韓國林學會誌 79(1) : 16-20. 1990.
Jour. Korean For. Soc. 79(1) : 16-20. 1990.

Relationships between Soil-Site Properties and Bamboo (*Phyllostachys bambusoides*) Growth[1]

Young Gwan Chung[2] and Carl W. Ramm[3]

土壤의 理化學的 特性과 대나무 生長과의 關係[1]

鄭永觀[2] · Carl W. Ramm[3]

ABSTRACT

Canonical correlation analysis was used to relate 17 soil-site variables to bamboo diameter, height, and internodal characteristics. The first canonical correlation was highly significant, explained much of the variance in both sets of variables, and the canonical variates made sense biologically. Surface soil depth, total nitrogen and percent organic matter had high positive correlations with the first soil-site canonical variate. Clay content (%) and cation exchange capacity were negatively correlated with the first soil-site canonical variate. Only 8 of predictor variables were considered relevant for predicting bamboo growth.

Key words : *bamboo, canonical correlation, multivariate analysis, redundancy, soil-site relationships.*

要 約

17個의 土壤因子와 대나무의 直徑, 稈高 그리고 節間 生長量과의 關係를 究明하기 위하여 正準相關分析을 實施하였다. 첫번째 正準相關은 매우 有意的 이었고, 變數의 2組合에서 分散은 크고, 그리고 正準變量은 生物學的으로 意義가 있었다.

따라서 表層土深, 全窒素 그리고 有機物含量등은 첫번째 土壤性質의 正準變量과 매우높은 正의 相關을 나타내었다. 그리고 粘土含量과 置換性容量은 負의 相關으로 나타났다. 17個 變數 가운데 8個 變數만이 대나무의 生長을 豫測하는데 關係되는 것으로 思料된다.

INTRODUCTION

This study investigated the relationships between physiochemical properties of soil and bamboo growth using canonical correlation analysis. The objectives were to determine which variables were responsible for the variation in bamboo growth, and to evaluate the usefulness of canonical correlation analysis in analyzing soil-site relationships. The use and interpretation of canonical correlations is briefly described.

In Korea, the bamboo, which is botanically consisted of 5 genera and 54 species, is produced from 5,360ha (estimated production about 35,000 bundles). The self-sufficiency of the wood reaches to only 20 percent, thus, the most of the wood is imported from abroad (Chung, 1959 ; Lee, 1985).

One of the most important agricultural problems to be solved is how to stabilize the fluctuation of the productivity as well as to enhance the yield.

[1] 接受 1989年 11月 1日 Received on November 12, 1989.
[2] 慶尙大學校 農科大學 College of Agriculture, Gyeongsang National University, Chinju, Korea
[3] Department of Forestry, Michigan State University, East Lansing, Michigan, U.S.A.

In agricultural point of view, site selection and adequate management of fertilizer should be of great importance, because the fluctuation of the growth due to soil properties of Korean hill country, of which soil characterized as high acidity and infertile.

In this study, therefore, a close observation of soil-site properties and related effect on the growth has been made.

METHODS

Canonical correlation analysis (CCA) is used for joint analysis of two sets of variables, typically where the variables in one set are recognized as predictors and the variables in the other set are recognized as responses. CCA may be viewed as a general regression of all dependent variables upon all independent variables. The objective is to find linear composites (the canonical variates) from each set of variables such that the correlation between the composites is maximized (Clark and Hosking, 1986).

The canonical variates, linear combinations of their respective data sets, are derived in pairs, one from the predictor data set (soil-site variables) and one from the outcome data set growth characteristics of bamboo). In other words, a linear combination (U_1) of the soil-site variables is found that is highly correlated with growth characteristics of bamboo. At the same time, a linear combination (V_1) of the bamboo variables is found that is highly correlated with the soil-site variables. U_1 and V_1 would be the first pair of canonical variates if they have the highest correlation of all potential linear combinations. The second pair of canonical variates (U_2 and V_2) would have the second largest correlation, and so on. The number of pairs of canonical variates is equal to the number of variables in the smaller data set.

Canonical correlations are the product-moment correlation coefficients between canonical variates U_i and V_i. They have many properties in common with simple correlation coefficients, such as a range from -1 to $+1$. The primary difference between the two is that the canonical correlation is a maximized correlation. The *k*th canonical correlation can be interpreted as the multiple correlation coefficient between, say, U_i and the variables in the response data set. The square of the first canonical correlation is the proportion of the variance in the first canonical variate U_1 (or V_1) that is accounted for by the first canonical variate V_1 (or U_1).

The magnitude of the canonical correlations are partially dependent upon the ratio of the total combined number of variables (p) to the sample size N (Gittins, 1985). As p approaches N, the canonical correlations - especially the first correlation - approach one. CCA, therefore, requires a large sample size relative to the total number of variables measured to accurately reflect the relationship between data sets.

There is not universal agreement as to whether the canonical coefficients or the canonical loadings are more suitable for interpretation. The canonical coefficients, similar to regression coefficients, are affected by multicollinearity. If multicollinearity is present, it is difficult to evaluate the relative contribution of the variables (Clark and Hosking, 1986). The numerical values of the canonical coefficients appear to be dependent upon the variables selected from each data set, as well as their scale (Gittins, 1985). For these reasons, many prefer to use the canonical loadings, the correlations of the original variables with their canonical variates. Canonical loadings tend to be more stable than raw or standardized canonical coefficients when variables are added or deleted and under replicate sampling (Gittins, 1985). On the other hand canonical loadings will show only linear relationships in the data, and the product-moment correlation coefficient is extremely sensitive to data outliers.

Redundancy analysis can help establish the number of 'biologically' significant linear relationships between the prediction and the response data set. Redundancy is the proportion of variance in data set 1 (outcome variables) that is accounted for by the i^{th} composite of data set 2 (predictor variables). Redundancy is not symmetrical, due to the fact that each of the sets has differential amount of intercorrelations; there is also a redundancy index that shows the proportion of the variance in set 2 that is accounted for by the i^{th} composite of set 1. It is important to consider redundancies as well as canonical correlations.

CCA may be used for inference testing or less formally for exploratory data analysis. For estimation and significance testing, CCA has the following assumptions: the data is continuous (i. e., interval scale); there are linear relationships between the two sets of variables; the correlation (or covariance) matrix is of full rank; and the population from which the data sets were derived has a multivariate normal distribution. For exploratory data analysis, normal distributions are not required. Linear relationships and continuous data, however, are still necessary conditions.

MAIERIALS

Data on 4 measurements of bamboo growth and 17 soil-site factors were collected from 107 sample plots in Cheonnam and Gyeongnam provinces, Korea. The 107 plots were selected at random across climatic regions. The plots were 20m x 20 m in size; all bamboo plants within the plot were measured. The 17 soil-site predictor variables were measured by soil analysis methods of the Forest Research Institute in Korea (FRIK, 1983). Bamboo and soil-site data were aggregated into plot averages. The data used in the canonical analysis is composed of two sets of variables:

a) outcome (dependent) variables:

DBH = average diameter at breast height (cm)
HT = average total height (m)
LID = largest internodal diameter (cm)
LIL = longest internodal length (cm)

b) predictor (independent) variables:

ALT = altitude (m)
SLP = percent slope
SSD = surface soil depth (cm)
G% = gravel content (%)
S% = sand content (%)
Si% = silt content (%)
C% = clay content (%)
pH = soil pH (H : 0 1 : 5)
OM = organic matter (%)
totN = total nitrogen (%)
P_2O_5 = available P_2O_5 (ppm)
CEC = cation exchange capacity (me/100g)
K = exchangeable K^+
Na = exchangeable Na^+
Ca = exchangeable Ca^{++}
Mg = exchangeable Mg^{++}
SO_4 = SO_4

RESULTS

Table 1 presents the canonical correlations, squared canonical correlations and the eigenvalues for the data set. Note that the second canonical variate pair accounts for less than 50% of the shared variability, and that the third and fourth canonical pairs account for even less. The sum of eigenvalues (Table 1) may be used to define the total variability in the data set. By this reckoning, the first pair of canonical variates accounts for 71% of the total variability; the first and second pair combined account for almost 89% of the total variability. The conclusion to be drawn from Table 1 is that the first canonical pair explains the majority of the variability in the data sets.

Table 1. Canonical correlation, squared canonical correlations and eigenvalues

Canonical pair	Canonical correlation	Squared canonical correlation	Eigenvalues
1	0.8831	0.7799	3.5442
2	0.6831	0.4666	0.8747
3	0.5509	0.3035	0.4357
4	0.3378	0.1141	0.1288

Table 2 presents the likelihood ratio tests for the hypothesis that the canonical correlation in each row (and all smaller correlations) are equal to zero. This test is not robust to deviations from normality. Based on these results, the remaining interpretation will focus on only the first two canonical variates. It should be emphasized that statistical significance does not infer biological significance : the linear combinations that form the canonical variates are selected to maximize correlation, not interpretation.

Tabel 2. Likelihood ratio tests that canonical correlations equal zero

Canonical Correlation	Likelihood ratio	Approximate F
1	0.0724	4.7583**
2	0.3291	2.4496**
3	0.6170	1.6019*
4	0.8859	0.8190

* =significant at alpha=0.05
**=significant at alpha=0.01

Our interpretation of the canonical analysis will be based on the canonical loadings. Kimberly (1981) suggests considering only those canonical loadings above 0.5 ; this threshold value will be used for our interpretation.

The first canonical variate associated with bamboo growth (PROD 1) may be considered a measure of the physical dimensions of bamboo. Average dbh, height and largest internodal diameter have high correlations with PROD 1 (Table 3) and the remaining variable, longest internodal length, has a loading well above the threshold level.

Table 3. Correlations (loadings) between bamboo growth and their canonical variates

Original	Canonical variates			
variables	*PROD* 1	*PROD* 2	*PROD* 3	*PROD* 4
DBH	0.8802	-0.3543	0.2997	0.0994
HT	0.9910	0.0891	0.0229	0.0971
LID	0.8756	-0.3476	0.2970	0.1559
LIL	0.6253	0.3986	0.6679	0.0631

The first soil-site canonical variate, SOIL 1, may be considered a measure of soil fertility or site productivity. There are large positive correlations between SOIL 1 and surface soil depth, total nitrogen, percent organic matter, and available P_2O_5 (Table 4). Other variables having correlations with SOIL 1 greater than 0.5 include percent silt content and exchangeable Ca++. Percent clay content and cation exchange capacity have high negative correlations with SOIL 1. These are related in that the greater the clay content of the soil the greater the cation exchange capacity.

Table 4. Correlations (loadings) between the soil -site variables and their canonical variates

Original	Canonical variates			
variables	SOIL 1	SOIL 2	SOIL 3	SOIL 4
ALT	0.1284	0.2429	-0.3083	0.0148
SLP	0.2081	-0.2623	-0.2856	0.2643
SSD	0.9136	-0.0299	-0.0365	0.1666
G%	0.3543	0.1451	0.0654	0.2881
S%	-0.1363	0.3117	-0.1774	0.1843
Si%	0.5440	0.0185	0.2183	-0.0181
C%	-0.5639	-0.5230	-0.0086	-0.2623
pH	0.1884	0.2074	-0.2200	0.4234
OM	0.7764	-0.3244	0.1966	0.1158
totN	0.7983	-0.2468	0.0339	-0.0826
P_2O_5	0.6613	0.0085	0.4488	-0.0870
CEC	-0.5258	-0.4172	-0.1223	0.2086
K+	0.4657	0.3219	-0.0115	-0.1309
Na+	0.0680	-0.2589	0.2477	0.0613
Ca++	0.5562	0.0987	0.0968	0.3707
Mg++	0.3442	0.0389	0.3349	0.3265
SO_4	0.2813	0.1354	0.1555	0.1539

Lee (1985) reported that degree of slope, altitude, topography and aspect were highly related with bamboo diameter and height growth.

The second canonical variate associated with bamboo growth has low canonical loadings, especially for average height. This indicates low correlation between the canonical variate PROD 2 and the outcome variables (DBH, HT, LID, LIL). This variate may be related to the physical dimensions of the longest internodal length and dbh. The contradiction in correlation and the results of the likelihood ratio test indicate that redundancy analysis is necessary.

Table 5 presents the redundancy analysis for the productivity variables and for the soil factor variables. Note that, for both data sets, the amount of variability explained by the second canonical variate is less than 5 percent. This would argue that although the correlation between the second pair of canonical variates is highly significant, the relationship is not relevant.

Two aspects have to be considered when evaluating redundancy. First, the amount of interrelationship within a set will influence the amount of variance to be explained. Thus we will

Table 5. Redundancy analysis

canonical *variate*	Proportion of standardized variance of bamboo variables : explained by PROD	explained by SOIL
1	0.7286	0.5683
2	0.1033	0.0482
3	0.1562	0.0474
4	0.0119	0.0014
canonical *variate*	Proportion of standardized variance of soil-site variables : explained by SOIL	explained by PROD
1	0.2571	0.2005
2	0.0650	0.0303
3	0.0455	0.0138
4	0.0484	0.0055

expect different contributions from each set of canonical variates. Second, if there are different numbers of variables in the different sets, there are of different levels of explanation for each set of variables. Extracting a set number of components will utilize more variance from the set with the samller dimension than from the larger set. Therefore, the total redundancy of the smaller set with the larger set will always be greater than the total redundancy of the larger set with the smaller set (Clark and Hosking, 1986). This can be seen in Table 5. The proportion of variance in set 1, bamboo growth (4 variables), that is accounted for by the two canonical variates in set 2, the soil-site variables, is 61.6% (e.g., 0.5683 +0.0482). The proportion of variance in the set 2 that is accounted for by the two canonical variates in set 1 is only 23.1%.

CONCLUSIONS

Canonical correlation analysis was useful in evaluating relationships between two sets of data. Of the 17 predictor variables initially measured, only 8 can be considered relevant for predicting bamboo growth. These variables are surface soil depth, percent silt content, percent clay content, percent organic matter, total nitrogen, available P_2O_5, cation exchange capacity and exchangeable Ca^{++}. Of the 8 variables, the results suggest concentrating on surface soil depth, percent organic matter and total nitrogen.

LITERATURE CITED

1. Chung, H.P. 1959. Studies on the Bamboo in Korea. J. Korean Bot. Soc. 2 : (1) : 14-21.
2. Clark, W.A.V. and P.C. Hosking, 1986. Statistical Method for Geographers. John Willey & Sons, 518pp.
3. Forest Research Institute in Korea. 1983. Essential in Practice of Standard Forest Experiment, the Method of Soil Analysis. 189-198pp.
4. Gittins, R. 1985. Canonical Analysis : A Review with Applications in Ecology. Springer -Verlag, New York. 351pp.
5. Johnston, R.J. 1978. Multivariate Statistical Analysis in Geography - a primer on the general linear model- Longman Group Limited, London. 280p.
6. Kimberly, G.S. 1981. Canonical correlation analysis and its use in wildlife habitat studies. Page 249. The Use of Multivariate Statistics in Studies of Wildlife Habitat. David E. Capen (ed.), USDA Forest Service, General Technical Report RM-87.
7. Lee, S.H. 1985. A Study on the Distribution of Bamboo Forest, Their Site and Growth in Korea. M.S. Thesis. Graduate School of Seoul National University. 42pp.