# 9

# The spatial analysis of site phosphate data

C. E. Buck*

W. G. Cavanagh†

C. D. Litton‡

## 9.1 Introduction

During the last fifteen years field survey has become an increasingly prevalent method of investigating the archaeology of regions varying in size from individual sites to complete landscapes. This increase has occurred for many reasons with perhaps the most apparent being:

1. the theoretical developments within archaeology, in particular the stress on adequate sampling techniques;
2. considerations of conservation, with the surface collection of artefacts, soil etc being less destructive than excavation;
3. limited finance, as field survey projects are in general cheaper than excavation projects.

Although initially concerned with the recording of surface features and collection of surface artefacts, field survey has now grown to encompass techniques such as aerial photography, soil resistivity, magnetometry, soil magnetic susceptibility and soil phosphate analysis.

Our attention is focussed on the last, soil phosphate, but in combination with results using other methods. Arrhenius 1938 has summarised work in Sweden to show that enhanced soil phosphate content is often coincident with sites of known archaeological activity. It is decomposition of organic matter in the soil which causes enhanced phosphate; as such, evidence of its presence can complement the information obtained

---

* Department of Mathematics,
University of Nottingham
Nottingham NG7 2RD

† Department of Archaeology,
University of Nottingham
Nottingham NG7 2RD

‡ Department of Mathematics,
University of Nottingham
Nottingham NG7 2RD

from other techniques. In comparison with geophysical techniques the value of soil phosphate analysis in field survey has been relatively slowly realised, now that its complementary nature is appreciated however it is becoming increasingly routine.

Four levels of discrimination have been applied to the use of archaeological soil phosphate analysis:

1. *prospection*—the location of sites within a wide area in which archaeological activity is suspected but not previously pin-pointed;
2. *site survey*—the investigation of the extent of archaeological sites already known or suspected to exist;
3. *location of features*—the surveying of known archaeological sites to ascertain the location within them of phosphate rich features (usually as part of excavation);
4. *investigation of features*—phosphate analysis of soil from features can reveal the position of organic remains otherwise undetected (for example the position of a body in a grave).

In this paper we demonstrate a method of statistical analysis developed for data from site survey ((ii) above), although application to other types of data is envisaged in the future.

## 9.2 Display and analysis of phosphate data

de G. Sieveking *et al.* 1973 reported the development and successful use of a rapid, portable, quantitative method for soil phosphate analysis. However selection of a suitable technique for the display, analysis and hence interpretation of the data has proved problematic. A wide range of different methods have been used for displaying the data for example: proportional circles (Dietz 1957), white, half shaded and solid shaded circles (Craddock 1980), varying density of cross hatching (Pare & Nebelsick 1981) and contour maps (Hamond 1983). In general these types of plots are subjective, with the areas of high or low concentration being seen as more or less important depending upon the plotting levels selected. This means that reliable interpretation is extremely difficult and requires much expertise.

With the increasingly large number of sets of phosphate data from archaeological site survey, it seems necessary to develop a method to allow the location of zones of high and low values within the survey area (allowing for a typical coarse survey grid and high noise level within the data) in a reproduceable and objective manner. This can be achieved using the Bayesian change-point method described below. The method is designed for use with gridded data of the sort commonly collected in site field survey and considers each point in turn in row and column transects.

## 9.3 Change-point analysis for a transect

In general terms, change-point analysis is a statistical method used to determine where, if anywhere, in a sequence of observations, a change occurs in the value of some parameter (or parameters) of the underlying statistical model. Details of the maximum likelihood approach may be found in Hinkley 1971; Pettitt 1979 discusses the nonparametric approach; the Bayesian approach is developed in Smith 1975 and Booth & Smith 1982. Cavanagh *et al.* 1985, Cavanagh *et al.* 1988 give examples of its application to some archaeological situations.

We suppose that the ln(phosphate concentration) values, denoted by $y_1, y_2, \dots, y_n$, taken at regular intervals across a transect are realisations of some underlying random process such that in the absence of any change in the structure of the process, the joint density of $y_1, \dots, y_n$ would have the form

$$p(y_1, \dots, y_n|) = \prod_{i=1}^{n} p(y_i|),$$

where is a vector of parameters.

The process is said to have a change-point at $r$ $(1 \leq r < n)$ if there exists ${}_1$ and ${}_2({}_1 \neq {}_2)$ such that

$$p(y_1, \dots, y_n|r, {}_1, {}_2) = \prod_{i=1}^{r} p(y_i|{}_1) \prod_{i=r+1}^{n} p(y_i|{}_2).$$

The process is said to have change-points at $r$ and $s$ $(1 \leq r < s < n)$ if there exists ${}_3({}_3 \neq {}_2)$ such that

$$p(y_1, \dots, y_n|r, s, {}_1, {}_2, {}_3) = \prod_{i=1}^{r} p(y_i|{}_1) \prod_{i=r+1}^{s} p(y_i|{}_2) \prod_{i=s+1}^{n} p(y_i|{}_3).$$

If ${}_1 = {}_3$, the structure of the process before $r$ is the same as that after $s$ and we shall assume this to be the case in this paper.

Let $M_{rs}$ denote the model that assumes changes at $r$ and $s$ $(1 \leq r < s < n)$. Let $M_{rn}$ denote the model with one change at $r$ $(1 \leq r < n)$ and let $M_{nn}$ denote the model with no changes. For a particular area ${}_1$ and ${}_2$ will be unknown, but we may express our prior information about them by $p({}_1, {}_2)$. Then we have

$$p(y_1, \dots, y_n|M_{rs}) = \int\int p(y_1, \dots, y_n|r, s, {}_1, {}_2)p({}_1, {}_2)d_{12} \qquad (9.1)$$

and for the model with no change

$$p(y_1, \dots, y_n|M_{nn}) = \int\int p(y_1, \dots, y_n|)d. \qquad (9.2)$$

Given the observations $y_1, \dots, y_n$ inferences about the change-point(s) are equivalent to inferences about the possible models $M_{rs}$. Using Bayes theorem the posterior probability of model $M_{rs}$ is given by

$$p(M_{rs}|y_1, \dots, y_n) \propto p(y_1, \dots, y_n|M_{rs})p(M_{rs}), \qquad (9.3)$$

where $p(M_{rs})$ denotes the prior probability of model $M_{rs}$.

In the context of soil phosphate analysis, Cavanagh *et al.* 1988 have shown that it is not unreasonable to assume that the ln(phosphate concentration) both from zones of high and low phosphate concentration have a normal distribution, but with different means. Furthermore, it also seems not too unrealistic to assume that the variances for the two types of zones are equal. In other words, if $y_i$ is a background value (low level) then $y_i \sim N(\mu_1, \tau^{-1})$, whereas if it is from the high level zone $y_i$ $N(\mu_2, \tau^{-1})$ where $\tau$ is the precision (the inverse of the variance). Thus for the two change model $M_{rs}$ $(1 \leq r < s < n)$

$$p(y_1, \ldots, y_n | r, s, \mu_1, \mu_2, \tau) = \left(\frac{\tau}{2\pi}\right)^{n/2} \exp\left\{-\frac{\tau}{2}\left[\sum_{i=1}^{r}(y_i - \mu_1)^2 + \sum_{i=r+1}^{s}(y_i - \mu_2)^2 + \sum_{i=s+1}^{n}(y_i - \mu_1)^2\right]\right.$$

Similar but slightly simpler expressions will hold for the one change and no change models.

Let $p(i)$ $i = 0, 1, 2$ be the prior probability of $i$ changes in a transect. Let $p(r|1)$ and $p(r, s|2)$ be the prior probabilities of models $M_{rn}$ and $M_{rs}$ conditional on 1 and 2 changes respectively. Then, using (1), (2) and (3) and a normal-gamma prior for $\mu_1$, $\mu_2$ and $\tau$, it is possible to calculate the posterior probabilities, $p(M_{rs}|y_1, \ldots, y_n)$, of the various models. (The algebra involved is straightforward but the resulting expressions are too complicated to be reproduced here. We refer the reader to Broemeling 1985, Chapter 7 for the details.) From these posterior probabilities it is possible to calculate the posterior probabilities of the number of changes and hence to make inferences about the number of changes present, if any, and their location.

Let the posterior probability of $i$ changes be denoted by $q(i)$ $i = 0, 1, 2$. Then

$$q(0) = p(M_{nn}|y_1, \ldots, y_n);$$

$$q(1) = \sum_{r=1}^{n-1} p(M_{rn}|y_1, \ldots, y_n);$$

and

$$q(2) = \sum_{1 \leq r < s < n} p(M_{rs}|y_1, \ldots, y_n).$$

For instance, if $q(0)$ is greater than the sum of $q(1)$ and $q(2)$, then we infer that there are no changes present in the transect in which case all the transect could be either at the high level or at the low level. On the other hand if $q(0)$ is less than the sum of $q(1)$ and $q(2)$, then at least one change is more likely than no change. In this case the number of changes is indicated by the larger of $q(1)$ and $q(2)$. In the two change case the positions of the changes may be inferred from the mode of $p(M_{rs}|y_1, \ldots, y_n)$. A similar argument may be applied to the one change case.

In a practical situation there remains the problem of how to specify the prior densities and probabilities. Possible methods include the use of values from similar areas previously analysed, the use of samples taken in the general vicinity of the area under consideration, or the use of a vague prior to represent little or no prior knowledge. The latter is the approach that we will adopt, although it does entail one further difficulty. As Booth & Smith 1982 point out, by comparing model $M_{nn}$ (no change) with the other models we are dealing with models of different dimensionality. To overcome this, they suggest that the posterior probability of no change should be multiplied by $(\frac{3}{2})^{\frac{1}{2}}$ and that the posterior probabilities of no change, one and two changes be renormalised. The effect of this is to make the detection of a change (or changes) more difficult.

## 9.4 Data collected over a regular grid

In the previous section we have described how to detect changes along a transect. If we now consider data collected over an $M$ by $N$ grid, we analyse each row of $N$ data points separately. Thus in each row we can make inferences about the number and position of any changes. This analysis is now repeated for each of the $N$ columns of the grid.

Using the row analysis and using thresholds in the posterior probabilities we have partitioned the grid into three categories, namely, high phosphate level, low phosphate level or undetermined. Likewise when using the columns. However it is unlikely that the partitions will be the same and we suggest that the information from both analyses may be combined in either of the following two ways.

**The intersection principle** If a position in the grid is found to be at an high level by both methods then the position should be classified as high. Otherwise it is classified as low.

**The union principle** Any position considered to be high by *either* the row analysis or column analysis is classified to be at the high level, the remaining positions have a low level.

Having found, by either the intersection or union principle, zones of high phosphate, it is then possible to reassess the zones by an iterative procedure. To do so let us focus attention on the zones formed by using the intersection principle. We re-analyse each row using the same change-point method as before except that we use as our prior information for $\mu_1$, $\mu_2$ and $\tau$, values that can be determined from the site data but omitting the data from the row under study. This is carried out separately for each row and then repeated for the columns. Using the intersection principle a new partition of the region into zones is found. This is repeated until there is no change in the zones. A similar method could be applied to zones found by the union principle.

## 9.5 Example

We illustrate our methodology with an example from the 1987 season of the Laconia Survey in Greece. The site chosen for the analysis has current reference number LS 165 and, from the pottery found, is judged to be of the Roman period. Soil samples were taken at 10m intervals over a 16 by 16 grid. The raw phosphate concentration readings in mgP/100g of soil are given in Fig. 9.1. (For various reasons that need not concern us here data were not available at nine of the sampling positions.)

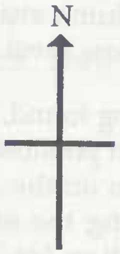
For our initial row and column analyses, we have taken

$$\begin{aligned} p(0) &= 0.5, \quad p(1) = p(2) = 0.25; \\ p(r|1) &= (n-1)^{-1} \text{ for } 1 \le r < n; \\ \text{and } \; p(r,s|2) &= 2(n-1)^{-1}(n-2)^{-1} \text{ for } 1 \le r < s < n, \end{aligned}$$

where $n$ is the number of actual observations in the row or column under consideration. For the parameters $\mu_1$, $\mu_2$ and $\tau$ we have used a vague prior as described earlier.

In Figs. 9.2(a) and (b) are displayed the zones of high or low phosphate concentration found by using the row and column analyses respectively. In Figs. 9.2(c) and (d) are

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 77 | 57 | 55 | 31 | 37 | 45 | 59 | 64 | 64 | 55 | 73 | 32 | 17 | 62 | 108 | 121 |
| 59 | 55 | 34 | 45 | 41 | 47 | 38 | 28 | 45 | 53 | 33 | 48 | 52 | 80 | 101 | 112 |
| 45 | 66 | 62 | 34 | 38 | 48 | 32 | 44 | 62 | 80 | 60 | 27 | 60 | 50 | 75 | 108 |
| 55 | 40 | 41 | 66 | 36 | 68 | 55 | 45 | 21 | 80 | 66 | 88 | 91 | 88 | 83 | 91 |
| 59 | 57 | 80 | 71 | 19 | 80 | 60 | 60 | 60 | 62 | + | + | 166 | 77 | + | 68 |
| 60 | 68 | 75 | 85 | 57 | 44 | 30 | 62 | 38 | 91 | + | + | 68 | 77 | + | 59 |
| 48 | 73 | 101 | 80 | 47 | 64 | 41 | 34 | 47 | 71 | 62 | 116 | 60 | 73 | 52 | 294 |
| 68 | 80 | 50 | 121 | 131 | 64 | 59 | 47 | 77 | 68 | 143 | 66 | 32 | 50 | 55 | 50 |
| 71 | 71 | 71 | 91 | 80 | 68 | 57 | 75 | 73 | 77 | 60 | 34 | 47 | 50 | 50 | 101 |
| 57 | 125 | 91 | 136 | 83 | 68 | 71 | 83 | 62 | 104 | 62 | 62 | 45 | 59 | 41 | 27 |
| 60 | 83 | 94 | 108 | 80 | 88 | 66 | 71 | 27 | 75 | 80 | 77 | 34 | 57 | 30 | 71 |
| 55 | 66 | 94 | + | 88 | 116 | 83 | 77 | 44 | 41 | 59 | 41 | 57 | 55 | 47 | 48 |
| 53 | 77 | 91 | 108 | 73 | 108 | 85 | 83 | 53 | 33 | 75 | 23 | 60 | 57 | 47 | 36 |
| 57 | 71 | 75 | 80 | 73 | 85 | 85 | 73 | 53 | 131 | 57 | 38 | 64 | 38 | 55 | 71 |
| 62 | 47 | 68 | 80 | 97 | 91 | 77 | 77 | 52 | 41 | 27 | 68 | 68 | 71 | 75 | 66 |
| 64 | 55 | 59 | 73 | 62 | 73 | 83 | 59 | 36 | 37 | 57 | 68 | + | + | 108 | 83 |



Raw phosphate concentration readings in mg P/100 g of soil taken at 10 m intervals from site LS 165 of the Laconia Survey, Greece.
Missing data values are indicated by + .

Figure 9.1

the results of the union and intersection rules applied to the row and column results. In Figs. 9.2(e) and (f) are the results of the union and intersection rules after iteration has taken place and convergence has been obtained.

## 9.6 Discussion

Study of Figs. 9.2(a) to (f) shows that the analysis has proved successful in supplying an objective indication of the extent of the area of high phosphate concentration. We would argue that the result illustrated in Fig. 9.2(f) best represents this, and that the results from the union of rows and columns (see Fig. 9.2(e)) include an ambiguous lower level, neither 'on-site' nor 'off-site', which might be identified with the halo effect recognised elsewhere on the evidence of sherd data (Bintliff & Snodgrass 1985).

The question that arises though, is how best to interpret these plots? Phosphate data arising from field survey of this type should, we feel, not be interpreted in isolation. In this context local topography, tile and sherd counts, soil magnetic susceptibility data and geophysical results should be combined to produce a complementary definition of the site. Bearing this in mind we shall attempt crudely to put the results from LS 165 into context.

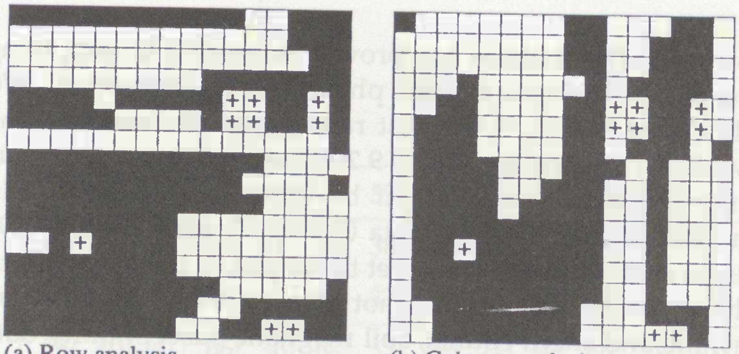
The survey team interpreted the site as that of a small Roman farmhouse or villa; with the main density of sherd and tile covering approximately 100 square metres. Returning to Fig. 9.2(f), we believe that the area of high phosphate (approximately 5000 square metres) in the south west part of the grid is associated with the archaeological activity. The less well defined area of high phosphate concentration in the north east probably arises from modern activity associated with a water source still in use today.

Attempting to interpret this evidence by comparison with excavated sites requires that we turn to evidence from the Western Empire as no small rural Roman villas have been excavated in Greece. In the Western Empire a modest villa might cover some 100 square metres, out-buildings might increase that to 700 square metres, and courtyards, middens and spaces between buildings give an overall extent of approximately 2000 square metres. Large villa complexes, of course, can encompass a considerably greater area. Nevertheless high phosphate concentration extends over approximately twice the area we might crudely expect for a site of the type identified by the survey team. It is possible that this spread is due to post-depositional processes of dispersion.

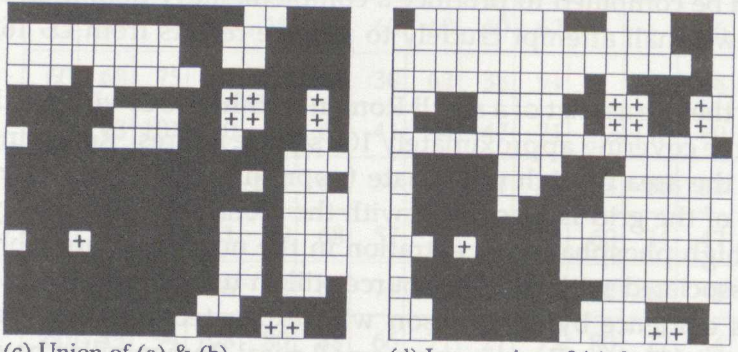
Although this discussion may seem rather premature, its purpose is to underline firstly, the need for the sort of context and confirmatory analysis as outlined above, and secondly, the need to make our archaeological definitions ever more explicit. We wish to stress however, that the pottery finds seem to indicate the minimum definition of a site; a definition which ignores those organic remains which have decayed and are not visible on the surface, but are preserved as an increase in soil phosphate concentration.

## 9.7 Conclusion

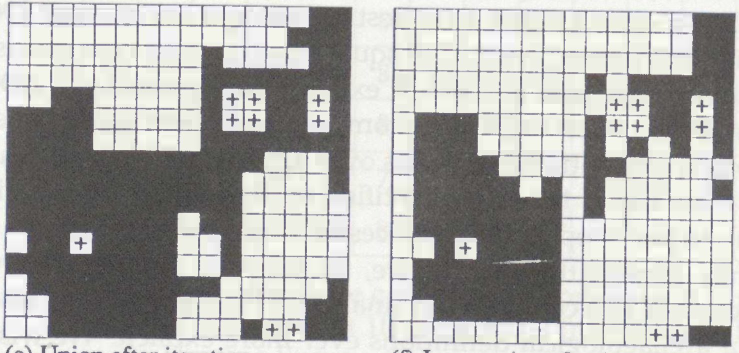
The use of Bayesian change-point analysis has proved a successful method for objectively distinguishing high and low zones of phosphate concentration for data collected as part of site field survey. It now remains to integrate the results with those from other site surveys and even excavation, and to study the effectiveness of applying change-point analysis to other sets of field survey data. Furthermore we would like



(a) Row analysis (b) Column analysis



(c) Union of (a) & (b) (d) Intersection of (a) & (b)



(e) Union after iteration (f) Intersection after iteration

KEY

■ High phosphate level

□ Low phosphate level

⊞ Missing data

Figure 9.2

to investigate the possibilities of applying other image processing techniques which do not ignore the local correlation structure to these data sets.

## Acknowledgements

This work has been carried out under the auspices of a UK Science and Engineering Research Council project grant (reference : GB/E/30263).

## References

ARRHENIUS, O. 1938. "Den Gotlandska Akerjordens fosfathalt", *Sveriges Geologiska Undersokning*, 32 (2): 1–15.

BINTLIFF, J. L. & A. M. SNODGRASS 1985. "The Cambridge/Bradford Boeotian expedition: the first four years", *Journal of Field Archaeology*, 12: 123–161.

BOOTH, N. B. & A. F. M. SMITH 1982. "A Bayesian approach to retrospective identification of change-points", *Journal of Econometrics*, 19 (1): 7–22.

BROEMELING, L. D. 1985. *Bayesian Analysis of Linear Models*. Dekker, New York.

CAVANAGH, W. G., S. HIRST, & C. D. LITTON 1988. "Soil phosphate, site boundaries and change-point analysis", *Journal of Field Archaeology*, p. (to appear).

CAVANAGH, W. G., R. R. LAXTON, & C. D. LITTON 1985. "An application of change-point analysis to the shape of prehistoric corbelled domes; I. The maximum likelihood method". *in* Voorrips, A. & Loving, S. H., (eds.), *P.A.C.T. 11, The Pattern of the Past*, pp. 191–203. Council of Europe, Stasbourg.

CRADDOCK, P. T. 1980. "The soil phosphate survey at the Newark Road subsite, Fengate". *in* Pryor, F., (ed.), *Excavation at Fengate, Peterborough, England: the third report*, Archaeology Monograph 6, pp. 213–217. Royal Ontario Museum.

DE G. SIEVEKING, G., I. H. LONGWORTH, M. J. HUGHES, A. J. CLARK, & A. MILLETT 1973. "A new survey of Grime's Graves - first report", *Proceedings of the Prehistoric Society*, 39: 182–218.

DIETZ, E. F. 1957. "Phosphorus accumulation in soil of an Indian habitation site", *American Antiquity*, 22 (4): 405–409.

HAMOND, F. W. 1983. "Phosphate analysis of archaeological sediments". *in* Reeves-Smyth, T. & Hamond, F. W., (eds.), *Landscape Archaeology in Ireland*, British Series 116, pp. 47–80. British Archaeological Reports, Oxford.

HINKLEY, D. V. 1971. "Inference about the change-point from cumulative sum tests", *Biometrika*, 58 (3): 509–523.

PARE, C. F. E. & L. D. NEBELSICK 1981. "The phosphate survey", *Proceedings of the Society of Antiquaries of Scotland*, 111: 146–147.

PETTITT, A. N. 1979. "A non-parametric approach to the change-point problem", *Applied Statistics*, 28: 126–135.

SMITH, A. F. M. 1975. "A Bayesian approach to inference about a change-point in a sequence of random variables", *Biometrika*, 62 (2): 407–16.