IDS/TP/8

TREND SURFACE ANALYSIS (MX23)

by

M. Norton-Griffiths

and

L. Pennycuick

Technical Paper No. 8

INSTITUTE FOR DEVELOPMENT STUDIES
UNIVERSITY OF NAIROBI
P.O. Box 30197
Nairobi, Kenya.

July 1974

This paper stems from the authors' work in the Serengeti Ecological Monitoring Programme of the Serengeti Research Institute, sponsored by the African Wildlife Leadership Foundation, Washington, D.C.

Any views expressed in this paper are those of the authors. They should not be interpreted as reflecting the views of the Serengeti Research Institute, the African Wildlife Leadership Foundation, the Institute for Development Studies or the University of Nairobi.

IDS/TP/8

TREND SURFACE ANALYSIS

MX23

INTRODUCTION

Trend surface analysis has long been used by geographers, geologists and ecologists for fitting surfaces to data recorded at sample points scattered within a two dimensional sample space. (see bibliography). The sample points are located by x- and y- coordinates, z being the value of the variable measured at each point. The data for trend surface analysis therefore consist of a number of x- y- coordinates each with an associated z value. The objective of the analysis is to fit the 'best' surface through the z values in order to reveal the basic patterns of gradients and contours within the sample space.

Trend susrface analysis is essentially a three-dimensional multiple regression. In an ordinary multiple regression (one dimensional curve fitting), increasingly higher order polynomials are calculated to give lines of increasing complexity that pass through the data points (linear, quadratic, cubic, etc.). The curve of 'best' fit is found from an analysis of variance. Trend surface analysis proceeds in exactly the same way. Successively higher order polynomials are calculated from the x- y- coordinates to give the equations for surfaces of increasing complexity that pass through the z values. An analysis of variance is also used to find the surface of 'best' fit.

PROGRAMME STRUCTURE

The programme reads in the following data:

(i) the maximum size of the sample space under consideration,
(ii) a text heading for identification of all output and
(iii) a series of x- y- coordinates, each with an associated z value.

The programme goes through a number of stages of analysis. At each stage it calculates the trend and component for linear, quadratic, cubic, quartic and quintic surfaces and at each stage outputs the values of the regression coefficients and the surfaces. The surfaces are output in the form of the calculated values of z for each of the grid coordinates specified by the maximum size of the sample space under consideration. Finally, an analysis of variance table is output, and the programme then returns to read in more data.

The scheme of the analysis is shown in Figure 1. It is easiest to think entirely in terms of sums of squares. The total sums of squares (Σx^2 - $(\Sigma x)^2/N$) are those of the original set of z values, N being the number of sample points. As each higher order polynomial is calculated an increasing percentage of these sums of squares is accounted for. Each 'trend' is made up from a number of components, represented by the new terms added at each stage, and at each stage the new component accounts for some of the residual sums of squares of the previous stage. The percentage of the total sums of squares accounted for by each trend gives the 'goodness of fit' of the surface. The analysis of variance table is based on these sums of squares, and from it the significance of each trend can be determined. The surface of best fit is given by the last component which significantly reduces the residual sums of squares from the previous stage in the analysis.

DATA INPUT (Figure 2)

Control Cards

The form in which this programme will be available has not yet been decided upon. Users will therefore have to contact the University of Nairobi Computer Centre to find out which control cards are required.

A Series of Data

A series of data for this programme is defined as a number of sets of x- y- coordinates, each with an associated z value, on which a trend surface analysis is to be carried out. Each series of data requires user to specify the maximum size of the sample space as well as a text heading to identify all output for that series. The data follows after these two specifications. The programme will process any number of series of data. A 'last card' is required to tell the programme when to stop. All data are input on 80 col punch cards.

Data Preparation

A series of data consists of a number of sets of x- y- coordinates, each with an associated z value. It is assumed that the x- and y- axes are at right angles to each other, and that the origin is in the top left hand corner.

The x- coordinate therefore gives the distance of a sample point along the horizantal x axis, while the y- coordinate gives its distance along the vertical y axis (See Figure 3.)

The maximum permissable value of x is 25, and the maximum permissable value of y is 29.

Card 1: This specifies the maximum size of the surface that is to be output (maximum x value is 25, maximum y value is 29). The x value is put in cols. 1 and 2, and the y value in cols. 3 and 4. Both values are right justified, see the second example in Figure 2 where the 9 is in col. 2 and the 8 is in col. 4.

Card 2: This is a text heading that is used to identify all outputs from a series of data. All 80 cols. may be used, for both characters and numbers.

Data Cards: These cards contain sets of x- and y- coordinates, each with and associated z value. A maximum of ten sets may be contained on any one data card. Each set consists of seven columns on the card, in which:

cols. 1 & 2 contain the x value, right justified, maximum value 25
cols. 3 & 4 contain the y value, right justified, maximum value 29
cols. 5,6 & 7 contain the z value, right justified. No decimal point is allowed, although the value can be negative, in which case it is preceeded by a minus sign.

As many data cards are filled in as are required. When the last set of x-, y- and z values have been filled in, a -1 is put into the columns that would have contained the next x- value, even if this requires a separate data card (See Figure 2.) This -1 tells the programme that all the data for one series has been read in.

The next series of data starts with a Card 1, followed by as Card 2 and subsequent data cards. Any number of series of data may be input.

Last Card: When all the series of data have been prepared, a card with -1 in cols. 1 & 2 is required. This last card tells the programme to stop (See Figure 2.)

Notes on Data Preparation

Maximum Size of x- and y-: This is an unfortunate constraint caused by the available core store. In large sample spaces the x- y- coordinates have to refer to grid squares rather than to precise grid coordinates. For example, in the 250 km x 280 km Serengeti Ecosystem, the grid coordinates of rain gauges referred to 10 x 10 km grid squares. In smaller sample spaces the grid squares can be 5 x 5 km, 1 x 1 km, etc.

The x- and y- Axes: The maximum /specified size of the x- and y- axes defines the smallest rectangle that contains all the sample points at which observations were made. The programme uses this information to output the different surfaces by calculating the value of z for each of the specified x- y- coordinates.

Although it is conceptually simplest to have the x axis oriented west-east and the y axis north-south, this is not essential. All that is required is that the two axes are at right angles to each other, and that the origin is in the top left hand corner. Their orientation with respect to geographic north is irrelevant.

Maximum Size of z: The maximum range of z is from 999 to -99 (the minus has to be entred in the data, while a + is implicit) . Data may therefore have to be transformed to fit this range. Decimals are not permitted, so transformations may again be required.

No Observations: It is not necessary to observe all possible sample points within the defined sample space. Points that are not observed are not entered into the data. However, points that are observed, and where the value of z was 0, must be entred.

OUTPUT

The first output contains the text heading and:

N: the total number of sample points,

Mean Density: mean value of the parameter z at those points,

Variance: the variance of those z values.

Regression Coefficients and Surfaces

For the linear, quadratic, cubic, quartic and quintic analyses, the programme outputs:

(a) the text heading

(b) the constant followed by the regression coefficients, in E format

The identity of the regression coefficients is shown by b_1, b_2, b_3 ....b_{20} in Figure 1. Thus for the linear equation, the identities are:

$z = \text{constant} + b_1.x + b_2.y$ (the constant and b_1 and b_2 being output),

while for the quadratic the identities are:

$z = \text{constant} + b_1.x + b_2.y + b_3.x^2 + b_4.xy + b_5.y^2$ (the constant, and $b_1...b_5$ being output) and so on.

The E format is interpreted as follows :

.0.2174	E 03	=	217.4
-0.6240	E 01	=	-6.24
-0.4414	E-01	=	-0.04414

and so on.

(c) the surfaces

After each set of regression coefficients the appropriate surface is output. The x and y axes are written, and the value of z at each of the x-y coordinates. The z is calculated from the multiple regression equation for the surface. The maximum size of the surface is that specified by the user.

The Analysis of Variance Table

This is the most important part of the trend surface analysis. It is identified by the text heading, and appears once all the regression coefficients and surfaces have been output.

(a) column headings

source: Gives the source of the sums of squares.

sums of squares:

Gives the total sums of squares, and the sums of squares 'explained' by the various trends and components.
Total sum of squares is that of the original set of z values; linear trend sum of squares = variation explained by linear trend (i.e. by the linear equation); linear residual sum of squares = variation not explained by linear trend; thus
linear trend + linear residual = total sum of squares,
similarly for all higher order trends and components. Reference to Figure 1 will assist in interpreting the sum of squares.

coefficient of determination:

Gives the % of the variation explained by and added by a trend or component. e.g. coeff. of det. for linear trend $= \frac{\text{linear sum. sq.}}{\text{total sum. sq.}} \times 100\%$

correlation coefficient:

$= \sqrt{\text{coefficient of determination}/100}$

thus, linear trend correlation coefficient

$\sqrt{\frac{\text{linear trend sum squares}}{\text{total sum of squares.}}}$

DF (degrees of freedom)

Gives the degrees of freedom for each sum of squares; total DF = N-1; trend DF = number of terms in x and/or y in trend surface equation (e.g. cubic trend DF = 9 for there are nine terms x, y, x^2 xy, y^2, x^3, x^2y, xy^2, y^3. Constant is not counted).

Component DF = number of terms <u>added</u> for that order polynomial (e.g. cubic component has 4 DF, for 4 terms are added x^3, x^2y, xy^2, y^3);

Residual DF = total DF - trend DF.

Mean Square:

The mean squares are found by dividing the sum of squares by their degrees of freedom. Thus, mean square = sum of squares/DF

e.g. quintic component mean square $= \frac{\text{quintic component sum of sq.}}{\text{quintic component DF}}$

F : the variance ratio

F = mean square of trend or component / residual mean square,

e.g. cubic component F = cubic component mean square/cubic residual ms,

e.g. quintic trend F = quintic trend ms/ quintic trend residual ms.

(b) interpreting the analysis of variance table

The objective of the analysis of variance table is to be able to test the significance of each trend and to decide upon the surface of best fit.

(i) the significance of trends

The F ratios for he trends can be looked up in Tables of F for the appropriate degrees of freedom. Some, or all, of the trends may be significant. It may also happen that only one will be significant, and it could be any one of them. Don't be suprised by anything!

(ii) the significance of components
A component's ratio may be looked up in tables of F in the same way. A significant trend may not have a significant component associated with it.

(iii) the surface of best fit
This is decided on by the last component that is significant. For example, one might find that all the trends are significant, but only the quadratic and cubic components are significant. The cubic surface would thus be the surface of best fit, for this was the last component that significantly reduced the residual sums of squares from the previous stage. Although the quartic and quintic trends account for more of the total sums of squares, the gain in information is spurious.

(iv) no significant trends
It sometimes happens that no trends are significant, in which case bad luck. This means that there are no systematic variations across the sample space. One could always try the data again by transforming it in some way to reduce the total sums of squares (e.g. log all the z values, or take their cube roots).

RESIDUAL VALUES

The residual values from a trend surface analysis can be found by substracting the observed values of z at the sample points from the calculated values of z output for the surface of best fit. In general it will be found that the residual values from the outlying sample points will be less than those in the centre. (See Vl below) Nonetheless, the spatial distribution of the residual values can be of interest. They indicate points where local effects are marked, and they can also indicate areas where events are not in keeping with the general trends across the sample space. A trend surface of these points can then be carried out if necessary. Alternatively, the residual values themselves can be submitted to trend surface analysis. For example, in a trend surface analysis of the average annual rainfall recorded at 61 rain gauges within the Serengeti Ecosystem, the residuals showed:

(i) high positive values at gauges located near hill(e.g. a local effect of the hills causing higher rainfall than the basic overall trend would suggest)and

(ii) low and/or negative residual values from gauges on the Serengeti Plains. This indicated a general rain shadow effect from the Ngorongoro highlands. A separate trend surface analysis was subsequently carried out on the gauges from the Serengeti Plains.

INTERPRETATION OF THE SURFACE OF BEST FIT

Once the surface of best fit is decided upon, it can be used to draw contour maps, or to measure gradients along transects cutting across the sample space. Two points must be borne in mind when interpreting the surface of best fit. Firstly, the whole objective of the analysis is to reveal the systematic nature of contours and gradients across a sample space. This means that the revealed trends are relevant to the whole sample space and may therefore be fairly inaccurate representations of trends within small areas of the main sample space. As has already been pointed out, the analysis of the residual values can pinpoint locations or areas where non-systematic events may be occuring.

The second important point to remember is that polynomials, especially higher order ones, have the bad habit of 'fixing' on the outlying sample points. It is for this reason that the residual values from the outlying points tend to be smaller than those from within. It also means that the values of z calculated for points lying outside the sample points may be very wild, and may be completely meaningless.

For example, Figure 3 shows the location of 24 sample points within a sample space defined by x = 10 and y = 10. The polynomials will tend to fix on the outlying sample points, and the calculated values of z for the area x = 1 to 3 and y = 1 to 5 can be expected to be meaningless. The maximum value of x- and y-, defined by the user in this programme, gives the size of the rectangle that contains all the sample points. This does not necessarily mean that the calculated surface will be relevant to all parts of the rectangle. It is safest, therefore, to apply the calculated surface only to an area defined by a line that joins up the outlying sample points.

APPLICATIONS

The selected bibliography given below will indicate the most widespread applications of trend surface analysis. Trend surface analysis is also now being applied to wildlife ecological research in East Africa, some of the most useful applications being:-

(i) rainfall maps - by fitting surfaces through annual (or other) rainfall amounts recorded at rain gauges. The surface of best fit is used to draw rainfall contour maps, i.e. isohyet maps.

(ii) density distribution of large mammal species - from density data recorded on systematic or random flight lines. Trend surface analysis reveals the patterns of density distribution across National Parks and Game Reserves.

(iii) Patterns of distribution of other enviromental variables, e.g. fire frequency, greenness of grass, change in water content of rivers, sequence of grass growth, woodland density, drainage density, etc. Distribution patterns of large mammals, and the changes in their distribution patterns, be related to these variables by comparing the calculated surfaces.

(iv) response surfaces - instead of x- and y- axes representing space, they can represent the values of other variables, e.g. grass greenness and grass height. The response surfaces of different large mammals to these two variables can then be measured by trend surface analysis, giving some indications of ecological separation. This approach can be used with component scores, factor scores and discriminant scores.

BIBLIOGRAPHY

1. Chorley, R.J. and Haggett, P. "Trend Surface Maping in Geographical Research." Transactions of the Institute of British Geographers. 37 1965, pp. 47 - 67.

2. Doornkamp, J.C. and King, C.A.M. Numerical Analysis in Geomorphology: an Introduction. London, Edward Arnold, 1971.

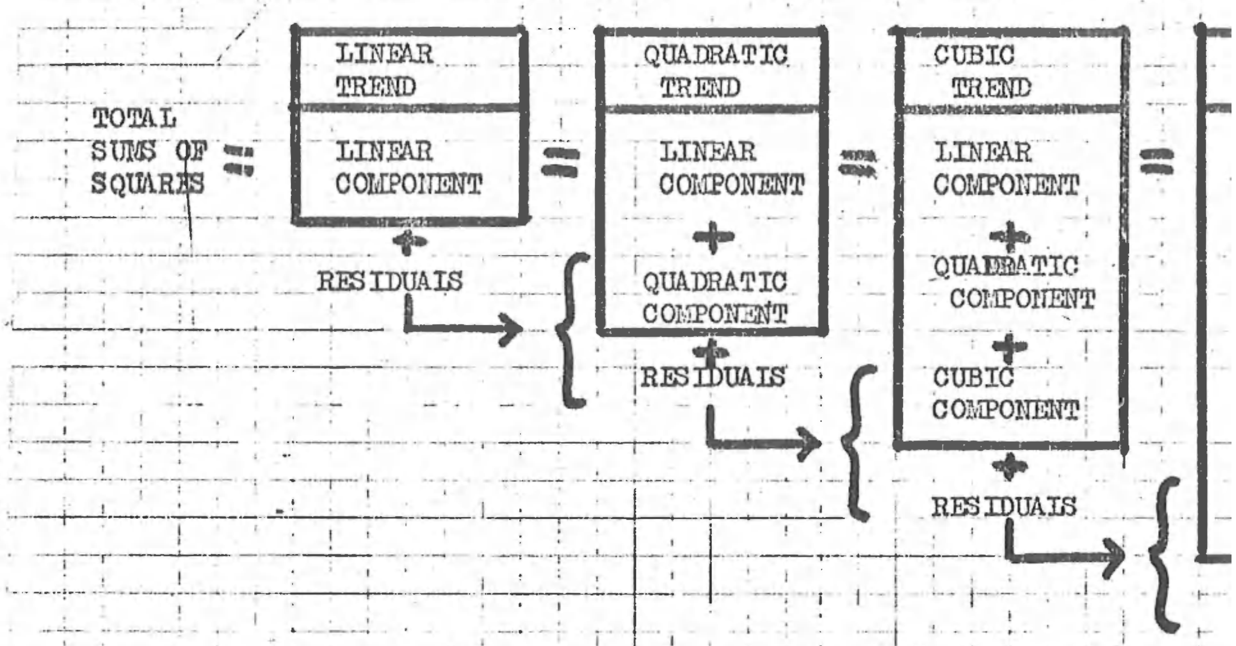
3. Gittins,R. "Trend Surface Analysis of Ecological Data." Journal of Ecology. 56 (3) 1968.

4. King, L.C. "Trend Surface Analysis of Central Pennine Erosion Surfaces." Transactions of the Institute of British Geographers. 47 1969, pp. 47 - 61.

5. Merrian, D.F. and Sneath, P. "Quantitative Comparison of Contour Maps." Journal of Geophysical Research. 71 (4) 1966, pp. 1105 - 1115.

6. Norton-Griffiths, M. "The Number and Distribution of Large Mammals in Ruaha National Park, Tanzania." East African Wildlife Journal. Forthcoming.

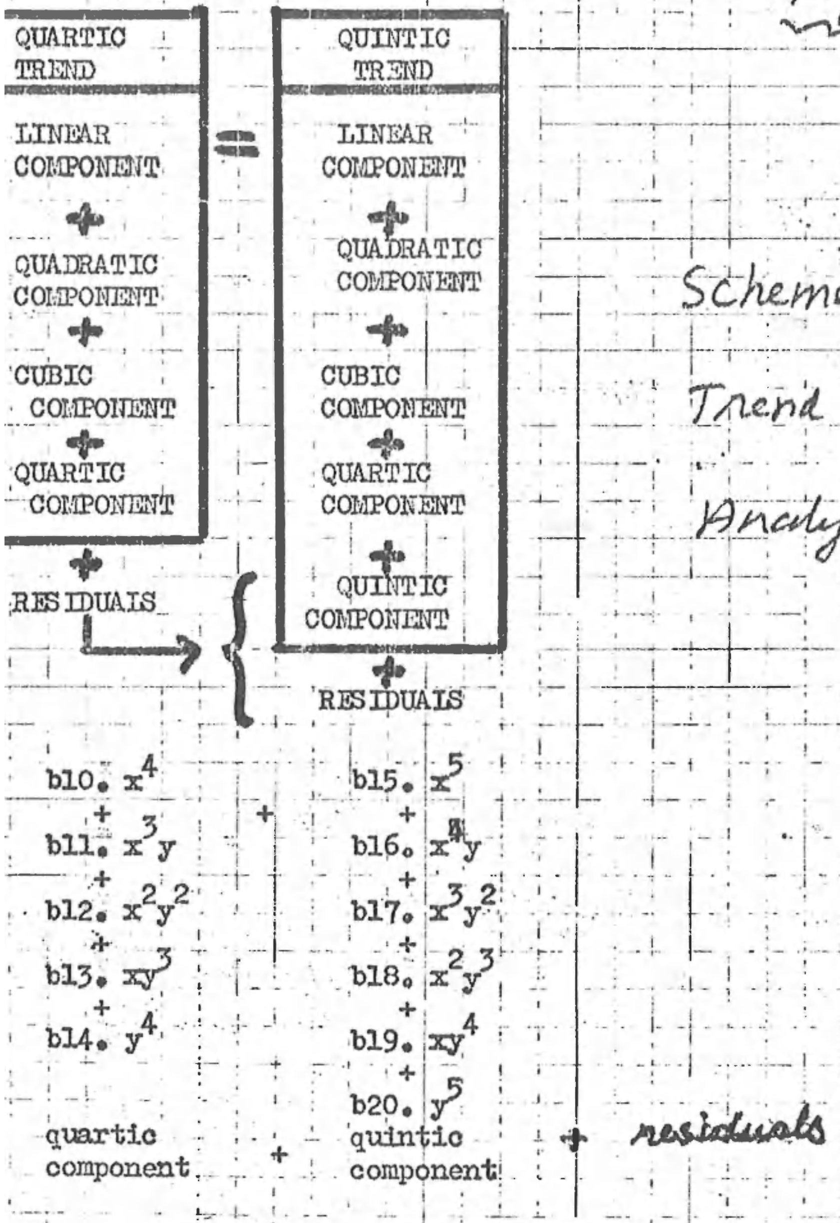
7. Norton-Griffiths, M., Herlocker, D. and Pennycuick, L. "The Patterns of Rainfall and Climate in the Serengeti Ecosystem." East African Wildlife Journal. Forthcoming.



CONSTANT + $b1.\ X + b2.\ y$ + $b3.x^2 + b4.\ xy + b5.\ y^2$ + $b6.\ x^3 + b7.\ x^2y + b8.\ xy^2 + b9.\ y^3$ +

CONSTANT + linear component + quadratic component + cubic component +

Fig 4 IDS/TP8



Scheme of Trend Surface Analysis

ICL Fig. 2

IDS/TP 8

International Computers (East Africa) Limited

80 COL. DATA SHEET

JOB: Data for MX23

DATE:

PAGE No. OF

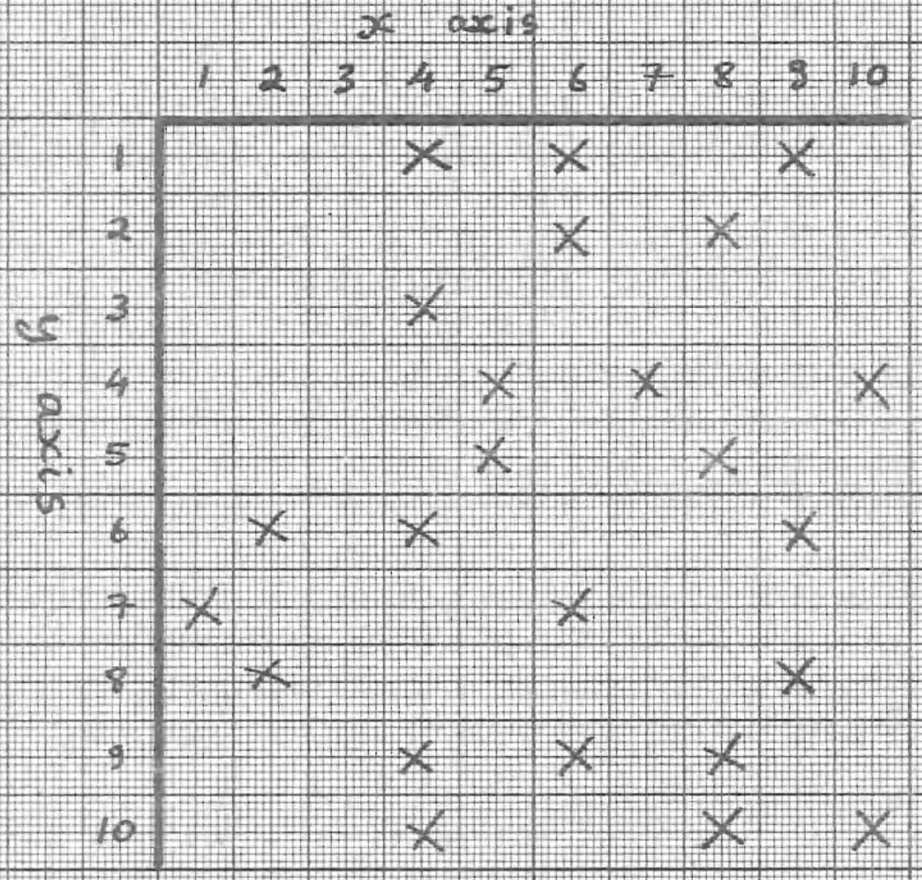
```
                                                   sets of x-, y- and z
2528                                                                    Card 1
RAINFALL IN FEBUARY 1973 SERENGETI (CM)          x y z  x y z           Card 2
16 7 8210 7 121314106 16 91631522 311015 7021121251610 931216 931311125
 114  513 7  82017  7 7 9 5812 9 641619 621619 48 7161251216125 311 70  data    Series 1
1422  5 413  8 815  9 4 9 5516201251521 751118 67-1                     cards
 9 8                                                                    card 1
NACK PLOT 7 (ONE METER GRID)                                            card 2
 1 8 14 9 7-26 1 3  0 4 8 15 8 1 -1 2 4 -2 6 7 14 1 7  3 2 8 13 9 4 12  data    Series 2
-1                                                                      cards
-1                                                                      LAST CARD
```

this indicates end of a series of data

end of series of data

last card

Fig 3



X = location of a sample point