J. Hyg., Camb. (1974), **73**, 15
Printed in Great Britain

On the use of contour maps in the analysis of spread of communicable disease*

By MICHAEL SPLAINE and ALAN P. LINTOTT
Zambia Operational Research Group, P.O. Box 172, *Kitwe, Zambia*
and JUAN J. ANGULO
Computer Center, Emory University, Atlanta, Ga., 30322, *USA*

(*Received* 17 *October* 1973)

SUMMARY

The co-ordinates of the dwellings where cases of variola minor (alastrim) occurred during a small epidemic were used in a worked example of contour mapping of disease spread. The contoured variable was the date of onset, relative to an arbitrary base date, of the case introducing the disease into each of twenty-two households. Three contour maps prepared with slightly different computer programmes or dates exhibited similar concentric loops whose centres were close to the first infected household. The average rate of spread of the disease was estimated by regression of the number of days to onset of the first case in the household on the average distance from an arbitrary origin to the relevant contour line. The calculated average rate of spread was 1·22 metres per day. An additional map was contoured using the cumulative number of cases as the contoured variable, relative to the onset of the example epidemic.

INTRODUCTION

Contour maps have long been used for pictorial representation of the relations between geographical co-ordinates of given places and the corresponding values of a dependent variable. Application of contour maps to metereological, geological, topographical and geographical topics are well known. Fairly recently, contour maps have been used in Ecology, for instance, for studying geographical variation of characteristics of animal and plant species (Adams, 1970; Kiester, 1971). More rarely, contour mappings of disease have dealt with geographical locations of endemic foci (Hopps, 1969). It seems that the dynamics (time and space distributions) of epidemic spread of diseases transmitted from person to person have not been examined by contour mapping, in spite of the interesting possibilities of this approach. An attempt was made to analyse, by contour mapping, the spread of variola minor during a small epidemic. The results are reported as a worked example, in the hope that they may stimulate application to larger outbreaks with various modes of disease spread.

* Please address reprint requests to Dr Angulo, at his present address: Instituto Adolfo Lutz, C.P. 7027, São Paulo, S.P., Brazil.

https://doi.org/10.1017/S0022172400023780 Published online by Cambridge University Press

Table 1. *Basic data for the contour-map analysis of the spread of variola minor in Vila Guarani*, 1956

Geographic co-ordinates of dwellings with cases* (x cm.)	(y cm.)	Identification number of the household with cases	Number of cases in the household	Cumulative number of cases†	Date of onset of the first case in the household‡
7·5	6·7	1	1	1	18
7·5	6·7	2	4	5	40
7·5	6·7	3	1	6	56
7·6	4·7	4	2	8	59
7·4	9·2	5	3	11	84
8·1	5·5	6	3	14	72
11·4	8·5	7	2	16	109
10·9	3·7	8	1	17	93
17·7	5·9	9	1	18	93
12·5	9·0	10	4	22	111
3·8	4·1	11	2	24	108
8·5	5·2	12	1	25	94
4·9	13·7	13	11	36	105
4·8	4·1	14	6	42	110
0·7	12·7	15	2	44	111
6·8	7·3	16	1	45	112
10·6	4·6	17	1	46	113
5·6	13·9	18	1	47	117
2·5	4·3	19	2	49	124
4·4	7·8	20	3	52	138
4·4	7·8	21	1	53	160
4·4	7·8	22	1	54	178

* Actual measurements on a map at a scale of 1 in 4270.

† Relative to the case introducing the disease into the district.

‡ The date of onset has been converted to the number of days from an arbitrary base of 1 March 1956.

MATERIALS AND METHODS

Epidemic data

The basic data for the application of the example are the published time and space distributions of cases of variola minor (alastrim) during the epidemic occurring, in 1956, in Vila Guarani, a semi-rural school district of the city of São Paulo, state of São Paulo, Brazil (Rodrigues-da-Silva, Rabello & Angulo, 1963; Angulo, Rodrigues-da-Silva & Rabello, 1968). The dates of onset were thoroughly investigated and even the chain of contagion could be reconstructed in detail with an unusually high degree of credibility (Angulo, Rodrigues-da-Silva & Rabello, 1968). The dwellings with one or more cases were located on a map prepared from an aerophotogrammetric plot, after actual visits to these dwellings.

Manipulation of data for computation

The data included in the published epidemic curve, chain of contagion and geographical distribution of dwellings with cases (Rodrigues-da-Silva *et al.* 1963; Angulo *et al.* 1968) were manipulated for application of computer-plotting procedures. For convenience of calculations, the calendar dates were converted to the number of days from an arbitrary base of 1 March 1956. That was the month when variola minor was introduced into the essentially closed community where the example epidemic occurred (Rodrigues-da-Silva *et al.* 1963). The original scale of the aerophotogrammetric plot of the Vila Guarani district, 1 in 2000, was reduced in the published map (Angulo *et al.* 1968) to 1 in 4270. A grid with squares of 1 cm. side was superimposed on the reduced map. The vertical and horizontal co-ordinates of the centre of each dwelling with one or more cases were measured and they appear in Table 1. This table also shows the household identification number, the number and cumulative number of cases occurring in each household and the converted date of onset of the first case in the household, hereinafter called the date of onset.

The fact that, in two instances, three households were living in a single house, although separated by internal partitions, is recognized in the original publication (Angulo *et al.* 1968) as well as in the present paper's Figs. 1 and 3. In these figures, the date of onset appears separately for each household. On the other hand, Fig. 2 presents a single averaged value for the three households in the above two instances (Households 1, 2 and 3 and Households 20, 21 and 22). The date of onset is the value contoured in each of Figs. 1 to 4, while the value contoured in Fig. 5 is the cumulative number of cases (Table 1).

The dependent variable (date of onset) together with the independent variables (the co-ordinates) were used as input to the computer-contouring package used by Splaine, Lintott & Barclay (1970). The IBM Computer Programme 'Numerical Surface Techniques and Contour Map Plotting' (see the Appendix to the present report) was employed in an IBM computer fitted with an IBM 1627 model 2 Plotter. Later, contouring by the same computer, using the Calcomp G.P.C.P. programme package (see Appendix) and a Calcomp Plotter, was tried for comparison purposes. Finally, the cumulative number of cases was used as the input to the Calcomp programme.

RESULTS

Contour mapping

Four contour maps were produced by the computer. Fig. 1 shows the contour map produced by using the contouring package used by Splaine *et al.* (1970). It exhibits three almost concentric loops in the centre but no readily discernible pattern outside this area. Fig. 2 exhibits the contour map produced by the same computer using the Calcomp G.P.C.P. package. This package contours to the specified map boundary regardless of sparseness of data. Also, the dates of onset in Households 1, 2 and 3 are shown as a single averaged value. The same applies to Households 20, 21 and 22. As said before, each of these groups of three households lived in a single house with internal partitions separating the house into three household dwellings. Fig. 3 shows the contours obtained with the same G.P.C.P. package, but with a set of printing instructions somewhat different from that used for producing the map from Fig. 2.

Fig. 5 shows the contours obtained with the Calcomp G.P.C.P. package, when

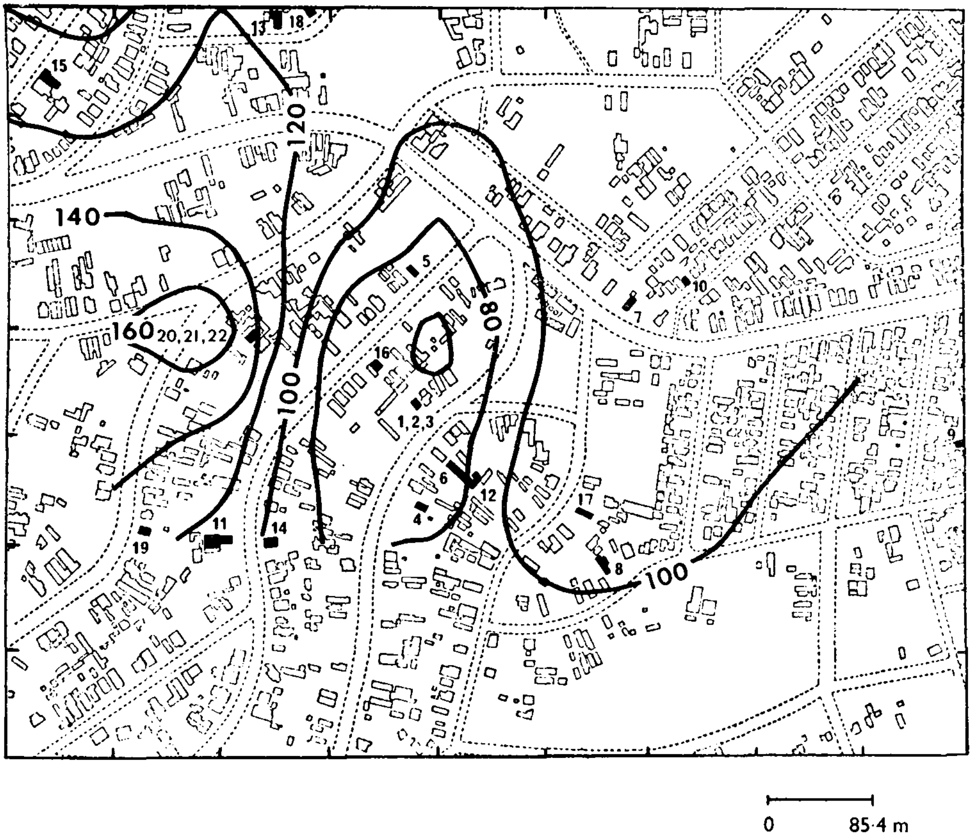

Fig. 1. Contour lines superimposed on an aerophotogrammetric map of Vila Guarani, a school district where an epidemic of variola minor occurred in 1956. The blackened polygons are the dwellings where cases occurred and their numbers correspond to those in Table 1. The contour value of each household with cases is that appearing in the last column of Table 1. The computer programme was the IBM N.S.T.C.M.P. package.

the dependent variable was the cumulative number of cases (Table 1). This number of cases is an indication of the progress or course of the epidemic. The contouring pattern is essentially the same as that provided by the cumulative time intervals to onset of the first case of the epidemic (Figs. 1–4). It should be mentioned that the later plots (those from Figs. 3 and 5, for instance) were generated by increasing the number of neighbouring data points used for the determination of mesh point values, thus increasing the degree of smoothing of the original data.

In a few places on the maps obtained (e.g. Household 8 in Fig. 1), a contour can be seen to lie on the 'wrong' side of the position of the household dwelling. Such occurrences are due to the averaging and smoothing procedures employed in the programmes and in no case do they invalidate the general interpretative picture produced. With only 22 data points available, it is not surprising that procedures which differ in detail should produce localized differences in graphical

https://doi.org/10.1017/S0022172400023780 Published online by Cambridge University Press

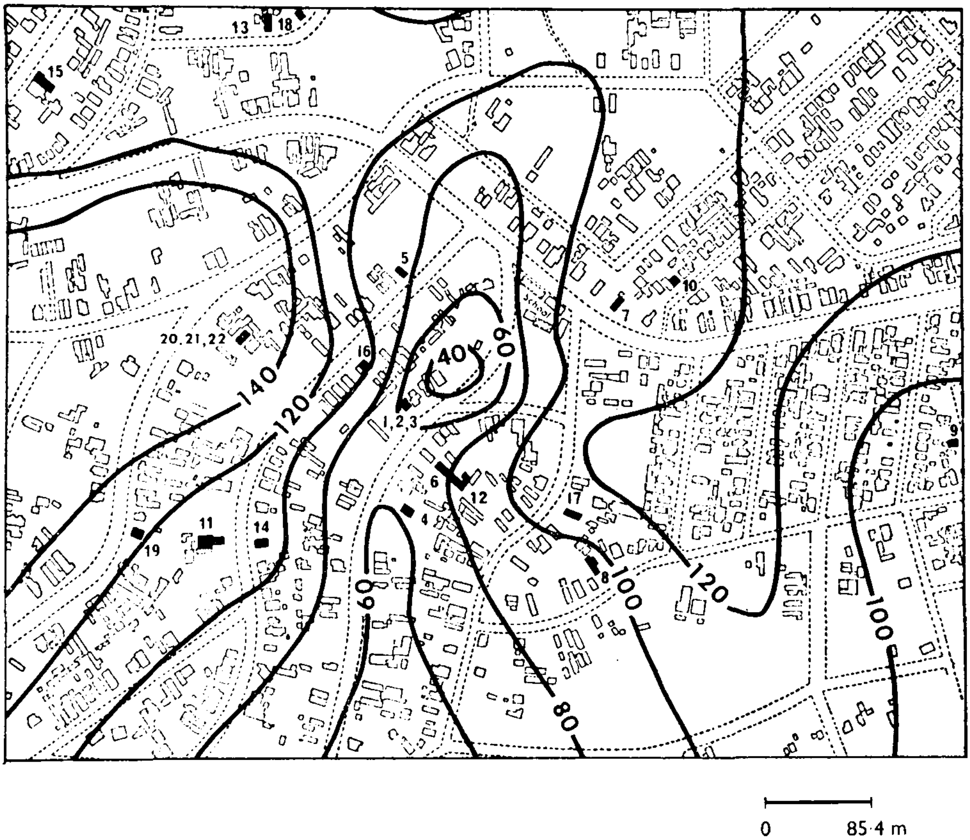

Fig. 2. Same as in Fig. 1, except that the computer programme is the Calcomp G.P.C.P. package. Also, the contour values of Households 1, 2 and 3 were averaged since they lived, separated by internal partitions, in a single house. The same applies to Households 20, 21 and 22.

interpretation. It is not the purpose of this paper to analyse the relative merits of these programmes, but merely to point out that, in general, the maps follow the same over-all pattern. While contour maps display interesting general trends, caution must be exercised in drawing specific conclusions from the detailed configuration of the contours.

Estimation of the rate of disease spread

It seems practicable to estimate an average rate of spread of variola minor from the contour map of say, Fig. 3. On this contour map, an arbitrary origin was chosen, nominally at the centroid of the innermost loop. From this centroid, sixteen radial lines were drawn with uniform angular spacing. The distances (in cm.) from the origin to the contour line were then read off by ruler, along each of the radial lines, provided that the contour value (number of days to onset) continued to increase along each radial line. All such measured distances are shown in Table 2. These measurements may be considered as estimates of the average radius of each particular contour of date of onset, in the case of a closed

https://doi.org/10.1017/S0022172400023780 Published online by Cambridge University Press

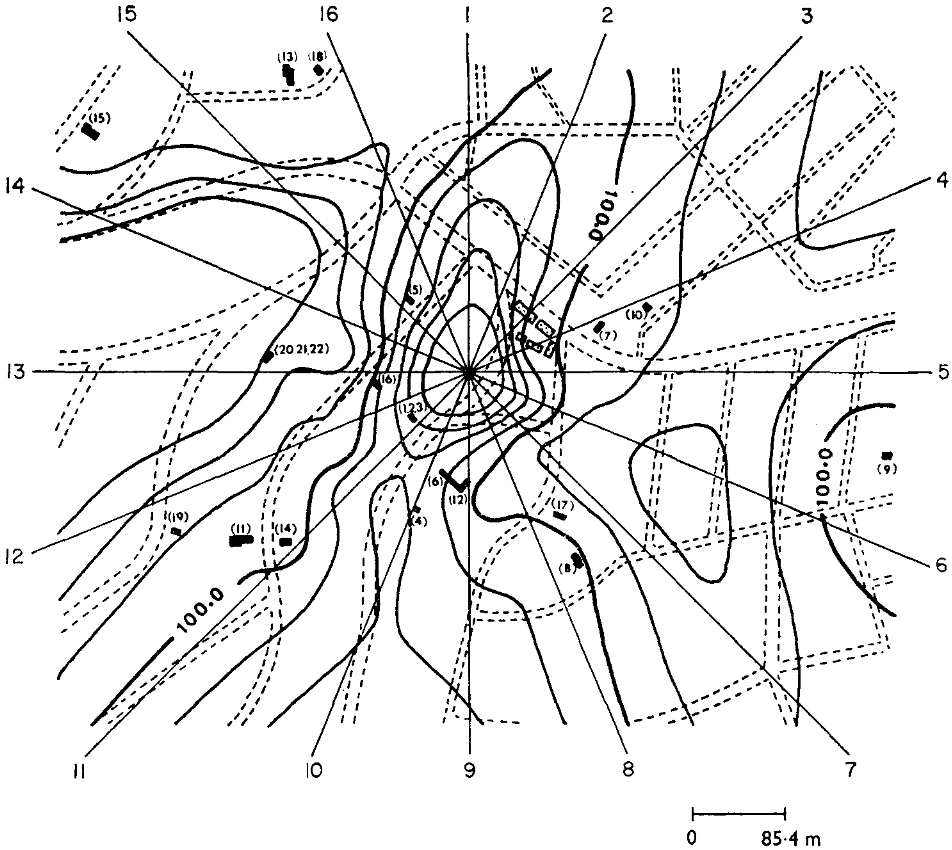

Fig. 3. Same as in Fig. 2, except that the computer programme had different instructions for plotting the data and more detail was obtained. Also, the contour values are not averaged for either Households 1, 2 and 3 or Households 20, 21 and 22. From an arbitrary centre, radial lines have been drawn to estimate the rate of spatial spread of variola minor according to the procedure described in the text.

contour. Average distances (in cm.) were then plotted against the number of days to onset, at intervals of 20 days (Fig. 4). A reasonably straight line was obtained. It was arbitrarily chosen to regard this average radius as the mean distance covered by the disease in the number of days indicated by the numerical value attached to the contour line. The precise definition breaks down in the case of open contours. Nevertheless, it may be assumed that the distances so measured are in some way representative of the rate of spread. Under these circumstances, the reasonably straight line obtained in Fig. 4 suggests a constant linear rate of spread.

Using the data in Table 2, the equation of the regression line for predicting the distance from the origin (Fig. 3), for a given time in days, is

$$\text{distance in centimetres} = 0{\cdot}1874 + 0{\cdot}02861 \times \text{time in days}.$$

Thus, the average rate of spread of variola minor was 0·0286 cm. per day on the already published map (Angulo *et al.* 1968) and on Fig. 3. For a more suitable

https://doi.org/10.1017/S0022172400023780 Published online by Cambridge University Press

Table 2. *Estimation of the average rate of spatial spread of variola minor in Vila Guarani*, 1956, *by distances measured from an arbitrary origin to intersection with contour lines along radial lines**

Radial line no.†	Contour line value 20	40	60	80	100	120	140	160
1	1·35	2·35	3·40	4·15	4·90	—	—	—
2	1·30	1·80	3·05	4·90	—	—	—	—
3	0·95	1·20	1·45	2·05	3·90	7·35	—	—
4	0·80	0·95	1·15	1·35	2·20	4·95	7·75	—
5	0·85	1·05	1·25	1·55	2·10	3·80	—	—
6	0·90	1·15	1·50	1·80	2·10	2·60	4·05	—
7	0·95	1·20	1·40	1·60	1·80	2·30	—	—
8	0·90	1·15	1·30	1·60	1·90	—	—	—
9	1·00	1·25	1·50	1·90	—	—	—	—
10	1·05	1·40	1·95	—	—	—	—	—
11	1·10	1·45	1·95	4·05	4·85	—	—	—
12	1·00	1·35	1·65	2·05	2·50	4·50	7·50	—
13	0·85	1·15	1·40	1·65	1·95	2·30	3·30	4·55
14	0·85	1·10	1·40	1·65	2·00	2·40	2·90	3·85
15	0·90	1·20	1·50	1·95	2·55	3·10	3·75	—
16	1·05	1·40	1·95	2·70	3·50	4·70	—	—
Mean	0·9875	1·322	1·737	2·330	2·788	3·800	4·875	4·200

* Figures in the columns correspond to centimetres actually measured on the published map of Vila Guarani (Angulo, Rodrigues-da-Silva & Rabello, 1968). This map was reduced from an aerophotogrammetric plot. See Fig. 3.

† From Fig. 3. The contour-line values correspond to dates of onset of the first case in the household, after conversion of the calendar date to a number of days from base of 1 March 1956. March was the month when variola minor was introduced into Vila Guarani (Rodrigues-da-Silva, Rabello & Angulo, 1963).

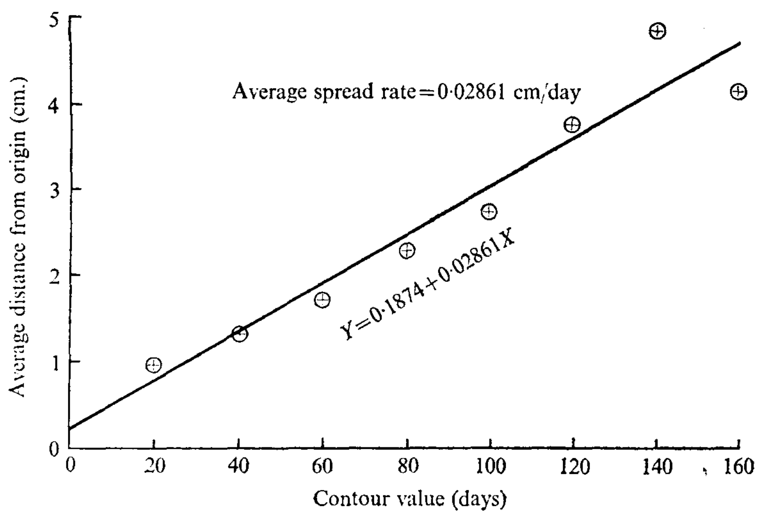

Fig. 4. Regression of the average distance from the arbitrary origin appearing in Fig. 3 on the contour values of the households in the last column of Table 1.

https://doi.org/10.1017/S0022172400023780 Published online by Cambridge University Press

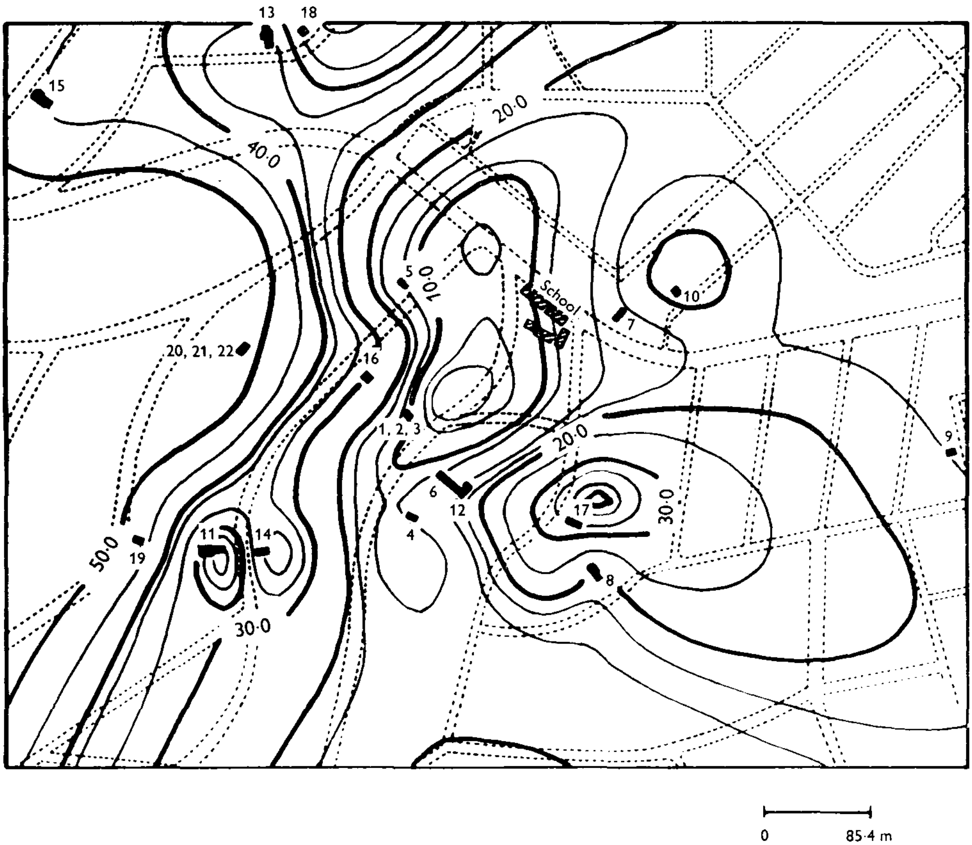

Fig. 5. Same as in Figs. 1 to 4, except that the contoured variable is the cumulative number of cases (Table 1). The computer programme is the Calcomp G.P.C.P. package. The contoured value for the building housing Households 1, 2 and 3 is the cumulative number of cases for Household 3. For the building housing Households 20, 21 and 22, the contoured value is the cumulative number of cases of Household 22.

presentation of these results, they have been corrected according to the scale of the map (1 in 4270). Thus, the estimated average rate of spread should read: 1·22 metres per day.

DISCUSSION

The results obtained by the use of contour mapping are presented only as a worked example, with no claim as to statistical rigor nor as a definitive formulation of the problem of estimation of the rate of spread of a communicable disease. In this regard, the example epidemic only provided small numbers. On the other hand, the procedure devised for estimating the average rate of spread provides an objective measure of this rate, based on a scattered set of observations.

Some loops are in spatial association with attacked dwellings, while no definite loop is found close to other dwellings with cases. This is seemingly so because, when contouring data which have considerable variability about some underlying pattern, one is bound to find sets of closed contours associated with some, at least, of the data points. This behaviour is inherent since all points do not lie on the 'underlying' surface pattern. Thus, no special significance should be attached to loops associated with single data points. Differences in dates of onset between households occupying different parts of the same building are a serious embarrassment to the contouring method of analysis. The latter method, of its very nature, cannot make much sense from such data.

If the number of cases in each household was approximately the same and if the intervals between successive dates of onset were approximately constant, the contouring programme would not find any significant difference between data sets formulated for date of onset on the one hand, and for cumulative number of cases on the other. The data from Vila Guarani conform in some respects with the above hypotheses, so it is not surprising to note general similarities between the respective contour maps. The differences between the maps are merely a measure of the random departures of the individual data items from the hypotheses.

With reservations due to the limitations of the example data and procedures, it may be said that the use of contour mapping for studying the spatial spread of variola minor disclosed interesting general trends. For instance, a fairly uniform spread of the disease from a central point. This perhaps justifies the kind of calculations presented above. However, the detailed shapes of the contours are a function of a mathematical procedure based on only twenty-two observations (households). Therefore, no specific meaning should be attached to any one small region of the contour map, particularly to loops associated with a single data point. With these reservations in mind, it may be pointed out that the centre of the concentric loops, evident in all three first contour maps obtained, is no more than 65 m. distant from the location of the first infected household.

It is known that, at least partially, variola minor spread among the students at the school (Angulo, Rodrigues-da-Silva & Rabello, 1964). Reconstruction of the chain of contagion and inspection of the dates of onset of cases in households yielded a highly suggestive evidence of the role of the district school as a nodal centre for the spread of variola minor in Vila Guarani district (Angulo *et al.* 1968). However, the close proximity of the 'centre' to the location of the school (no more than 100 m.) is believed to be fortuitous. In effect, there is no reason to *suppose that the date of infection of a student at the school could be correlated* with the distance between the infected student's home and the school. It must be concluded that, if the fairly uniform outward spreading pattern shown by the contour maps has any meaning at all, the pattern must largely represent the spread of the disease by contacts other than those at the school. In this regard, the epidemic in Vila Guarani was found to consist of summation of smaller, almost independent outbreaks occurring in sub-areas, the dwellings of the households where the disease was introduced by a single member (Angulo, Rodrigues-da-Silva & Rabello, 1967, 1968).

If the school was the sole means of spread of the disease, a formless jumble of contour lines would be expected. If house-to-house was the sole means of spread, fairly regular concentric contours over the whole map would be expected. For a mixture of the two means of spread, a pretty irregular pattern would be expected, but with some concentric contours present. The patterns obtained for the Vila Guarani district are consistent with this expectancy.

It is felt that contour maps drawn by the procedure used in the present study are interesting from an interpretational point of view for the Vila Guarani district. The value of such maps for epidemic areas with very different patterns of children attending school remains to be investigated. Estimation of the rate of spread through contour mapping suggests that while variola minor was spreading in Vila Guarani, its linear rate of spread was fairly uniform. Although the procedure used to obtain these results seems to be reasonable, the validity of its estimates is unknown.

Meyers (1949) found very variable and slow rates for the spread of measles from the origin to the various other areas of a health-centre district of New York City. In his intuitive study, this worker employed distances from the centre of the area where the epidemic started and the time to reach a peak in the various other areas of the district. Meyers also suggested that the elementary schools were important in the spread of the epidemic. Slow rate of geographic spread of measles was also noted by Stocks (1930).

One of the authors (J.J.A.) gratefully acknowledges the facilities provided by Dr W. Buell Evans.

REFERENCES

ADAMS, R. P. (1970). Contour mapping and differential systematics of geographic variation. *Systematic Zoology* **19**, 385.

ANGULO, J. J., RODRIGUES-DA-SILVA, G. & RABELLO, S. I. (1964). Variola minor in a primary school. *Public Health Reports, Washington* **79**, 355.

ANGULO, J. J., RODRIGUES-DA-SILVA, G. & RABELLO, S. I. (1967). Spread of variola minor in households. *American Journal of Epidemiology* **86**, 479.

ANGULO, J. J., RODRIGUES-DA-SILVA, G. & RABELLO, S. I. (1968). Sociologic factors in the spread of variola minor in a semi-rural school district. *Journal of Hygiene* **66**, 7.

HOPPS, H. C. (1969). Computer-produced distribution maps of disease. *Annals of the New York Academy of Sciences* **161**, 779.

KIESTER, A. R. (1971). Species density of North American amphibians and reptiles. *Systematic Zoology* **20**, 127.

MEYERS, J. (1949). Study of the geographical and time progression of a measles epidemic in the Mott Haven Health Center District, New York City. *American Journal of Public Health* **39**, 1446.

RODRIGUES-DA-SILVA, G., RABELLO, S. I. & ANGULO, J. J. (1963). Epidemic of variola minor in a suburb of São Paulo. *Public Health Reports, Washington* **78**, 165.

SPLAINE, M., LINTOTT, A. P. & BARCLAY, G. P. T. (1970). Pictorial representation of the distribution of the sickle-cell trait by contours drawn by a computer-controlled $X-Y$ plotter. *Annals of Human Genetics* **34**, 51.

STOCKS, P. (1930). The mechanism of a measles epidemic. *Lancet* i, 796.

APPENDIX

The IBM Computer Programme 'Numerical Surface Techniques and Contour Map Plotting'

The first stage in the IBM computer programme estimated the value of the date of onset for every intersection of a square grid within the required boundaries of the published map. For each 1 cm. square enclosed by adjacent grid points, a plane was constructed in the space defined by the three dimensions: longitude, latitude and date of onset. If there were any observations within a 1 cm. square, the plane was forced to pass through the centroid of those observations. The attitude of the plane was then chosen to give the least-squares fit with the nearest observation in each octant outside that square. The relevant value of the date of onset was inversely weighted by distance from the square. If there was no observation within the 1 cm. square, the plane was fitted by least squares to the above eight external observations without being forced through an internal centroid of observations. When the equation to the fitted plane had been obtained, the programme calculated the implied values of the date of onset in each corner of the 1 cm. square of the grid. The procedure was repeated until the whole map area had been dealt with. Thus, except at the periphery, where there were only two values and the corners, where there was only one value, there were four calculated values for each grid intersection point. These values were averaged to obtain a single value for subsequent use.

In a second stage, the programme calculated a path for each required contour within each square in turn. It then issued the necessary instructions to the X–Y plotter for drawing the contour lines. Each contour path was calculated as follows. Any contour line within a square usually intersected with two of the square sides. These intersection points were found by linear interpolation between adjacent corners of the square. The two relevant adjacent 1 cm. squares were then inspected to locate the next nearest points of intersection of the named contour with a grid mesh line. The programme had thus four points to consider. A circle was fitted to the first three points of the set and a second circle was fitted to the last three points of the set. The two circles intersected within the original square and an intermediate curve was constructed within that square. It was a weighted mean of the two circles. The X–Y plotter plotted this intermediate curve. If any other required contours lay within that square, they were dealt with similarly. The whole procedure was then repeated until all grid squares had been dealt with. Contour values were printed at suitable intervals. Towards the edges of the mapped area, contour plotting was inhibited when the number of observations lying outside a given grid square fell below a critical value.

This description of the essential details of the IBM programme should convey some idea of the trouble to which it is necessary to go in order to achieve a sound, graphical interpretation of a number of discrete data points. The alternative programme which was available (Calcomp G.P.C.P.) used a set of procedures which were different in detail, but which led to similar contour maps.

Calcomp G.P.C.P. Contour Generation

The programme initially generates function values over a user-defined rectangular grid using an approximating process that consists of two basic operations:

(*a*) tangent plane or gradient determination; and

(*b*) the extension of this information to generate grid value.

Tangent plane generation for each input point begins with the selection of n (usually $n = 8$, but may be user-controlled) neighbouring data points that are closest to the point in question. The plane must then pass through the value of the function at the control point and be of such orientation as to minimize the sum of the angles that it makes with the vectors or lines to the neighbouring points. These angles are, of course, weighted by a function of distance of the neighbouring point to the control point. In this manner, a tangent plane (or gradient) is evaluated for each input (control) point.

Grid generation, like tangent plane generation, begins by the selection of the n neighbouring data points (now with gradients) closest to the mesh point in question. The gradient or tangent plane information for each point is evaluated at the mesh point location and the evaluation is weighted inversely on the basis of the distance from the mesh point to the various data points. The surface approximation process is complete after this process has been applied to all mesh points.

Before drawing the contours, the G.P.C.P. package refines each grid cell by dividing each cell into a subgrid. A third order interpolation in X and Y defines the contours crossing each grid cell but at the same time preserves the gradient of the function across cell boundaries.