Measuring and Modeling Crop Losses at the Field Level

L. V. Madden

Assistant professor, Department of Plant Pathology, Ohio Agricultural Research and Development Center and the Ohio State University, Wooster 44691.

Salaries and research support provided by State and Federal Funds (especially a USDA Crop Loss Grant) appropriated to the Ohio Agricultural Research and Development Center, the Ohio State University. Journal Article 150-82.

A more detailed list of literature references is available from the author.

I thank S. S. Mendiola for assistance with the preparation of this manuscript, and R. E. Gingery, J. K. Knoke, and K. Leonard for helpful suggestions.

Accepted for publication 24 May 1983.

It is a paradox that there are few reliable estimates of crop losses due to plant diseases, even though plant pathology has developed as a separate science largely because of those losses (15,23,30). This lack of loss data may have retarded the progress of plant protection as much as any other single factor (23).

The concept of Integrated Pest Management (IPM) developed in the 1970s, but one of the biggest limitations to implementing IPM programs has been the paucity of information on crop losses (2). Total loss of a crop due to all pests has often been "estimated" at greater than 100% (36). For example, in their review of the literature, Pimentel et al (36) found that "estimates" of apple losses due to diseases, insects, and weeds totaled 126%! It is no longer acceptable to merely guess the loss due to pests in general, and diseases in particular. Optimal controls, or combination of controls, cannot be determined without reliable loss data. Clearly, no IPM program will be acceptable without valid crop loss estimates.

This paper presents reviews and summaries of selected concepts in measuring, modeling, and predicting losses caused by plant diseases with emphasis on the types of experiments necessary to obtain data useful for developing loss models. Empirical models are described, evaluated, and then compared.

MEASURING CROP LOSSES

Crop loss is defined as the measurable reduction in quantity and quality of yield (16,45). Zadoks and Schein (45) provide a thorough description of the categories of yield as well as the many forms of direct and indirect losses caused by plant diseases. Yield can be represented and analyzed in original units (eg, volume or weight) or as a percent (proportion) of a disease-free control. Percent data will not give the same results as data with units (25). Unless indicated otherwise, only quantity losses due to disease or diseases in combination with crop and environmental variables will be discussed in this paper.

Designing experiments. James (16), and James and Teng (23) presented a detailed review of the methodology for measuring crop losses. Some of their suggestions are elaborated herein, especially as to how they relate to experimental designs. Five factors should be considered in designing experiments for measuring crop loss (16). First, the experiments should be conducted in more than one location, and preferably in areas where the crop is normally grown and the disease is important. In some circumstances, however, conducting experiments where the pathogen is not found permits more thorough control over the level of disease intensity. For example, we are determining losses of maize due to maize dwarf mosaic virus (MDMV) in plots in northern Ohio where the virus is usually not present in order to have control plots nearly free of disease (J. K. Knoke, R. Louie, and L. V. Madden, *unpublished*).

Second, experiments should be performed for more than one year. Although the number of years necessary to achieve precise crop loss estimates is uncertain, James (16) recommends three. More years may be needed with a highly variable environment.

Third, experiments should be conducted with the major cultivars grown in the region of interest. If possible, these cultivars should represent a range of susceptibilities to the pathogen.

Fourth, multiple levels of a treatment should be applied to potentially achieve multiple levels of crop loss rather than the traditional two-level experiments of "infected" and "healthy" treatments which are still common today. Multiple levels of crop loss can be obtained by varying inoculation or protection methods, crop cultivars, or locations (16). Plants (plots) can be inoculated at different *times* and *frequencies*, and the concentration of the inoculum can be altered. Because it is difficult to obtain uniform and repeatable levels of disease intensity with variation in inoculation, much of the crop loss research still relies on natural infection and variation in other treatments to produce multiple levels of disease intensity. Protectant chemicals can be used to vary the level of disease, although tests should be performed to determine if the chemicals affect yield in the absence of disease (16). Other forms of protection can also be used, such as oil sprays to prevent infection by stylet-borne viruses (46). Isolines of the crop that differ only in genes for susceptibility to a given pathogen can be used to vary disease intensity, but such isolines are not very common. Multiple levels of diseases may be obtained by planting in geographical areas that vary in the prevalence of a particular pathogen. In this approach, unfortunately, the effects of disease intensity are confounded with effects of several other factors at each location, such as weather and soil properties.

Fifth, disease intensity often should be used as a covariable in the analysis *instead* of the preassigned treatment level. Since intensity for most diseases can be represented on a continuous scale, and resulting disease intensity may be very different from that intended, the most powerful statistical analyses utilize this information to derive precise crop loss estimates.

Experimental design restrictions. There are many restrictions on experiments for determining crop losses. Plot size and shape, interplot distance, and number of replications should be selected to produce results representative of losses in commercial fields. Few crop loss researchers have evaluated these factors. Preliminary experiments should be conducted to determine sample sizes needed to produce results with a certain level of precision (19). Often there is a conflict between the desired level of precision and that obtainable in the experiment within realistic limits of time and resources.

The three most commonly used experimental designs for determining the effects of disease intensity on crop losses are: randomized complete block (RCB), split plot, and Latin square. Each design has advantages and disadvantages (13,35). Latin squares are the most limited because they are not very powerful with less than five treatment levels, but become unmanageable with more than eight (35). The split plot design, originally developed for agronomic applications, may or may not be a satisfactory alternative to the very common RCB. To compare these two designs, consider an experiment with four replications (R), 10 cultivars (C), and five levels of disease intensity (D). Table 1

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. § 1734 solely to indicate this fact.

contains the degrees of freedom breakdown for a RCB design, split plot design with D as the subplot, or split plot with C as the subplot. The number of degrees of freedom (df) for the appropriate error term is one measure of the power of the analysis of variance (ANOVA) tests (13). As is usually the case, assume that the interaction of R with the whole plot serves as the error term for the whole plot, and the other interactions with R are nonexistent. With RCB, there is only one error term which has 147 df. With D as the subplot, there are 120 df for testing D effects, and only 27 df for testing C effects. With C as the subplot, there are 135 df for testing the effect of C, and only 12 df for testing D. Clearly, on the basis of df alone, RCB is the best approach for testing the significance of C and D. Split plot designs, however, have advantages. If randomization is restricted because it is impossible to randomly assign all levels of all factors to locations in a field, then split plots may be necessary. Additionally, adjacent plots or rows may be much more homogeneous than separated plots or rows. This increases the sensitivity of the split-plot analysis, even though the df are lower. If a split plot is used, the division of subplots should be chosen to provide the higher df for the factor of greatest interest. If evaluation of cultivars is the main concern, then C should be the subplot. If disease effects are the main concern, then D should be the subplot. These same arguments can be used for other more complicated experimental designs.

PREDICTORS OF CROP LOSS

Crop loss in its entirety is a function of disease, insects, weeds, other pests, and environment. For the limited purposes of this review, crop loss can be considered a function of the disease epidemic plus an error term, where the error term accounts for the unexplained variability. Disease-loss models use one or more epidemic characteristics to predict crop loss. In the statistical jargon, these characteristics are called predictors, carriers, or independent variables, even though they are seldom independent (34,35).

The graphical depiction of a typical epidemic, ie, a disease progress curve, is presented in Fig. 1, and shows several epidemic characteristics including the final level of disease (X_f), disease at any given time (X_t), initial amount of disease (X_0), time of epidemic onset (t_0), rate of increase (slope) at any time (r), and area under the disease progress curve (*ADPC*). Models for crop loss can be categorized as those using single or multiple independent variables. Models with a single predictor, either disease intensity at a particular time or the time at which a certain level of disease is reached, are called *critical-point* models (16,23). Zadoks and Schein (45) pointed out the limitations of using these single predictors. The greatest error lies in the assumption that all disease progress curves reaching the same level at a particular time will cause the same crop loss. This assumption is seldom true. Nevertheless, in a survey of papers from 1948 to 1982 dealing with empirical modeling of crop loss as a function of disease, I found ~57% considered only a single independent variable; ~77% of the papers considered a single predictor or compared single to multiple predictors. These papers demonstrated adequate to excellent success using single predictors. A partial explanation for this success is that disease at a particular time is actually a summation or integration of the change in disease levels throughout an epidemic (26,29). Disease at time t is highly correlated with disease at time $t-1$ (26). By considering only one predictor, a researcher is, in essence, using a great deal of information about a given epidemic.

The other single predictor that has often been used is *ADPC*, which integrates the level of disease for the entire epidemic. A similar predictor is the weighted disease average of Hills et al (14). Approximately nine percent of the crop loss modeling papers in the survey considered *ADPC* or the weighted disease average as the main predictor; ~20% of the papers used *ADPC* with other predictors.

Multiple predictors in crop loss models include: levels of disease intensity at several times; other epidemic characteristics such as time of epidemic onset and rate of increase; and disease levels in combination with other crop characteristics. Models that utilize disease levels, or changes in disease levels at several times have been called *multiple-point* models (16). Multiple-point models are usually more accurate than critical-point models because of their more thorough description of the epidemic. However, the number of predictors may become too large and cumbersome to use, especially for survey situations. Only ~7% of the surveyed crop loss modeling papers considered multiple-point models alone; ~20% of the papers considered multiple point and other epidemic or crop characteristics.

TABLE 1. Comparison of degress of freedom (df) for a randomized complete block (RCB) and two versions of a split-plot design for an experiment with four replications (R), 10 cultivars (C), and five levels of disease intensity (D)

Design[a]					
RCB		Split-plot[b]		Split-plot[c]	
Factor	df	Factor	df	Factor	df
R	3	R	3	R	3
C	9	C	9	D	4
D	4	Error (C)	27	Error (D)	12
C × D	36	D	4	C	9
		C × D	36	C × D	36
Error	147	Error	120	Error	135
Total	199	Total	199	Total	199

[a] D and C are considered fixed effects; interaction of R with whole plot equals the whole plot error; other interactions with R are considered nonexistent.
[b] C as whole plot; D as subplot.
[c] D as whole plot; C as subplot.

PROCEDURES FOR MODELING CROP LOSS

Regression analysis. Of the many possible ways of representing crop loss as a function of disease epidemics, linear regression is by far the most popular. Regression analysis is a very powerful tool in situations in which a dependent variable (eg, yield and crop loss) varies with an independent variable or variables in a systematic fashion, and the scattering of observations around the curve follows a statistical relationship (35). The linear regression model can be represented by:

$$Y_i = \Sigma_j^P (B_j Z_{ij}) + u_i \quad (1)$$

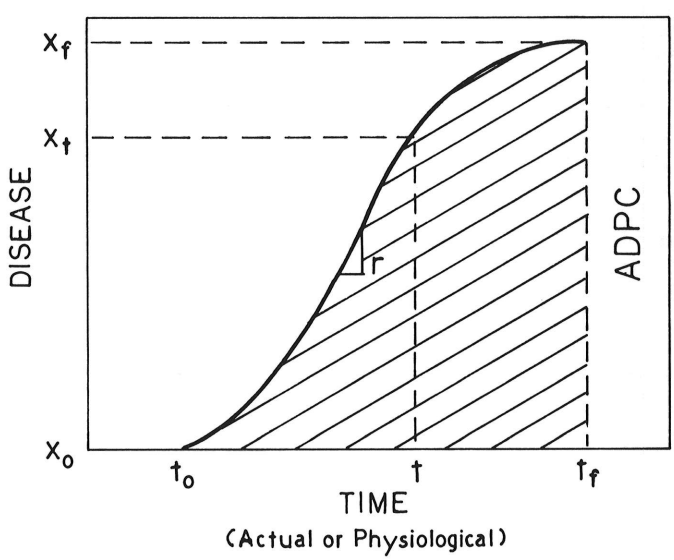

Fig. 1. Hypothetical disease progress curve depicting several curve characteristics. X_0 = initial amount of disease; X_f = final level of disease; X_t = disease at time t; r = rate of disease increase; and $ADPC$ = area under the disease progress curve. See text for details.

in which: Y_i is the dependent variable for the i-th sampling unit (eg, plant, row, or plot); Z_{ij} is the j-th independent variable for the i-th sampling unit; B_j is the j-th parameter, reflecting the weight given to Z_{ij} in constitution of Y_i; p is the number of independent variables; and u_i is the i-th error term (unexplained variability). Examples of Z are the epidemic characteristic discussed above. The single predictor form of equation 1 is:

$$Y_i = B_0 + B_1 Z_{i1} + u_i \quad (2)$$

in which Z_{i1} can represent: X (disease intensity) or a transformation of X at a single time, $ADPC$, or time at which a certain X is reached. Models with the unknown parameters replaced by the estimates from the data are presented without the i subscripts and error term, and are called prediction equations. An example of a critical-point model is the equation derived by Gregory et al (7) for crop loss of corn (L = crop loss) in relation to the severity of southern corn leaf blight at the dough stage (X_d):

$$L = 0.69\ X_d. \quad (3)$$

In this case, $B_0 = 0$ and $B_1 = 0.69$. For every 1% increase in X, loss increased by 0.69%.

The multiple predictor form of equation 1 can be written as:

$$Y_i = B_0 + B_1 Z_{i1} + \cdots + B_p Z_{ip} + u_i \quad (4)$$

in which the Zs represent levels of disease intensity, changes in disease intensity, or other epidemic and/or crop characteristics. An example of this type of model is the equation of Burleigh et al (4) for loss (in percent) of wheat yield due to stem rust:

$$L = 5.38 + 5.53\ X_2 - 0.33\ X_5 + 0.50\ X_7 \quad (5)$$

in which X_2, X_5, and X_7 are the levels of stem rust intensity (X) at three growth stages of wheat (4). The use of a measure of disease intensity with a crop characteristic is exemplified by the model of Scott and Hollins (38) for crop loss (L) of wheat:

$$L = -0.19 + 0.41\ Z_2 + 0.17\ Z_1 \quad (6)$$

in which: Z_2 is a lodging index and Z_1 equals the percent of shoots with severe eyespot disease.

With many diseases that are measured as incidence only (proportion of plants infected), regression analysis provides a useful mechanism for estimating the yield per infected and healthy plant. The following model can be used when, as is usually the case, it is more convenient to harvest entire plots (rows) rather than individual plants. The model for yield per plot (Y) can be written as:

$$Y_i = B_1 Z_{i1} + B_2 Z_{i2} + u_i \quad (7)$$

in which: Z_1 is the number of infected plants per plot; Z_2 is the number of healthy (disease-free) plants per plot; B_1 and B_2 equal the yield per infected and healthy plant, respectively. Because there can be no yield when there are no plants, B_0 equals zero. Once the parameters are estimated, they can be compared with a t-test (35) to determine if there is a difference in yield between infected and healthy plants. Compensation of plants for each other can be tested by estimating the parameter for the interaction of Z_1 and Z_2 (ie, B_{12}) and testing whether B_{12} is significantly different from zero. If B_{12} is not different from zero, the data do not support the hypothesis of compensation. We are using this approach to determine crop losses of maize due to strains of MDMV.

Regression model evaluation. To properly evaluate or verify a model, one should be aware of the statistical assumptions inherent in the procedure used to develop the model. The us are assumed to be normally and independently distributed at each level of Z with mean zero and constant variance σ^2 (35). The us are estimated by the residuals (observed minus predicted Ys), which can be evaluated graphically, and also with formal tests (11,35). The residuals are the keys to model evaluation; a model may be inappropriate if a plot of residuals versus the predicted Ys follows a nonrandom pattern (35). Nonconstant error variance and outliers can be determined with the same plot. Non-normality can be evaluated with a plot of the residuals versus their "normal" probability scores (11,35). Formal tests are available for all of these evaluations although often the graphical presentations are more revealing.

The significance of the relationship between Y and the collection of Zs is evaluated with an F-test; individual B parameters can be tested with t- or F-tests (35). The square root of the residual mean square, the standard error about the line (S), can be used as a measure of precision. Precision is inversely related to the standard error. Since the size of S is a property of the units of Y, often a unitless measure of precision is needed. The coefficient of determination (R^2), which ranges from 0 to 1, is a unitless measure of the proportion of explained variability, and is directly related to precision. Most disciplines have rules-of-thumb regarding minimum acceptable values of R^2. In well-controlled lab studies, R^2 is expected to be above 0.95 or even 0.99. For crop loss studies, I have found values above 0.80 to be excellent, and above 0.90 to be unusual, especially when data from all replications are used in the analysis. With linear models, R^2 must increase or remain constant as additional variables are added to a model, even if the additional variables are not significant (35). The coefficient of determination adjusted for degrees of freedom (R_a^2), on the other hand, will decrease if nonsignificant variables are added to a model. R_a^2 is, therefore, an excellent metric for comparing linear models with a different number of Zs. Some crop loss researchers use this measure of precision (6). R_a^2 need not be used when only one Z is being considered.

Choosing which independent variables to use in a linear regression model from a large set of possible variables is a complex problem with no unique solution. With the advent of high-speed computers with large memories, many automated, stepwise procedures for variable selection were developed and are now commonly used. Most statistical packages have more than one procedure for variable selection (5,10). I have found it rare for any two of these stepwise procedures to result in the same subset of independent variables. In the last few years, interactive computing systems have become more common, and so have interactive statistical programs for choosing independent variables for a regression model (11,37). These techniques permit a fuller appraisal of the data and can result in a better final subset of variables, but they do require more knowledge of statistics than earlier automated methods. However, blind acceptance of results of automated programs without statistical advice is a very dangerous practice.

When the predictors, eg, levels of disease intensity at several times during an epidemic, are highly correlated with each other, estimated parameters will have large variances and may appear to be nonsignificant. Estimated parameters could even have signs opposite of that expected on theoretical grounds. Two solutions to this problem are the use of *ridge regression* and *principal components* regression. Ridge regression became popular in the 1970s as a means of reducing the variance of estimated parameters by producing (slightly) biased parameters (32). A few programs are available for this type of analysis; we are using a Macro in the MINITAB system (37) for this purpose. I find ridge regression to be very subjective and do not recommend it to anyone who does not have a strong background in applied statistics. Principal components regression will be discussed in the multivariate statistics section below.

Nonlinear regression. Madden et al (28,29) used a nonlinear regression model to relate crop loss to disease intensity. Nonlinear models cannot be written in the form of equation 1. Instead, with most nonlinear models the parameters appear as exponents or are multiplied or divided by other parameters. The model of Madden et al (28), a form of the Weibull cumulative distribution, can be written as:

$$L_i = 1 - \exp(-((X_i - d)/b)^s) + u_i \quad (8)$$

in which: X_i is disease intensity at one time, L_i is crop loss

(represented as a proportion), and *d, b*, and *s* are parameters. (For interpretations of these parameters, see [28].) Another useful nonlinear model, which was suggested by Seinhorst (39), is presented here as a prediction equation (ie, no *u* or i subscripts) with different symbols:

$$y = m + (1-m)N^{(Z-t)} \quad (9)$$

in which: y is relative yield (proportion), Z is the population density of nematodes, and N, m, and t are parameters. N is the nematode damage potential, t is the tolerance below which no crop loss occurs, and m is the minimum relative yield at the highest population density of nematodes.

Estimating parameters of nonlinear models is computationally complex and may not even work. Large data sets are usually required and good initial "estimates" of the parameters are necessary (26). If the estimated parameters are highly correlated (>0.95), they may compensate for each other and, therefore, the estimates may not correspond to the optimal solution. I believe that nonlinear models should be used only when there are good theoretical reasons for doing so.

Path analysis. Path analysis uses regression techniques, but is based on more refined assumptions and allows more refined interpretation than the methods described above. Although used successfully in research in sociology and psychology, path analysis has rarely been used in phytopathology. It is a statistical technique for assessing the causal order among variables in a system closed to outside influences (44). In other words, assuming cause-effect responses, one uses this analysis to test assumptions of *direct* and *indirect* effects of variables. Conceptually, a path analysis is presented as a path diagram with the assumed causal order represented by the direction of arrows. As an example, consider the regression model of Scott and Hollins (equation 6) (38). Their equation does not show explicitly the direct and indirect causes of crop loss. One possible relationship among these three variables is presented in Fig. 2. In keeping with path analysis practices, equation 6 is presented in Fig. 2 with all variables expressed as *Z*s. Eyespot may directly affect loss, or the disease may affect lodging which, in turn, may affect loss. Lodging from other causes would also affect loss. Standardized coefficients (parameters) from ordinary, least-squares regression are used to estimate the path coefficients (*r*s), which show the degree of direct and indirect effects. In Fig. 2, the *e*s are called latent variables and are analogous to the *u*s of regression. Scott and Hollins (38), unfortunately, did not give all the regression information necessary to calculate the path coefficients. This is not intended as criticism of their work; the example was chosen only to suggest a useful statistical technique.

Plant pathogens probably cause numerous direct and indirect effects on losses of many crops. Path analysis is a formal

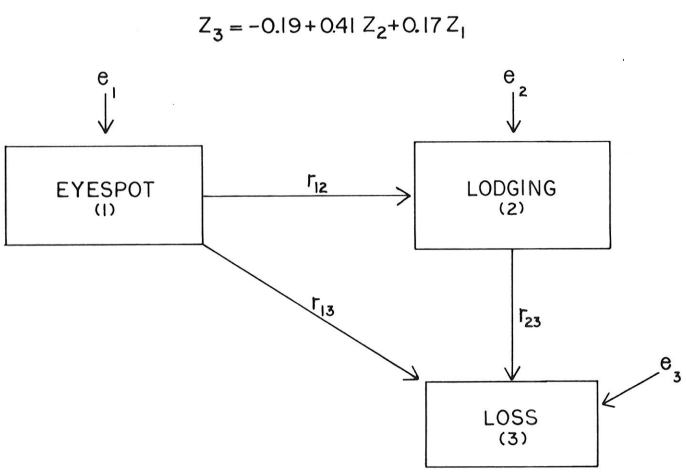

Fig. 2. Path diagram depicting the effect of eyespot (*Pseudocercosporella herpotrichoides*) and lodging on loss of wheat. Derived from Scott and Hollins (38).

mechanism for testing hypotheses about such effects. For instance, Hampton (8) used path analysis to assess the yield components of beans that were infected by bean yellow mosaic virus and bean common mosaic virus. He hypothesized that these viruses reduce the number of pods per plant, which, in turn, affects the number of seeds per pod and the seed weight, and also that the number of seeds per pod influences the seed weight. He was able to use path analysis to test these hypotheses and compare infected with healthy beans.

Multivariate statistics. For analysis of systems with many correlated variables (predictors or dependent variables) multivariate statistics are often useful. Although multivariate statistics, by definition, are concerned with multiple dimensions (9,33), there is not always a clear-cut separation of these techniques from univariate ones. Multiple regression and correlation analysis often are grouped with univariate techniques even though they deal with multiple variables. Multivariate statistics is a vast and complicated field of study and no adequate summary can be presented here. Only potential uses of these techniques for crop loss research will be discussed.

Principal components and factor analysis are two methods for reducing the dimensionality of a data set by finding linear combinations of the original variables (27,33). Since most variability can be explained with fewer, *independent* components (or factors) than the original data, effective dimension reduction is accomplished. The new components are independent of each other because of mathematical requirements, and therefore interpretations often are improved (9,27,33). Formal statistical tests seldom are performed with principal components or factor analysis.

Although principal component analysis is a useful technique in its own right, it is often used as an initial step to another form of analysis. For instance, estimated regression parameters are very imprecise when the predictors are highly correlated. A solution to this problem, other than ridge regression, is to produce "new" independent variables consisting of the principal components which account for most of the variability among the predictors. The multiple regression model can be written as:

$$Y_i = \Sigma_j^P (B_j C_{ij}) + u_i \quad (10)$$

in which all terms are defined as before and C_{ij} is the j-th (out of p) principal component for the i-th sampling unit (31). Interpretation of the parameter B_j is based on which variables are heavily loaded in the j-th principal component. Wiese (43) used this approach to relate yield of peas to several crop and environmental variables. The technique can be performed by several statistical computing packages, including BMDP (5).

Canonical correlation is a multivariate statistical technique for determining the relationship between two sets of variables. In a sense, it is a generalization of regression and correlation analysis. A function(s) of one set of variables (eg, *Y*s) is related to a function(s) of another set of variables (eg, *Z*s). For example, one set of variables could be yield components of wheat (tillers per unit, kernels per head, weight per kernel) and the other set could be levels of leaf rust at different times and/or environmental variables. Canonical correlation determines the association between the two sets of variables by defining separate linear combinations of the two sets, and then associating the linear combinations with correlation coefficients (33). Stynes (40) used this technique to relate root pathogens of wheat to soil properties.

If the *Z*s are replaced by indicator codes for treatment levels, canonical correlation reduces to multivariate analysis of variance (MANOVA) (9). MANOVA can also be looked upon as a generalization of ANOVA in which univariate variables are replaced by sets (vectors) of variables. This allows an analysis of yield components, for example, at levels of inoculation, chemicals, etc. Most plant pathologists analyze each variable separately. However, yield components are correlated with each other (6), and this reduces the efficiency of ANOVA. Unfortunately, MANOVA does not provide convenient mechanisms for testing differences among the treatment means. MANOVA also may be difficult or impossible with some complicated experimental designs. I find that a satisfactory compromise is to first perform a MANOVA, where

appropriate, and if certain factors or their interactions are significant, then to use ANOVA and contrasts of the means for each variable separately.

Most crop loss researchers have been reluctant to use multivariate statistics because of the rigor of matrix notation and algebra that it involves. Plant pathologists need not become proficient in multivariate or other branches of statistics, but should develop good working relationships with applied statisticians who can help design and analyze experiments. These statisticians should be considered full-fledged members of a research team.

Model validation. Once a model is chosen to represent crop loss, the overriding question is: How valid is the model, or how well does it actually behave in use (34)? Testing a model on the data used to develop it (ie, verification) will almost certainly overestimate the model's performance, because the procedure used to develop the model makes greatest use of all idiosyncrasies of those particular data (34). On the average, the apparent precision (as possibly measured by R^2 or S) will be greater than the true precision of the model for the whole population of values.

Ideally, the validity of a model should always be assessed. Mosteller and Tukey (34) described two levels of validation: simple cross-validation, and double cross-validation. Simple cross-validation tests the model with data different from those used to estimate the parameters and choose the model form. Sometimes, if enough observations are available, half or a smaller subset of the data can be omitted from the initial analysis, and the resulting model can be used to predict the omitted values. An R^2 can then be calculated for the fit of the model to these omitted data. An expansion of this approach, often called the "jack-knife" procedure, is to set aside one observation, estimate the model parameters and test the prediction of the omitted point, return this omitted point to the data set, remove the next value, and continue as before. An R^2 can be calculated based on the predictions of the missing values. Note that every prediction is for an observation not included in the analysis that serves as a basis for the prediction. Crop loss data sets are seldom large enough for elimination of a substantial fraction of the points for validation purposes. Thus, "jack-knifing" is an excellent approach for validating crop loss models, especially with today's high-speed computers.

A higher level of validation, double cross-validation, involves testing a model with data collected separately from those used in developing the model (34). Ideally, these data should be collected after the model parameters were estimated. If this is not possible, the validation data should be "hidden" so that they have no influence on the model builder (34). Crop loss researchers seldom perform double-cross validation. When validation studies are attempted, parameter estimates and the functional form of the model(s) often vary (6,29).

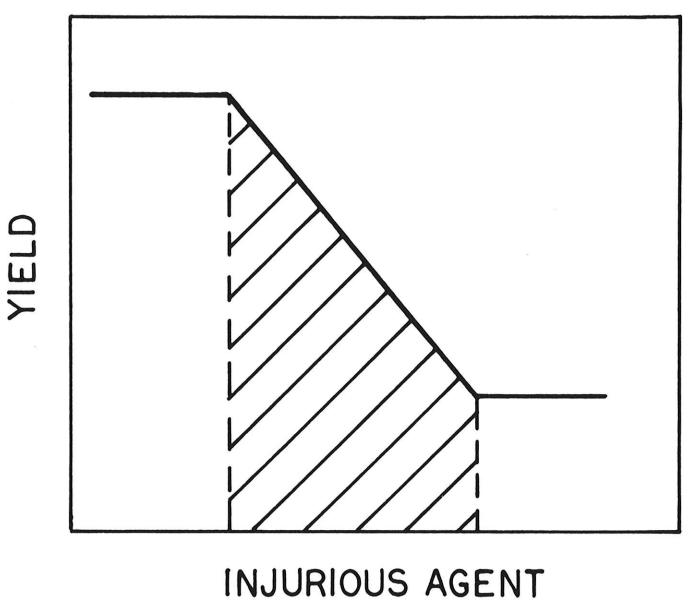

Fig. 3. Theoretical relationship between yield and level of an injurious agent, such as a plant disease. Shaded region is the part of the curve that is typically analyzed and modeled. Redrawn from Tammes (41); the scale is arbitrary.

Theoretical considerations. Very few of the models for crop loss due to diseases at the field level consider the physiology of yield reduction (1), probably because most crop loss researchers are not strongly oriented toward physiology. Physiological models can be very complex and require more inputs than are practical in field situations. A complete understanding of the process of crop loss will not be possible, however, without considering the physiological response of crops to disease (3).

Tammes (41) suggested a theoretical relationship between yield and the levels of an injurious agent, such as a plant disease. Fig. 3 is redrawn from his original article. The relationship is loosely based on physiological considerations. Properties of this relationship under a given cropping regime include an upper limit of yield (= attainable yield), a lower limit of yield for which an increase in disease does not cause a further reduction in yield, a threshold level of disease below which there is no measurable change in yield, and the middle part of the curve where there is a significant correlation between yield and disease. Although some of the models account for these theoretical limits (eg, equations 8 and 9), most models deal only with the middle (shaded) portion of Fig. 3 and assume no limits to upper and lower yield. Incorporation of the theoretical considerations of Tammes (41) into methods of disease loss estimation entails conducting thorough field studies to obtain several levels of disease intensity, and also increased modeling research. Nevertheless, these considerations are necessary for a fuller understanding of crop loss.

DISCUSSION AND CONCLUSIONS

Plant pathologists are only beginning to understand the complex relationship between plant diseases and crop losses. During the past decade, crop loss research has progressed from the obscure concern of a few scientists to a major speciality within plant pathology. The American Phytopathological Society now has a disease loss committee and PHYTOPATHOLOGY has a section on Disease Detection and Losses. Horsfall and Cowling (15) and James and Teng (23) have given excellent historical accounts of the determination of losses due to plant diseases in field plots. James (16–22) has made especially important contributions to the science of crop loss assessment. In an elegant and pioneering series of papers in the 1970s, James and co-workers (16–22) showed that crop loss assessment and modeling are, indeed, of scientific merit.

Methods for measuring and quantifying disease intensity in single plants and fields (23,24,42) and for determining and predicting crop losses are improving. Nevertheless, there still are few cases in which loss or yield can be described with the precision and repeatability necessary to make predictions on a large scale basis. Furthermore, few models have been developed for losses caused by more than one disease, or by diseases in combination with other pests. Until reliable predictive capabilities are achieved and multiple diseases are considered, plant pathologists will be unable to make much of a contribution to the estimation of yields on a worldwide basis (12).

LITERATURE CITED

1. Allison, J. C. S. 1970. On the physiology, estimation and forecasting of crop losses due to pests and diseases. Rhodesia Agric. J. 74:15-18.
2. Apple, J. L. 1977. The theory of disease management. Pages 79-101 in: Plant Disease, An Advanced Treatise. Vol. I. J. G. Horsfall and E. B. Cowling, eds. Academic Press, New York. 465 pp.
3. Boote, K. J. 1983. Coupling pests to crop growth simulators to predict yield reductions. Phytopathology 73:1581-1587.
4. Burleigh, J. R., Roelfs, A. P., and Eversmeyer, M. G. 1972. Estimating damage to wheat caused by *Puccinia recondita tritici*. Phytopathology 62:944-946.
5. Dixon, W. J., and Brown, M. B., editors. 1979. Biomedical computer programs, P-Series. University of California Press, Berkeley. 880 pp.
6. Fried, P. M., MacKenzie, D. R., and Nelson, R. R. 1981. Yield loss

caused by *Erysiphe graminis* f. sp. *tritici* on single culms of 'Chancellor' wheat and four multilines. Z. Pflanzenkrankh. Pflanzenschutz 88:256-264.
7. Gregory, L. V., Ayers, J. E., and Nelson, R. R. 1978. Predicting yield losses in corn from southern corn leaf blight. Phytopathology 68:517-521.
8. Hampton, R. O. 1975. The nature of bean yield reduction by bean yellow and bean common mosaic viruses. Phytopathology 65:1342-1346.
9. Harris, R. J. 1975. A Primer of Multivariate Statistics. Academic Press, New York. 332 pp.
10. Helwig, J. T., and Council, K. A., editors. 1979. SAS User's Guide. Statistical Analysis Systems (SAS) Institute, Inc., Raleigh, NC. 494 pp.
11. Henderson, H. V., and Velleman, P. F. 1981. Building multiple regression models interactively. Biometrics 37:391-411.
12. Hickman, J. R. 1983. Global crop condition assessment. Phytopathology 73:1597-1600.
13. Hicks, C. R. 1973. Fundamental Concepts in the Design of Experiments. Holt, Rinehart, and Winston, New York. 349 pp.
14. Hills, F. J., Chiarappa, L., and Geng, S. 1980. Powdery mildew of sugar beet: Disease and crop loss assessment. Phytopathology 70:680-682.
15. Horsfall, J. G., and Cowling, E. B. 1978. Pathometry: The measurement of plant disease. Pages 119-136 in: Plant Disease, An Advanced Treatise. Vol. II. J. G. Horsfall and E. B. Cowling, eds. Academic Press, New York. 436 pp.
16. James, W. C. 1974. Assessment of plant diseases and losses. Annu. Rev. Phytopathol. 12:27-48.
17. James, W. C., Callbeck, L. C., Hodgson, W. A., and Shih, C. S. 1971. Evaluation of a method used to estimate loss in yield of potatoes caused by late blight. Phytopathology 61:1471-1476.
18. James, W. C., Jenkins, J. E. E., and Jemmett, J. L. 1968. The relationship between leaf blotch caused by *Rhynchosporium secalis* and losses in grain yield of spring barley. Ann. Appl. Biol. 62:273-288.
19. James, W. C., and Shih, C. S. 1973. Size and shape of plots for estimating yield losses from cereal foliage diseases. Exp. Agric. 9:63-71.
20. James, W. C., Shih, C. S., Callbeck, L. C., and Hodgson, W. A. 1973. Interplot interference in field experiments with late blight of potato (*Phytophthora infestans*). Phytopathology 63:1269-1275.
21. James, W. C., Shih, C. S., Hodgson, W. A., and Callbeck, L. C. 1972. The quantitative relationship between late blight of potato and loss in tuber yield. Phytopathology 62:92-96.
22. James, W. C., Shih, C. S., Hodgson, W. A., and Callbeck, L. C. 1973. A method for estimating the sucrose in marketable tubers caused by potato late blight. Am. Potato J. 50:19-23.
23. James, W. C., and Teng, P. S. 1979. The quantification of production constraints associated with plant diseases. Pages 201-267 in: Applied Biology. Vol. IV. T. H. Coaker, ed. Academic Press, New York.
24. Lindow, S. E. 1983. Estimating disease severity of single plants. Phytopathology 73:1576-1581.
25. MacKenzie, D. R., and King, E. 1980. Developing realistic crop loss models for plant diseases. Pages 85-89 in: Assessment of Losses which Constrain Production and Crop Improvement in Agriculture and Forestry. P. S. Teng and S. V. Krupa, eds. University of Minnesota, Minneapolis. 327 pp.
26. Madden, L. V. 1980. Quantification of disease progression. Prot. Ecol. 2:159-176.
27. Madden, L. V., and Pennypacker, S. P. 1979. Principal component analysis of tomato early blight epidemics. Phytopathol. Z. 95:364-369.
28. Madden, L. V., Pennypacker, S. P., Antle, C. E., and Kingsolver, C. H. 1981. A loss model for crops. Phytopathology 71:685-689.
29. Madden, L. V., Pennypacker, S. P., and Kingsolver, C. H. 1981. A comparison of crop loss models. Phytopathol. Z. 101:196-201.
30. Main, C. E. 1977. Crop destruction—The raison d'être of plant pathology. Pages 55-78 in: Plant Disease, An Advanced Treatise. Vol. I. J. G. Horsfall and E. B. Cowling, eds. Academic Press, New York. 465 pp.
31. Mandel, J. 1982. Use of the singular value decomposition in regression analysis. Am. Stat. 36:15-24.
32. Marquardt, D. W., and Snee, R. D. 1975. Ridge regression in practice. Am. Stat. 29:3-20.
33. Morrison, D. F. 1976. Multivariate Statistical Methods. McGraw-Hill, New York. 415 pp.
34. Mosteller, F., and Tukey, J. W. 1977. Data Analysis and Regression. Addison-Wesley Publishing Co., Reading, MA. 588 pp.
35. Neter, J., and Wasserman, W. 1974. Applied Linear Statistical Models. Richard D. Irwin, Inc., Homewood, IL. 842 pp.
36. Pimentel, D., Krummel, J., Gallahan, D., Hough, J., Merrill, A., Schreiner, I., Vittum, P., Koziol, F., Back, E., Yen, D., and Fiance, S. 1978. Benefits and costs of pesticides used in U.S. food production. BioScience 28:772-783.
37. Ryan, T. A., Joiner, B. L., and Ryan, B. F. 1980. Minitab Reference Manual. The Pennsylvania State University, University Park. 138 pp.
38. Scott, P. R., and Hollins, T. W. 1978. Prediction of yield loss due to eyespot in winter wheat. Plant Pathol. 27:125-131.
39. Seinhorst, J. W. 1965. The relation between nematode density and damage to plants. Nematologica 11:137-154.
40. Stynes, B. A. 1980. Synoptic methodologies for crop loss assessment. Pages 166-175 in: Assessment of Losses which Constrain Production and Crop Improvement in Agriculture and Forestry. P. S. Teng and S. V. Krupa, eds. University of Minnesota, Minneapolis. 327 pp.
41. Tammes, P. M. L. 1961. Studies of yield losses. II. Injury as a limiting factor of yield. Neth. J. Plant Pathol. 67:257-263.
42. Teng, P. S. 1983. Estimating and interpreting disease intensity and loss in commercial fields. Phytopathology 73:1587-1590.
43. Wiese, M. V. 1980. Comprehensive and systematic assessment of crop yield determinants. Pages 262-269 in: Assessment of Losses which Constrain Production and Crop Improvement in Agriculture and Forestry. P. S. Teng and S. V. Krupa, eds. University of Minnesota, Minneapolis. 327 pp.
44. Wright, S. 1960. Path coefficients and path regressions: Alternative or complementary concepts? Biometrics 16:189-202.
45. Zadoks, J. C., and Schein, R. D. 1979. Epidemiology and Plant Disease Management. Oxford University Press, New York. 427 pp.
46. Zitter, T. A., and Simons, J. N. 1980. Management of viruses by alteration of vector efficiency and by cultural practices. Annu. Rev. Phytopathol. 18:289-310.