USING PORTFOLIO THEORY TO EVALUATE THE REGIONAL DEPLOYMENT OF TRANSFERRING TREE SEEDS UNDER UNCERTAIN FUTURE CLIMATES

Jia Qi

A Graduate Thesis

Submitted in partial fulfillment of the requirements

For the degree of

Master of Science in Forestry

Faculty of Natural Resources Management

Lakehead University

Sept, 2010

Lakehead
UNIVERSITY

OFFICE OF GRADUATE STUDIES

NAME OF STUDENT:

DEGREE AWARDED:

ACADEMIC UNIT:

TITLE OF THESIS:

This thesis has been prepared

under my supervision

and the candidate has complied

with the Master's regulations.

Signature of Supervisor

Date

ABSTRACT

Qi, J. 2010. Using portfolio theory to evaluate the regional deployment of transferring tree seeds under uncertain future climates. Master of Science in Forestry, Lakehead University. Advisor, Dr. Kevin Crowe.

Key words: black spruce, climate change, adaptive variation, tree improvement, reforestation, seed transfer, portfolio optimization, focal point seed zones.

In this study, the problem of deploying seed from multiple tree improvement orchards to multiple sites in an environment of uncertain climate change is addressed. A modeling approach is designed and applied with the objective providing a robust solution, i.e., transferred seed sources perform well across multiple climatic scenarios. The approach involves two steps. First, the focal point seed zone method is employed to predict seed deployment zones under multiple future climate scenarios. Next, a portfolio model is applied to minimize the risk of maladaptation of the transferred seed under multiple climatic scenarios.

The method was applied using black spruce (*Picea mariana*) field data from 7 sites using 24 seed sources from the Great Lakes area. The focal point seed zone method generated deployment zones for 24 seed sources over three 30-years periods under 12 predicted future climate scenarios. Next, the optimization procedure searched for eligible sites that can receive improved seed sources from 7 provenances considering 12 different climatic scenarios. The portfolio model also produced the optimal composition of candidate seed sources at each eligible site. Sensitivity of the solutions to different emission scenarios is compared. Finally, geographic representations of results were illustrated in Geographic Information System. It was concluded that this modeling framework provides a useful approach for decision-makers to address the problem of deploying seed at regional scale, such that the risk of climatic maladaptation is minimized.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

This study was funded by a National Sciences and Engineering Research Council (NSERC). I would like to thank Dan McKenney of the Great Lake-St. Lawrence Forestry Centre for providing climate data.

I would like to offer my sincerest thanks to my committee members. Dr. William Parker has provided critical help with seed zone modeling. Mr. Paul Charrette has provided a very helpful review of my thesis. Thank you to my external examiner Dr. Greg O'Neill for a thoughtful review of my thesis. Special thanks to Jane Parker for help me with editing my manuscript.

My deepest thanks to my advisor Dr. Kevin Crowe. Your advice, support and encouragement have motivated and inspired me. Your guidance has provided me with opportunity to overcome difficulties in this study.

Finally, I would like to thank my family and friends for their ongoing concern, support and encouragement.

1. Introduction

Trees are vital economic resources for the forest industry and also important ecological components of biodiversity. One approach to satisfying the increasing global demand for forest products and the requirement for ecological conservation is to improve the yield per hectare through tree improvement (Daniels 1984; Li et al. 1999; Pijut et al. 2007). In the last fifty years, many studies on tree improvement (e.g., tree breeding, genetic modification, and molecular level DNA modification) have shown that trees can be selected bred successfully for desirable properties. Selected trails include: 1) enhanced productivity, 2) improved quality, and 3) improved resistance to disease and pests (Pijut et al. 2007). Because of the promising economic and social benefits of a tree improvement program, many countries have made significant investments into tree improvement programs; e.g., Japan (Satoo 1960; McKeand & Kurinobu 1998), India (Chandha & Patnik 1990), Canada (Weisgerber & Sindelar 1992), and China (Shen et al. 2007; Su et al. 2003).

While improved trees have demonstrated superior growth potential and enhanced productivity, these properties are expected to occur exclusively within clearly defined climatic envelopes, from which such trees have been selected and for which they have been bred. Hence, the prospect of dramatic climatic change has forced a significant problem upon those with a sunk cost in tree improvement programs; namely, where should these improved seeds be deployed, at minimal risk, in an environment of

uncertain climatic change? This, in short, is the problem to be addressed in this thesis. It is a problem of:

(a) *assisted migration* which is to occur in

(b) an environment of *uncertain climatic change*; such that

(c) the *risk of maladaptation is minimized.*

Each element of the problem will now be described.

a) It is a problem of assisted migration

Theoretically, species may be expected to respond to climate change in one of three-ways: adaption, niche-tracking or extinction (Peterson et al. 2005). Since changes in climate are expected to be rapid, neither adaption nor natural migration may be sufficient to retain high productivity of genetically improved trees if locally adapted populations continue to be used (Davis & Shaw 2001; Davis et al. 2005). Hence, to avoid potential reductions in productivity as a result of climate change, artificial assisted migration (i.e. seed transfer) may be required to ensure adequate growth and adaptation of improved trees in future scenarios under climate change (Rehfeldt et al. 1999).

Seed transfer studies date from the beginning of the 20th century. Currently, seed transfer models are extensively employed to predict the suitable areas in which seed sources may be deployed based on the analyses of provenance trials (Campbell 1986; Campbell & Sugano 1987). To establish these transfer models, seeds from multiple provenances are planted and observed in provenance tests, and the relationship between observed growth variables and climatic variables of the test sites is determined to

quantify the climatic tolerance of each provenance (Parker & vanNiejenhuis 1996; Lesser & Parker 2006).

In these seed transfer studies, the least significant difference (LSD) values can be applied to quantify the minimum phenotypic difference required to distinguish among populations (Rehfeldt 1982). Parker & vanNiejenhuis (1996) used the LSD to quantify the adaptive variation for white spruce in the focal point seed zone method. The LSD can be also used to guide the maximum distances that seed should be moved without incurring unacceptable levels of maladaptation (Bower & Aitken, 2008).

To apply these models to seed transfer, a dataset that represents the appropriate future climate condition is indispensable. But there are significant uncertainties accompanying estimates of future climates. The cause of this uncertainty is inherent within the approaches used to estimate future climates.

b) In an environment of uncertain climatic change

Models used to predict climatic changes are computer simulations of the circulation of heat in the atmosphere and oceans; hence they are referred to as "coupled" atmosphere-ocean general circulation models (AOGCMs). Several uncertainties should be borne in mind when interpreting the output of these models. The first uncertainty comes from the selection of the particular AOGCM—for there are several, and they differ in their outputs. The Intergovernmental Panel on Climate Change (IPCC) has evaluated approximately 20 AOGCMs by comparing observed and reproduced past climate changes. Though the IPCC concluded that these models are able, in general, to provide credible simulation of climate (at least in large scales), the report also concludes

that no single model can be considered best; they all have errors in different respects (Randall et al. 2007; Meehl et al. 2007).

The second source of uncertainty arises from the anthropogenic cause of climatic change (Parry et al. 2007). The effect of anthropogenic factors (e.g., the quantity of green house gas emitted) affecting climate in the future is difficult to predict, because the amount of greenhouse gases emitted depends on the policies of countries and the development of regional economies. Such uncertainties are reflected in the discrepancies in Special Report on Emission Scenarios (SRES), which is a set of plausible representations of the future development of emissions of substances that are based on a coherent and internally consistent set of assumptions about driving forces and their key relationships (Nakicenovic & Swart 2000; IPCC 2007b).

The third source of uncertainty arises from the spatial scale at which predicted climatic changes are estimated. Although current AOGCMs can provide credible predictions of future climate change on a large scale, details of regional conditions may involve too many factors for computation. Hence, the accuracy of those models will decrease when the desired spatial resolution increases (Raper & Giorgi 2005; Randall et al. 2007).

c) Minimizing the risk of maladaptation

Handling the "deep uncertainties" in climate change prediction is the key to solving the problems of risks in seed transfer planning. Lempert et al. (2004) generalized and compared features of two different approached to address climate-change uncertainties. One method, called predict-then-act, characterizes uncertainties using

probability distributions. These distributions represent the likelihood of alternative future scenarios. Having quantified such distributions, one then selects the scenario with the highest probability of a desirable response. In contrast, Lempert's second approach, referred to as "assess-risk-of-policy", uses multiple feasible policy options to generate a combined, robust strategy, which is insensitive to the uncertainties (Lempert et al. 2004); i.e., a strategy is selected that will perform well across all feasible future scenarios. Lempert et al. (2004) suggest that the robust strategy is best for planned adaptation to climatic change, since it is not feasible to assign probabilities to any of the future climate scenarios predicted by the AOGCM's. Hence, our problem will be one of formulating a decision support model that provides a robust solution for the deployment of seed under multiple climatic futures.

Objective of Research

The objective of this research is to formulate and evaluate a decision support model that can be used to minimize the risk of transferring seed from multiple sources (e.g., improved tree seeds) to multiple destinations. In effect, to formulate a model that extends the approach designed by Crowe and Parker (2008) who solved the problem of selecting multiple seed sources for regeneration at one site.

The significance of this innovation is that the model can be applied at the regional scale to evaluate:

i. how well a region can be covered, at minimal risk, by seed from a particular set of improved seed sources; and

ii. which set of seed sources will be most commonly used in covering a region at minimal risk.

In addition, the sensitivity of the solutions to assumptions on future CO_2 emissions and parameters constraining the acceptable limits of maladaptation can be explored.

This modeling approach will be evaluated by applying it to a case study; the problem of finding a robust solution to deploying black spruce (*Picea mariana*) , from seven hypothetical improved seed sources, to the entire province of Ontario, under 12 different climatic scenarios.

2. Literature review

This research requires reviewing literature in the following fields:

a) principles of predicting climatic change and its biological impacts;

b) the scientific methods used to support assisted migration; and,

c) decision modeling for planned adaptation to climatic change.

2.1 Principles for predicting climatic change and its biological impacts.

In reviewing literature on this topic, I will address the following questions:

Upon what evidence do we conclude that climate is changing?

How are projections of future climates formed?

How do we estimate the impact of climatic change on vegetation?

2.1.1 Evidence for Climatic Change

There is strong evidence of significant change of the climate on the earth, both globally and regionally, and ranging in duration from years to millennia (Karl & Trenberth 2005). The evidence ranges from paleoclimatic proxy measurements (i.e., study result from ice sheets, tree rings and rocks) to modern instrumental records. The reliability of such estimates is typically coarser in remote past periods and higher after 1850, when the surface temperature began to be measured globally (Karl & Trenberth 2005; IPCC 2008). Digital technology is available for collecting and analysing millions

of individual meteorological records from the last decades of the twentieth century. In addition, the application of space-borne sensors has provided comprehensive climate change images over the last 30 years (Hulme 2005). Thus, a multidimensional view of climate change during the last 150 years is possible (Jones et al. 1999).

Reconstruction of climate change over the last century shows that earth's mean surface temperature has risen over the last 150 years (Jones et al. 1999). The annual mean surface air temperature of the earth is the most widely used indicator to describe the global climate (Hulme 2005). According to measurements in the reference period 1961-1990, 14° C has been widely acknowledged as the average measure to describe temperature state and variation (Jones et al. 1999). Over the last 100 years (1906-2005), global average surface temperatures have risen by 0.74˚C ± 0.18˚C when estimated by a linear trend (Trenberth et al. 2007). Two 20-year periods, 1925-1944 and 1978-1997, are the greatest warming periods in the last century (Jones et al. 1999).

Changes in the heat balance of the Earth have also influenced the hydrological cycle. Evaporation from the ocean has intensified, leading to an overall increase in precipitation amounts (Hulme 2005). However, coupled with the atmospheric cycle, the redistribution of precipitation is not uniform, i.e., some regions become wetter while others become drier. The forth assessment from the IPCC has reported significantly increased precipitation in eastern parts of North and South America, northern Europe and northern and central Asia, while drying has been observed in the Sahara, the Mediterranean, southern Africa and parts of southern Asia (IPCC 2007b). Since the 1970s, precipitation has generally increased over the surface areas north of 30˚N, but has

a downward trend in the tropics, and droughts have become more common in tropical and subtropical regions (Trenberth et al. 2007).

Since systematic observation and recording of global climate occurred after 1850, it is impossible to directly access the climate change in the remote past. Investigation of historic climate relies on proxies which include ocean and lake sediments, tree rings, corals and ice cores, etc. (Overpeck et al. 2005). Fossils, geochemistry of ancient sediment, and dendroclimatology are the main sources of these proxies. Paleoclimate studies rely on multiple sources of proxies so that the results can be cross-verified (IPCC 2007c).

The study of paleoclimate indicates that the warming of the last decades of the 20th century is exceptional in magnitude and speed. The rate of increase in GHGs was at least five times faster over the period of 1960 to 1999 than any other 40-year period during the past 2,000 years prior to the industrial era (Jansen et al. 2007). The comparison implies there may be climatic threats to vegetation in the next 100-200 years. Overpeck et al. (2005) summarised three types of threats: 1) natural abrupt climate change; 2) anthropogenic climate change (including sea level rise); and 3) probability of abrupt climate shifts triggered by anthropogenic change. Thus, it can be said that current global warming and future climate change is unprecedented.

2.1.2 How are projections of future climates formed?

The most common approach to predicting climatic change is through the use of mathematical simulation models. These models are based on the physical, chemical and

biological properties of the components of the climatic system, their interactions, and feedback processes.

General circulation models (GCM) are used for climate simulation. There is a hierarchy of construction for these models, and they vary in comprehensiveness and complexity (Raper & Giorgi 2005). These models range from global scale simulations with coarse resolution to regional high spatial resolution models (Raper & Giorgi 2005).The simple application can be an atmospheric general circulation model (AGCM) or an ocean general circulation model (OGCM).

Since climate changed under the interaction among many components, it is necessary to consider the effects of the atmosphere, the land, the ocean, and the cryosphere together. Thus, it is necessary to couple an AGCM to an OGCM for more exact simulation. The resulting models are called atmosphere ocean general circulation models (AOGCM), which are the most widely used models for future climate projection (Sun & Hansen 2003).

AOGCMs are capable of representing climatic change in three dimensions. Gridded AOGCMs divide the atmosphere and oceans into a number of cells horizontally and vertically. The spatial resolutions of AOGCMs are described as horizon resolution (e.g. 10 x 10 km^2/ grid) and vertical resolution (e.g. 20 levels). AOGCMs are usually global models; the sub-grid-scale process can provide finer resolution details when it operates on a small area (IPCC 2007a; Raper & Giorgi 2005). Another approach to explicitly simulate regional climate evolution is to apply a regional climate model (RCM) instead of an AOGCM. RCM is similar to AOGCM, but it focuses on only a

small portion of the AOGCM and provides finer resolution and more precise climate simulation than an AOGCM. Table 1 generalizes features of several common GCMs which were evaluated in the IPCC Fourth Assessment Report (IPCC 2007a).

Table 1. Selected models and specs from the Fourth Assessment Report of IPCC

Model ID,Vintage	Sponsor(s)	Atomosphere Top Resolution References	Ocean Resolution Z Coord., Top BC References	Coupling Flux Adjustments References	Land Soil,Plants, Routing References
UKMO-HadGEM1,2004	Hadley Centre for Climate Prediction and Research/Met Office, UK	Top=5hPa 2.5˚×3.75˚ L19 Pope et at.,2000	1.25˚×1.25˚ L20 Depth, rigid lid Gordon et al., 2000	No adjustments Gordon et al., 2000	Layer, canopy, routing Cox et al., 1999
CGCM3.1(T47),2005	Canadian Centre for Climate Modeling and Analysis, Canada	top = 1 hPa T47 (~2.8˚×2.8˚)L31 McFarlane et al., 1992; Flato, 2005	1.9˚×1.9˚ L29 Depth, rigid lid Pacanowski et al., 1993	heat, fresh water Flato, 2005	layers, canopy, routing Verseghy et al., 1993
CGCM3.1(T63),2005		top = 1 hPa T63 (~1.9˚×1.9˚)L31 McFarlane et al., 1992; Flato, 2005	1.9˚×1.4˚ L29 Depth, rigid lid Flatoand Boer, 2001; Kim et al., 2001	heat, fresh water Flato, 2005	layers, canopy, routing Verseghy et al., 1993
CSIRO-MK3,2004	Commonwealth Scientific and Industrial Research Organisation (CSIRO) Atmospheric Research, Australia	top = 4.5 hPa T63 (~1.9˚×1.9˚)L18 Gordon et al., 2002	0.8˚×1.9˚ L31 Depth, rigid lid Gordon et al., 2002	no adjustments Gordon et al., 2002	layers, canopy Gordon et al., 2002
MRI-CGCM2.3.2,2003	Meteorological Research Institute, Japan	top = 0.4 hPa T42(~2.8°×2.8°)L30 Shibata et al., 1999	0.5°–2.0° × 2.5° L23 depth, rigid lid Yukimoto et al., 2001	heat, freshwater, momentum (12°S–12°N) Yukimoto et al., 2001; Yukimoto and Noda, 2003	layers, canopy, routing Sellers et al., 1986; Sato et al., 1989
GFDL-CM2.0,2005	U.S. Department of Commerce/National Oceanic and Atmospheric Administration (NOAA)/Geophysical Fluid Dynamics Laboratory (GFDL), USA	top = 3 hPa 2.0° ×2.5° L24 GFDL GAMDT, 2004	0.3°–1.0° × 1.0° depth, free surface Gnanadesikan et al., 2004	no adjustments Delworth et al., 2006	bucket, canopy, routing Milly and Shmakin, 2002; GFDL GAMDT, 2004
GFDL-CM2.1,2005		top = 3 hPa 2.0° ×2.5° L24 GFDL GAMDT, 2004 with semi-Lagrangian transports	0.3°–1.0° × 1.0° depth, free surface Gnanadesikan et al., 2004	no adjustments Delworth et al., 2006	bucket, canopy, routing Milly and Shmakin, 2002; GFDL GAMDT, 2004

A study of 15 AOGCMs shows that these models have different advantages, e.g. some have a high accuracy in temperature prediction, while some may be good at precipitation simulation (Covey et al. 2003; Lambert & Boer 2001). The IPCC (2007a)

also evaluated AOGCMs' prediction against observed and reproduced past climate, and compared prediction among models. Although, it was concluded that in general the models provide credible simulation of climate (at least in large spatial-temporal scales), no single model could be the best, i.e., they all have errors in different respects (Randall et al. 2007; Meehl et al. 2007).

The Special Report on Emissions Scenarios (SRES) presents emission scenarios developed by Nakicenovic and Swat (2000) and is used as the basis for climate projections (Ehhalt et al. 2001). Forty different emission scenarios are characterised by distinctly different levels of population, and economic and technological development. These scenarios are derived from four entirely different storylines (A1, A2, B1 and B2) that describe how the world might develop.

The A1 storyline assumes a world of very rapid economic growth, global population that peaks in mid-century and declines thereafter, and the rapid introduction of new and more efficient technologies. In contrast, the A2 storyline describes a very heterogeneous world with a continuously increasing population and fragmented, slowly advancing economics and technologies.

The B1 storyline describes a convergent world with the same global population that peaks in mid-century and declines thereafter (as in the A1 storyline) but with rapid change in economic structures toward a service and information economy. The B2 storyline describes a world in which the emphasis is on local solutions to economic, social and environmental sustainability. It is a world with continuously increasing global population (at a rate lower than in A2) and with intermediate levels of economic

development, and less rapid and more diverse technological change than in the B1 and A1 storylines (IPCC 2007b).

2.2 How do we estimate the impact of climate change on vegetation?

A basic concept used in predicting the impact of climate change on vegetation is *ecological niche*. Ecological niche is defined as a geographical range and habitat that a species or population can or does occupy (Kimmins 2004). This term was first defined as a species' geographic distribution determined by a suite of environmental factors (Grinnell 1917; Grinnell 1924; Peterson et al. 2005). Hutchinson (1957) redefined ecological niche as comprising those environmental conditions within which a species can survive and grow (Hutchinson 1957).

A bioclimatic envelope constitutes the climatic component of the fundamental ecological niche (or 'climatic niche') (Pearson & Dawson 2003). Based on the foundation of ecological niche theory, one can draw an outline of a species' or population's distributional range by modeling its bioclimatic niche or "climate envelope." It should be noted, however, that the niche-based bioclimatic model considers only climatic variables, and no other environmental factors that may have relationship with the distribution of a species (Pearson & Dawson 2003). Given the future climate scenarios, niche-based bioclimatic models are able to project the potential distribution range of a particular species or population.

Two terms, fundamental niche and realized niche, are used to distinguish whether a distribution of a species was established based on the limitation of biotic or abiotic factors (Hutchinson 1957). This distinction is important in the context of

bioclimatic modeling, particularly with regard to the methodologies used to characterize bioclimatic envelopes (Pearson & Dawson 2003). Bioclimatic models based on the empirical relationship between observed species distributions and environmental variables are realized niches. Such models correlate climatic variables against observed distribution, on the assumption that the best indicator of a species' climate requirements can be based on its current distribution (Huntley et al. 1995; Peterson et al. 2002; Peterson et al. 1999). Some other bioclimatic models describe the fundamental niche by interpreting physiological limitation mechanisms under species' climate requirements. (Sykes et al. 1996).

Early climatic envelope modeling used simple correlations between climate variables and observed distribution, such as the works of Johnston (cited in Pearson & Dawson 2003). Huntley et al. (1995) used a weighted regression to model the relationship between eight species' phytogeographic patterns and present climate variables, and a major shift was projected under a double CO_2 concentration scenario (Huntley et al. 1995). Other advanced models apply algorithms to determine the relationship between species distribution and environmental factors. For example, Peterson et al. (2002) employed artificial neural networks (ANN) to characterize bioclimatic envelops based on observed species distributions and five environmental variables. Applying genetic algorithms and museum specimen occurrence data, Peterson et al. (2002) developed ecological niche models for 1870 species occurring in Mexico and projected them onto two climate surfaces modeled for 2055 (Peterson et al. 2002).

A concise summary of the different approaches used in designing climate envelope models is presented by Peterson et al. (2005):

1) a model may apply point or gridded data;
2) the selection of independent variables can be extensive or constrained, (based on assumptions regarding which environmental dimensions will be most relevant to limiting species' or population's distributions)
3) niche modeling approaches may
    i. fit a model using a predetermined statistical approach,
    ii. relax assumptions regarding the form of the relationship, or
    iii. may avoid any assumptions whatsoever about the form of the relationship involved.

The reliability of these climate envelope models has been questioned. Pearson & Dawson (2003) listed the following three criticisms of bioclimatic modeling: 1) it may ignore biotic interactions which impact species distributions; 2) it may be defective in considering the altering of evolutionary change on species distribution; 3) it may not account for species dispersal, but instead may focus on predicting the potential range of organisms under climate change (Pearson & Dawson 2003).

In conclusion, many examples have indicated that bioclimate envelop models are capable of providing perhaps the best available guide for policy making at the current time. They have been usefully employed to identify possible magnitude of future changes to distributions and to suggest which species, habitats and populations are most at risk from climate change (Pearson & Dawson 2003).

Bioclimatic envelope models have been used to predict the impact of climatic change on the distributions of tree species. Iverson & Prasad (1998) projected 80 tree

species' responses under climate change and predicted that nearly half of them have potential of a northern shift of at least 100km. Iverson *et al.* (2007) also mapped 134 tree species from the eastern United States for potential response under six climate projections and predicted that approximately 66 species would expand their distributions while 54 species would lose at least 10% of their suitable habitat.

Shafer et al. (2001) conducted a similar study to project potential distributions of western North American tree and shrub taxa under future climate scenarios. Three GCMs (HADCM2, CGCM1, and CSIRO) were applied to project species distribution changes. In addition, GHGs and aerosol variation were considered in these simulation models. The results indicate: 1) that potential range shifts could be large for many tree and shrub taxa. 2) that shifting trends are not limited to northward distributional expansion, but also include southward shifts of the existing ranges of a few species; and, 3) that fragmented distributions of some species may intensify, while the simulated potential distributions of other species may expand (Shafer et al. 2001).

Using a climatic envelope model, McKenney et al. (2007) predicted distributional shifts for 130 North American tree species under 6 future climate scenarios (3 GCMs coupled with 2 emission scenarios). The study contrasted the predictions of two future scenarios: a full-dispersal scenario and a no-dispersal scenario. The former scenario involved a decreased distribution of 12% while the latter involved a 58% decrease. In addition, northward shifts of 700 km and 330 km respectively were predicted averaged across all species (Mckenney et al. 2007).

Climate envelope models were also used to predict changes of tree-line in the Swedish Scandes (Kullman 2001; Kullman 2003; Kullman 1996). It was estimated that a a distributional shift of 100-165 m upslope would occur during the 21st century as a result of global warming.

On a population scale, intraspecific populations generally respond to climatic change in a manner similar to species. However, rapid climate change and fragmented landscapes have overwhelmed the ability of populations to adapt or migrate for many plant populations (Jump & Penuelas 2005). Neither adaptation nor migration may be sufficient for tracking the niche which they currently dominate (Davis & Shaw 2001). Species with small populations, fragmented ranges, or low fecundity may face extirpation (Aitken et al. 2008).

2.2.1 Scientific Methods of Assisted Migration

The assisted migration of seeds must occur within a defined seed zone. A Seed zone is defined as a geographic area within which genotype × environment (G×E) interaction is minimised (Campbell & Sorensen 1978). The G×E interaction refers to the phenotypic effect of interactions between genes and the environment. Within a seed zone, the G×E interactions show no differentiation to the seed source of a specific species (Rehfeldt 1983a); thus, seed zone is also defined as the spatial region where seed can be collected or deployed.

The region used for collecting seeds is referred to as a seed procurement zone, and the region used for planting seeds is referred to as a seed deployment zone. A number of provenance tests have shown that a seed grows best in its local

provenance (Parker 1992). Hence, reforestation efforts typically deploy seed within its local seed zone to minimize maladaptation. This method, however, limits the suitable range of reforestation for a specific seed sources; e.g., improved seeds (Parker 1992).

Seed zones for seed transfer can be delineated by models based on patterns of genetic variation related to the environment of the provenance. Campbell (1978) developed the first seed zone model by regressing 6 principal components, which were derived from 16 phenological and morphological variables in common-garden trials for 115 sources in Oregon, against each provenance's geographic variables. This was the first attempt to use genetic variation modeling. Environmental factors such as temperature and moisture were not considered in this model.

Rehfeldt (1982) applied a similar procedure to model population differentiation of western larch (*Larix occidentalis*) based on growth potential, phenology, and cold-hardiness variables from 2-year seedling trials which were regressed against the geographic location and ecological characteristics of the provenance's climate. This study demonstrated elevation to be the strongest predictor of seed source performance (Rehfeldt 1982).

The work of Campbell and Rehfeldt defined the strategy of seed transfer models; i.e., regression-based empirical models, in which the information of tree growth is generalized and regressed against the environmental variables. This regression model is then used to represent adaptive variation. In later studies, least significant difference (LSD) was introduced to quantify the variation difference of the target-area from the provenance (Rehfeldt 1982). Rehfeldt (1983, 1990) also suggested using continuous

seed zones instead of discrete ones. Seed zones were developed for Douglas-fir (*Pseudotsuga menziesii*) in central Idaho and western Montana by regressing adaptive differences against ecological and geographic variables (Rehfeldt 1983b; Rehfeldt 1983a).

Since the 1980's, advanced computer performance and geographic information systems (GIS) enabled the application of models to large-scale spatial calculations. Based on the previous seed transfer models, Parker (1992) designed a site-specific seed zone delineation procedure, named the "focal point seed zone" method. This approach treats an individual site (to be reforested) as a focal point, and generates unique seed procurement zones for any seed source (Parker 1992).

Five principal steps have been generalized in the focal point seed zone method (Parker 1992): 1) determine the area of interest and intensively sample provenances from the area; 2) use short-term common garden trials to assess growth potential and phenological characteristics; 3) perform principal component analysis (PCA) to summarize the observed variables into a few axes which are capable of representing major genetic variation; 4) use GIS to create a three-dimensional trend-surface to express the variation of principal component axes; and, 5) delineate individual focal point seed zones by constructing a contour map for each principal component. This approach was initially tested in delineating seed zones for jack pine (*Pinus banksiana* Lamb.) on a site in northern Ontario. Results of delineated zones demonstrate that the focal point seed zone method can help in transferring seed sources from their regional seed zones with minimal maladaptation (Parker 1992).

Parker and Niejenhuis (1994) also applied the focal point seed zone method to an adaptive variation study of black spruce (*Picea mariana)* provenances in north-western Ontario. Twenty-five growth and phenological variables were derived from a two-season common greenhouse trial. An analysis of variance (ANOVA) procedure identified 18 variables, which showed significant differences among provenances. Results show that the environmental variables were good predictors for describing black spruce needle flushing date and growth potential of height (Parker et al. 1994).

Parker and Niejenhuis (1996) produced a series of focal point seed zone maps for black spruce from northwestern Ontario using a version of the original modified focal point seed zone method proposed in 1992. Two improvements were made in this new study:

1) Principal component analysis (PCA) was introduced to summarize provenance growth and phenological variables. Thus, the number of dependent variables was minimised, while variation of biological information was maximised.
2) Standard deviation from the focal point for each PCA axis was calculated to quantify adaptive variation difference.

With these improvements, it became practical to retrieve seed zones from contour maps of PCA axes (Parker & vanNiejenhuis 1996). The application of canonical correlation analysis as an alternative to regression based focal point seed zones was presented by Lesser and Parker (2006). White spruce seed zones derived from canonical correlation were compared with regression-based seed zones. The comparison showed

that both models can describe adaptive variation by the environmental variables. However the canonical correlation produced finer resolution in some areas, presumably because of the statistical efficiency of the algorithm (Lesser & Parker 2006).

In response to rapid climate change, some studies have used the focal point seed zone method to predict future adaptive seed zones (Thomson & Parker 2008; Thomson et al. 2009). The latest trend is combining these models with other mathematical models (such as optimization models) to constitute decision support systems (Crowe & Parker 2005; Crowe & Parker 2008).

Aitken et al. (2008) explained constraints on migration and adaptation for modeling adaptive zones on population scale. Populations may face extirpation if they fail to adapt under climate change, because they are unlikely to migrate quickly enough under rapid climate change. Population response curves, therefore, should be used to predict the maximum extent to which seed can be moved under climate change (Aitken et al. 2008). Such population response curves, which are similar to the focal point seed zone models, relate population survival and growth of planted seedlings to geographic or climatic distances between provenances and common garden locations.

Thompson and Parker (2008) applied the focal point seed zone method to predict Jack pine growth response under climate change. Population response curves for the next 60 years suggest that future temperature increases are expected to cause a northward shift of the optimal habitat by approximately 2° from 46°N and 47°N of its current optimal habitat (Thomson & Parker 2008). Next, Thompson et al. (2009) applied the same method to study growth response of Black spruce under climate change. The

study suggested central sources in Great Lakes area are currently growing at or close to optimum and will be negatively affected by increased future temperatures, while eastern sources will benefit from warmer environments with climate change. Southern sources are required to be transferred to cooler environments, and the effects of global warming may cause significant decline of growth and the potential extirpation of local populations (Thomson et al. 2009).

2.3 Decision modeling for planned adaptation to climatic change.

A wide variety of decision tools have been applied to solve the problem of planned adaptation resulting from climate change. Conventional decision-making approaches, such as cost-effectiveness analysis, cost-benefit analysis, tolerable windows and/or safe landing approach, have been developed (Toth & Mwandosya 2001).

Modern portfolio theory provides risk-return analysis that addresses the problem of uncertain climate futures. Modern portfolio modeling is a robust method for planning under uncertainty, which was first introduced in Markowitz's (1952) landmark paper, "Portfolio Selection". A portfolio model is concerned with creating a budget constraint for an optimal composition of assets characterized by different returns with different levels of risks. Decision options are represented by a probability distribution of expected returns, and risks are estimated based on the covariance of expected returns. The decision rule is to choose a portfolio which offers the highest expected return at the same (or lower) level of risk, or to choose the portfolio with the lowest risk with the same (or higher) expected return (Markowitz 1992) .

Figge (2004) proposed a bio-folio, which applies portfolio model to biodiversity conservation. Recognizing the similarity of risk-return relations between biodiversity and investment assets, bio-folio assumed that biodiversity is a social asset, and the diversification of such an asset is related to the risk. Figge suggests that the portfolio model can be applied to manage a biodiversity portfolio by weighing the (expected) return of a portfolio of genes, species or ecosystems with the (expected) risk (Figge 2004).

Crowe and Parker (2008) presented a decision support system for planning adaptation under climate change, based on the principles of modern portfolio theory that minimized risk and maximized return in adaptation of seed sources under uncertainties over several future climate scenarios (Crowe & Parker 2008a). This study used provenance trials of 127 white spruce seed sources using the focal point seed zone method to assess adaptive variation under changing climatic conditions. The study presented an efficient frontier which indicated the optimal solution that minimized risk while satisfying the desired level of expected return.

3. Methods

Figure 1 illustrates the conceptual framework of our modeling. Producing input data for the seed portfolio model requires the application of several models; hence, before proceeding with a detailed description of the seed portfolio model, I will present a description of:

a) the inputs required by this model, and

b) the methods by which these inputs are produced.

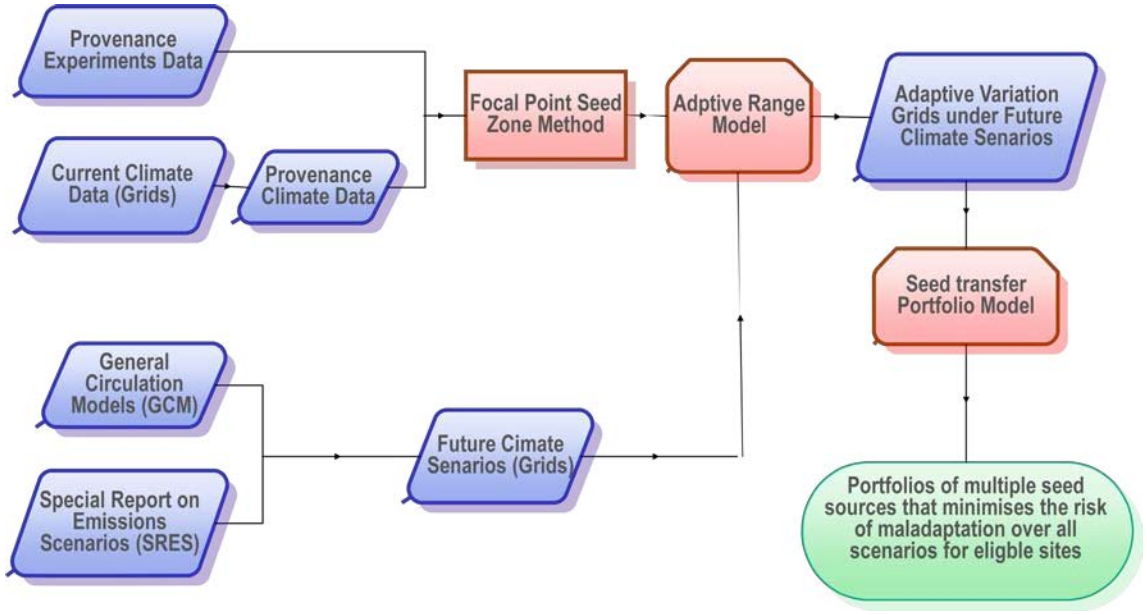

Figure.1. Framework by which multiple models and datasets provide input to the seed portfolio model.

a) Input Required by the Seed Portfolio Model

The seed portfolio model requires information on each "candidate site" within a region. A candidate site is a site that may be eligible for receiving seed sources from its provenance. The information required for each site is a) how well adapted a given seed source is estimated to be at that candidate site, under b) each of several future climatic scenarios.

b) Methods by which Inputs are Produced

As illustrated in Figure 1, the inputs for the seed portfolio model are produced using two major modeling methods: i) climate change models; and ii) the focal point seed zone method. The climate change models provide estimates of what the climatic envelope will be like at a given candidate site in the future; and ii) the focal point seed method estimates how well adapted a given seed source will be to a given climatic envelope.

We will not review in detail how the climate change models were built and executed; we use output freely distributed by multiple modeling teams from around the world whose work has already been extensively peer-reviewed (e.g., CCCMA, CGCM, CSIRO35, and HADCM). The focal point seed zone method, however, does warrant detailed description; we built our own model using this method and applied it to a unique set of data.

3.1 Focal point seed zone method

The focal point seed zone method is a statistical modeling procedure designed to delineate continuous seed procurement zones using least significant differences (LSD) to measure the adaptive variance between any given "focal point" and any other given point in a region (Parker 1992, 1996). Two sources of data are required: 1) growth data of provenance experiments derived from common garden trials, and 2) current climate data of each provenance derived from climate grids.

The focal point seed zone method is quite complex; hence we will describe it by referring to Figure 2.

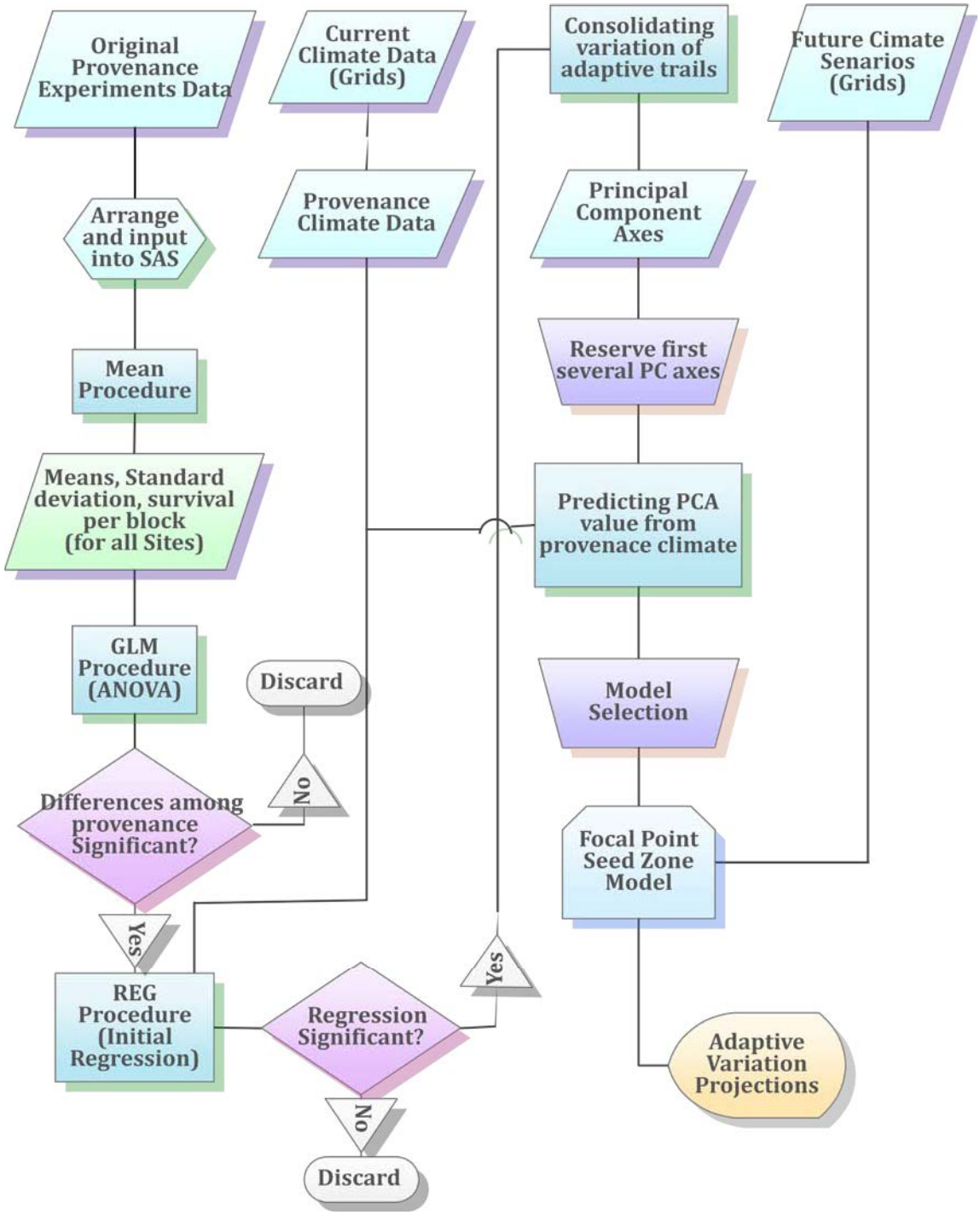

Figure. 2. Technical flow of focal point seed zone method.

Figure 2 can be divided into four major steps:

1) Collection of growth data
2) Collection of climatic data
3) Statistical analysis of growth and climatic data.
4) Production of a seed zone map.

Each step will now be described.

First, growth data are collected; *i.e*., data from provenance trials (e.g., average height, average diameter at breast height) are arranged. These data are based on the growth of different provenances at different regional common garden tests. The purpose of collecting these data is to assess the *different responses* of *different provenances* under *different regional climate conditions.*

When growth data have been collected from these provenance tests, one problem is that the tested provenances are not uniform, *i.e*., not all the provenances have been tested in all test sites. To address this problem before statistical analysis is performed, a group of provenances is preselected with maximized quantity and also tested in as many gardens as possible. When few common garden trials are involved, a simple manual listing can be used to select the optimal group for modeling; but when the records contain more than five gardens, it is necessary to run a computer-assisted enumeration program to list all possible combinations of tested provenances for selection.

The second main step in developing a focal point seed zone system is the collection of climae data for these provenances. Conventional methods to obtain

provenance climate data employ climate records from the nearest meteorological stations; but data obtained in this manner are very coarse. Data of finer resolution can be obtained through the application of geographical information systems to regional and global climate observations; i.e., these data can be interpolated into the spatial grid files. Based on validated climate models, the resolution of these grids can be as fine as 1 square kilometer. The advantage of using climate-grids is that small climatic differences between adjacent provenances can be identified. Furthermore, climate-grids can provide species variation gradient maps under current climate condition by coupling an established response model with climate grids.

The third step in the focal point seed zone method is to perform statistical analysis to determine whether significant relationships exist between any of the growth variables and the climatic variables. First, the mean and standard deviation of all growth variables, and percentage of survival of provenances, are calculated for each block. Next, a two-way analysis of variance test (ANOVA) is performed separately for growth variables of each test site using the PROC GLM (General linear model) procedure in Statistics Analysis System (SAS®) to detect any significant differences of growth variables among provenances.

The main premise of the focal point seed zone modeling approach is that tree growth variables manifest significant differences among provenances. Dependent variables that show insignificant differences among provenances are eliminated from the following procedures. Thus, the remaining variables are the informative variables for describing growth responses of various seed sources to different climate conditions.

All remaining growth variables are regressed against climatic variables to determine whether the relationship is significant. A linear regression procedure, combined with maximum R-Square improvement selection, is applied. During this procedure, each growth variable is regressed against all climatic variables one by one to find the best one-variable model. Only a growth variable with a significant regression is retained. In this way, the procedure determines whether these variables are indeed responsive to climatic variables. Since the focus is how trees respond to different climate conditions, variables that reflect growth response to other factors are filtered out.

Principal components analysis is then applied to summarize the main components of variation in the remaining growth variables. The first several axes which have an accumulated variance greater than a certain threshold (e.g., 80% or 85%) are used to represent original growth variables. Therefore, the number of predictors for describing adaptive variation can be reduced to three or fewer.

An LSD value for each principal component can then be obtained by using the weighted average LSD value of the original variables. The raw LSD value (α=0.05) of each original variable can then be determined from the ANOVA test and divided by the standard deviation of the variable to express the LSD as a number of standard deviations. Standardized LSD values for each original variable are then multiplied by their respective variable loadings (absolute value of eigenvectors), summed, and divided by the sum of the absolute loadings again to produce a weighted average LSD. Multiplied by principal components scores, these weighted LSD values are capable of converting standard deviation values to LSD values, which are then applied to quantify the magnitude of adaptive variation of provenances (Crowe & Parker 2005).

In a final regression, normalized provenance factors for each principal component axes are regressed against climatic variables by multiple stepwise linear regressions. To obtain valid results, multiple stepwise linear regressions with different significant levels are performed to compare the results between groups. The R-square selection method for multiple linear regressions can also be applied to find subsets of climatic variables that best predict principal component axes by linear regression in the given sample.

The fourth major step in the focal point seed zone method is the production of a seed zone map. A focal point seed zone map presents unique seed procurement areas for any specific provenance (seed source). To generate a focal point seed zone map for any specific seed source, geographic location of the provenance is used to identify scores of PC-axes for this seed source. PC-LSD maps are then adjusted with corresponding scores of PC axes to represent the adaptive variation pattern of the specific seed source. Finally, a focal point seed zone map is produced by overlaying contoured grids of each PC-LSD map and intersection contoured intervals are used to predict potential procurement regions.

3.2 The portfolio optimization model

The portfolio optimization model formulated for this research has the objective of minimizing the total risk of all portfolios of seeds to be transferred *from* a set of seed sources *to* a set of eligible sites across a region. The "risk" of each site's portfolio is measured by the expected covariance in adaptaptive suitability (measured in LSD) of all

pairs of seed sources (within the portfolio) to one another across multiple feasible climatic futures. In effect, by minimizing the covariance of expected adaptation one seeks a robust solution to the problem of selecting provenances in an uncertain future; *i.e*, a set of seed sources that will perform well across each of the many plausible climatic futures. The mathematical formulation of the model is presented below.

Indices and sets

i, I = index and set of eligible sites

j,J = index and set of candidate seed sources

Parameters

$COV_{ijj'}$ = the covariance in return (mean LSD) at site i , across the mean of all climatic scenarios, between sources j and j'

R_{ij} = the expected return (mean LSD) for source j at site i

LSD_{MAX} = the maximal acceptable maladaption level

VAR_{MAX} = the maximal allowable fluctuation of LSD over time

Decision variables

X_{ij} = the proportion of seed source j at site i

Objective function

Minimize:

[1] $$\sum_{i \in I} \sum_{j \in J} \sum_{j' \in J} X_{ij} X_{ij'} Cov_{ijj'}$$

subject to

[2] $$\sum_{j \in J} X_{ij} = 1 \qquad \forall\, i \in I$$

[3] $$\sum_{j \in J} X_{ij} R_{ij} \leq LSDmax \qquad \forall\, i \in I$$

The objective function [1] of the model is to minimize the total covariance existing between all selected pairs of sources at all sites within the region. Equation [2] ensures that the sum of the proportion of seed sources at each site equals 100%. Equation [3] ensures that the expected adaptive suitability of each portfolio to each site not exceed the maximal acceptable level of maladaptation (measured in LSDs).

Figure 3 illustrates how the optimization model will be used. Here, one can observe three steps in applying the model: 1) preliminary selection of sites eligible for receiving seed; 2) execution of the optimization model; and 3) analyses of the results and mapping.

After the focal point seed zone model is used to estimate how well adapted each site within a region might be to each seed source (under multiple climatic futures), it immediately becomes evident that *some* sites are not suitable for receiving any of the designated seed sources. Hence, some candidate sites must be removed from the data set before the model is executed. Sites are removed based on two criteria. Either a) the mean

LSD for a given seed source at a given site under a given scenario is too high (i.e., it exceeds LSD_{MAX}; or b) the variance of expected LSD of a given seed source on a given site, within a given scenario, is too great (i.e., it exceeds VAR_{MAX}).

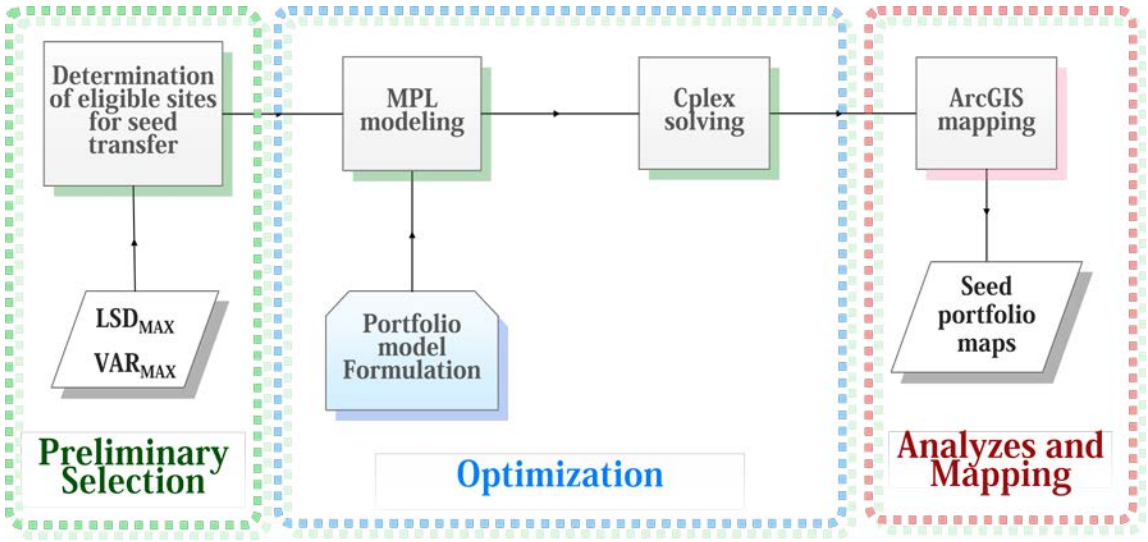

Figure. 3. Technical flow of optimization procedure.

The second step in the optimization process is the execution of the model. The model is built using mathematical programming software (MPL®); and, since the objective function is non-linear, CPLEX's barrier algorithm is used to find the optimal solution.

The final step in optimization procedure is analysis and mapping. Here, we are interested in using the model to help support decisions on allocating seed at the regional scale; i.e., we are interested in addressing the following questions:

a) Which sites are covered with minimal risk, and which are not?

b) Which seed sources are used most often at the regional scale and which are not?

c) How sensitive are these solutions to different assumptions on the future quantities of CO_2 emissions?

4. Case Study

We applied our modeling approach to the problem of deploying 7 black spruce seed sources to 924 candidate sites in the province of Ontario. There are 4 elements of this case study requiring documentation:

1) Source of provenance data;
2) Preparation of provenance data;
3) Source of current climate data;
4) Source of future climate data.

4.1 Source of provenance data

In 1967, a nation-wide black spruce provenance study was initiated by the Canadian Forest Service. Twenty field experiments were established across the species range between 1973 and 1977. Three hundred twenty seven seed sources were collected and planted in multiple test sites. Height and diameter at breast height were measured in 20 sites at the age of 36-year in 2005. Figure 4 presents the distribution of provenances and experimental sites. Growth data of black spruce were gathered from 20 test sites consisting of 327 provenances, which covered the entire black spruce range in Canada. Most of the seed sources were tested in several adjacent sites and none were tested at all sites.

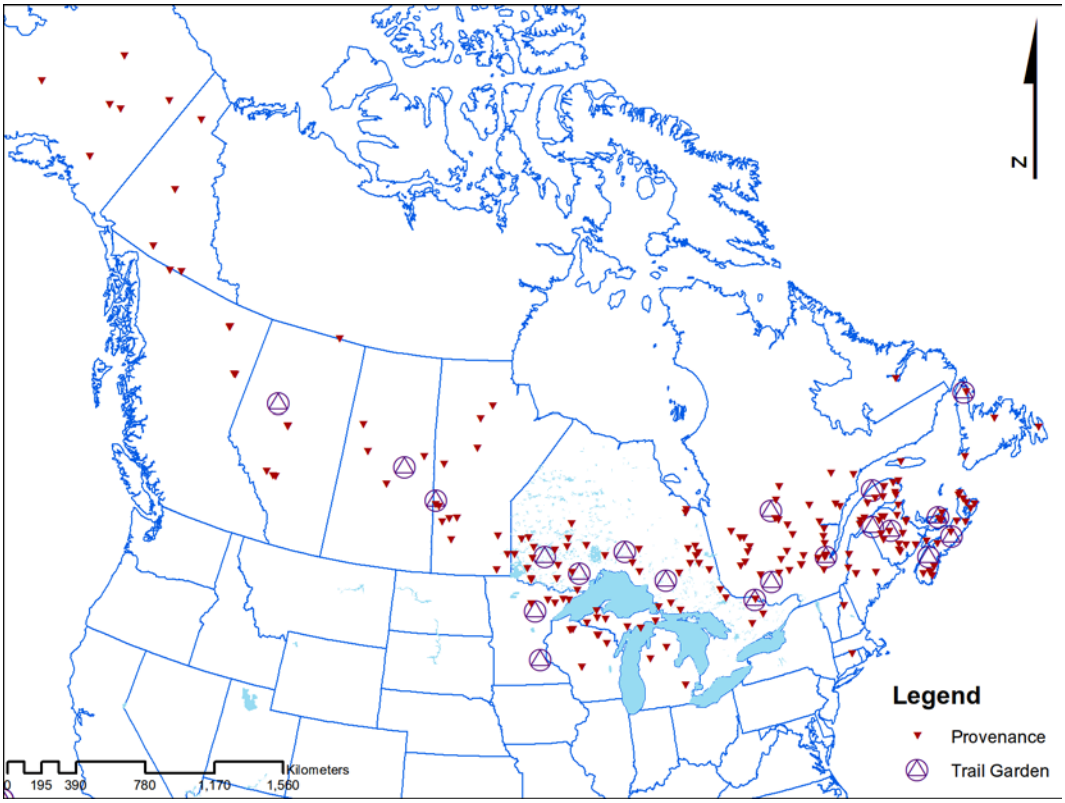

Figure. 4. Distribution of provenances and experimental sites of the black spruce provenance study in Canada.

4.2 Preparation of provenance data

For the purpose of establishing focal point seed zone models, growth data of a group of seed sources in the study area had to be selected from the 327 provenances. To precisely assess how the populations are adapted to different climate conditions, we selected those provenances should have tested on the most sites. Twenty test sites produced about 612,645 combinations (i.e., pairs of test sites) of test sites in this study. To find the optimal combination of sites on which enough provenances were tested, a computer program was written to enumerate and select the optimal combinations to establish the focal point seed zone models. Throughout this program, all feasible

combinations were listed and sorted into groups of the different numbers of test site combinations.

Table 2 lists the first five optimal candidate combinations for each group. The site combination with 24 provenances tested (*a* in Table 2) and the combination with 34 provenances tested (*b* in Table 2) qualified, since both test sites and provenances from these two candidate combinations covered the study area. The set of 7 test sites has an additional test site located in eastern Ontario which was capable of providing more information about black spruce's growth response under climate conditions of this region. Thus, growth data of these 24 provenances were extracted for focal point seed zone modeling. Figure 5 illustrates distribution of the 24 provenances and 7 test sites. Table 6 in appendix describes the geographic locations of these test sites.

Table 2. Selection of best 5 in each number of combination groups of enumeration results.

Number of test site	Test Site	Number of tested provenances
10	3,4,7,10,12,13,14,16,18,20	5
9	3,4,7,10,12,13,16,18,20	6
9	4,7,10,12,13,14,16,18,20	6
9	3,4,7,10,12,13,14,16,18	5
9	3,4,7,10,12,13,14,16,20	5
9	3,4,7,10,12,13,14,18,20	5
8	3,4,7,10,13,14,16,20	8
8	4,7,10,12,13,16,18,20	7
8	3,4,7,10,12,13,16,18	6
8	3,4,7,10,12,13,16,20	6
8	3,4,7,10,12,13,18,20	6
7	3,7,10,12,14,16,18	24 (*a*)
7	3,4,7,10,13,16,20	9
7	4,7,10,13,14,16,20	9
7	3,4,7,10,13,14,16	8
7	3,4,7,10,13,14,20	8
6	3,7,10,12,16,18	34 (*b*)
6	3,7,10,12,14,18	26
6	3,7,12,14,16,18	25
6	7,10,12,14,16,18	25
6	3,7,10,12,14,16	24

Note: Test sites and geographic locations are listed in Table 6.

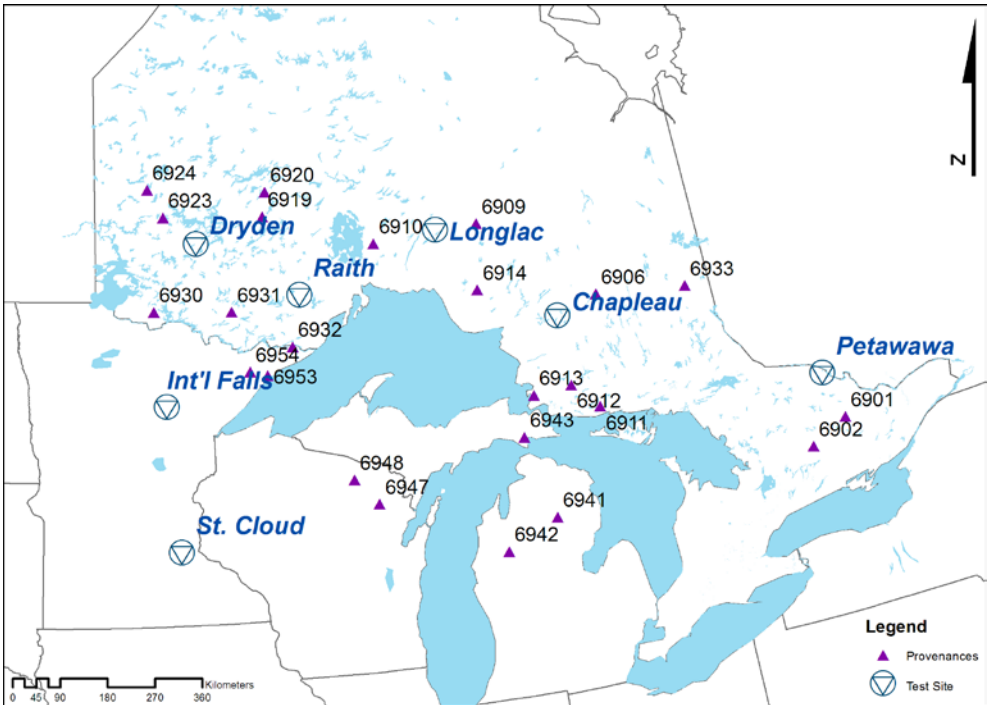

Figure. 5. Distribution of 7 test sites and 24 provenances for modeling.

4.3 Source of current climate data

Current climate data (interpolations of observed data, representative of 1950-2000) were provided by WORLDCLIM, which is a set of global climate grids with high spatial resolution (1km) (Hijmans et al. 2005). Data were available for downloading at www.worldclim.org. Current climate data of North America were downscaled from world scale original data, which contain 36 climate variables: average monthly minimum temperatures, average monthly maximum temperatures, and monthly precipitation. All data were downloaded and downscaled for estimating the climate of the provenances.

4.4 Source of future climate data

Four GCMs (CGCm31, CSIRO-mk35, MIROC3.2 and NCARC), coupled with three emission scenarios (A1b, A2 and B1) for three periods (2011-2040, 2041-2070,

and 2071-2100), totalling 36 scenarios of future climate data, were provided by Dan McKenney of the Canadian Forestry Service, Sault Ste. Marie, Ontario (McKenney et al. 2009). The dataset were provided as gird files, which have covered our study area. The resolution of these grids is 10 km. Thirty-six climate variables, average monthly minimum temperatures, average monthly maximum temperatures, and monthly precipitation for each month, were obtained for every projection as 10km resolution grid file covered the study region.

5. RESULTS

The results are presented in two sections:

1) The results from applying the focal point seed zone method to estimate how well adapted the set of 7 seed sources will be to a set of sites across the province of Ontario, under current and multiple future climatic scenarios;
2) The portfolio optimization results.

5.1 Results from Applying the Focal Point Seed Zone Method

5.1.1 Adaptation to Current Climate

a) Statistical Results

The proportion of total variation explained by each of the principal components is shown in Figure 6. Principal components axes 1 to 4 cumulatively explained 79.78% variation in original growth variables. The remaining PC axes contributed relatively little to explain the variation, accounting for about 3% on average.

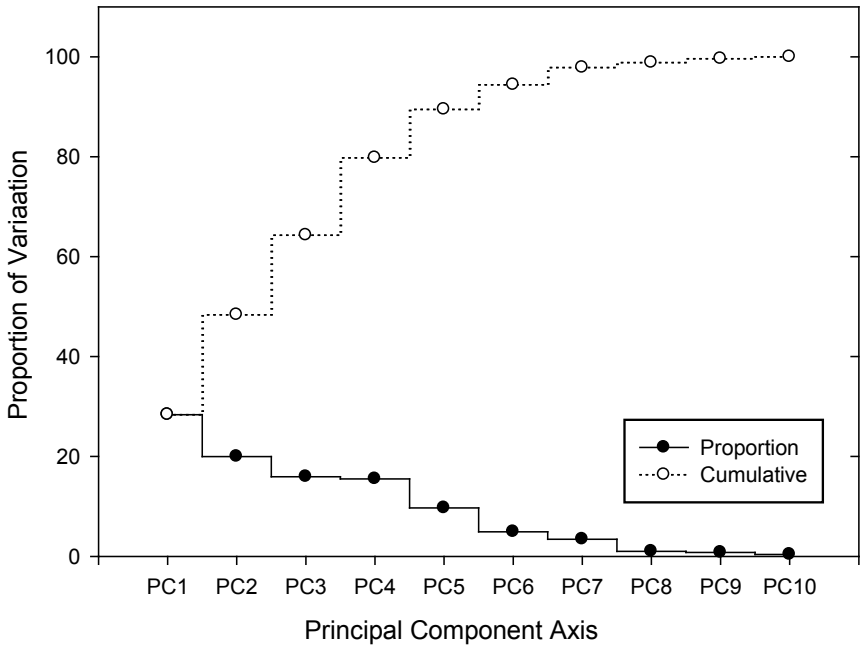

Figure. 6. Proportion and cumulative variation in growth variables explained by principal components axes.

The first four PC-axes explained nearly 80% of the total variation; thus, they have been applied for the following multiple linear regression. Details of the PCA for these four axes are presented in Table 3. Principal component 1 (PC1) explained 28.37% of the total variation, followed by principal component 2 (PC2) which explained 19.98%. Finally, principal component 3 (PC3) and 4 (PC4) explained 15.93 and 15.50 percent of the total variation respectively.

Growth variables in site 3, site 7, and site 10 demonstrated the strongest positive eigenvectors for PC1, while growth variables in site 12, site 16, and site 18 exhibited negative eigenvectors for PC1. Geographic locations of sites 3, 7 and 10 are generally north of sites 12, 16 and 18, thus suggesting that PC1 represents growth potential limitation related to the latitude.

Table 3. Eigenvalue, proportion and cumulative of variation explained, eigenvectors of first four PC axes.

	PC1	PC2	PC3	PC4
Eigenvalue	2.8367	1.9981	1.5930	1.5496
Proportion	0.2837	0.1998	0.1593	0.1550
Cumulative	0.2837	0.4835	0.6428	0.7978
Eigenvectors	PC1	PC2	PC3	PC4
Site 3 average height	0.3944	0.3138	-.4067	-.1683
Site 3 average DBH	0.3568	0.3828	-.4121	-.0766
Site 7 average height	0.4361	-.1019	0.1092	0.4770
Site 7 average DBH	0.3476	0.0024	0.0864	0.6032
Site 10 average height	0.4362	-.0939	0.3663	-.3110
Site 10 average DBH	0.3976	-.1131	0.2663	-.4685
Site 12 average DBH	-.2138	0.3354	0.2293	0.0104
Site 14 average DBH	0.0333	0.3084	0.5111	0.1713
Site 16 average DBH	-.0929	0.4630	-.1208	0.1213
Site 18 average DBH	-.0480	0.5487	0.3325	-.1192

Multiple regression procedures were performed to determine regression models for predicting adaptive variation of black spruce. The maximal R-square selection method for multiple linear regressions was applied to list several best regression models for each PC axis. For each PC axis, these models are divided into several groups, which used different number of independent variables. By evaluating R-square, tolerance and significance of each independent variable for each model, four best regression models for modeling PC axes of variation are determined and presented in Table 4. Each of the selected models was significant ($P < 0.05$) in explaining the variation in principal components factors scores, and each of the entered variables demonstrated significant t-values (< 0.05).

Table 4. The best multiple linear regression models predicted by regression of PC1, PC2, PC3, and PC4 against each of 36 climate variables.

Dependent Variable	Independent Variables	Parameter Estimate	Parameter		Model	
			P > t	Tolerance	*P* > F	R^2
PC1	Intercept	21.7033	.0016	.	.0003	.5954
	MaxT08	-0.7372	.0040	0.2741		
	MinT04	0.3752	.0247	0.2014		
	Prec04	-0.0669	.0010	0.5648		
PC2	Intercept	12.37991	.0001	.	.0002	.5631
	Prec07	-0.07657	.0002	0.8295		
	Prec09	-0.01552	.0003	0.8295		
PC3	Intercept	6.7664	.0048	.	.0167	.3227
	MaxT03	-0.5795	.0086	0.1948		
	MinT02	0.3497	.0047	0.1948		
PC4	Intercept	9.0904	.0038	.	.0136	.4677
	MaxT04	-0.4949	.0016	0.4899		
	Prec02	-0.0601	.0119	0.3194		
	Prec07	-0.1270	.0018	0.2687		
	Prec08	0.0880	.0148	0.6531		

R-squares of regression models for PC1 and PC2 are desired; both of them were greater than 0.55. PC4 shows a better regression than PC3 in both model significance and variable tolerance. When considered together, the R-square value, model significance, and variable tolerance indicate that climatic variables have weak correlation to explain PC3. Therefore, the regression model for PC3 was excluded from calculating the general predicted factor scores (i.e., the total weighted principal component scores of LSD) to express adaptive variation in one single variable.

The LSD scores of PC1, PC2 and PC4 are therefore weighted according to the proportion of total variation explained by each principal component, and then summed to generate the new variable – total weighted LSD (equation [7]).

$$PC_1 = 21.7034 - 0.7372MaxT_{08} + 0.3751MinT_{04} - 0.0669Prec_{04} \quad [4]$$

$$PC_2 = 12.3799 - 0.0766Prec_{07} - 0.0667Prec_{09} \quad [5]$$

$$PC_4 = 9.0904 - 0.4949MaxT_{04} - 0.0601Prec_{02} - 0.1270Prec_{07} + 0.088Prec_{08} \quad [6]$$

$$Total\ weighted\ LSD = 0.2837\mathrm{LSD}_{PC1} + 0.1998\mathrm{LSD}_{PC2} + 0.1550\mathrm{LSD}_{PC4} \quad [7]$$

b) Mapped Results of Current Adaptation

The predicted factor scores for PC1, PC2 and PC4, are presented in the following figures as maps, where the factor scores for each grid are expressed as units of standard deviation (SD).

The first principal component grid (Figure 7) reveals that growth potential of black spruce is greatest in south-eastern Ontario, and decreases as it moves northward. The color gradient represents the distance of a location from the average performance of 24 provenances on PC1 factor scores.

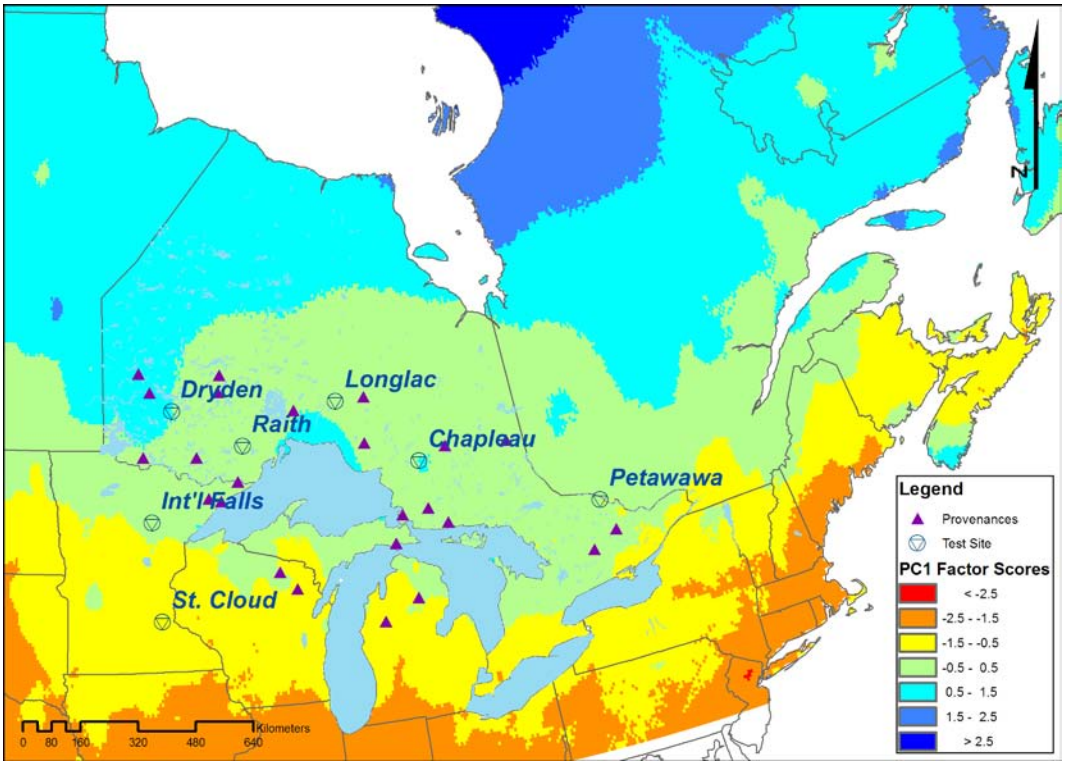

Figure. 7. Predicted factor scores for PC1 based on regression of current climate data.

Factor scores for principal component 2, represents the growth constraint of precipitation on growth. Figure 8 indicates that the best adaptive areas extend from the lake shore and southeast coast to the interior regions.

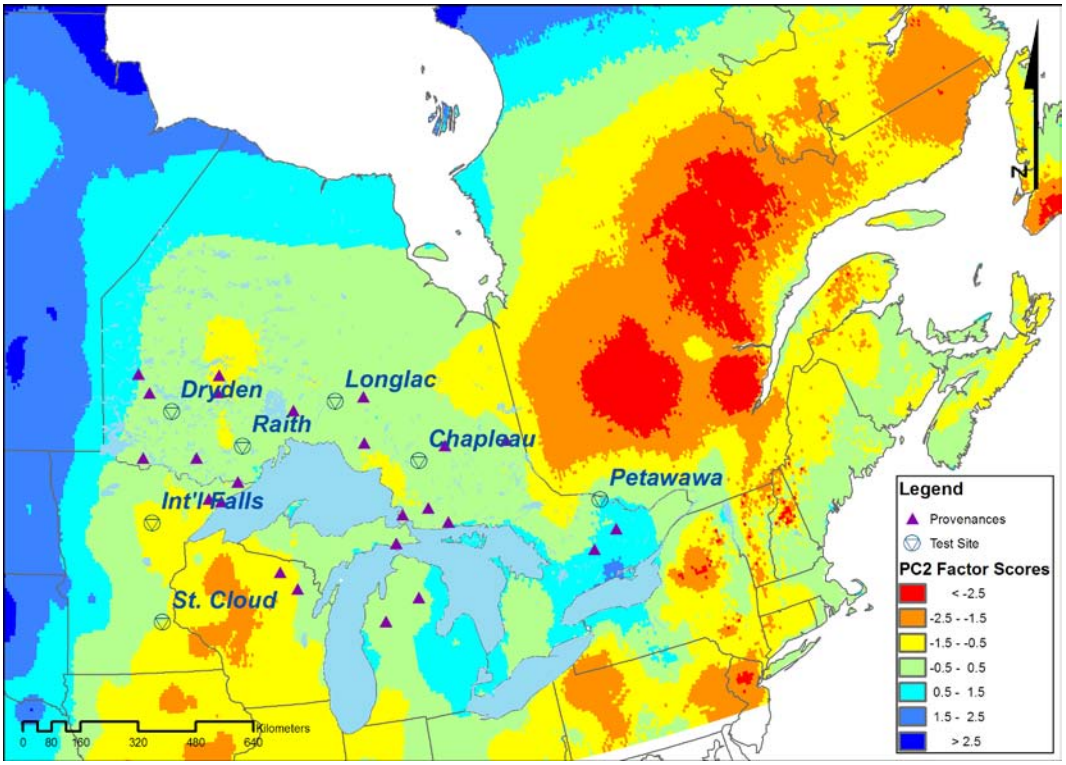

Figure. 8. Predicted factor scores for PC2 based on regression of current climate data.

Factor scores for principal component 4 synthesize the effects of monthly maximal temperatures and precipitations (Figure 9). Southern Ontario, Michigan, Wisconsin are the best adapted regions in this map. Compared with the maps of factor scores in the first two, areas that have extreme LSD values for both positive and negative sides are appear as regular latitudinal gradient belts in southern and northern extensions. This indicates that PC4 represents a more restricted limitation to black spruce growth of temperature and precipitation than other PC axes.

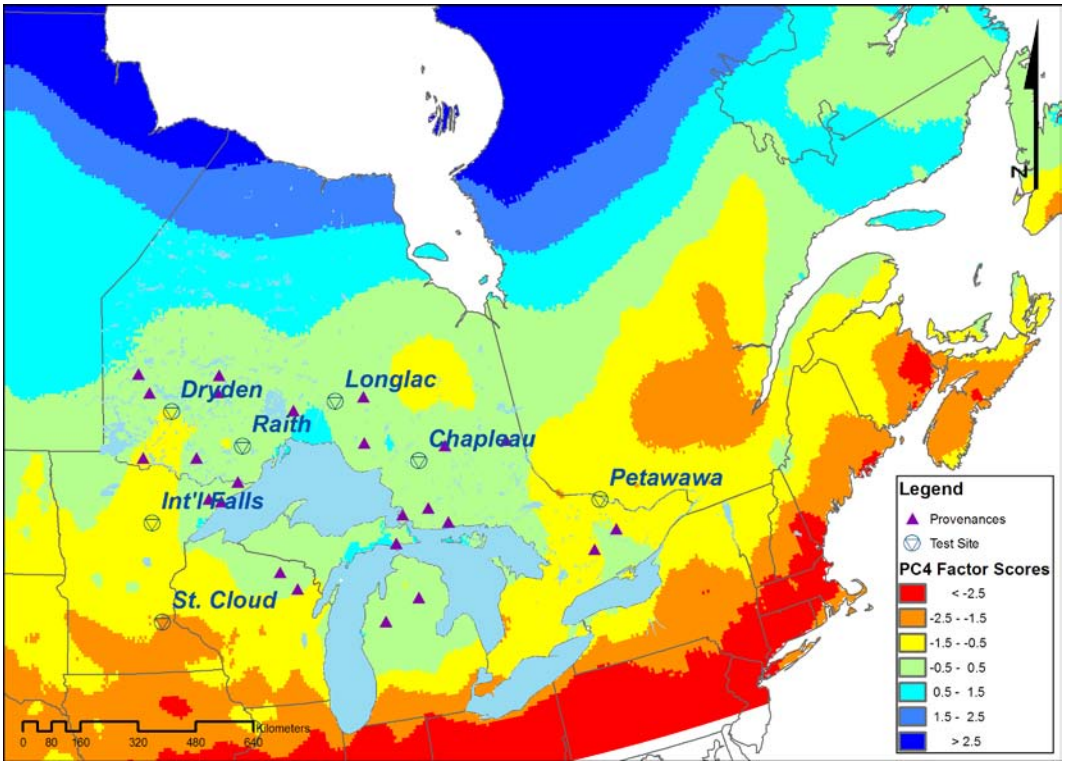

Figure. 9. Predicted factor scores for PC4 based on regression of current climate data.

Figure10 presents a grid of predicted factors for the summation of weighted principal component scores; i.e., the total weighted predicted factor scores of principal components 1, 2, and 4. Green areas are the best adapted for black spruce, within which LSD is less than ± 0.5.

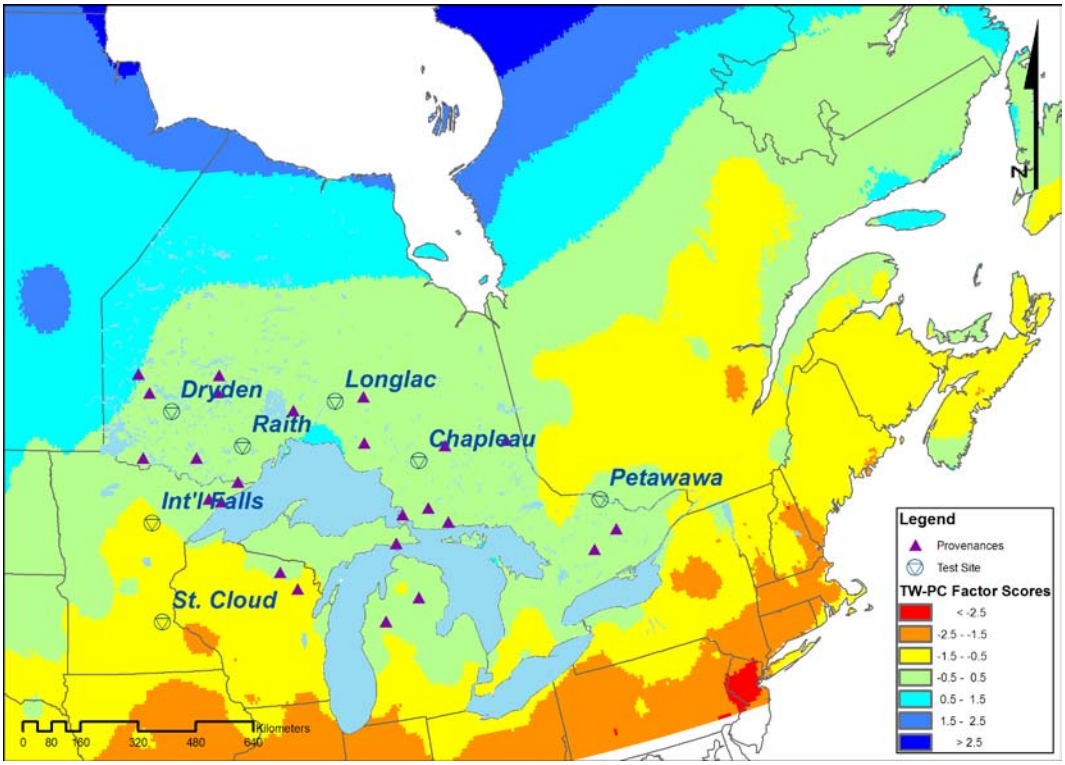

Figure. 10. Predicted total weighted LSD based on regression of current climate data.

5.1.2 Adaptation to Future Current Climate Scenarios

Based on the regression model (equation 7) and the future climate predictions (obtained from GCMs coupled with SRESs), a set of maps was generated to represent adaptive variation patterns of black spruce under multiple possible climate conditions. Four climate models (CGCM31, CSIROmk35, MIROC, NCARC) each coupled with 3 emission scenarios (A1b, A2, B1) for three 3 periods (2011-2040, 2041-2070, 2071-2100) were used to describe plausible future climates. Hence, 36 projections of the adaptive variation pattern maps were produced. These maps are presented in Figures 11 to 46. Below is an index of these 36 maps:

CGCM31 × A1b ------ Figure 11 - 13
CGCM31 × A2 ------ Figure 14 - 16
CGCM31 × B1 ------ Figure 17 - 19
CSIROmk35 × A1b ------ Figure 20 - 22
CSIROmk35 × A2 ------ Figure 23 - 25
CSIROmk35 × B1 ------ Figure 26 - 28
MIROC × A1b ------ Figure 29 -31
MIROC × A2 ------ Figure 32 - 34
MIROC × B1 ------ Figure 35 - 27
NCARC × A1b ------ Figure 38 - 40
NCARC × A2 ------ Figure 41 - 43
NCARC × B1 ------ Figure 44 - 46

In reviewing these 36 projections, several dominant patterns emerge; namely:

1) most scenarios present a northward shifting of current adaptive zones;
2) the rates of this northward shifting varies by scenario;
3) most current adaptive zones are forecast to shrink along the southern edge and expand along the northern edge; and
4) fragmentation, or even disappearance of many zones of adaptation, is quite common in future scenarios.

Details supporting these trends are summarized in Table 5.

Table 5. Variation comparisons of predicted 1.0 LSD adaptive zones for average score of 24 seed sources under 12 future climate scenarios.

Climate model	Emission Scenario	General trend	Rate of change	Southern range edge	Northern range edge	Adaptive zone of Provenance
CGCm31	A1	Slow northward shift	Constant rate	Fast shrinking	Slight expanding	Fragmented around the Grate Lake
	A2	Slow northward shift	Accelerating	Fast shrinking	Fast expanding	Disappeared, Shrinking
	B1	Shrinking	Constant rate	Fast shrinking	Shrinking	Fragmented
CSIROmk35	A1	Fast northward shift	Decelerating	Fast shrinking	Fast expanding	Fragmented, disappeared
	A2	Very fast northward shift	Accelerating	Very fast shrinking	very fast expanding	Fragmented, disappeared
	B1	Very fast northward shift	Decelerating	Slow shrinking	Slow expanding	Disappeared
MIROC	A1	Northward shift	Constant rate	Shrinking	Expanding	Disappeared
	A2	Northward shift	Accelerating	Shrinking	Expanding	Fragmented,
	B1	Northward shift	Constant rate	Shrinking	Expanding	Fragmented, mostly disappeared
NCARC	A1	Stay in situ	Constant rate, slow	Shrinking	Slight expanding	Southern area disappeared
	A2	Northward shift	Accelerating	Fast shrinking	Fast expanding	Disappeared
	B1	Stay in situ	Constant rate	expanding in west, shrinking in east	Shrinking in west, expanding in east	Partial disappeared

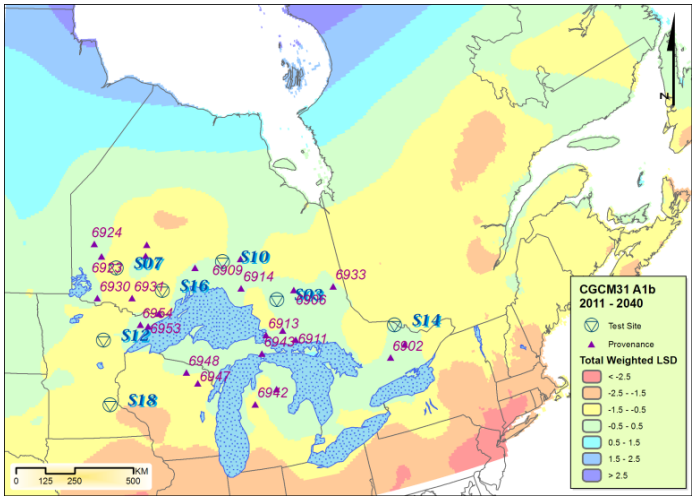

Figure. 11. Predicted total weighted LSD during 2011 - 2040 based on CGCM31 coupled with SRES A1b.

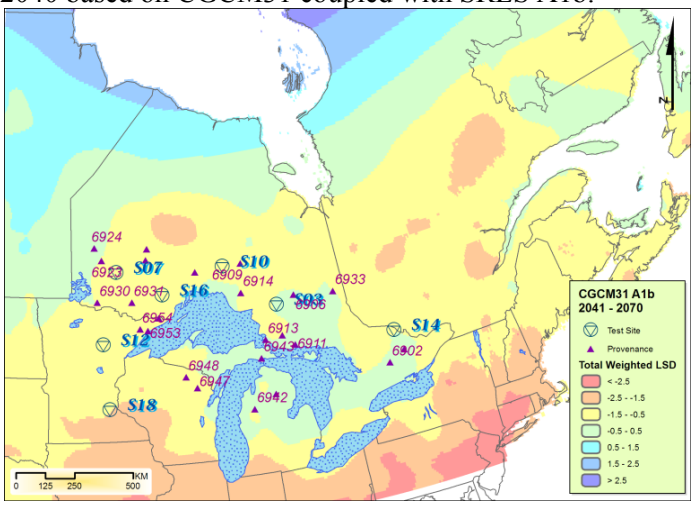

Figure. 12. Predicted total weighted LSD during 2041 - 2070 based on CGCM31 coupled with SRES A1b.

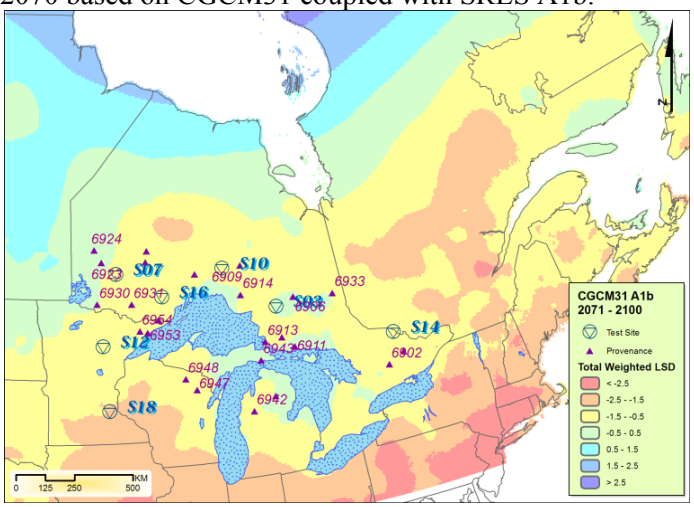

Figure. 13. Predicted total weighted LSD during 2071 - 2100 based on CGCM31 coupled with SRES A1b.

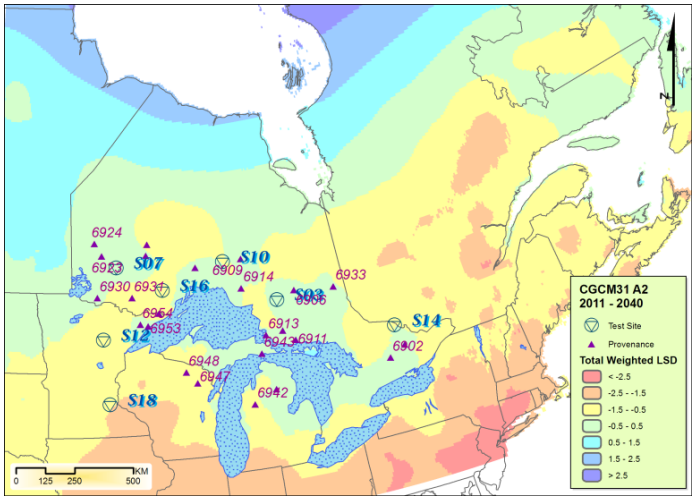

Figure. 14. Predicted total weighted LSD during 2011 - 2040 based on CGCM31 coupled with SRES A2.

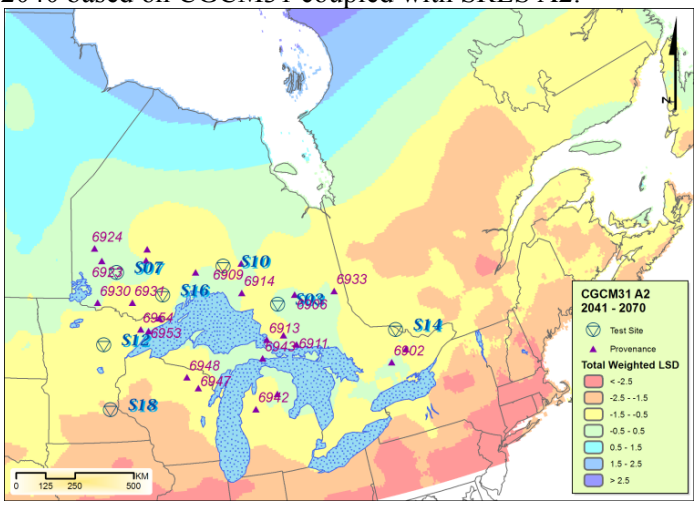

Figure. 15. Predicted total weighted LSD during 2041 - 2070 based on CGCM31 coupled with SRES A2.

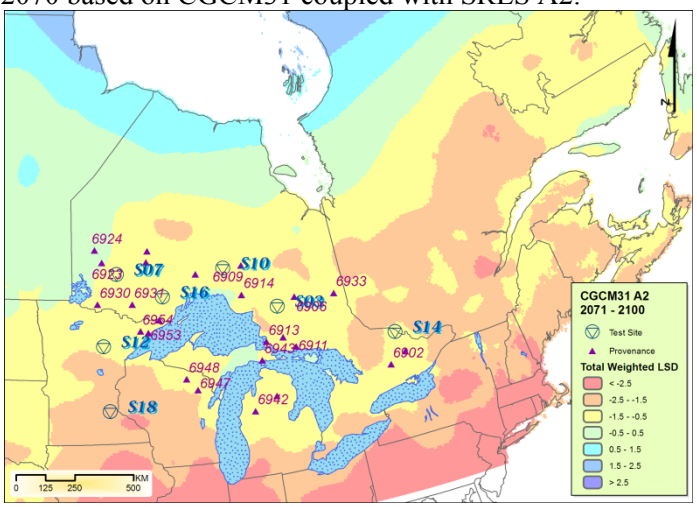

Figure. 16. Predicted total weighted LSD during 2071 - 2100 based on CGCM31 coupled with SRES A2.

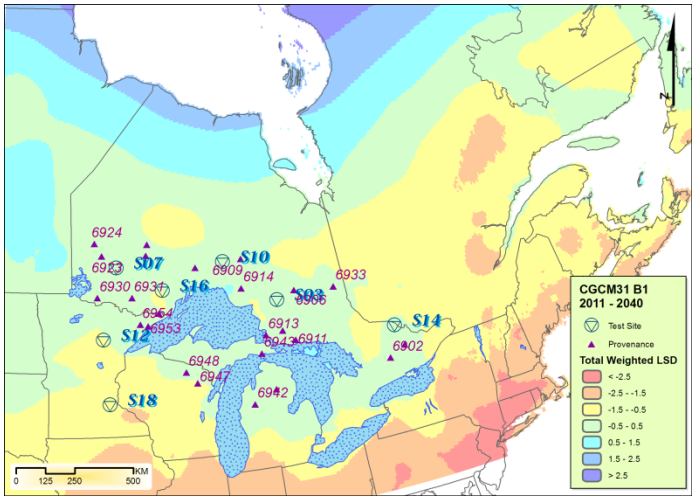

Figure. 17. Predicted total weighted LSD during 2011 - 2040 based on CGCM31 coupled with SRES B1.

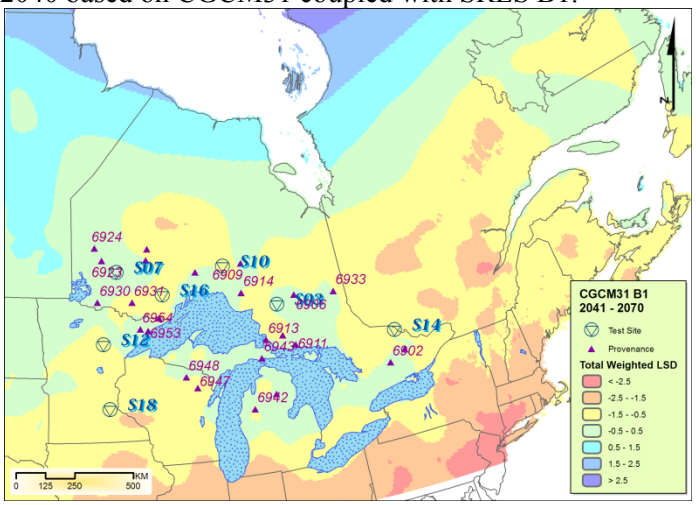

Figure. 18. Predicted total weighted LSD during 2041 - 2070 based on CGCM31 coupled with SRES B1.

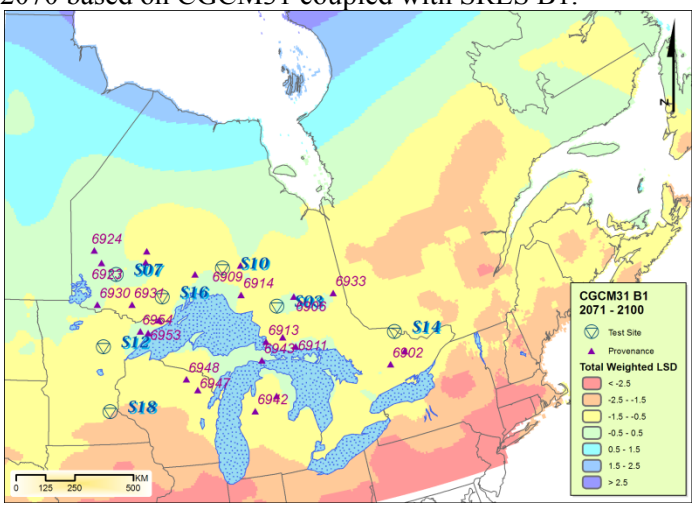

Figure. 19. Predicted total weighted LSD during 2071 - 2100 based on CGCM31 coupled with SRES B1.

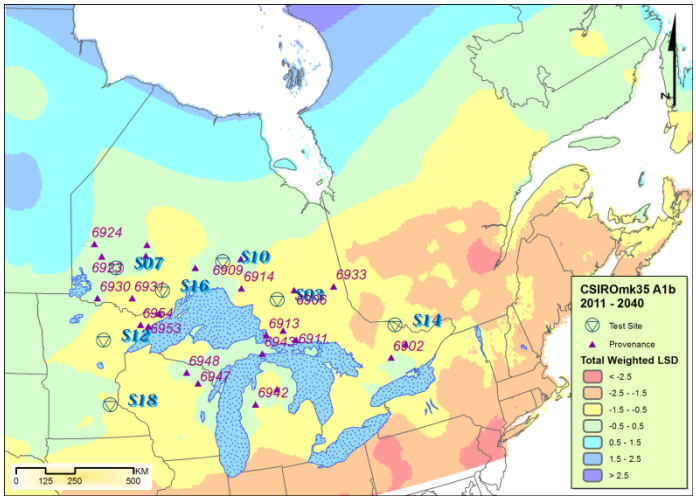

Figure. 20. Predicted total weighted LSD during 2011 - 2040 based on CSIROmk35 coupled with SRES A1b.

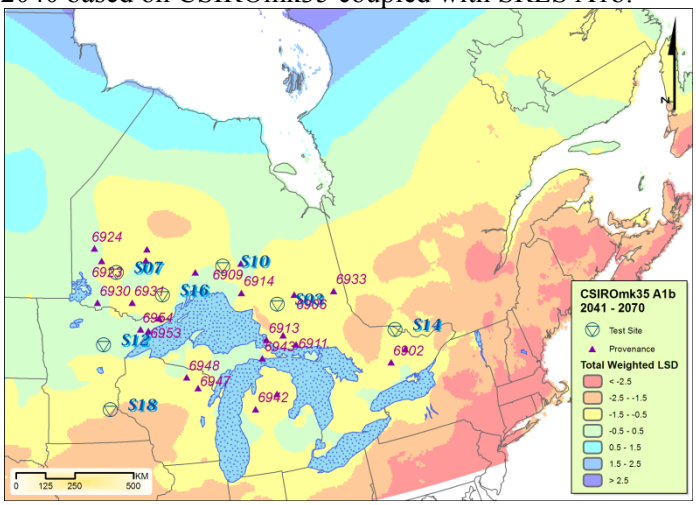

Figure. 21. Predicted total weighted LSD during 2041 - 2070 based on CSIROmk35 coupled with SRES A1b.

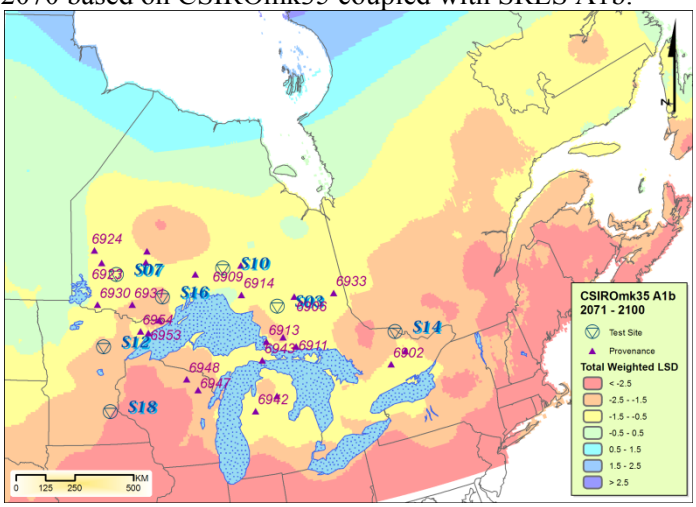

Figure. 22. Predicted total weighted LSD during 2071 - 2100 based on CSIROmk35 coupled with SRES A1b.

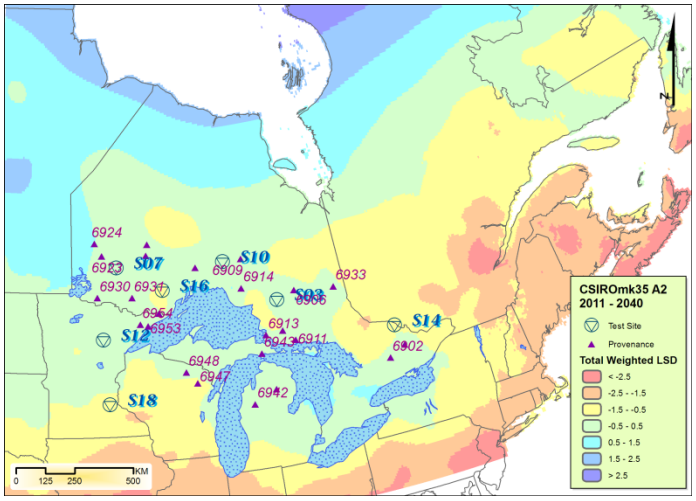

Figure. 23. Predicted total weighted LSD during 2011 - 2040 based on CSIROmk35 coupled with SRES A2.

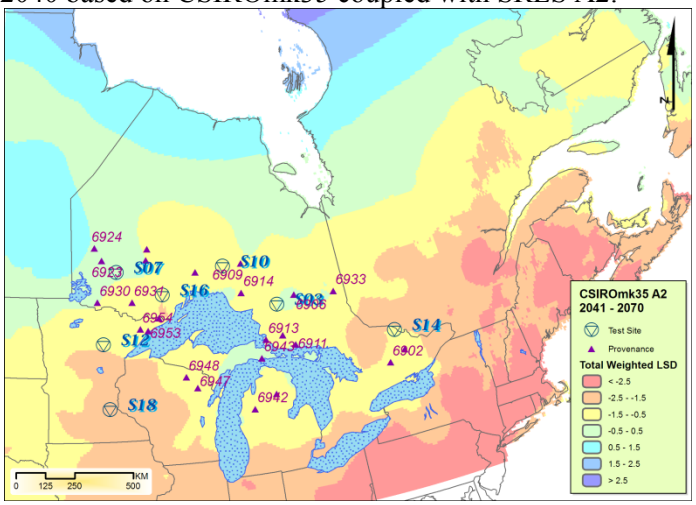

Figure. 24. Predicted total weighted LSD during 2041 - 2070 based on CSIROmk35 coupled with SRES A2.

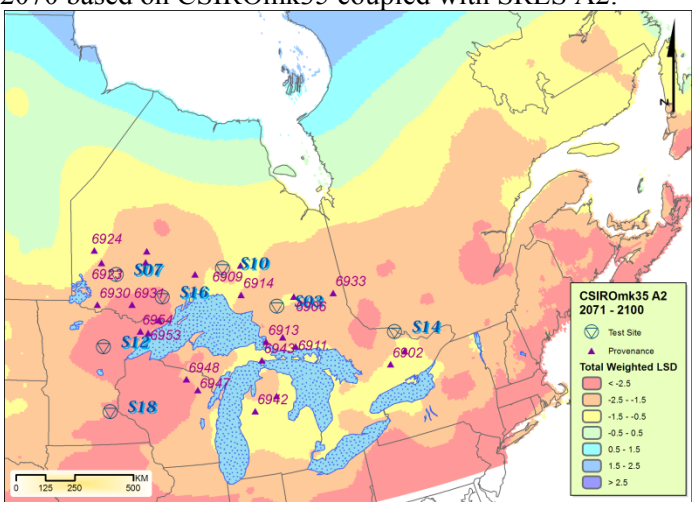

Figure. 25. Predicted total weighted LSD during 2071 - 2100 based on CSIROmk35 coupled with SRES A2.

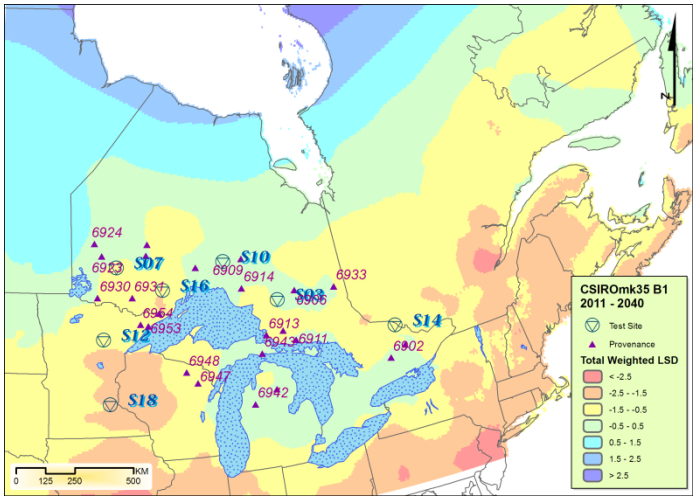

Figure. 26. Predicted total weighted LSD during 2011 - 2040 based on CSIROmk35 coupled with SRES B1.

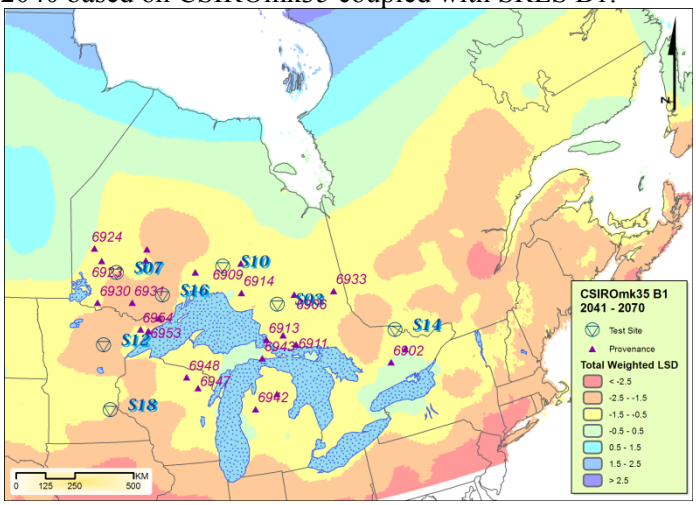

Figure. 27. Predicted total weighted LSD during 2041 - 2070 based on CSIROmk35 coupled with SRES B1.

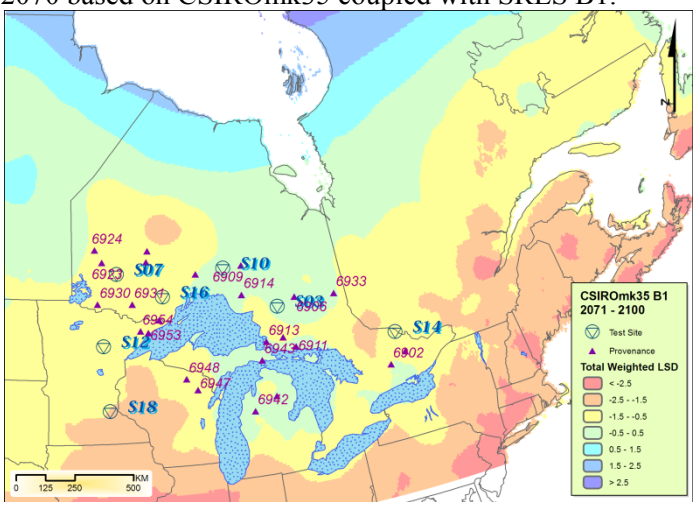

Figure. 28. Predicted total weighted LSD during 2071 - 2100 based on CSIROmk35 coupled with SRES B1.

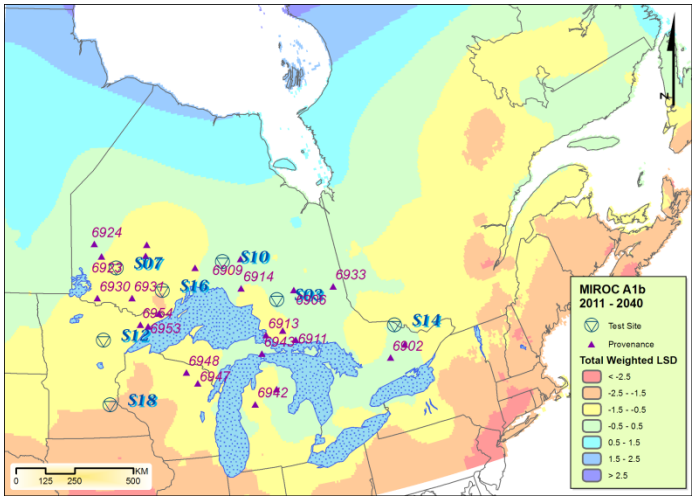

Figure. 29. Predicted total weighted LSD during 2011 - 2040 based on MIROC coupled with SRES A1b.

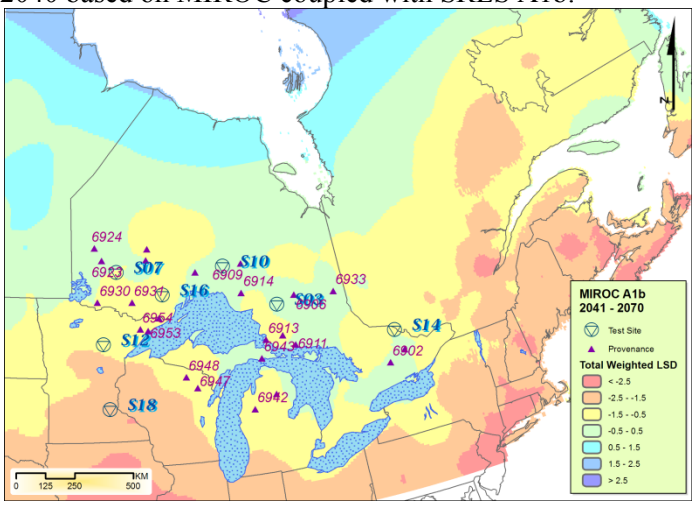

Figure. 30. Predicted total weighted LSD during 2041 - 2070 based on MIROC coupled with SRES A1b.

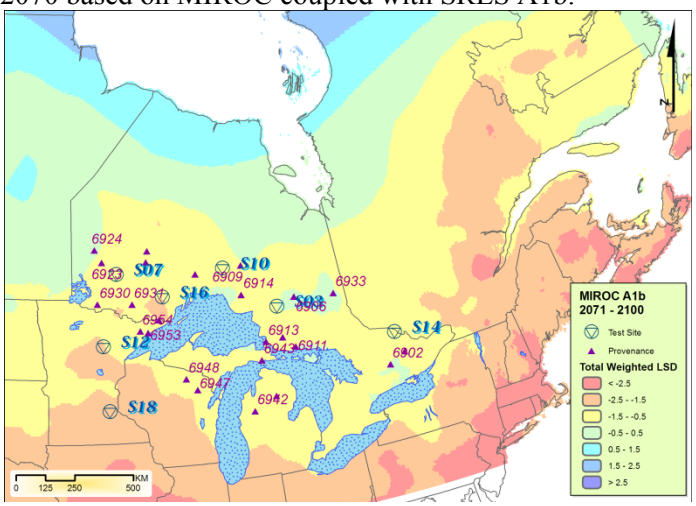

Figure. 31. Predicted total weighted LSD during 2071 - 2100 based on MIROC coupled with SRES A1b.

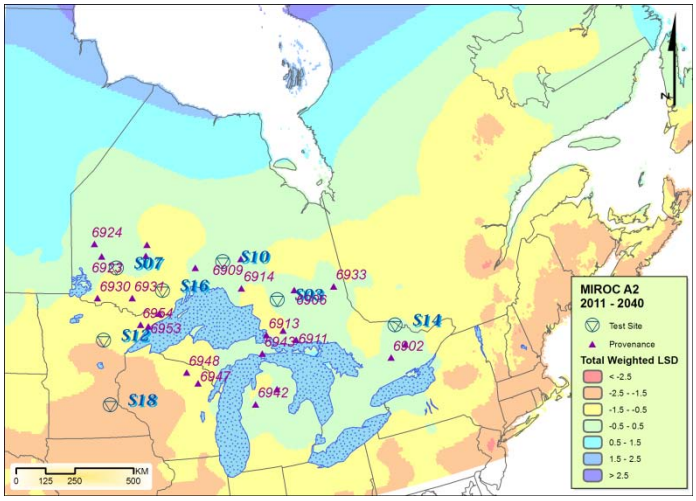

Figure. 32. Predicted total weighted LSD during 2011 - 2040 based on MIROC coupled with SRES A2.

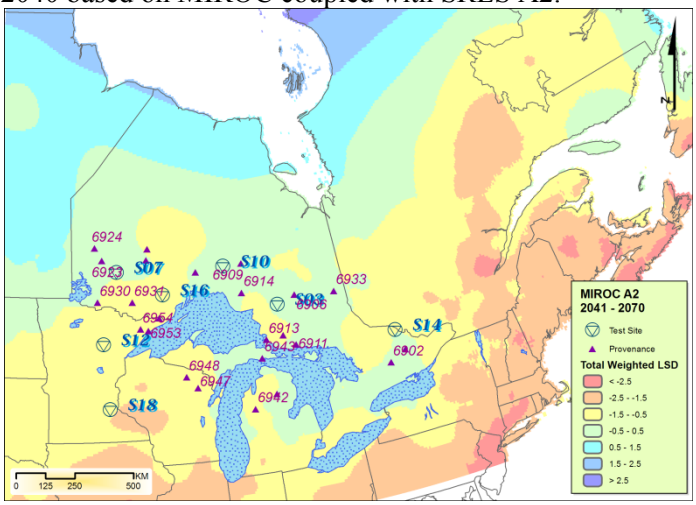

Figure. 33. Predicted total weighted LSD during 2041 - 2070 based on MIROC coupled with SRES A2.

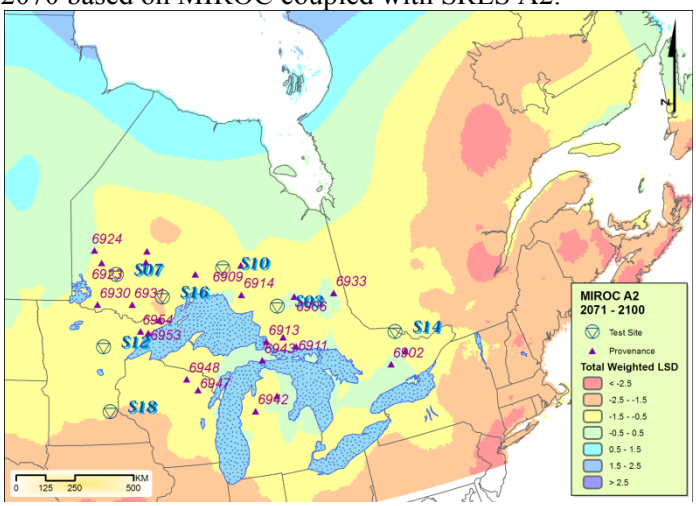

Figure. 34. Predicted total weighted LSD during 2071 - 2100 based on MIROC coupled with SRES A2.

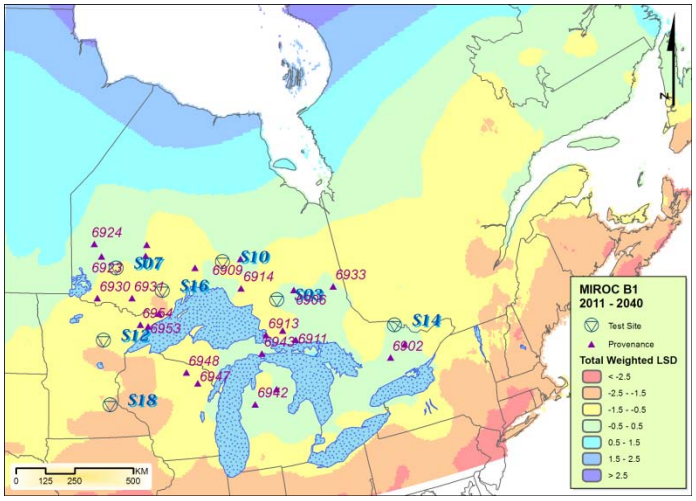

Figure. 35. Predicted total weighted LSD during 2011 - 2040 based on MIROC coupled with SRES B1.

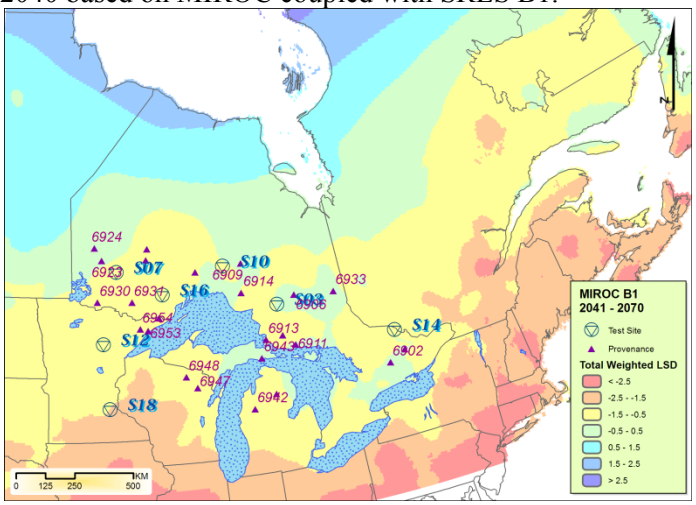

Figure. 36. Predicted total weighted LSD during 2041 - 2070 based on MIROC coupled with SRES B1.

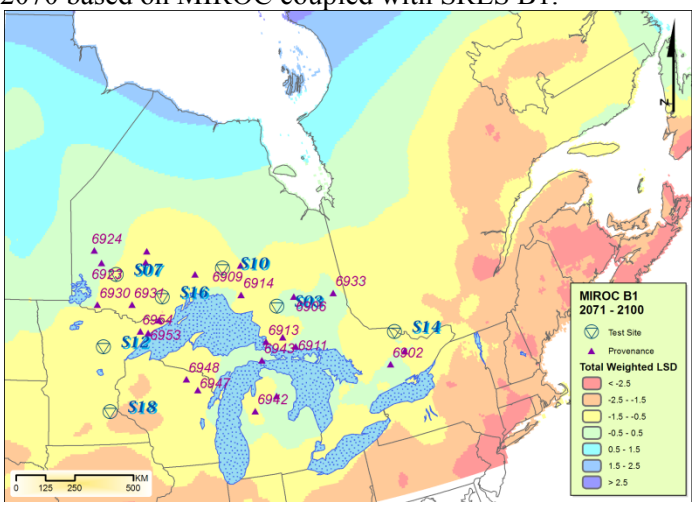

Figure. 37. Predicted total weighted LSD during 2071 - 2100 based on MIROC coupled with SRES B1.

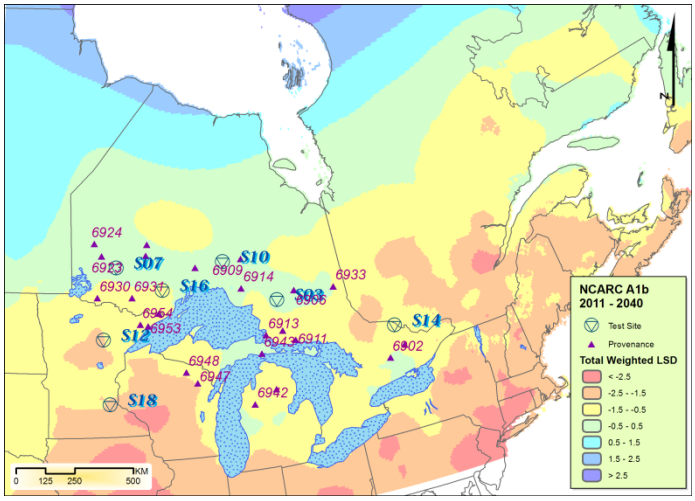

Figure. 38. Predicted total weighted LSD during 2011 - 2040 based on NCARC coupled with SRES A1b.

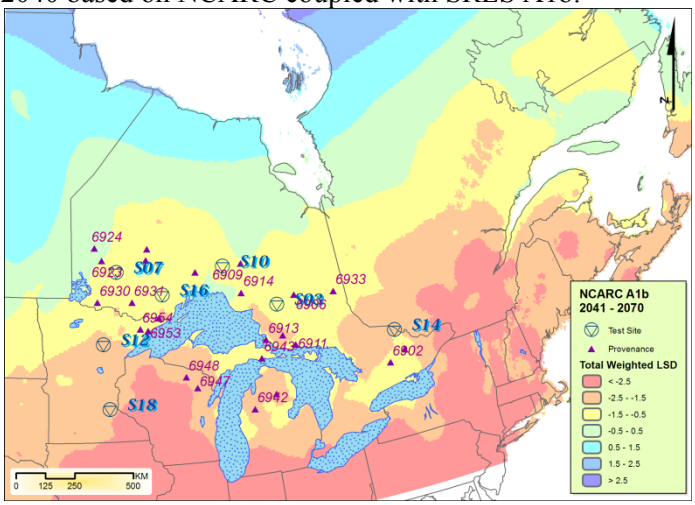

Figure. 39. Predicted total weighted LSD during 2041 - 2070 based on NCARC coupled with SRES A1b.

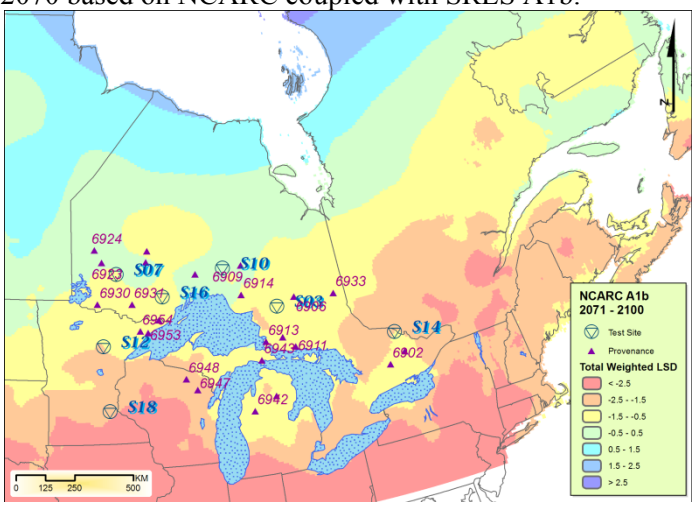

Figure. 40. Predicted total weighted LSD during 2071 - 2100 based on NCARC coupled with SRES A1b.

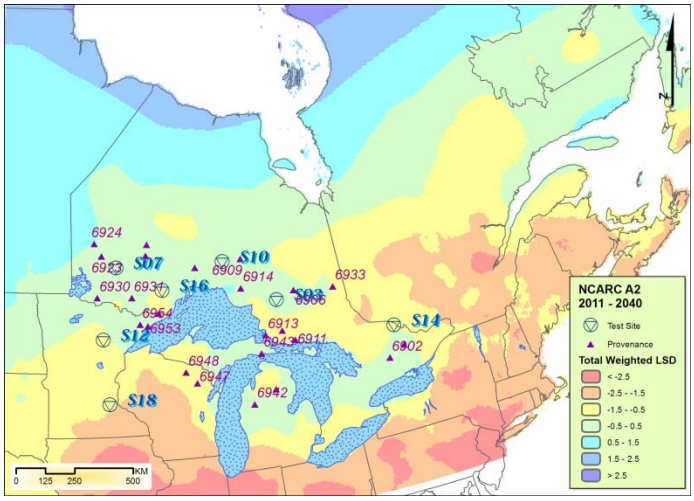

Figure. 41. Predicted total weighted LSD during 2011 - 2040 based on NCARC coupled with SRES A2.

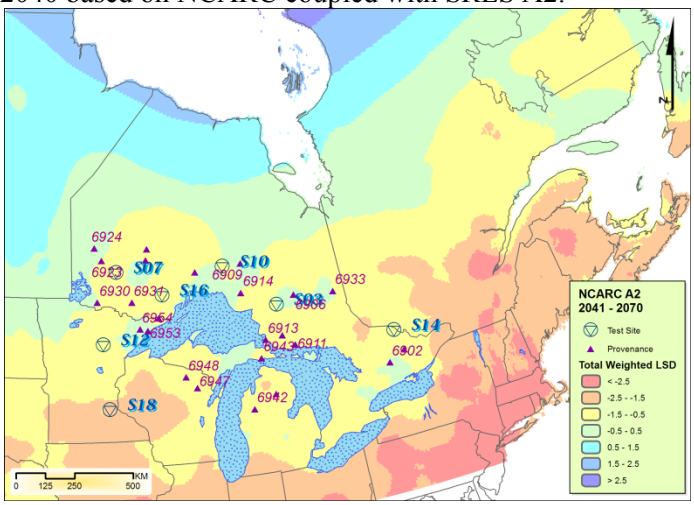

Figure. 42. Predicted total weighted LSD during 2041 - 2070 based on NCARC coupled with SRES A2.

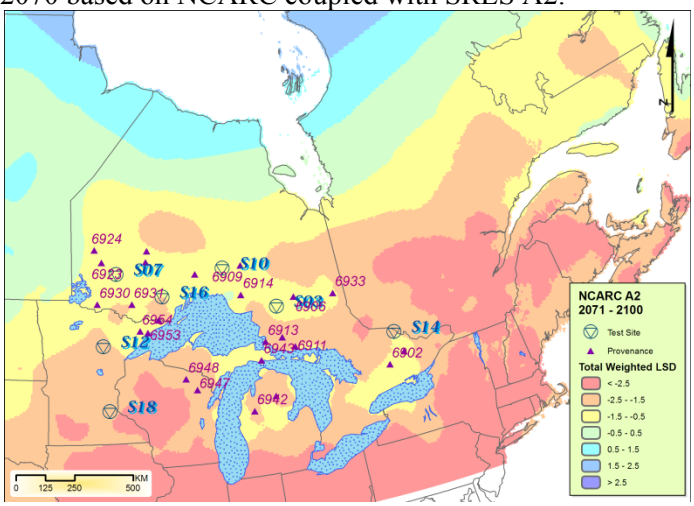

Figure. 43. Predicted total weighted LSD during 2071 - 2100 based on NCARC coupled with SRES A2.

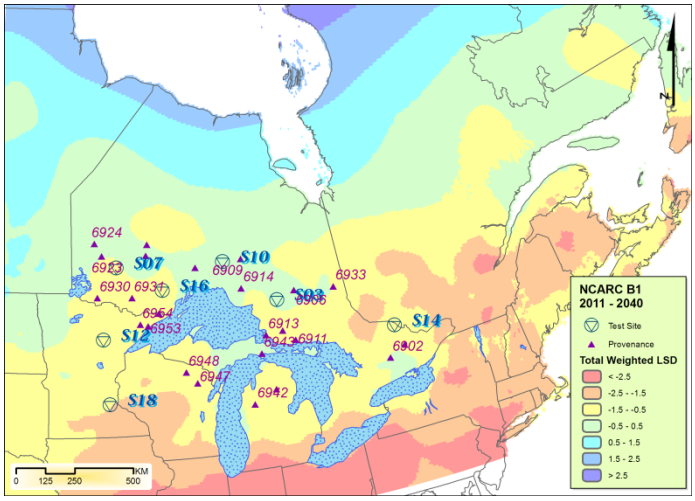

Figure. 44. Predicted total weighted LSD during 2011 - 2040 based on NCARC coupled with SRES B1.

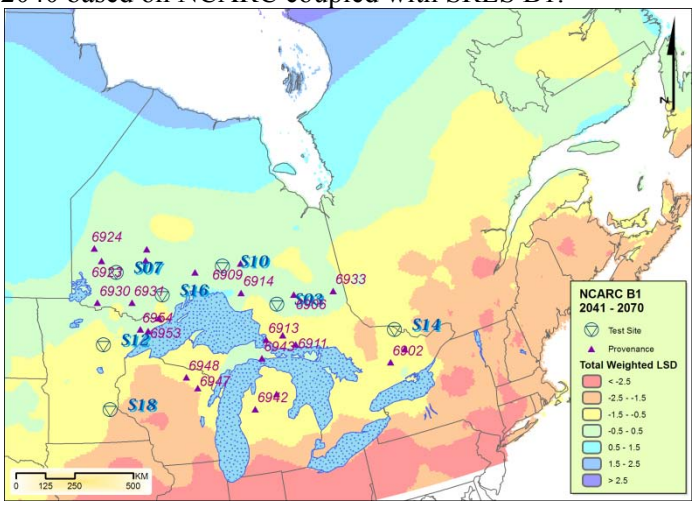

Figure. 45. Predicted total weighted LSD during 2041 - 2070 based on NCARC coupled with SRES B1.

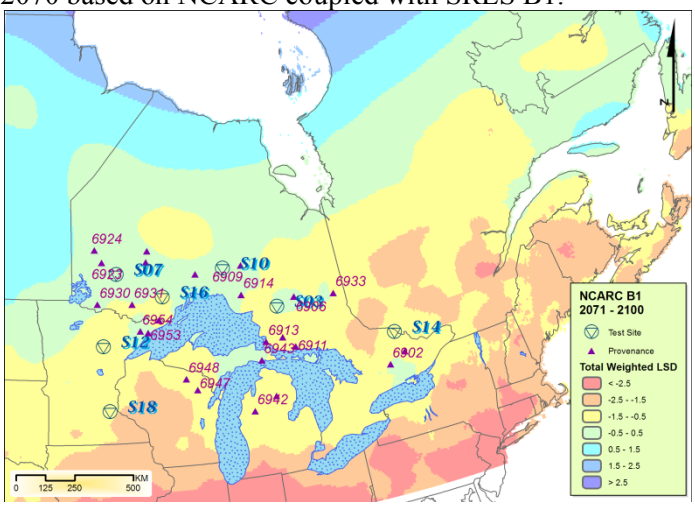

Figure. 46. Predicted total weighted LSD during 2071 - 2100 based on NCARC coupled with SRES B1.

5.2 Results from Applying the Portfolio Optimization Model

In designing the portfolio model for this problem, the objective was to allocate seed sources to sites such that each site has a reduced risk of maladaptation across multiple climatic futures. In applying the model, several questions were pursued to support this objective, namely:

1. How well can the region be covered?
2. What seed sources are most commonly used in providing this coverage?
3. How sensitive are the solutions to particular scenarios?

5.2.1 How well can the selected seed sources cover the region?

In evaluating the coverage of the region by the 7 seed sources, we wished to examine any trade-off that might exist between the total area that can be covered by these seeds versus the quality (measure of maladaptation) of this coverage.

This sensitivity analysis was performed by placing different upper bounds on the estimated value of adaptive unsuitability (LSD_{MAX}) and different upper bounds on the variance of unsuitability (VAR_{MAX}). These results are presented in figures 47 and 48.

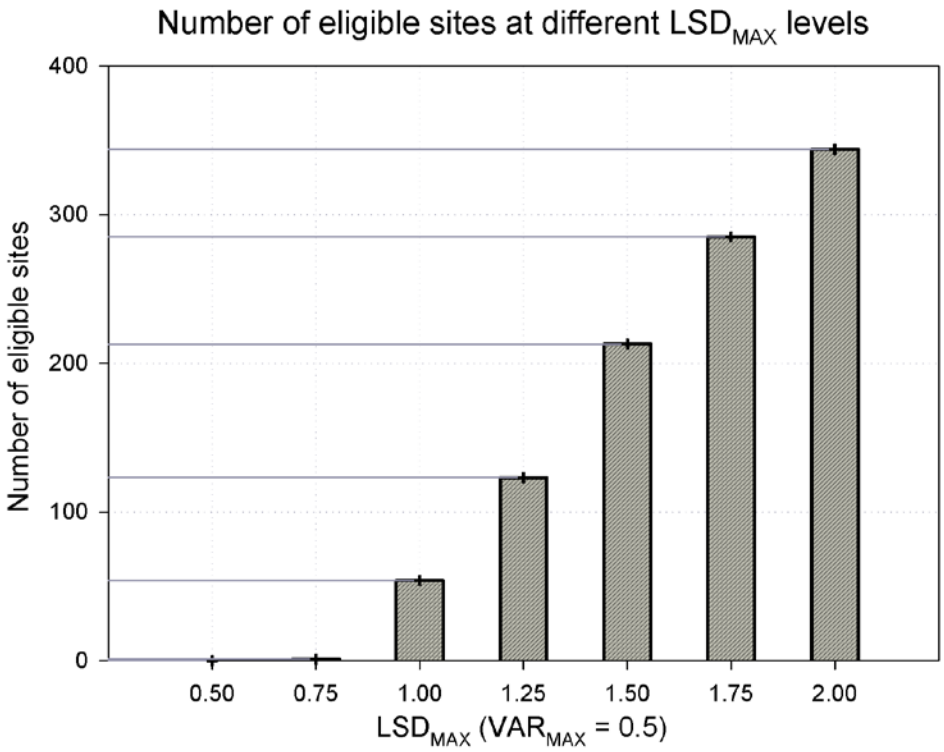

Figure. 47. Achieved number of eligible sites at different LSD_{MAX} levels.

Figure 47 shows that the coverage of the region is almost linearly sensitive to changes in the LSD thresholds.

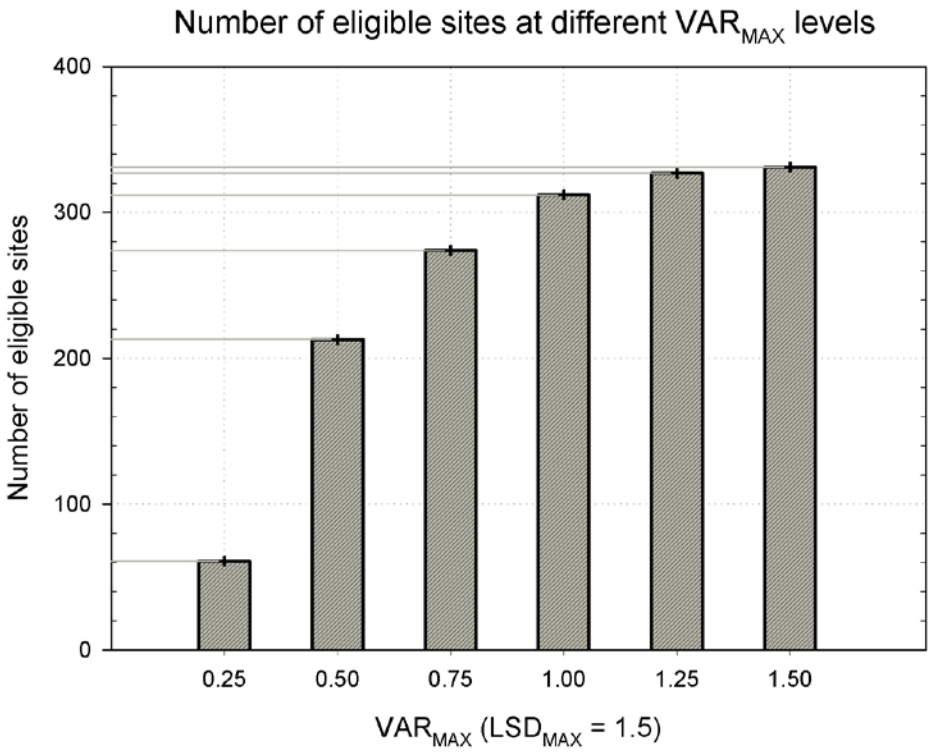

Figure. 48. Achieved number of eligible sites at different VAR_{MAX} levels.

Figure 48 indicates that diminishing returns (in terms of greater coverage of a region) sets in after relaxing the VAR_{MAX} beyond 0.5.

The spatial results of these sensitivities are mapped and illustrated in Figures 49 to 54. The observable trend from relaxing the LSD_{MAX} is a southward expansion of the eligible regions.

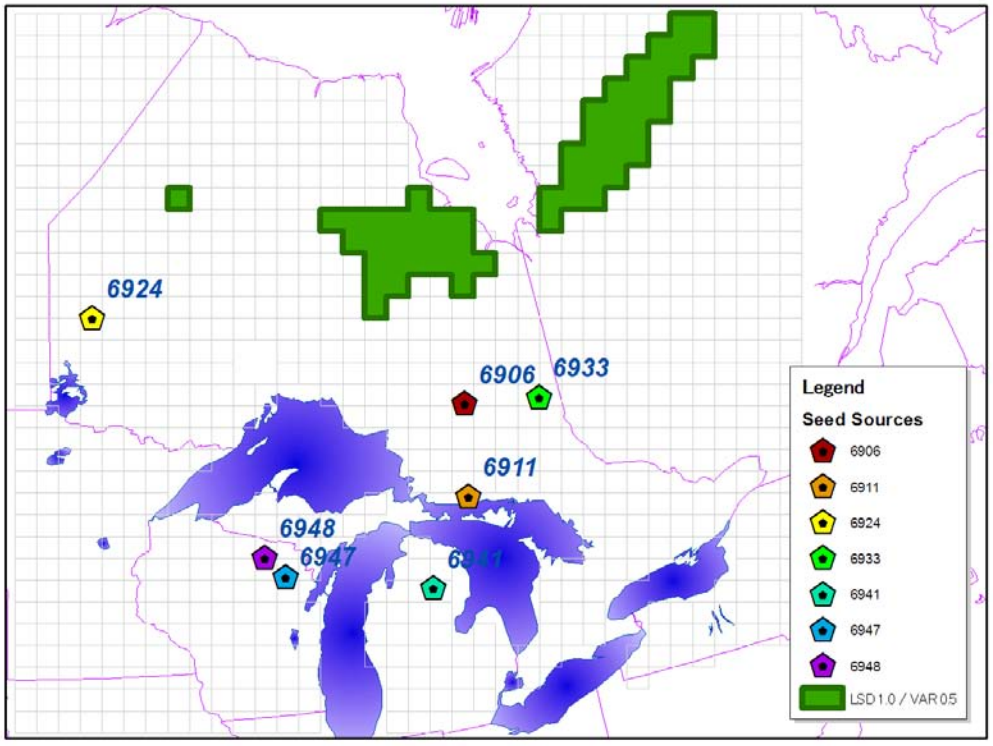

Figure. 49. Coverage of 54 sites at 1.0 LSD_{MAX} and 0.5 VAR_{MAX}.

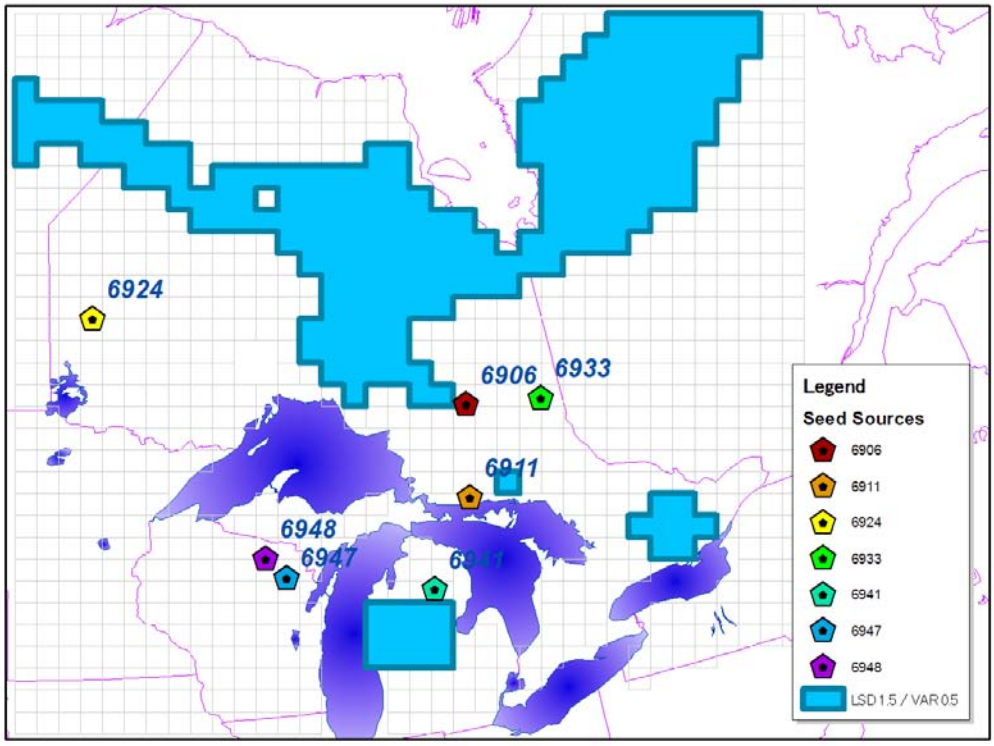

Figure. 50. Coverage of 213 sites at 1.5 LSD_{MAX} and 0.5 VAR_{MAX}.

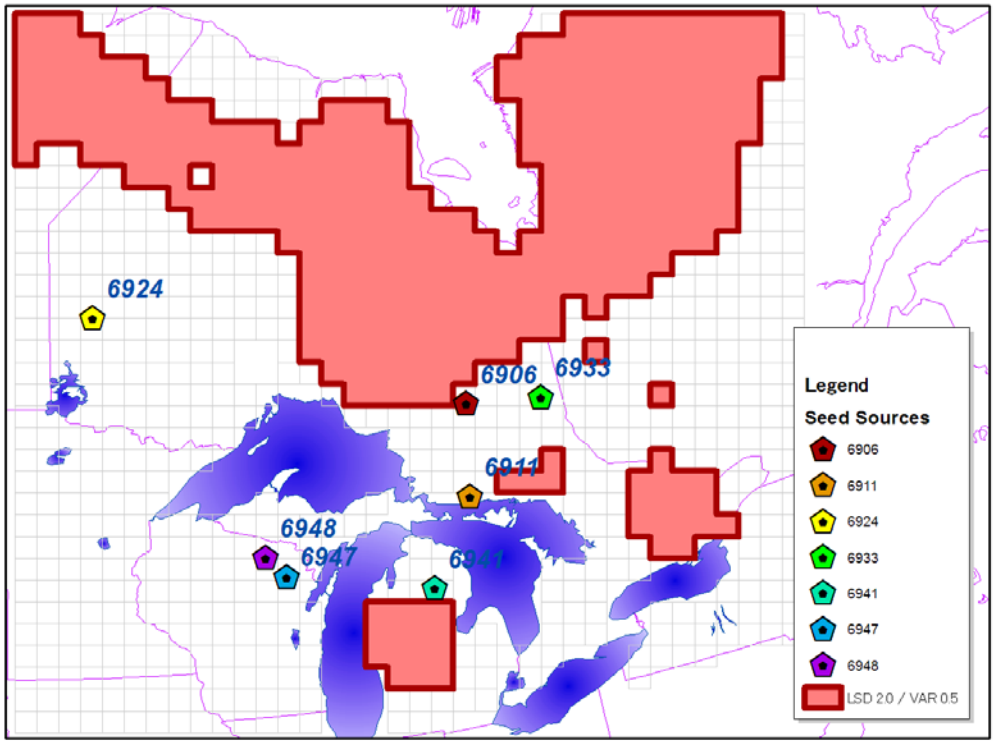

Figure. 51. Coverage of 344 sites at 2.0 LSD_{MAX} and 0.5 VAR_{MAX}.

In Figures 52 to 54, the spatial effect of relaxing the threshold on variance is illustrated. Here, we see scattered new sites appearing in the south as the constraint VAR_{MAX} is relaxed. This implies that many southern sites may be excluded due to their high risk of fluctuated climate change rather than their maladaptation level for the candidate seed sources.

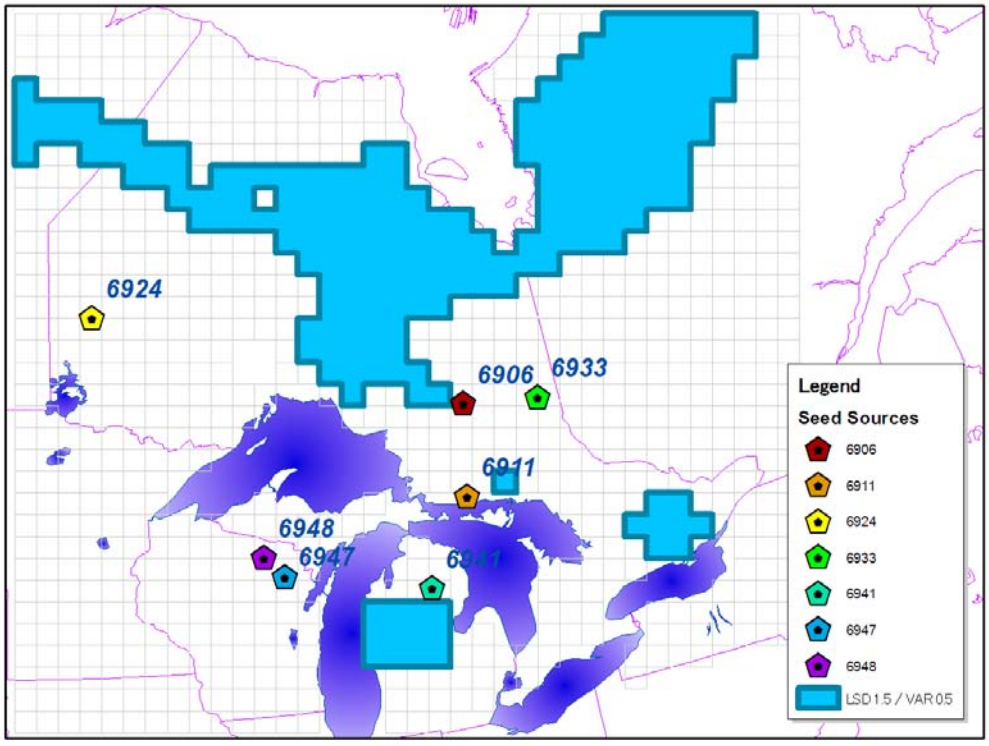

Figure. 52. Coverage of 213 sites at 1.5 LSD_{MAX} and 0.5 VAR_{MAX}.

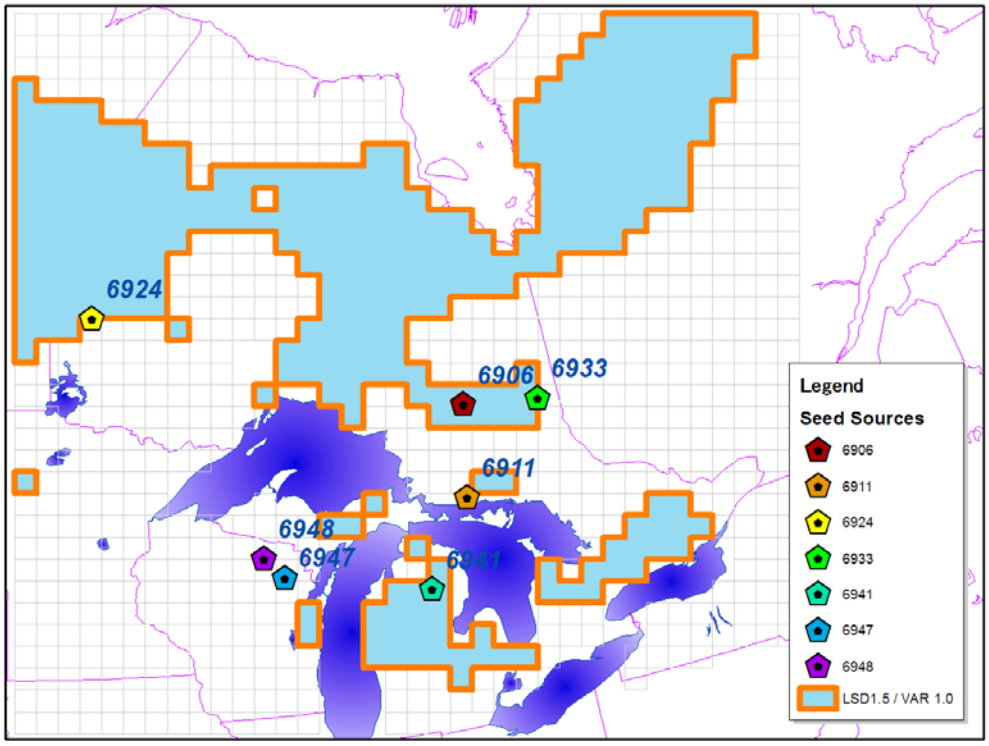

Figure. 53. Coverage of 312 sites at 1.5 LSD_{MAX} and 1.0 VAR_{MAX}.

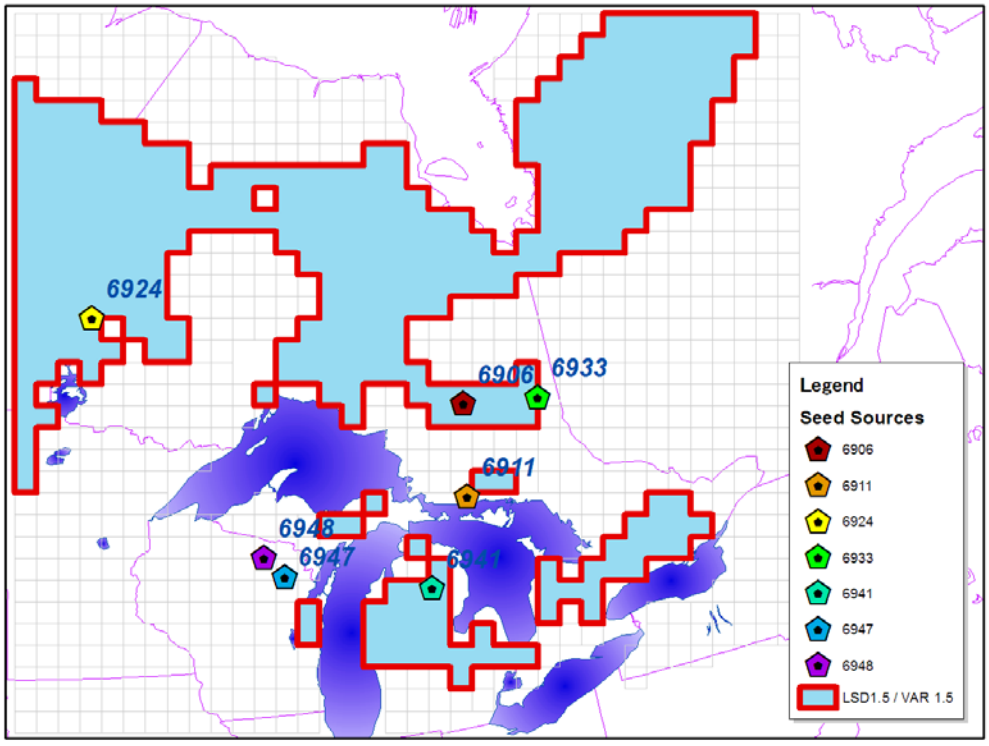

Figure. 54. Coverage of 322 sites at 1.5 LSD_{MAX} and 1.5 VAR_{MAX}.

5.2.2 Which sources are most deployed?

Figure 55 presents the utility (proportion of a seed source used in total regional coverage) of each seed source evaluated at three different LSD_{MAX} levels. The triangles represented average utilities of the candidate seed sources. Figure 55 indicates that seed source #6947 is the most valuable one which obtained 33% average coverage over all three schemes. Seed source #6924 ranked second place, which obtained 26% average coverage. Then the seed source #6948 and #6906 achieved about 13% average coverage and ranked third and fourth place respectively.

From a view of geographic locations of these seed sources, seed source #6947 is located in the south of the study area, while seed source #6924 is located in the north of the study area. These two seed sources mainly occupied the southern and northern

sites in the eligible transfer region. The other sources were mixed and filled in the gaps between the north and south.

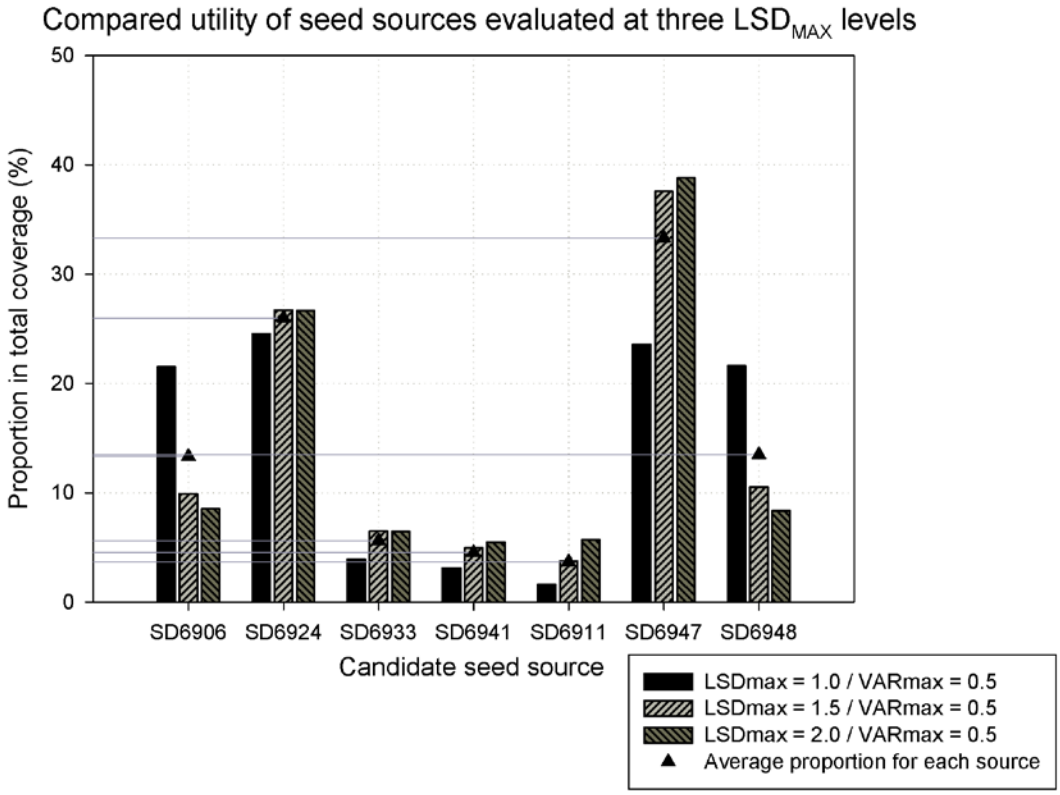

Figure. 55. Compared utility of seed sources evaluated at three LSD_{MAX} levels

Figures 56 and 57 (below) illustrate the orchard seed sources expected to be adapted in each tile of the study area in future climates. In Figure 56, the solution is constrained by an LSD_{MAX} of 1.0 and in Figure 57 the solution is constrained by an LSD_{MAX} of 1.5.

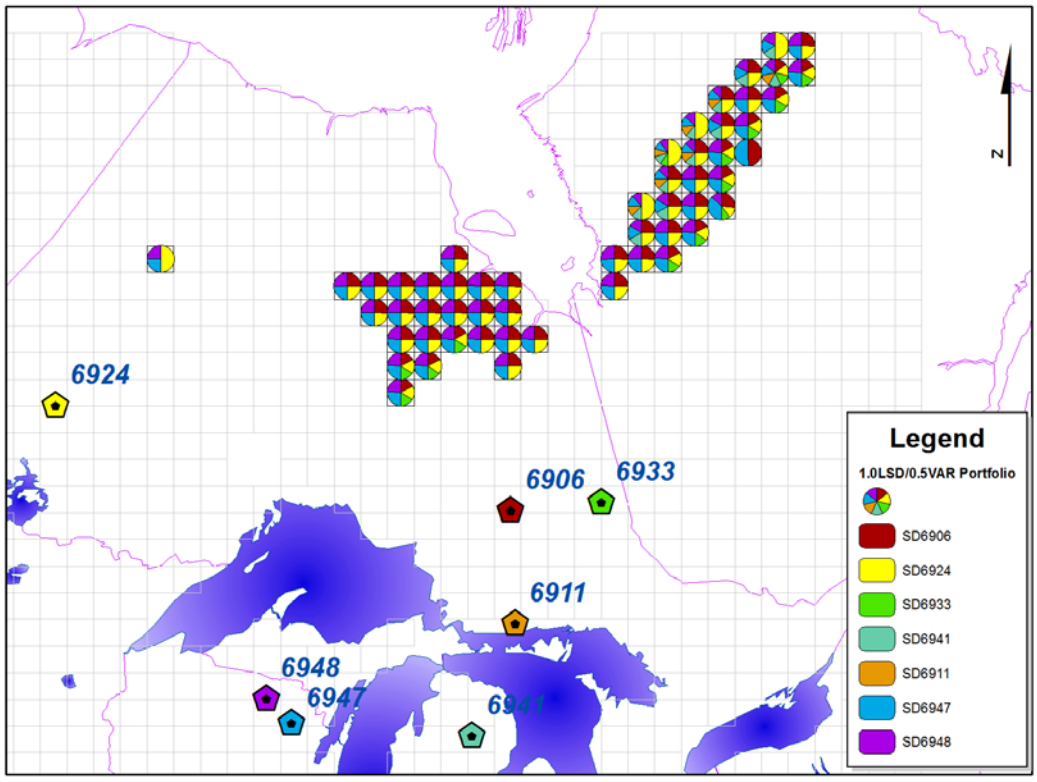

Figure. 56. The seed source portfolios of the optimal solution solved at 1.0 LSD_{MAX} and 0.5 VAR_{MAX}.

The northern and southern edges of the expanded area in Figure 57 were mainly occupied by seed sources #6924 and #6947. The center of the belt in Figure 57 is comprised of seed similar to those found in Figure 56 (i.e., by sources #6906, #6924, #6947 and #6948). The expanded region in the south (in Figure 57) is mainly occupied by the southern source (#6947) and a tiny addition of other sources.

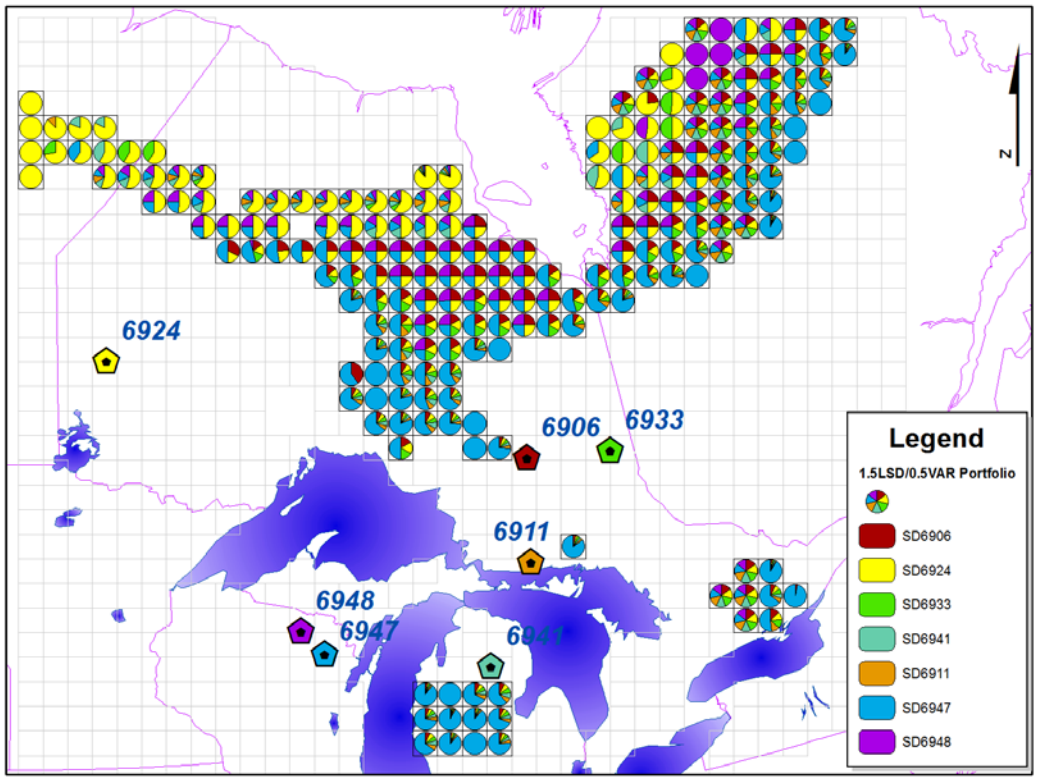

Figure. 57. Seed source portfolios of the optimal solution solved at 1.5 LSD_{MAX} and 0.5 VAR_{MAX}.

5.2.3 How acute is the trade-off between risk versus expected return?

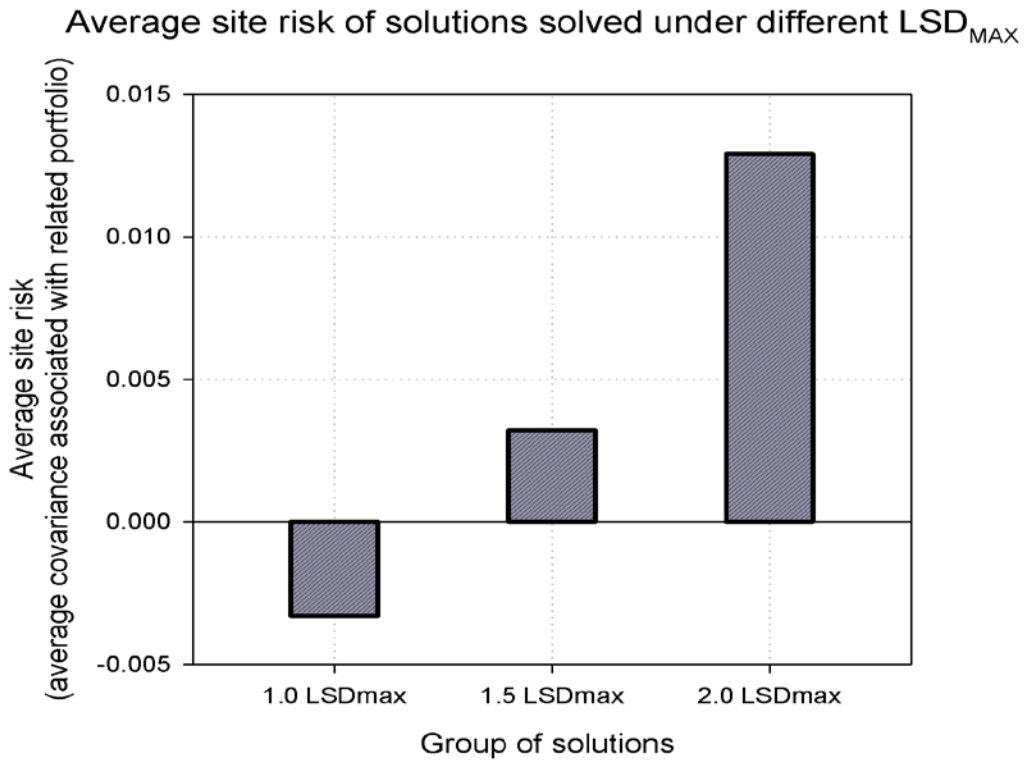

Figure. 58. Average absolute site risk of the optimal solutions solved under different LSD_{MAX} levels.

The sensitivity analysis was designed to explore whether the optimal solutions solved by the portfolio model varied greatly under different CO_2 emission scenarios applied. The three coupled emission scenarios applied in this study represent high, medium, or low CO_2 evolution in a plausible future.

Figures 59 to 61 presented eligible transfer regions and related portfolio of seed sources at each site, based on the solutions solved using different CO_2 emission scenarios. In these three trials, the threshold of LSD_{MAX} was fixed at 1.5, and the VAR_{MAX} was fixed at 0.5. There were 251 sites eligible for transfer under high CO_2 emission scenario (Figure 59), 391 sites eligible for transfer under medium CO_2

emission scenario (Figure 60), and 425 sites legible for transfer under low CO_2 emission scenario (Figure 61). Hence, our results indicate that high CO_2 emission scenarios will result in a significant decline in the eligible area for transferring the candidate sources. In addition, the high CO_2 emission scenarios prevent large southern area from being used.

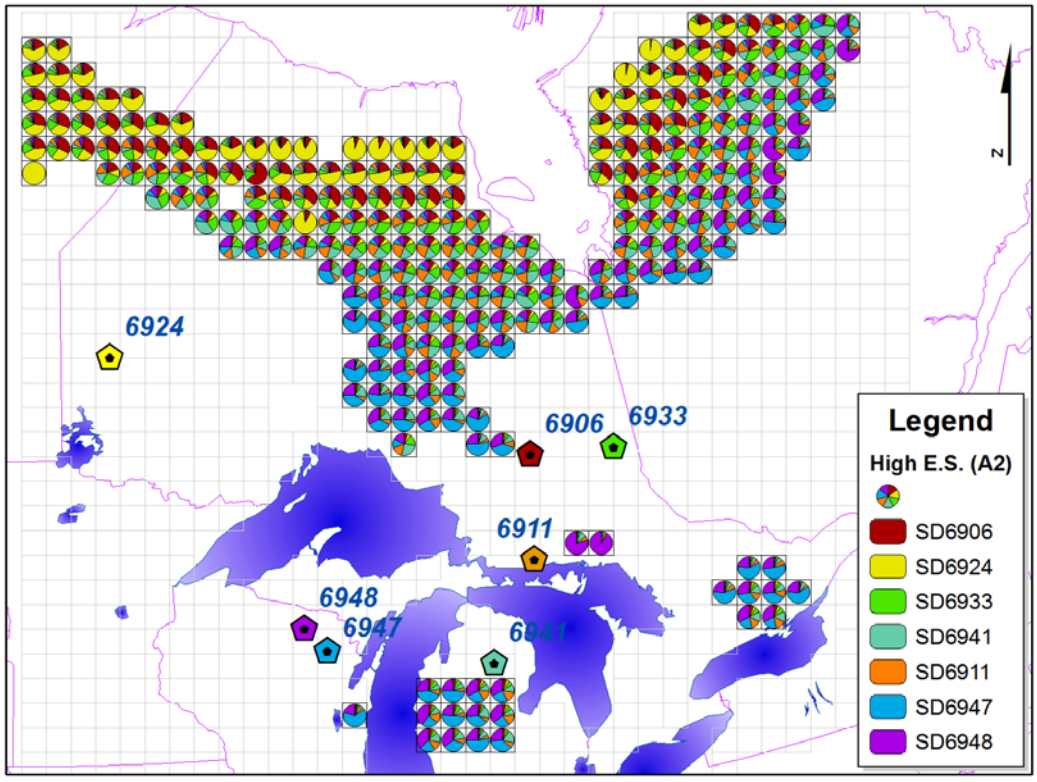

Figure. 59. Seed sources portfolios of the optimal solution solved based on high CO_2 emission scenarios.

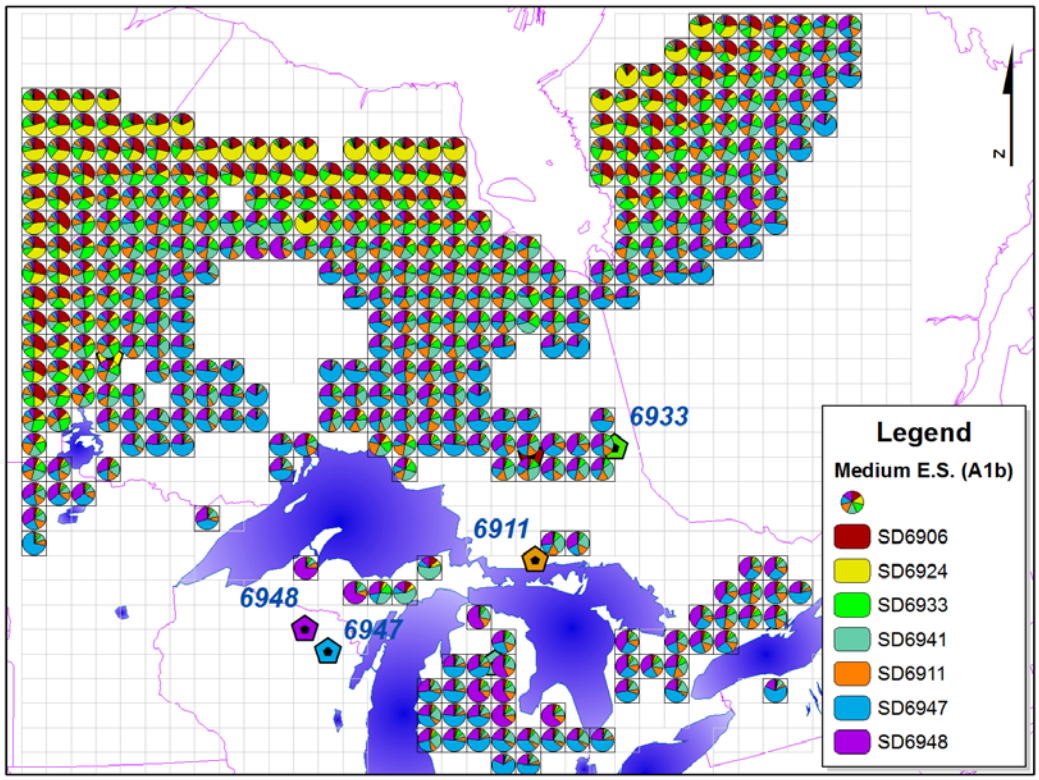

Figure. 60. Seed sources portfolios of the optimal solution solved based on medium CO2 emission scenarios.

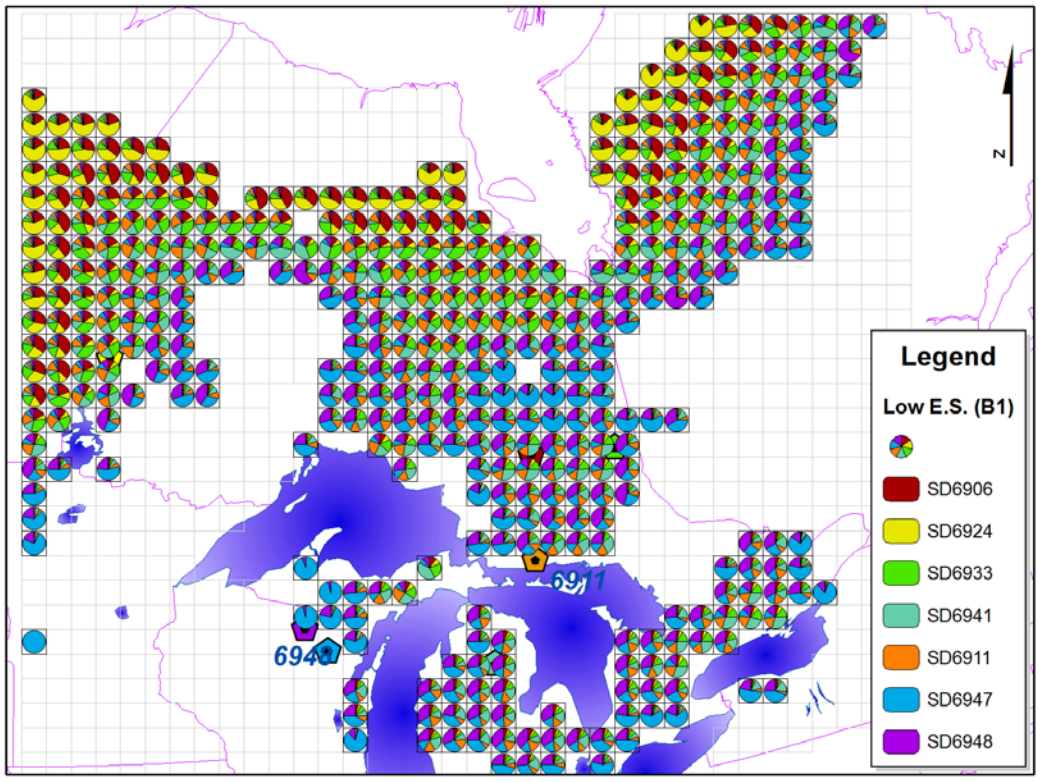

Figure. 61. Seed sources portfolios of the optimal solution solved based on low CO2 emission scenarios.

The portfolio of seed sources at each seven site in these three solutions illustrates the same trend that southern seed sources are optimally used in the southern part of the eligible region while northern seed sources are optimally used in the northern area.

Figure 62 summarises the utility of seven seed sources in the solutions based on these different emission scenarios. By comparing the utility within each source, northern seed sources (e.g., #6906 and #6924) are used more frequently when high CO_2 emissions are predicted. When low CO_2 emissions are predicted, the southern seed sources (e.g., #6947 and #6948) will be used more. The triangles represented the average utility of the seed source over different CO_2 emission scenarios. Our conclusion is that

the utilities of these candidate seed sources are not sensitive to various CO_2 emission scenarios.

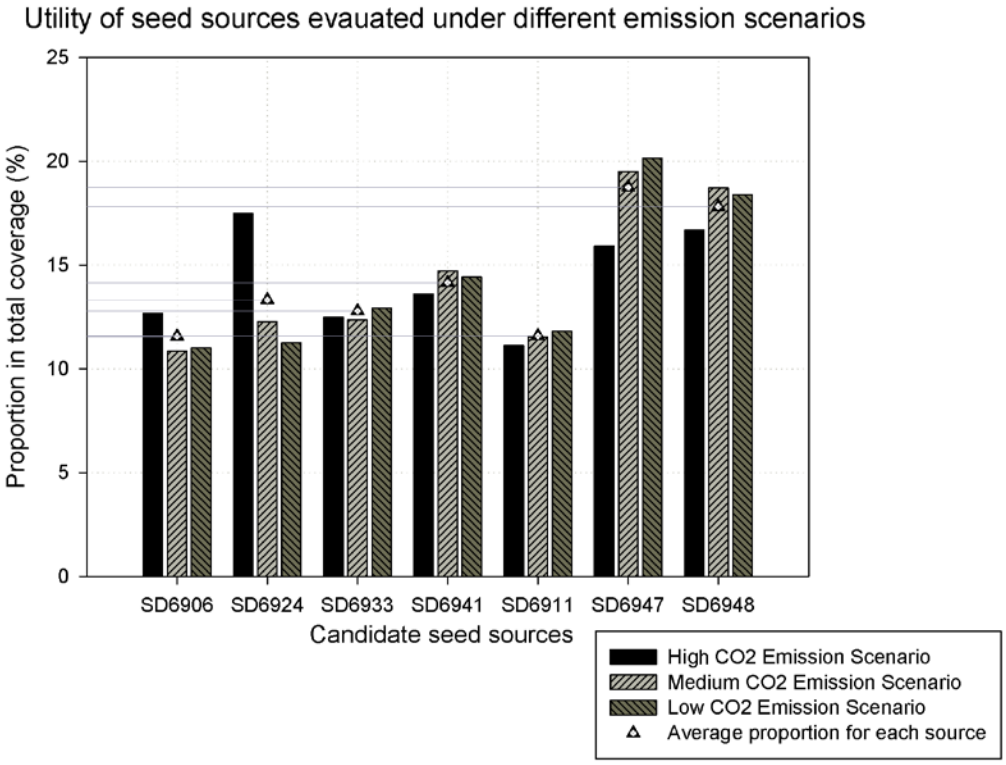

Figure. 62. Seed sources utilities evaluated under different CO_2 emission scenarios.

6. Discussion

In this study, we applied the climate change impact model and the portfolio model to evaluate the deployment schemes of seed transfer with minimized risk resulting from uncertain future climates. Based on the predictions of 4 climate models, coupled with 3 emission scenarios, the climate-change impact model revealed that there were multiple zones where provenances might best be adapted. Hence, the portfolio model was applied to minimize the risk of seeds deployed in response to multiple plausible future climates. Analysis of the portfolio optimization indicated that the coverage of eligible region for receiving seed sources is linearly sensitive to changes in the LSD thresholds. By comparing the utilities of applied candidate seed sources, two seed sources (#6924 and #6947) are used more than others. Sensitivity analysis on solution response to different CO_2 emission scenarios revealed that the solution is most sensitive to higher CO_2 emission scenarios. In addition, the utility of candidate seed sources is stable under various assumptions of CO_2 emission scenarios. In this discussion, we will address three points that concern the extension of this modeling prototype toward practical implementation.

First, in introducing this problem, we expressed concern that a decision modeling approach is needed to guide seed movement from orchards under climate change. In order to address the operational feasibility of our modeling approach, we now discuss the assumptions underlying how one defines climatic envelopes for orchard seed sources, given that the seeds it produces are from different provenances themselves, and thereby represent slightly different climatic envelopes.

In our case study, we used only one location of provenance to define the climatic envelope of each orchard seedlot. This was a simplification we indulged in because the case study was not intended for any practical application; but only to illustrate and evaluate the potential usefulness of this modeling prototype. But if we were to apply this model to develop real breeding zones, how would we define the climatic envelopes of the orchard seeds? We suggest two plausible approaches. The first, and less precise approach, would be to average the climate values of the entire breeding zone in defining the climatic envelope. The second, and more precise approach, would be to average climate variables from only those sites where parental material was gathered in producing the seed to be deployed.

A second concern regarding the practical implementation of this modeling approach concerns problem-size. What practical problems might this modeling approach confront if the case study contained 200 seed sources instead of seven?

One potential problem arising from increased problem size would be computational; i.e., would it take significantly longer to generate optimal solutions if the problem were much larger? We suspect not; for although the objective function is quadratic, the decision variables are not binary. Computational feasibility would only become a concern with increased problem size if binary decision variables were added to the formulation.

A second possible problem with increased problem size arises from implementation; i.e., how feasible would it be to collect and plant portfolios of 150 different seed sources? Clearly, an upper bound on the number of seeds that one can

practically include in a portfolio must be addressed at some point; but, in the current formulation of the optimization model, no such upper bound exists. A formulation for such a constraint would be:

Let Z_{ij} = 1 if site i receives seed from source j, 0 otherwise.

[8] $X_{ij} \leq Z_{ij}$ $for\ each\ i\ and\ j$

[9] $X_{ij} \geq .10Z_{ij}$ $for\ each\ i\ and\ j$

[10] $Z_{ij}\ is\ binary$

In the above formulation, equation [8] ensures that Z_{ij} equals 1 if site *i* receives seed from source *j*. Equation [9], ensures that, if source *j* is to be a part of the portfolio at site *j*, then it must comprise at least 10% of that portfolio. In this way, an upper bound on the number of seed sources to be used in a given portfolio can be constrained. Unfortunately, equation [10] requires that each Z_{ij} variable be binary. Hence, in very large problems (i.e., problems with more than 10,000 binary decision variables) computational infeasibility becomes a possible problem. Should computational infeasibility occur, one would be forced to revert to the formulation without an upper bound on the number of seed sources, and simply select the top seed sources found in the solution.

A final question that needs to be addressed regarding the practical implementation of this model is the interpretation of the risk values. For example, when examining the trade-off between risk versus return, one can form a biological interpretation of the return value (maladaptation, measured in LSD's) based on the

empirical data of the provenance tests; but, it is not immediately clear how one should understand, biologically, any given covariance value. The general idea, of course, is that a portfolio of seeds with a high negative covariance tends to perform well across all scenarios; but there is no numerical necessity that this be the case. Hence, we suggest that, before any practical implementation of a solution occur, the decision-maker check and see how well *each* future climatic scenario is covered by the solution.

7. Conclusion

The objective of this research was to develop a modeling framework for the deployment of seed, from multiple improved seed sources, in an environment of uncertain climatic change. A robust solution to this problem was sought by using portfolio optimization. In the case study, we showed how this can be applied at the regional scale.

We conclude that this framework provides a useful advance in supporting decisions on the seed deployment problem in an environment of uncertain climate—in general; and in particular, it is useful in evaluating:

a) which areas can be covered with the least risk;

b) which seed sources will be in highest demand; and

c) how sensitive the solutions are to assumptions on future CO_2 emissions.

Literature Cited

Aitken, S.N., Yeaman, S., Holliday, J.A., Wang, T.L., and Curtis-McLane, S. 2008. Adaptation, migration or extirpation: climate change outcomes for tree populations. Evolutionary Applications. 1(1): 95-111.

Bower and Aitken. 2008. Ecological genetics and seed transfer guidelines for Pinus albicaulis (Pinaceae). American Journal of Botany. 95:66-76.

Campbell, R.K. 1986. Mapped genetic-variation of douglas-fir to guide seed transfer in southwest Oregon. Silvae Genetica. 35(2-3): 85-96.

Campbell, R.K. and Sorensen, F.C. 1978. Effect of test environment on expression of clines and on delimitation of seed zones in douglas-fir. Theoretical and Applied Genetics. 51(5): 233-246.

Campbell, R.K. and Sugano, A.I. 1987. Seed zones and breeding zones for sugar pine in southwestern Oregon. USDA Forest Service Pacific Northwest Research Station, Research Paper (383): U1-U18.

Chandha, K.M. and Patnik, S.S. 1990. Country report, India. In Proc. Asia Pacific regional workshop on tree breeding and propagation.

Covey, C., AchutaRao, K.M., Cubasch, U., Jones, P., Lambert, S.J., Mann, M.E., Phillips, T.J., and Taylor, K.E. 2003. An overview of results from the Coupled Model Intercomparison Project. Global and Planetary Change. 37(1-2): 103-133.

Crowe, K.A. and Parker, W.H. 2005. Provisional breeding zone determination modeled as a maximal covering location problem. Canadian Journal of Forest Research. 35(5): 1173-1182.

Crowe, K.A. and Parker, W.H. 2008. Using portfolio theory to guide reforestation and restoration under climate change scenarios. Climatic Change. 89(3-4): 355-370.

Daniels, J.D. 1984. Role of Tree Improvement in Intensive Forest Management. Forest Ecology and Management. 8(3-4): 161-195.

Davis, M.B. and Shaw, R.G. 2001. Range shifts and adaptive responses to Quaternary climate change. Science. 292(5517): 673-679.

Davis, M.B., Shaw, R.G., and Etterson, J.R. 2005. Evolutionary responses to changing climate. Ecology. 86(7): 1704-1714.

Ehhalt, D., Prather, M., Dentener, F., Derwent, R., Dlugokencky, E., Holland, E., Isaksen, I., Katima, J., Kirchhoff, V., and Matson, P. 2001. Atmospheric Chemistry and Greenhouse Gases. In Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovenmental Panel on Climate Change. Cambridge University Press.

Figge, F. 2004. Bio-folio: applying portfolio theory to biodiversity. Biodiversity and Conservation. 13(4): 827-849.

Grinnell, J. 1917. Field tests of theories concerning distributinal control. The American Naturalist, Vol. 51, No. 602, pp. 115-128.

Grinnell, J. 1924. Geography and evolution. Ecology, Vol. 5, No. 3, pp. 225-229.

Hijmans, R.j., Cameron, S.E., Parra, J.L., Jones, P.G., and Jarvis, A. 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology. 25(15): 1965-1978.

Hulme, M. 2005. Recent Climate Trends. *In* Climate Change and Biodiversity. Yale University Press, New Haven & London. pp. 31-40.

Huntley, B., Berry, P.M., Cramer, W., and McDonald, A.P. 1995. Modelling present and potential future ranges of some European higher plants using climate response surfaces. Journal of Biogeography. 22(6): 967-1001.

Hutchinson, G.E. 1957. Concluding remarks. Cold spring harbor symposium on quantitative biology. 22(2). pp. 415-457.

IPCC 2007a. Climate Change 2007: The Phsical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M.Tignor and H.L. Miller (eds.)]. Cambridge University Press.

IPCC 2007b. Summary for Policymakers. In: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press.

IPCC 2007c. Technical Summary. In: Climate Change 2007: Mitigation. Contribution of Working Group III to the fourth Assessment Reportof the Intergovernmental Panel on Climate Change [B. Metz, O.R. Davidson, P.R. Bosch, R. Dave, L.A. Meyer (eds)]. Cambridge University Press.

IPCC 2008. Climate Change 2007: Synthesis Report. Contribution of Working Group I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Intergovernmental Panel on Climate Change.

Iverson, L.R. and Prasad, A.M. 1998. Predicting abundance of 80 tree species following climate change in the eastern United States. Ecological Monographs. 68(4): 465-485.

Iverson, L.R., Prasad, A., and Schwartz, M. W. 1999. Modeling potential future individual tree-species distributions in the eastern United States under a climate change scenario: a case study with *Pinus virginiana*. Ecological Modelling. 115(1): 77-93.

Iverson, L.R., Prasad, A.M., Matthews, S.N., and Peters, M. 2008. Estimating potential habitat for 134 eastern US tree species under six climate scenarios. Forest Ecology and Management. 254(3): 390-406.

Jansen, E., Overpeck, J., Briffa, K.R., Duplessy, J.-C., Joos, F., Masson-Delmotte, V., Olago, D., Otto-Bliesner, B., Peltier, W.R., Rahmstorf, S., Ramesh, R., Raynaud, D., Rind, D., Solomina, O., Villalba, R., and Zhang, D. 2007.

Palaeoclimate. In: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Miller (eds.)]. Cambridge University Press.

Jones, P.D., New, M., Parker, D.E., Martin, S., and Rigor, I.G. 1999. Surface air temperature and its changes over the past 150 years. Reviews of Geophysics. 37(2): 173-199.

Jump, A.S. and Penuelas, J. 2005. Running to stand still: adaptation and the response of plants to rapid climate change. Ecology Letters. 8(9): 1010-1020.

Karl, T.R. and Trenberth, K.E. 2005. What Is Climate Change? *In* Climate Change and Biodiversity. Yale University Press, New Haven & London. pp. 15-28.

Kimmins, J.P. 2004. Forest Ecology: A foundation for sustainable forest management and environmental ethics. Third Edition ed. Pearson Education, Inc., Upper Saddle River, New Jersey.

Kullman, L. 1996. Rise and demise of cold-climate *Picea abies* forest in Sweden. New Phytologist 134(2): 243-256.

Kullman, L. 2001. 20th century climate warming and tree-limit rise in the southern Scandes of Sweden. Ambio 30(2): 72-80.

Kullman, L. 2003. Recent reversal of Neoglacial climate cooling trend in the Swedish Scandes as evidenced by mountain birch tree-limit rise. Global and Planetary Change. 36(1-2): 77-88.

Lambert, S.J. and Boer, G.J. 2001. CMIP1 evaluation and intercomparison of coupled climate models. Climate Dynamics 17(2-3): 83-106.

Lempert, R., Nakicenovic, N., Sarewitz, D., and Schlesinger, M. 2004. Characterizing climate-change uncertainties for decision-makers - An editorial essay. Climatic Change 65(1-2): 1-9.

Lesser, M.R. and Parker, W.H. 2006. Comparison of canonical correlation and regression based focal point seed zones of white spruce. Canadian Journal of Forest Research. 36(6): 1572-1586.

Li, B., McKeand, S., and Weir, R. 1999. Tree improvement and sustainable forestry - impact of two cycles of loblolly pine breeding in the U.S.A. Forest Genetics. 6(4). pp. 229-234.

Markowitz, H.M. 1952. Portfolio selection. The Journal of Finance. 7(1): 77-91.

Markowitz, H.M. 1992. A Theory of Practice - A Citation-Classic Commentary on Portfolio Selection, Efficient Diversification of Investments by Markowitz,H.M. Current Contents/Arts & Humanities. (19): 18.

McKeand, S. and Kurinobu, S. 1998. Japanese tree improvement and forest genetics. Journal of Forestry. 96(4): 12-17.

McKenney, D., Papadopol, P., Siltanen, M., and Lawrence, K. 2009. High-Resolution climate change scenarios for North America. Canadian Forest Service, Great Lakes Forestry Centre.

Mckenney, D.W., Pedlar, J.H., Lawrence, K., Campbell, K., and Hutchinson, M.F. 2007. Potential impacts of climate change on the distribution of North American trees. Bioscience 57(11): 939-948.

Meehl, G.A., Stocker, T.F., Collins, W.D., Friendlingstein, P., Gaye, A.T., Gregory, J.M., Kitoh, A., Knutti, R., Murphy, J.M., Noda, A., Raper, S.C.B., Watterson, I.G., Weaver, A.J., and Zhao, Z.-C. 2007. Global Climate Projections. In: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M.Tignor and H.L. Miller (eds.)]. Cambridge University Press.

Nakicenovic, N. and Swart, R. 2000. Special Report on Emissions Scenarios. A Special Report of Working Group III of the Intergovernmental Panel on Climate Change. Cambridge University Press.

Overpeck, J., Cole, J., and Bartlein, P. 2005. A "Paleoperspective" on Climate variability and Change. *In* Climate Change and Biodiversity. Yale University Press, New Haven & London. pp. 91-108.

Parker, W.H. 1992. Focal Point Seed Zones - Site-Specific Seed Zone Delineation Using Geographic Information-Systems. Canadian Journal of Forest Research. 22(2): 267-271.

Parker, W.H. and vanNiejenhuis, A. 1996. Regression-based focal point seed zones for *Picea mariana* from northwestern Ontario. Canadian Journal of Botany. 74(8): 1227-1235.

Parker, W.H., vanNiejenhuis, A., and Charrette, P. 1994. Adaptive Variation in *Picea-Mariana* from Northwestern Ontario Determined by Short-Term Common Environment Tests. Canadian Journal of Forest Research. 24(8): 1653-1661.

Parry, M.L., Canziani, O.F., Palutikof, J.P., and Co-authors 2007. Technical Summary. Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press.

Pearson, R.G. and Dawson, T.P. 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? Global Ecology and Biogeography. 12(5): 361-371.

Peterson, A.T., Ortega-Huerta, M.A., Bartley, J., Sanchez-Cordero, V., Soberon, J., Buddemeier, R.H., and Stockwell, D.R.B. 2002. Future projections for Mexican faunas under global climate change scenarios. Nature. 416(6881): 626-629.

Peterson, A.T., Soberon, J., and Sanchez-Cordero, V. 1999. Conservatism of ecological niches in evolutionary time. Science. 285(5431): 1265-1267.

Peterson, A.T., Tian, H., Martinez-Meyer, E., Soberon, J., Sanchez-Cordero, V., and Huntley, B. 2005. Modeling Distributional Shifts of Individual Species and Biomes. *In* Climate Change and Biodiversity. Yale University Press, New Haven & London. pp. 211-321.

Pijut, P.M., Woeste, K.E., Veugadesan, G., and Michler, C.H. 2007. Technological advances in temperate hardwood tree improvement including breeding and molecular marker applications. In Vitro Cellular & Developmental Biology-Plant 43(4): 283-303.

Randall, D.A., Wood, R.A., Bony, S., Colman, R., Fichefet, T., Fyfe, J., Kattsov, V., Pitman, A., Shukla, J., Srinivasan, J., Stouffer, R.J., Sumi, A., and Lovejoe, T.E. 2007. Cilmate Models and Their Evaluation. In: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M.Tignor and H.L. Miller (eds.)]. Cambridge University Press.

Raper, S.C.B. and Giorgi, F. 2005. Climate Change Projections and Models. *In* Climate Change and Biodiversity. Yale University Press, New Haven & London. pp. 199-210.

Rehfeldt, G.E. 1982. Differentiation of Larix Occidentalis Populations from the Northern Rocky Mountains. Silvae Genetica. 31(1): 13-19.

Rehfeldt, G.E. 1983a. Adaptation of *Pinus-contorta* Populations to Heterogeneous Environments in Northern Idaho. Canadian Journal of Forest Research. 13(3): 405-411.

Rehfeldt, G.E. 1983b. Ecological Adaptations in Douglas-Fir (Pseudotsuga-Menziesii Var Glauca) Populations .3. Central Idaho. Canadian Journal of Forest Research. 13(4): 626-632.

Rehfeldt, G.E., Ying, C.C., Spittlehouse, D.L., and Hamilton, D.A. 1999. Genetic responses to climate in Pinus contorta: Niche breadth, climate change, and reforestation. Ecological Monographs 69(3): 375-407.

Satoo, T. 1960. A brief report on forest tree improvement in Japan. 7th Northeastern Forest Tree Improvement Conference. Burlington,Vermont August 8-19. pp. 25-27.

Shafer, S.L., Bartlein, P.J., and Thompson, R.S. 2001. Potential changes in the distributions of western North America tree and shrub taxa under future climate scenarios. Ecosystems. 4(3): 200-215.

Shen, X.H., Zhang, D.M., Li, Y., and Zhang H.X. 2007. Temporal and Spatial Change of the Mating System Parameters in a Seed Orchard of *Pinus tabulaeformis* Carr. from http://www-genfys.slu.se/staff/dagl/Umea07/Proceedings/.

Su, X.H., Zhang, B.Y., Huang, Q.J., Huang, L.J., and Zhang X.H. 2003. Advances in tree genetic engineering in China. China Science and Technology Press, Quebec City, Canada.

Sun, S. and Hansen, J.E. 2003. Climate simulations for 1951-2050 with a coupled atmosphere-ocean model. Journal of Climate. 16(17): 2807-2826.

Sykes, M.T., Prentice, I.C., and Cramer, W. 1996. A bioclimatic model for the potential distributions of north European tree species under present and future climates. Journal of Biogeography 23(2): 203-233.

Thomson, A.M. and Parker, W.H. 2008. Boreal forest provenance tests used to predict optimal growth and response to climate change. 1. Jack pine. Canadian Journal of Forest Research. 38(1): 157-170.

Thomson, A.M., Riddell, C.L., and Parker, W.H. 2009. Boreal forest provenance tests used to predict optimal growth and response to climate change: 2. Black spruce. Canadian Journal of Forest Research. 39(1): 143-153.

Toth, F.L. and Mwandosya, M. 2001. Decision-making Frameworks. Climate Change 2001: Mitigation. Cambridge University Press.

Trenberth, K.E., Jones, P.A., Bojariu, R., Easterling, D., Klein Tank, A., Rahimzadeh, F., Renwick, J.A., Rusticucci, M., Soden, B., and Zhai, P. 2007. Observations: Surface and Atmospheric Climate Change. In: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Miller (eds.)]. Cambridge University Press.

Weisgerber, H. and Sindelar, J. 1992. IUFRO's Role in Coniferous Tree Improvement - History, Results, and Future-Trends of Research and International-Cooperation with European Larch (Larix-Decidua Mill). Silvae Genetica. 41(3): 150-161.

Appendix 1: Geographic locations of test sites

Table 6. Geographic locations of test sites in black spruce provenance study.

Site index	Province / State	Test site name	Longitude	Latitude
1	Alberta	Peace River	-116.664	56.003
2	New Brunswick	Black Brook	-67.457	47.352
3	Ontario	Chapleau	-83.431	47.964
4	Quebec	Chibougamau	-74.209	50.023
5	Nova Scotia	East Dalhousie	-64.790	44.748
6	Prince Edward Island	Dromore	-62.845	46.280
7	Ontario	Dryden	-92.484	49.923
8	New Brunswick	Fredericton	-66.389	46.782
9	Quebec	Lac Ste. Ignace	-66.304	49.017
10	Ontario	Longlac	-86.161	49.754
11	Manitoba	Mafeking	-101.396	52.698
12	Minnesota	Int'l Falls	-93.409	47.187
13	Quebec	Mont-Laurier	-75.845	46.601
14	Ontario	Petawawa	-77.400	45.974
15	Nova Scotia	Pleasant Valley	-62.697	45.109
16	Ontario	Raith	-89.877	48.936
17	Newfoundland	Roddickton	-56.204	50.969
18	Minnesota	St. Cloud	-93.190	44.730
19	Saskatchewan	Prince Albert	-104.401	54.238
20	Quebec	Valcartier	-71.531	46.859

Appendix 2: Focal point seed zone program

```
libname BScom01e
'D:\Thesis.090106\SubTopic.3.Com.3,7,10,12,14,16,18\BS+';
run;

data BScom01e.data;
infile
'D:\Thesis.090106\SubTopic.3.Com.3,7,10,12,14,16,18\InputData.csv' dlm
= ',' ;
input SITE PROV BLK TREE HGT DBH;
run;

proc sort data = BScom01e.data;
by SITE PROV;
run;

proc means data = BScom01e.data mean std n nmiss ;
title 'Seika.-.[BlackSpruce.Com.<3,7,10,12,14,16,18>].MeansProc';
by SITE ;
var HGT DBH;
run;

proc means data = BScom01e.data mean std;
title 'Seika Prov PReview';
class Prov;
var HGT DBH;
output out = BScom01e.PRovSel mean = ProvAvgHgt ProvAvgDBH Stddev =
ProvStdHgt ProStdDBH;
proc print data = BScom01e.PRovSel;
option pagesize = 1000;
run;

data BScom01e.PRovSel;
set BScom01e.PRovSel;
V = ProvAvgDBH * ProvAvgDBH  * 0.25 * 3.14 * (ProvAvgHgt*100 +300) *
0.41 ;
proc print data = BScom01e.ProvSel;
run;

proc means data = BScom01e.data mean std n nmiss noprint;
title 'Seika.-.[BlackSpruce.Com.<3,7,10,12,14,16,18>].MeansProc';
by SITE;
class PROV;
class BLK ;
var HGT DBH;
output out = BScom01e.Mean mean = AvgHgt AvgDBH Stddev = StdHgt StdDBH
n = ctObsHGT ctObsDBH nmiss = ctMissHGT ctMissDBH Skewness = SkewHgt
SkewDBH Kurtosis = KurtHgt KurtDBH;
proc print data = BScom01e.mean;
Title "Proc Means Results";
option pagesize = 1000;
run;

data BScom01e.mean;
set Bscom01e.mean;
```

```
SURR = ( ctObsHGT + ctObsDBH ) / (ctObsHGT + ctObsDBH + ctMissHGT +
ctMissDBH  );
drop ctObsHGT ctObsDBH ctMissHGT ctMissDBH;
run;

proc sort data = BScom01e.mean;
by _TYPE_ PROV SITE BLK;
proc print;
Title "Refined Means Results ^ SURR added ^";
run;

proc glm data = BScom01e.data outstat = BScom01e.TwowayANOVARESULT;
title 'Two-Way ANOVA for HGT / DBH by SITE ';
by SITE;
class PROV BLK;
model HGT DBH = PROV BLK PROV * BLK ;
means PROV PROV*BLK / lsd alpha = .05 lines;
proc print data = BScom01e.twowayANOVARESULT;
run;

data BScom01e.twowayANOVARESULTkeepPROV;
set BScom01e.twowayANOVARESULT;
if _SOURCE_ = "ERROR"  then delete;
if _SOURCE_ = "BLK" then delete;
if _SOURCE_ = "PROV*BLK" then delete;
proc print data =  BScom01e.twowayANOVARESULTkeepPROV;
title "Two-way ANOVA RESULTs for HGT / DBH by SITE ^ show PROV ^";
run;

data BScom01e.twowayANOVARESULTkeepUNSIG;
set BScom01e.twowayANOVARESULT;
if PROB = .  then delete;
if PROB < 0.05 then delete;
proc print data =  BScom01e.twowayANOVARESULTkeepUNSIG;
title "Two-way ANOVA for HGT / DBH by SITE ^show UNSIG^";
run;

data BScom01e.OnewayANOVAdata;
set BScom01e.mean;
if _TYPE_ = 0 then delete;
if _TYPE_ = 1 then delete;
if _TYPE_ = 2 then delete;

proc print;
Title 'Two-way Anova for Avg STDDev Hgt/DBH SURR ';
run;

proc glm data = BScom01e.OnewayANOVAdata;
title 'Two-way Anova Results for Avg STDDev Hgt/DBH SURR  ';
class SITE PROV;
model AvgHgt AvgDBH stdHgt stdDBH SURR = SITE PROV SITE *PROV ;

run;
proc glm data = BScom01e.OnewayANOVAdata;
title 'One-way Anova Results for Avg STDDev Hgt/DBH SURR  ';
class PROV;
model AvgHgt AvgDBH stdHgt stdDBH SURR = PROV;
```

```
by SITE;
run;

data BScom01e.HajimeRegData;
set BScom01e.mean;
if _TYPE_ = 0 then delete;
if _TYPE_ = 1 then delete;
if _TYPE_ = 3 then delete;
drop BLK _TYPE_ _FREQ_  ;
drop StdHgt StdDBH SkewHgt SkewDBH KurtHgt KurtDBH SURR;
If Site = 3 Then do;
      S3AvgHgt = AvgHgt ;
      S3AvgDBH = AvgDBH;
      END;
If Site = 7 Then do;
      S7AvgHgt = AvgHgt ;
      S7AvgDBH = AvgDBH;
      END;
If Site = 10  Then do;
      S10AvgHgt = AvgHgt ;
      S10AvgDBH = AvgDBH;
      END;
If Site = 12 Then do;
      S12AvgHgt = AvgHgt ;
      S12AvgDBH = AvgDBH;
      END;
If Site = 14 Then do;
      S14AvgHgt = AvgHgt ;
      S14AvgDBH = AvgDBH;
      END;
If Site = 16 Then do;
      S16AvgHgt = AvgHgt ;
      S16AvgDBH = AvgDBH;
      END;
If Site = 18 Then do;
      S18AvgHgt = AvgHgt ;
      S18AvgDBH = AvgDBH;
      END;
drop SITE AvgHgt AvgDBH;
run;

proc means data = BScom01e.HajimeRegData noprint;
Title "BKSPC.procMeans.[Sum]->HajimeRegData ";
class Prov;
var   S3AvgHgt S3AvgDBH S7AvgHgt S7AvgDBH S10AvgHgt S10AvgDBH
      S12AvgHgt S12AvgDBH s14AvgHgt S14AvgDBH S16AvgHgt S16AvgDBH
      S18AvgHgt S18AvgDBH;
output out = BScom01e.HajimeRegDataRF
sum = S3AvgHgt S3AvgDBH S7AvgHgt S7AvgDBH S10AvgHgt S10AvgDBH
      S12AvgHgt S12AvgDBH S14AvgHgt S14AvgDBH S16AvgHgt S16AvgDBH
      S18AvgHgt S18AvgDBH;
run;

data BScom01e.HajimeRegDataRF;
set BScom01e.HajimeRegDataRF;
if _Type_ = 0 then delete;
drop _Type_ _Freq_;
```

```
run;

data BScom01e.ProvClimate;
infile
'D:\Thesis.090106\SubTopic.3.Com.3,7,10,12,14,16,18\SAS\ProvExtractClim
ate.txt' dlm = ',' LRECL = 1000 ;
input PROV LONG LAT ALT MaxT01 MaxT02 MaxT03 MaxT04 MaxT05 MaxT06
MaxT07 MaxT08 MaxT09 MaxT10 MaxT11 MaxT12
       MinT01 MinT02  MinT03 MinT04 MinT05 MinT06 MinT07 MinT08 MinT09
MinT10 MinT11 MinT12
       Prec01 Prec02 Prec03 Prec04 Prec05 Prec06 Prec07 Prec08 Prec09
Prec10 Prec11 Prec12;
run;

data Bscom01e.HajimeRegDataRF;
MERGE BScom01e.HajimeRegDataRF BScom01e.ProvClimate;
by PROV;
run;

proc print data = BScom01e.HajimeRegDataRF;
Title "Data display for Hajime Regression ";
run;

Proc REG data = BScom01e.HajimeRegDataRF;
Title "Hajime REG TEST";
model S3AvgHgt S3AvgDBH S7AvgHgt S7AvgDBH S10AvgHgt S10AvgDBH
      S12AvgHgt S12AvgDBH S14AvgHgt S14AvgDBH S16AvgHgt S16AvgDBH
      S18AvgHgt S18AvgDBH
      =
      LONG LAT ALT MaxT01 MaxT02 MaxT03 MaxT04 MaxT05 MaxT06
      MaxT07 MaxT08 MaxT09 MaxT10 MaxT11 MaxT12 MinT01 MinT02
      MinT03 MinT04 MinT05 MinT06 MinT07 MinT08 MinT09 MinT10
      MinT11 MinT12 Prec01 Prec02 Prec03 Prec04 Prec05 Prec06
      Prec07 Prec08 Prec09 Prec10 Prec11 Prec12
/ selection = maxr start =1 stop = 1;
run;

proc Princomp data = BScom01e.HajimeRegDataRF
out = BScom01e.FinalRegData PREFIX = PC STD N = 10;
var   S3AvgHgt S3AvgDBH S7AvgHgt S7AvgDBH S10AvgHgt S10AvgDBH
      S12AvgDBH S14AvgDBH S16AvgDBH S18AvgDBH;
title "PCA Results";
proc print data = BScom01e.FinalRegData;
title "Print Final Reg Data";
run;

proc Reg data = Bscom01e.FinalRegData;
model PC1 PC2 PC3 PC4 PC5
      =
      LONG LAT ALT MaxT01 MaxT02 MaxT03 MaxT04 MaxT05 MaxT06
      MaxT07 MaxT08 MaxT09 MaxT10 MaxT11 MaxT12 MinT01 MinT02
      MinT03 MinT04 MinT05 MinT06 MinT07 MinT08 MinT09 MinT10
      MinT11 MinT12 Prec01 Prec02 Prec03 Prec04 Prec05 Prec06
      Prec07 Prec08 Prec09 Prec10 Prec11 Prec12
      / Selection = STEPWISE TOL ;
title 'Final REG / Stepwise (SLE = 0.5) ';
run;
```

```
proc Reg data = Bscom01e.FinalRegData;
model  PC1 PC2 PC3 PC4 PC5
      =
      LONG LAT ALT MaxT01 MaxT02 MaxT03 MaxT04 MaxT05 MaxT06
      MaxT07 MaxT08 MaxT09 MaxT10 MaxT11 MaxT12 MinT01 MinT02
      MinT03 MinT04 MinT05 MinT06 MinT07 MinT08 MinT09 MinT10
      MinT11 MinT12 Prec01 Prec02 Prec03 Prec04 Prec05 Prec06
      Prec07 Prec08 Prec09 Prec10 Prec11 Prec12
      / Selection = STEPWISE TOL ;
title 'Final REG / Stepwise (SLE = 0.15) ';
run;

proc Reg data = Bscom01e.FinalRegData;
model  PC1 PC2 PC3 PC4 PC5
      =
      LONG LAT ALT MaxT01 MaxT02 MaxT03 MaxT04 MaxT05 MaxT06
      MaxT07 MaxT08 MaxT09 MaxT10 MaxT11 MaxT12 MinT01 MinT02
      MinT03 MinT04 MinT05 MinT06 MinT07 MinT08 MinT09 MinT10
      MinT11 MinT12 Prec01 Prec02 Prec03 Prec04 Prec05 Prec06
      Prec07 Prec08 Prec09 Prec10 Prec11 Prec12
      / Selection = STEPWISE SLE = 0.05 TOL ;
title 'Final REG / Stepwise (SLE = 0.05) ';
run;

Proc Reg data = BScom01e.FinalRegData;
model PC1 PC2 PC4
      =
      LONG LAT ALT MaxT01 MaxT02 MaxT03 MaxT04 MaxT05 MaxT06
      MaxT07 MaxT08 MaxT09 MaxT10 MaxT11 MaxT12 MinT01 MinT02
      MinT03 MinT04 MinT05 MinT06 MinT07 MinT08 MinT09 MinT10
      MinT11 MinT12 Prec01 Prec02 Prec03 Prec04 Prec05 Prec06
      Prec07 Prec08 Prec09 Prec10 Prec11 Prec12
      /selection = RSQUARE best = 5 start = 1 stop = 4 ;
Title "Final REG / RSQUARE (best = 5 Var = 1-4)";
run;

proc Reg data =  BScom01e.FinalRegData;
Title 'PC REG Details';
model PC1 = MaxT07 Prec04     /TOL;
model PC1 = MaxT07 Prec03     /TOL;
model PC1 = MaxT08 Prec04     /TOL;
model PC1 = Prec04 Prec10     /TOL;
model PC1 = MaxT08 Prec03     /TOL;
model PC1 = MaxT08 MaxT09 Prec03/TOL;
model PC1 = MaxT08 MaxT09 Prec04/TOL ;
model PC1 = MaxT05 MaxT07 Prec04/TOL ;
model PC1 = MaxT08 MinT04 Prec04/TOL ;
model PC1 = MaxT08 MaxT10 Prec04/TOL ;
model PC1 = MaxT05 MaxT08 MaxT09 Prec04/TOL;
model PC1 = MaxT08 MaxT10 MinT05 Prec04/TOL ;
model PC1 = MaxT05 MaxT08 MaxT10 Prec04/TOL ;
model PC1 = MaxT04 MaxT08 MaxT09 Prec04/TOL ;
model PC1 = MaxT01 MaxT02 MinT12 Prec03/TOL     ;

model PC2 = Prec07 Prec09     /TOL;
model PC2 = Prec06 Prec09     /TOL;
model PC2 = Prec03 Prec09     /TOL;
```

```
model PC2 = Prec04 Prec09     /TOL;
model PC2 = MinT11 Prec09     /TOL;
model PC2 = ALT Prec07 Prec09       /TOL;
model PC2 = Prec04 Prec06 Prec09    /TOL;
model PC2 = Prec04 Prec07 Prec09    /TOL;
model PC2 = MinT07 Prec06 Prec09    /TOL;
model PC2 = Prec06 Prec09 Prec10    /TOL;
model PC2 = ALT Prec04 Prec07 Prec09/TOL;
model PC2 = ALT MinT10 Prec06 Prec09/TOL;
model PC2 = ALT MinT11 Prec06 Prec09/TOL;
model PC2 = ALT Prec03 Prec07 Prec09/TOL;
model PC2 = ALT Prec07 Prec09 Prec11/TOL;

model PC4 = MinT04 MinT10     /TOL;
model PC4 = MinT05 MinT10     /TOL;
model PC4 = MinT05 MinT09     /TOL;
model PC4 = MinT02 MinT04     /TOL;
model PC4 = MaxT05 Prec09     /TOL;
model PC4 = ALT MinT04 MinT10 /TOL;
model PC4 = MinT03 MinT04 MinT10/TOL;
model PC4 = MaxT05 MinT04 MinT10/TOL;
model PC4 = MinT04 MinT10 Prec08/TOL;
model PC4 = ALT MinT05 MinT10 /TOL;
model PC4 = MaxT10 Prec02 Prec03 Prec07/TOL;
model PC4 = MaxT03 Prec02 Prec03 Prec07/TOL;
model PC4 = MaxT04 Prec02 Prec03 Prec07/TOL;
model PC4 = MaxT02 Prec02 Prec03 Prec07/TOL;
model PC4 = MaxT04 Prec02 Prec07 Prec08/TOL;
run;
```

Appendix 3: Portfolio model example

```
TITLE
      SeedSourcePortfolioSelection;

OPTION
      ModelType = Quadratic

INDEX

      i :=1..54   ;     ! index eligible sites
      j :=1..7    ;     ! index 7 candidate seed sources
      k :=1..7    ;     ! index 7 candidate seed sources

DATA

      R[i,j]            :=DATAFILE("Rij.csv");


      COV[i,j,k]  :=DATAFILE("COVijk.csv");


VARIABLE

      PSD[i,j];

MODEL

      Min TotalRis =
COV[1,1,1]*PSD[1,1]*PSD[1,1] + COV[1,1,2]*PSD[1,1]*PSD[1,2] +
COV[1,1,3]*PSD[1,1]*PSD[1,3] + COV[1,1,4]*PSD[1,1]*PSD[1,4] +
COV[1,1,5]*PSD[1,1]*PSD[1,5] + COV[1,1,6]*PSD[1,1]*PSD[1,6] +
COV[1,1,7]*PSD[1,1]*PSD[1,7] +
…
…
COV[52,7,1]*PSD[52,7]*PSD[52,1] + COV[52,7,2]*PSD[52,7]*PSD[52,2] +
COV[52,7,3]*PSD[52,7]*PSD[52,3] + COV[52,7,4]*PSD[52,7]*PSD[52,4] +
COV[52,7,5]*PSD[52,7]*PSD[52,5] + COV[52,7,6]*PSD[52,7]*PSD[52,6] +
COV[52,7,7]*PSD[52,7]*PSD[52,7] ;

SUBJECT TO

      EQ2CELLPortfolio [i] -> EQ2:
      SUM(j : PSD[i,j]) = 1;

      EQ3CELLReturn [i] -> EQ3:
      SUM(j : PSD[i,j]*R[i,j]) <= 0.65 ;

      !LowBound [i] -> EQ4LB:
      !SDS >= Pmin;

      !UpBound [i] -> EQ4UB:
      !SDS <= Pmax;

      !LowBound2 [i] -> EQ5LB2:
      !SDS >= Pmin * SEL;
```

```
        !BinaryLink [i] -> EQ6BL:
        !SEL >= SDS;

        !BINARY SEL;

END
```