THE FACTOR ANALYSIS OF MULTIVARIATE DATA SYSTEMS[1]

Robert T. Eckert and Robert D. Westfall[2]

Because of recent advances in data acquisition and analysis systems, large sets of measurements may now be generated. Most work published to date on large (or multivariate) data systems has been limited to uni- or bivariate methods of analysis. Such methods, however, are inadequate for screening relationships among variables or for determining whether sets of measurements relate to such experimental conditions as family relationships, cultural treatments, or natural environmental conditions. A number of multivariate techniques which overcome these inadequacies are now available. Although the theoretical basis for these techniques had been developed in the early 1900's, access to large memory computer systems has only recently made application of these techniques practicable.

We present here some considerations on the analysis of relationships among measurements: first discussing some drawbacks in univariate techniques and then examining the use of a particularly appropriate multivariate technique, factor analysis, using a study of white pine weevil - Norway spruce relationships as an example.

DISCUSSION OF THE METHODS

The analysis of relationships among measurements generally begins with simple correlations. However, variables included in the analysis may overlap in the information they contain and therefore may bring about spurious correlations. Thus, interpretation of large numbers of simple correlations can be very difficult.

Partial correlation and multiple regression analyses have been used to further sort out information represented in simple correlations. Both of these techniques test the relationship between any two variables after the linear effects of the remaining variables in the analysis have been removed (Kempthorne, 1969). With small numbers of variables, this property can aid in analysing relationships; but when numbers of variables are large, collinearities among these variables can render the analysis ambiguous or misleading.

Factor analysis, however, lacks the above deficiencies. Rather than substracting out linear effects, these effects are divided into separate modes

[1] Based, inpart, on portions of a dissertation submitted by the senior author, in partial fulfillment of the requirements for the Master of Science degree at the SUNY College of Environmental Science and Forestry, Syracuse, New York.

[2] Department of Botany, Ohio State University, Columbus, Ohio; and Research Associate, Department of Silviculture and Forest Influences, SUNY College of Environmental Science and Forestry, Syracuse, New York, respectively.

Under the factor model, each trait or measure (Xi) is represented as a linear function of a small number of common-factor variates and a single specific variate (Morrison, 1967). In other words, variation in each measurement is broken down into two major components:

1. the variation unique or specific to a measurement

2. the variation associated with other measurements in the sample (covariation or communality).

The factor model may be represented by the following:

$$X_1 = a_{11}F_1 + a_{12}F_2 + \ldots \ldots a_{1m}F_m + e_1$$

$$X_2 = a_{21}F_1 + a_{22}F_2 + \ldots \ldots a_{2m}F_m + e_2$$

.

.

.

$$X_p = a_{p1}F_1 + a_{p2}F_2 + \ldots \ldots a_{pm}F_m + e_p$$

Where:

$X_1, \ldots X_p$ = standardized random variables

$F_1, \ldots F_m$ = standard common factor variates or "factors"

$a_{11}, \ldots a_{pm}$ = loadings reflecting the importance of the mth factor in composition on the pth response variate (X_p)

$e_1, \ldots e_p$ = specific variates reflecting the uniqueness of the pth response variate (X_p)

In the computation of factors, the matrix of correlations among variables is first reduced to represent only common variation: that is, unities in the diagonal vector of the correlation matrix are replaced by communalities. Of the number of approaches which may be used to calculate communalities, the one used here reduced the original matrix to an image covariance matrix (Kaiser, 1963). In this method, each variable is regressed against the remaining variables and the resulting coefficients of multiple determination (R^2's) are inserted as communalities.

Characteristic roots and vectors are then extracted from the reduced correlation matrix. These vectors are then rotated according to one of a number of criteria. The criterion used here is that of simple structure, whereby two conditions are established: (1) any given variable will have only a few true relationships (factors) and (2) no one factor will be represented by more than a small proportion of the variables. Thus if these conditions are met, the rotated factor matrix will have the following characteristic form:

1. Only a few of the relationships (loadings) on any given factor will be of substantial size.

2. A given row of the factor matrix, containing loadings for a variable, should have nonzero loadings with only a few factors.

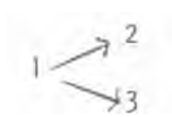
3. Any two factors should exhibit a different pattern of high and low loadings.

In orthogonal simple structure (where the factor vectors are uncorrelated), loadings express correlations between variables and factors. The sum of the squared loadings for any variable will equal the communality for that variable and the sum of the squared loadings for any factor will equal the variation (in standardized units) described by that factor.

If, however, factor vectors are correlated (or oblique), three matrices will describe the relationships present among variables: 1. A matrix of correlations among factors; 2. A factor pattern matrix, containing loadings or levels of importance between factors and variables; 3. A factor structure matrix, containing correlations between factors and variables (Cattell, 1965). The factor pattern matrix is the most useful of the three in describing relationships between variables.

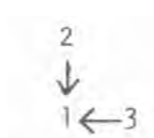
MATERIALS AND METHODS

In order to illustrate interpretations of relationships, using the methods of analysis discussed above two types of three-variate relationships were simulated. The first, the common cause type, is illustrated directly below.



As seen above, variables 2 and 3 are each linearly related to variable 1. This relationship is often used to illustrate applications of partial correlation analysis.

In the second relationship, shown below, a single variable (1) is a linear function of the other two (2 and 3). Variables 2 and 3 are uncorrelated. This will be referred to as the two independent causes type.



An arbitrary linear function for each relationship was used to generate the respective data. Twenty-four (24) and thirty-two (32) "observations" were made respectively for the former and the latter relationships given above.

To further illustrate interpretations of the three types of correlation analysis (simple and partial correlations and factor analysis), data from a white pine weevil - Norway spruce study were used. The focus of this study was to examine associations of weevil damage in Norway spruce with weevil activities on the trees and tree morphology (Eckert, 1974).

A heavily weeviled 10-year-old Norway spruce stand in Pack Forest, Warrensburg, New York, was chosen for the study. One hundred forty-one trees in five strips were measured in 1968 for height, length of the previous year's leader, mid-point diameter of the previous year's leader, and frequency of weeviling over the past eight years. In addition, average daily weevil activities on each sample tree were summed for the period, May through June, 1968. These were: the number of non-mating weevils on the leader or incidence, the number of weevils in mating position on the leader, and the number of ovipositing females of the leader. In September of that year, the current year's weevil damage was classified according to three levels: (1) no damage, (2) the presence of feeding punctures, some larval development, but no emergence, and a living leader, (3) larval emergence, the leader killed.

Length measurements were logarithmically transformed; weeviling frequencies were arcsine transformed.

Simple correlation matrices were calculated for the three data sets. Partial correlations were then estimated using the following formula:

$$r_{ij.kl...} = \frac{-a_{ij}}{a_{ii}a_{jj}}$$

where aij, aii, and ajj are the off-diagonal and the two diagonal elements, respectively, of the inverse simple correlation matrix (Steele and Torrie, 1960).

In the factor analysis of the data, each simple correlation matrix was then reduced to an image covariance matrix, from which characteristic roots and vectors were extracted (Veldman, 1967). The characteristic vectors were first rotated to Varimax orthogonal simple structure (Veldman, 1967). Then the factor vectors were hand-rotated to oblique simple structure, using the reference vector method (Comrey, 1973).

RESULTS AND DISCUSSION

Analysis of the Simulated Relationships.--Variables 2 and 3 in the simulated common cause relationship are shown from the simple correlations (Table 1) to be significantly related. However, the partial correlations show that this relationship is due almost entirely to their common relationship with variable 1.

The same inferences may be made from the factor analysis of this relationship, although these may not be readily apparent from the simple structure patterns. From the factor pattern matrix (Table 1), we found that variables 2 and 3 are associated with the first factor, while variable 1 alone is associated with the second. Without being aware of the relationship between the two factors, one might initially infer that variables 2 and 3 are independent of 1. However, the two factors are highly correlated, indicating that variable 1 independently varies and affects, in part, the variation in both variables 2 and 3.

Simple correlations, partial correlations and factors of the two independent causes relationship are shown in Table 2. Note that although variables 2 and 3 in the simple correlation matrix are uncorrelated, their partial correlation is highly negative. This has occurred because partial correlations are covariance adjusted, Thus there is a very small, but significant, residual negative relationship between variables 2 and 3.

The factor analysis directly relates to the system shown in the simple correlations: (1) the variation in measurement 1 is influenced by that in measurement 2 and 3, (2) variables 2 and 3 are independent. Here oblique rotation would not improve the factor structure.

Analysis of the Norway spruce - white pine weevil relationship.--The simple correlations (Table 3) indicate, with the exception of diameter, a number of significant relationships among the tree morphological, weevil activity, and damage measurements. However, because of the size of the correlation matrix, the number of types of associations among the measurements and the relationships among these types are difficult to sort out.

Factor analysis indicates that, at least in these data, the correlations may be reduced to four general associations (or factors) (Table 4): a measure of tree morphology (factor 1), an association between weevil activity on a tree and damage to that tree (factor 2), an association between tree height or growth rate and damage (factor 3), an association between weevil mating and weeviling frequency on a tree (factor 4). Ninety-nine percent of the variation common to the eight variables in the study was described by these four factors.

The matrix presented in Table 4 gives only the number of associations; the factors shown are mathematically independent. However, the relatively large number of nonzero (greater than + 0.10) loadings in each factor indicates that the factors may be correlated and thus may be further rotated. In fact, oblique rotation is effective in this case: the character of each factor remains unchanged, even though factor structure is simpler (Table 5), and a number of the factors are moderately to highly correlated (Table 6).

A number of inferences can be made from these results, particularly with the aid of the partial correlations (as in the simulated examples above).

1. Leader diameter is unrelated to weevil damage except through its relationships with height growth (factor 1).

2. The most vigorous trees in the stand are most frequently damaged. However, the relatively moderate correlations between the tree vigor - damage factor (3) and the weevil activity - damage factors (2 and 4) indicate that vigor makes only a moderate to small contribution to the variation in damage to the stand.

3. Weevil activity in the stand occurred on those trees which had been damaged that year (factor 2).

Partial correlations were particularly helpful in sorting out relationships between variables in factors 2 and 4, the factor pair most highly correlated (Table 6). These relationships are illustrated in Figure 1. Only those variables loaded over an absolute value of 0.2500 were included in the figure. In addition, lines connect only variable pairs which have significant partial correlations. Therefore, none of the variables loaded within either of the two factors are partially correlated.

From this figure, one can infer that the cause of the relationships among incidence, oviposition, and current damage was through mating. Therefore, mating was the primary reason for the weevils presence on the tree, mated females oviposited on the same tree, and the tree was damaged.

We should point out, in summary, that partial correlations should be used with caution: because of the properties of the technique, some correlations may appear spurious. In addition, there should be reasonable a *priori* justification for the relationships described through the techniques discussed above. However, we hope we have shown that when used appropriately, simple and partial correlations and factor analysis together provide power in descriptions of multivariate relationships.

LITERATURE CITED

Cattell, R. B. 1965. Factor analysis: An introduction to essentials. II. The role of factor analysis in research. Biometrics 21:405-435.

Comrey, A. L. 1973. A First Course in Factor Analysis. Academic Press, New York.

Eckert, R. T. 1974. White pine weevil response to host morphological and chemical factors and to micrometeorological conditions in a Norway spruce stand. M. S. Thesis. SUNY Col. Environmental Sci. and For., Syracuse, New York.

Kaiser, H. F. 1963. Image analysis. In: Harris, C. W. (Ed.), Problems in Measuring Change, Univ. Wisconsin Press, Madison, Wis.

Kempthorne, O. 1969. An Introduction to Genetic Statistics. Iowa State University Press, Ames, Iowa.

Morrison, D. F. 1967. Multivariate Statistical Methods. McGraw Hill Book Company, New York.

Steele, R. G., and J. H. Torrie. 1960. Principles and Procedures of Statistics. McGraw Hill, Inc., New York.

Veldman, D. J. 1967. Fortran Programming for the Behavioral Sciences. Holt, Rinehart and Winston, New York.

Table 1.--Simple correlations, partial correlations, and factors of the simulated common cause relationship.

Simple Correlations

Variable	2	3
1	0.8231**	0.8674**
2		0.6830**

Partial Correlations

Variable	2	3
1	0.6345**	0.7359**
2		-0.1095

Oblique Factors[a]

	Factor	
Variable	1	2
1	0.0121	0.9340
2	0.8510	-0.0295
3	0.8636	-0.0049

[a] Correlation between factors 1 and 2 = 0.91**

** Indicates significance at $p \leq 0.01$

Table 2.--Simple correlations, partial correlations, and factors of the simulated two independent causes relationship.

Simple Correlations

Variable	2	3
1	0.6735**	0.6868**
2		-0.0308

Partial Correlations

Variable	2	3
1	0.9562**	0.9577**
2		-0.9184**

Orthogonal Factors[a]

	Factor	
Variable		
1	0.6873	0.6650
2	0.0190	9.9458
3	0.9470	-0.0090

[a] Correlation between factors 1 and 2 = 0.00.

** Indicates significance at $p \leq 0.01$.

Table 3.--Simple correlations among measures of weevil activity, Norway spruce morphology and weevil damage.

Variable	Mating	Oviposition	Tree Height	Previous Leader Length	Previous Leader Diameter	Current Damage	Frequency Weeviling
Incidence	0.6688*	0.2446**	0.1627	0.2695**	0.0543	0.4104**	0.4348**
Mating		0.3451**	0.2093*	0.2927**	0.0539	0.4698**	0.3163**
Oviposition			0.0401	0.1001	-0.0190	0.3247**	0.2242**
Tree Height				0.7821**	0.1960*	0.4459**	0.4105**
Previous Leader Length					0.4566**	0.4557**	0.3888**
Previous Leader Diameter						0.0941	0.1235
Current Damage							0.5392**

*, ** $p \leq 0.05$ and $p \leq 0.01$, respectively.

Table 4.--Orthogonal factors of the Norway Spruce - white pine weevil association.

R^{2a}	Variables	Factors 1	2	3	4
50.4	Incidence	0.0949[b]	0.6777	0.1285	0.1370
51.6	Mating	0.0335	0.4233	0.1659	0.5551
15.5	Oviposition	0.0239	0.3761	0.0314	0.1103
64.7	Tree height	0.5874	0.1584	0.5226	0.0630
70.8	Length previous leader	0.2679	0.0907	0.7741	0.1703
27.6	Leader Diameter	0.5158	0.0225	0.0926	0.0358
44.3	Damage	0.1577	0.4772	0.4105	0.1466
38.0	Frequency of weeviling	0.0685	0.2505	0.4007	0.3905

[a] The coefficient of multiple determination derived from regressing each variable against those remaining.

[b] Factor loading or correlation between a factor and a variable. Variables which each load highly with a given factor are correlated.

Table 5.--Oblique factor pattern of the Norway Spruce - white pine weevil association.

		Factors			
R^{2a}	Variables	1	2	3	4
50.4	Incidence	-0.0318[b]	-0.7484	0.0414	-0.0690
51.6	Mating	-0.0305	-0.0527	-0.0023	0.7545
15.5	Oviposition	-0.0264	0.3737	-0.0295	0.0267
64.7	Tree height	0.4035	0.2144	0.4512	0.0614
70.8	Length previous Leader	-0.0673	-0.0296	0.8421	0.0867
27.6	Leader Diameter	0.5510	0.0541	-0.0550	0.1975
44.3	Damage	-0.0706	0.4857	0.3925	-0.0390
38.0	Frequency of weeviling	-0.1047	-0.0910	0.3492	0.4689

[a] The coefficient of multiple determination.

[b] Factor loading. For oblique factors this represents the degree of association between a variable and a factor.

Table 6.--Correlations among oblique factors.

Factors	2	3	4
1	-0.1382	0.4640	-0.1989
2		0.2772	0.8104
3			0.3766

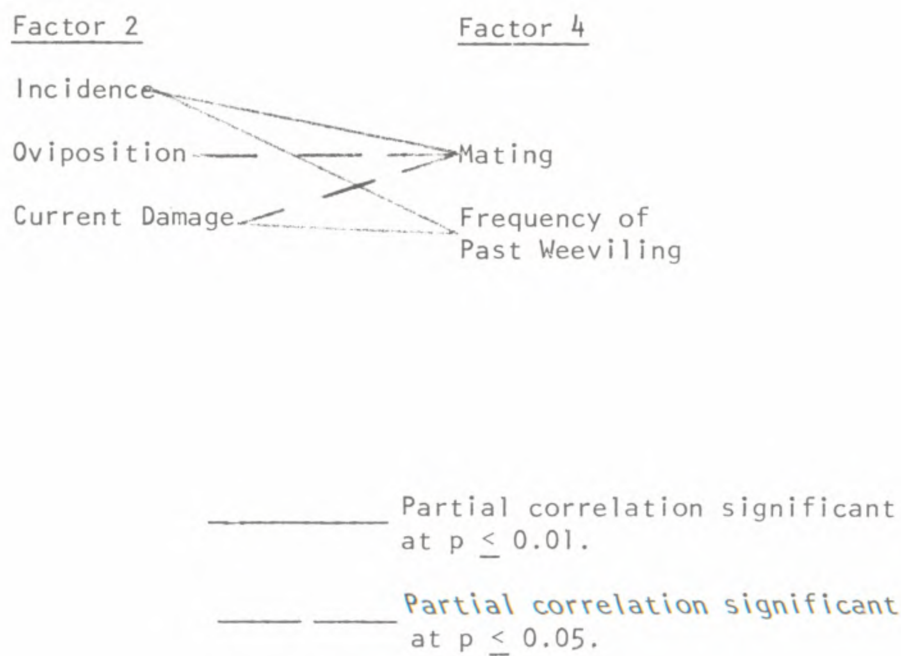


Figure 1.--Relationships among variables on Oblique Factors 2 and 4.

DISCUSSION

<u>Haskell</u> - I wonder if, in your selection, you recommend on the selective state one per family or two per family? In other words, will you make more gain allowing for some inbreeding versus the more gain you can get selecting more in a family?

<u>Morgenstern</u> - In the selection I'm concerned about the long-term effect of inbreeding and, therefore, mentioned to select only one tree per family.

<u>Haskell</u> - But you actually could make more gains in growth rate if you allow selecting more, for example, two or three within a family.

<u>Morgenstern</u> - Well, I realize this. It would be possible to keep track of ancestry, but this can become very complicated. You would have to consider the alternatives. Basically, I would select only one tree per family just to be on the safe side.

<u>Gerhold</u> - You mentioned some problems of interpreting data concerning variances and heritability estimates. I think you have some similar problems in considering biases that may influence the <u>means</u> of the families, and maybe this is even a more serious problem. For example, do you have random sampling of the pollen from populations, are there differences in the proportion of half-sibs among families, and so on. These can influence the means as well as the variances.

<u>Morgenstern</u> - Yes, I agree.

<u>Gerhold</u> - So, it's not clear in my own mind whether wind pollinated families are advantageous, or whether it's worth taking the extra time and effort for making the controlled pollination.

<u>Morgenstern</u> - Well, my concern is mainly to get the program underway, and I realize that there are other problems. Eventually, when the seed orchards have been established, it will be better to work with controlled crosses only. At this stage, when we have such a lack of data, plus-tree selection seems to be a barrier and the foresters are not willing to go ahead and select anything less than the best trees. They would like to do a perfect job; and as a result of this common attitude, the number of selections has been very small in some of the programs that I have seen. That's why we want to accelerate the program and these progeny tests would be a good way to start.

<u>Gerhold</u> - In addition to this, there is the possible trade-off between the increased number of families you may be able to handle through wind pollinated tests versus more definite information from controlled matings.

Schreiner - I was wondering whether there is any information on comparison of progenies of the same mother trees from seed collected in different years. Silen's project may include that. There could be wide progeny fluctuations from year to year where the interbreeding population is not restricted to the deme in sensu Carter-Mayr (defined by Langlet in the 1967 IUFRO Proc.).

Morgenstern - The best way to deal with this problem is by collecting in good seed years, and seed years in black spruce are more easily predictable than in other species.

Schreiner - Has anyone else been concerned with this problem?

Morgenstern - There is no information for black spruce, but some for Norway spruce. In this species, in some years only the early- or late-flushing trees have flowered as Sarvas in Finland found some years ago.

Demeritt - Kris, in your 7 x 7 diallel in black spruce did you break out a maternal reciprocal effect?

Morgenstern - Yes.

Demeritt - That could be one reason why your heritability is quite low or lower than your open-pollinated or the control-pollinated one-year-old material. The heritability could be inflated by maternal effects that one year. I think Kriebel found in eastern white pine that the general combining ability is about the same magnitude as the maternal-reciprocal effect at one, two, and three years.

Morgenstern - As shown in my paper in Silvae Genetica, the maternal and reciprocal effect was not as important as that. Probably the within-plot variance is a big factor accounting for some of the heritability differences shown on the slide. In the diallel cross, general combining ability was largest for such characters as germination rate and height growth but not for other characters like survival. Perhaps these results have some generality.

Kung - Did you run both the Q & R analysis from your data?

Eckert - No.

Kung - I think if you run both Q-analysis and R-analysis together, you will have grouping of trees and the relationship between the grouping of trees and the characteristics of weevil attack. I think it is important to run factoring on trees as well as factoring on weevil attack. By running both Q- and R- analysis you can get information which is worth double your money spent.

Eckert - Yes, that's something that we probably ought to be doing, I agree.

Morgenstern - The information obtained through this analysis should be related to previous knowledge on weevil damage, The "factors" can be interpreted and listed in the analysis. What kind of background information was used to evaluate this type of analysis? In factor analysis, meaningless correlations can be obtained. The variables initially chosen will determine the interpretation.

Eckert - Yes, this is true, but I spent quite a lot of time in the field with this study. The selection of individual variables is important in a factor analytic study, although, one of the uses of factor analysis is determining which variables are related to underlying modes of variation not completely described by one variable alone. Also, there are some techniques which estimate the relative importance of factor loadings based upon the proportion of R^2 for an individual variable attributable to any one factor. It is true that the interpretation of these factors is our own; however, it is based upon my field experiences and the results of techniques described by Cattel and Comrey for factor rotation. The confirmation of a hypothesis derived from a factor analytic study, or indeed any study, is substantiation in the field with further work. Factor analysis is an aid to thinking, it's not a proof, by any means,

Morgenstern - For example, in your analysis, weeviling was not related to tree size. This was a surprise to me because in some earlier studies weeviling was related to leader diameter.

Eckert - Another possibility might be to measure stem diameter and leader diameter with tree height. This would probably yield a more realistic tree volume factor. However, in this study, I was interested in leader characteristics. Leader volume would be a better description of Factor 1.

Feret - Can you elucidate what the advantages of this sort of analysis are in relation to step-wise discriminate analysis?

Eckert - They go hand in hand. Basically, discriminant analysis is one way multivariate analysis of variance. You must group data before analysis, then discriminant analysis will give you the relative importance of variables in discriminating among groups. Factor analysis requires no groupings and enables you to identify underlying modes of variation which, for example, may be similar to the discriminant functions you would obtain by doing discriminant analysis of the same data. Which analysis you use depends upon your objectives and your data.

Kung - To me, the clustering of variables and observations can be easily seen on a factor analysis which is very uncommon in step-wise regression. Peter was asking what are the advantages of factor analysis over step-wise regression?

Eckert - Did you say discrimination?

Feret - Yes, discriminate analysis.

Kung - To me the clustering of your data is the best feature of the factor analysis.

Eckert - Yes, and I want to point out again the problem of colinearity in regression analysis.

Kung - It goes R-analysis, Q-analysis. You can cluster your data in both directions.

Gerhold - A couple of comments, rather than questions. I think you may have implied that inbreeding is equal to inbreeding depression. Many of us are accustomed to thinking that way because after selfing we see serious inbreeding depression in many instances but with milder forms of inbreeding I'm not sure we ought to think of it in quite those terms. The other comment pertains to the separation of functions. You talked about separating the seed production function from the selection or testing function. I want to suggest separating two other functions, just during the process of analysis, and interpretation of data these are the processes of producing information and of selecting progenitors. It might be useful initially to think of them separately and then combine them in some fashion.

Weir - Dr. Gansel's work shows what is currently known about inbreeding as you develop crosses among half-sibs and full-sibs, of course, selfing being extreme. It would be a pretty straight line depression and I think the idea there was that the depression at half-sib mating level was serious. However, I do agree that we tend to equate inbreeding with depression and maybe we shouldn't. One thing is clear to me that we better learn how to live with some sort of inbreeding because you can only go through so many cycles, and unless you start with an infinitely large population, eventually you are going to be restricted and you're going to be forced into some sort of relatedness, co-ancestry if you wish, build-up. Everything we do in selection just seems to magnify this difficulty.

Gerhold - I have the impression that some tree breeders think we should permit less inbreeding than actually must occur in nature, That's an intuitive basis for my remark.

Weir - Yes, well here again consider the allowable amount of inbreeding you let build up in the breeding population versus what you allow in production orchards. While it is true that there may be a background level of inbreeding in nature that gives us our current wild seed situation, if we can avoid that in the artificial situation of the seed orchard, I think we would be that much better off in terms of realized gain.

Gansel - My comment on that particular question is that while there is a direct line in your inbreeding value of your F (depression of growth) it's true, but when you figure out exactly what the amount of inbreeding depression you get in, for example, a 30 or 40 clone orchard with about 3 or 4 half-sibs in there, it's so infinitesimal that you really may not be noticing it and that you do get considerably more value by selection 1 or 2 or 3 half-sibs or full-sibs or something similar to that rather than just eliminate them because you know that they are related.

Weir - Under assumption of random mating, I would agree and I think that maybe people have misunderstood our situation in that we're not avoiding use of related trees in a production orchard. However, when we do put them in, we try to additionally help out the situation by spacially separating them as you would ramets of the same clone. But my point in this unrelatedness is if you want to retain as much flexibility as possible you need maximum unrelatedness. The single pair mating is all you have to do to get completely unrelated families. That's all--anything you do beyond that is purely for selection differential and that type of thing is not going to give you any increase in unrelatedness. So it is a real factor that we are going to be dealing with for some time to come.

Westfall - I would like to point out some differences among the types of multivariate analysis. I think there are three types that are going to be particularly useful in genetics research. One Pete pointed out and that was discriminate function analysis. That's really no more than a multivariate one-way analysis of variance -- that's all it is. Vectors are lined in the direction of maximum discrimination between the groups. Now, this is very similar to the paper that Warren gave yesterday except that he extracted essentially 99 possible discriminate functions along 360°. From that analysis, you could chose those vectors which best discriminated among the Groups. A discriminate analysis has a constraint on that but it does a similar sort of thing. The second type of analysis is canonical correlation analysis and that's no more than a multivariate regression analysis, relating a multivariate set of dependent variables with a multivariate set of independency variables. It has some properties that I think regression analysis can't give us, at least not the way it's being treated now. The vectors out of a canonical correlation analysis are aligned in maximum correlation between the dependent variables and the independent variables. These are the most important vectors. And the last is the one we presented and that is factor analysis. That's nothing more than analysis of covariance, looking at the relationships among the variables themselves independent of dependency - whether they are independent or dependent is really irrelevant. That's all I have to say in response to that.

Gabriel - I wanted to ask what you had for moisture content in the sealed vials.

Farmer - In the vial? Our moisture content there, and this is based on pretty limited work, is a little over 10 percent.

Gabriel - My other question is have you tried any of these stored pollens on natural flowers?

Farmer - I haven't. Donovan Forbes has used some of the pollen that has been stored for one year successfully.

Gabriel - How did the seed sets compare with your germination figures?

Farmer - His seed set was adequate, but I don't think we can make a direct comparison of seed set and percent germination. All I know is that he has gotten good seed set with the stored pollen that he's used.

Kung - You have mentioned a standard treatment and you have shown that the range for germination went from 30 to 97 percent, but do you have figures on the percentage of pollen germinating on the female flower? And what medium you have used is comparable to that percentage? I won't accept a standard medium because you'll overestimate or underestimate the germination of pollen on the female flower.

Farmer - This does not represent germination percentage on a flower at all. It gives you an indication whether you've got live pollen or not. I would say that if you got 10 or 20 percent germination in in-vitro tests, you have pollen that is adequate to use in control crossing; but I wouldn't try to, based on my data, tell you anything about what kind of germination you get on the flower.

Kung - It is important that you try to find out a special medium which will yield germination percentage comparative to the female flower as a standard medium, instead of saying that a medium containing 95 percent water, 5 percent sucrose, and 5 ppm baron is your standard medium.

Farmer - We've gotten good germination on a simple medium here, and you probably could use plain agar and get a fairly reasonable estimate of viability. But we do note that there is some effect of boron; this has been found in a number of other species. We tested stored pollen on a number of media, but we did not test stored pollen on plain agar and water as we did some of the fresh pollen. It would be interesting to find out what kind of germination percent you get on these minimal kinds of media after storage periods. It would also be very interesting to determine, as Bill Gabriel has pointed out, what kind of seed set we're getting with stored pollen.

Rauter - When you tested your pollen you said that you had gotten your pollen tube elongation in about six hours. Does this hold true with your stored pollen?

Farmer - Yes, this is pretty much the same.

Rauter - There's no reaction or anything?

Farmer - In fact, you can check the stuff in 1 or 2 hours and tell whether you've got it or not. Very, very rapid growth.

Feret - Were there any genotype storage interactions?

Farmer - There was nota statistically significant clone x storage interaction in this particular test, but we did have one clone that seemed to have a greater reduction in viability than other clones after storage. Perhpas more intensive sampling would confirm a genotype x storage interaction.

Rauter - Just a comment here. We collected some of our materials for the last two years in January and February. Previously, we were doing it in April and May and we found we were getting good results. When we tried it in January and February and put it in the greenhouse, it did take an extremely long period of time for it to root and we didn't get very good results. When we put it in cold storage for three or four months and brought it out in May or June and rooted it, it seemed almost to have a tendency to go further ahead. It just seems to meet whatever dormancy requirements it has.

Farmer - How is the water quality you used for your cutting?

Garrett - It was never checked, so I really can't say.

Farmer - Is it tap water?

Garrett - The water we use in our greenhouses comes from an artesian well. The water has not been tested, but I know that it's pretty hard.

Dorn - You said you got up to 88 percent success. Did you mean successfully transplanted or just rooted?

Garrett - These are trees with good root systems on them. I assumed the study was going to end there but it didn't -- they've been out-planted in the field and as of August 1, they look good and healthy.

Miller - I wonder if it would be possible to report data like this in terms of "growth-degree-days" or something especially when it's collected in the growing season so we could relate it from one part of the country to another. Something like June 2 really doesn't tell us very much in Wisconsin.

Garrett - No, and it doesn't tell us very much either because June 2 one year is extremely different from June 2 the next year.

Miller - That's what I mean but "growth-degree-days" should be able to tie it down.

Garrett - That's why I put May in quotes as best collection time because if the same conditions are met the first of June the next year, I think the best rooting will be obtained in June and not earlier by the calendar.

Miller - You know the agricultural people now use "growth-degree-days" instead of saying corn will ripen in 60 days.

Garrett - I think that's what we're trying to say to you.

Farmer - I am interested in the stage of growth that shoots are in when cuttings are taken. By late May in our area, new shoots are elongated. You can't get good cuttings out of them. Are cuttings taken before your shoot elongation?

Garrett - Yes, the best success is just before shoot elongation.

Gabriel - If we are talking about the physiological condition of the tree at the time of collection of the cuttings, would you think that this would be the influential factor in the proper time to take your cuttings rather than the date?

Garrett - I think that's what Bob, Dick, and I have all been trying to get accross.

Gabriel - He was talking about the height. It is internal conditions that I'm talking about, not about the external appearances

Garrett - We really don't have any quick way of checking a tree's chemistry or other internal characteristics and saying OK today is the day to take cuttings. We have to look at external conditions of the buds, know something about the weather conditions, and have a lot of faith.

Gabriel - Well, this is the problem we're stuck with in sugar maple. If you figure this out, we'd be very happy to hear about it.

Garrett - OK.

Clausen - Have you any idea why they use this very complicated rooting medium? If would be very difficult for anybody else ever to reproduce it. I doubt that he could and that could have a lot of effect on trying to reproduce results. Also, I am curious about why he did not try to root the cuttings with no hormone treatment since to use nothing at all seems to be the current trend.

Garrett - I would like to be able to answer for him, but I can't.

Fowler - Are your cuttings developing into fairly normal types.

Garrett - Yes, I would say they are very normal. They're all upright cuttings, and we have not experienced any topophytic effects.

Weir - I've got a couple of comments and a question in the middle, too. One, I think you are to be congratulated. We've never been able to equal Mother Nature in the South with our controlled pollination, and it seems like you're in good shape there. For my own information, I'm curious why did you paint the pollen on instead of using something like a syringe to inject it?

Sayward - We were working with a breeding arboretum slightly smaller than one acre in size, and so far our pollen yield has been very low. At times we were working with as little as 2 cc's of pollen, and it was usually a question of making it go far enough. For the sources for which we had a lot of pollen, we used the same method all the way rather than switching back and forth. We found we could go reasonably fast if we kept the paint brush in a large soda straw fastened to each pollen vial. When that particular pollen source was to be used, we picked out the vial-brush complex, cut open the bag, brushed the pollen onto the flowers, resealed the bag by folding and stapling, tagged the branch, and moved on to the next bag or tree.

Weir - One other comment. I noted your prescription for irrigation was so much after an inch of rainfall, just a suggestion--something we found very successful in an operational scale in our seed orchards in the south is a tensiometer, the actual brand name of this thing is an Irrometer and these, once they get calibrated properly, have been very reliable. They are permanently placed in the soil, or you can take them out for the winter in this climate. It gives a reading on moisture stress, moisture tension, in the soil and it's a very good gauge to determine irrigation needs. I just wondered if you had tried any; and if not, I suggest it as something you might think about.