permea+log10(PatchDist + 1):InPatch, family=binomial(logit))
summary(Mod4)
lrt14<-lrtest(Mod1,Mod4)


## make table###
models<-c("MaxMod",paste("Mod",c(1:4),sep=""))
formulas<-sapply(models,FUN=formula)
test<-c("None","lrt12","lrt13","lrt14")
lrt<-list(lrt12,lrt13,lrt14)
pvalue<-c("N/A",sapply(lrt,function(x) x$'Pr(>Chisq)'[2]))
LogLik<-c(logLik(Mod1),sapply(lrt,function(x) x$'LogLik'[2]))
SimpResults<-data.frame(models[2:5],as.character(formulas[2:5]),test,LogLik,pvalue)
###################################################################################

#Find minimal acceptable model for habitat suitability with glmer

library(lme4)
data3<-read.table("DataWithoutOther.txt")
glmer1<-glmer(intBT~LC+log10(PatchDist+1):InPatch+Permea
+(1|ID_1),family=binomial(logit),data=data3)
glmer2<-glmer(intBT~LC+log10(PatchDist+1):InPatch +(1|ID_1),family=binomial(logit),data=data3)
lrt12<-lrtest(glmer1,glmer2)
glmer3<-glmer(intBT~LC+Permea +(1|ID_1),family=binomial(logit),data=data3)
```

```
lrt13<-lrtest(glmer1,glmer3)
glmer4<-glmer(intBT~log10(PatchDist+1):InPatch+Permea +(1|ID_1),family=binomial(logit),data=data3)
lrt14<-lrtest(glmer1,glmer4)
#Create table of model simplification results
models<-c(paste("glmer",c(1:4),sep=""))
formulas<-sapply(models,FUN=formula)
test<-c("None","lrt12","lrt13","lrt14")
lrt<-list(lrt12,lrt13,lrt14)
pvalue<-c("N/A",sapply(lrt,function(x) x$'Pr(>Chisq)'[2]))
LogLik<-c(logLik(glmer1),sapply(lrt,function(x) x$'LogLik'[2]))
SimpResults<-data.frame(models[1:4],as.character(formulas[1:4]),test,LogLik,pvalue)

#Try glmm with penalised quazi likelihood
library(MASS)
library("nlme")
correl = corSpher(value = c(100,0.5), form = ~xpoint+ypoint|IDf, nugget = TRUE, fixed = F)
correl = Initialize(correl, data = data3)
fitlc5<-glmmPQL(intBT ~LC+log10(PatchDist+1):InPatch+Permea, random = ~1|IDf, correlation=correl,
data=data3, family=binomial)
summary(fitlc5)
#compare to glm
Mod1<-glm(intBT~Permea+log10(PatchDist + 1):InPatch+LC, family=binomial(logit))
summary(Mod1)

##Correlograms##
library("ncf", lib.loc="C:/Users/Sarah/Documents/R/win-library/3.0")

Correlog3<- spline.correlog(x = xpoint,
                    y = ypoint,
                    z = resid(glmer1,data=data3), xmax = 400)

Correlog2<- spline.correlog(x = xpoint,
                    y = ypoint,
                    z = resid(Mod1,data=data3, type="pearson"), xmax = 400)

Correlog<- spline.correlog(x = xpoint,
                   y = ypoint,
                   z = resid(fitlc5,data=data3, type="pearson"), xmax = 400)


##All Distances##

Correlog2a<- spline.correlog(x = xpoint,
                    y = ypoint,
                    z = residuals(Mod1,data=data3, type="pearson"), xmax = F)
```

```
Correloga<- spline.correlog(x = xpoint,
                          y = ypoint,
                          z = resid(fitlc5,data=data3, type="pearson"), xmax = F)

Correlog3a<- spline.correlog(x = xpoint,
                           y = ypoint,
                           z = resid(glmer1,data=data3), xmax = F)


par(mfrow=c(3,2),cex.main=1, mai=c(0.7,0.7,0.2,0.2))


plot.spline.correlog(Correlog2a)
text(33000,0.9,label="A")

plot.spline.correlog(Correlog2)
text(400,0.9,label="B")

plot.spline.correlog(Correlog3a)
text(33000,0.9,label="C")

plot.spline.correlog(Correlog3)
text(400,0.9,label="D")

plot.spline.correlog(Correloga)
text(33000,0.9,label="E")

plot.spline.correlog(Correlog)
text(400,0.9,label="F")

#Create plots of model predictions
library(lme4)
library(ggplot2)
fm1 <- glmer1
newdat <- expand.grid(
  LC=c("AllForest","BuiltImp","BuiltPerv","Swamp")
  , PatchDist=c(1:300)
  , Permea=c("High","Low","MediumLow","Variable")
  , InPatch=c(0,1)
  , intBT = 0
)
mm <- model.matrix(terms(fm1),newdat)
newdat$intBT <- mm %*% fixef(fm1)
pvar1 <- diag(mm %*% tcrossprod(vcov(fm1),mm))
tvar1 <- pvar1+VarCorr(fm1)$ID_1[1]
tvar1 <-
  newdat <- data.frame(
```

```
    newdat
    , plo = newdat$intBT-1.96*sqrt(pvar1)
    , phi = newdat$intBT+1.96*sqrt(pvar1)
    , tlo = newdat$intBT-1.96*sqrt(tvar1)
    , thi = newdat$intBT+1.96*sqrt(tvar1)
  )
#probability
newdatp<-sapply(newdat[5:9],FUN=function(x) {exp(x)/(1+exp(x))}, simplify=TRUE)
newdat[5:9]<-newdatp
##Plot LC levels when Permea=High
newdatPFS<-subset(newdat,Permea=="High"& InPatch==1&(LC=="AllForest"|LC=="Swamp"))
newdatPBI<-subset(newdat,Permea=="High"&
InPatch==0&(LC=="BuiltImp"|LC=="BuiltPerv")&PatchDist==1)
g0P <- ggplot(newdat,aes(colour=LC))+theme_minimal()+ coord_cartesian(xlim=c(-
10,300),ylim=c(0,0.9))+
  geom_line(data=newdatPFS,aes(x=PatchDist, y=intBT))+
  labs(list(x="Distance to Patch Edge",y="Probability Buckthorn is Present"))
g1P<-g0P + geom_line(data=newdatPFS,aes(x =PatchDist , y = phi),linetype=2)+
  geom_line(data=newdatPFS,aes(x =PatchDist , y = plo),linetype=2)+
  geom_errorbar(data=newdatPBI, aes(x = PatchDist, y =intBT ,ymin = plo, ymax = phi,width=10))+
  geom_point(data=newdatPBI,aes(x =PatchDist , y = intBT))

g2P<-g0P + geom_line(data=newdatPFS,aes(x =PatchDist , y = thi),linetype=2)+
  geom_line(data=newdatPFS,aes(x =PatchDist , y = tlo),linetype=2)+
  geom_errorbar(data=newdatPBI, aes(x = PatchDist, y =intBT ,ymin = tlo, ymax = thi,width=10))+
  geom_point(data=newdatPBI,aes(x =PatchDist , y = intBT))
##Plot Permea Levels when LC=AllForest
newdatL<-subset(newdat,LC=="AllForest"& InPatch==1)
g0L <- ggplot(newdatL, aes(x=PatchDist, y=intBT, colour=Permea))+geom_line()+theme_minimal()+
  coord_cartesian(xlim=c(-10,300),ylim=c(0,0.9))+
  labs(list(x="Distance to Patch Edge",y="Probability Buckthorn is Present"))
g1L<-g0L + geom_line(aes(x =PatchDist , y = phi),linetype=2)+ geom_line(aes(x =PatchDist , y =
plo),linetype=2)
g2L<-g0L + geom_line(aes(x =PatchDist , y = thi),linetype=2)+ geom_line(aes(x =PatchDist , y =
tlo),linetype=2)
#create multiplot function
multiplot <- function(..., plotlist=NULL, file, cols=1, layout=NULL) {
  require(grid)

  # Make a list from the ... arguments and plotlist
  plots <- c(list(...), plotlist)

  numPlots = length(plots)

  # If layout is NULL, then use 'cols' to determine layout
  if (is.null(layout)) {
    # Make the panel
```

```
    # ncol: Number of columns of plots
    # nrow: Number of rows needed, calculated from # of cols
    layout <- matrix(seq(1, cols * ceiling(numPlots/cols)),
                     ncol = cols, nrow = ceiling(numPlots/cols))
  }

  if (numPlots==1) {
    print(plots[[1]])

  } else {
    # Set up the page
    grid.newpage()
    pushViewport(viewport(layout = grid.layout(nrow(layout), ncol(layout))))

    # Make each plot, in the correct location
    for (i in 1:numPlots) {
      # Get the i,j matrix positions of the regions that contain this subplot
      matchidx <- as.data.frame(which(layout == i, arr.ind = TRUE))

      print(plots[[i]], vp = viewport(layout.pos.row = matchidx$row,
                                      layout.pos.col = matchidx$col))
    }
  }
}
multiplot(g1P,g1L,cols=1)
################################################################################

##Leave one out cross-validation for the habitat suitability model based on script
##written by Alexander Brenning

library(lme4)
data3<-read.table("DataWithoutOther.txt")
attach(data3)

data3$PD<-log10(PatchDist+1)*InPatch
detach(data3)
attach(data3)
fit = glmer(intBT~LC+PD+Permea +(1|ID_1),family=binomial(logit),data=data3)
summary(fit)
as.data.frame(coef(fit))


my.model = function(formula, data) {

  glmer(intBT~LC+PD+Permea +(1|ID_1),family=binomial(logit),data=data)
}
```

```
my.predict = function(object, newdata) {


  pred=model.matrix(terms(object),newdata)%*% fixef(object)

  # Now convert logit to prob:
  pred = exp(pred) / (1+exp(pred))
  return(pred)
}


library(sperrorest)

# model formula; it isn't really used because formula is hard-coded in my.model:
fo = intBT ~LC+PD+Permea

my.error = function (obs, pred) {
  require(ROCR)
  if (is.factor(obs)) {
    pos = levels(obs)[2]
    neg = levels(obs)[1]
  } else {

    # if your response variable y is coded numerically, 1 represents an event,
    # and 0 represents a non-event (no buckthorn):
    pos = 1
    neg = 0
  }
  predobj = prediction(pred, factor(obs==pos))
  auroc = performance(predobj, measure = "auc")@y.values[[1]]
  err = list(auroc = auroc)
  err$error = mean((obs == pos) != (pred >= 0.5))
  err$accuracy = 1 - err$error
  return(err)
}

res = sperrorest(formula = fo, data = data3, coords = c("xpoints","ypoints"),
                model.fun=my.model, pred.fun = my.predict,
                smp.fun = partition.factor, smp.args = list(fac="IDf"),
                err.fun = my.error, err.pooled = TRUE, err.unpooled = FALSE)
summary(res$pooled.error)
##################################################################################

### Create table of timeseries with transition as 0 if a cell stayed uninvaded and 1 if invaded###
#Import data for number of neighbours with buckthon and the age in cells
dat<-read.csv("neighwbttimeseries.csv")
agedat<-read.csv("ageincelltimeseries2.csv")
```

```
n2dat<-read.csv("neighwbttimeseries2ndorder.csv")
dat$LC<-agedat$LandCover
n2dat$LC<-agedat$LandCover
dat<-subset(dat,LC!="Other")
n2dat<-subset(n2dat,LC!="Other")
agedat<-subset(agedat,LandCover!="Other")
agedat$xpoint<-data3$xpoint
agedat$ypoint<-data3$ypoint
data3<-read.table("DataWithoutOther.txt")
data3$LC<-factor(data3$LC)
glmer1<-glmer(intBT~LC+log10(PatchDist+1):InPatch+Permea
+(1|ID_1),family=binomial(logit),data=data3)
###add HS to table###
dummy<-model.matrix(intBT~LandCover-1,agedat)
dummy2<-model.matrix(intBT~Permea-1,agedat)
attach(agedat)
PD<-log10(PatchDist+1)*In_Patch
alldummy<-data.frame(intBT,dummy,dummy2,PatchDist,In_Patch,PD)
detach(agedat)
rm(PD)
attach(alldummy)
cf = fixef(glmer1)
intercept = cf[1]
cf = cf[-1]

pred = intercept + cf["LCBuiltImp"]*alldummy$LandCoverBuiltImp +
cf["LCBuiltPerv"]*alldummy$LandCoverBuiltPerv+
  cf["LCSwamp"]*alldummy$LandCoverSwamp+ cf["PermeaLow"]*alldummy$PermeaLow+
cf["PermeaMediumLow"]*alldummy$PermeaMediumLow+
  cf["PermeaVariable"]*alldummy$PermeaVariable+ cf["log10(PatchDist + 1):InPatch"]*alldummy$PD

pred = exp(pred) / (1+exp(pred))
oth<-ifelse(agedat$LandCover=="Other",0,1)
hs1<-pred*oth
agedat$hs<-hs1
detach(alldummy)

###Create table with NA for still has buckthorn in, 1 for changed from no buckthorn to buckthorn and 0
for
###stayed no buckthorn###
fcn<-function(x,y) ifelse(x>0 & y>0,NA,ifelse(x>0 & y<=0,1,0))
transd<-lapply(X=agedat[69:121],FUN=fcn,y=agedat[70:122])
transdf<-data.frame(transd)
transdf2<-transdf[1]
for (i in 1:54){
  transdf2<-data.frame(transdf2,transdf[(1+(54*i))])
}
```

```
transdf2<-data.frame(agedat$TARGET_FID,transdf2)
write.table(transdf2,"timeseriestranistions.txt")

trans<-read.table("timeseriestranistions.txt")

###Make table with transition as response variable###
tsdata<-
data.frame(rep(1,length(trans[,1])),trans[,1],dat[,4],trans[2],dat[,14],n2dat[,14],agedat[,59],agedat[,125
],dat[,8],n2dat[,67],agedat[,41],agedat[,67],agedat[,123],agedat[124])
colnames(tsdata)<-
c("year","TARGET_FID","ID_1","trans","neigh1","neigh2","LC","hs","FREQUENCY","nhood2","PatchDist",
"InPatch","xpoint","ypoint")
z<-
data.frame(rep(2,length(trans[,1])),trans[,1],dat[,4],trans[,3],dat[,15],n2dat[,15],agedat[,59],agedat[,125
],dat[,8],n2dat[,67],agedat[,41],agedat[,67],agedat[,123],agedat[124])
colnames(z)<-
c("year","TARGET_FID","ID_1","trans","neigh1","neigh2","LC","hs","FREQUENCY","nhood2","PatchDist",
"InPatch","xpoint","ypoint")
tsdata<-rbind(tsdata,z)
###automate it###
for (i in 3:53){
  z<-
data.frame(rep(i,length(trans[,1])),trans[,1],dat[,4],trans[,(i+1)],dat[,(i+13)],n2dat[,(i+13)],agedat[,59],ag
edat[,125],dat[,8],n2dat[,67],agedat[,41],agedat[,67],agedat[,123],agedat[124])
  colnames(z)<-
c("year","TARGET_FID","ID_1","trans","neigh1","neigh2","LC","hs","FREQUENCY","nhood2","PatchDist",
"InPatch","xpoint","ypoint")
  tsdata<-rbind(tsdata,z)
}
none<-ifelse(tsdata$neigh1<0,0,1)
tsdata$neigh1<-tsdata$neigh1*none
none2<-ifelse(tsdata$neigh2<0,0,1)
tsdata$neigh2<-tsdata$neigh2*none2
tsdata = tsdata[ !is.na(tsdata$trans) , ]
tsdata<-subset(tsdata,LC!="Other")
tsdata<-subset(tsdata,FREQUENCY==9)
tsdata$hsc<-(tsdata$hs-mean(tsdata$hs))
write.table(tsdata,"tstransglmmdat2.txt")

tsdata<-read.table("tstransglmmdat2.txt")
###################################################################################

##Create spread model using results from habitat suitability model#
#load required packages
library(lme4)
library("ggplot2", lib.loc="~/R/win-library/3.1")
library(geoR)
```

```
#create multiplot function
multiplot <- function(..., plotlist=NULL, file, cols=1, layout=NULL) {
  require(grid)

  # Make a list from the ... arguments and plotlist
  plots <- c(list(...), plotlist)

  numPlots = length(plots)

  # If layout is NULL, then use 'cols' to determine layout
  if (is.null(layout)) {
    # Make the panel
    # ncol: Number of columns of plots
    # nrow: Number of rows needed, calculated from # of cols
    layout <- matrix(seq(1, cols * ceiling(numPlots/cols)),
                     ncol = cols, nrow = ceiling(numPlots/cols))
  }

  if (numPlots==1) {
    print(plots[[1]])

  } else {
    # Set up the page
    grid.newpage()
    pushViewport(viewport(layout = grid.layout(nrow(layout), ncol(layout))))

    # Make each plot, in the correct location
    for (i in 1:numPlots) {
      # Get the i,j matrix positions of the regions that contain this subplot
      matchidx <- as.data.frame(which(layout == i, arr.ind = TRUE))

      print(plots[[i]], vp = viewport(layout.pos.row = matchidx$row,
                                      layout.pos.col = matchidx$col))
    }
  }
}
#load timeseries data created with timeseriestransition script
tsdata2<-read.table("tstransglmmdat2.txt")
#remove cells that do not have full number of second order neighbours
tsdata2<-subset(tsdata2,nhood2==21)
tsdata2$ID_1<-as.factor(tsdata2$ID_1)
#Remove data for the most recent 10 years
tsdata5<-subset(tsdata2,year>10)
#Run model#
tsglmer<-
glmer(trans~I(neigh1>0)+neigh1+I(neigh2>0)+neigh2+hsc+year+(1|ID_1),family=binomial,data=tsdata5)
summary(tsglmer)
```

```
## Create Semivariograms
#get residuals from model
res<-resid(tsglmer)
tsdata6<-data.frame(tsdata5,res)
#create subset for years 20 to 10
tsdata6<-subset(tsdata6,year<20)
par(mfrow=c(1,1),cex.main=1,mar=c(5.1,4.1,2.1,2.1))
#Caluculate the spatial distance between all pairs of cells
dists<-dist(c(tsdata6$xpoint,tsdata6$ypoint))

breaks<-seq(from=0,to=400,by=25)
coords<-matrix(c(tsdata6$xpoint,tsdata6$ypoint),ncol=2)

v2<-variog(coords =coords , data = tsdata6$res, uvec=25*(0:12))
plot(v2,type="b",xlab="Distance",ylab="Semivariance")
plot(seq(0,400),seq(-1.2,1.2,length.out=401),type="n",
    xlab="Distance",ylab="Semivariance")
lines(v2$u,v2$v+v2$sd*1.96)
lines(v2$u,v2$v-v2$sd*1.96)
points(v2$u,v2$v)


##Serial Autocorrelation##
#Randomly select subset of smapling areas#
ID<-levels(tsdata5$ID_1)
csam<-as.list(sample(ID,size=5))
res<-resid(tsglmer)
t1<-0
tsres<-data.frame(tsdata5,res,t1)
fun<-function(x){subset(tsres,ID_1==x)}
csub<-lapply(csam,FUN=fun)
csub<-rbind(csub[[1]],csub[[2]],csub[[3]],csub[[4]],csub[[5]])
#Create "coordinates in time" (0,year)
tscoords<-matrix(c(csub$t1,csub$year),ncol=2)

TF<-unique(csub$TARGET_FID)
series<-lapply(TF,FUN=function(x){subset(csub,TARGET_FID==x)})
tscoordsl<-lapply(series,function(x){matrix(c(x$t1,x$year),ncol=2)})
tsvl<-NULL
tsvmv<-NULL
#Calculate semivariogram for each cell then take the average
for (i in 1:456){
  tsv<-variog(coords=tscoordsl[[i]],data=series[[i]]$res)
  tsvmv<-rbind(tsvmv,tsv$v)
  tsvl<-c(tsvl,tsv)
}
```

```
tsvmeanv<-apply(tsvmv,MARGIN=2,FUN=mean)
par(mgp=c(2,1,0),mar=c(4.1,3.1,1,1.5))
plot(tsvl[[1]],tsvmeanv, xlab="Lag Time in Years",ylab="Mean Semivariance")

##Create Graphs of Predictions##
fm1 <- tsglmer
newdat <- expand.grid(
  hsc=c(seq(-0.25,0.5,0.01))
  , year=c(1)
  , neigh1=c(0,1,4,8)
  , neigh2=c(0,1,4,8)
  , trans = 0
)
mm <- model.matrix(terms(fm1),newdat)
newdat$trans <- mm %*% fixef(fm1)
pvar1 <- diag(mm %*% tcrossprod(vcov(fm1),mm))
tvar1 <- pvar1+VarCorr(fm1)$ID_1[1]
tvar1 <-
  newdat <- data.frame(
    newdat
    , plo = newdat$trans-1.96*sqrt(pvar1)
    , phi = newdat$trans+1.96*sqrt(pvar1)
    , tlo = newdat$trans-1.96*sqrt(tvar1)
    , thi = newdat$trans+1.96*sqrt(tvar1)
  )
#probability
newdatp<-sapply(newdat[5:9],FUN=function(x) {exp(x)/(1+exp(x))}, simplify=TRUE)
newdat[5:9]<-newdatp
##Plot neigh1 levels when neigh2=0 and year=1
newdatn1<-subset(newdat,neigh2==0)
g0a <- ggplot(newdatn1, aes(x=hsc, y=trans, colour=as.factor(neigh1)))+geom_line()+
  labs(list(x="Centred Habitat Suitability",y="Probability of Transition\nto Buckthorn Present",
          colour="Number of\n1st Order\nNeighbours"))
g1a<-g0a + geom_line(aes(x =hsc , y = phi),linetype=2)+ geom_line(aes(x =hsc , y = plo),linetype=2)
g1a<-g1a+theme_minimal()
g2a<-g0a + geom_line(aes(x =hsc , y = thi),linetype=2)+ geom_line(aes(x =hsc , y = tlo),linetype=2)
g2a<-g2a+theme_minimal()
multiplot(g1,g2,cols=1)
##Plot neigh2 levels when neigh1=0 and year=1
newdatn2<-subset(newdat,neigh1==0)
g0b <- ggplot(newdatn2, aes(x=hsc, y=trans, colour=as.factor(neigh2)))+geom_line()+
  labs(list(x="Centred Habitat Suitability",y="Probability of Transition\nto Buckthorn Present",
          colour="Number of\n2nd Order\nNeighbours"))
g1b<-g0b + geom_line(aes(x =hsc , y = phi),linetype=2)+ geom_line(aes(x =hsc , y = plo),linetype=2)
g1b<-g1b+theme_minimal()
g2b<-g0b + geom_line(aes(x =hsc , y = thi),linetype=2)+ geom_line(aes(x =hsc , y = tlo),linetype=2)
g2b<-g2b+theme_minimal()
```

```
multiplot(g1,g2,cols=1)

##Redo predictions with year variable neigh=0 hsc=0

fm1 <- tsglmer
newdat <- expand.grid(
  hsc=c(0)
  , year=c(1:54)
  , neigh1=c(0)
  , neigh2=c(0)
  , trans = 0
)
mm <- model.matrix(terms(fm1),newdat)
newdat$trans <- mm %*% fixef(fm1)
pvar1 <- diag(mm %*% tcrossprod(vcov(fm1),mm))
tvar1 <- pvar1+VarCorr(fm1)$ID_1[1]
tvar1 <-
  newdat <- data.frame(
    newdat
    , plo = newdat$trans-1.96*sqrt(pvar1)
    , phi = newdat$trans+1.96*sqrt(pvar1)
    , tlo = newdat$trans-1.96*sqrt(tvar1)
    , thi = newdat$trans+1.96*sqrt(tvar1)
  )
#plot confidence
newdatp<-sapply(newdat[5:9],FUN=function(x) {exp(x)/(1+exp(x))}, simplify=TRUE)
newdat[5:9]<-newdatp
newdatc<-newdat
g0c <- ggplot(newdat, aes(x=year, y=trans))+geom_line()+
  labs(list(x="Year",y="Probability of Transition\nto Buckthorn Present"))
g1c<-g0c + geom_line(aes(x =year , y = phi),linetype=2)+ geom_line(aes(x =year , y = plo),linetype=2)
g1c<-g1c+theme_minimal()
g2c<-g0c + geom_line(aes(x =year , y = thi),linetype=2)+ geom_line(aes(x =year , y = tlo),linetype=2)
g2c<-g2c+theme_minimal()
multiplot(g1c,g2c,cols=1)

##Graph of trans vs neigh1
fm1 <- tsglmer
newdat <- expand.grid(
  hsc=c(0)
  , year=c(1)
  , neigh1=c(0:8)
  , neigh2=c(0)
  , trans = 0
)
mm <- model.matrix(terms(fm1),newdat)
newdat$trans <- mm %*% fixef(fm1)
```

```
pvar1 <- diag(mm %*% tcrossprod(vcov(fm1),mm))
tvar1 <- pvar1+VarCorr(fm1)$ID_1[1]
tvar1 <-
  newdat <- data.frame(
    newdat
    , plo = newdat$trans-1.96*sqrt(pvar1)
    , phi = newdat$trans+1.96*sqrt(pvar1)
    , tlo = newdat$trans-1.96*sqrt(tvar1)
    , thi = newdat$trans+1.96*sqrt(tvar1)
  )
#plot confidence
newdatp<-sapply(newdat[5:9],FUN=function(x) {exp(x)/(1+exp(x))}, simplify=TRUE)
newdat[5:9]<-newdatp
newdatd<-newdat
g0d <- ggplot(newdatd, aes(x=neigh1, y=trans))+geom_line()+
  labs(list(x="Number of First Order Neighbours",y="Probability of Transition\nto Buckthorn Present"))
g1d<-g0d + geom_line(aes(x =neigh1 , y = phi),linetype=2)+ geom_line(aes(x =neigh1 , y =
plo),linetype=2)
g1d<-g1d+theme_minimal()
g2d<-g0d + geom_line(aes(x =neigh1 , y = thi),linetype=2)+ geom_line(aes(x =neigh1 , y = tlo),linetype=2)
g2d<-g2d+theme_minimal()
multiplot(g1a,g1b,g1c,g1d,cols=1)

##Create cumulative residual plots##
tsglm4<-
glm(trans~I(neigh1>0)+neigh1+I(neigh2>0)+neigh2+hsc+year+ID_1,family=binomial(logit),data=tsdata5)
summary(tsglm4)
library("gof", lib.loc="C:/Users/Sarah/Documents/R/win-library/3.0")
crtsglm5<-cumres(tsglm4,variable=c("Number of First Order Neighbours"=tsdata5$neigh1,"Number of
Second Order Neighbours"=tsdata5$neigh2,"Centred Habitat
Suitability"=tsdata5$hsc,"Year"=tsdata5$year))
xlabs<-c("Predicted","Number of First Order Neighbours","Number of Second Order
Neighbours","Centred Habitat Suitability","Year")
for (i in 1:5){
  plot(crtsglm4,idx=i,ylab="Cumulative Residuals",xlab=xlabs[i])
}

## Create GAM of neigh1 and neigh2 variables##
library("mgcv")
gam1<-gam(trans~s(neigh1,k=9)+s(neigh2,k=13)+hsc+year+ID_1,family=binomial(logit),data=tsdata5)
summary(gam1)
plot(gam1)
par(mfrow=c(2,1),mgp=c(2,1,0))
plot(gam1,select=1,xlab="Number of First Order Neighbours",ylab="Contribution to Fitted Values")
plot(gam1,select=2,xlab="Number of Second Order Neighbours",ylab="Contribution to Fitted Values")

### Graph of increase in buckthorn numbers over time ###
```

```
agedat<-read.csv("ageincelltimeseries2.csv")
btt<-0
btpy<-c(0,0)
for (i in 0:54){
  sub<-subset(agedat, intBT==1 & Age>=i)
  btt<-nrow(sub)
  btpy<-rbind(btpy,c(i,btt))
}
btpy<-btpy[-1,]
btpy<-data.frame(btpy,2013:1959)
lbtpy<-log((btpy$numbt/3323)/(1-(btpy$numbt/3323)))
names(btpy)<-c("yearbp","numbt","years")
btpy2<-rbind(subset(btpy,years==2013),subset(btpy,years<2003))
plot(btpy2$years,btpy2$numbt,xlab="Year",ylab="Number of Cells Occupied",pch=16)
points(subset(btpy,years>=2003&years!=2013)[,3],subset(btpy,years>=2003&years!=2013)[,2])
points(subset(btpy,years==2013)[,3],subset(btpy,years==2013)[,2],pch=16)
######################################################################################

##leave one out cross validation for spread model based on script written by Alexander Brenning
data3<-tsdata2
attach(data3)

fit =
glmer(trans~I(neigh1>0)+neigh1+I(neigh2>0)+neigh2+hsc+year+(1|ID_1),family=binomial,data=tsdata5)

summary(fit)
as.data.frame(coef(fit))


my.model = function(formula, data) {
  ### note that I am hard-coding the formula and the name of the cluster variable, CID,
  ### please change as needed:
  glmer(trans~I(neigh1>0)+neigh1+I(neigh2>0)+neigh2+hsc+year+(1|ID_1),family=binomial,data=data)
}

my.predict = function(object, newdata) {

  pred = model.matrix(terms(object),newdata)%*% fixef(object)
  # Now convert logit to prob:
  pred = exp(pred) / (1+exp(pred))
  return(pred)
}


library(sperrorest)

# model formula; it isn't really used because formula is hard-coded in my.model:
```

```
fo = trans~I(neigh1>0)+neigh1+I(neigh2>0)+neigh2+hsc

my.error = function (obs, pred) {
  require(ROCR)
  if (is.factor(obs)) {
    pos = levels(obs)[2]
    neg = levels(obs)[1]
  } else {

    # if your response variable y is coded numerically, 1 represents an event,
    # and 0 represents a non-event (no buckthorn):
    pos = 1
    neg = 0
  }
  predobj = prediction(pred, factor(obs==pos))
  auroc = performance(predobj, measure = "auc")@y.values[[1]]
  err = list(auroc = auroc)
  err$error = mean((obs == pos) != (pred >= 0.5))
  err$accuracy = 1 - err$error
  return(err)
}


res = sperrorest(formula = fo, data = data3,
                model.fun=my.model, pred.fun = my.predict,
                smp.fun = partition.factor, smp.args = list(fac="ID_1"),
                err.fun = my.error, err.pooled = TRUE, err.unpooled = FALSE)
summary(res$pooled.error)
detach(data3)
#########################################################################################

##Simulation of buckthorn growth and management##

#Install and load required packages
library("ggplot2")
library("plotrix")
library("simecol")
library("raster")
library("lme4")
#load data to make model
tsdata5<-read.table("tsdataless10.txt")
#load raster of simulation patch habitat suitability from ArcMap
sphs<-raster("simpatchhs1")
#convert to matrix
sphsm<-as.matrix(sphs)
#subtract mean hs to centre
sphsc<-sphsm-0.2839319
```

```
#run model
tsglmer<-
glmer(trans~I(neigh1>0)+neigh1+I(neigh2>0)+neigh2+hsc+year+(1|ID_1),family=binomial,data=tsdata5)
summary(tsglmer)
#extract coefficients
cf = fixef(tsglmer)
intercept = cf[1]
cf = cf[-1]
#matrix with 1 for areas with hs>0
x2<-ifelse(sphsm==0,0,1)
#2nd order neighbourhood
wdist2<-matrix(c(0,1,1,1,0,
                1,0,0,0,1,
                1,0,0,0,1,
                1,0,0,0,1,
                0,1,1,1,0),nrow=5)
#initial conditions
initial2<-matrix(0,nrow=38,ncol=80)
#create function to split the site in to the desired number of divisions for incremental management
#f is the number of divisions
incremental<-function(f) {mp<-NULL
                       c<-1
                       r<-0
                       b<-0
                       y<-0
  for (i in 1:f) {
    if (f==1) a<-x2 else a<-0
    while(b<1133/f) {
      if (f==1) break
      c<-c+ifelse(r==38,1,0)
      if (c>80) break
      r<-r+ifelse(r==38,-37,1)
      y<-x2[r,c]
      a<-c(a,y)
      b<-sum(a)
      }
    mpi<-a
    mpi[]<-i
    mp<-c(mp,mpi[2:length(mpi)])
    b<-0
    a<-0
    y<-0
    }
  mp<-matrix(mp,nrow=38)
}

#Create a model where buckthorn is managed incrementally
```

```
#Where f is the number of divisions and manage is the likelihood that buckthorn will be removed when
#a cell is managed (Removal success),year is the number of years before the present, intercept and cf
#are coefficients from the spread model

CA4 <- new("gridModel",
        main = function(time, init, parms) {
          z <- init
          n1<-eightneighbours(z)
          n2<-neighbours(z,wdist=wdist2)
          mp<-incremental(f)
          pgen <- parms$intercept + parms$cf["I(neigh1 > 0)TRUE"]*n1*ifelse(n1==0,0,1)+
parms$cf["neigh1"]*n1+parms$cf["I(neigh2 > 0)TRUE"]*n2*ifelse(n2==0,0,1)+parms$cf["neigh2"]*n2
+parms$cf["hsc"]*parms$hsc+parms$cf["year"]*parms$year
          pgen = exp(pgen) / (1+exp(pgen))
          zgen <- ifelse(z == 0 &
                    runif(z) < pgen, 1, 0)
          zsurv <-
ifelse(eval(parse(text=paste("mp+1==time|",paste("mp+",paste(seq(1,24,1)*f,sep=""),"==time","|",sep=
"",collapse=""),
                      "mp+",as.character(25*f),"==time",sep="",collapse="")))& (runif(z) <manage),0, 1)
          (zgen+z)*zsurv*x
        },
        parms = list(cf=cf, intercept=intercept,hsc=sphsc,x=x2,f=1,year=1,manage=0.8),
        times = c(from = 1, to = 30, by = 1),
        init = initial2,
        solver = "iteration"
)

CA4 <- sim(CA4)
plot(CA4,delay=200)

#Create a model where buckthorn is managed at different frequencies, yman= frequency of
# management in years, manage=likelihood that buckthorn will be removed when a cell is managed
CA2 <- new("gridModel",
         main = function(time, init, parms) {
          z <- init
          n1<-eightneighbours(z)
          n2<-neighbours(z,wdist=wdist2)
          pgen <- parms$intercept + parms$cf["I(neigh1 > 0)TRUE"]*n1*ifelse(n1==0,0,1)+
parms$cf["neigh1"]*n1+parms$cf["I(neigh2 > 0)TRUE"]*n2*ifelse(n2==0,0,1)+parms$cf["neigh2"]*n2
+parms$cf["hsc"]*parms$hsc+parms$cf["year"]*parms$year
          pgen = exp(pgen) / (1+exp(pgen))
          zgen <- ifelse(z == 0 &
                    runif(z) < pgen, 1, 0)
          zsurv <- ifelse(time%%yman==0
                    & (runif(z) <manage),0, 1)
          (zgen+z)*zsurv*x
```

```
        },
        parms = list(cf=cf, intercept=intercept,hsc=sphsc,x=x2,manage=0.8,yman=5,year=1),
        times = c(from = 1, to = 100, by = 1),
        init = initial2,
        solver = "iteration"
)
CA2<-sim(CA2)
### make graphs of average number of cells occupied in each year after 1000 iterations###
CAmeangraph<-function(model,times,parms,initial,plottitle){
  parms(model)<-parms
  init(model)<-initial
  times(model)<-times
  model<-sim(model)
  sCA2o<-data.frame(lapply(model@out,sum))
  colnames(sCA2o)<-c(paste(1:as.numeric(times[2]),sep=","))
  levelmanage<-sCA2o
  for (i in 1:1000) {
    model<-sim(model)
    sCA2o<-data.frame(lapply(model@out,sum))
    colnames(sCA2o)<-c(paste(1:as.numeric(times[2]),sep=","))
    colnames(levelmanage)<-c(paste(1:as.numeric(times[2]),sep=","))
    levelmanage<-data.frame(rbind(levelmanage,sCA2o))
  }
  levelmanagemean<-apply(levelmanage,MARGIN=2,FUN=mean)
  levelmanageci<-apply(levelmanage,MARGIN=2,FUN=sd)*1.96/sqrt(1001)

return(plotCI(x=1:as.numeric(times[2]),y=levelmanagemean,uiw=levelmanageci,pch=20,gap=0,main=plo
ttitle,xlab="Year",ylab="Number of Buckthorns"))
}
#If you would like to output future graphs to a jpeg use: jpeg("filepath/Rplot%03d.jpeg")
#jpeg("CAmeangraphs/Rplot%03d.jpeg")

#Iterate simulation for values of manage ranging from 0 to 1 by 0.1 with management frequency set to
every 4 years
CA2plots=NULL
for (i in seq(0,1,0.1)) {
  CA2ploti<-CAmeangraph(CA2,list(from=1,to=50,by=1),
                    list(cf=cf, intercept=intercept,hsc=sphsc,x=x2,manage=i,yman=4,year=1),
                    initial2,paste(i*100,"% Removal Every 5 Years",sep=""))
  CA2plots<-c(CA2plots,CA2ploti)
}
#plot the number of cells occupied over 50 years at different levels of manage (ie removal success)
par(mar=c(5.1, 4.1, 4.1, 5.1))
plot(seq(0,50,length.out=50),seq(0,1200,length.out=50),type="n",xlab="Year",ylab="Number of Cells
Occupied")
cols<-rainbow(11)
for (i in seq(2,22,2)){
```

```
  lines(CA2plots[[1]],CA2plots[[i]],col=cols[i/2])
}
legend("topright",inset=c(-0.3,0.05),legend=c(paste(seq(0,1,0.1),sep=",")),
      lty=1,col=cols,cex=0.74,bty="n",y.intersp=0.7,title="Removal\nSuccess",xpd=TRUE)
##plot the mean over 4 years to smooth graph##
e<-NULL
for (i in seq(2,22,2)){
  d<-colMeans(matrix(CA2plots[[i]][1:48],nrow=4))
  e<-c(e,max(d))
  lines(seq(4,48,4),d,col=cols[i/2])
}
#Maximum mean number of occupied cells for each level of removal success
btnlr<-e

#Iterate simulation for values of yman (ie management frequency) ranging from 1 to 1 by 10 with
#removal success set to 0.8
CA2plotsy<-NULL
for (i in 1:10) {
  CA2ploti<-CAmeangraph(CA2,list(from=1,to=50,by=1),
                      list(cf=cf, intercept=intercept,hsc=sphsc,x=x2,manage=0.8,yman=i,year=1),
                      initial2,paste(i*100,"% Removal Every 5 Years",sep=""))
  CA2plotsy<-c(CA2plotsy,CA2ploti)
}
#plot the number of cells occupied over 50 years at different levels of yman (ie management frequency)
par(mar=c(5.1, 4.1, 4.1, 5.1))
plot(seq(0,50,length.out=50),seq(0,1200,length.out=50),type="n",xlab="Year",ylab="Number of Cells
Occupied")
cols<-rainbow(10)
for (i in seq(2,20,2)){
  lines(CA2plotsy[[1]],CA2plotsy[[i]],col=cols[i/2])
}
legend("topright",inset=c(-0.4,0.05),legend=c(paste(seq(1,10,1),"Years",sep=" ")),
      lty=1,col=cols,cex=0.75,bty="n",y.intersp=0.7,title="Management\nFrequency",xpd=TRUE)
##plot the mean of number of years between management##
e<-NULL
for (i in seq(2,20,2)){
  b<-seq(from=i/2,to=((50%/%(i/2))*(i/2)),by=i/2)
  c<-CA2plotsy[[i]][1:((50%/%(i/2))*(i/2))]
  d<-colMeans(matrix(c,nrow=(i/2)))
  e<-c(e,max(d))
  lines(b,d,col=cols[i/2])
}
#Maximum mean number of occupied cells for each level of management frequency
btnf<-e

#Not Used#
# CA2plotsya<-NULL
```

```
# for (i in seq(1,50,5)) {
#  CA2ploti<-CAmeangraph(CA2,list(from=1,to=50,by=1),
#                     list(cf=cf, intercept=intercept,hsc=sphsc,x=x2,manage=0.8,yman=5,year=i),
#                     initial2,paste(i*100,"% Removal Every 5 Years",sep=""))
#  CA2plotsya<-c(CA2plotsya,CA2ploti)
# }
# par(mar=c(5.1, 4.1, 4.1, 5.1))
# plot(seq(0,50,length.out=50),seq(0,1200,length.out=50),type="n",xlab="Year",ylab="Number of Cells
Occupied")
# cols<-rainbow(10)
# for (i in seq(2,16,2)){
#   lines(CA2plotsya[[1]],CA2plotsya[[i]],col=cols[i/2])
# }
# legend("topright",inset=c(-0.5,0.05),legend=c(2014-seq(1,50,5)),
#       lty=1,col=cols,cex=0.75,bty="n",y.intersp=0.7,title="Date",xpd=TRUE)

#Iterate simulation for values of f (ie number of subsections) ranging from 1 to 10 by 1 with
#removal success set to 0.8
CA2plotsinc<-NULL
for (i in seq(2,10,1)) {
  CA2ploti<-CAmeangraph(CA4,list(from=1,to=50,by=1),
                      list(cf=cf, intercept=intercept,hsc=sphsc,x=x2,f=i,year=1,manage=0.8),
                      initial2,paste(i*100,"% Removal Every 5 Years",sep=""))
  CA2plotsinc<-c(CA2plotsinc,CA2ploti)
}
#plot the number of cells occupied over 50 years at different levels of f (ie number of subsections)
par(mar=c(5.1, 4.1, 4.1, 5.1))
plot(seq(0,50,length.out=50),seq(0,1000,length.out=50),type="n",xlab="Year",ylab="Number of Cells
Occupied")
cols<-rainbow(10)
for (i in seq(2,18,2)){
  lines(CA2plotsinc[[1]],CA2plotsinc[[i]],col=cols[i/2])
}
legend("topright",inset=c(-0.39,0.05),legend=c(seq(2,10,1)),
      lty=1,col=cols,cex=0.75,bty="n",y.intersp=0.7,title="Number\nof Subsections",xpd=TRUE)

e<-NULL
for (i in seq(2,20,2)){
  b<-seq(from=i/2,to=((50%/%(i/2))*(i/2)),by=i/2)
  c<-CA2plotsinc[[i]][1:((50%/%(i/2))*(i/2))]
  d<-colMeans(matrix(c,nrow=(i/2)))
  e<-c(e,max(d))
  lines(b,d,col=cols[i/2])
}
#Maximum mean number of occupied cells for each level of management frequency
btnd<-e
```

```
##Effort = amount of cells surveyed*0.5+cells managed*1
## =1133 * times managed*0.5 + sum(CAdiff)

#function to extract results of simulations
CAoutmean<-function(model,times,parms,initial,plottitle){
  parms(model)<-parms
  init(model)<-initial
  times(model)<-times
  model<-sim(model)
  sCA2o<-data.frame(lapply(model@out,sum))
  colnames(sCA2o)<-c(paste(1:as.numeric(times[2]),sep=","))
  Simout<-sCA2o
  for (i in 1:1000) {
    model<-sim(model)
    sCA2o<-data.frame(lapply(model@out,sum))
    colnames(sCA2o)<-c(paste(1:as.numeric(times[2]),sep=","))
    colnames(Simout)<-c(paste(1:as.numeric(times[2]),sep=","))
    Simout<-data.frame(rbind(Simout,sCA2o))
  }
  Simoutci<-apply(Simout,MARGIN=2,FUN=sd)*1.96/sqrt(1001)
  return(Simoutmean<-apply(Simout,MARGIN=2,FUN=mean))
}

##Calculate effort while Varying Management Frequency, removal success=0.8##
CAsums=NULL
CAeffsf<-NULL
for (i in seq(1,10,1)) {
  CAsums<-CAoutmean(CA2,list(from=1,to=50,by=1),
                    list(cf=cf, intercept=intercept,hsc=sphsc,x=x2,manage=0.8,yman=i,year=1),
                    initial2,paste(i*100,"% Removal Every 5 Years",sep=""))
  CAsums<-c(0,CAsums)
  CAsums5<-as.numeric(CAsums[seq(0, length(CAsums), i)])
  CAsums6<-as.numeric(CAsums[seq(i+1, length(CAsums), i)])
  if (length(CAsums5)>length(CAsums6)) CAsums5<-CAsums5[-length(CAsums5)]
  CAdiff<-CAsums5-CAsums6
  CAsumdiff<-sum(CAdiff)
  CAeff<-1133*0.5*(50/i)+CAsumdiff
  CAeffsf<-c(CAeffsf,CAeff)
}

plot(1:10,CAeffsf,xlab="Management Frequency",ylab="Effort",type="l")

## Calculate effort while Varying removal success, yman=4##
CAsums=NULL
CAeffslr<-NULL
for (i in seq(0.1,1,0.1)) {
```

```
  CAsums<-CAoutmean(CA2,list(from=1,to=50,by=1),
                    list(cf=cf, intercept=intercept,hsc=sphsc,x=x2,manage=i,yman=4,year=1),
                    initial2,paste(i*100,"% Removal Every 5 Years",sep=""))
  CAsums<-c(0,CAsums)
  CAsums5<-as.numeric(CAsums[seq(0, length(CAsums), 4)])
  CAsums6<-as.numeric(CAsums[seq(5, length(CAsums), 4)])
  if (length(CAsums5)>length(CAsums6)) CAsums5<-CAsums5[-length(CAsums5)]
  CAdiff<-CAsums5-CAsums6
  CAsumdiff<-sum(CAdiff)
  CAeff<-1133*0.5*(50/4)+CAsumdiff
  CAeffslr<-c(CAeffslr,CAeff)
}

plot(seq(0.1,1,0.1),CAeffslr,xlab="Removal Success",ylab="Effort",type="l")

## Calculate effort while Varying size of divisions, removal success=0.8##
CAsums=NULL
CAeffsd<-NULL
for (i in seq(2,10,1)) {
  CAsums<-CAoutmean(CA4,list(from=1,to=50,by=1),
                    list(cf=cf, intercept=intercept,hsc=sphsc,x=x2,f=i,year=1,manage=0.8),
                    initial2,paste(i*100,"% Removal Every 5 Years",sep=""))
  CAsums<-c(0,CAsums)
  CAsums5<-as.numeric(CAsums[seq(1, length(CAsums), 1)])
  CAsums6<-as.numeric(CAsums[seq(2, length(CAsums), 1)])
  if (length(CAsums5)>length(CAsums6)) CAsums5<-CAsums5[-length(CAsums5)]
  CAdiff<-CAsums5-CAsums6
  CAsumdiff<-sum(CAdiff)
  CAeff<-1133*0.5*(50/4)+CAsumdiff
  CAeffsd<-c(CAeffsd,CAeff)
}

plot(seq(2,10,1),CAeffsd,xlab="Number of Subsections",ylab="Effort",type="l")
#Extra plots#
# plot(seq(1,10,1),seq(5000,25000,length.out=10),type="n",xlab="Management
Level",ylab="Management Effort")
# lines(seq(2,10,1),CAeffsd,col="blue")
# lines(seq(1,10,1),CAeffslr,col="red")
# lines(1:10,CAeffsf,col="green")
# legend("topright",inset=c(-0.56,0.05),legend=c("Number of\nDivisions","Likelihood
of\nRemoval","Frequency of\nManagement"),
#
lty=1,col=c("blue","red","green"),cex=0.72,bty="n",y.intersp=1.5,title="Management\nVariable",xpd=TR
UE,seg.len=1)
# twoord.plot(lx=0:10,ly=btnlr,rx=1:10,ry=CAeffslr)
# twoord.plot(lx=1:10,ly=btnf,rx=1:10,ry=CAeffsf)
# twoord.plot(lx=2:10,ly=btnd,rx=2:10,ry=CAeffsd)
```

```
#Plot summarizing the effort and number of occupied cells for different management scenarios
btnf2<-c(0,btnf)
btnd2<-c(0,btnd)
sumbtn<-data.frame(btnlr[-1],btnf,btnd2)
CAeffsd2<-c(0,CAeffsd)
sumeff<-data.frame(CAeffslr,CAeffsf,CAeffsd2)
plot(seq(6000,15000,length.out=1201),0:1200,type="n",xlab="Effort",ylab="Final Number of Occupied
Cells")
lines(sumeff[,1],sumbtn[,1])
points(sumeff[,1],sumbtn[,1])
text(sumeff[,1],sumbtn[,1], seq(0.1,1,0.1), cex=0.6, pos=3,font=2)
lines(sumeff[,2],sumbtn[,2],lty=2)
points(sumeff[,2],sumbtn[,2],pch=2)
text(sumeff[,2],sumbtn[,2], 1:10, cex=0.6, pos=4,font=2)
lines(sumeff[2:10,3],sumbtn[2:10,3],lty=3)
points(sumeff[2:10,3],sumbtn[2:10,3],pch=3)
text(sumeff[2:10,3],sumbtn[2:10,3], 2:10, cex=0.6, pos=4,font=2)
```