TECHNICAL REPORT 127

An Analysis of a Site Index – Biogeoclimatic Ecosystem Classification (SIBEC) Analysis

2019

BRITISH COLUMBIA

127

An Analysis of a Site Index – Biogeoclimatic Ecosystem Classification (SIBEC) Analysis

Gord Nigh

BRITISH COLUMBIA

The use of trade, firm, or corporation names in this publication is for the information and convenience of the reader. Such use does not constitute an official endorsement or approval by the Government of British Columbia of any product or service to the exclusion of any others that may also be suitable. Contents of this report are presented for discussion purposes only. Funding assistance does not imply endorsement of any statements or information contained herein by the Government of British Columbia. Uniform Resource Locators (URLs), addresses, and contact information contained in this document are current at the time of printing unless otherwise noted.

ISBN 978-0-7726-7881-2 – Print version
ISBN 978-0-7726-7882-9 – Digital version

Citation
Nigh, G. 2019. An analysis of a Site Index – Biogeoclimatic Ecosystem Classification (SIBEC) analysis. Prov. B.C., Victoria, B.C. Tech. Rep. 127. www.for.gov.bc.ca/hfd/pubs/Docs/Tr/Tr127.htm

Gord Nigh
B.C. Ministry of Forests, Lands, Natural Resource Operations and Rural Development
P.O. Box 9512, Stn. Prov. Govt.
Victoria, BC V8W 9C2

Copies of this report may be obtained, depending upon supply, from:
Crown Publications, Queen's Printer
2nd Floor, 563 Superior Street
Victoria, BC V8W 9V7
1-800-663-6105
www.crownpub.bc.ca

For more information on other publications in this series, visit www.for.gov.bc.ca/scripts/hfd/pubs/hfdcatalog/index.asp

© 2019 Province of British Columbia

When using information from this report, please cite fully and correctly.

ABSTRACT

The Site Index – Biogeoclimatic Ecosystem Classification (SIBEC) model allows site index, which is a measure of site productivity, to be estimated from Biogeoclimatic Ecosystem Classification site series and species. SIBEC site index estimates and their standard errors are published online; however, a simplistic approach to calculating these standard errors was taken. New growth intercept models were developed from stem analysis data and were applied to SIBEC data collected in the Coastal Western Hemlock biogeoclimatic zone, very dry maritime subzone, 01 site series (CWHxm2/01) for Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco var. *menziesii*). A more appropriate approach to calculating the standard error of the mean site index estimate was derived and applied to the SIBEC data. The formula for the standard error requires knowledge of the covariances between growth intercept model predictions, which are unknown except when only one model is used to make the predictions. The results indicated that the corrected standard errors were substantially lower than the standard errors calculated using the current method. A typical SIBEC site index estimate is made up of data from growth intercept models, site index models, and stem analysis. Combining these data sources greatly complicates the standard error calculations, which makes their computation even more difficult than when only growth intercept models are used to estimate site index.

ACKNOWLEDGEMENTS

I thank Peter Ott of the British Columbia Ministry of Forests, Lands, Natural Resource Operations and Rural Development for many discussions that helped me better understand the problem and work out the solution. I also thank Graham Hawkins, also with the British Columbia Ministry of Forests, Lands, Natural Resource Operations and Rural Development, for valuable review comments.

CONTENTS

Abstract iii
Acknowledgements iii
1 **Introduction** 1
2 **Data** 2
2.1 Growth Intercept Model Development 2
2.2 SIBEC Model Development 2
3 **Analysis** 2
3.1 Growth Intercept Model 2
3.2 SIBEC Model 3
3.3 Standard Error of the Estimated Mean 3
3.4 Application of the Standard Error of the Estimated Mean 5
4 **Results** 5
5 **Discussion** 7
6 **Conclusions** 9
Literature Cited 9

APPENDIX

1 Derivation of var $(\widehat{si}_i)$ and cov $(\widehat{si}_i, \widehat{si}_j)$ 10

TABLES

1 Parameter estimates, their variances and covariances, and the residual variance for the growth intercept models from age 1 to 50 6
2 Standard errors of the estimated mean site index and intermediate calculations from Equation 8 for three scenarios 7

1 INTRODUCTION

The Biogeoclimatic Ecosystem Classification (BEC) system in British Columbia groups similar ecosystems into categories of a hierarchical classification system. At the site level, the most common category is the site series. Site series have similar vegetation potential and environmental properties, particularly soil moisture and soil nutrient regime (Green and Klinka 1994). This uniformity of climate and soil properties suggests that sites within a site series have similar productivity levels. This, in turn, has led to the development of the Site Index – Biogeoclimatic Ecosystem Classification (SIBEC) model for estimating site index. A SIBEC model predicts site index from tree species and site series.

Site index is a measure of forest site productivity, and in British Columbia it is defined as the height of a site tree at age 50 (all ages in this report are breast height ages). The height growth of site trees reflects the potential productivity of a site. Site trees can be identified using the guidelines in the *SIBEC Data and Sampling Standards* (B.C. Ministry of Forests and Range 2009).

In addition to the SIBEC model, site index is commonly estimated with growth intercept (e.g., Nigh 1997) or site index (e.g., Bruce 1981) models. Growth intercept models are of the form:

$$si = f(h, A; \boldsymbol{b})$$

and site index models are of the form:

$$h = g(si, A, \boldsymbol{b})$$

where si = site index (m), h = height (m), A = age (yr), and $\boldsymbol{b}$ is a vector of model parameters, which is unique for each model. For the growth intercept models, the parameter values depend on the age of the sample tree and are available only for $A \leq 50$. More details are provided by Nigh (1997). Site index models predict height rather than site index, and most site index models in British Columbia cannot be mathematically inverted to predict site index from height and age. A numerical routine must be employed to find the site index that gives the stated height at the specified age.

The SIBEC model is developed by establishing SIBEC sample plots (B.C. Ministry of Forests and Range 2009) and determining the site series and height and age of the site tree(s). Site index is predicted with either a growth intercept model or a site index model, or, less commonly, by performing a stem analysis on the site tree and measuring the site index directly. The second-approximation SIBEC site index estimate[1] for a given species and site series is the mean site index from sample plots that are derived from a combination of stem analysis data, growth intercept model predictions, and/or site index model predictions. Some work has been done to capture the within–site series variability in site index, but it met with little success (Nigh 2010). Therefore, calculating the mean site index is the standard method of analysis.

1 First-approximation SIBEC site index estimates were derived from ANOVA models and reported as mid-points of site index classes (Mah and Nigh 2003).

The SIBEC site index values are published online.[2] In addition to the mean site index, the sample size and the standard error of the mean site index are reported for the second-approximation site index estimates. The standard error of the mean is based on the predicted or observed (for stem analysis data) site indices. For site index estimates that are based partially or wholly on model estimates, the reported standard error of the mean has some deficiencies, as will be shown. This report derives and demonstrates a more appropriate standard error of the estimated mean for coastal Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco var. *menziesii*).

2 DATA

The growth intercept models and the SIBEC site index estimate developed in this study were based on data from two projects and consist of plots with one tree each.

2.1 Growth Intercept Model Development

The publication on the growth intercept models for coastal Douglas-fir (Nigh 1997) does not report the covariances between the parameter estimates, which, as will be shown, are needed to develop the standard error of the estimated mean. Consequently, new growth intercept models were developed with the Nigh (1997) data set to obtain the missing information. The growth intercept models developed in this study are not identical to the Nigh (1997) models due to slight differences in the data sets, but they are very similar. In total, 47 Douglas-fir stem analysis plots were available for analysis. The age and height of the sample trees were measured from the pith nodes, starting at the node below breast height (age 0) up to the node at age 50. The site index associated with a sample tree is the height of the tree at age 50.

2.2 SIBEC Model Development

A new SIBEC site index estimate was developed for Douglas-fir in the Coastal Western Hemlock biogeoclimatic zone, very dry maritime subzone variant, 01 site series (CWHxm2/01). This estimate is based solely on the site indices from the above-mentioned growth intercept models to demonstrate the calculation of the standard error of the estimated mean. Thirty-five Douglas-fir stem analysis plots were established in the CWHxm2/01 site series (Nigh 2010). After stem splitting, the heights of the sample trees at ages 0–50 were recorded as input to the growth intercept models.

3 ANALYSIS

3.1 Growth Intercept Model

The growth intercept model is:

$$si_i = 1.3 + e^{\beta_{0,A}} \times gi_{i,A}{}^{\beta_{1,A}} + \varepsilon_i \tag{1}$$

where si_i is the site index (m) for tree i, $\beta_{0,A}$ and $\beta_{1,A}$ are model parameters, which depend on the age A (yr) ($1 \leq A \leq 50$) of the sample tree, $gi_{i,A}$ (cm/yr) is the growth intercept for tree i at age A, and ε_i is the random error term for

2 www2.gov.bc.ca/gov/content/environment/plants-animals-ecosystems/ecosystems/sibec

tree i with the usual nonlinear regression assumptions of normality, homoscedasticity, and independence (Seber and Wild 1989). The growth intercept is given by:

$$gi_{i,A} = 100 \times \frac{h_{i,A} - 1.3}{A - A_{i,p}} \tag{2}$$

where $h_{i,A}$ is the height (m) of sample tree i at age A and

$$A_{i,p} = \frac{1.3 - h_{i,0}}{h_{i,1} - h_{i,0}} \tag{3}$$

$A_{i,p}$ is the proportion of height growth for tree i between ages 0 and 1 that occurred below breast height. Fitting model (1) to the Nigh (1997) data set resulted in 50 sets of model parameters, corresponding to ages 1–50. The predicted site indices from the growth intercept models are given by Equation 4:

$$\hat{si}_i = 1.3 + e^{b_{0,A}} \times gi_{i,A}{}^{b_{1,A}} \tag{4}$$

where $\hat{si}_i$ is the predicted site index for tree i, and $b_{0,A}$ and $b_{1,A}$ are the estimated model parameters, which depend on the age A.

3.2 SIBEC Model

A SIBEC site index estimate is the mean site index of SIBEC sample trees for a given species and site series. The estimate is calculated with data that are either observed site indices (from stem analysis trees of age ≥ 50) or predicted site indices (from typical SIBEC plots and stem analysis trees whose age is < 50). The current method of calculating SIBEC site index estimates ($\overline{si}$) and their standard error ($s_{\overline{si}}$) are given by Equations 5 and 6, respectively:

$$\overline{si} = \frac{1}{m}\sum_{i=1}^{m} si_i \tag{5}$$

$$s_{\overline{si}} = \sqrt{\frac{1}{m \times (m-1)}\sum_{i=1}^{m}\left(si_i - \overline{si}\right)^2} \tag{6}$$

where si_i is the observed or predicted site index for tree i, and m is the number of sample trees. These estimates are correct when the site index is observed, such as when all the SIBEC data are from stem-analyzed trees older than age 50. However, typically for SIBEC plots, most site indices are predicted, not observed. The correct SIBEC estimate ($\overline{si}$) is still calculated as in Equation 5 but with the predicted site index ($\hat{si}_i$) instead of the observed site index because the predicted site index is an unbiased estimator of the true site index. However, the predicted site index includes a prediction error, and the model parameter estimates have a variance and covariance, all of which should be accounted for in the calculation of $s_{\overline{si}}$ in Equation 6.

3.3 Standard Error of the Estimated Mean

The formula for the standard error of the estimated mean (the SIBEC site index estimate) when the data consist of growth intercept model predictions, which accounts for the above-mentioned deficiencies, is derived here.

The error in the predicted site index for tree i (ε_i) is given by Equation 7:

$$\varepsilon_i = si_i - \widehat{si}_i \tag{7}$$

where si_i is the true (but unknown) site index, and $\widehat{si}_i$ is the predicted site index. If we condition on the m available sample trees as well as the fitted growth intercept models, the standard error of the SIBEC site index estimate is the square root of the variance of the mean of the errors, ε_i. This variance is derived in Equation 8.

$$\begin{aligned}
\operatorname{var}\left(\frac{1}{m}\sum_{i=1}^{m}\varepsilon_i\right) &= \frac{1}{m^2}\operatorname{var}\left(\sum_{i=1}^{m}\varepsilon_i\right) \\
&= \frac{1}{m^2}\operatorname{var}\left(\sum_{i=1}^{m} si_i - \sum_{i=1}^{m}\widehat{si}_i\right) \\
&= \frac{1}{m^2}\left[\sum_{i=1}^{m}\operatorname{var}\left(si_i\right) + \sum_{\substack{i,j=1\\ i\neq j}}^{m}\operatorname{cov}\left(si_i, si_j\right) + \sum_{i=1}^{m}\operatorname{var}\left(\widehat{si}_i\right) + \sum_{\substack{i,j=1\\ i\neq j}}^{m}\operatorname{cov}\left(\widehat{si}_i, \widehat{si}_j\right) - 2\sum_{i,j=1}^{m}\operatorname{cov}\left(si_i, \widehat{si}_j\right)\right] \\
&= \frac{1}{m^2}\left[\sum_{i=1}^{m}\operatorname{var}\left(si_i\right) + \sum_{i=1}^{m}\operatorname{var}\left(\widehat{si}_i\right) + \sum_{\substack{i,j=1\\ i\neq j}}^{m}\operatorname{cov}\left(\widehat{si}_i, \widehat{si}_j\right)\right]
\end{aligned} \tag{8}$$ [3]

where m is the number of observations (SIBEC sample trees), $\operatorname{var}(si_i)$ is the variance of the true site index for tree i, $\operatorname{var}(\widehat{si}_i)$ is the variance of the predicted site index for tree i, and $\operatorname{cov}(\widehat{si}_i, \widehat{si}_j)$ is the covariance between the predicted site indices for trees i and j. $\operatorname{Var}(si_i)$ is estimated with the residual variance (mean squared error) from the regression analysis for the growth intercept model corresponding to the age of tree i. An approximation to $\operatorname{var}(\widehat{si}_i)$ based on a Taylor series expansion is given by Equation 9 (see Appendix 1 for a derivation of this formula):

$$\begin{aligned}
\operatorname{var}(\widehat{si}_i) \approx\ & (e^{b_{0,A}} \times gi_{i,A}{}^{b_{1,A}})^2 \times \operatorname{var}(b_{0,A}) + 2 \times (e^{b_{0,A}} \times gi_{i,A}{}^{b_{1,A}}) \times \\
& [e^{b_{0,A}} \times gi_{i,A}{}^{b_{1,A}} \times \ln(gi_{i,A})] \times \operatorname{cov}(b_{0,A}, b_{1,A}) + \\
& [e^{b_{0,A}} \times gi_{i,A}{}^{b_{1,A}} \times \ln(gi_{i,A})]^2 \times \operatorname{var}(b_{1,A})
\end{aligned} \tag{9}$$

where $\operatorname{var}(b_{0,A})$ is the variance of parameter $b_{0,A}$, $\operatorname{cov}(b_{0,A}, b_{1,A})$ is the covariance between parameters $b_{0,A}$ and $b_{1,A}$, $\operatorname{var}(b_{1,A})$ is the variance of parameter $b_{1,A}$ (all the preceding values are obtained from the fitting of the growth intercept model [Equation 1]), and $\widehat{si}_i$ is the predicted site index for tree i.

The covariances between predicted site indices (the last term in Equation 8: $\operatorname{cov}(\widehat{si}_i, \widehat{si}_j)$) are generally unknown. However, if the ages (A) of the sample

3 $\operatorname{Cov}(si_i, \widehat{si}_j) = 0$ since the true site indices are independent of the predicted site indices (when the model is developed from different data) and $\operatorname{cov}(si_i, si_j) = 0$ for all $i \neq j$ since the SIBEC sample trees are independent observations.

trees are all the same, then these covariances (see Appendix 1) are given by Equation 10:

$$\begin{aligned}\mathrm{cov}(\hat{si}_i, \hat{si}_j) \approx\ & e^{b_{0,A}} \times gi_{i,A}{}^{b_{1,A}} \times e^{b_{0,A}} \times gi_{j,A}{}^{b_{1,A}} \times \\ & \{\mathrm{var}(b_{0,A}) + [\ln(gi_{i,A}) + \ln(gi_{j,A})] \times \\ & \mathrm{cov}(b_{0,A}, b_{1,A}) + \ln(gi_{i,A}) \times \ln(gi_{j,A}) \times \mathrm{var}\,(b_{1,A})\}\end{aligned} \tag{10}$$

3.4 Application of the Standard Error of the Estimated Mean

The SIBEC estimate for the CWHxm2/01 site series and its standard error were calculated from the Nigh (2010) data for three scenarios: (1) all trees were assumed to be 10 years old (largest variance scenario), (2) a more likely scenario in which the trees have a variety of ages between 10 and 49 years, and (3) all trees were assumed to be 49 years old (smallest variance scenario). For the second scenario, the ages were randomly chosen from a discrete uniform distribution between 10 and 49. The upper and lower limits of age (10 and 49) were chosen because sample trees younger than age 10 are not suitable for SIBEC modelling (B.C. Ministry of Forests and Range 2009), and trees at age 50 do not require a model to estimate site index because their measured heights are the site index. The growth intercept (Equation 2) for these trees was then calculated. For each of the three scenarios, the site index for each tree was predicted with Equation 4 using the appropriate parameter estimates and the growth intercept. The standard error of the estimate was calculated with Equation 8. Since the covariance of the model estimates cannot be obtained for the second scenario (tree ages randomly chosen), for demonstration purposes this covariance was assumed to be the average of the covariances for the age 10 and 49 scenarios.

A bootstrap approach was used to analyze the SIBEC estimate and its standard error. One thousand bootstrap samples of 35 observations were drawn randomly with replacement from the data for the three scenarios. The mean SIBEC estimate and its standard error were calculated with Equations 4 and 8, respectively, for each sample and then averaged over all 1000 bootstrap samples.

4 RESULTS

The results from the re-fitting of the growth intercept models are presented in Table 1. The table shows the parameter estimates, their variances and covariances, and the residual variance.

Table 2 presents averages from the 1000 bootstrap samples. The mean site index and its standard error for the three scenarios (i.e., all trees are at age 10, trees have randomly chosen ages between 10 and 49, and all trees are at age 49) are shown in Table 2, as are the intermediate calculations for the standard error (i.e., the three terms in Equation 8).

The mean and variance of the measured site indices from the 35 trees in the Nigh (2010) data set are 34.57 m and 9.304, respectively, giving a standard error of the mean of 0.516. These values were calculated with Equations 5 and 6. Doing these same calculations with the predicted site indices instead of the measured site indices gives means of 34.91, 35.10, and 34.66, and standard errors of the mean of 0.666, 0.534, and 0.506 for the age 10 data, random age data, and age 49 data, respectively.

TABLE 1 *Parameter estimates, their variances and covariances, and the residual variance (s^2) for the growth intercept models from age (A) 1 to 50*

A (yr)	$b_{0,A}$	$b_{1,A}$	var($b_{0,A}$)	cov($b_{0,A}$, $b_{1,A}$)	var($b_{1,A}$)	s^2
1	1.31509457	0.55037526	0.06805853	−0.01699510	0.00427001	20.31836645
2	0.96366813	0.62340122	0.06494052	−0.01581408	0.00387044	15.99688748
3	0.96131489	0.61544253	0.05515192	−0.01323155	0.00319088	13.98810473
4	0.92281382	0.61832160	0.05235333	−0.01242684	0.00296470	13.02462339
5	0.84958294	0.63098718	0.04605124	−0.01084047	0.00256441	11.10531635
6	0.75808458	0.64758160	0.04154814	−0.00970036	0.00227534	9.51322461
7	0.61562002	0.67718551	0.03820170	−0.00886714	0.00206698	8.02543872
8	0.51596484	0.69646052	0.03858223	−0.00890597	0.00206402	7.61630564
9	0.41026503	0.71832838	0.03744210	−0.00861095	0.00198783	6.96035518
10	0.27975778	0.74675186	0.03679891	−0.00844408	0.00194440	6.33919913
11	0.18680345	0.76605857	0.03615319	−0.00827347	0.00189959	5.88607133
12	0.08480722	0.78814078	0.03596297	−0.00821655	0.00188308	5.50250944
13	0.01246800	0.80400310	0.03594823	−0.00820625	0.00187890	5.28679372
14	−0.06348316	0.82055337	0.03455880	−0.00788078	0.00180229	4.89183691
15	−0.08818260	0.82549584	0.03410858	−0.00777140	0.00177569	4.77971967
16	−0.14753928	0.83863948	0.03226802	−0.00734801	0.00167789	4.39294734
17	−0.20432166	0.85149043	0.03074053	−0.00699905	0.00159781	4.05444105
18	−0.25478273	0.86311459	0.02974275	−0.00677264	0.00154619	3.82533415
19	−0.28013129	0.86897386	0.02779976	−0.00633028	0.00144519	3.54841641
20	−0.30964052	0.87646823	0.02527488	−0.00575973	0.00131591	3.19547365
21	−0.32862589	0.88100279	0.02348732	−0.00535322	0.00122319	2.94230432
22	−0.36021154	0.88900787	0.02163468	−0.00493505	0.00112854	2.67807701
23	−0.36700825	0.89106398	0.01969972	−0.00449578	0.00102858	2.45139460
24	−0.37963751	0.89464497	0.01782396	−0.00407053	0.00093193	2.21175296
25	−0.38824311	0.89750516	0.01653535	−0.00377980	0.00086618	2.04755567
26	−0.40080088	0.90140463	0.01568243	−0.00358881	0.00082332	1.93012974
27	−0.40071530	0.90230321	0.01493123	−0.00342025	0.00078542	1.84425182
28	−0.41601001	0.90655097	0.01429821	−0.00327787	0.00075331	1.75178517
29	−0.44870494	0.91501796	0.01319574	−0.00302825	0.00069663	1.58826749
30	−0.46287491	0.91915859	0.01247642	−0.00286588	0.00065989	1.49002883
31	−0.45963654	0.91951192	0.01157622	−0.00266216	0.00061369	1.38842459
32	−0.48517529	0.92672826	0.01055185	−0.00243003	0.00056096	1.25238292
33	−0.50756631	0.93306551	0.00908217	−0.00209411	0.00048399	1.07003420
34	−0.52742065	0.93871786	0.00790891	−0.00182560	0.00042240	0.92603851
35	−0.55264423	0.94575083	0.00730755	−0.00168894	0.00039126	0.84520852
36	−0.55626111	0.94773504	0.00626158	−0.00144888	0.00033604	0.72508717
37	−0.57620702	0.95355339	0.00531282	−0.00123087	0.00028583	0.61008603
38	−0.58573802	0.95681396	0.00417917	−0.00096925	0.00022531	0.47921398
39	−0.60344472	0.96217285	0.00358859	−0.00083335	0.00019397	0.40874310
40	−0.61820425	0.96679977	0.00308965	−0.00071837	0.00016741	0.34955683
41	−0.64331565	0.97390580	0.00255700	−0.00059530	0.00013890	0.28577656
42	−0.65142899	0.97694469	0.00216454	−0.00050452	0.00011786	0.24111684
43	−0.65424150	0.97891506	0.00178327	−0.00041620	0.00009736	0.19848975
44	−0.67368086	0.98482614	0.00133861	−0.00031286	0.00007328	0.14791415
45	−0.68330453	0.98835717	0.00094797	−0.00022184	0.00005203	0.10469563
46	−0.70361584	0.99475598	0.00065989	−0.00015468	0.00003634	0.07235555
47	−0.71377426	0.99873546	0.00048364	−0.00011355	0.00002672	0.05281892
48	−0.71515713	1.00045004	0.00029696	−0.00006982	0.00001645	0.03246258
49	−0.72074393	1.00322198	0.00013564	−0.00003194	0.00000754	0.01481468
50	−0.71511970	1.00319848	0.00009594	−0.00002262	0.00000534	0.01053303

TABLE 2 *Standard errors of the estimated mean site index and intermediate calculations from Equation 8 for three scenarios (age = 10, age is randomly chosen between 10 and 49, and age = 49). Values shown are averages from 1000 bootstrap samples. Symbol (^) signifies a predicted value.*

Age (yr)	$\sum \mathrm{var}(si_i)$	$\sum \mathrm{var}(\widehat{si}_i)$	$\sum \mathrm{cov}(\widehat{si}_i, \widehat{si}_j)$	SE
10	221.872	7.112	172.359	0.572
Random	72.124	2.109	86.379[a]	0.362[a]
49	0.519	0.014	0.398	0.028

a Covariance cannot be calculated. Values shown are an estimate based on averaging the covariances from the age 10 and age 49 data.

5 DISCUSSION

The results include material that may not be readily apparent.

- The mean site indices (i.e., the values that become the SIBEC site index estimate) of the measured site indices and the predicted site indices are effectively the same, although the predicted mean site indices are somewhat higher than the measured mean site index. For most applications, the differences are insignificant. This provides evidence that the growth intercept models for coastal Douglas-fir that were developed in this study give reasonable site index predictions.
- The standard error of the estimated mean site index calculated with Equation 6 (i.e., the current method) is approximately the same for the measured site index (0.516) as for the three scenarios with the predicted site indices (0.666, 0.534, and 0.506). It may, therefore, be tempting to assume that the current method of calculating the standard error of the estimated mean is reliable for predicted site indices, but, as mentioned previously, the standard errors for the predicted site indices are deficient in some aspects.
- $\mathrm{Var}(\widehat{si}_i)$ represents variation in the estimated site index due to the model parameters being estimates themselves and does not include variation around the conditional mean (i.e., the model). The variation around the conditional mean is estimated with $\mathrm{var}(si_i)$. In this case, the former is much smaller than the latter. The variance around the conditional mean is unknown but is assumed to be the same as for the model fitting data.
- The correlations between predicted site indices are non-zero because all the growth intercept models were developed from the same data set, which resulted in any peculiarities or deviations from the true relationship between site index and the growth intercept being reflected in all the model predictions. Had these individual models been fitted with independent data sets (i.e., a completely different data set for each model), these covariances would be zero. However, this way of fitting the models is cost-prohibitive.
- The growth intercept models clearly become more accurate at ages closer to 50; all variance and covariance terms in Equation 8 decrease as age increases. This result is intuitive. It is much easier to predict the height of a tree at age 50 from its height at age 49, say, than it is from its height at age 10.

- The covariances between model predictions are a significant source of variation and should not be ignored in the calculation of the standard error of the mean. Despite their importance, these covariances are, unfortunately, difficult to obtain.
- The current method of calculating the standard error of the estimated mean ignores some important sources of variation and covariation. Its formula is not contained in Equation 8, which indicates that more than just a simple adjustment is required to correct the standard errors. Therefore, consideration should be given to discontinuing reporting the standard error when the SIBEC site index estimate is based wholly or even partially on model predictions.

The approaches for calculating the standard error of the mean presented in Equations 6 and 8 are both rooted in model-based inference. The simpler approach in Equation 6 assumes that we have little knowledge about how the site indices (si_i) are generated. However, if we know that the site indices are estimated with a model, then we can take the more appropriate approach to calculating the standard error presented in this report (Equation 8).

One approach to eliminating the need for cross-model correlations is to fit one model across all ages instead of 50 models (one for each age). This was done for lodgepole pine (*Pinus contorta* var. *latifolia* Dougl. ex Loud.) (Nigh 1995). One drawback of doing this is the added complexity in calculating the variances and covariances because of the additional parameters in the model. Furthermore, this does not help when SIBEC estimates are based on site indices that are derived from sources other than growth intercept models.

Growth intercept models were developed as a preliminary step to determining the SIBEC site index estimate. Generally, these models are developed once and then remain static unless there is some impetus to re-fit them, such as to improve their fit or (in this case) to obtain additional information about the models. Therefore, the errors associated with the models are also static. If the models are refitted, the SIBEC site index estimate will also change, as will its standard error, and these should be re-calculated.

A typical SIBEC sample will contain site indices from a mixture of data sources: growth intercept model predictions, site index model predictions, and measured site indices from stem analysis data. When the site indices are predicted from a site index model, the standard error of the mean site index cannot be derived as shown in this report because the predicted site indices are usually obtained numerically with an iterative technique rather than being predicted directly from a model. Hence, the variance and covariances of the estimated site indices are unknown. Given that the covariances between growth intercept model predictions are unknown, except in specific cases, and that variances and covariances from the site index models are unobtainable, standard errors of the means usually cannot be calculated using the techniques in this report.

Standard statistical tests, such as the *t* test, and confidence intervals formed from the standard error of the mean are predicated on the data being independently and identically distributed. These tests and confidence intervals are therefore not valid when using the standard error of the predicted mean derived in this report, even if the covariances could be determined.

The formula for the standard error of the mean is not specific to the growth intercept models. With appropriate modifications, the formulae can be applied in other situations where a model is used to predict a parameter of interest.

6 CONCLUSIONS

The current approach to calculating the standard error of a SIBEC site index has some deficiencies unless all the data in the calculation of the mean consist of measured site indices. The standard error may be difficult to obtain because the covariances between the site index predictions are required in the calculations. Except in the simplest case where all site index estimates are from one growth intercept model, it is difficult to obtain the necessary covariances. Obtaining variances and covariances becomes awkward when a mix of growth intercept models and site index models is required to estimate site index. Statistical tests and confidence intervals based on the calculated variance cannot be reliably constructed due to a violation of statistical assumptions of constant variance and independence among the data.

LITERATURE CITED

British Columbia Ministry of Forests and Range. 2009. SIBEC: sampling and data standards. Victoria, B.C. www2.gov.bc.ca/assets/gov/environment/plants-animals-and-ecosystems/ecosystems/sibec-documents/standards.pdf

Bruce, D. 1981. Consistent height-growth and growth-rate estimates for remeasured plots. For. Sci. 27:711–725.

Green, R.N. and K. Klinka. 1994. A field guide to site identification and interpretation for the Vancouver Forest Region. B.C. Min. For., Res. Br., Victoria, B.C.

Mah, S. and G.D. Nigh. 2003. SIBEC site index estimates in support of forest management in British Columbia. B.C. Min. For., Res. Br., Victoria, B.C. Tech. Rep. 004. www.for.gov.bc.ca/hfd/pubs/Docs/Tr/Tr004.pdf

Neter, J., W. Wasserman, and M.H. Kutner. 1990. Applied linear statistical models. Irwin, Boston, Mass.

Nigh, G. 1995. Variable growth intercept models for lodgepole pine in the Sub-Boreal Spruce biogeoclimatic zone, British Columbia. B.C. Min. For., Res. Br., Victoria, B.C. Res. Rep. 02.

_____. 1997, A growth intercept model for coastal Douglas-fir. B.C. Min. For., Res. Br., Victoria, B.C. Res. Rep. 10.

_____. 2010. A closer look at site index – biogeoclimatic site series correlations: Douglas-fir in the Coastal Western Hemlock zone, xm2 variant, 01 site series. For. Chron. 86:477–483.

Seber, G.A.F. and C.J. Wild. 1989. Nonlinear regression. John Wiley & Sons, Toronto, Ont.

The growth intercept model is:

$$\widehat{si}_i = 1.3 + e^{b_{0,A}} \times gi_{i,A}{}^{b_{1,A}}$$

and for the purposes of these derivations it is linearized by a first-order Taylor series expansion around the point $b_{0,A} = \hat{b}_{0,A}$ and $b_{1,A} = \hat{b}_{1,A}$, where $\hat{b}_{0,A}$ and $\hat{b}_{1,A}$ are the realized parameter estimates for parameters $b_{0,A}$ and $b_{1,A}$, respectively. For clarity of presentation, let $g = e^{\hat{b}_{0,A}} \times gi_{i,A}{}^{\hat{b}_{1,A}}$, then:

$$\begin{aligned}
\widehat{si}_i &= 1.3 + g \approx 1.3 + g + g \times (b_{0,A} - \hat{b}_{0,A}) + g \times \ln(gi_{i,A}) \times (b_{1,A} - \hat{b}_{1,A}) \\
&= 1.3 + g - g \times \hat{b}_{0,A} + g \times b_{0,A} - g \times \ln(gi_{i,A}) \times \hat{b}_{1,A} + g \times \ln(gi_{i,A}) \times b_{1,A} \\
&= 1.3 + g \times [1 - \hat{b}_{0,A} - \ln(gi_{i,A}) \times \hat{b}_{1,A}] + g \times b_{0,A} + g \times \ln(gi_{i,A}) \times b_{1,A} \\
&= c + g \times b_{0,A} + g \times \ln(gi_{i,A}) \times b_{1,A}
\end{aligned}$$

where $c = 1.3 + e^{\hat{b}_{0,A}} \times gi_{i,A}{}^{\hat{b}_{1,A}} \times [1 - \hat{b}_{0,A} - \ln(gi_{i,A}) \times \hat{b}_{1,A}]$

The term c is a constant for fixed $gi_{i,A}$. Now, approximations to var($\widehat{si}_i$) and cov($\widehat{si}_i$, $\widehat{si}_j$) can be derived using formulae for the basic properties of variances and covariances (Neter et al. 1990):

$$\begin{aligned}
\mathrm{var}(\widehat{si}_i) &\approx g^2 \times \mathrm{var}(b_{0,A}) + 2 \times g^2 \times \ln(gi_{i,A}) \times \mathrm{cov}(b_{0,A}, b_{1,A}) + g^2 \times \ln(gi_{i,A})^2 \times \mathrm{var}(b_{1,A}) \\
&= g^2 \times [\mathrm{var}(b_{0,A}) + 2 \times \ln(gi_{i,A}) \times \mathrm{cov}(b_{0,A}, b_{1,A}) + \ln(gi_{i,A})^2 \times \mathrm{var}(b_{1,A})]
\end{aligned}$$

For the covariances, there are model predictions for two trees (indexed by i and j). These trees have ages A_i and A_j, and their site indices are predicted by:

$$\begin{aligned}
\widehat{si}_i &= 1.3 + e^{b_{0,A_i}} \times gi_{i,A_i}{}^{b_{1,A_i}} \\
\widehat{si}_j &= 1.3 + e^{b_{0,A_j}} \times gi_{i,A_j}{}^{b_{1,A_j}}
\end{aligned}$$

Again, for clarity, let $g_i = e^{\hat{b}_{0,A_i}} \times gi_{i,A_i}{}^{\hat{b}_{1,A_i}}$ and $g_j = e^{\hat{b}_{0,A_j}} \times gi_{j,A_j}{}^{\hat{b}_{1,A_j}}$. Then based on the above Taylor series expansion and properties of covariances:

$$\begin{aligned}
\mathrm{cov}(\widehat{si}_i, \widehat{si}_j) &\approx \mathrm{cov}[c_i + g_i \times b_{0,A_i} + g_i \times \ln(gi_{i,A_i}) \times b_{1,A_i}, c_j + g_j \times b_{0,A_j} + g_j \times \\
&\qquad \ln(gi_{j,A_j}) \times b_{1,A_j}] \\
&= g_i \times g_j \times \mathrm{cov}(b_{0,A_i}, b_{0,A_j}) + g_i \times g_j \times \ln(gi_{j,A_j}) \times \mathrm{cov}(b_{0,A_i}, b_{1,A_j}) + g_i \times \ln(gi_{i,A_i}) \times \\
&\quad g_j \times \mathrm{cov}(b_{1,A_i}, b_{0,A_j}) + g_i \times \ln(gi_{i,A_i}) \times g_j \times \ln(gi_{j,A_j}) \times \mathrm{cov}(b_{1,A_i}, b_{1,A_j}) \\
&= g_i \times g_j \times [\mathrm{cov}(b_{0,A_i}, b_{0,A_j}) + \ln(gi_{j,A_j}) \times \mathrm{cov}(b_{0,A_i}, b_{1,A_j}) + \ln(gi_{i,A_i}) \times \\
&\quad \mathrm{cov}(b_{1,A_i}, b_{0,A_j}) + \ln(gi_{i,A_i}) \times \ln(gi_{j,A_j}) \times \mathrm{cov}\,(b_{1,A_i}, b_{1,A_j})]
\end{aligned}$$

where c_i and c_j are the values of c for trees i and j, respectively.

The covariances between parameters for two models associated with the two trees are required to solve this equation but are not available. However, if both trees have the same age ($A_i = A_j = A$), then the covariance formula can be simplified and solved.

$$\begin{aligned}\mathrm{cov}(\widehat{si}_i, \widehat{si}_j) \approx\ & e^{\hat{b}_{0,A}} \times gi_{i,A}{}^{\hat{b}_{1,A}} \times e^{\hat{b}_{0,A}} \times gi_{j,A}{}^{\hat{b}_{1,A}} \times [\mathrm{cov}(b_{0,A}, b_{0,A}) + \ln(gi_{j,A}) \times \\ & \mathrm{cov}(b_{0,A}, b_{1,A}) + \ln(gi_{i,A}) \times \mathrm{cov}(b_{1,A}, b_{0,A}) + \ln(gi_{i,A}) \times \\ & \ln(gi_{j,A}) \times \mathrm{cov}(b_{1,A}, b_{1,A})] \\ =\ & e^{\hat{b}_{0,A}} \times gi_{i,A}{}^{\hat{b}_{1,A}} \times e^{\hat{b}_{0,A}} \times gi_{j,A}{}^{\hat{b}_{1,A}} \times \{\mathrm{var}(b_{0,A}) + [\ln(gi_{i,A}) + \ln(gi_{j,A})] \times \\ & \mathrm{cov}(b_{0,A}, b_{1,A}) + \ln(gi_{i,A}) \times \ln(gi_{j,A}) \times \mathrm{var}(b_{1,A})\}\end{aligned}$$

Estimates of the variances and covariances in the above formulae are obtained from the fitting of the growth intercept models.