Generic Graphics for Uncertainty and Sensitivity Analysis

R. M. Cooke
Dept. Mathematics, TU Delft, The Netherlands

J. M. van Noortwijk
HKV Consultants, Lelystad, The Netherlands

ABSTRACT: We discuss graphical methods which may be employed generically for uncertainty and sensitivity analysis. This field is rather new, and the literature reveals very little in the way of theoretical development. Perhaps it is the nature of these methods that one simply 'sees' what is going on. Apart from statistical reference books [Cleveland, 1993] which focus on visualizing data, the main sources for graphical methods are software packages. Our focus is visualization to support uncertainty and sensitivity analysis. A simple problem for illustrating four generic graphical techniques, namely tornado graphs, radar plots, multiple scatter plots, and cobweb pots. Uncertainty analysis codes, with or without graphic facilities, were benchmarked in a recent workshop of the technical committee Uncertainty Modelling of the European Safety and Reliability Association. The report [Cooke 1997] contains descriptions and references to the codes, as well as simple test problems. An extended version of this paper will appear in [Saltelli appearing].

1. A SIMPLE PROBLEM

The following problem has been developed for the Cambridge Course for Industry, Dependence Modeling and Risk Management [Cambridge Course for Industry 1998], and it serves to illustrate the generic techniques. Suppose we are interested in how long a car will start after the headlights have stopped working. We build a simple reliability model of the car consisting of three components:
the battery (bat), the headlight lampbulb (bulb), the starter motor (strtr). The headlight fails when either the battery or the bulb fail. The car's ignition fails when either the battery or the starter motor fail. Thus considering bat, bulb and strtr as life variables:
headlite = min(bat, bulb), ignitn = min(bat, strtr). The variable of interest is then ign-head = ignitn - headlite. Note that this quantity may be either positive or negative, and that it equals zero whenever the battery fails before the bulb and before the starter motor.

We shall assume that bat, bulb and strtr are independent exponential variables with unit expected lifetime. The question is, which variable is most important to the quantity of interest ign-head?

2. TORNADO GRAPHS

Tornado graphs are simply bar graphs arranged vertically in order of descending absolute value. The spreadsheet add-on Crystal Ball performs uncertainty analysis and gives graphic output for sensitivity analysis in the form of tornado graphs (without using this designation). After selecting a 'target forecast variable', in this case ign-head, Crystal Ball shows the rank correlations of other input variables and other forecast variables as in Figure 1.

The values, in this case rank correlation coefficients, are arranged in decreasing order of absolute value. Hence the variable strtr with rank correlation 0.56 is first, and bulb with rank correlation -0.54 is second, and so on. When influence on the target variable, ign-head, is interpreted as rank correlation, it is easy to pick out the most important variables from such graphs. Note that bat is shown as having rank correlation 0 with the target variable ign-head. This would suggest that bat was completely unimportant for ign-head.

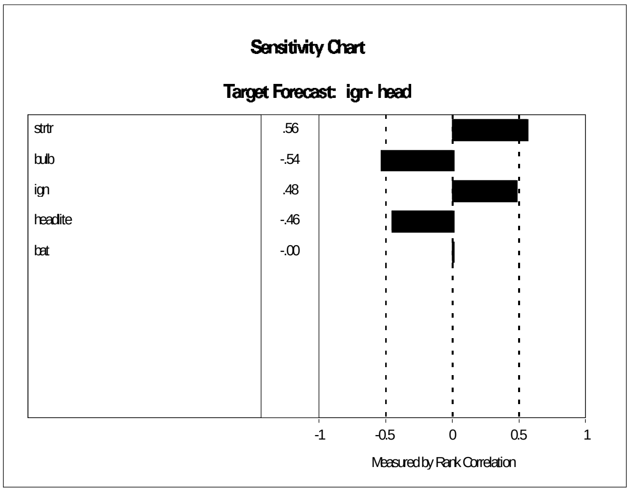

Figure 1

3. RADAR GRAPHS

Radar graphs provide another way of showing the information in Figure 1. Figure 2 shows a radar graph made in EXCEL by entering the rank correlations from Figure 1. Each variable corresponds to a ray in the graph. The variable with the highest rank correlation is plotted furthest from the midpoint, and the variable with the lowest rank correlation is plotted closest to the midpoint. The real value of radar plots lies in their ability to handle a large number of variables. Figure 3 is a radar plot showing six variables of interest (radiation dose to various organs) in different colors, against 161 explanatory variables [Goossens et al 1997]. In A3 it is possible to view all this information at one time.

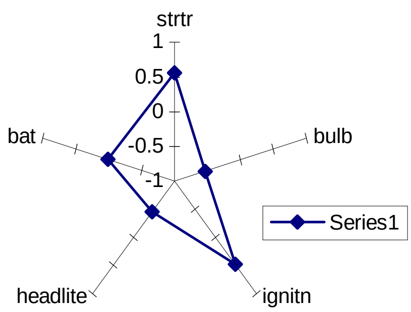

Figure 2

4. MULTIPLE SCATTER PLOTS

Statistical packages support varieties of scatter plots. For example, SPSS provides a multiple scatter plot facility. Simulation data produced by UNICORN for 1000 samples has been read into SPSS to produce the plot matrix shown in Figure 6. We see pairwise scatter plots of the variables in our problem. The first row, for example, shows the scatter plots of ign-head and, respectively, bat, bulb, strtr, ignitn and headlite.

Figure 4 contains much more information than Figure 1. Let (a,b) denote the scatter plot in row a and column b, thus (1,2) denotes the second plot in the first row with ign-head on the vertical axis and bat on the horizontal axis. Note that (2,1) shows the same scatter plot, but with bat on the vertical and ing-head on the horizontal axes.

Although Figure 1 suggested that bat was unimportant for ign-head, Figure 4 shows that the value of bat can say a great deal about ign-head. Thus, if bat assumes its lowest possible value, then the values of ign-head are severely constrained. This reflects the fact that if bat is smaller than bulb and strtr, then ignitn = headlite, and ign-head = 0. From (1,3) we see that large values of bulb tend to associate with small values if ign-head; if bulb is large, then the headlight may live longer than the ignition making ign-head negative. Similarly, (1,4) shows that large values of strtr are associated with large values of ign-head. These facts are reflected in the rank correlations of Figure 1.

In spite of the above remarks, the relation between *rank* correlations depicted in Figure 1 and the scatter plots of Figure 4 is not direct. Thus, bat and bulb are statistically independent, but if we look at (2,3), we might infer that high values of bat tend to associate with low values of bulb. This however is an artifact of the simulation. There are very few very high values of bat, and as these are independent of bulb, the corresponding values of bulb are not extreme. If we had a scatter plot of the *rank* of bat with the *rank* of bulb, then the points would be uniformly distributed on the unit square.

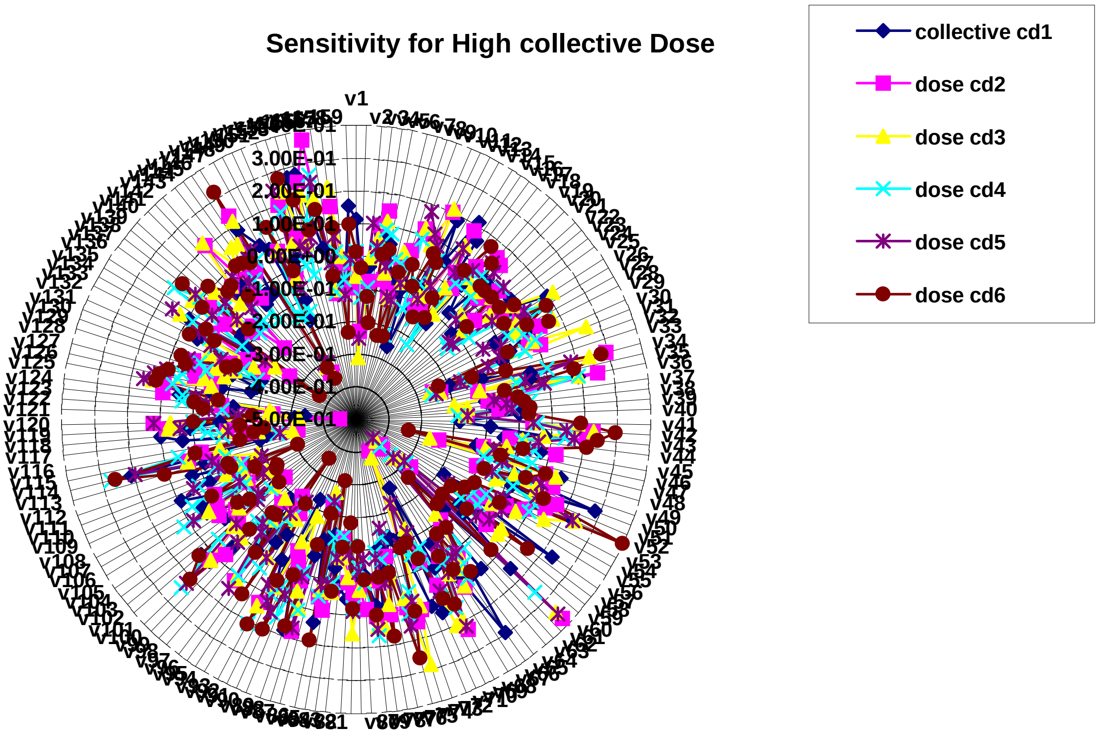

Figure 3

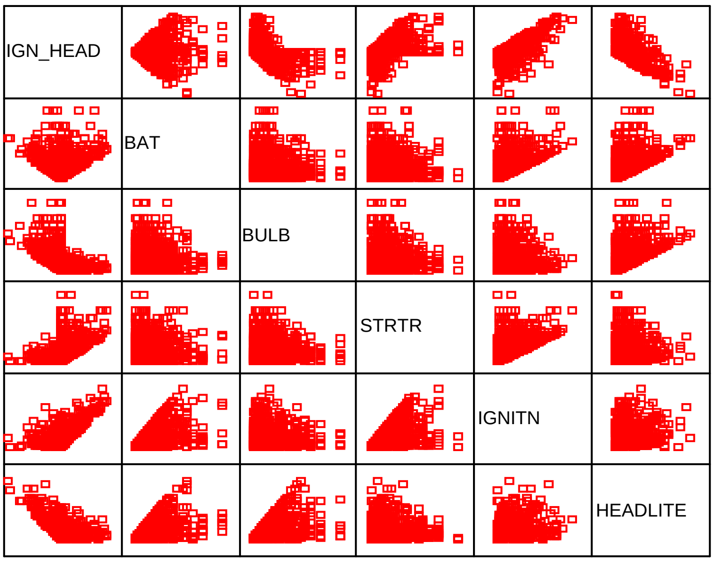

Figure 4

Figure 5 shows an overlay scatter plot. Ign-head is depicted on the vertical axis, and the values for bat, bulb and strtr are shown as blue, green and red points respectively, when viewed in color. Figure 5 is a superposition of plots (1,2), (1,3) and (1,4) of Figure 5. However, Figure 5 is more than just a superposition. Inspecting Figure 5 closely, we see that there are always a red, green and blue point corresponding to each realized value on the vertical axis. Thus at the very top there is a green point at ign-head = 238 and bulb slightly greater than zero. There are blue and red points also corresponding to ign-head = 238. These three points correspond to the same sample. Indeed, ign-head attains its maximum value when strtr is very large and bulb is very small. If a value of ign-head is realized twice, then there will be two triples of blue-green-red points on a horizontal line corresponding to this value, and it is impossible to resolve the two separate data points. For ign-head = 0, there are about 300 realizations.

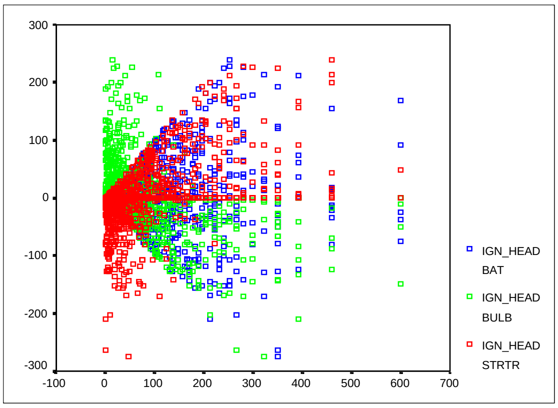

Figure 5

The distribution underlying Figure 4 is six dimensional. Figure 4 does not show this distribution, but rather shows 36 two-dimensional projections from this distribution. Figure 5 shows more than a collection of two dimensional projections, as we can sometimes resolve the individual data points for bat bulb, strtr and ign-head, but it does not enables us to resolve all data points. The full distribution is shown in cobweb plots.

5. COBWEB PLOTS

The uncertainty analysis program UNICORN contains a graphical feature that enables interactive visualization of a moderately high dimensional distribution. Our sample problem contains six random variables. Suppose we represent the possible values of these variables as parallel vertical lines. One sample from this distribution is a six-vector. We mark the six values on the six vertical lines and connect the marks by a jagged line. If we repeat this 1000 times we get Figure 6 below. We can recognize the exponential distributions of bat, bulb and strtr. Ignitn and headlite, being the minimum of independent exponentials, are also exponential. Ign-head has a more complicated distribution. The graphs at the top are the ‘cross densities’; they show the density of line crossings midway between the vertical axes. The role of these in depicting dependence becomes clear when we transform the six variables to ranks or percentiles, as in Figure 7:

Line shading is introduced to identify the four quartiles of the left most variable. A number of striking features emerge when we transform to the percentile cobweb plot. First of all, there is a curious hole in the distribution of ign-head. This is explained as follows. On one third of the samples, bat is the minimum of {bat, bulb, strtr}. On these samples ignitn=headlite and ign-head = 0. Hence the distribution of ign-head has an atom at zero with weight 0.33. On one-third of the samples strtr is the minimum and on these samples ign-headlite is negative, and on the remaining third bulb is the minimum and ign-head is positive. Hence, the atom at zero means that the percentiles 0.33 up to 0.66 are all equal to zero. The first positive number is the 67-th percentile.

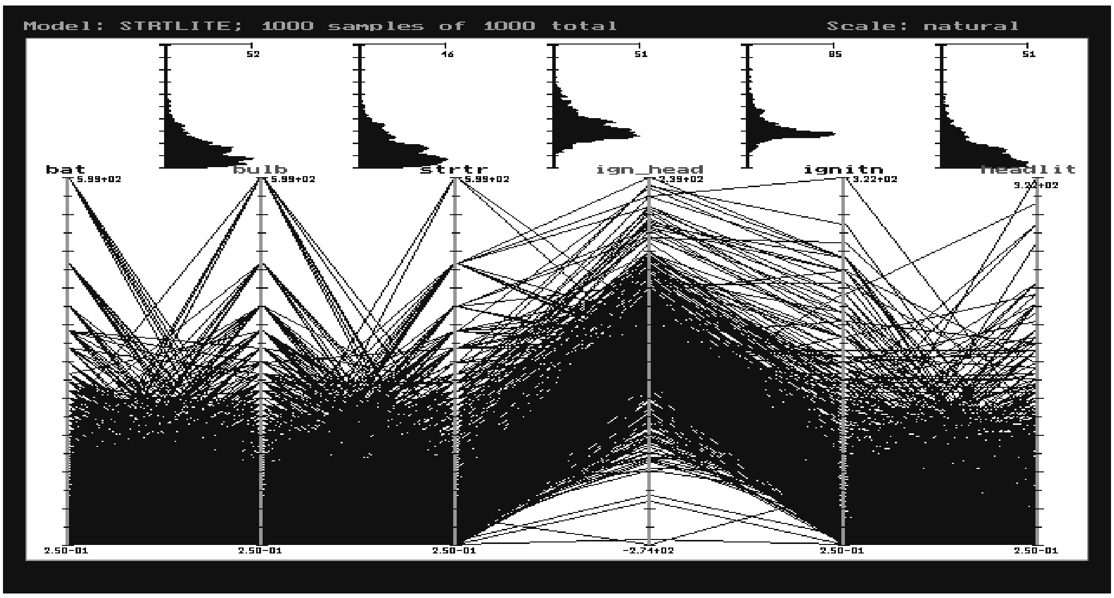

Figure 6

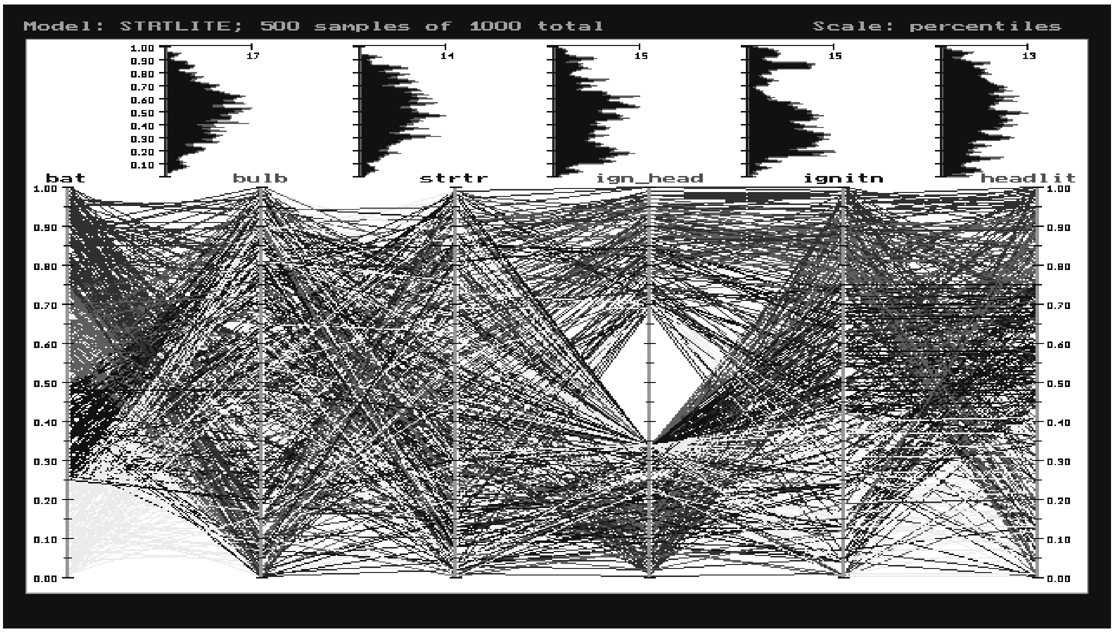

Figure 7

Note the cross densities in Figure 7. One can show the following for two adjacent continuously distributed variables X and Y in a percentile cobweb plot[1]:

- If the rank correlation between X and Y = 1 then the cross density is uniform
- If X and Y are independent (rank correlation 0) then the cross density is triangular
- If the rank correlation between X and Y = -1, then the cross density is a spike in the middle.

Intermediate values of the rank correlation yield intermediate pictures. The cross density of ignitn and headlite tends toward uniform, and the rank correlation between these variables is 0.42.

Cobweb plots support interactive conditionalization; that is, the user can define regions on the various axes and select only those samples which intersect the chosen region. Figure 8 shows the result of conditionalizing on ign-head = 0, Notice that if ign-head = 0, then bat is almost always the minimum of {bat,bulb,strtr}, and ignitn is almost always equal to headlite. This is reflected in the conditional rank correlation between ignitn and headlite almost equal to 1. We see that the conditional correlation as in Figure 8 can be very different from the unconditional correlation of Figure 7. From Figure 8 we also see that bat is almost always less than bulb and strtr.

Cobweb plots allow us to examine local sensitivity. Thus we can say, suppose ign-head is very large, what values should the other variables take? The answer is gotten simply by conditionalizing on high values if ign-head. Figure 9 shows conditionalization on high values of ign-head, Figure 10 conditionalizes on low values. If ign-head is high, then bat, strtr and ignitn must be high. If ign-head is low then bat, bulb and headlite must be high. Note that bat is high in one case and low in the other. Hence, we should conclude that bat is very important both for high values and for low values of ign-head. This is a different conclusion than we would have drawn if we considered only the rank correlations of Figures 1 and 2

These facts can also be readily understood from the formulae themselves. Of course the methods come into their own in complex problems where we cannot see these relationships immediately from the formula. The graphical methods then draw our

[1] These statements are easily proved with a remark by Tim Bedford. Notice that the cross density is the density of X+Y, where X and Y are uniformly distributed on the unit square. If X and Y have rank correlation 1, then X+Y is uniform [0,2]; if they have rank correlation -1 then X + Y = 1; if X and Y are independent then the mass for X+Y=a is proportional to the length of the segment x+y=a.

attention to patterns which we must then seek to understand.

REFERENCES

Cambridge Course for Industry, 1998, *Dependence Modelling and Risk Management*, University of Cambridge, Cambridge.

Cleveland, W.S. *Visualizing Data, 1993*, Marray Hill, New Jersey: AT&T Bell Laboratories.

Cooke, R.M. 1995, UNICORN: Methods and code for uncertainty analysis, AEA Technology, Cheshire.

Crystal Ball, version 4.0, 1996, Users Manual, Decisioneering Inc. Aurora, Colorado.

European Safety and Reliability Association Technical Committee Uncertainty Modelling, 1997, *Benchmark Workshop* (Cooke, R.M. ed), Delft.
EXCEL for Windows 95, 1995, Microsoft.

Goossens, L.H.J., J.D. Harrison, B.C.P. Kraan, R.M. Cooke, F.T. Harper, and S.C. Hora, *Probabilistic Accident Consequence Uncertainty Analysis: uncertainty assessment for internal dosimetry*, NUREG/CR-6571, EUR 16773, SAND98-0119, vol. 1, prepared for U.S. Nuclear Regulatory Commission, Commission of the European Communities, 1997.

Saltelli, A. (appearing), *Mathematical and Statistical Methods for Sensitivity Analysis of Model Output*, Wiley, New York.

SPSS for Windows, 8.00, 1997, SPSS.

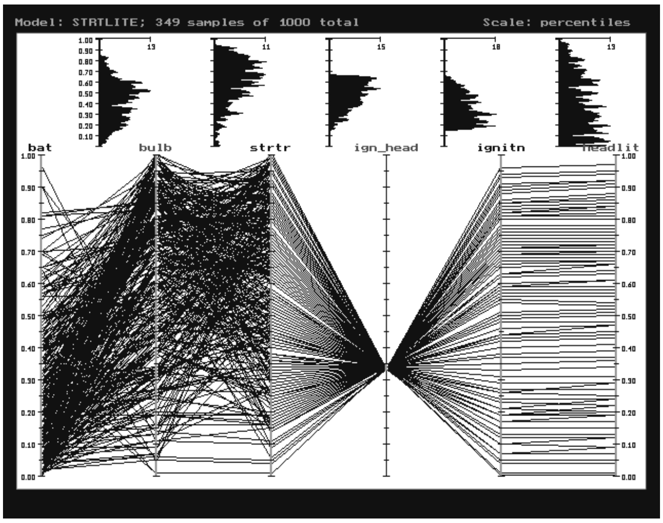

Figure 8

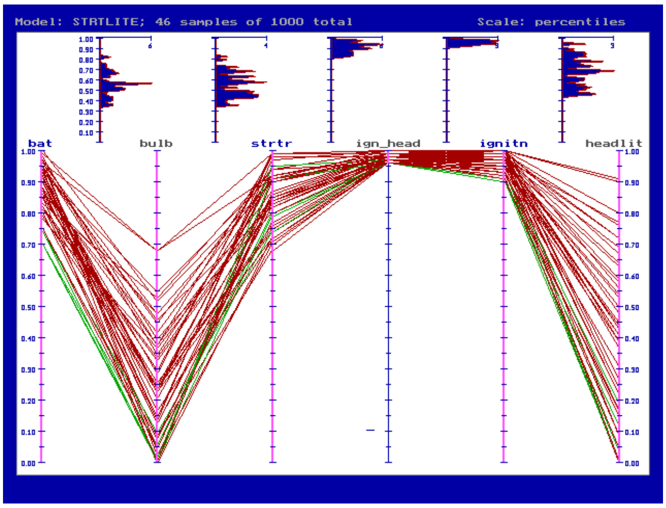

Figure 9

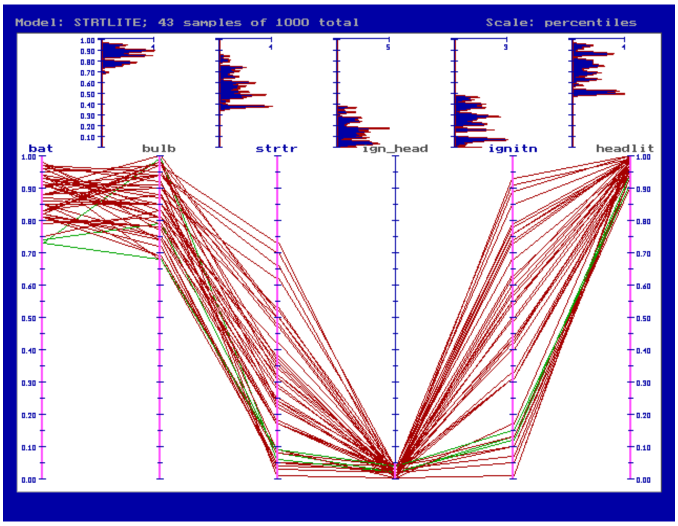

Figure 10