Data Management for Long-term Field Trials

D P Blevins and S R Colbert
North Carolina State Forest Nutrition Cooperative
Department of Forestry
North Carolina State University
Raleigh, North Carolina
USA

Abstract

Results from long-term field trials are a necessary basis for decision-making in forest management. Because of the complexity and quantity of data collected from these trials, careful attention must be given to potential errors and their effect on the analysis and interpretation of results. The accuracy and precision of data from long-term field studies can be greatly improved by implementation of a data management programme designed to reduce the errors associated with all aspects of trial establishment and measurement. The four components of such a data management system are 1) documentation of all procedures involved with study establishment and measurement, 2) elimination of errors during data collection, 3) identification and correction (where possible) of bias and mistakes, and 4) creation of a practical and well-documented data storage format.

Key words: *data management, documentation, data collection, error-checking, data storage*

Introduction

Measurement of long-term field trials, including response to silvicultural treatments, provides a necessary basis for decision-making in forest management. The reliability of management decisions based on field trial results depends on the accuracy and precision of the data collected. Long-term trials are often established with no consideration for procedures designed to minimise errors associated with data collection and processing.

Several characteristics of long-term studies make data management particularly important. First, they represent a tremendous investment in time and money. Second, a large quantity of diverse and complex data is collected. Third, measurements are collected over an extended period by various individuals, often using different instruments. Finally, turnover of personnel is often high.

Data Management

A sound data management programme is designed to reduce the errors associated with all aspects of field trial measurement, including data collection, entry, analysis and storage. The principal objective is to avoid loss of measurement data and creation of ambiguous, useless datasets. Data management systems have four general components: documentation, data collection, error-checking and data storage.

DOCUMENTATION

Documentation of all procedures involved with study establishment, maintenance and measurement is the first step in any data management system. Complete documentation is an essential part of avoiding data loss or ambiguity. Examples of activities requiring detailed description include study establishment (eg location, plot size, past management, numbering system, assumptions), installation of treatments (data, equipment, personnel, weather conditions), initial and subsequent measurement (eg units, instruments, personnel, dates), and disturbances to the original plot layout or deviations from remeasurement schedule. Carefully planned and documented photographs or slides are especially useful for recording field procedures.

DATA COLLECTION

Many of the errors associated with data collection can be avoided by proper study design, site selection, plot layout, and treatment installation (eg Curtis 1983, Synnott 1979). Special consideration should be given to identifying measurement variables adequate to address hypotheses, as well as document stand conditions critical to interpretation of results. Errors of concern during data collection include inaccurate instrument reading or use, incorrect instrument calibration, measuring the wrong individual, and erroneous plot identification codes. Well-designed, prepared tally sheets are instrumental in reducing data collection errors. Hand-held, electronic data recorders can also be used to avoid errors that occur during transcription from tally sheets and to expedite data processing. Effective use of data recorders requires development of programs designed for specific applications by field technicians, rather than by computer programmers unfamiliar with fieldwork.

Tally sheets or data recorder programs should contain as much information as possible (eg study identification, plot codes, tree numbers) to reduce the amount of data recorded in the field. Where available, previous measurements (eg diameter, height) should be included to allow comparison and correction of errors in the field. Damage and mortality codes should include all identified possibilities, as well as allow flexibility for unanticipated problems. Space should also be provided for comments pertaining to problems encountered during measurement.

ERROR-CHECKING

Following collection, field data should be checked immediately for errors so that variables can be remeasured if necessary. Errors may be classified as accidental, bias, mistakes or sampling (Schreuder *et al.* 1993). Accidental errors are unavoidable, resulting from inconsistent environmental conditions, rounding off, and limitations or deficiencies of instruments or methods, and are assumed to be compensating. Bias, or systematic error, results from selection or favouring (possibly unintentionally) of one outcome over another. Mistakes are due to human carelessness, casualness or fallibility. Sampling error is a random component resulting from uncontrolled variables and the intrinsically uncertain nature of measurement which cannot normally be assessed.

For purposes of data management, error-checking is generally concerned with locating and eliminating errors due to bias and mistakes. Proper training, instrument maintenance, and quality control during data collection procedures can do much to reduce these types of error. Bias and mistakes can be located using statistical software to check all observations in a dataset. Specific items to identify include errors in the frequency of plot codes due to mistakes in transcription of plot identification, illogical combinations of values for certain variables, negative growth over a period of time, and extreme or out-of-range values. Out-of-range values can be identified by plotting individual values or relationships between certain variables or by using statistical tests for outliers.

Once identified, the cause of a specific error must be investigated before a correction is considered. The error should be checked against the original tally sheet, wherever possible, to see whether it was indeed measured, rather than only transcribed incorrectly. Incorporation of initial error-checking procedures in electronic data recorder programs allows identification and verification of suspect measurements in the field. Measurement bias can often be corrected due to its systematic nature. Mistakes, in contrast, are more difficult to eliminate since they are more often random in character. Changes to data should be made only if the source of error can be verified adequately.

DATA STORAGE FORMAT

A standard data storage format must be developed at the beginning of any long-term study. The data storage format should allow access by many types of software to expedite data exchange and be flexible to accommodate additional information if required (Sweet and Byrne 1990). For example, almost all software programs can correctly input ASC11 text, which stores no formatting characters, and add characters as needed. Any documentation important to data analysis and interpretation should be located within the final dataset, preferably at the top. Examples of this type of information include file name, study title, data title, measurement data, total number of observations, last edit date, data format and variable definitions, value definitions, and calculations. It is usually possible for a software package to read the data and ignore the documentation. In addition, the data storage format itself should be well documented, as well as easy to use and interpret.

Conclusion

Development of long-term field trials requires careful consideration of data management procedures designed to minimise errors associated with data collection, processing, analysis and interpretation. The tremendous investment represented by long-term field trials necessitates establishment of a data management programme to ensure precision and accuracy of measurements and applicability of results.

References

Curtis, R O. 1983. Procedures for establishing and maintaining permanent plots for silvicultural and yield research. United States Department of Agriculture, General Technical Report. PNW-155.

Schreuder, H T, Gregoire, T G and Wood, G B. 1993. *Sampling Methods for Multiresource Forest Inventory.* John Wiley & Sons, New York.

Sweet, M D and Byrne, J C. 1990. A standardized data structure for describing and exchanging data from remeasured growth and yield plots. United States Department of Agriculture Forest Service, General Technical Report. INT-271.

Synnott, T J. 1979. A manual of permanent plot procedures for tropical rain forests. *Tropical Forestry Papers* ***14***. Commonwealth Forest Institute, University of Oxford, Oxford.