SPATIAL PROCESSES

MODELS & APPLICATIONS

A D Cliff & J K Ord

p Pion Limited, 207 Brondesbury Park, London NW2 5JN

ISBN 0 85086 081 4

Printed in Great Britain by Page Bros (Norwich) Limited

1.6.2 Tests of significance

Whenever $I_{s\text{-}t}$ reduces to any of the measures discussed in earlier sections, the corresponding normal approximations will apply. In general, we need more structure upon the $\{y_{ij}\}$ to be able to make statements about the form of the distribution, even for large n. Nevertheless, normality may often prove a reasonable approximation.

However, when the $\{w_{ij}\}$ and the $\{y_{ij}\}$ are both binary and the number of nonzero values for each is small (specifically, both S_0/n and T_0/n should stay finite as n goes to infinity), then the distribution of $\frac{1}{2}I_{s\text{-}t}$ is approximately Poisson with mean $S_0T_0/2n(n-1)$. The factor of 2 in the denominator is included to avoid double counting of the joins. This result is demonstrated in section 2.4.4.

1.7 Examples

1.7.1 Map pattern analysis

The BW join-count statistic given in equation (1.6) has been evaluated for the various distributions mapped in figures 1.3 and 1.4. Binary weights were used to test for spatial autocorrelation among contiguous areal units. The moments were evaluated under nonfree sampling [equations (1.33)–(1.34)]. The sampling distribution was assumed to be approximately normal, and the quantities, z, shown on the diagrams can be treated as normal deviates. On figure 1.3a, the high degree of spatial clustering of boroughs in standardised mortality ratio (SMR) category 1 is confirmed by the large deficit of BW joins. The map was reduced to two colours by giving a B coding to each borough in the SMR category (or categories) of interest, and to all others a W coding. The range of values for z when $n = 29$ is shown in figures 1.3(b)–1.3(d); a large negative z score is characteristic of strong positive autocorrelation with the BW statistic [figure 1.3(b)], a large positive value with a checkerboard pattern [figure 1.3(d)], and a value close to zero with a random pattern of B and W units [figure 1.3(c)]. Moderate spatial contagion is also detected for figure 1.4(b) (bronchitis deaths in Wales), but not for figure 1.4(a) (tuberculosis deaths). However, for Wales, n is only equal to 13 counties and, as we shall see in section 2.7, this makes the normal approximation to the sampling distribution of BW dubious.

1.7.2 Diffusion on graphs; measles in Cornwall

In a study of measles epidemics in Cornwall, 1966–70, Haggett (1976) reduced the map showing the twenty-seven local authority areas which report measles cases in Cornwall (see figure 1.7) to various binary graphs corresponding as closely as possible to given hypothetical diffusion processes. The graphs tried were

G-1 Local contagion assuming a spread of measles only between contiguous local authorities. Planar graph with 34 joins.

G-2 Wave contagion assuming spread by shortest paths from the Plymouth area, where measles is endemic. The disease is not endemic in any Cornish local authority. Planar graph with 28 joins.

G-3 Regional contagion assuming spread occurs only within two separate regional subsystems, one based on east Cornwall and one based on west Cornwall. Planar subgraphs with 32 joins.

G-4 Urban–rural contagion assuming spread within sets of urban and rural communities created as separate subgraphs. Nonplanar subgraphs with 181 joins.

G-5 Population size assuming spread down the population-size hierarchy from largest to smallest local authority. Nonplanar graph with 26 joins.

G-6 Population density assuming spread down the population-density hierarchy. Nonplanar graph with 26 joins.

G-7 Journey-to-work contagion assuming that (a) these flows provide a surrogate for spatial interaction between areas, and (b) measles spread followed high interaction links. Nonplanar graph with 58 joins.

To assess the relationship of each graph to the spread of measles epidemics, the 222 weekly maps, 1966–70, were translated into outbreak/no outbreak terms. Vertices on each of the graphs were coded *B* for outbreak or *W* for no outbreak; and the *BW* statistic discussed above was

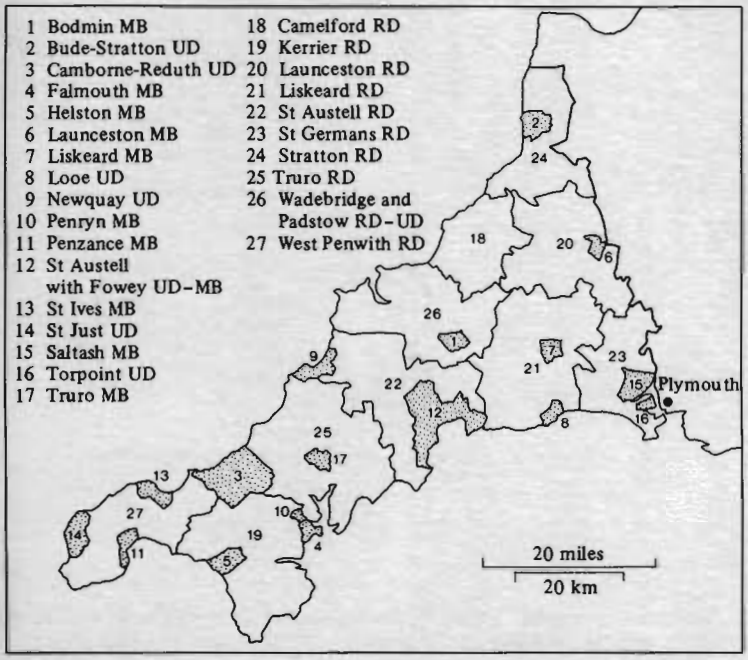


Figure 1.7. Cornish local authorities or General Register Office districts used to report disease data. Urban and metropolitan districts are stippled.

evaluated under nonfree sampling to measure the degree of autocorrelation (contagion) present in the graph. The greater the degree of correspondence between any given graph and the transmission path followed by the diffusion wave, the larger (negative) should become the z-score for BW joins.

Figure 1.8 shows the results for a forty-week epidemic starting in 1966, week 40. The epidemic peak is shown by the vertical pecked lines on the graphs. The population-size/density hierarchies are extremely important during the early build-up phase of the epidemic, whereas after the epidemic

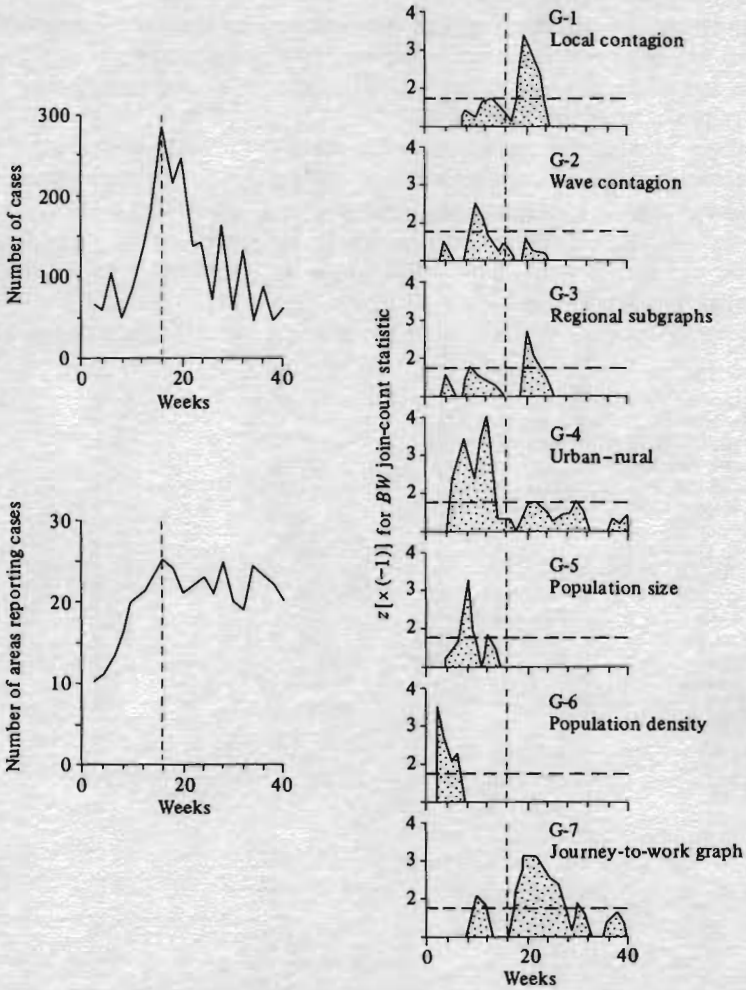


Figure 1.8. Diffusion on graphs. Values of the BW statistic [x (−1) and in standardised score form] for a measles epidemic in Cornwall, 1966 (week 40)–1967 (week 27). For definitions of graphs G-1 to G-7, see text. Vertical pecked lines indicate epidemic peak; horizontal pecked lines correspond with the $\alpha = 0{\cdot}05$ significance level in a one-tailed test of significance. Source: Haggett (1976, page 141).

peak there is a marked switch to the locally contagious spatial spread graphs G-1, G-3, and G-7. The results suggest a simple diffusion model dominated by central place (size) effects in the early stages and localised spatial spread once the epidemic is fully established. This links back to the discussion of figure 1.1 earlier in the chapter.

1.7.3 Space-time interactions; cholera in the metropolis

The weekly death rate from cholera in each of the London boroughs is given in table 1.2. Their locations and identity numbers appear in figure 1.9. Using these data, we tested for spatial autocorrelation amongst contiguous boroughs in each week of the epidemic. The coefficient, I, defined in equation (1.15) was evaluated with $w_{ij} = 1$ if boroughs i and j shared a common boundary, and $w_{ij} = 0$ otherwise. The significance of I was tested using the randomisation assumption [equations (1.37) and (1.39)]. The results are shown in figure 1.10(a). The maps illustrate the spatial pattern of deaths for four weeks of the epidemic. Weeks 8 and 23 are associated with the build-up and fade-out phases of the epidemic, which was at its peak in week 20. Reference back to figure 1.2(b) indicates that the source in week 8 is in the area served by the Southwark and Vauxhall Water Company. That company's area continued to suffer extensively throughout the epidemic, which eventually faded out in the East End.

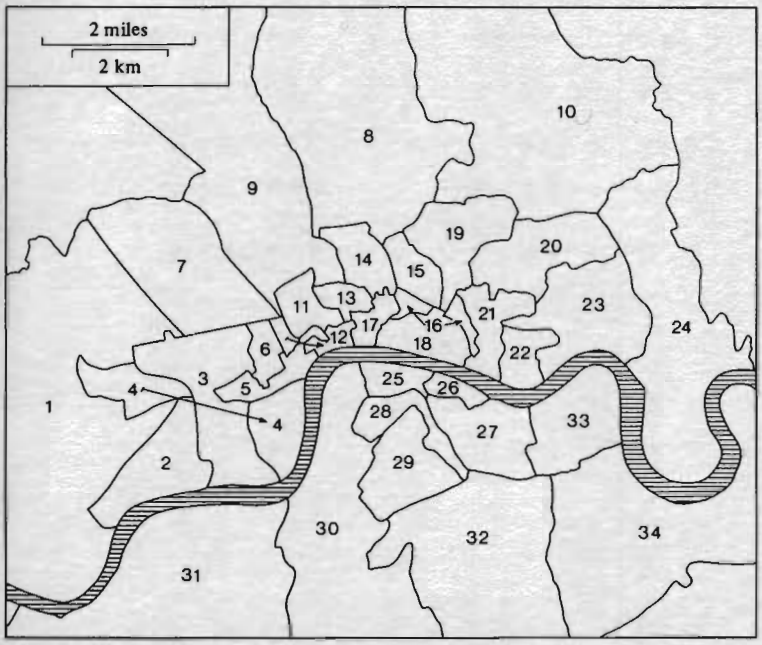


Figure 1.9. Locations and identity numbers of the London boroughs given in table 1.2.

Table 1.2. Weekly deaths from cholera in thirty-four districts of metropolitan London, April 28–November 24, 1849. Source: PP 1850, XXI, page 221.

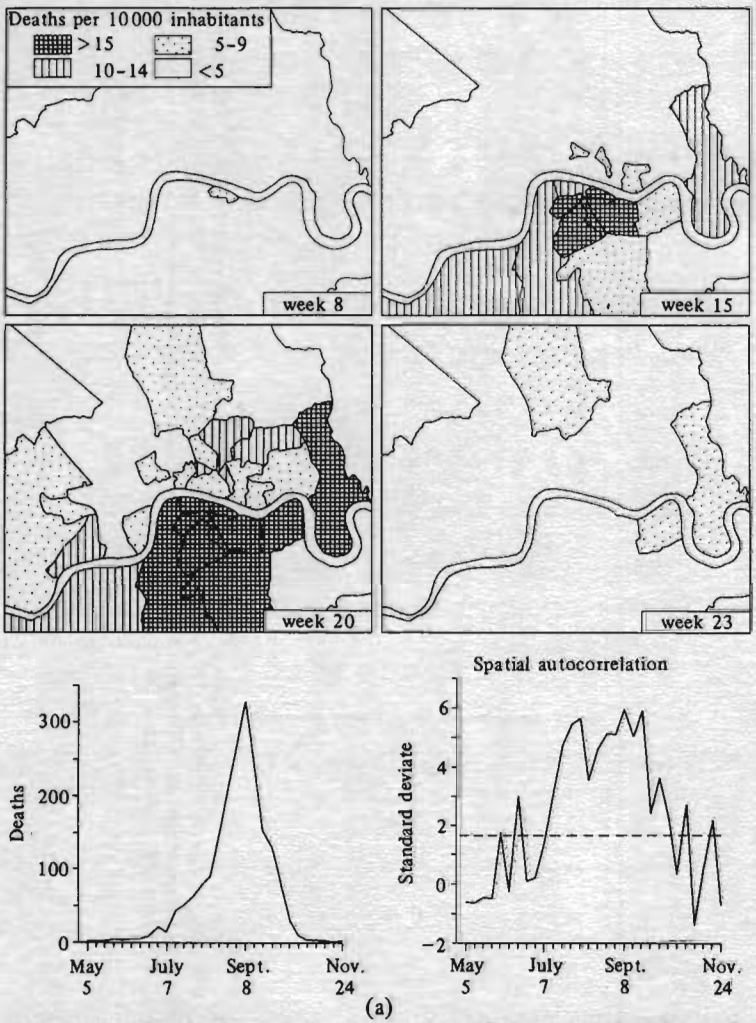
District	April 28	May 5	May 12	May 19	May 26	June 2	June 9	June 16	June 23	June 30	July 7	July 14	July 21	July 28	August 4	August 11	August 18	August 25	September 1	September 8	September 15	September 22	September 29	October 6	October 13	October 20	October 27	November 3	November 10	November 17	November 24	Total deaths	1841 population	Deaths per 10000
1						1				1	1	2	4	4	5	11	12	13	23	35	40	11	12	12	6							193	61326	31
2						1						4	3	14	12	5	20	28	29	48	39	12	14	5	2	1		1				238	40179	59
3						1	1				1	8	4	3	13	9	6	11	10	18	21	5	7	8	2							128	66552	19
4					1		1	1	3	13	4	10	30	20	17	17	34	36	75	56	42	25	27	10	3	2			1			428	56712	75
5			1									2	3	1	8	6	7	8	10	9	13	8	5	2	2	1	1	2			1	90	25091	36
6											1	2	2		1	6	4	6	11	12	6	5	4	1								55	37398	15
7				1	2			1	1	1	3	11	4	6	10	7	24	34	41	51	28	14	10	5	1	1				1	1	258	138164	19
8							1	1	2	3	2	3	2	6	4	14	15	22	15	32	15	33	35	19	8	3			1			189	55690	34
9		1						2	2	3	1	1	1	10	8	19	20	45	54	56	45	36	13	9	1		2	1				327	129763	25
10					1		2		2		1	2		3	4	3	7	13	10	16	7	6	8	11	5		1		1			103	37121	27
11												5	13	6	15	35	28	28	30	27	45	8	6	1		1						248	54292	46
12							1			1		1	4	13	7	17	22	19	15	21	17	10	2	4	2	1				1		158	43598	36
13						1	3	1		1	4	12	9	5	8	8	12	21	9	20	15	4	2	3								138	44461	31
14										2	1	2	2	5	4	6	16	10	18	15	20	10	6	3	1							121	56708	21
15							1			2	1	3	8	5	11	8	13	24	22	28	33	9	9	2	3							177	49829	36
16								1	1		1	5	3	9	13	14	18	20	22	28	17	14	3	8	1	2	1	2				183	39655	46
17			1				3	1	4	22	19	20	18	35	32	18	40	35	45	42	43	21	19	4	2	2						437	29142	150
18										8	5	7	15	17	3	7	18	26	38	20	21	12	3									200	55920	36
19		1				2				1	1	4	6	17	17	21	98	121	139	109	91	52	31	29	12	5	3	1				759	83432	91

Table 1.2 (continued)

District	April 28	May 5	May 12	May 19	May 26	June 2	June 9	June 16	June 23	June 30	July 7	July 14	July 21	July 28	August 4	August 11	August 18	August 25	September 1	September 8	September 15	September 22	September 29	October 6	October 13	October 20	October 27	November 3	November 10	November 17	November 24	Total deaths	1841 population	Deaths per 10000
20									1		1	5	5	10	16	35	125	127	128	96	91	36	28	15	7	5	3		1			734	74088	99
21	1						2	2	3	4	2	7	14	20	30	28	45	55	74	58	48	39	17	20	4	2	2	1				478	71765	67
22								1	3	4		11	7	9	16	16	17	10	15	27	18	6	5	8	1		4					178	41350	43
23								1	4	12	9	11	35	22	31	24	55	58	64	59	49	30	17	11	3	4	1					500	90687	55
24									2	2	1	8	43	26	17	8	17	22	27	41	33	20	15	8	5	1	2					298	31122	96
25									3	4	9	14	14	32	46	47	57	52	65	75	49	29	15	6	6	1			1			526	32975	160
26								6	2	3	5	4	23	26	38	28	28	25	41	44	33	16	10	6	2							340	19837	171
27							3	11	7	6	13	23	64	64	67	56	53	53	70	101	72	32	9	10	6	1			1			773	34947	221
28								1		4	3	16	51	70	112	62	73	57	77	109	108	58	10	8	7	3	1	1				831	46644	178
29							1	1	1	6	5	13	53	66	86	55	65	65	93	157	137	66	15	15	2				1			861	54606	158
30								2	1	10	18	49	106	111	143	84	96	89	179	278	234	117	50	24	7	2	1		1	1		1565	112927	139
31						1				1	1	4	13	19	24	48	32	18	29	45	28	10	3	5	7							288	26337	109
32							1		5	3	2	12	37	43	43	32	33	35	56	79	53	16	9	7		3	1					468	37964	123
33								1	1	5	24	37	37	31	14	25	27	21	20	40	32	18	7	4		2						346	13917	249
34		1				2	1	6	1	2	8	12	28	34	37	29	72	51	75	50	73	44	26	17	8		1					578	55212	105

Key to districts: 1 Kensington, 2 Chelsea, 3 St George, Hanover Square, 4 Westminster, 5 St Martin-in-the-Fields, 6 St James, Westminster, 7 Marylebone, 8 Islington, 9 Pancras, 10 Hackney, 11 St Giles, 12 Strand, 13 Holborn, 14 Clerkenwell, 15 St Luke, 16 East London, 17 West London, 18 London City, 19 Shoreditch, 20 Bethnal Green, 21 Whitechapel, 22 St George-in-the-East, 23 Stepney, 24 Poplar, 25 St Saviour, 26 St Olave, Southwark, 27 Bermondsey, 28 St George, Southwark, 29 Newington, 30 Lambeth, 31 Wandsworth, 32 Camberwell, 33 Rotherhithe, 34 Greenwich.

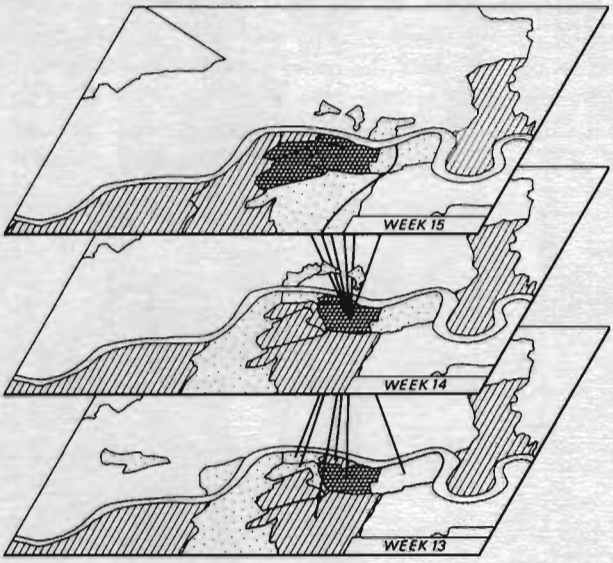
The left-hand graph shows the number of deaths recorded in each week of the epidemic, and the right-hand graph plots the value of the I statistic in standard-deviate form. The pecked line is the $\alpha = 0{\cdot}05$ significance level in a one-tailed test of positive spatial autocorrelation. The degree of spatial contagion waxes and wanes in remarkable sympathy with the epidemic curve, and is indicative of substantial areal concentration of high-risk and low-risk zones when the epidemic was in full swing.

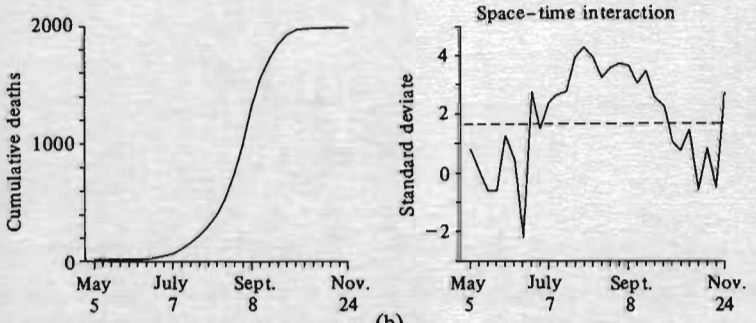


(a)

Figure 1.10. Deaths from cholera in London, 1849, and results of tests for spatial autocorrelation. (b) Cumulative deaths and results of space-time interaction analysis The area served by the Southwark and Vauxhall Water Company is shown on figure 1.2(b).

To try to determine whether the epidemic was dominated by the *in situ* rise and fall of death rates in the areas served by different water companies, or by a spatial diffusion process, we evaluated the space–time index, $I_{s\text{-}t}$, given in equation (1.44). The $\{w_{ij}\}$ were defined as above and we set $y_{ij} = 1$ if borough i was above the mean death rate per 10000 people at time t and area j, $j \neq i$, was also above the mean at $t-1$; $y_{ij} = 0$ otherwise. $I_{s\text{-}t}$ was computed and tested for significance using equations (1.45)–(1.46) for all pairs of maps during the epidemic, and we are thus searching for a relationship between high death rates in area i at t, and high death rates in boroughs contiguous to i at $t-1$; that is, a first order relationship for a space–time spatial diffusion process. The results are illustrated in figure 1.10(b). The set of maps shows a three-week sequence. The bonds show the space–time relationships examined for the cross-hatched area on the week-13 map, and may be compared with figure 1.5. The left-hand





(b)

Figure 1.10 (continued)

graph indicates that the cumulative number of deaths with time has the logistic shape characteristic of most diffusion and epidemic processes, and the right-hand graph is a plot of $I_{s\text{-}t}$ in standard-deviate form (vertical axis) against time. $I_{s\text{-}t}$ is significant during the epidemic peak, which shows that *in situ* processes and diffusion are both important during the peak, whereas the strength of the diffusion processes weakens when the number of cases is low.

To check this interpretation, we constructed the following model. Let $x_{i,t}$ be the death rate per 10000 people from cholera in borough i at time t. Then we postulate

$$X_{i,t} = a + b_1 x_{i,t-1} + b_2 \sum_{j \in J} w_{ij} x_{j,t-1} + \epsilon_{i,t} \tag{1.46}$$

in the manner of figure 1.5. The summation is over the j boroughs contiguous to i. Although there are problems posed by zero death rates in some boroughs during the build-up and fade-out phases of the epidemic, we fitted the model by ordinary least squares using a stepwise regression procedure. Interest centres upon the order in which variables were entered and the levels of explanation provided. If *in situ* growth is more important, we would expect variable 1 (coefficient b_1) to be entered first, whereas if a diffusion process is dominant, we would expect variable 2 (coefficient b_2) to appear first.

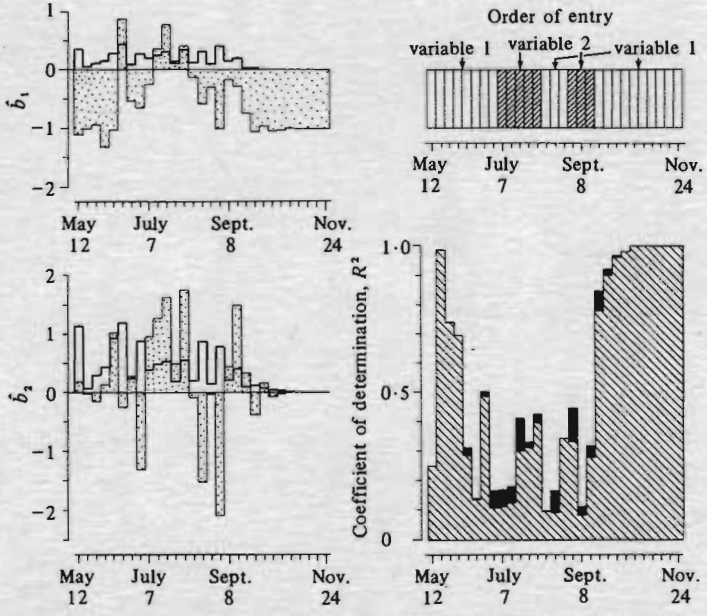


Figure 1.11. Results of stepwise regression analysis of 1849 cholera data. Graphs give the values of the coefficients of the variables, their standard errors, order of entry of variables, and values of R^2.

The results are shown in figure 1.11 for all pairs of maps during the epidemic. The left-hand diagrams give the estimated values of b_1 and b_2 (stipple) and their standard errors (solid line). The top right-hand diagram shows whether variable 1 or variable 2 was entered first for each week. The bottom right-hand diagram gives the values of R^2 (coefficient of determination) after one variable has been entered (cross-hatched) and after the second step (black). High levels of 'explanation' are achieved at the start and end of the epidemic, where the process is dominated by the purely temporal autoregressive component (variable 1). This is largely because of the zero death rates reported by several boroughs during these phases of the epidemic. Where continuous data records are available (beginning of June to early September), levels of R^2 for the model are not high. Space–time interaction (variable 2) is more important in the immediately preepidemic and postepidemic peak periods (July and late August), and temporal autocorrelation is dominant at the epidemic peak.

Given that the quality of the water supply is similar in neighbouring areas and the infectious nature of the disease, these patterns of spread seem reasonable. Indeed, the *in situ* growth may be interpreted as the local spread (within the recording unit) which is likely to predominate when the overall numbers of new cases are low and the disease occurs in isolated pockets.

1.8 Conclusions

In this chapter, the concepts of spatial, temporal, and space–time autocorrelation have been outlined, and various measures of autocorrelation in these domains have been defined. We have indicated that the sampling distributions of the measures approach normality as the size of the study area increases, and tests of significance for the coefficients have been proposed based upon this fact. Formulae for the location and scale parameters of the coefficients have been given. To illustrate the use of the methods, we have analysed the map patterns formed by the incidence of various diseases—measles in Cornwall, 1969–70; cholera in London, 1849; and tuberculosis and bronchitis in Wales, 1959–63—and have tried to indicate how interpreting the maps in the light of the statistical evidence can give insights into the underlying processes shaping the geographical patterns.

As noted above, normal distribution theory is essential to tests of hypotheses using the coefficients. Accordingly we shall turn, in the next chapter, to consider the circumstances under which this theory holds.