

## **LOOSENING THE CLAMP: AN EXPLORATORY GRAPHICAL APPROACH TO THE CLIMATE LEAF ANALYSIS MULTIVARIATE PROGRAM**

**W.A. Green**

### **ABSTRACT**

The Climate Leaf Analysis Multivariate Program (CLAMP) is an established methodology for physiognomic analysis of dicot leaf floras. This paper uses a meta-analysis of four studies that provide CLAMP data on 245 floras from Asia, Africa, and North and South America to demonstrate the application of a new analytical methodology for the exploration of the relationship between leaf morphology and environment. This methodology involves the application of a generalized "pairs" plot or scatter plot matrix (SPLOM), a form of graphical analysis for multivariate data. It is compared with the results from regression, hierarchical cluster analysis, principle components analysis, and canonical correspondence analysis. Analysis of the available data using pairs plots reveals extensive multiple covariation among explanatory leaf physiognomic variables and identifies sources of systematic error that eigenvector ordination methods tend to conceal. Pairs plots provide a supplementary method for analyzing complex multivariate data on leaf physiognomy and contribute to biological understanding of leaf–environment interactions. Because pairs plots allow more flexible investigation of multivariate data than existing eigenvector and regression-based approaches, they may prove useful not only for analyzing CLAMP data, but also in exploring multiple covariation in other complex paleontological data sets.

W.A. Green. Department of Geology and Geophysics, Yale University, P.O. Box 208109 Yale Station, New Haven, Connecticut 06520 U.S.A. [walton.green@yale.edu](mailto:walton.green@yale.edu)

**KEY WORDS:** leaf physiognomy, Climate Leaf Analysis Multivariate Program (CLAMP), exploratory graphical analysis, pairs plots, Scatter Plot Matrix (SPLOM)

---

### **INTRODUCTION**

As paleontological data become more quantitative, multivariate, complex, and voluminous, the choice of tools for data analysis acquires a greater

influence over the biological and geological conclusions that are drawn from a given body of data. Either the data must be processed, summarized, its dimensionality reduced, and its details

PE Article Number: 9.2.9A

Copyright: Paleontological Society September 2006

Submission: 10 June 2006 Acceptance: 4 August 2006

Green, W.A., 2006. Loosening the CLAMP: An Exploratory Graphical Approach to the Climate Leaf Analysis Multivariate Program. *Palaeontologia Electronica* Vol. 9, Issue 2; 9A:17p, 1.63MB; [http://palaeo-electronica.org/paleo/2006\\_2/clamp/index.html](http://palaeo-electronica.org/paleo/2006_2/clamp/index.html)

obscured; or we need new tools to handle the presentation and publication of larger, more complex data sets. In this paper, one such new tool—the "pairs" plot—is suggested as a way of improving the standard procedures for examining the relationship between leaf architecture and environment.

The Climate Leaf Analysis Multivariate Program (CLAMP) is a method of analyzing fossil leaf assemblages or floras (specifically those deriving from the woody dicotyledonous component of forest ecosystems) by quantifying a set of significant morphological or architectural leaf variables and relating these variables (averaged over the flora) to climate parameters. Such a procedure also allows estimation of ancient climate parameters by uniformitarian extrapolation of patterns found in the distribution of leaf attributes in modern vegetation (Wolfe 1993, 1995, Wolfe and Spicer 1999).

This general notion, which is sometimes referred to as "leaf physiognomy", has been accepted since the early twentieth century when Bailey and Sinnott (1915) pointed out the strong correlation between the temperature in which modern forests grow and the proportion of the species that compose them that have "entire" (i.e., untoothed) leaves. From this observed correlation in the modern world, a determination of the percentage of species with entire margins in a fossil flora allows the estimation of the temperature in which it grew. With the introduction of computers that could handle algorithmic classification and ordination of multivariate data, Wolfe (1993) proposed a multivariate method of coding leaves (originally based on 29, but later updated to 31 variables) that was intended both to improve the precision of temperature estimates over the univariate linear regressions that had preceded it and to allow the estimation of other climatic variables.

In addition to temperature, other variables that have been more-or-less successfully estimated using CLAMP data are precipitation (e.g., Wilf et al. 1998), and moist enthalpy, which can be used to calculate paleo-elevation (e.g., Forest et al. 1999). The linear relationships between leaf size and precipitation and between leaf physiognomy and moist entropy are less clear than the relationship between leaf margins and temperature. As Kennedy (1998) points out: "It appears that CLAMP provides a relatively accurate estimation of temperature, but only a general estimation of precipitation." Other variables (including many related to timing of changes in temperature and precipitation, like growing season precipitation, or warm

month mean temperature) have been studied less extensively.

Despite the relatively widespread application of CLAMP methods, its procedures have been criticized as overly complex and no more informative than simple regression models (Wilf 1997, Wilf et al. 1998). Nevertheless it provides the only well-known procedure for collecting multivariate data on leaf morphology, and in certain contexts has become a standard way of determining ancient terrestrial climatic parameters. Therefore the focus is almost exclusively on the CLAMP method, though some of the issues identified may also apply to other recent leaf physiognomic studies like the "digital" approach of Huff et al. (2003) and Royer et al. (2005). Much of the debate about the advantages of the CLAMP method over various regression models centers around statistical details: the goal has been to maximize the "explanatory power" of the method and minimize the standard errors of the temperature estimates that it provides. This may not, however, be the best way to choose an analytical methodology, because we have no satisfactory mechanistic explanation of the relationships between most leaf morphological characters and climatic variables. Thus we are by definition engaged in data analysis: that is, we are trying to determine what measured quantities signify and to design empirical models to predict them, not trying to test models based on theory against real data. Minimizing analytic error and maximizing explained variance produce a model that best explains a given set of data. Whether this model will ever explain any other data, be of practical predicative utility, or suggest fruitful lines of future inquiry, is a very different question.

In the following consideration of the available data, several issues with the CLAMP method that ought to be addressed become apparent. The focus throughout is on the analytical choices made, not on the collection of raw data: for the purposes of this discussion, it is assumed that the matrix of CLAMP scores is a relatively good reflection of the woody dicot leaf forms present in a living flora. No coding scheme is perfect, but the CLAMP method is the only such coding method that has been widely applied. In contrast, the statistical methods for analyzing CLAMP data represent only a small fraction of the available procedures for multivariate data analysis. Therefore, it seems necessary to explore to what extent the results of a CLAMP analysis are sensitive to the analytical methods chosen and to inherent biases in the data. Are the eigenvector and regression techniques that are

generally applied to these data appropriate? What other techniques should be tried? In short: how can we improve the methods used to analyze CLAMP data? The alternative or supplementary analytical program proposed by this paper is based on graphical data analysis using pairs plots and seems to show substantial advantages over eigenvector approaches for exploring the relationships between dicot leaves and the environments in which they grow. Though the focus of this examination is specifically aimed at one type of paleobotanical data, the general issue of whether exploratory graphical analysis is more appropriate than data-reduction is applicable to many other paleontological data sets.

## MATERIALS AND METHODS

The data available for this analysis come from four studies. The first and largest, WOLFE173<sup>1</sup>, is an updated version of the data published by Wolfe (1993), sometimes referred to as "CLAMP 3B" (e.g., Jacobs 2002). It is available on the web as two Excel spread sheets (.xls files) containing respectively the morphological and environmental data for 173 floras, points representing which are colored black in the figures in this paper. For the 103 floras that were published in 1993, the number of species in each flora and its latitude, longitude, and elevation were typed in from Wolfe (1993). The geographical distribution of this data set is mainly restricted to the continental United States and Japan, though there are a few floras from Alaska and continental east Asia. These data are available in the Appendix, and include i) the climatic variables and ii) the morphological leaf scores.

The second data set, JACOBS, is from Jacobs (1999, 2002) who gives CLAMP scores and associated environmental data for 30 floras in tropical Africa. This study used the original 29-variable coding scheme, so there are two variables with all values missing. Points from JACOBS are colored red in the figures in this paper. The third data set, GREGORY, is from Gregory-Wodzicki (2000), provides CLAMP scores and environmental data for 12 floras in Bolivia and is colored green here. Finally, Kowalski (2002) provides CLAMP scores and environmental data for 30 floras in tropical South America, which are represented by blue points here.

The data in JACOBS, GREGORY, and KOWALSKI were scanned in from tables in the cited publications, processed for automatic text-recognition, and then hand-edited and spot checked for accu-

racy. The data were read into the open-source program R (R Development Core Team 2004) from tab-delimited text files, which are available in the supplementary data archive and preprocessed so that all studies were in comparable form. The code used is given in the script file in the data archive. The data matrices are not printed because all the data have appeared in print before.

The completed data set consists of 245 floras and is stored as a series of data frames in R with the suffix -ALL for the raw frames containing both morphological and environmental data, and a suffix -CLAMP for the cleaned CLAMP scores. The complete data set is ALL, and the supplementary material typed in from Wolfe (1993) is a separate data frame called wolfe1993. Stranks (1996) provides additional data from New Zealand that have not yet been processed.

The 31 physiognomic variables described in Wolfe (1993, 1995) are listed in Table 1. Unfortunately, unless growing season precipitation is taken to be the same as annual precipitation, the only environmental variable that appears in all four data sets is mean annual temperature (MAT), so our comparison of different studies is restricted to a single response variable. This is unfortunate because the main point of applying a multivariate framework is to elicit information about multiple response variables. Because little of the true uncertainty in a temperature estimate comes from analytic error in the explanatory variables (this contention is defended below), it is highly unlikely that a multivariate framework will really improve temperature prediction much, however much of it can be made to reduce residual error of the regression.

## RESULTS

Exploring the relationship between a single pair of variables is simple and intuitive. For instance, as is well-known, a plot of the percentage of untoothed leaves (P) against MAT shows a strong linear relationship. Figure 1 shows this relationship by plotting all available CLAMP data along with associated least-squares regression lines. The regression line for all the data is shown as a dotted line and limiting the regression to the floras for which information is available on the total number of species coded does not change the line perceptibly. The points are colored to show which study they came from and the thicker, colored lines show the results that are obtained when separate regressions are performed for each study. These regression lines are cropped to the extreme ranges

1. Additional data files are available from the PE website at this URL: [http://palaeo-electronica.org/2006\\_2/clamp/index.html](http://palaeo-electronica.org/2006_2/clamp/index.html)

**Table 1.** Variables in 31-dimensional Matrices.

	<b>VARIABLE NAME FROM Wolfe (1993)</b>	<b>ABBREV. USED HERE</b>	<b>COMMENTS</b>
1	Lobed	Lobd	
2	No.Teeth	Entr	
3	Regular.teeth	TReg	
4	Close.teeth	TClS	
5	Round.teeth	TRnd	
6	Acute.teeth	TAcu	
7	Compound.teeth	TCmp	
8	Nanophyll	ZNan	Missing in JACOBS
9	Leptophyll.1	ZLe1	
10	Leptophyll.2	ZLe2	
11	Microphyll.1	ZMi1	
12	Microphyll.2	ZMi2	
13	Microphyll.3	ZMi3	
14	Mesophyll.1	ZMe1	
15	Mesophyll.2	ZMe2	
16	Mesophyll.3	ZMe3	Missing in JACOBS
17	Emarginate.apex	AEmg	
18	Round.apex	ARnd	
19	Acute.apex	AAcu	
20	Attenuate.apex	AAtn	
21	Cordate.base	BCor	
22	Round.base	BRnd	
23	Acute.base	BACu	
24	L.W.< 1:1	Rl1	
25	L.W.1-2:1	Rb12	
26	L.W.2-3:1	Rb23	
27	L.W.3-4:1	Rb34	
28	L.W.> 4:1	Rgt4	
29	Obovate	SObo	
30	Elliptic	SElp	
31	Ovate	SOvt	

of each data set, which are shown by colored bars near the edges of the plot.

Above and to the right of the main bivariate plot are histograms showing the marginal distributions of each variable. Note that both of these distributions are polymodal, probably because of

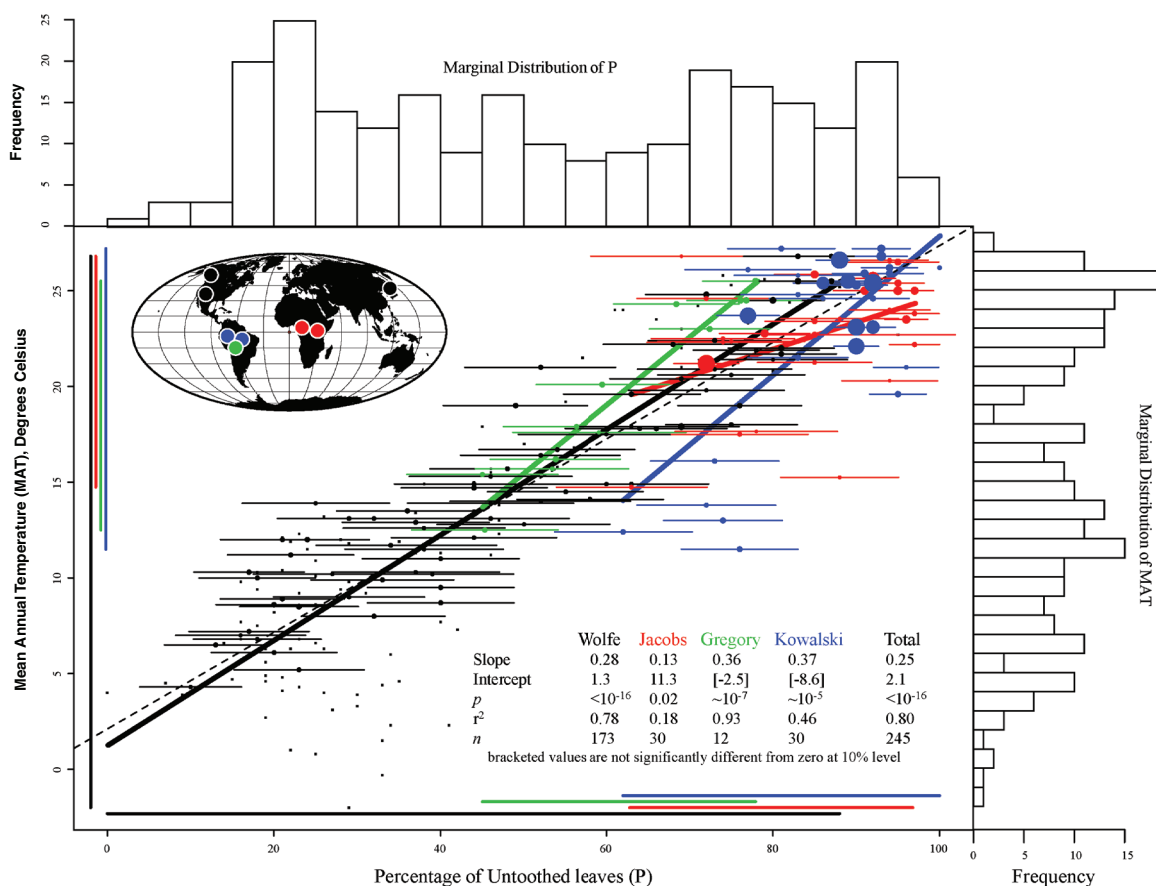
irregular geographical sampling: there are relatively few floras representing the intermediate temperatures because the latitudes that would supply them (the horse latitudes) are kept dry by Hadley circulation and therefore have not provided as fertile a source for "appropriate" floras to study.

It is evident from this univariate exploration that the "study effect" (the effect on the regression line of which study the data is drawn from) is important, though it cannot be determined from this representation whether it is due to poor repeatability of the coding or whether it is caused by spatial autocorrelation. In this regard, note how the slope of the regression line in the two South American studies (GREGORY and KOWALSKI) is very similar, though the intercepts differ, while JACOBS's African data have a slope that is quite different from that found in the other three (predominantly New World) studies. This phenomenon of slope having a greater spatial autocorrelation than intercept has also recently been pointed out by Mosbrugger et al. (2005). This incomparability of models based on training sets from different regions has also been frequently remarked upon (Stranks 1996, Jacobs 2002, Spicer et al. 2004, Greenwood et al. 2004), but with equal frequency has been ignored when citing binomial sampling errors or standard deviations as if they were true uncertainties.

To expand our consideration from one explanatory and one response variable to 31 explanatory and one response is not trivial. Perhaps the simplest solution is the reduction of all 31 explanatory variables to a single distance metric. Clustering the available 245 floras hierarchically shows imperfect clustering by study (the "blocks" of color) in Figure 2, which shows dendrograms produced by an agglomerative hierarchical algorithm using the Euclidean distance metric under two clustering procedures (single-linkage and complete linkage), with different properties. (Single linkage clustering finds "spherical" clusters of objects in  $n$ -space; complete linkage finds strings of closely-connected objects.)

With a few exceptions (e.g., Traiser 2004) such clustering procedures have not been used extensively in leaf physiognomy, perhaps because they produce no explicit models or quantitative estimates of independent variables, but merely give a visual display of similarities among floras. From such a display, we can nevertheless qualitatively conclude that the studies do cluster together, but not without noise.

Much more prevalent—perhaps even ubiquitous among explicit considerations of CLAMP



**Figure 1.** Comparative bivariate plot of temperature (MAT) against percentage untoothed leaves (P) for 245 floras and four studies, color-coded by study. Least-squares regression lines, marginal distributions, regression statistics, and the approximate geographical locations of the study areas are also shown. Points for which data are available on how many species are represented have error bars showing plus and minus the binomial sampling error. The size of the point plotted is also proportional to the sample size, except in the case where data are missing. The same colors representing the four studies are used in all subsequent figures.

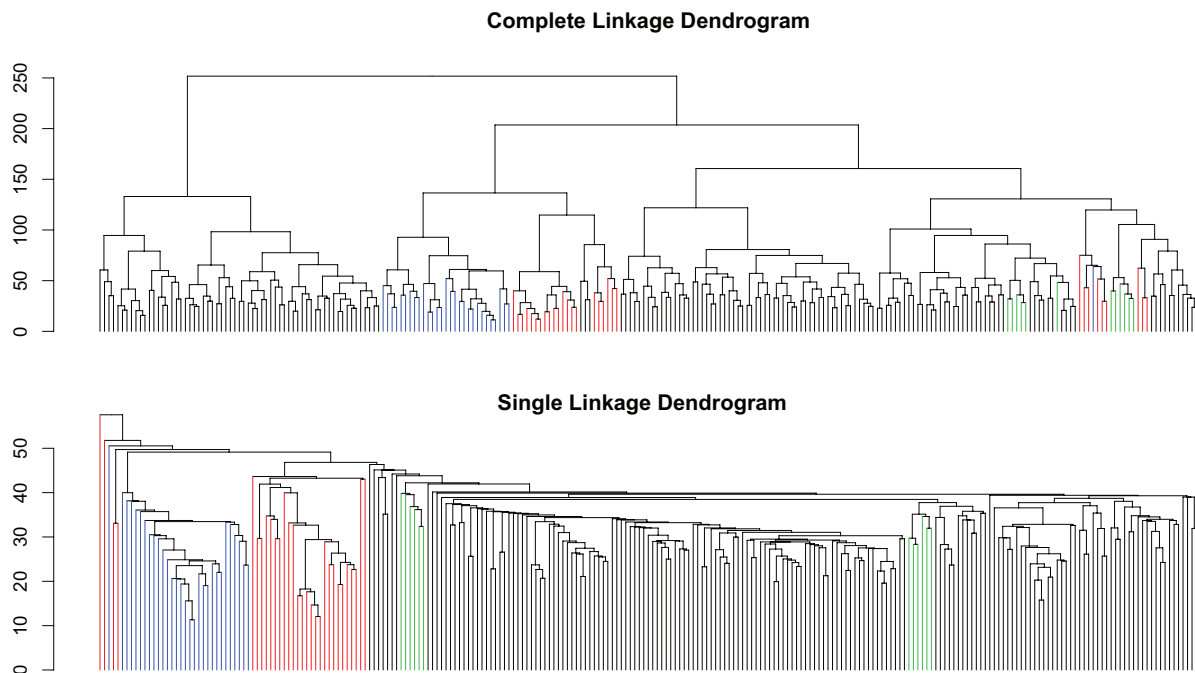
data—are eigenvector techniques for rotating multivariate vector spaces and re-projecting data along a few major axes of variation. The simplest and most general of these is principle components analysis (PCA). Originally Wolfe (1993) relied on correspondence analysis, and then (in 1995) switched to canonical correspondence analysis, or CCA (Ter Braak 1986). Both methods were specifically designed for comparison of environmental data with species distributions and have become fashionable in community ecology. Ter Braak (1986) makes it very clear, however, that:

The rationale of the technique [CCA] is derived from a species packing model wherein species are assumed to have Gaussian (bell-shaped) response surfaces with

respect to compound environmental gradients. (Ter Braak 1986, p. 1168)

and that

The vital assumption is that the response surfaces of the species are unimodal, the Gaussian (bell-shaped) response model being the example for which the methods performance is particularly good. For the simpler case where species-environment relationships are monotone, the results can still be expected to be adequate in a qualitative sense....The method would not work if a large number of species were distributed in a more complex way, e.g., bimodally. Ter Braak 1986, p. 1177



**Figure 2.** Complete linkage (top) and single linkage (bottom) agglomerative dendrograms colored by study showing (imperfect) clustering by study and area for 245 floras clustered according to the 31 CLAMP variables.

There are *a priori* reasons to expect species to have unimodal or linear distributions along environmental gradients, but this logic does not necessarily hold for morphological variables like the proportion of leaves with attenuate apices. As can be seen from the pairs plots presented below, some of these relationships between morphological variables and temperature are arched or parabolic. Therefore the theoretical applicability of CCA to morphological variables averaged over floras on a continental or global scale is highly questionable. Like many other statistical methods, CCA is also vulnerable to non-linearity and multiple colinearity:

When the data are collected over a sufficient habitat range for species to show nonlinear, nonmonotonic relationships with environmental variables, it is inappropriate to summarize these relationships by correlation coefficients or to analysis the data by techniques that are based on correlation coefficients, such as canonical correlation analysis.

Ter Braak 1986, p. 1167

and

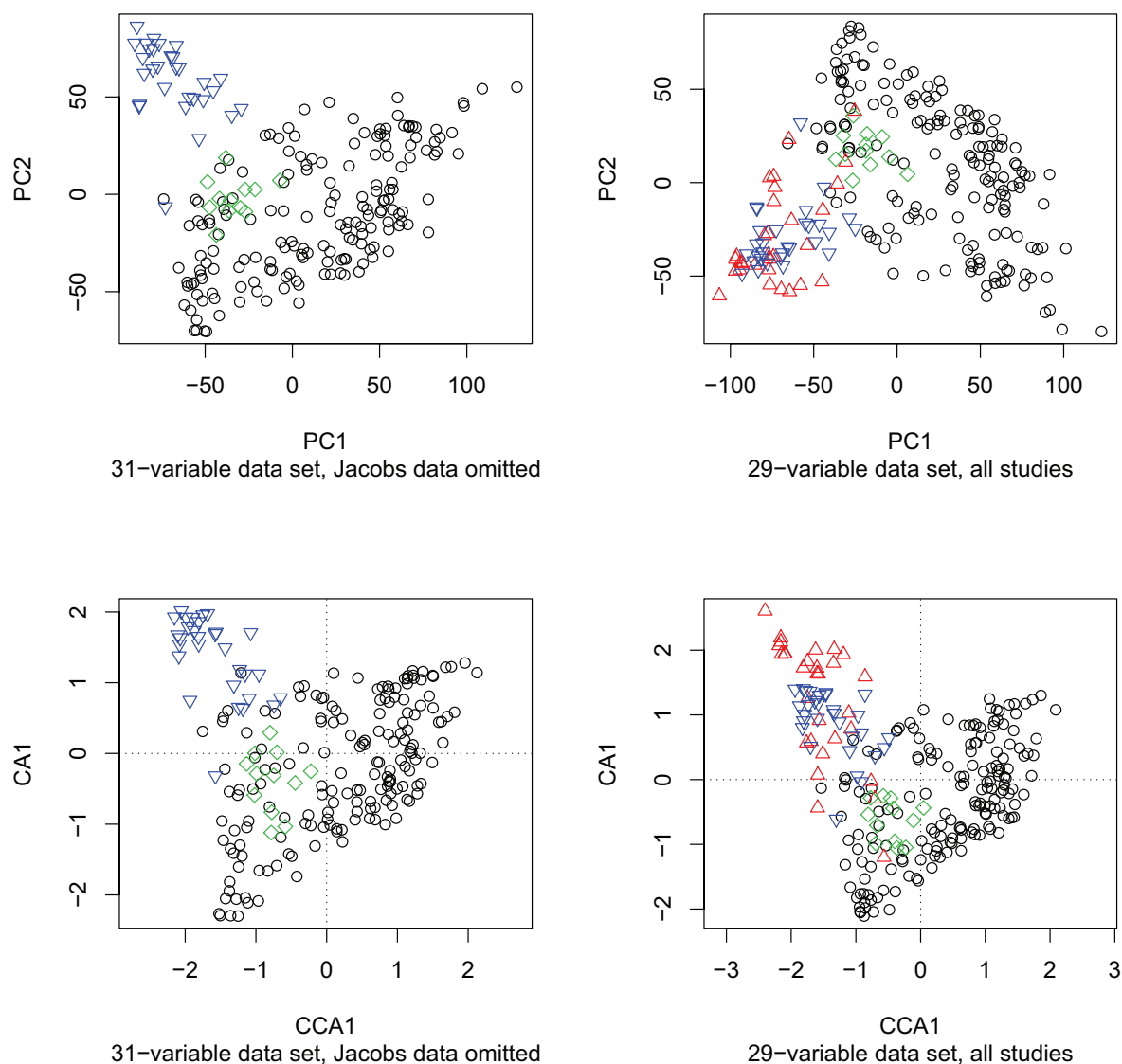
When the environmental variables are strongly correlated with each other—for

example, simply because the number of environmental variables approaches the number of sites—the effects of different environmental variables on community composition cannot be separated out and, consequently, the canonical coefficients are unstable. This is the multicollinearity problem.

Ter Braak 1986, p. 1170f.

Many statistical procedures—including simple linear regression—work in practice even when their assumptions are unrealistic, so this alone would not invalidate the application of CCA to CLAMP data. The argument made here is not that CCA produces incorrect results, but merely that the ubiquitous application of it to CLAMP data may be evidence of excessive analytical rigidity.

Most publications explicitly using CLAMP have followed Wolfe's lead even to the point of using Excel spreadsheet macros in the files that can be downloaded from the CLAMP website and a commercial program called CANOCO (Lep and milauer 2003) specifically designed to perform CCA. In fact CCA is now available in many general-purpose statistical packages, including three different implementations for R, and therefore continued reliance on compiled, proprietary software



**Figure 3.** Principle components analysis biplots (top) and canonical correspondence analysis biplots (bottom) of the data. The left pair of biplots uses all 31 variables present in three of the studies; the right pair reduces the number of variables to 29 and shows all four studies. Note the basic similarity between the PCA and CCA plots: the scatters of points are viewed from different angles, but the relationships between studies are similar.

seems additional evidence of methodological canalization. As Figure 3 illustrates, moreover, the analysis of CLAMP data is not a case in which PCA and CCA produce significantly different results: the top two bivariate plots are principle components; the bottom two are canonical correspondence axes.

Note how in the upper left quadrant of Figure 3, the bivariate plot of the first two principle compo-

nents clearly discriminates KOWALSKI from the other two studies. When JACOBS is added, however, (the upper right quadrant of Figure 3) the gap seems much less distinct. This illustrates how sensitive this form of analysis is to sampling. The more data that are added, the harder it is to discriminate clusters that looked distinct when there were fewer points. Although only PCA and CCA have been tested, it is difficult to imagine a realistic situation in

which other related eigenvector methods would lead to radically different interpretations of CLAMP data, although they may—like PCA and CCA—differ in the exact values they produce.

Both in the case of hierarchical cluster analysis and eigenvector analysis, it is apparent that the study effect contributes some structure to the data but by no means determines them. Formally, we could also use the multivariate *t* test (Hotelling's  $T^2$ ) to check the pairwise null hypotheses of multivariate equivalence of means. In all 4-chosen-2 = 6 cases we are forced to reject the null with *p*-values less than  $10^{-6}$ . In simple terms, the studies could not possibly all be equivalent. Note that the Hotelling code (see data archive) can not deal with missing values, so the number of variables had to be reduced to 29 in the three cases out of the six pairwise comparisons where variables were missing. Whatever the statistical logic, the data are clearly affected by the source from which they were obtained, though it cannot be determined with the available information whether this is due to the studies being in different geographic regions or whether people actually code leaves differently.

A dendrogram reduces 31 variables to a single distance metric, eigenvector methods reduce 31 variables to a few principle components, of which two are shown in Figure 3. What about the remainder of the variables? One response is: the first two principle components account for a large proportion of the variance, so the other variables do not matter much. This seems to be a limited way of looking at the process of data analysis: if only the axes of maximum variance are of interest, then why collect multivariate data? Multivariate data are often collected to answer more than one question, and a variable that answers a particular question (like the presence of teeth answering questions about temperature) may say nothing about another question (regarding, for instance, plant growth form). To choose variables exclusively from mathematical criteria like variance maximization seems to abdicate the responsibility for interpreting results biologically.

One way of looking at more variables is called a scatter plot matrix (SPLOM; Basford and Tukey 1999) or, more simply, a pairs plot. Figure 4 shows all the pairwise relationships between the original two variables plotted in Figure 1 (P and MAT), the first two principle components (PC1 and PC2), the first canonical component (CA1, the primary axis corresponding to the matrix of sites), and the first constrained canonical component (CCA1, the pri-

mary axis corresponding to the environmental matrix).

As can be seen, a pairs plot allows the plotting of a very large number of multivariate data in a compact form. The question then arises: what is the value added by eigenvector methods of data reduction if it is possible to plot and examine the raw data themselves? In Figure 5, all the 31 explanatory variables and MAT are presented in this pairwise fashion, with additional details as described in the figure caption.

This is a very concentrated way of presenting data; it plots  $32 \times 245 = 7840$  two-digit numbers, the equivalent in characters of about twelve and a half manuscript pages of text. Each of the small plots above the matrix diagonal is a similar bivariate plot showing the relationship between two of the 32 variables. Thus the scatter plot in the 32nd column and 2nd row of the pairs plot is a reduced version of Figure 1; it is simply the bivariate plot of P against MAT. The second and 32nd of the diagonal cells also correspond to the marginal histograms in Figure 1. The shadings below the diagonal are obtained by performing four two-sided hypothesis tests for each cell:

$H_0$ : slope of the least squares regression line = 0

$H_0$ : Pearson's product-moment correlation coefficient = 0

$H_0$ : Spearman's rank order correlation coefficient = 0

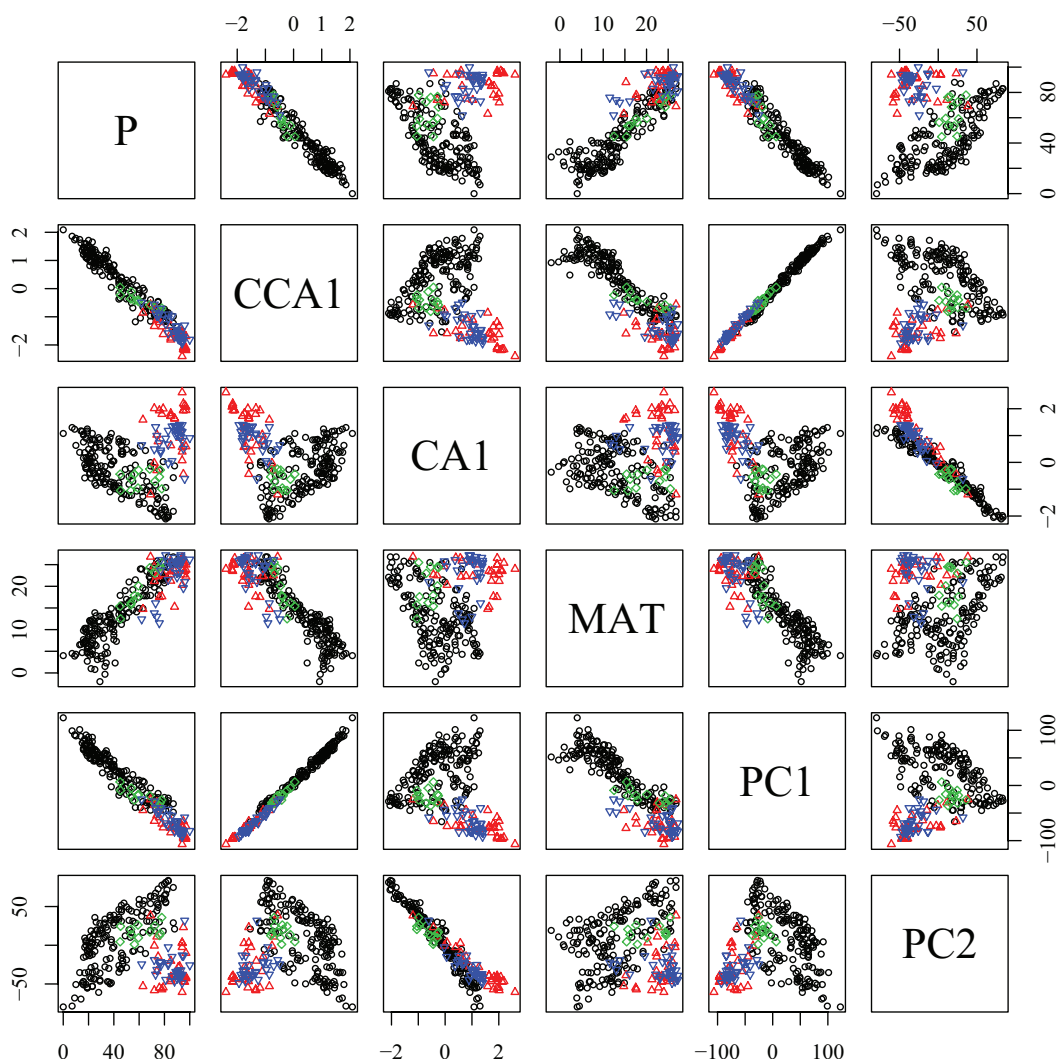
$H_0$ : Kendall's rank order correlation coefficient = 0

In each case the alternative hypothesis is the equivalent inequality.

The cell is colored white if the mean of the three correlation coefficients is positive and if all tests reject; black if the mean of the three correlation coefficients is negative and all tests reject; and medium grey if all tests fail to reject. If some but not all of the tests reject, the cell is colored light grey or dark grey depending on the sign of the mean of the correlations of the tests that are significant. In all cases the color of the text is black if the mean of the three correlation coefficients is positive and white otherwise. All tests are made at the level  $\alpha = 0.05 / \text{number of comparisons}$ , where the number of comparisons is  $(\text{number of variables}) * (\text{number of variables} - 1) / 2$ , i.e., the 5% level with Bonferroni correction for multiple comparisons.

This representation of the data allows us to examine complex multiple-covariation among the explanatory variables in detail. For instance, compare the second column with the third-through-sev-





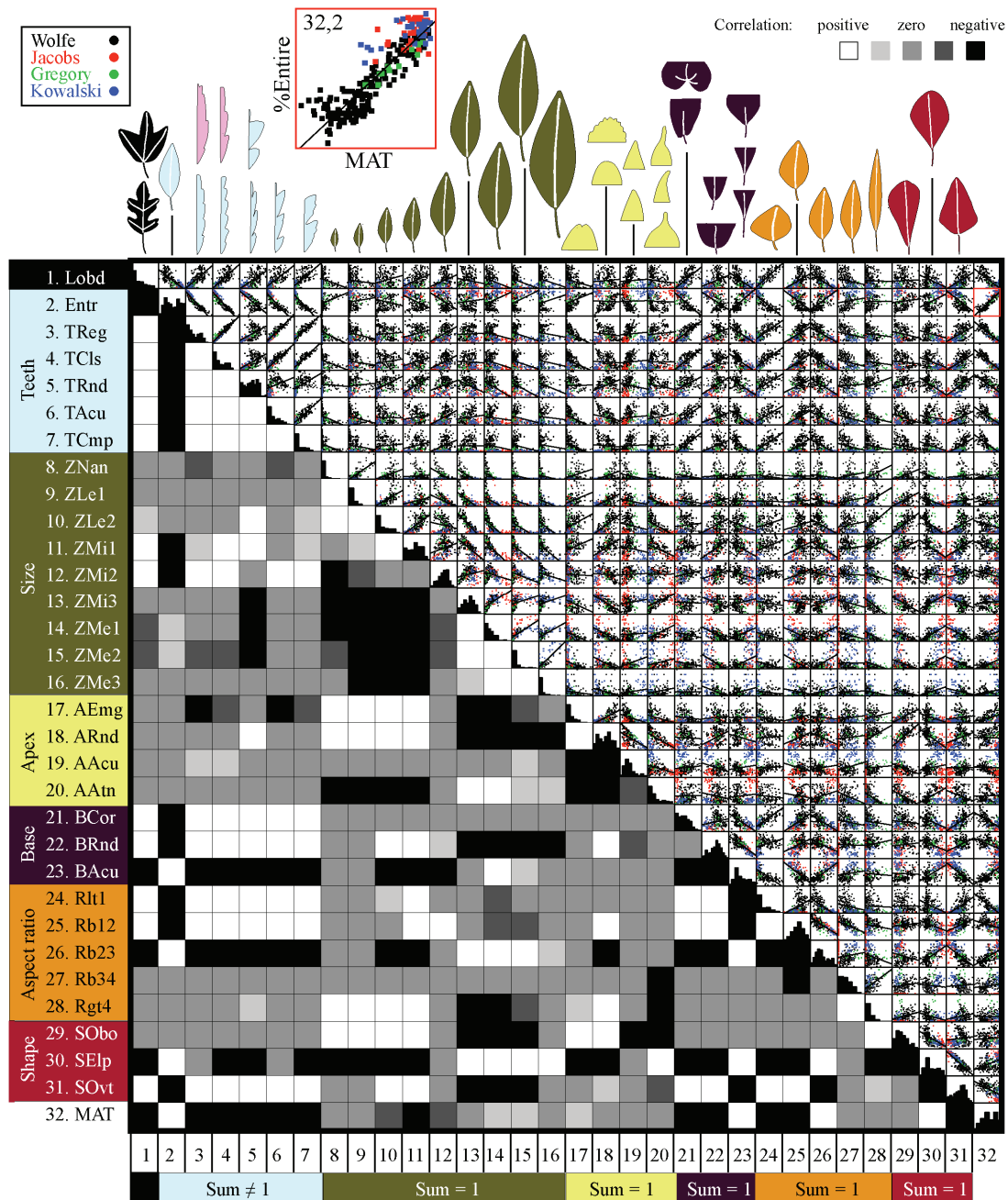
**Figure 4.** Traditional pairs plot as used in Basford and Tukey (1999). This is simply a matrix of scatter plots showing the relationships between each pair of a set of variables. All the data here are restricted to the original 29 CLAMP variables. In particular, note the similarity between PC1 and CCA1, demonstrating how little it matters which eigenvector method is chosen. Plots above the diagonal are inverses of the plots in the lower diagonal. Note that the plots in the third row, second column and sixth row, fifth column are exact duplicates of the right-hand two plots in Figure 3.

enth column block. They are inverses of each other, as they must be, because the second column represents the percentage of species lacking teeth and the third-through-seventh columns give the percentage of species with particular types of teeth. Another interesting block of covarying values is provided by the leaf sizes: columns and rows 8 through 18. Here, the smallest four leaf size categories are all strongly positively correlated with each other as are the largest four leaf size categories, while there is a strong negative correlation between the small and large blocks. Only the middle three size categories are not strongly collinear.

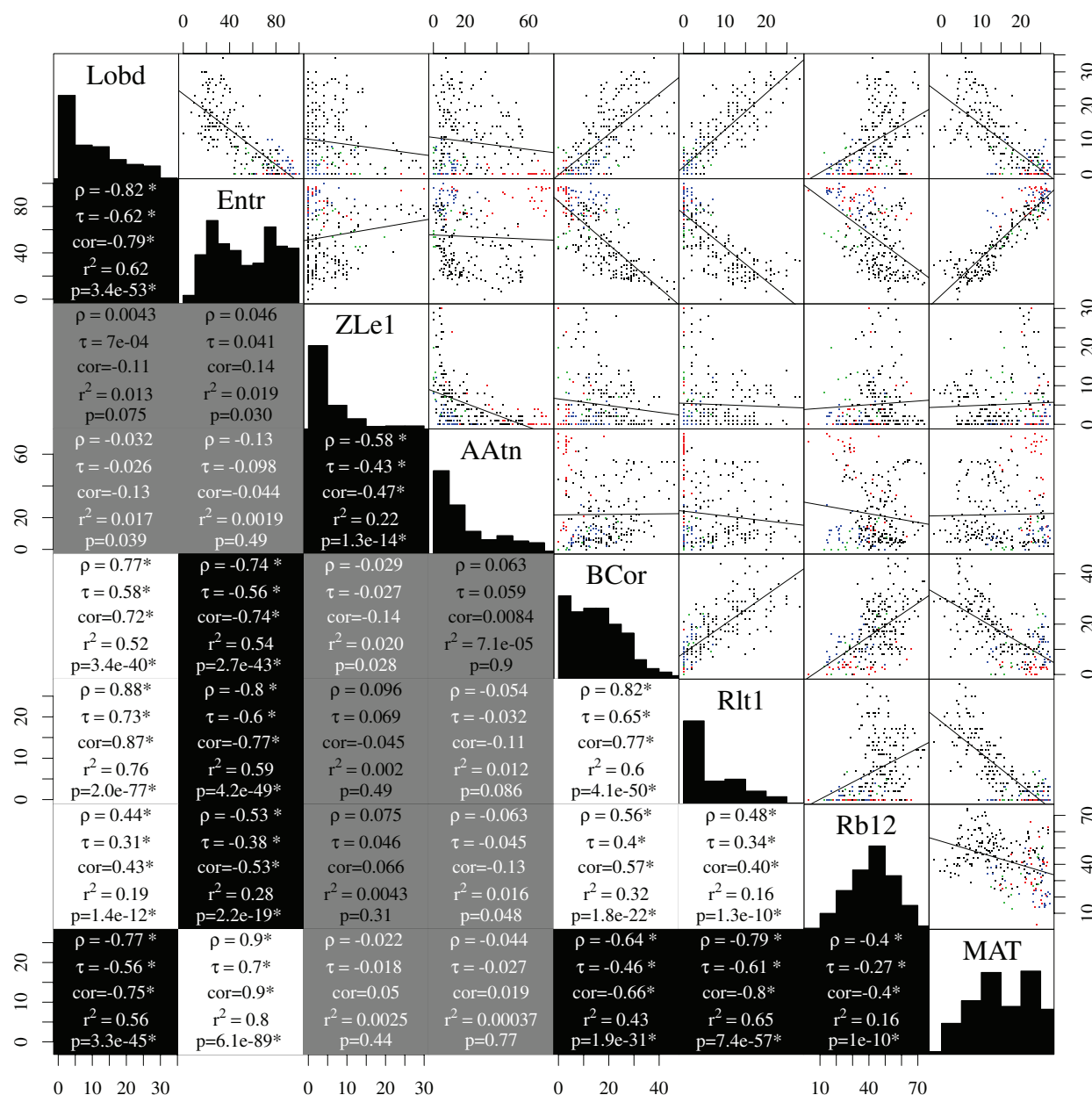
Graphical display of this sort of data makes the strong covariation among the variables apparent and indicates that any statistics calculated from them that assume independence should be treated with caution.

At such a small scale, it can be difficult to see details of the scatter plots, so Figure 6 shows another pairs' plot of a subset of the variables.

In this plot, in addition to the scatter plot matrix above the diagonal, the numbers in the blocks below the diagonal give all of the pairwise correlation coefficients (the two common non-parametric correlation coefficients, Spearman's  $\rho$  and



**Figure 5.** Elaborated pairs plot of 31 explanatory and one response variable. The variables are represented along the top of the matrix by the figures from Wolfe (1993) that were originally used to illustrate them, and along the left side by their abbreviated names. In two cases—"teeth regular" and "teeth close"—the line drawings colored red also show the opposite of the character being coded, "teeth irregular" and "teeth distant". Groups of related variables are shown by the colored bars along the top and left and those groups of related variables that are constrained to sum to 1 are so marked along the bottom of the matrix. Above the diagonal are scatter plots like the example shown in Figure 1; below the diagonal, the white, black, and grey squares show whether the relationships between the variables are statistically significant. The white squares show a strong positive relationship between the variables, the black a strong inverse relationship, and the intermediate shades of grey show weaker or absent relationships. Along the diagonal are plotted histograms of each of the 32 variables (the marginal distributions for the bivariate plots).



**Figure 6.** Pairs plot of selected variables.  $\rho$  is Spearman's rank order correlation coefficient;  $\tau$  is Kendall's rank order correlation coefficient;  $\text{cor}$  is the ordinary (Pearson's) product-moment correlation coefficient; and the  $r^2$  and  $p$ -values relate to the least-squares fitted line. Where the associated correlation tests (and in the case of the  $p$ -value, the regression) are significant at the 5% level, the statistics are followed by asterisks. The blocks are white with black text if the mean of the three correlations is positive and all four of the tests are significant, and they are light grey with black text if some, but not all of the tests are significant. The blocks are colored black and dark grey with white text in the reverse scenarios. If none of the tests are significant, the blocks are a medium grey and the sign of the mean of the three correlations is shown only by the color of the text printed on them: black for a positive and white for a negative correlation.

Kendall's  $\tau$  as well as the ordinary product-moment correlation coefficient) and the  $r^2$  and  $p$ -values for the least-squares fitted line.

This reduced set of variables could have been selected by a stepwise multiple regression procedure with formal rules for adding or subtracting variables from a model based on information criteria. This has been done and it is easy to produce models in which all terms are significant, with  $r^2$  as high as 0.8 and  $p < 10^{-16}$ . Since there are several such stepwise techniques, all of which produce about the same quality of model from different variables, there is no reason to choose one model over another, and the variables for this smaller plot were selected based on their perceived interest instead of on a formal stepwise procedure.

This particular choice of explanatory variables happens to produce a multiple linear least-squares model for MAT with of 0.86 and  $p < 10^{-16}$ ; adding a factor to the model showing which study each observation came from as an additional explanatory variable increases the  $r^2$  significantly to 0.88.

These models could still be improved further by continued analysis (for instance, interactions, non-linear terms, and variable transformations were not even tried). It does not seem valuable, however, to spend time massaging a model until the issues that the univariate case brought up (like incomparability of studies) are taken care of.

## DISCUSSION

The CLAMP method can be criticized at two levels: data collection and data analysis. This paper is concerned primarily with improving the methods of analysis, but there are a few problems with the analyses that are based in the data collection.

First is the fact that the raw (by species) scores have, except in doctoral dissertations (Stranks 1996, Kennedy 1998) seldom, or never, been published. This means that some of the most important and interesting questions about phylogenetic distribution of leaf morphological variables and the differences among plants of different growth form (habit) can not be asked. Some recent studies like that by Kennedy et al. (2002) have not even published the CLAMP scores averaged by flora, but have only printed biplots of eigenvector loadings. This form of presentation is so highly processed and incorporates so many assumptions that it makes interpretation of the results difficult and reanalysis of the data impossible.

Second, Wolfe's (1993) selection of characters was explicitly based on preliminary eigenvector analyses, for instance:

During one stage of the study, I expanded the character set to include about 20 character states additional to [the original 29]. Judging from eigenvalues and percent of total variance accounted for, these characters either added nothing or even lowered both eigenvalues and percent variance Wolfe 1993, p. 20.

To reject potentially interesting variables from a coding scheme on the basis of low eigenvalues and percent of variance accounted for is to allow the statistical horse to bolt: you proceed quickly, but have little control over the direction you are traveling. Among the character states rejected were those relating to compound leaves, spinose teeth, and inrolled or thickened margins, all of which have clear mechanical adaptive significance. If we hope to obtain ecological or environmental data from leaf physiognomy as well as information about climate, such character states should be retained.

The same criticism goes for lumping together characters like "teeth round" and "teeth appressed" merely because "combining the states produced both higher eigenvalues and percent variance. (Wolfe 1993, p. 24).

Thirdly, the description of some of the character states seems ambiguous. Though it is not possible without a comparative study to say for certain that interpretations of the character descriptions would vary, it is not clear whether "0.25 if the teeth are both regular and irregular and some leaves have teeth" (Wolfe 1993, p. 24) should be interpreted as "0.25 if the teeth are regular and/or irregular and some leaves have teeth" or as "0.25 if the teeth are all regular or all irregular and some leaves have teeth."

In the case of deeply lobed leaves, the leaf size is supposed to be scored from a single lobe, but the aspect ratio and shape still refer to the overall leaf, whereas in the case of a compound leaf the leaflet is what is scored for size, aspect ratio, and shape. This is particularly problematic in genera like *Rosa* in which a plant can have compound, deeply lobed, and simple leaves on the same branch.

Fourthly, the scored variables are divided into sections relating to common topics. Leaf size, for instance is coded as proportions of leaves falling into nine size classes. The scores in some of these sections, like leaf size, aspect ratio, or shape must

sum to one while the scores in the section describing teeth and lobation do not have a constant sum. This means that the presence of teeth is implicitly weighted more heavily in the overall description of the flora than, for instance, the leaf size, and it is not clear that any normalization procedure can correct this bias. The restriction of groups of variables to a constant sum introduces dependence and implicit weighting that are hidden by eigenvector analyses but made apparent by a graphical approach (see Figure 5).

Despite these issues with the process of coding, no morphological coding scheme could be ideal, and these criticisms of CLAMP are offered in a spirit of improving what is the only such system currently available in the published literature. In particular, Wolfe's original article (1993) was much more broadly focused than some subsequent publications: a discussion of axes of variation other than those corresponding to temperature and precipitation made it not only a contribution to paleoclimatology but also ecologically and botanically interesting.

More important than these problems in the coding are the true uncertainties associated with the estimation of paleoclimatic variables. It is generally accepted that the leaf physiognomy of a flora indicates the general climate regime of the area in which it grew: "tropical," "sub-tropical," or "temperate," and "wet" or "dry". These are categories that not only any botanist, but many laymen would recognize from simple leaf silhouettes. Beyond this there remains doubt as to the degree of precision and reliability that leaf physiognomy can provide, but there has been relatively little general discussion of what causes the real uncertainty in the procedure of estimating ancient environmental parameters from leaf morphology.

It is noticeable that two doctoral dissertations that have examined CLAMP data in detail are less sanguine about the errors associated with the methodology than most published articles. Stranks (1996) cautions that "the method is still in a developmental stage with many questions remaining unanswered" (Stranks 1996, p. 122) and "that a relationship exists between physiognomy and climate is clear. Whether it can successfully be applied to fossil floras in order to extract climate and altitude, however remains to be resolved" (Stranks 1996, p. 124). Though she does not use the term "spatial autocorrelation," she correctly observes that "the response of southern hemisphere sites in general cannot be compared to those of northern hemisphere sites" (Stranks 1996,

p. 124). Greenwood et al. (2004) support this contention. Kennedy (1998) lists several sources of potential error and admits that "qualitative sources of error, such as subjectivity in morphotyping and taphonomic bias, could potentially introduce large amounts of uncertainty into palaeoclimatic interpretations" (Kennedy 1998, p. 20). In contrast to this conservatism, many publications suggest that: "CLAMP...is a powerful paleoclimate proxy with the ability to yield quantitative data on past temperatures, precipitation, growing season length, and humidity, as well as enthalpy" (Spicer et al. 2005, p. 429).

Some of the sources of error that must be dealt with are, in rough, increasing order of relative importance or difficulty of quantification:

1. Binomial sampling error. This is the simple and well-understood error associated with the random selection with replacement of  $n$  leaves out of a population of which a proportion  $P$  have untoothed margins. If this selection is repeated many times, the standard error of  $P$  should approach  $\sqrt{P(1-P)/n}$ . This imposes a minimum error on the order of a few degrees with floras of about 30 species. In floras that have many more species (e.g., >100), the binomial error becomes insignificant (Wilf 1997).
2. Repeatability of coding. At this stage, it is not clear what errors may be produced by different people coding the same floras, so this potential source of error is not readily distinguishable from spatial autocorrelation or the study effect discussed above. Future work will invest this source of error using blind experiments.
3. Spatial autocorrelation and irregular sampling. The current sampling distribution is very poor, but can be improved by collecting more samples where they are lacking, by gridding the available locality data on a raster and applying statistical tools spatially, and by creating spatially distributed artificial floras from species range data as has been done by Traiser et al. (2005). Unfortunately, climate station data are seldom or never available from exactly the same places as floras are collected. Up to a point, this can be addressed by appropriate methods of interpolation, but errors introduced by microclimatic variation and patchy species distributions may continue to remain problematic.
4. Inherent time-averaging. This is not an issue if

MAT is the only dependent variable, but MAT is a grossly time-averaged quantity that will be perceived differently if data on, e.g., mean monthly temperatures are compared across studies. It is easy to illustrate how dramatically plants have evolved to respond to the *timing* of temperature change: CAM plants open their stomata during the night when it is cool and transpiration is reduced. As soon as one calculates average daily temperatures—much less monthly or yearly means—from an hourly record, one loses the ability to explain an entire evolutionary strategy that allows thousands of species of plants to exist. This is an extreme example, but the more general point that different temporal scales will affect the significance of variables like temperature must be taken into consideration.

5. Other sources of noise (elevation, microclimate, disturbance, soil type, systematics, taphonomy, etc.). All of these variables are known to be of importance at particular spatial and temporal scales, and must be considered. Is the sample skewed by collection of more low-altitude floras than high-altitude? Do secondary-growth (recently cleared) forests respond differently than primary forests? In the absence of clear answers to these questions about systematic biases, calculation of a stochastic binomial sampling error becomes nearly irrelevant.
6. Uniformity through time. How far back in time can spatial patterns observed in the modern day be extrapolated? This is a broad question facing all methods of reconstructing ancient climates; a simple criterion that is often implicitly invoked is that a method must work increasingly well as it approaches the present; hence error must increase as we go back in time.

The error figures usually associated with estimates of mean annual temperature (MAT) from leaf morphology are usually one- or two-standard deviation analytic errors calculated by assuming only binomial sampling error or normally distributed stochastic variation in the explanatory variables and then propagating this error through a regression line. When the number of species increases much beyond a typical 30, these analytical errors are dramatically reduced, which has led to the appearance in the literature of, for instance, temperature estimates of plus or minus a few degrees (Burnham et al. 2001, Kowalski 2002, Kennedy et al. 2002). Even errors of under a degree have

appeared, which as Miller et al. (in press) point out is incompatible with a rigorous error analysis of the relationship between P and MAT.

Errors 4–6 may be ultimately unquantifiable and uncorrectable, but there is abundant evidence that the issue of spatial autocorrelation can be handled. Work by Thompson et al. (1999) provides graphical tools for plotting floras in ecological space and Traiser (2004) and Traiser et al. (2005) give spatially distributed leaf physiognomic data from synthetic floras for the whole continent of Europe. In concert with the sort of exploratory data analysis that is presented here, these techniques may make it possible, not only to improve estimates of terrestrial paleoclimates, but also to extract additional types of data about how environments and plant ecosystems have changed through time.

## CONCLUSIONS

A graphical exploratory investigation of CLAMP data reveals further serious and unaddressed statistical issues with the standard procedures used to analyze such data. *Exploration* and *estimation* are different goals. If the only utility envisioned for fossil leaf floras is the production of ever-more-precise but possibly inaccurate climate estimates, then the current methods of publication and analysis of CLAMP data are satisfactory. In order, however, to understand the ways in which plant leaves respond to environmental stimuli in the context of real communities, we need application of looser, more flexible tools for data analysis, an appraisal of uncertainty that accounts for systematic bias and unquantifiable noise as well as trivial stochastic errors, and the publication of raw data in a form that can be compared between studies. Graphical techniques like pairs plots are effective methods of exploratory analysis of multivariate data, but theories of biological interest like mechanistic models of leaf response to environmental variables cannot be tested against such data unless the standard forms in which the data are currently published are extended to include the raw (by species) scores.

Leaf morphology remains a valuable and under-exploited source of multivariate data. The CLAMP method is not ideal, but it is the best source of data currently available. It can give satisfactory results if the data it produces are published and analyzed appropriately. From a biological as opposed to a strictly paleoclimatological perspective, appropriate analysis consists of taking the “climate” out of CLAMP and allowing multivariate data

on leaf architecture to illuminate broader ecological questions. The pairs plot as a tool for graphical exploratory analysis provides information on complex covariation among leaf-physiognomic variables, and allows evaluation of systematic errors in CLAMP data, neither of which can be done with eigenvector methods of data reduction or with hierarchical clustering. This has the potential to make multivariate leaf-physiognomic data interesting not only to paleoclimatologists, but also to plant morphologists and functional ecologists. Moreover, the exploratory graphical approach advocated here may prove valuable in other paleontological data sets where current analyses obscure interesting detail in complex, multivariate data.

### ACKNOWLEDGMENTS

Funding for this work was provided by a Yale University Fellowship and by NSF Grant DEB-0235736. Thanks to L.J. Hickey, I.M. Miller, D. Royer, A. Seilacher, C. Traiser, and several anonymous reviewers for comments on the manuscript.

### REFERENCES

- Bailey, I.W. and Sinnot, E.W. 1915. A botanical index of Cretaceous and Tertiary climates. *Science*, 41: 831-834.
- Basford, K.E. and Tukey, J.W. 1999. *Graphical Analysis of Multiresponse Data: Illustrated with a Plant Breeding Trial*. London: Chapman and Hall.
- Burnham, R.J., Pitman, N.C.A., Johnson, K.R., and Wilf, P. 2001. Habitat-related error in estimating temperatures from leaf margins in a humid tropical forest. *American Journal of Botany*, 88(6):1096-1102.
- Forest, C.E., Wolfe, J.A., Molnar, P., and Emanuel, K.A. 1999. Paleothermometry incorporating atmospheric physics and botanical estimates of paleoclimate. *Geological Society of America Bulletin*, 111(4):497-511.
- Greenwood, D.R., Wilf, P., Wing, S.L., and Christophel, D.C. 2004. Paleotemperature estimation using leaf-margin analysis: Is Australia different? *Palaeos*, 19:129-142.
- Gregory-Wodzicki, K.M. 2000. Relationships between leaf morphology and climate, Bolivia: implications for estimating paleoclimate from fossil floras. *Paleobiology*, 26(4):668-688.
- Huff, P.M., Wilf, P., and Azumah, E.J. 2003. Digital future for paleoclimate estimation from fossil leaves? Preliminary results. *Palaeos*, 18:266-274.
- Jacobs, B.F. 1999. Estimation of rainfall variables from leaf characters in tropical Africa. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 145:231-250.
- Jacobs, B.F. 2002. Estimation of low-latitude paleoclimates using fossil angiosperm leaves: examples from the Miocene Tugen Hills, Kenya. *Paleobiology*, 28(3):399-421.
- Kennedy, E.M. 1998. *Cretaceous and tertiary mMegaflores from New Zealand and their climate signals*. Unpublished Ph.D. Thesis, Open University.
- Kennedy, E.M., Spicer, R.A., and Rees, P.M. 2002. Quantitative palaeoclimate estimates from Late Cretaceous and Paleocene leaf floras in the northwest of the South Island, New Zealand. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 184:321-345.
- Kowalski, E.A. 2002. Mean annual temperature estimation based on leaf morphology: a test from tropical South America. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 188:141-165.
- Lep, J. and Milauer, P. 2003. *Multivariate Analysis of Ecological Data Using CANOCO*. Cambridge: Cambridge University Press.
- Miller, I.M., Brandon, M., and Hickey, L.J. (in press). Reading the leaves of the Methow Basin (Washington State): Paratropical signature supports a far-traveled Baja BC. *Earth and Planetary Science Letters*.
- Mosbrugger, V., Traiser, C. and Klotz, S. 2005. "Synthetic Floras"—a flexible tool in large scale vegetation analysis. *XVII International Botanical Congress Abstracts*, p. 157.
- R Development Core Team 2004. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Royer, D.L., Wilf, P., Janesko, D.A., Kowalske, E.A., and Dilcher, D.C. 2005. Correlations of climate and plant ecology to leaf size and shape: Potential proxies for the fossil record. *American Journal of Botany*, 92(7):1141-1151.
- Spicer, R.A., Herman, A.B., and Kennedy, E.M. 2004. Foliar Physiognomic Record of Climatic Conditions during Dormancy: Climate Leaf Analysis Multivariate Program (CLAMP) and the Cold Month Mean Temperature. *Journal of Geology*, 112:685-702.
- Spicer, R.A., Herman, A.B., and Kennedy, E.M. 2005. The Sensitivity of CLAMP to Taphonomic Loss of Foliar Physiognomic Characters. *Palaeos*, 20(5):429-438.
- Stranks, L. 1996. *Physiognomic and taphonomic studies in New Zealand and Australia: implications for the use of palaeobotany as a tool for palaeoclimate estimation*. Unpublished Ph.D. Thesis, University of Oxford, Oxford.
- Ter Braak, C.J.F. 1986. Canonical correspondence Analysis: a new eigenvector technique for multivariate direct gradient analysis. *Ecology*, 67: 1167-1179.
- Thompson, R.S., Anderson, K.H. and Bartlein, P.J. 1999. Atlas of Relations Between Climatic Parameters and Distributions of Important Trees and Shrubs in North America. *U.S. Geological Survey Professional Paper*, 1650a&b.

- Traiser, C. 2004. *Blattphysiognomie als Indikator für Umweltparameter: Eine Analyse rezenter und fossiler Floren*. Unpublished Ph.D. Thesis, Eberhard-Karls Universität Tübingen.
- Traiser, C., Klotz, S., Uhl, D., and Mosbrugger, 2005. Environmental signals from leaves—a physiognomic analysis of European vegetation. *New Phytologist*, 166:465-484.
- Wilf, P. 1997. When are leaves good thermometers? A new case for Leaf Margin Analysis. *Paleobiology*, 23(3):373-390.
- Wilf, P., Wing, S.L., Greenwood, D.R. and Greenwood, C.L. 1998. Using fossil leaves as paleoprecipitation indicators: An Eocene example. *Geology* 26(3):203–206.
- Wolfe, J.A. 1993. A method of obtaining climatic parameters from leaf assemblages. *U.S. Geological Survey Bulletin*, 2040, 73 pp.
- Wolfe, J.A.. 1995. Paleoclimatic estimates from Tertiary leaf assemblages. *Annual Review of Earth and Planetary Sciences*, 23:119-142.
- Wolfe, J.A. and Spicer, R.A. 1999. Fossil Leaf Character States: Multivariate Analysis. In Jones, T.P. and Rowe, N.P. (eds.), *Fossil Plants and Spores: Modern Techniques*, Geological Society, London, 233-239.



## APPENDIX

This file is an attempt to document the script, data, and functions (in the R graphics/statistics programming language) used in the analyses I carried out for the article 'Loosening the CLAMP', which is to be electronically published in *Palaeontologia Electronica* in 2006. All citations should refer to this article.

In addition to the article and this README file, there are<sup>1</sup>:

a file called Rscript.txt in the R programming language

a data/ directory with 5 flat tab-delimited ascii tables,

a functions/ directory with three functions called by Rscript.txt

Like any software, this is a work-in-progress, so I recommend that anyone who intends to replicate my results contact me for newer and better-documented versions of these files. For anyone with modest familiarity with the R language, how-

ever, it should be possible to run the file Rscript.txt after making the necessary modifications for the platform on which it is to be run. In addition to setting a filepath in the indicated places, it will be necessary to source() the functions in the functions/ directory. All the data have been published before (in the cited publications) but are provided here for the first time in digital, machine-readable form. Note that the files were optically scanned or hand-typed and DO contain errors. The first part of the script file is devoted to laundering the data and cleaning some of these errors up, but any use of the data tables without running Rscript.txt should expect to find errors in the data. Nor is there any guarantee that my script has caught all the errors to be found....

Queries and bug reports to walton.green@yale.edu (Walton A. Green, Department of Geology, Yale University, P.O. Box 208109 Yale Station, New Haven, Connecticut, 06520). All data and computer code are intended for free distribution under copyleft protection like GNU General Public License v. 2 or later.

1. Files are available online at [http://paleo-electronica.org/paleo/2006\\_2/clamp/index.html](http://paleo-electronica.org/paleo/2006_2/clamp/index.html)