Earl,

I propose to submit this to FOREST SCIENCE
Will you review this and send me a letter of authorization?
Lee

NOTE ON THE FITTING OF
LOCAL VOLUME FUNCTIONS USING ORTHOGONAL POLYNOMIALS

by

Lee C. Wensel
Assistant Professor

and

Jan Van Roessel
Assistant Specialist

School of Forestry and Conservation
University of California
Berkeley

This work was carried out under contract with the California State Division of Forestry as part of an inventory analysis project for the Latour State Forest. All computer work was done at the University of California's Computer Center.

MAY BE PRINTED IN HILGARDIA SOON
LATEST WORD
6/64
EJ

Introduction

A "local" volume equation simply expresses a functional relationship between the volume of a tree and its diameter at breast height. It differs from a "standard" volume equation in that the effects of varying tree height and form are expressed implicit~~y~~ly by the sample trees selected rather than by an equation including terms for these factors. In a sense then, the effects of tree height and form are "averaged out" by the sampling process.

Local volume equations are necessary components of many inventory analysis programs since often tree diameter is the only tree size variable recorded in the inventory. In many cases, a local volume table is constructed using relatively small amounts of data to establish a diameter-height relationship and then the nominal tree volumes for each diameter class are obtained from a standard volume table for that species (or form class). These volumes are then smoothed in some way and tabulated. These tables may be adjusted with experience. Although these kinds of tables are very subjective in nature they are used widely because of the costs associated with building more objective tables.

The problem treated here is to transform these tabulated volumes into equation form so that they can be used in an inventory analysis program. In particular, this paper is an outgrowth of the process developed to transform forty (40) separate local volume tables into equation form for use in a series of inventory analysis programs written to process the continuous forest inventory for the Latour State Forest near Reading (Redding), California. The volume tables used to construct the volume functions are in part from the U.S. Forest Service (1956) and, for species not represented in these tables, from tables prepared by the Staff at the Latour State Forest.

Desirable properties of the local volume function

In addition to desiring the functional values to be approximately equal to the tabulated volumes, there are a number of desirable properties that the local volume equation should possess. Firstly, since usually a large number of equations are involved for any inventory (one for each species or species group), the type of equation selected should be general enough so that the same fitting procedure can be used for each.

The second desirable property is that, evaluated at 1/10 - inch intervals, it should be monotonic increasing. That is, volume increases as diameter increases.

Lastly, the equation should not be negative in the diameter range that it will be used over. Since the values of the equation at the lower end of the range approach zero, it is quite possible that an irregular trend here could (and often does) cause the fitted curve to dip below the axis yielding negative volumes.

Since the straight line satisfies the above criteria it should be apparent that it is quite possible that no equation can be found that both satisfies the above criteria and also has an acceptable standard error. What is presented here is a systematic method of fitting a family of equations to a set of data and then choosing the "best" equation from this family.

The m^{th} degree polynomial

The obvious equation that meets the criterion of being completely general in application is the m^{th} degree polynomial

$$v_j = \sum_{i=0}^{m} \beta_i d_j^i + e_j \qquad (1)$$

where v_j is the volume of trees in the j^{th} diameter class of d_j inches, the β_i $(i=0,1,2,\ldots,m)$ are the regression coefficients, n is the number of powers included, and e_j is the difference between the tabulated and functional volumes. It is important to recognize that the v_j are not random variables and hence the usual statistics of multiple regression are not applicable in the classical sense.

This function can be applied generally using the usual procedures of multiple linear regression. However, there are two computational difficulties involved: (1) the high correlation of adjacent powers often leads to numeric singularity in the cross product matrix which must be inverted for the solution, and (2) since the computed coefficients are not independent, the entire set of coefficients must be recalculated for changes in the model.

The m^{th} degree polynomial consisting of orthogonal polynomials

Because of the difficulties encountered with fitting conventional polynomials, and for other reasons that will follow, orthogonal polynomials were resorted to. The entire procedure that follows is operational in a FORTRAN program called VOLTAB which is available for distribution.

The use of orthogonal polynomials implies that each of the powers of the variable d_j in the above model is replaced by a polynomial function as follows:

$$v_j = \sum_{i=0}^{m} \theta_I p_i(d_j) + e_j \qquad (2)$$

where $p_i\ (d_j)$ indicates the polynomial function of power i for d_j and the θ coefficients are similar to the β coefficients in equation (1). Let s_i be the unbiased estimate of θ_i.

Orthogonal polynomials can be obtained by means of the Gram-Schmidt scheme. This amounts to finding an orthogonal basis for the vectors[1/]

$$D,\ D^2,\ D^3, \ldots, D^m.$$

However, the quantities $d^3_i, \ldots, d^m_i$ can become extremely large or small causing great inaccuracies in the calculations. Forsythe (1967) suggests the use of a three-term recurrence scheme that has the advantage that the large values can be avoided altogether. Moreover, recursion schemes are very adaptable to computer applications. Also, when the diameters are scaled down to approximately a (-2,2) interval, the orthogonal polynomial values will all be within this range, no matter what the degree of the polynomial model may be. Then numerical problems, due to the size of the powers of the diameters when using the conventional model, are entirely avoided as overflow conditions cannot be created.

1/ All arrays will be designated by capital letters. All vectors are column vectors and the scalar product of two vectors will be denoted by (A,B).

This three-term recurrence scheme is as follows:

$$p_o(d_j) = 1$$

$$p_1(d_j) = d_j p_o(d_j) - \alpha_1$$

$$p_2(d_j) = d_j p_1(d_j) - \alpha_2 p_1(d_j) - \beta_1 p_o(d_j) \qquad (3)$$

$$\vdots$$

$$p_i(d_j) = d_j p_{i-1}(d_j) - \alpha_1 p_{i-1}(d_j) - \beta_{i-1} p_{1-2}(d_j)$$

where the α and β constants are defined by:

$$\alpha_i = \frac{\sum_{j=1}^{n} d_j p_{i-1}^2(d_j)}{\sum_{j=1}^{n} p_{i-1}^2(d_j)} \quad \text{and} \quad \beta_1 = \frac{\sum_{j=1}^{n} p_i^2(d_j)}{\sum_{j=1}^{n} p_{i-1}^2(d_j)} \qquad (4,5)$$

or in vector notation

$$\alpha_i = \frac{(D\, P_{i-1}, P_{i-1})}{(P_{i-1}, P_{i-1})} \quad \text{and} \quad \beta_j = \frac{(P_i, P_i)}{(P_{i-1}, P_{i-1})} \,. \qquad (4', 5')$$

The vectors P_i $(i=1,2,\ldots,m)$, consisting of the orthogonal polynomial function values, have been chosen to be orthogonal, that is

$$\sum_{j=1}^{n} p_i(d_j)\; p_k(d_j) = 0 \qquad (i \neq k) \qquad (6)$$

or in vector notation

$$(P_i, P_k) = 0 \qquad (i \neq k)$$

where P_i denotes the vector of polynomial values of power i.

Now if P represents a matrix with the vectors $P_1, P_2, \ldots, P_m$ as columns, then this matrix will be an orthogonal matrix, so that P'P will be a diagonal matrix since the diagonal elements are zero by condition (6).

This reduces the normal equations to a set of independent equations of the type

$$s_i \sum_{j=1}^{n} p_i^2(d_j) = \sum_{j=1}^{n} p_i(d_j)v_j \qquad i=1,2,\ldots,m \qquad (6')$$

which greatly simplifies the computations involved.

In the formula for the solution of the least squares normal equations

$$S = (P'P)^{-1}P'V \qquad (7)$$

where V represents the vector of volumes with n elements, $(P'P)^{-1}$ will also be a diagonal matrix with rank m, its elements being simply the reciprocals of the elements of the matrix P'P. Therefore, matrix inversion is unnecessary in solving for the coefficient vector S. This can also be seen from equation (6).

This can be interpreted geometrically as follows: least squares estimation consists of projecting the observation vector onto the space spanned by the columns of the matrix P. Since in the above case the columns of P are vectors coinciding with the orthogonal axes by means of which the column space of P can be defined, the V vector can be projected separately onto each of these column vectors, and the sum of these individual projections represents the estimate of the mean vector of the hypothesized model. The general formula for the projection of one vector onto another is

$$b_j = \frac{\sum_{I=1}^{n} c_i a_i}{\sum_{i=1}^{n} a_i^2} a_j \qquad j=1,2,\ldots,n$$

or in vector notation

$$B = \frac{(C,A)}{(A,A)} A$$

where B is the projection of A onto C, and both vectors have n elements.

Applying the expression for the scalar constant in the above formula to the projection of V onto the columns of P, we obtain the following expression for the elements of the vector S:

$$s_i = \frac{\sum_{j=1}^{n} v_j p_i(d_j)}{\sum_{j=1}^{n} p_i^2(d_j)} \qquad i=1,2,\ldots,m$$

or in vector notation

$$s_i = \frac{(V,P_i)}{(P_i,P_i)} \qquad i=1,2,\ldots,m$$

Identical coefficients are obtained by multiplying the matrix $(P'P)^{-1}$ with the matrix P'V as in equation (7) above. The result can be obtained directly from equation (6').

Selection of degree of orthogonal polynomial

How many terms(powers) should be included in the model (2) in practice? Certainly, since we want the functional values of the volume to agree as closely as possible to the volumes assigned to trees of the respective classes, a large number of terms might be suggested. However, it is hardly worthwhile to include terms that add little or nothing to the relationship. The development of the previous section suggests the method for decoupling equation (2).

Since $\sigma^2 \, (P'P)^{-1}$ represents the covariance matrix of the vector S, and the off-diagonal elements will be zero, it can be concluded that the elements of S are uncorrelated and independent. Hence, the individual elements of S can be tested for significance and rejected on an individual basis through a proper test without invalidating the rest of the S vector.

If the variance of s_i is estimated by

$$\hat{\sigma}^2 = \frac{\sum_{j=1}^{n} [v_j^2 - \sum_{i=1}^{m} s_i^2 \, P_i^2 \, (d_j)]}{n-m}$$

then the null hypothesis $H_0: \theta_1 = 0$ can be tested using the statistic

$$t_{n-m} = \frac{s_m}{\sqrt{\hat{\sigma}^2 / (P_i, P_i)}} .$$

Conclusions

The advantages of the use of orthogonal polynomials are thus the following:

1) The matrix inversion for the solution of the normal equation is simplified and expedited through the diagonal character of the matrix to be inverted.

2) A model with a maximum number of power terms can be computed since the terms are additive and independent. Thus, the proper model can be determined through backward testing of the significance of the expression coefficients, beginning with the highest power computed.

3) A model with any number of powers can be computed without difficulties arising through high correlation between adjoining powers and numerical difficulties from extremely small or large numbers, as in a conventional polynomial model.

4) The above conditions allow testing a great number of models with a simple algorithm thus allowing the selection of a volume equation model which will not give declining values or negative values in the projected range and at the beginning of the table respectively.

The Solution, the VOLTAB computer program

The input to the program for each volume table to be produced consists of a table card indicating species, site, Scribner or cubic volume, the range over which the data are to be fitted and the range over which the fit has to be extended. This card is followed by a set of cards containing the volume data of the sample tree or volume data from the hand-fitted table (as in the Latour case) for even inch diameters.

The program performs the following tasks:

1) An eight power model is fitted for each table.

2) The number of statistically significant terms in this model is determined by performing a test on the coefficients.

3) A check is made to see whether the optimum statistical model contains negative or decreasing values, and if so, the next best model is selected, etc. If no model can be found without negative or decreasing volumes, the model with the smallest number of these values is selected and a warning is printed in the table summary. Whenever negative values at the beginning of the range are less than the calculated standard deviations for the table, the negative values are replaced by zeros.

Finally, for each table a summary is printed, presenting the information of the table heading card, the number of powers for the optimum fit and the number of powers for a table with a permissible range. Corresponding standard deviations are also given.

Forsythe, A. C. 1967. Generation and use of orthogonal polynomials for data fitting with a digital computer. SIAM Journal (5)

U. S. Forest Service. 1956. Tenth-inch volume tables for the commercial conifer species of California. Forest Survey, California Forest and Range Experiment Station.