Vanclay,J.K'97

Getting the most out of your permanent plot data

Jerome K. Vanclay
Center for International Forestry Research, Indonesia

Abstract

A catalogue of ideas for graphical analyses of growth data is presented, in the hope of stimulating the more imaginative analyses. Graphs can be particularly revealing, because the human eye is good at detecting patterns. Suggestions are given to make graphs more effective, and analyses more insightful.

Introduction

In this paper, I want to stimulate some more innovation in the analysis of permanent plot data. I was motivated to collate these suggestions recently after refereeing some manuscripts which I felt did not do justice to the data at hand. Analyses of permanent plot data should not merely follow accepted procedures copied from previously published work; rather it should involve careful consideration of the objectives of the study, and of the potential and limitations of the data at hand. In many cases, graphical approaches are the most revealing, and I urge greater use of such techniques, for preliminary screening of data, as a supplement to formal statistical tests, and as a convincing way to communicate findings. I will not comment on statistical procedures (these have been adequately dealt with elsewhere, e.g., Warren 1981, 1986), and I will confine my comments largely to graphical techniques for preliminary data exploration prior to more formal statistical analyses, and for illustrating findings in publications.

Revisiting objectives

Perhaps the first step in any analysis is to clarify the objectives. This applies to the analysis of any of any data, whether from temporary or permanent plots, experimental or passive monitoring plots (Vanclay 1994*a*). It applies to the objectives of the analysis in question, and to the objectives of the plot system.

Ideally, the objectives of the plot installation will be consistent with those of the planned analysis. This need not always be the case, and useful conclusions may be drawn from plots established for purposes disparate from the analyses in question, but careful consideration of possible outcomes is warranted if the objectives are not congruent. The critical questions are:

- to what extent may results be influenced by the design of the plot system?
- will results merely be artefacts, or can the plot data cast light on the issue at hand?
- are the available data sufficient to detect the phenomenon of interest, if it exists?

These questions impinge on both the plot system, and on the analysis at hand. It is appropriate to revisit the objectives of the plot system during any analysis of plot data, particularly since many plot systems are established with rather vague objectives. Clearly, care must be taken to ensure that any enhancements proposed for the plot system do not confound observations (Vanclay 1994*a*), but analyses of plot data may reveal forest types and conditions that warrant further sampling (e.g., Beetson *et al.* 1992), or may reveal additional attributes that should be measured on existing plots.

The analysis at hand should also reflect the objectives, not merely follow a standard proforma. This requires clear objectives, ideally, stated as testable hypotheses. I have addressed this issue before (Vanclay 1992, 1994*b*), as have many others (e.g., Lund *et al.* 1992, Rennolls and Gertner 1993), but it remains one of the greatest weaknesses in many analyses, so is worth revisiting.

Checking for errors

Deliberate analyses remain amongst the best ways to detect errors in a database, but a great deal of time may be lost in this way. It is preferable to commence an analysis with careful checks for errors; this avoids many pitfalls, including spurious results and wasted time. Doug Sheil (1995) recently presented an excellent review of procedures to detect errors and inconsistencies in long-term data from permanent plots in tropical moist forests. I make no attempt to summarize his work; I prefer that you consult the source.

Graphical analyses

Modern spreadsheets and other computer packages make it easy to graph data, and offer a many options to customize graphics. Unfortunately, these features are often abused, so that they detract from, rather than contribute towards a fuller understanding of the data. Edward Tufte (1983) devotes an entire book to this topic (a classic book, beautifully illustrated: I recommend it). However, his message is simple and well argued: maximize the information:ink ratio by focusing on the information and keeping the graphics simple. The object of a graphic is to stimulate the reader to think about the implication of the graphic (e.g., for forest management), not to wonder how the graphic was produced. Many of the special effects (e.g., 3-dimensional appearance, hatching, etc.) available in computer packages may produce the latter, rather than the former reaction. While the appearance of graphics is particularly important in the presentation of results, it is also an important consideration in the analysis, as too many "gimmicks" may conceal, rather than reveal information.

I am not proposing that graphical approaches should be used to the exclusion of other techniques. Rather, I propose that they offer an important supplement to standard statistical techniques, especially in the preliminary data exploration phase of analysis. One enduring advantage of graphical approaches is the ability to illustrate the shape of a relationship, and thus indicate suitable functional relationships and transformations for use in statistical models: this is why graphical inspection of raw data and residuals is a standard statistical technique (see e.g., Weisberg 1985, Vanclay 1994*a*). In addition to their role in supporting statistical analysis, graphs offer an efficient way to convey information to readers.

Examining a single set of data

What can be done to reveal the nature of a forest stand to the reader of a scientific paper? Summary
ables with lots of numbers may appear rigorous, but are often tedious (and boring) to interpret, while
ketches of the stand profile (see e.g., Oldeman 1990, for many examples) can be lively, but rather
ubjective. Fortunately, much quantitative data can be graphed to make it more accessible. With
magination, almost any aspect of a forest stand can be illustrated in a graphical way, but I shall
onfine my attention to two aspects: the species composition, and the stand structure.

pecies richness, or biodiversity, is currently topical, but is not always meaningful (Hurlbert 1971,
'anclay 1996). Many indices have been proposed, but nearly all have some limitations (MacGurran
988). Species counts appeal in their simplicity, but can be misleading as they reveal little about the
mpling effort of local species distribution (Mawdsley 1996). Thus the presentation of information
n the biodiversity of a site is not straight forward, and careful thought on the matter is warranted.

e charts are often used to illustrate the dominance of the most abundant species, commonly
hibiting stem numbers, basal areas, or "importance" (the mean of relative number and relative
sal area), but the implication is not always clear, as the outcome may depend on sample size or
acement. One way to offer supporting information regarding the sampling effort is to accompany
e charts with graphs showing cumulative species numbers plotted against sampling effort (e.g., area
number of trees sampled). Experiment by taking sub-plots or tree numbers in different orders, and
e how this influences the shape of the relationships. Try scaling the axes or the data, to see if a
are-root or logarithmic transformation suggests a straight-line relationship: this may support your

contention that no asymptote has been reached, or may suggest mathematical relationships that m later be fitted to the data.

Stand structure is often quantified as a stand table showing the number of stems in each of seve metric size classes (i.e., fixed-interval classes, e.g., 10-20, 20-30, ... cm dbh). The adequacy of suc summary depends much on the size of the sample and the number of classes. In many cases, it m be more appropriate to use deciles (i.e., 10% of total tree numbers in each class), rather than met classes, but this may make it difficult to interpret graphs, and negates the utility of concepts such de Liocourt's q (i.e., the ratio of numbers in successive classes; e.g., Philip 1994). Korsgaard (19 argued that it is more informative to graph stand basal area (instead of stem numbers) within e size class. He observed that natural dipterocarp forests in Malaysia tend to maintain approximat equal basal areas in each class, and felt that the harvesting history of a stand could be inferred fr the distribution of basal areas within size classes.

An efficient alternative may be to plot the cumulative numbers or basal areas of trees (commenc from the largest) against the tree size. This is analogous to the traditional stand table and Korsgaar table of basal areas respectively, but is independent of class sizes and less dependent on sample s so may have greater utility.

Another attribute often used to describe forests is the dominant height, often defined as the m height of a specified number (e.g. 50/ha) of the fattest trees. However, the choice of the fattest rat than the tallest, and the arbitrary selection of a predefined number of trees (e.g. 50/ha), may influe results. A more informative alternative is to plot the running mean tree height against tree ra where trees may be ranked by height or by diameter. This alternative may offer substantially m information, especially in mixed forests, as for example, stands with emergents will exhibit mark different trends than trees with a more uniform canopy.

Finally, when examining commercial aspects such as timber volumes, think carefully about quality of the various components contributing to the estimates. Some questions that shoul considered include:

1. Do commercial species differ in form to such an extent that species-specific volume equat should be used, or can a general equation be used for all species?
2. Has sufficient account been taken of the various factors that may lead to a reduction in vol including but not limited to inaccessible areas, buffer zones, logging damage, stem defects, et
3. What volume is being predicted: phytomass, total stem volume, sawlog volume, veneer tim etc?

It is inevitable that some approximations are needed in the analysis of tropical forest data, and requires no apologies, but does demand clear descriptions of the assumptions made and the presented.

Comparing data from different treatments or places

When more than one set of data is involved, the analysis may differ from the previous case number of ways, depending on the situation. Three situations are of interest, namely:

1. *replications*, where conditions are assumed comparable, and the variability of responses interest;
2. *treatments*, where conditions have been altered in a known way, and differences in responses are of interest;
3. *monitoring*, where differences in responses are observed, and inferences about ch conditions are of interest.

In the case of field forestry experiments, these three situations may represent an unattainable ideal, but they serve to illustrate the need to explore similarities and differences in the basic conditions underlying the data.

Replications should be identical in as many aspects as possible, and any factors that could vary should be investigated to see what contribution they make to the variance between replications. With treatments, a limited number of factors should be varied in a controlled way, while all other factors remain the same (as with replications). Of particular concern are factors not under experimental control that may vary with treatments, and thus *confound* the results (e.g., if insect defoliators are more prevalent in fertilized plots, no growth response may be visible, because it is obscured by the effect of defoliation). With monitoring systems, we need to know how all the conditions change, so that we can identify possible causes, and can be aware of possible confounding factors.

These caveats apply to all data comparisons, whether they relate to different treatments, different places or different years. The analyst has the responsibility to clarify what differences and what similarities exist, so that an objective assessment of the probable causes and possible confounding factors can be made. This may done using graphical and regression analyses, but it may also be useful to illustrate the distribution of the data in the data space defined by the two most influential factors. This is an analogue of the issue of supplementary sampling (e.g., Beetson *et al.* 1992), and the same exploratory techniques may be used. An alternative is to calculate the principal components (excluding the response variable), and to examine the distribution of data within the data space created by the first two principal components (PCs). If the first PC captures most of the variation (relative to the second PC), there is a real danger of confounding, and further investigation is warranted.

Clearly, it is important to understand the data, and to know how the various data subsets differ. One way to gain such an understanding is to graph all the data on a plot-by-plot basis (discussed above), and then to make graphs on pair-wise or group-wise basis to see how plots differ. Pair-wise graphs may be most useful in cases where stand-level data takes the form of distributions (e.g., stand tables), and the data from two or three plots may be included on the same graph using different symbols to indicate the origin of the data. Group-wise graphs may be most useful in cases where stand-level data can be summarized into a single number (e.g., site productivity, stand basal area, etc.), and may include data from many plots, especially when graphed against any of the factors that differ greatly among plots (e.g., graph site productivity against rainfall and elevation; stand basal area against time since last harvest, etc.). In both cases, visual impressions can be confirmed with standard statistical tests, such as F-tests on the residual variance about regressions (see any statistical text).

Comparing time series data

Time series data are analogous, since time, rather than place, has changed, but the also offer some particular challenges, since "everything is connected to everything else". Take for instance, tree growth: changes in tree growth rates over time may be attributed to increased age, increased tree size, increased competition, to a combination of these factors, or to other factors. Amongst the other factors is the important question: how can one be sure that the changes observed are due to environmental change, not to procedural changes?

In even-aged single-species plantations, the stand-level changes in basal area and tree numbers are of some interest, but in the mixed tropical forest, individual tree characteristics may be easier to interpret. Three components of growth and change are of interest: diameter increment, mortality, and recruitment. Mortality and recruitment are difficult to deal with, since data are rarely of sufficient number or quality to provide good graphs of mean rates by stand density and by tree size (in the case of recruitment). However, it is worth experimenting with what data are available, and with graphs of the predicted values from statistical models fitted to the data.

There are several ways to appraise diameter increments. One useful way to gain an overall idea of growth relationships is to graph mean diameter increment within each of several species and size classes versus tree size (dbh) and stand basal area. These may be followed by graphs of individual tree increments versus tree size and stand basal area to gain more detailed insights.

Special considerations when several factors vary

As the number of variables increase, more care needs to be taken, as it becomes more and more difficult to understand possible interactions. One way to screen for possible confounding is to compile a scatterplot matrix (cf. correlation matrix). Another way is to examine the principal components of the regressor variables: if the data are orthogonal, all components will explain an equivalent share of the variance, but with non-orthogonal data, the amount of variance explained by a PC may diminish quickly with its rank.

Synthesis

My attempt to stimulate ideas for more imaginative analyses of data comprises three simple components

1. experiment with alternatives, especially visual ones, because the eye is good at detecting patterns;
2. keep it simple, so that the noise (and embellishment) does not detract from the signal;
3. supplement the graphics with statistical tests to confirm or reject what you see.

References

Beetson, T., Nester, M. and Vanclay, J.K., 1992. Enhancing a permanent sample plot system in natural forests. *The Statistician* 41:525-538.

Hurlbert, S.H., 1971. The nonconcept of species diversity: a critique and alternative parameters. *Ecology* 52:577-586.

Korsgaard, S., 1992. An analysis of growth parameters and timber yield prediction, based on research plots in the permanent forest estate of Sarawak, Malaysia. Council for Development Research, Denmark, 120 pp.

Lund, H.G., Päivinen, R. and Thammincha, S.(eds) 1992. Remote Sensing and Permanent Plot Techniques for World Forest Monitoring, Proceedings of IUFRO S4.02.05 Wacharakitti International Workshop, 13-17 January 1992, Pattaya, Thailand, 271 pp.

Magurran, A.E., 1988. *Ecological Diversity and its Measurement.* Princeton University Press, NJ, 179 pp.

Mawdsley, N., 1996. The theory and practice of estimating regional species richness from local samples. *In*: D.S. Edwards et al. (eds) *Tropical Rainforest Research - Current Issues*. Kluwer, pp. 193-213.

Oldeman, R.A.A., 1990. *Forests: Elements of Silvology*. Springer, Berlin, 624 pp.

Philip, M.S., 1992. *Measuring Trees and Forests.* CAB International, Wallingford UK, 310 pp.

Rennolls, K. and Gertner, G., 1993. The optimal design of forest experiments and forest surveys, Proceedings of IUFRO S4.11 Conference, 10-14 September 1991,University of Greenwich, UK, 333 pp.

Sheil, D., 1995. A critique of permanent plot methods and analysis with examples from Budongo Forest, Uganda. *Forest Ecology and Management* 77:11-34.

Tufte, E.R., 1983. *The Visual Display of Quantitative Information*. Graphics Press, Cheshire CT, 197 pp.

Vanclay, J.K., 1992. Permanent plots for multiple objectives: defining goals and resolving conflicts. *In*: H.G. Lund, R. Päivinen and S. Thammincha (eds) Remote Sensing and Permanent Plot Techniques for World Forest Monitoring. Proceedings of IUFRO S4.02.05 Wacharakitti International Workshop, 13-17 January 1992, Pattaya, Thailand, pp. 157-163.

Vanclay, J.K., 1994*a*. *Modelling Forest Growth and Yield: Applications to Mixed Tropical Forests.* CAB International, Wallingford, UK, 312 pp.

Vanclay, J.K., 1994*b*. Resource Inventory for Land-Use Planning. *In*: Seminar on Land-Use Planning and Land Tenure to Secure the Permanent Forest Estate. International Tropical Timber Organization, Report SRS-11, pp. 7-16.

Vanclay, J.K., 1996. Towards more rigorous Assessment of Biodiversity. Monte Verita conference, in press.

Warren, W.G., 1981. Basic statistical methods in forestry research: use, misuse and prognosis. In Proc. XVII IUFRO Congress, Japan, Div. 6, pp. 108-10.

Warren, W.G., 1986. On the presentation of statistical analysis: reason or ritual. *Can. J. For. Res.* 16:1185-1191.

Weisberg, S., 1985. *Applied Linear Regression*, 2nd ed. Wiley, NY, 324 pp.