

# Comparison of canonical correlation and regression based focal point seed zones of white spruce

Mark R. Lesser and William H. Parker

**Abstract:** The focal point seed zone methodology determines spatially explicit areas of adaptive similarity for any selected geographic point and is used to match seed sources and planting sites. A total of 127 seed sources (provenances) of white spruce (*Picea glauca* (Moench) Voss) from Ontario and western Quebec were established at a greenhouse and in six field trials throughout Ontario. Growth and phenological variables were measured over three growing seasons. Two focal point seed zone methodologies were employed: (i) using models derived from principal components analysis (PCA) of biological response variables followed by multiple linear regression against climate variables and (ii) using models derived from canonical correlation analysis (CANCOR). While both approaches use climate data to model adaptive variation, CANCOR reduces the number of steps in the analysis by simultaneously finding the relationships of biological and climatic variables that maximize the covariance between the two data sets. Although more of the variation in adaptive biological traits was actually described by climate variables using the PCA–regression approach, this method produced intuitively less realistic patterns. Both methods showed similar overall geographic trends, but the CANCOR method had a finer resolution, especially in southern Ontario, presumably due to statistical efficiency; growth was modeled by all climate variables.

**Résumé :** La méthode de détermination des zones semencières à partir de points focaux permet de délimiter de façon spatialement explicite les zones d'adaptation similaire pour n'importe quel point géographique sélectionné. Cette méthode est utilisée pour associer les sources de graines et les sites de plantation. Les auteurs ont étudié 127 sources de graines (provenances) d'épinette blanche (*Picea glauca* (Moench) Voss) de l'Ontario et de l'ouest du Québec établies en serre et dans six tests au champ à travers l'Ontario. Des variables reliées à la croissance et la phénologie ont été mesurées pendant trois saisons de croissance. Deux méthodes de détermination des zones semencières à partir de points focaux ont été utilisées : (i) des modèles découlant de l'analyse en composantes principales des variables de réponse biologique, suivie de la régression linéaire multiple avec des variables climatiques et (ii) des modèles découlant de l'analyse des corrélations canoniques. Bien que les deux approches utilisent les données climatiques pour modéliser la variation liée à l'adaptation, l'analyse des corrélations canoniques réduit le nombre d'étapes en identifiant simultanément les relations entre les variables biologiques et climatiques qui maximisent la covariance entre les deux ensembles de données. Quoiqu'une plus grande partie de la variation des caractères biologiques indicatifs de l'adaptation ait pu être prise en compte par les variables climatiques avec l'analyse en composantes principales et la régression, cette méthode a produit des patrons intuitivement moins réalistes. Les tendances géographiques étaient généralement similaires avec les deux méthodes, mais la résolution obtenue avec l'analyse des corrélations canoniques était meilleure, particulièrement dans le sud de l'Ontario, vraisemblablement à cause de l'efficacité statistique. La croissance a été modélisée par toutes les variables climatiques.

[Traduit par la Rédaction]

## Introduction

White spruce (*Picea glauca* (Moench) Voss) has a trans-continental range in North America and is found extensively throughout both the Boreal and Great Lakes – St. Lawrence forest regions of Ontario (Rowe 1972; Nienstaedt and

Zasada 1990). White spruce possesses a high level of genetic variation (Nienstaedt and Teich 1972; Hamrick et al. 1992). While the majority of this variation has been found within provenances, significant amounts of variation have evolved between provenances for a wide variety of traits (Khalil 1985; Blum 1988; Li et al. 1993; Lesser and Parker 2004). Determination of the resulting pattern of adaptive variation is important from an ecological perspective to use best-adapted seed. Understanding patterns of variation is also important to the forest industry and to our understanding of biodiversity in white spruce at the genetic level.

Species-specific seed zones should be developed based on genetic information obtained from provenance testing or other genetic experiments (Morgenstern 1996). The focal point seed zone concept (Parker 1992; Parker and van

Received 13 June 2005. Accepted 3 February 2006. Published on the NRC Research Press Web site at <http://cjfr.nrc.ca> on 1 June 2006.

M.R. Lesser and W.H. Parker.<sup>1</sup> Faculty of Forestry and the Forest Environment, Lakehead University, 955 Oliver Road, Thunder Bay, ON P7B 5E1, Canada.

<sup>1</sup>Corresponding author (e-mail: [bill.parker@lakeheadu.ca](mailto:bill.parker@lakeheadu.ca)).

Niejenhuis 1996a, 1996b) builds on work by Rehfeldt (1984, 1991, 1994) and Campbell (1986) to characterize and map multivariate patterns of variation. The focal point approach delineates a unique spatially explicit seed zone for any selected point based on demonstrated adaptive variation for the species rather than by management jurisdictions or untested climate–species relationships. This definition of seed zone is similar to the definition given by van Buijtenen (1992) for seed deployment zones, where the zone for any given planting location is characterized by a similar pattern of variation in traits with a significant climate–trait correlation. Focal point seed zones have previously been developed for black spruce (*Picea mariana* (Mill.) BSP) and jack pine (*Pinus banksiana* Lamb.) in northwestern Ontario (Parker and van Niejenhuis 1996a, 1996b) using principal components analysis (PCA), multiple linear regression, and GIS.

The purpose of this study was to develop focal point seed zones that best model the observed patterns of adaptive variation in white spruce across Ontario and western Quebec (Lesser and Parker 2004). The previously used focal point methodology (Parker and van Niejenhuis 1996a, 1996b) was compared to an alternate approach using canonical correlation analysis (CANCOR). By its nature CANCOR is statistically a more appropriate method and highly suitable for ecological applications such as seed zone development (Gittins 1985; Westfall 1992).

## Materials and methods

### Test establishment and data collection

A total of 127 white spruce seed sources (here the term “seed sources” is used synonymously with populations and provenances) from across Ontario and western Quebec were obtained from several cooperatives and seeded in Jiffy pot 3065-140's in January 2002 in the Lakehead University greenhouse. Seed source collections came from wild stands (see Fig. 3) and comprised five or more open-pollinated families from within a 1 km<sup>2</sup> area. Families were bulked into population samples for the trials. Detailed provenance location information is given in Table A1.

Five field trials and a Lakehead University greenhouse trial were established in June and July 2002. A sixth field trial was established in fall 2003 at Angus, Ontario, using the seedlings from the greenhouse trial. Each trial consisted of three completely randomized blocks with 10 single-tree plot replications of all 127 seed sources. Field trial locations from west to east are Dryden, Kakabeka, Longlac, Englehart, Angus, and Petawawa (see Fig. 3).

Data were collected over three field seasons (2002–2004) for 94 biological variables. Seedling height and survival were measured in all 3 years. Root-collar diameter was measured in 2003 and 2004. Bud flush (six stages) and bud set (five stages) were scored in 2003. For a more complete description of scoring stages and field methods refer to Lesser and Parker (2004). Each measured trait was considered a separate variable in the analysis by trial site and year. The exception was 2002 height, which was treated as a single variable across all trial locations, as it reflects greenhouse growing conditions and not field trial conditions.

### Climate data

Climatic data for the period 1961–1990 were obtained from Dr. Dan McKenney, Canadian Forest Service, Landscape Analysis and Application Section, Great Lakes Forestry Centre (McKenney 2004). Canada-wide grids along with point data for the 127 provenance locations were provided for 67 climate variables (Table 1). The grid scale is approximately 6 min. Maximum monthly temperature, minimum monthly temperature, and monthly precipitation constituted 36 of these variables. The remaining 31 variables were derived using the BIOCLIM–ANUCLIM and SEEDGROW prediction systems. These variables consisted of growing degree-days, temperature and precipitation amounts by quarter, and growing period, along with growing season length, start time, and end time. These variables rather than the primary climate variables may be more closely related to potential vegetation community responses (Mackey et al. 1996). More complete definitions of individual climate variables can be found in Lesser and Parker (2004) and Mackey et al. (1996).

### Screening process

Two criteria were used to select which of the 94 biological variables would be used to determine focal point seed zones. First, significant differences (genetic variation) had to be present among sources. Second, the observed variation had to correspond to a climatic or geographic variable. Analysis of variance was used to detect significant differences ( $p \leq 0.05$ ), and the intraclass correlation coefficients were calculated to determine how much of the expressed variation could be attributed to genetic variation expressed among seed sources. Simple linear regressions of the selected biological variables were run based on provenance mean values, against climatic and geographic variables, to determine whether differences were attributable to climatic or geographic factors.

### PCA–regression analysis

Following the screening process, provenance mean values of the 57 retained biological variables were analyzed using PCA. PCA summarizes the main components of variation in a data set. PCA was run using the Princomp procedure in SAS (SAS Institute Inc. 2000). Eigenvalues were examined to determine which of the PC axes would be retained, and analysis of the eigenvectors showed each variable's contribution to each axis.

Normalized provenance factor scores were calculated for the three main axes of variation. These factor scores were then used as new summary variables in multiple linear regressions against the 67 climate variables. Multiple regressions were run using the  $R^2$  regression procedure in SAS (SAS Institute Inc. 2000). To avoid issues of correlated independent variables and overfitting of the model, predictor variables with tolerances less than 0.1 and nonsignificant  $t$  values ( $p \leq 0.05$ ) were eliminated, and the model was refitted (Wilkinson et al. 1992). The regression equations were then used to model the three main PCA axes by conversion to spatial data using GIS. Predicted scores for each axis were reproduced as contoured geographic grids using grid algebra in the Grid subpackage of ArcGIS (ESRI 2002). These grids summarized the spatial pattern of adaptive variation.

Table 1. Range, units of measure, and canonical correlations and coefficients for canonical variates 1–3 for 67 climate variables.

Climate variable	Range		Units	Canonical correlation			Canonical coefficient		
	Min.	Max.		1	2	3	1	2	3
Mean diurnal range	8	13	°C	-0.284	0.119	-0.281	4.228	-3.453	-1.927
Isothermality	0.2	0.2	—	0.021	0.034	0.109	-0.177	0.129	-0.275
Temperature seasonality	3	5	°C	-0.289	0.113	-0.362	4.025	-20.148	-21.006
Max. temperature warmest period	19	26	°C	0.206	-0.093	0.258	-4.386	15.764	0.520
Min. temperature coldest period	-28	-11	°C	0.317	-0.118	0.373	-22.893	28.163	-8.075
Temperature annual range	34	52	°C	-0.298	0.105	-0.347	-17.028	25.207	5.015
Mean temperature wettest quarter	-6	19	°C	-0.111	-0.091	0.014	0.471	-0.187	1.027
Mean temperature driest quarter	-18	19	°C	0.141	-0.012	0.382	0.064	-0.224	-0.110
Mean temperature warmquarter	13	19	°C	0.302	-0.122	0.310	-4.325	14.530	-3.257
Mean temperature coldest quarter	-19	-5	°C	0.320	-0.120	0.374	13.547	-35.783	-9.971
Annual precipitation	548	1118	mm	0.105	-0.101	0.171	48.387	-22.723	44.688
Precipitation of wettest period	78	116	mm	-0.072	-0.072	-0.149	0.354	3.555	1.995
Precipitation of driest period	20	72	mm	0.219	-0.070	0.320	-0.934	-1.856	2.716
Precipitation seasonality	12	50	mm	-0.212	0.098	-0.430	0.325	-0.146	0.119
Precipitation of wettest quarter	226	335	mm	-0.104	-0.044	-0.234	0.177	-3.425	-2.666
Precipitation of driest quarter	69	223	mm	0.209	-0.094	0.322	1.548	5.594	0.116
Precipitation of warmest quarter	201	309	mm	-0.170	-0.058	-0.300	-3.106	-3.195	-4.921
Precipitation of coldest quarter	69	303	mm	0.189	-0.056	0.325	-6.587	-11.946	-20.531
End of growing season	285	325	Julian day	0.327	-0.093	0.334	12.252	6.163	-1.860
No. of days in growing season	157	217	days	0.329	-0.116	0.357	0.000	0.000	0.000
Total precipitation for period 1	84.1	229	mm	0.218	-0.076	0.292	8.675	2.257	5.071
Total precipitation for period 2	82.2	125.8	mm	-0.142	-0.060	-0.290	10.857	-24.640	15.736
Total precipitation for period 3	348.7	608	mm	0.191	-0.162	0.135	-86.704	172.558	-90.247
Total precipitation for period 4	259	506.8	mm	0.216	-0.155	0.183	82.597	-165.984	90.025
GDD <sup>a</sup> above base temperature for period 3	946	1895	degree-days	0.315	-0.120	0.331	3.655	15.661	-6.801
Annual mean temperature	0	6	°C	0.326	-0.128	0.366	34.531	-43.267	-65.649
Annual min temperature	-7	1	°C	0.329	-0.129	0.362	-197.859	-185.517	-27.117
Annual max temperature	5	11	°C	0.312	-0.122	0.363	-13.611	-80.375	-7.650
Mean temp period 3	11	14	°C	0.273	-0.120	0.274	15.825	17.163	-4.443
Jan. mean monthly min. temperature	-28	-10	°C	0.321	-0.115	0.372	31.731	29.726	13.073
Feb. mean monthly min. temperature	-26	-11	°C	0.330	-0.121	0.365	20.369	29.420	2.410
Mar. mean monthly min. temperature	-19	-6	°C	0.336	-0.130	0.378	13.249	22.137	16.507
Apr. mean monthly min. temperature	-7	0.9	°C	0.342	-0.135	0.369	21.829	22.697	-4.845
May mean monthly min. temperature	0	6	°C	0.321	-0.116	0.369	13.026	5.604	4.770
June mean monthly min. temperature	5	11	°C	0.317	-0.132	0.327	5.562	8.430	12.621
July mean monthly min. temperature	8	14	°C	0.317	-0.124	0.295	8.148	0.661	0.471
Aug. mean monthly min. temperature	7	13	°C	0.327	-0.127	0.326	13.121	9.833	8.757
Sept. mean monthly min. temperature	2	10	°C	0.308	-0.120	0.327	3.489	1.916	13.018
Oct. mean monthly min. temperature	-2	5	°C	0.315	-0.123	0.306	-0.603	8.868	3.185
Nov. mean monthly min. temperature	-12	0	°C	0.306	-0.139	0.346	20.362	27.895	-3.162
Dec. mean monthly min. temperature	-22	-6	°C	0.311	-0.131	0.362	28.929	25.008	7.174

**Table 1** (concluded).

Climate variable	Range		Units	Canonical correlation			Canonical coefficient		
	Min.	Max.		1	2	3	1	2	3
	Jan. mean monthly max. temperature	-14		-2	°C	0.310	-0.116	0.385	-2.144
Feb. mean monthly max. temperature	-11	-2	°C	0.322	-0.119	0.371	-2.517	8.087	1.969
Mar. mean monthly max. temperature	-3	3	°C	0.314	-0.127	0.349	2.931	8.148	-1.202
Apr. mean monthly max. temperature	5	11	°C	0.302	-0.122	0.362	1.694	4.512	7.737
May mean monthly max. temperature	13	19	°C	0.236	-0.106	0.293	-5.127	8.420	-0.477
June mean monthly max. temperature	16	23	°C	0.223	-0.100	0.212	2.906	0.981	8.298
July mean monthly max. temperature	19	26	°C	0.210	-0.088	0.253	-0.058	-29.984	8.375
Aug. mean monthly max. temperature	19	25	°C	0.301	-0.088	0.339	7.601	-2.919	6.646
Sept. mean monthly max. temperature	13	20	°C	0.316	-0.116	0.348	-16.579	4.107	1.112
Oct. mean monthly max. temperature	7	13	°C	0.322	-0.133	0.335	0.793	12.904	3.509
Nov. mean monthly max. temperature	-4	6	°C	0.310	-0.126	0.367	-0.586	-1.589	6.249
Dec. mean monthly max. temperature	-12	0	°C	0.304	-0.118	0.382	-4.815	15.453	5.618
Jan. mean monthly temperature	23	113	mm	0.113	-0.021	0.315	-9.068	7.024	0.823
Feb. mean monthly temperature	19	76	mm	0.235	-0.054	0.326	-5.726	3.486	-2.910
Mar. mean monthly temperature	28	75	mm	0.208	-0.079	0.302	-7.442	-0.417	-5.128
Apr. mean monthly temperature	33	75	mm	0.221	-0.128	0.265	-9.428	0.342	-7.368
May mean monthly temperature	49	87	mm	0.184	-0.147	0.113	-1.477	-1.371	-5.043
June mean monthly temperature	64	107	mm	-0.137	-0.093	-0.318	-1.747	1.431	-1.758
July mean monthly temperature	58	106	mm	-0.200	0.033	-0.351	-4.792	2.968	-3.888
Aug. mean monthly temperature	68	105	mm	-0.111	-0.126	-0.122	-2.465	2.488	-2.273
Sept. mean monthly temperature	62	112	mm	-0.172	-0.024	-0.162	-2.785	-0.817	-3.116
Oct. mean monthly temperature	38	107	mm	0.023	-0.107	0.038	-2.751	2.451	-4.680
Nov. mean monthly temperature	29	116	mm	0.104	-0.108	0.259	-8.025	4.570	-7.384
Dec. mean monthly temperature	25	116	mm	0.207	-0.081	0.312	-6.103	8.724	-1.277

<sup>a</sup>GDD, growing degree-days.

In the final stage of the focal point seed zone procedure, an Arc Macro Language (AML) computer program was used to produce a unique seed zone for white spruce at any given point within the study area based on data from the 127 seed source locations. Contoured geographic grids for each of three axes were overlaid and intersected using the INTERSECT command in ArcGIS (ESRI 2002). The three grids were given equal weighting in the intersection process. The resulting grid identified areas of similarity in terms of standard deviations from the source location (Parker and van Niejenhuis 1996a, 1996b). Zones of decreasing adaptive similarity from the focal point were identified by lighter shading patterns. All grids were produced in ArcGIS 8.3 (ESRI 2002).

### CANCOR

Provenance means for the 57 retained biological variables, along with the provenance values for the 67 climatic variables, were entered into CANCOR (SAS Institute Inc. 2000). CANCOR considers both sets of data (biological and climatic) simultaneously and selects linear functions that maximize the covariance between the variable sets. By this procedure CANCOR finds linear functions of each variable set that maximize the correlation between the sets. As in PCA, the successive pairs of canonical variates are uncorrelated, or orthogonal, to each other (Thompson 1984).

Canonical variates were assessed for level of significance using an *F* test ( $p > 0.05$ ). Significant variates were retained for further analysis. Correlations of variables from each variable set and each canonical variate were calculated to determine which variables were contributing the most to each of the canonical variates. For each of the significant variates the standardized canonical coefficients for the climate variable set were calculated. The canonical coefficients are essentially equivalent to partial regression coefficients in multiple linear regressions (Gittins 1985). Canonical coefficients from one variable set can therefore be used to predict values of the other variable set for each of the canonical variates in question.

The climatic canonical coefficients were used to model the biological scores for the three significant canonical variates. The three models were converted to spatial data using the Grid subpackage of ArcGIS 8.3 (ESRI 2002). Each of the 67 supplied climate variable grids (McKenney 2004) was multiplied by its respective coefficient, and the 67 resulting grids were then summed, producing a single grid that represented the predicted biological pattern of variation for each respective canonical variate.

To facilitate comparison with PCA–regression based focal point seed zones, CANCOR grids were standardized using the provenance point means and standard deviations. The resulting standardized grids summarized the spatial pattern of adaptive variation in relation to the 67 climate variables. The computer program written to produce focal point seed zones for the PCA–regression method was adapted to intersect and overlay the three CANCOR grids, showing areas of similarity in standard deviations from the focal point.

For demonstration purposes, focal point seed zones were produced using both the PCA–regression and CANCOR methods for four points across the study area. Methods were compared based on visual interpretation of the resulting seed

zones and quantitatively by comparing the cumulative amount of variation explained by each method. For the PCA–regression approach, this was calculated as the sum of the products of variation explained by each principal component and its coefficient of determination ( $R^2$ ) from the associated multiple regression model. For the CANCOR approach, the cumulative proportion of variance that the climatic variables explained in the biological data set was determined by canonical redundancy analysis (SAS Institute Inc. 2000).

## Results

### PCA–regression analysis

Fifty-seven biological variables were retained from the double screening process used to select the variables exhibiting adaptive variation. The first three axes from PCA were retained (Table 2) for modelling patterns of adaptive variation and described 54.5% of the total. PC1 described 34% of the total variation. PC2 and PC3 described an additional 12.5% and 8%, respectively. The remaining 54 PC axes (results not shown) had low eigenvalues (less than 3.5) and individually contributed little to the accounted for variation.

Principal components are uncorrelated (orthogonal) by definition and represent the structure of multivariate data (Morrison 1990). In this case the axes reflect the influence of different categories of variables with different biological significance. PC1 mainly represented growth potential, as seen by the relatively large positive eigenvectors associated with growth variables (Table 2). PC1 was also strongly determined by Englehart survival variables, the final greenhouse elongation variable (day 70), and greenhouse bud flush variables. The first four greenhouse elongation variables showed a negative relationship with PC1, indicating that the opposite of growth potential was being expressed in the early stages of greenhouse growth; however, the final elongation measurement at day 70 had a positive correlation with PC1 of the same magnitude as that with field-trial growth variables. This result indicates that the fastest growing sources during the early weeks of the greenhouse study tended to grow slower in later weeks. We hypothesize that this result likely reflects bud-flush timing in the greenhouse. The northern sources flushed earlier and began elongation earlier, before eventually being surpassed by faster growing, but later flushing, southern sources.

PC2 was strongly determined by phenological traits at the five field trials. Bud set and bud flush variables showed relatively large positive relationships with PC2, with the exception of later-stage bud flush variables at the Longlac trial, which were weaker, but still positive (Table 2). PC3 showed a relatively large negative relationship with greenhouse bud flush variables. PC3 also showed a strong positive relationship with the first four greenhouse elongation variables. The day 70 elongation variable had a much weaker correspondence. This result, coupled with the large positive correspondence of the day 70 elongation variable to PC1, suggests that the pattern of growth initiation in the greenhouse is essentially uncorrelated with the two main components of variation, that is, growth and phenology in the field.

PC1 factor scores were fit to a model ( $R^2 = 27.1\%$ ) containing the following climatic variables: precipitation in the



**Table 2.** Summary of principal components analysis for principal components (PC) 1–3.

	PC1	PC2	PC3
Eigenvalue	19.37	7.10	4.63
Percent variaton	33.98	12.5	8.13
Cumulative variation	33.98	46.4	54.56
Eigenvector			
Dryden bud flush stage 2	0.06	0.16	0.05
Dryden bud flush stage 3	0.07	0.17	0.04
Dryden bud flush stage 4	0.06	0.21	-0.02
Dryden bud flush stage 5	0.05	0.17	-0.01
Dryden bud flush stage 6	0.04	0.18	0.05
Longlac bud flush stage 2	0.11	0.13	-0.13
Longlac bud flush stage 3	0.13	0.11	-0.15
Longlac bud flush stage 4	0.14	0.09	-0.17
Longlac bud flush stage 5	0.12	0.03	-0.21
Longlac bud flush stage 6	0.10	0.00	-0.17
Greenhouse bud flush stage 2	0.14	-0.01	-0.19
Greenhouse bud flush stage 3	0.17	-0.02	-0.21
Greenhouse bud flush stage 4	0.17	-0.02	-0.19
Greenhouse bud flush stage 5	0.11	0.00	-0.20
Greenhouse bud flush stage 6	0.10	0.02	-0.28
Dryden bud set stage 5	0.06	0.26	0.12
Kakabeka bud set stage 3	0.05	0.24	0.07
Kakabeka bud set stage 4	0.11	0.24	0.13
Kakabeka bud set stage 5	0.10	0.21	0.14
Longlac bud set stage 4	-0.02	0.22	0.10
Longlac bud set stage 5	0.03	0.20	0.18
Englehart bud set stage 3	0.02	0.20	0.15
Englehart bud set stage 4	0.09	0.21	0.14
Englehart bud set stage 5	0.04	0.16	0.12
Petawawa bud set stage 3	0.01	0.17	0.06
Petawawa bud set stage 4	0.05	0.25	0.13
Height 2002	0.19	-0.15	0.06
Dryden height 2003	0.10	-0.04	-0.12
Kakabeka height 2003	0.17	-0.11	0.09
Longlac height 2003	0.19	-0.07	0.02
Englehart height 2003	0.18	-0.12	0.10
Petawawa height 2003	0.19	-0.05	0.09
Dryden height 2004	0.15	-0.07	-0.03
Kakabeka height 2004	0.18	-0.08	0.10
Longlac height 2004	0.17	-0.08	0.00
Englehart height 2004	0.19	-0.08	0.13
Petawawa height 2004	0.18	-0.03	0.07
Angus height 2004	0.17	-0.09	0.02
Dryden diameter 2003	0.11	0.00	-0.04
Longlac diameter 2003	0.18	0.01	0.06
Englehart diameter 2003	0.19	-0.07	0.13
Petawawa diameter 2003	0.17	-0.06	0.07
Dryden diameter 2004	0.16	-0.06	-0.05
Kakabeka diameter 2004	0.18	-0.11	0.08
Longlac diameter 2004	0.17	-0.02	0.05
Englehart diameter 2004	0.19	-0.03	0.16
Petawawa diameter 2004	0.17	-0.02	0.09
Angus diameter 2004	0.18	-0.05	0.07
Petawawa survival 2002	0.08	-0.09	0.00
Englehart survival 2003	0.13	0.00	0.11
Englehart survival 2004	0.14	0.00	0.11
Longlac survival 2004	0.02	-0.13	-0.03

**Table 2 (concluded).**

	PC1	PC2	PC3
Greenhouse elongation day 18	-0.11	-0.13	0.25
Greenhouse elongation day 22	-0.12	-0.14	0.29
Greenhouse elongation day 26	-0.11	-0.19	0.25
Greenhouse elongation day 30	-0.06	-0.24	0.20
Greenhouse elongation day 70	0.18	-0.11	0.09

wettest period (coincides with the growing season), August maximum temperature, and August precipitation (Table 3). The inclusion of two summer precipitation variables, along with a late-summer temperature variable, suggests the importance of moisture conditions during bud development for the following year’s growth potential. Mapped predicted factor scores showed a trend of greater growth potential occurring in the southeast portion of the study area and decreasing through central Ontario and into northern areas (Fig. 1a). Growth potential increases again to the southwest of Lake Nipigon, showing values similar to those in south central Ontario. A similar trend of increasing growth potential moving west across northwestern Ontario was reported for a regional study of jack pine (Parker and van Niejenhuis 1996a).

The best model ( $R^2 = 48.5\%$ ) for predicting PC2 factor scores contained only June minimum temperature (Table 3), reflecting the importance of late spring – early summer temperatures for phenological characteristics, especially bud-flush timing. The same geographic trend evident for PC1 (Fig. 1a) can also be seen for PC2 (Fig. 1b), with differences in western Quebec and the eastern shore of Lake Superior. Higher scores corresponding to later bud-flush timing in the spring and later bud-set timing in the late summer – early fall are located predominantly in the southern portions of the study area. The influence of the Algonquin Highlands in creating an environment similar to that of more northern areas can be seen on the grid in south central Ontario.

PC3 factor scores were predicted ( $R^2 = 28.0\%$ ) by annual precipitation, March maximum temperature, and October precipitation (Table 3). A strong north–south trend is evident in the mapped factor scores (Fig. 1c), closely resembling winter temperature patterns in Ontario (McKenney 2004).

**CANCOR**

The first three canonical variates (CVs) were significant ( $p < 0.05$ ) (Table 4). The total amount of covariance in the data sets explained by the three variates is 72% (Table 5). Bud-set variables all had relatively high positive correlations with CV1, ranging from 0.28 for Petawawa stage 3 to 0.13 for Englehart stage 3. Bud-flush variables showed a mixture of positive and negative correlations that were all relatively weak. Growth variables showed a wide range of both positive and negative correlations.

Correlations of the biological variables to CV2 showed a similar pattern, with no identifiable trend in variable categories that would give this variate a clear biological interpretation (Table 5). For this axis, as canonical scores increased growth trait values decreased and climatic variables with negative correlations had a positive relationship with growth.

CV3 showed a generally strong positive relationship to the bud-set variable category (Table 5). Bud-flush and growth

**Table 3.** Multiple regression models of principal component analysis factor scores against climate variables.

Dependant variable	$p > F$	Independent variables	Coefficient	Tolerance	$p > t$
Principal component 1 ( $R^2 = 27.1\%$ )	<0.0001	Constant	-8.572	—	<0.0001
		Precipitation wettest period	-0.043	0.348	0.0078
		August max. temperature	0.287	0.882	<0.0001
		Aug. precipitation	0.068	0.376	<0.0001
Principal component 2 ( $R^2 = 48.45\%$ )	<0.0001	Constant	-3.393	—	<0.0001
		June min. temperature	0.400	1	<0.0001
Principal component 3 ( $R^2 = 27.96\%$ )	<0.0001	Constant	2.986	—	0.0005
		Annual precipitation	-0.011	0.165	<0.0001
		March max. temperature	0.427	0.541	<0.0001
		Oct. precipitation	0.080	0.219	<0.0001

**Table 4.** Correlations, eigenvalues, proportions, significance levels, and variance explained by the climate variables for canonical variates 1–3.

Canonical variate	Canonical correlation	Eigenvalue	Proportion	Cumulative covariance	Approx. $F$ value	$p > F$	Cumulative proportion explained by climate variables
1	0.999	754.446	0.443	0.443	1.360	<0.0001	0.0175
2	0.998	300.067	0.176	0.620	1.230	<0.0001	0.1100
3	0.997	172.887	0.102	0.721	1.140	0.008	0.1310

variable categories had a range of both positive and negative correlations. This can be interpreted as meaning that as canonical scores increase the timing of field bud set tends to increase, while height and diameter at some field sites tends to decrease.

Correlations of the climate variables with CV1 reflect a strong influence of temperature-related variables (Table 1). Monthly temperature variables all showed relatively high positive correlations. Overall, precipitation-related variables had lower correlations, with variables from the summer months and those associated with the growing season generally being negative. Correlations with CV2 overall were much weaker, with the highest correlation being negative for total precipitation in period 3. Only 7 of the 67 variables showed a positive relationship. The strongest correlation for CV3 was -0.43 for precipitation seasonality. Again, monthly temperature variables all showed relatively high positive correlations, with monthly precipitation variables relatively lower. The summer month precipitation variables (June–September) all showed negative correlations. Overall, correlations do not give any clear interpretation of how any given “set” of climate variables interacts with a set, or category, of biological variables, pointing to the complexity of the relationship between climate and biological variables.

Canonical coefficients for the climate variables were used to model the biological variable scores for each canonical variate (Table 1). These coefficients, or weights, reflect the association of each variable after the influence of all other variables in the set have been removed (Gittins 1985). While in principle the coefficients can be used as an indication of the effects and direction that variables have, interpretation is more difficult and not as reliable as using the correlation values for such purposes. This issue is a result of the drastically different magnitudes of scale between climate variables (Gittins 1985). While the coefficients are individually not interpretable to any great degree, the standardized grids developed by modeling the climate variables based on their

coefficients show meaningful trends. A clear north–south trend is shown in the grid for canonical variate 1, with scores generally decreasing with movement northwards (Fig. 2a), off the east coast of Lake Superior and from the Algonquin Highlands in central Ontario. This grid reflects winter temperature patterns and shows a strong resemblance to the PC3 grid (Fig. 1c). The trend for canonical variate 2 grid is latitudinal in the east and longitudinal in the west (Fig. 2b). Higher scores are found in northeastern areas, with scores decreasing with movement south. Northwestern Ontario shows similar scores to those in more southern areas. Lakeshore effects are evident along both the north and eastern shores of Lake Superior. There is also a noticeable effect, once again, in the central Ontario area caused by the Algonquin Highlands region. There is a strong parallel between the grid for canonical variate 2 and the grid for PC2 (Fig. 1b). Both grids show similar trends in the southeast to north central and the northwest to north central regions; however, the scores in the two grids show opposite polarity. The grid for canonical variate 3 (Fig. 2c) emphasizes both the lakeshore effect of Lake Superior and the highland effect of the Algonquin area that were seen in the first two variates. Overall the trend is primarily longitudinal.

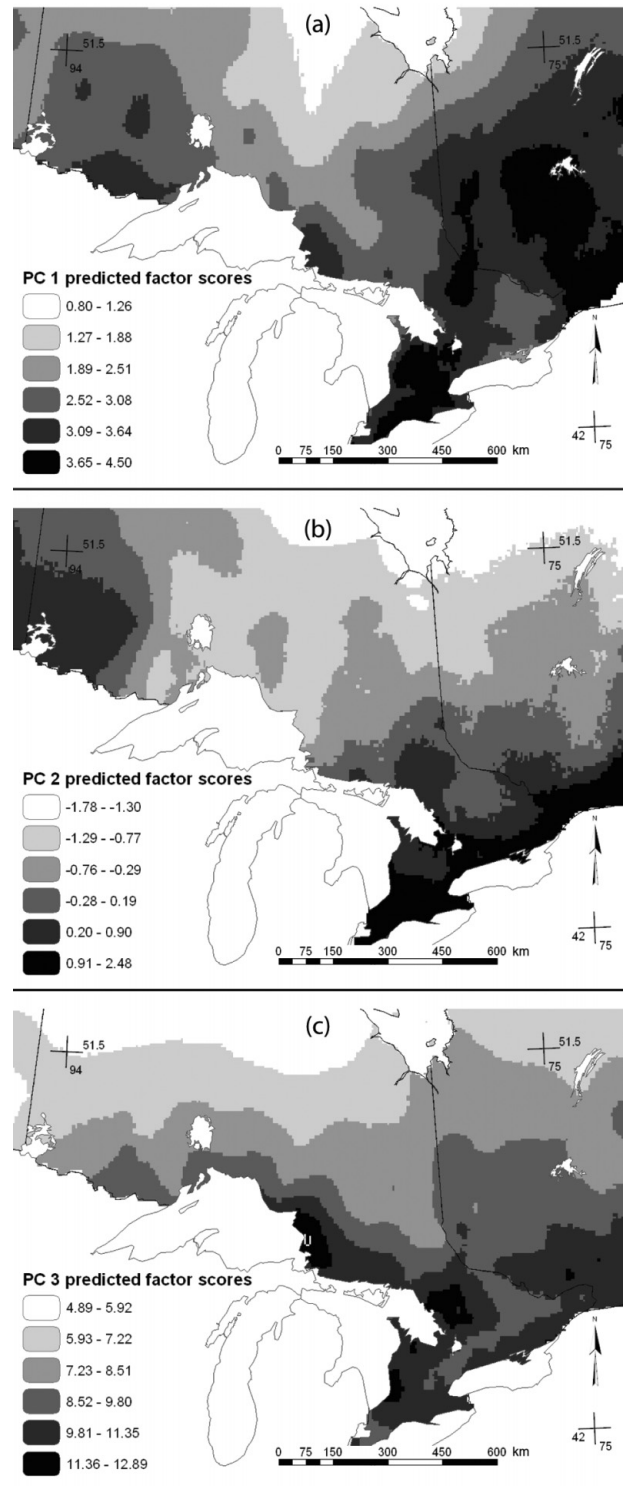
#### Focal point seed zones

Overall, focal point seed zones developed from CANCOR appear quite similar to regression-based seed zones. The same lakeshore effects and the effect of the Algonquin Highlands are readily apparent in both sets of seed zone maps (Fig. 3). Both methodologies show a transition at approximately 47°N as focal points move from northeastern Ontario into southern Ontario and Quebec. This transition corresponds to the boreal Great Lakes – St. Lawrence forest region transition in eastern Ontario (Rowe 1972). Focal points located north of 47°N generally show little acceptable area to the south, with southern points showing little to no acceptable area north of that latitude.

**Table 5.** Correlations of the biological variables to canonical variates 1–3.

Biological variable	Canonical variable		
	1	2	3
Dryden bud flush stage 2	0.004	-0.280	-0.118
Dryden bud flush stage 3	0.093	-0.266	-0.091
Dryden bud flush stage 4	0.026	-0.169	0.054
Dryden bud flush stage 5	0.120	-0.120	0.053
Dryden bud flush stage 6	0.102	-0.141	0.094
Greenhouse bud flush stage 2	0.032	-0.082	-0.002
Greenhouse bud flush stage 3	-0.091	-0.086	0.015
Greenhouse bud flush stage 4	-0.039	-0.128	-0.040
Greenhouse bud flush stage 5	-0.070	-0.121	-0.031
Greenhouse bud flush stage 6	-0.120	-0.124	0.020
Longlac bud flush stage 2	0.017	-0.028	0.159
Longlac bud flush stage 3	0.017	-0.095	0.184
Longlac bud flush stage 4	0.032	-0.094	0.110
Longlac bud flush stage 5	-0.058	-0.023	0.063
Longlac bud flush stage 6	-0.004	0.040	0.046
Dryden bud set stage 5	0.164	0.09	0.176
Englehart bud set stage 3	0.135	0.000	0.185
Englehart bud set stage 4	0.138	-0.156	0.214
Englehart bud set stage 5	0.233	-0.242	-0.127
Kakabeka bud set stage 3	0.156	-0.081	0.235
Kakabeka bud set stage 4	0.232	-0.294	0.214
Kakabeka bud set stage 5	0.264	-0.157	0.212
Longlac bud set stage 4	0.230	-0.121	0.223
Longlac bud set stage 5	0.191	-0.116	0.214
Petawawa bud set stage 3	0.281	0.073	0.055
Petawawa bud set stage 4	0.224	0.074	0.116
Angus height 2004	0.061	-0.169	-0.156
Dryden height 2003	0.022	-0.035	-0.102
Dryden height 2004	0.053	-0.025	-0.235
Englehart height 2003	0.002	-0.297	-0.129
Englehart height 2004	0.110	-0.339	-0.103
Height 2002	-0.036	-0.227	-0.183
Kakabeka height 2003	-0.096	-0.299	-0.033
Kakabeka height 2004	-0.032	-0.328	0.065
Longlac height 2003	0.017	-0.217	-0.089
Longlac height 2004	-0.062	-0.217	-0.052
Petawawa height 2003	0.103	-0.328	0.007
Petawawa height 2004	0.155	-0.320	-0.037
Angus diameter 2004	0.107	-0.180	0.023
Dryden diameter 2003	0.033	-0.248	0.128
Dryden diameter 2004	0.055	-0.017	-0.027
Englehart diameter 2003	0.134	-0.348	-0.010
Englehart diameter 2004	0.162	-0.367	0.009
Kakabeka diameter 2004	-0.026	-0.398	0.048
Longlac diameter 2003	-0.035	-0.374	0.044
Longlac diameter 2004	0.000	-0.263	0.017
Petawawa diameter 2003	0.162	-0.279	0.023
Petawawa diameter 2004	0.211	-0.345	0.032
Englehart survival 2003	0.308	-0.237	-0.011
Englehart survival 2004	0.290	-0.222	0.031
Longlac survival 2004	-0.098	-0.054	-0.289
Petawawa survival 2002	-0.115	-0.137	-0.066
Greenhouse elongation day 18	0.020	-0.186	-0.133
Greenhouse elongation day 22	-0.060	-0.091	-0.107
Greenhouse elongation day 26	-0.103	-0.024	-0.103
Greenhouse elongation day 30	-0.080	-0.016	-0.072
Greenhouse elongation day 70	0.075	-0.174	-0.096

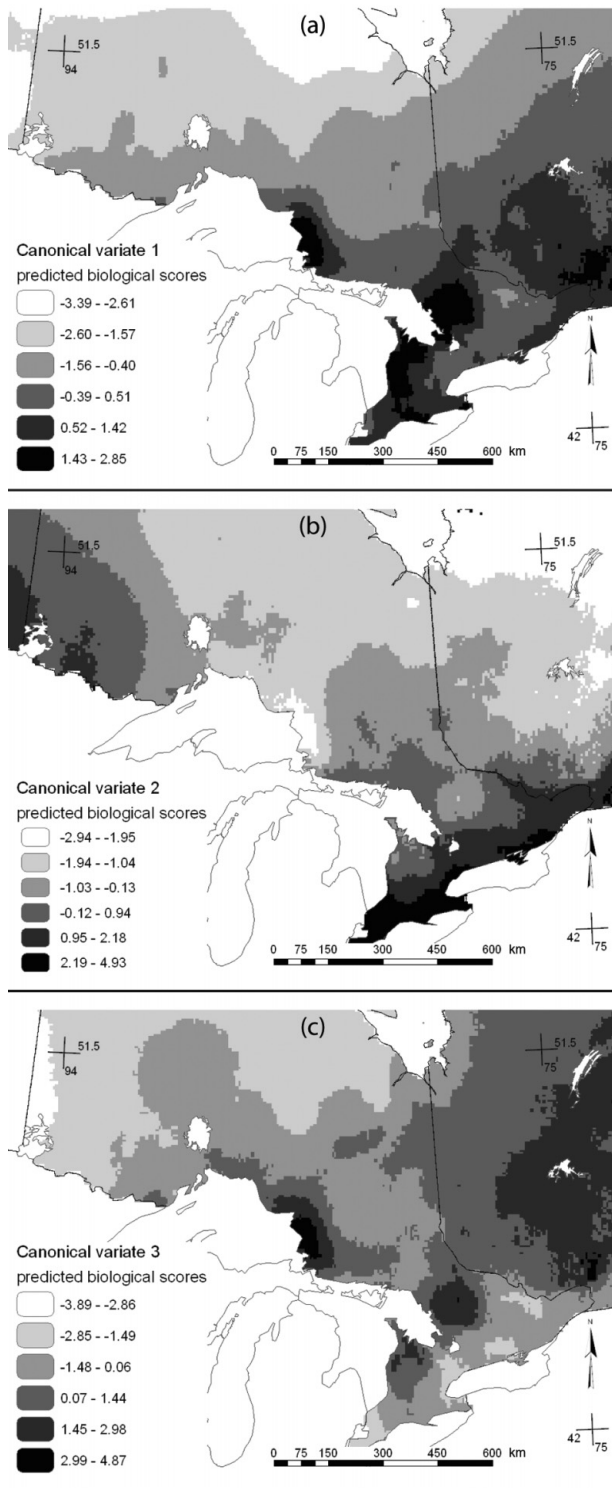
**Fig. 1.** Predicted factor scores based on 57 biological and 7 climatic variables for PC1 (a), PC2 (b), and PC3 (c).



While broad patterns of similarity are the same, there are also notable differences in the produced zones. For the two northern points, the clear “boundary” at approximately 50°N seen in the PCA–regression based zones is not apparent in the CANCOR zones. The CANCOR zones for the same northern points show zones extending north and south of 50°N but being more limited longitudinally (Fig. 3a–3d).



**Fig. 2.** Predicted standardized biological scores based on 57 biological and 67 climate variable canonical coefficients for canonical variate 1 (a), canonical variate 2 (b), and canonical variate 3 (c).



This longitudinal zonation parallels the north–south divides seen in Hills’s (1961) site subregions (e.g., 4S, 3W and 3E) in terms of area, but differs in terms of the floating boundaries inherent to the focal point seed zone method.

The examples in Fig. 3 illustrate a general pattern; that is, PCA–regression based zones are relatively coarser, and

CANCOR-based zones are relatively finer. The lakeshore effect of the north shore of Lake Superior is more pronounced in the CANCOR zones than in the regression-based zones for focal points located across northern Ontario (Fig. 3b). Generally, CANCOR-developed zones for southern focal points show less acceptable areas in the northwest as compared to regression-based zones for the same points. CANCOR-developed zones for southern points are also generally smaller than their regression-based counterparts and appear to be more heavily influenced by lakeshore effects and the Algonquin Highlands area. This result is not surprising, since PC1 accounts for growth traits at all trial locations, while CV1 has a greater association with southeastern trials. Southern CANCOR zones are also more fragmented than regression-based zones within the local area of the focal point, but show fewer remote disjunct areas of similarity (Fig. 3e–3h). These differences are due in part to the regression being a special case of CANCOR with only one dependent variable.

## Discussion

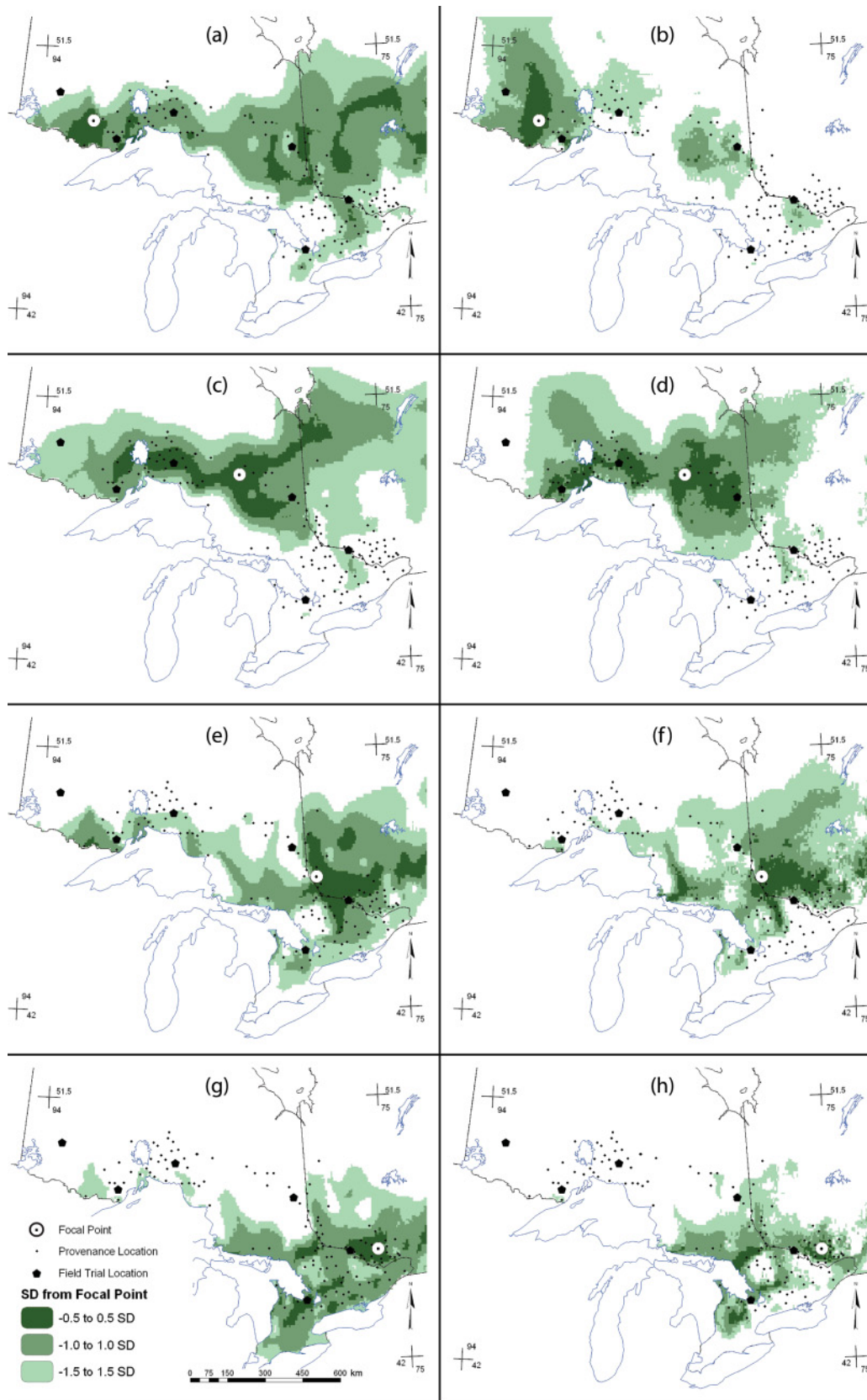
### Focal point seed zones

Generalized seed zones initially used in Ontario were based on a combination of Hills’s (1961) site regions and administrative boundaries. Site regions were determined based on geology, vegetation, climate, and soils. More recently a new approach was developed through climate analysis (Mackey et al. 1996). When compared to the focal point seed zones, these generic zones create zones that may be too specific in many cases and may not be specific enough in others.

Focal point seed zones are species specific and are based on the unique pattern of adaptive variation that each individual tree species will have within its range and within specific areas of that range. While these patterns will generally follow climatic, geological, and geographic trends in the landscape, a level of uncertainty is imposed by basing seed transfers on such generalizations without knowing the individual patterns of adaptation for a given species. Linear regressions were used to characterize the relationship of adaptive variation to climate. While nonlinear equations are often used to model this relationship (Rehfeldt 1991), these instances are typically in mountainous terrain. For the less variable terrain in Ontario, we felt that a linear model was appropriate. Previous studies in Ontario on jack pine and black spruce have both used linear models to explain the adaptive variation – climate relationship and have shown this to be an acceptable method (Parker 1992; Parker and van Niejenhuis 1996a, 1996b).

Focal point seed zones represent the areas of greatest adaptive similarity to the selected focal point. This approach is ideal in jurisdictions including Ontario, where local seed is considered best for reforestation efforts. Some have argued that the assumption of local being best adapted may not be true in some cases, possibly because of migration lag, evolutionary lag, or competitive exclusion (Rehfeldt et al. 1999). The focal point method indicates an indeterminate range of locations corresponding to the selected focal point based on the interaction of key climate variables. This range can be refined by narrowing the mapped contour intervals,

**Fig. 3.** A comparison of white spruce focal point seed zones created for four points: 49°N, 91°W (a, b); 49°N, 83°W (c, d); 47°N, 79°W (e, f); and 46°N, 76°W (g, h). PCA-regression based method seed zones are shown in the left-hand column (a, c, e, and g), and canonical correlation approach seed zones are shown in the right-hand column (b, d, f, and h). Shading indicates level of adaptive similarity in standard deviations to the focal point.



thus creating narrower, more specific zones from which to obtain seed collections (Parker and van Niejenhuis 1996a). However, while the identified sources are the most similar in terms of adaptive traits, focal point zones do not necessarily identify seed sources that will maximize growth potential at a given point.

Response functions developed for white spruce (based on 410 provenance series) by Cherry and Parker (2003) showed that increased growth will generally be achieved in northwestern Ontario by moving southern sources north. Although this approach may produce maximum yield for a given site, care must be taken to avoid planting maladapted seed. Hence, the use of focal point zones to delineate areas of similar adaptation, combined with response and transfer functions to identify the sources that will produce the greatest growth from within these zones, becomes a powerful strategy.

### PCA versus CANCOR-based zones

In spite of its ability to relate two sets of variables, CANCOR is often viewed sceptically by ecologists (Gittins 1985). CANCOR was first used in an ecological application by Austin (1968), who found the results to be unsatisfactory in comparison to results from other ordination techniques. Other studies, however, have shown that CANCOR can be used as an effective statistical tool in ecological applications (Pélissier et al. 2001; Glimaret-Carpentier et al. 2003). Westfall (1992) demonstrated the applicability of CANCOR in seed zone development, using it as the basis for seed zones of white fir (*Abies concolor* (Gord. & Glend.) Lindl. ex Hildebr.) in California. A previous attempt to use CANCOR as the basis of building focal point seed zones for black spruce in northwestern Ontario was viewed as less satisfactory than the PCA–regression based approach. The resulting zones contained more geographic discontinuities than the regression-based counterparts, such that, as with the results presented here, no biological interpretation of the axes was possible (Parker and van Niejenhuis 1996b).

When only the zone of greatest similarity ( $\pm 0.5$  standard deviation (SD)) is considered, the degree of similarity between focal point seed zones produced by the PCA–regression approach and CANCOR is very high. The fact that two independent statistical techniques produce such similar results strengthens the overall result, regardless of which approach is used. Even though both methods use the same data set to obtain the results, that they both produce such similar outcomes strengthens the conclusion that true patterns of adaptive variation are being identified and mapped.

When considering zones of lesser similarity ( $\pm 1.0$  and  $\pm 1.5$  SD), differences in the results from the two approaches become more apparent. A general longitudinal trend throughout northern portions of the study area is present in the CANCOR results, but less evident in the PCA–regression results, which better corresponds to actual observed patterns of adaptation (Lesser and Parker 2004) and patterns observed for black spruce and jack pine (Parker and van Niejenhuis 1996a, 1996b). Also, CANCOR-derived zones show fewer disjunct areas and more localized zones of similarity than the PCA–regression based zones over the central and northern parts of the study area. Greater fragmentation is seen in southern CANCOR zones; however, this

fragmentation may reflect finer scale weather patterns and be more accurate. Finally, lake shore and highland effects are seen more clearly in the CANCOR results. These effects are seen in the PCA–regression based results as well but do not persist across the landscape in the same way.

The two methodologies can be compared quantitatively by determining the cumulative variation explained by each approach. The sum of the products of  $R^2$  values and fractions of variation accounted for by each of the three PC axes totals 17%. Canonical redundancy analysis showed that the climate variables explained a cumulative 13% of the variation in the biological data set (Table 4). However, the PC components were determined based on magnitude rather than correspondence to climate data. Although the PCA–regression approach explains 4% more of the variation, the covariance between biological and climate variables is not maximized, since it is only based on 54.5% of the total variation explained by the first three PC axes. Hence, the greater explained variation may be misleading because of the two independent statistical steps used to obtain it. Furthermore, the portion of the variation that is not included in the regression step does not relate well to climate and if included would likely reduce the ability of the model to explain patterns of variation.

It is notable that the amount of variation explained by either the PCA–regression or CANCOR approach is less than 20%, indicating that most of the observed variation detected in this study is not adaptive and results from other sources, including experimental error. In a parallel study using canonical correlations of climate variables to summarize genecological variation in coastal Douglas-fir from Oregon and Washington, St. Clair et al. (2005) found that the first two canonical correlations accounted for 20% and 7% of the total trait variation. Presumably the bulk of the nonadaptive variation reflects within-population variation that was not measured in this study, but which has previously been demonstrated for white spruce in earlier work (Nienstaedt and Teich 1972).

### Genotype $\times$ environment interaction

By definition, genotype  $\times$  environment (G $\times$ E) interaction can only exist if more than one field trial is employed. With multiple field trials, a G $\times$ E interaction may exist and may be very complex in nature depending on the number and differences of the trial environments and the specific adaptations of the sources being tested.

The number of PCA axes needed to adequately describe adaptive variation in a series of provenance trials is partly related to G $\times$ E interaction, but the relationship is not clear. Even at a single site, PCA may produce as many orthogonal (uncorrelated) axes as there are response variables; for example, height, survival, and cold-hardiness traits might not be correlated for a group of seed sources at a single location and would require three axes to express the variation. Additional PCA axes will be needed to accommodate G $\times$ E interaction, that is, when rank changes exist among seed sources at different sites for the same response variable.

While the purpose of this study was not to determine the complexity of existing G $\times$ E interaction, the presence of the interaction drives the need for the focal point zone approach involving the overlay of different PCA-based geographic



grids. In this study significant interaction values were calculated for seed source  $\times$  test site for all response variables: height, diameter survival, bud flush, and bud set (results not shown). Nonetheless, our results suggest that the G $\times$ E interaction is not as important for growth traits, since all sites are correlated on PC1, but more important for bud flush, as indicated by different coefficients on PC2 (Table 2). The accurate prediction of adaptive zones by our approach strongly depends on accounting for G $\times$ E interaction with the chosen axes, and this representation in turn depends on the nature of the test sites. The ability to generalize adaptive suitability over a geographic area depends upon covering the range of environmental conditions represented by the chosen test sites.

With regard to the CANCOR approach, it is more difficult to determine the effect of G $\times$ E interaction, but a parallel trend to the PCA result is evident in CV2. All growth and survival variable correlations have the same negative sign, while bud flush and bud set have mixed signs on the same axis. Perhaps in some cases separate CANCORs per test site, or similar sites, might be justified, but synthesis of the separate analyses would be complex. Solution of this problem is beyond the scope of this study, but will provide the basis for a further research contribution.

## Conclusion

CANCOR offers an appealing statistical alternative to the PC–regression based methodology previously used to develop focal point seed zones. The former approach requires two independent steps, while CANCOR accomplishes the same result in one step. CANCOR simultaneously considers and relates the biological and climatic data sets together by maximizing the covariance between the two data sets (Gittins 1985). The CANCOR method retains all the climate variables, and the resulting seed zones are based on the weighted contributions of the full array. Thus, the orthogonal components are selected sequentially to give the best fit and optimize the predictive ability of the climate data. Information on relationships between biological and climate variables may be discarded by the independent PCA and regression analysis. The PCA segregates the biological data set into its various orthogonal components with no regard to their relationship to the climate data set, although there are exceptions to this when two or more variables coincidentally are jointly related to climate (Namkoong 1966). The approximate correspondence of the biological variables into growth potential (PC1), phenological timing (PC2), and greenhouse effects (PC3) allows a meaningful biological interpretation of the resulting patterns of variation, but is really an arbitrary mathematical determination that does not necessarily lend itself to explaining the true relationships between the biological and climate data sets (Gittins 1985). In the regression phase of the analysis, suitable equations are developed that, while seeking to explain the greatest amount of correlation between each PC axis and the climate array, may actually limit the explanatory power of the model.

## Acknowledgments

This work was funded primarily by an Ontario Living Legacy Trust Grant. Additional funding for this study was

provided by the Government of Ontario Interministerial Climate Change Committee. In-kind and additional financial support was received from Kimberly–Clark (now Neenah Paper), Weyerhaeuser, Tembec, Petawawa National Research Forest, the Ontario Tree Seed Plant, Bowater, and Greenmantle Forest Inc. Seed was provided by Dale Simpson (Canadian Forest Service (CFS)), Jean Beaulieu (CFS), Bob Sinclair (Ontario Ministry of Natural Resources (OMNR)), Ken Lennon (Neenah Paper), Janet Lane (Weyerhaeuser), and George Bruemmer (Tembec). Site preparation and upkeep were provided by Ken Lennon (Neenah Paper), Janet Lane (Weyerhaeuser), George Bruemmer (Tembec), Steve Deon (CFS), and Al Foley (OMNR). We also thank the Superior Woods Tree Improvement Association for providing financial support to the primary author. Further funding was provided by the Natural Sciences and Engineering Research Council of Canada (NSERC) in the form of a Discovery Grant awarded to the second author, an NSERC postgraduate scholarship to the primary author, and five NSERC undergraduate student research awards to summer field assistants. Thanks to Marilyn Cherry and Paul Gray (OMNR) for their involvement and support throughout this project. Thanks to Joan Lee for all her work in the greenhouse and other assistance she provided. Thanks to Al Stinson, Janet Lane, Cal Hartley, Megan Thompson, Paul Charrette, Randy Ford, Claire Riddell, Jenny Milson, and the many students who helped for their support and assistance. We also acknowledge the helpful comments and suggestions of Greg O'Neill and two anonymous reviewers.

## References

- Austin, M.P. 1968. An ordination study of a chalk grassland community. *J. Ecol.* **56**: 739–757.
- Blum, B.A. 1988. Variation in the phenology of bud flushing in white and red spruce. *Can. J. For. Res.* **18**: 315–319.
- Campbell, R.K. 1986. Mapped genetic variation of Douglas-fir to guide seed transfer in southwest Oregon. *Silvae Genet.* **35**: 85–96.
- Cherry, M., and Parker, W.H. 2003. Utilization of genetically improved stock to increase carbon sequestration. *Ont. For. Res. Inst. Min. Nat. Res. For. Res. Rep.* 160.
- ESRI. 2002. ArcGIS 8.3 [computer program]. Environmental Systems Research Institute Inc., Redlands, Calif.
- Gittins, R. 1985. Canonical analysis: a review with applications in ecology. Springer-Verlag, New York.
- Glimaret-Carpentier, C., Dray, S., and Pascal, J.P. 2003. Broad-scale biodiversity pattern of the endemic tree flora of the Western Ghats (India) using canonical correlation analysis of herbarium records. *Ecography*, **26**: 429–444.
- Hamrick, J.L., Godt, M.J.W., and Sherman-Broyles, S.L. 1992. Factors influencing levels of genetic diversity in woody plant species. *New For.* **6**: 95–104.
- Hills, G.A. 1961. The ecological basis for land-use planning. *Ont. Dept. Lands For. Res. Dept.* 46.
- Khalil, M.A.K. 1985. Genetic variation in eastern white spruce (*Picea glauca* (Moench) Voss) populations. *Can. J. For. Res.* **15**: 444–452.
- Lesser, M.R., and Parker, W.H. 2004. Genetic variation in *Picea glauca* for growth and phenological traits from provenance tests in Ontario. *Silvae Genet.* **53**: 141–148.



- Li, P., Beaulieu, J., Corriveau, A., and Bousquet, J. 1993. Genetic variation in juvenile growth and phenology in a white spruce provenance–progeny test. *Silvae Genet.* **42**: 52–60.
- Mackey, B.G., McKenney, D.W., Yang, Y., McMahon, J.P., and Hutchinson, M.F. 1996. Site regions revisited: a climatic analysis of Hills' site regions for the province of Ontario using a parametric method. *Can. J. For. Res.* **26**: 333–354.
- McKenney, D. 2004. Selected modeled climate data for point locations. Canadian Forest Service, Great Lakes Forestry Centre, Landscape Analysis and Application Section, Sault Ste. Marie, Ont.
- Morgenstern, E.K. 1996. Geographic variation in forest trees: Genetic basis and application of knowledge in silviculture. UBC Press, Vancouver.
- Morrison, D.F. 1990. Multivariate statistical methods. McGraw-Hill, New York.
- Namkoong, G. 1966. Statistical analysis of introgression. *Biometrics*, **22**: 488–502.
- Nienstaedt, H., and Teich, A. 1972. The genetics of white spruce. USDA For. Serv. Res. Pap. WO-15.
- Nienstaedt, H., and Zasada, J.C. 1990. *Picea glauca* (Moench) Voss. White spruce. In *Silvics of North America*. Vol. 1. Conifers. *Technical coordinators* R.M. Burns and B.H. Honkala. USDA Agric. Handb. 654. pp. 204–226.
- Parker, W.H. 1992. Focal point seed zones: site-specific seed zone delineation using geographic information systems. *Can. J. For. Res.* **22**: 267–271.
- Parker, W.H., and van Niejenhuis, A. 1996a. Seed zone delineation for jack pine in the former Northwest Region of Ontario using short-term testing and geographic information systems. Canadian Forest Service, Great Lakes Forestry Centre, Sault Ste. Marie, Ont. Technical Report NODA/NFP TR-35.
- Parker, W.H., and van Niejenhuis, A. 1996b. Regression-based focal point seed zones for *Picea mariana* from northwestern Ontario. *Can. J. Bot.* **74**: 1227–1235.
- Pélissier, R., Dray, S., and Sabatier D. 2001. Within-plot relationships between tree species occurrences and hydrological soil constraints: an example in French Guiana investigated through canonical correlation analysis. *Plant Ecol.* **162**: 143–156.
- Rehfeldt, G.E. 1984. Microevolution of conifers in the northern Rocky Mountains: a view from common gardens. In *Proceedings of the 8th North American Forest Biology Workshop*, Logan, Utah. *Edited by* R.M. Lanner. pp. 132–146.
- Rehfeldt, G.E. 1991. A model of genetic variation for *Pinus ponderosa* in the Inland Northwest (U.S.A.): applications in gene resource management. *Can. J. For. Res.* **21**: 1491–1500.
- Rehfeldt, G.E. 1994. Adaptation of *Picea engelmannii* populations to the heterogeneous environments of the Intermountain West. *Can. J. Bot.* **72**: 1197–1208.
- 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. *Ecol. Monogr.* **69**: 375–407.
- Rowe, J.S. 1972. Forest regions of Canada. *Can. For. Serv. Publ.* 1300.
- SAS Institute Inc. 2000. SAS system for Windows. SAS Institute Inc., Cary, N.C.
- St. Clair, J.B., Mandel, N.L., and Vance-Borland, K.W. 2005. Geneecology of Douglas fir in western Oregon and Washington. *Ann. Bot.* **96**: 1199–1214.
- Thompson, B. 1984. Canonical correlation analysis uses and interpretation. Sage Publications, Beverly Hills, Calif.
- van Buijtenen, J.P. 1992. Fundamental genetic principles. In *Handbook of quantitative forest genetics*. *Edited by* L. Fins, S.T. Freidman, and J.V. Brotschol. Kluwer Academic Press, Boston.
- Westfall, R.D. 1992. Developing seed transfer zones. In *Handbook of quantitative forest genetics*. *Edited by* L. Fins, S.T. Freidman, and J.V. Brotschol. Kluwer Academic Publishers, Boston. pp. 313–398.
- Wilkinson, L., Hill, M., and Vang, E. 1992. Systat: statistics. Version 5.2 [computer program]. Systat Inc., Evanston, Ill.

## Appendix A

**Table A1.** Seed source number, location, and geographic coordinates for 127 white spruce provenances in Ontario and Quebec.

No.	Location	Lat. (°N)	Long. (°W)	Elevation (m)
1	Cornwall	45.07	74.83	80
2	St-André Avellin	45.67	74.97	155
3	St-André Avellin	45.73	75.05	152
4	Camp 27	46.25	75.08	259
5	Thurso	45.62	75.23	100
6	Poupée	45.65	75.45	15
7	Lac Iroquois	46.03	75.57	213
8	Ruisseau Murphy	46.25	75.58	304
9	Val-Des-Bois	45.82	75.60	168
10	Augusta	44.83	75.63	100
11	Marlborough Township	45.12	75.80	90
12	Breckenridge	45.47	75.92	107
13	Wakefield	45.62	75.93	244
14	Bouchette	46.2	75.95	183
15	Aylwin	45.97	76.03	152
16	Grand-Remous	46.63	76.07	244
17	Antrim	45.32	76.18	121
18	Wyman	45.52	76.30	91
19	Lac Cayamant	46.15	76.33	274

**Table A1** (continued).

No.	Location	Lat. (°N)	Long. (°W)	Elevation (m)
20	Lac Du Faucard	46.85	76.35	305
21	Ladysmith	45.75	76.40	213
22	Lac Usborne	46.25	76.63	274
23	Renfrew	45.47	76.63	121
24	Silver Lake	44.82	76.68	180
25	Beachburg	45.68	76.80	137
26	Grove Creek	45.90	76.27	244
27	Riviere-Coulonge	46.35	76.87	274
28	Lac Cranson	45.83	76.95	122
29	Tyendinaga	44.33	77.13	107
30	Barrie	44.78	77.15	274
31	Sheenboro	45.97	77.25	152
32	Denbigh	45.08	77.28	305
33	Alice	45.77	77.28	150
34	Petawawa National Forestry Institute	45.98	77.45	160
35	Rolphon	46.17	77.67	183
36	Carlow	45.27	77.70	366
37	Marmora	44.55	77.75	229
38	Bancroft	45.1	77.97	396
39	Dummer	44.48	78.02	236
40	Anstruther Township	44.92	78.07	365
41	Haldimand	44.17	78.12	274
42	Whitney	45.53	78.27	396
43	Harvey	44.60	78.38	300
44	Canton Sébille	47.70	78.40	305
45	Lister	45.87	78.45	442
46	Canton Cameron	46.25	78.50	183
47	Osler	45.87	78.70	442
48	Lac Wawagosis	49.35	78.70	289
49	Lac Smith	46.72	78.83	335
50	Rutherglen	46.28	78.85	229
51	Baie Kelly	47.03	78.87	335
52	Mattawan Township	46.38	78.90	305
53	Eldon	44.47	78.92	280
54	Hindon Township	45.03	78.93	335
55	Canton Gaboury	47.33	79.00	305
56	Jocko Township	46.60	79.02	306
57	Lac Guay	47.20	79.03	305
58	Sinclair Township	45.47	79.08	370
59	Canton Mercier	46.78	79.12	305
60	Bonfield Township	46.23	79.13	245
61	Scott	44.12	79.18	290
62	Chisholm	46.13	79.27	275
63	Lac Hébécourt	48.53	79.30	224
64	Strong	45.78	79.42	381
65	Armour Township	45.62	79.42	300
66	Lac Labyrinthe	48.22	79.48	289
67	Notre-Dame-des-Quinze	47.58	79.50	213
68	Lorrain Township	47.25	79.52	240
69	Peck Township	45.48	78.75	460
70	Cobalt	47.03	79.68	306
71	McKellar	45.58	79.87	275
72	Englehart	47.87	79.92	215
73	East Mills	45.92	79.93	245
74	Erin	43.75	80.12	427
75	Osprey	44.35	80.33	503
76	Kirkland Lake	48.03	80.37	304
77	Bowman Township	48.48	80.42	290

**Table A1** (concluded).

No.	Location	Lat. (°N)	Long. (°W)	Elevation (m)
78	Bentinck	44.17	81.00	305
79	Clute 2	49.02	81.23	289
80	Pagwa	49.77	85.42	245
81	Fraserdale	49.03	81.58	215
82	Robb Township	48.58	81.62	290
83	St. Edmunds	45.25	81.63	206
84	Nairn Township	46.32	81.65	243
85	Gurney Township	49.05	82.25	215
86	Proctor	46.33	82.50	249
87	Cargill	49.3	82.70	289
88	Elizabeth Bay	45.83	82.75	191
89	Meldrum Bay	45.95	83.08	183
90	Arnott Township	49.62	84.58	275
91	Wawa	47.92	84.75	306
92	Bouchard	48.78	85.05	457
93	White River	48.62	85.32	305
94	Highway 11	49.77	85.47	236
95	Mobert Township	48.7	85.58	305
96	Strathearn	48.72	85.87	335
97	Manitouwadge	49.28	85.97	305
98	Caramat	49.6	86.15	305
99	Pic River	48.70	86.25	240
100	Kenogami	49.92	86.48	305
101	Nakina	50.2	86.78	335
102	False Creek	49.87	86.87	365
103	O'Sullivan	50.53	87.02	335
104	Long Lake	49.22	87.07	335
105	Maun/Anaconda Road	50.32	87.09	328
106	Eastnor	44.98	81.37	191
107	Terrace Bay	48.78	87.12	200
108	Grandpa Road	49.55	87.18	404
109	Jellicoe	49.7	87.42	365
110	Parks Lake	49.47	87.57	460
111	South Onaman River	50.03	87.65	305
112	Mountain Bay	48.91	87.77	195
113	Auden	50.15	87.88	335
114	Beardmore	49.55	88	365
115	Limestone	49.07	88.02	245
116	Nipigon	49.20	88.22	229
117	Stewart Lake	48.98	88.54	267
118	Chief Bay	49.05	89.05	275
119	Waweig Lake	50.15	89.12	305
120	Lakehead University woodlot	48.65	89.41	457
121	Pigeon River	48.02	89.65	306
122	Twist Lake	49.37	89.75	425
123	Shabaqua	48.62	89.90	410
124	Shebandowan	48.62	90.18	459
125	Upsala	49.07	90.52	489
126	Eva Lake	48.07	91.42	428
130	King	44	79.67	240