Principal components analysis of tree stem profiles[1]

PEDRO L. REAL
Universidad Austral de Chile, Facultad de Ciencias Forestales, Valdivia, Chile
JAMES A. MOORE
College of Forestry, Wildlife and Range Sciences, University of Idaho, Moscow, ID 83843, U.S.A.
AND
JAMES D. NEWBERRY
Potlatch Corporation, Lewiston, ID 83501, U.S.A.

Received March 1, 1989
Accepted July 14, 1989

REAL, P. L., MOORE, J. A., and NEWBERRY, J. D. 1989. Principal components analysis of tree stem profiles. Can. J. For. Res. **19**: 1538–1542.

The use of principal components analysis to study tree stem profiles was critically analyzed during 1085 destructively sampled Douglas-fir trees and 1260 simulated trees with known geometric shapes. Interpretation about the meaning of each principal component is provided and contrasted with others in the forestry literature. Nearly identical results with both the destructively sampled and simulated trees, along with certain theoretical consideratons, indicate that the principal components are related to tree form as opposed to tree profile or taper. Therefore, principal components analysis is a useful analytical tool for stratifying trees into different form groups.

REAL, P. L., MOORE, J. A., et NEWBERRY, J. D. 1989. Principal components analysis of tree stem profiles. Can. J. For. Res. **19** : 1538–1542.

L'utilisation de l'analyse en composantes principales pour étudier le profil de la tige des arbres a été évaluée à partir de 1085 tiges coupées de Sapin Douglas et de 1260 arbres simulés avec des formes géométriques connues. La signification de chaque composante principale est interprétée et comparée avec d'autres interprétations mentionnées dans la littérature forestière. Des résultats presque identiques avec les tiges coupées et simulées de même que certaines considérations théoriques indiquent que les composantes principales sont reliées à la forme plutôt qu'au frofil ou au défilement des arbres. Par conséquent, l'analyse en composantes principales est utile pour stratifier les arbres en différentes classes de forme.

[Traduit par la revue]

Introduction

The use of principal components analysis (PCA) for the study of tree stem profiles was first reported by Fries (1965). Fries studied birch and pine in Sweden and British Columbia using PCA and concluded, based on analysis of the first eigenvector, that both species have similar form in both locations. A similar analysis indicated that trees with shorter crowns showed less taper than those having longer crowns. Fries and Matern (1966) developed, with PCA, a system of taper equations for birch in Sweden. They interpreted the first principal component as the taper curve of the average tree, the second component as an indication of basal swelling, and the third component as an expression of form. Liu and Keister (1978) also used PCA to develop taper equation systems for loblolly pine and slash pine in Louisiana. They regarded the first principal component as defining the mean stem profile. The profiles of trees with different characteristics were also compared. Little difference among profiles was noted after examining the elements of the eigenvectors associated with the first principal component.

In these studies, PCA has been used to either build taper equation systems or explore differences in the profiles of different tree populations. As was concluded by Kozak and Smith (1966), we believe other methods of developing taper equations are preferrable to PCA. However, PCA appears to be a useful tool for differentiating among trees with different stem profiles. The key to using PCA for this purpose is attaching a biological or physical meaning to each principal component. Although there has been some consistency in attaching meaning to the principal components in these studies, they are not uniform and have not been adequately justified, in our opinion. Therefore, we chose to further explore the principal components obtained from tree profiles and attempted to attach justifiable meaning to them.

We explored PCA in three ways. First, certain theoretical aspects of PCA were considered. Next, analysis of data from simulated trees was conducted. The simulated trees allowed us to control sources of variation influencing the stem profile. Finally, Douglas-fir stem analysis data were analyzed so that conclusions reached from the theoretical considerations and simulated data could be verified with actual data.

Data

Sample trees

The development of a taper equation system (Real and Moore 1988) for second-growth Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco) in the Inland Northwest of the United States provided the actual stem analysis data used in this study. A total of 1085 trees were destructively sampled by the Intermountain Forest Tree Nutrition Cooperative. These trees were measured in 94 second-growth, even-aged stands of Douglas-fir (Moore 1984), which cover a broad range of ages, stand densities, sites, stand conditions, and geographic regions within the Inland Northwest. The distribution of the sample trees by diameter at 1.37 m (dbh) and total height is given in Table 1. The following variables

[1]Submitted as contribution No. 477 of the University of Idaho, College of Forestry, Wildlife and Range Sciences Experiment Station.

TABLE 1. Distribution (number) of the sample trees by diameter (dbh) and total height classes

dbh class (cm)	Total height class (m)					
	<15	15–21	21–27	27–34	>34	Total
12.7–25.4	54	164	9	—	—	227
25.4–38.1	14	387	259	13	1	674
38.1–50.8	—	20	75	69	2	166
>50.8	—	—	4	7	7	18
Total	68	571	347	89	10	1085

were measured for each tree: stump diameter (0.3 m), dbh, total height, diameter and height at the live crown base, and a variable number (between three and seven) of diameters and heights along the bole.

A matrix of diameters at the same position on each tree is needed for PCA. Because relative position is typically used for stem taper analysis and our trees were different heights, diameters at 3, 5, and 10% intervals to 90% of total height were estimated by quadratic interpolation from the actual diameters. A matrix of 1085 trees by 11 diameters was created with this procedure.

Simulated trees

The simulated trees used in this study were based on the following general model:

tree stem profile = f(tree form, tree taper) + random variation

where tree form is the characteristic shape of the tree profile, tree taper is the rate of narrowing in diameter relative to increasing height along a tree stem for a given form (Gray 1956), and random variation is any other factor unrelated to form or taper. The specific model used to generate the simulated stem profile data is

$$[1] \quad d = \beta_0 D \left| \frac{H - h}{H - \mathrm{BH}} \right|^{\beta_1}$$

where d is the diameter at height h, D is dbh, H is total tree height, h is any height along the tree bole, BH is breast height, and β_0, and β_1 are parameters related to taper and form, respectively. Ormerod (1973) introduced a model very similar to [1] for modeling stem profile for several tree species in British Columbia. Reed and Byrne (1985) used Ormerod's model as the basis for developing a variable-form stem profile system for red pine, jack pine, and white spruce in the upper Great Lakes region. Equation 1 was used by Newberry and Burkhart (1986) to construct variable-form and taper stem profile equations for loblolly pine. This model has parameters that are directly related to stem form and taper and has been shown to account for most of the variation in stem profile in the situations described above. Since we believe form and taper account for the vast majority of the variation in stem profile, we constructed the simulated tree data with [1] so that they varied in both form and taper.

A height–diameter model fitted to the sample trees was used to generate total heights for 2.5 cm diameter classes (12.7 to 63.5 cm or 5 to 25 in.). Twenty trees were constructed for each diameter class. Four different diameters for each diameter class were randomly obtained by assuming a uniform distribution over each class. For each of the four diameters, five different heights were assigned. The heights were assumed to be normally distributed with a mean equal to the estimate from the height–diameter model and a variance equal to the variance of prediction of the height–diameter model. Once the simulated tree diameters and heights were determined, each diameter and height pair was tripled and assigned one of three geometric shapes: cone, paraboloid, or neiloid. Equation 1 was then used to generate stem diameters at the same relative stem positions used with the actual sample trees. Thus, a matrix of 420 trees by 11 diameters was created for each geometric shape.

Methods and results

Theoretical aspects

Chatfield and Collins (1980) indicate that when all the original variables are positively correlated, the first principal component is a measure of "average size" or shape. Since diameters along a tree bole are, in most cases, positively correlated, the eigenvector associated with the first principal component gives the "average" profile of the trees being studied.

Two results from matrix algebra are also useful. First, the number of nonzero eigenvalues for the variance–covariance matrix is equal to the rank of the observation matrix (Graybill 1976; Morrison 1967). Second, the number of principal components is equal to the number of nonzero eigenvalues in the variance–covariance matrix (Morrison 1967). If we consider these results along with the stem profile model [1], further theoretical interpretations can be made about principal components of stem profile. If trees with a single stem form but different taper are considered, the stem profiles for these trees are linearly dependent, since each is a simple linear transformation (based on dbh) of the appropriate form. Therefore, the rank of the matrix of diameters from trees with a single form and different taper is equal to one. If, however, multiple tree forms are considered, those profiles from trees with the same form are linearly dependent, and hence, the rank of the diameter matrix in this case would be equal to the number of stem forms present. Since the number of principal components is equal to the rank of the diameter matrix, the number of principal components is equal to the number of stem forms in the population of trees.

From this, two points seem evident. First, the first principal component is the measure of average stem form. Second, the number of principal components is equal to the number of stem forms present in the tree population given model [1].

Simulated trees

Principal components analysis was applied to each of the three simulated tree data sets. The first principal component explains 100% of the total variance in each of the simulated diameter matrices. Therefore, each diameter matrix with one stem form has only one nonzero principal component. The plot of the eigenvector associated with the first principal component versus relative height for each geometric shape is given in Fig. 1. For clarity, the elements of each eigenvector have been joined by line segments and the value zero assigned to 100% of total height. Each figure duplicates the geometric shape of the diameter matrix from which they were derived.

Next, PCA was applied to a diameter matrix made up of all three geometric shapes. The results are provided in

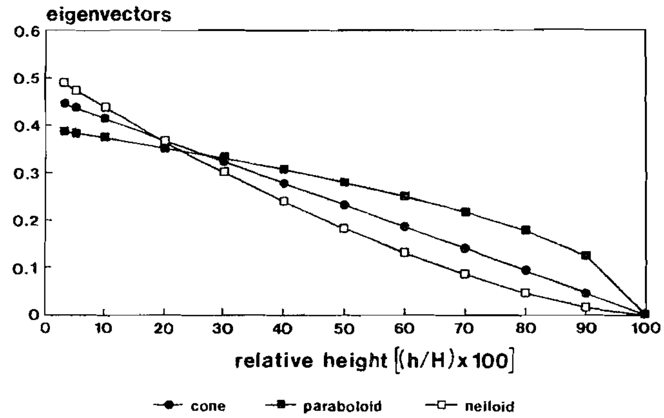

FIG. 1. Elements of the first eigenvectors versus relative height for simulated trees of three geometric shapes.

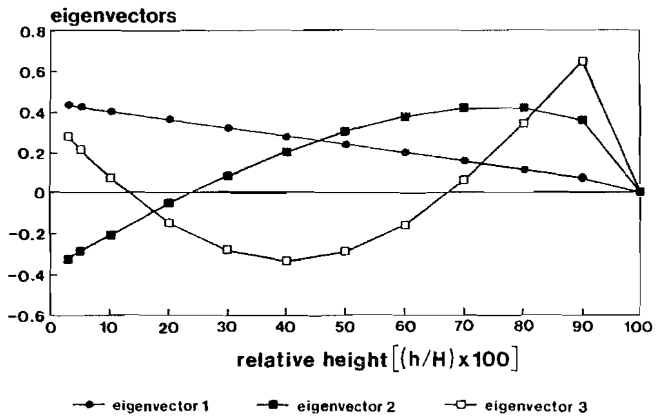

FIG. 2. Elements of the first three eigenvectors versus relative height for simulated trees of three geometric shapes equally divided in the same data set.

Table 2. There are three nonzero principal components associated with this diameter matrix. The first principal component is by far the most important since it alone explains approximately 98% of the variation in diameter. A plot of the eigenvectors associated with the three principal components versus relative height is shown in Fig. 2. The plot of the first eigenvector appears to be a combination of the three shapes included in this data set. The plots of the other two principal components do not have an easily interpretable pattern. The correlations between the elements of these eigenvectors and the stem diameters for the simulated trees are given in Table 3. The elements of the first eigenvector show strong correlations with all diameters aong the tree stem. The highest correlations are between 10 and 40% of total height. The elements of the second eigenvector are not as highly correlated with the stem diameters as are the elements of the first eigenvector. The highest correlations with the elements of the second eigenvector are associated with diameters near the top of the tree. The elements of the third eigenvector are only weakly correlated with diameters along the tree stem.

Sample trees

The results of PCA applied to the 1085 sample Douglas-fir trees are summarized in Tables 4 and 5. The first component is by far the most important, as it was with the simulated trees, since it alone explains 95.4% of the total variation. The second component accounts for 3.5% of the variation, and the remaining components account for about 1% of the variation. The elements of the first eigenvector show a similar correlation structure with sample tree bole diameters as was shown with the simulated trees. Again all the correlations are positive and the highest correlations are associated with bole diameters between 10 and 40% of tree height. The elements of the second eigenvector show a weaker correlation structure with bole diameters than the elements of the first eigenvector. The pattern of correlations is similar to the simulated trees, although the similarity is not as strong as with the first eigenvectors. The strongest correlations again are in the upper portion of the tree stem. The elements of the third eigenvector also show the same correlation pattern as did the elements of the third eigenvector from the simulated trees. Again the correlations are very weak.

A plot of the elements of the first three eigenvectors versus relative height is shown in Fig. 3. Clearly, the plot of the first eigenvector resembles the form of a tree. The plots of the second and third eigenvectors closely resemble the plots of the second and third eigenvectors from the simulated data.

Discussion and conclusion

Principal components analysis applied to the simulated and the Douglas-fir sample tree data sets gave results that are very similar. The plots of the elements of the first three eigenvectors from both data sets show almost identical patterns. The elements themselves are also very similar. Correlations between the elements of the eigenvectors and tree stem diameters are approximately the same in both sign and magnitude. We did not expect the results between the two data sets to be identical for two reasons: first, there is variation in the Douglas-fir data set that is not found in the simulated trees and second, the neiloid form included in the simulated data is not generally a form associated with Douglas-fir. Results from a combined cone–paraboloid data set were in many ways closer to the Douglas-fir data than the simulated data we presented here. We purposely chose the three-form data set so that there would be three nonzero eigenvalues. Based on the similarities between the two data sets, we contend that interpretations made about principal components based on theoretical considerations and the simulated data are applicable to the Douglas-fir sample trees. We also believe, because the simulated data are not species specific, these results are also applicable to other species with excurrent boles.

Therefore, based on the results of our analysis, the first principal component should be interpreted as a measure of average tree stem form. The second principal component indicates the variation in tree stem form not explained by the first component. The correlation analysis of both data sets with stem diameters indicates most of the variation in form occurs in the upper portion of the tree bole. The third component is not readily interpretable. Given the results of the correlation analysis, we are unsure as to the significance of this component. Here it is important to note that none of the eigenvectors appears to be related to stem taper as we define it, given the similarities between the simulated and real profile data.

These results are different from those previously reported in the literature. Previous studies have defined the first principal component as form (Fries 1965), taper curve of the

TABLE 2. Eigenvalues and percentage of the variance explained for each principal component for selected combinations of shapes (simulated trees)

Component	Eigenvalue	% variance explained	Cumulative	Shapes
First	1568.45	99.14	99.14	Cone-paraboloid
Second	13.64	0.86	100.00	Cone-paraboloid
First	1273.67	99.57	99.57	Cone-neiloid
Second	5.43	0.43	100.00	Cone-neiloid
First	1447.18	97.65	97.65	Paraboloid-neiloid
Second	34.86	2.35	100.00	Paraboloid-neiloid
First	1423.15	98.30	98.30	Cone-paraboloid-neiloid
Second	24.43	1.69	99.99	Cone-paraboloid-neiloid
Third	0.16	0.01	100.00	Cone-paraboloid-neiloid

TABLE 3. Correlations of the elements of the eigenvectors for the first three principal components with bole diameters for the simulated trees

Relative height[a]	Eigenvector 1	Eigenvector 2	Eigenvector 3
3	0.4388 (0.97)	−0.3250 (−0.05)	0.2803 (0.01)
5	0.4300 (0.97)	−0.2905 (−0.03)	0.2142 (0.00)
10	0.4083 (0.98)	−0.2076 (−0.03)	0.0760 (0.00)
20	0.3652 (0.99)	−0.0532 (−0.15)	−0.1476 (−0.02)
30	0.3227 (0.99)	0.0849 (0.29)	−0.2833 (−0.03)
40	0.2808 (0.97)	0.2046 (0.43)	−0.3364 (−0.04)
50	0.2393 (0.93)	0.3028 (0.55)	−0.2911 (−0.04)
60	0.1982 (0.87)	0.3762 (0.67)	−0.1613 (−0.03)
70	0.1571 (0.79)	0.4195 (0.76)	0.0598 (0.00)
80	0.1153 (0.71)	0.4213 (0.84)	0.3455 (0.04)
90	0.0710 (0.61)	0.3581 (0.89)	0.6487 (0.11)

NOTE: Correlation coefficients are in parentheses.
[a]Calculated as $(h/H) \times 100$.

TABLE 4. Eigenvalues and percentage of variance explained for each principal component (sample trees)

Component	Eigenvalue	% variance explained	Cumulative
First	43.197	95.42	95.42
Second	1.581	3.49	98.91
Third	0.318	0.70	99.61

average tree (Fries and Matern 1966) and mean stem profile (Liu and Keister 1978). Our results show that the first component is the average tree stem form and not profile or taper. The association of basal swelling with the second principal component (Fries and Matern 1966) is not supported by this analysis. The second component is most highly related to upper stem diameters in both the simulated and Douglas-fir sample tree data sets. Fries and Matern (1966) indicate the third component is an expression of stem form. We cannot interpret the third component in the context of stem profiles and are unsure about its significance. However, we do not believe it is an expression of stem form since the first component indicates form.

There are many analyses where it may be important to partition tree profiles into form and taper separately, since various biotic (for example, species, stand density, crown characteristics) and abiotic (for example, elevation, slope, soil substrate) factors may differentially affect stem form and taper. By stratifying trees into different groups, running PCA on the groups, and studying the first two components, groups with less form variation can be found. Based on our results, PCA is a useful analytical tool for grouping trees into different form strata, but is not useful for taper analysis.

TABLE 5. Correlations of the elements of the eigenvectors for the first three principal components with bole diameters for the sample trees

Relative height[a]	Eigenvector 1	Eigenvector 2	Eigenvector 3
3	0.4343 (0.98)	−0.3836 (−0.16)	0.4941 (0.10)
5	0.3975 (0.98)	−0.3456 (−0.16)	0.2327 (0.05)
10	0.3842 (0.99)	−0.1926 (−0.09)	−0.1098 (−0.03)
20	0.3476 (0.99)	−0.0660 (−0.04)	−0.3963 (−0.10)
30	0.3275 (0.99)	0.0486 (0.03)	−0.4012 (−0.10)
40	0.3039 (0.99)	0.1828 (0.11)	−0.3044 (−0.08)
50	0.2732 (0.97)	0.3160 (0.22)	−0.1316 (−0.04)
60	0.2335 (0.94)	0.4177 (0.32)	0.0848 (0.03)
70	0.1854 (0.90)	0.4397 (0.41)	0.2760 (0.12)
80	0.1307 (0.86)	0.3718 (0.46)	0.3411 (0.19)
90	0.0690 (0.80)	0.2242 (0.50)	0.2487 (0.25)

NOTE: Correlation coefficients are in parentheses.
[a]Calculated as $(h/H) \times 100$.

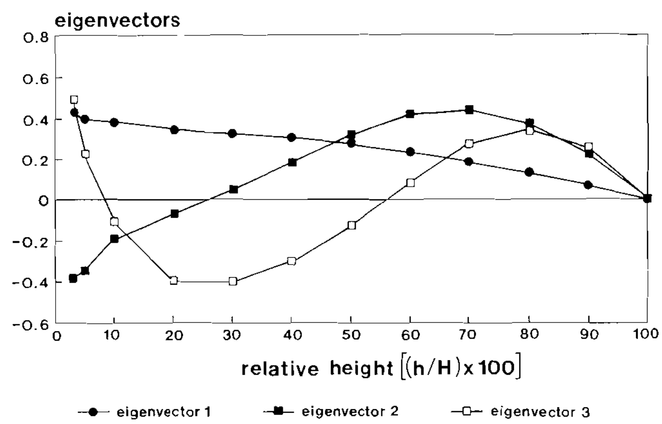

FIG. 3. Elements of the first three eigenvectors versus relative height for the actual trees.

Acknowledgements

This study was conducted while the senior author was Graduate Research Assistant at the University of Idaho. The support of the Intermountain Forest Tree Nutrition

Cooperative and Potlatch Corporation is greatly appreciated.

CHATFIELD, C., and COLLINS, A.J. 1980. Introduction to multivariate analysis. Chapman and Hall, New York.

FRIES, J. 1965. Eigenvector analysis show birch and pine have similar form in Sweden and in British Columbia. For. Chron. **41**: 133–139.

FRIES, J., and MATERN, D. 1966. On the use of multivariate methods for the construction of tree taper curves. IUFRO Advisory Group of Forest Statisticians Conference, Stockholm, Sweden. Department of Forest Biometry, Royal College of Forestry, Stockholm, Sweden. Res. Note No. 9. pp. 85–117.

GRAY, H.R. 1956. The form and taper of forest-tree stems. Oxford University, Imperial Forestry Institute, Oxford. Pap. No. 32.

GRAYBILL, F.A. 1976. Theory and application of the linear model. Duxburg Press, Wadsworth Publishing Co., Inc., Belmont, CA.

KOZAK, A., and SMITH, J.H.G. 1966. Critical analysis of multivariate methods for the construction of tree taper curves. For. Chron. **42**: 458–463.

LIU, C.J., and KEISTER, T.D. 1978. Southern pine stem form defined through principal components analysis. Can. J. For. Res. **8**: 188–197.

MOORE, J.A. 1984. Fourth annual report—intermountain forest tree nutrition cooperative. College of Forestry, Wildlife, and Range Sciences, University of Idaho, Moscow.

MORRISON, D.F. 1967. Multivariate statistical methods. McGraw-Hill, Inc., New York.

NEWBERRY, J.D., and BURKHART, H.E. 1986. Variable-form stem profile models for loblolly pine. Can. J. For. Res. **16**: 109–114.

ORMEROD, D.W. 1973. A simple bole model. For. Chron. **49**: 136–138.

REAL, P.L., and MOORE, J.A. 1988. An individual tree taper system for Douglas-fir in the Inland Northwest. *In* Forest Growth Modeling and Prediction. Proceedings of IUFRO Conference, August 24–28, 1987, Minneapolis, MN. *Edited by* A.R. Ek, S.R. Shifley, and T.E. Burk. USDA For. Serv. Gen. Tech. Rep. NC-120.

REED, D.D., and BYRNE, J.C. 1985. A simple, variable form volume estimation system. For. Chron. **61**: 87–90.