THE SPECTRAL ANALYSIS OF LINE PROCESSES

M. S. BARTLETT
UNIVERSITY COLLEGE, LONDON

1. The specification of line processes

In recent papers [2], [3], I have discussed the spectral analysis of point processes in one or more dimensions, showing that the degenerate character of such processes does not prevent spectral analysis techniques, already familiar with continuous processes being adapted to such processes. The question arises, somewhat analogously as in the case of spectral or other distribution functions themselves, whether other forms of degeneracy will be encountered in practice; and, if so, what procedures are possible. One class of process which does arise in various contexts is what I have termed a *line process* ([2], p. 295) in which the points of a point process are replaced, in two or more dimensions, by lines. The example given referred to a number of vehicles on a road, treated for simplicity as points in a one-dimensional continuum, and thus at any instant as a point process. If the points are considered at two instants of time we have a bivariate point process, but if the points are plotted continuously over time as another coordinate the process will consist of a number of lines. This example makes two things clear. First, the specification of the process is partly optional, for the same process is either a point process (in a coordinate x, say) developing in time, or a "static" two-dimensional line process in x and the time coordinate t. Such alternative representations are not exhaustive, for (as in dynamics) the velocity u could also be included if convenient as an additional coordinate, though of course this is not necessary, as u is always derivable in the other specifications. Second, the lines in the line process need not be straight, as when the vehicles are accelerating. Indeed, in any general mathematical specification the lines might not even possess tangents at any point, as in a collection of Brownian particles. We shall, however, for definiteness assume that derivatives exist, as in our example. Moreover, as in the case of point processes, only particular classes of processes can be statistically analyzed by standard techniques. In the case of point processes, spectral analysis requires stationarity (or the equivalent property in more than one dimension). When discussing the spectral analysis of line processes, we shall not only assume an appropriate stationarity property, but shall also for simplicity consider processes consisting merely of *straight* lines, though not necessarily of infinite extent. Such a process in two dimensions is sometimes useful as an idealized representation of the fibers in a sheet of paper.

It should be noted (Bartlett, [1], [4]) that, whether or not the lines are straight, the density functions for line processes corresponding to different representations will satisfy various relations. For example, with the first order densities

$$
\begin{aligned}
f_x(x, t) &= E\{d_x N(x, t)\}/dx, \qquad f_t(x, t) = E\{d_t N(x, t)\}/dt, \\
f_{x,u}(x, u, t) &= E\{d_{x,u} N(x, u, t)\}/dx\, du, \\
f_{t,u}(x, u, t) &= E\{d_{t,u} N(x, u, t)\}/dt\, du,
\end{aligned} \tag{1.1}
$$

we have

$$
f_x = \int f_{x,u}\, du, \qquad f_t = \int f_{t,u}\, du, \qquad f_{t,u} = |u| f_{x,u}. \tag{1.2}
$$

When $u \geqq 0$ for all possible u,

$$
f_t = \int u f_{x,u}\, du = \overline{u(x)} f_x, \tag{1.3}
$$

say,

$$
\overline{u(t)} f_t = \int u f_{t,u}\, du = \int u^2 f_{x,u}\, du \geqq \overline{u(x)}^2 f_x, \tag{1.4}
$$

whence $\overline{u(t)} \geqq \overline{u(x)}$, a result well known in the theory of traffic flow.

2. The spectra of line processes

In the case of point processes, their degeneracy implies that the corresponding spectral functions must be defined in a suitably extended sense. A similar extension will be necessary for processes which are strictly line processes, though the appropriate definition will depend on whether the lines are finite or infinite. Elsewhere I have shown (Bartlett, [4], §6.52) that a (straight) line process may conveniently be included as a degenerate example of the more general process

$$
X(\mathbf{r}) = \int \xi(\mathbf{s} - \mathbf{r})\, dN(\mathbf{s}), \tag{2.1}
$$

where $N(\mathbf{s})$ is some point process and $\{\xi(\mathbf{r})\}$ is a random function associated with each point event of $N(\mathbf{s})$ with the point as origin. The $\xi(\mathbf{r})$ are in general different realizations for each such point.

In the very special and purely random case of $N(\mathbf{s})$ a Poisson process and $\xi(\mathbf{r})$ zero except on an infinite line of random orientation, we find $f(\boldsymbol{\omega})$ varies as $1/\omega$, where $f(\boldsymbol{\omega})$ is the (unstandardized) spectrum of $X(\mathbf{r})$ and $\omega^2 = \omega_x^2 + \omega_y^2$. It is possible that direct measurement of the spectra of line (or near line) processes may be feasible in certain contexts; but in the analysis of line processes by digital computation it seems convenient to make use of any alternative representations to transform such a line process first to a convenient point process, and then to analyze this point process. Let us list some theoretical examples.

(i) In the case of the purely random family of straight lines $x \cos \theta + y \sin \theta = p$, the line process is equivalent to the Poisson point process in the infinite strip p from $-\infty$ to ∞, and θ from 0 to π (Kendall and Moran [5]).

(ii) For finite lines it is possible to think in terms of formula (2.1). Thus, if the lines are of fixed length 2ℓ, we may consider the two-dimensional point process of their centers, and an angle variable θ from 0 to π representing slope. (In the case of finite "arrows," that is, lines with directions, θ would vary from 0 to 2π.)

(iii) For finite lines of variable length $2L$, there will be an additional random variable L for each point. It is possible to think of L, or a transformation of it, as introducing a further dimension to the point process; but in practice, as such a point process would not in general be stationary even if the original process were, it seems preferable to specify L merely as an ancillary variable. The same procedure could apply to a set of particles (or vehicles) whose positions x and velocities u were given at a single instant t (or t and u at a single position x), giving rise to a one-dimensional point process with ancillary variable.

An interesting feature of the point process representations in (i) and (ii) is that the point process is specified on a particular coordinate structure which will affect its spectral function. Unlike p in (i) or $\mathbf{r}$ in (ii), θ is an angle variable with a Fourier *series* spectrum. The combined spectrum for p or $\mathbf{r}$ with θ will consequently be coefficients associated with a Fourier series, each coefficient of which will have the form of a spectral function for p or $\mathbf{r}$. (If the line process were specified in three dimensions, the single angle variable θ would be replaced by two angles θ and ϕ determining position on a unit sphere, with a corresponding series of coefficients associated with expansions in spherical harmonics (see, for example, Bartlett, [4], §6.53).

Another feature to notice is that the angle variable θ will be uniform for a completely random line process, but this does not apply to some transformed variable such as the slope $s = \tan \theta$, for which the density is

$$f(s) = \frac{1}{\pi} \frac{1}{1 + s^2}. \tag{2.2}$$

This raises the problem whether in some other example, such as the traffic situation with vehicle velocities, there is any advantage in transforming the ancillary variable to an angle variable by such a transformation as $\tan^{-1} s$, or more generally $\tan^{-1} (s - s_0)$. It might be worth exploring this possibility somewhat further, though the "nonstationarity" in general of the point process so extended, even if the transformation is carefully chosen, seems to make the use of spectral analysis less relevant with this device, as previously noted. It was felt that a simpler and more empirical incorporation of the ancillary variable in any spectral analysis was likely to be more informative, and the procedure adopted is discussed below.

3. The spectral analysis of line processes

The two numerical examples given for illustration will be

(a) an artificial, purely random, straight line process;

(b) a set of time instants at which vehicles passed a point on a road, together with their velocities as values of an ancillary variable.

Example (a) was chosen as a line process for which the representation (i) above was possible, as distinct from the more general representation (ii) which would have meant a more complicated spectral analysis. Similarly, example (b), while perhaps more immediately classifiable as a point rather than a line process, is similar but simpler than the first point process representation mentioned in (iii) for lines of variable length. However, the "periodogram" sums are defined below for both one- or two-dimensional point process representations, with either one angle variable θ or one ancillary variable U. In the case of an angle variable we write

$$J_s(\omega) = \sqrt{\frac{2}{n}} \sum_r e^{i(\omega' \mathbf{X}_r + s\theta_r)}, \tag{3.1}$$

where n is the number of points with (column) vector coordinates $\mathbf{X}_r$ for $r = 1, \cdots, n$ and ω' the (row) vector frequency (so that in two dimensions $\omega'\mathbf{X} \equiv \omega_1 x_1 + \omega_2 x_2$). In the case of an ancillary variable U, we consider the somewhat more empirical sum

$$J_U(\omega) = \sqrt{\frac{2}{n}} \sum_r e^{i\omega' \mathbf{X}_r} \delta U_r, \tag{3.2}$$

where δU_r stands for $U_r - \bar{U}$, $\bar{U}$ being the observed mean. For large n, the sampling properties of $J_U(\omega)$ will not be affected to the first order by the use of $\bar{U}$ in place of the true mean $E\{U\}$. It seems convenient to measure U from its mean, so that $J_U(\omega)$ is zero if U does not vary; and thus, $J_U(\omega)$ is kept as distinct as possible from the unmodified sum $J(\omega)$ (or $J_0(\omega)$ in (3.1) above).

Corresponding to equation (3.1), there will be a spectral function of the general form $f(\omega, s) \equiv \alpha_s(\omega)$, for $s = 0, 1, 2, \cdots$. For example, in the case of X_r representing the centers of lines of constant length, with θ measuring their angle of direction (0 to π), the assumption of independent θ gives

$$\begin{aligned} \alpha_s(\omega) &= 1 + f(\omega) E\{e^{is(\theta_r - \theta_s)}\} \\ &= 1 + f(\omega)\, \delta_{s,0}, \end{aligned} \tag{3.3}$$

say, where $1 + f(\omega)$ is the spectral function for $\mathbf{X}$ (standardized to unity for random $\mathbf{X}$) and $\delta_{s,0}$ is zero for even $s \neq 0$, and 1 for $s = 0$. For odd s, $\delta_{s,0} = 4/\pi^2 s^2$. In the simplified p, θ representation for purely random lines of infinite length, we may for convenience take the range 0 to ∞ for p and 0 to 2π for θ (instead of $-\infty$ to ∞ for p and 0 to π for θ). In this case we then have $\delta_{s,0} = 0$ for all integers s except $s = 0$.

In an alternative extreme nonrandom case where θ is constant, we should have $\delta_{s,0}$ in (2.5) equal to unity whatever s.

Consider next the sum in (2.4), which we replace in theoretical studies by

$$J'_U(\omega) = \sqrt{\frac{2}{n}} \sum_r e^{i\omega' \mathbf{X}_r} \Delta U_r, \tag{3.4}$$

where $\Delta U_r = U_r - E(U_r)$. Then for $I'_U(\omega)$ where $I_U(\omega) = J_U(\omega)J^*_U(\omega)$, and so forth, we obtain

$$E\{I'_U(\omega)\} = \frac{2}{n} \iint e^{i\omega' z} E\{dN(\mathbf{x})\Delta U(\mathbf{x}) dN(\mathbf{x}+y)\Delta U(\mathbf{x}+z)\}, \tag{3.5}$$

where $N(\mathbf{x})$ is the point process for $\mathbf{X}$, $U(\mathbf{x})$ is U_r at a point $\mathbf{X}_r$ for which $dN(\mathbf{x}) = 1$, and the integration is over the sample region containing the $\mathbf{X}_r$. If ΔU_r is independent of $N(\mathbf{x})$, there is no contribution to the integral except at $z = 0$, and $E\{I'_U(\omega)\} \to \lambda\sigma_u^2$ for all $\omega \neq 0$, where $\sigma_u^2 = E\{(\Delta U_r)^2\}$ and $E\{dN(\mathbf{x})\} = \lambda\, d\mathbf{x}$. More generally, we shall write

$$E\{dN(\mathbf{x})\Delta U(\mathbf{x}) dN(\mathbf{x}+z)\Delta U(\mathbf{x}+z)\} = \{\lambda\sigma_u^2\delta(\mathbf{z}) + \mu_u(\mathbf{z})\}\, d\mathbf{x}\, d\mathbf{z}. \tag{3.6}$$

If we write

$$E\{dN(\mathbf{x})dN(\mathbf{x}+\mathbf{z})\} - \lambda^2\, d\mathbf{x}\, d\mathbf{z} = \{\lambda\delta(\mathbf{z}) + \mu(\mathbf{z})\}\, d\mathbf{x}\, d\mathbf{z}, \tag{3.7}$$

the second term on the right-hand side of equation (2.8) can rise to $\sigma_u^2\mu(\mathbf{z})\, d\mathbf{x}\, d\mathbf{z}$ in the extreme case where U_r is perfectly correlated with U_s for points $\mathbf{X}_r$, $\mathbf{X}_s$ contributing to $\mu(\mathbf{z})$.

In order to examine further possibilities, let us consider the more general extended process

$$dM(\mathbf{x}) = dN(\mathbf{x})[1 + \xi\Delta U(\mathbf{x})], \tag{3.8}$$

where ξ is an arbitrary (possibly complex-valued) coefficient. Then $E\{dM(\mathbf{x})\} = \lambda\, d\mathbf{x}$, and for the complete covariance density $\nu_c(\mathbf{z})$ for $dM(\mathbf{x})$, we obtain

$$\lambda(1 + \xi\xi^*\sigma_u^2)\delta(\mathbf{z}) + \nu(\mathbf{z}), \tag{3.9}$$

say, where

$$\begin{aligned}\nu(\mathbf{z})\, d\mathbf{x}\, d\mathbf{z} = {} & \mu(\mathbf{z})\, d\mathbf{x}\, d\mathbf{z} + \xi^* E\{\Delta U(\mathbf{x}) dN(\mathbf{x}) dN(\mathbf{x}+\mathbf{z})\} \\ & + \xi E\{\Delta U(\mathbf{x}+\mathbf{z}) dN(\mathbf{x}) dN(\mathbf{x}+\mathbf{z})\} + \xi\xi^*\mu_u(\mathbf{z})\, d\mathbf{x}\, d\mathbf{z}.\end{aligned} \tag{3.10}$$

To demonstrate the nature of these functions in a particular one-dimensional clustering model, suppose ΔU in a cluster is associated with cluster size; for example, with traffic data large clusters might well be associated with low velocities if overtaking were difficult. For definiteness suppose the relation is linear, so that

$$E(\Delta U|r) = \beta[r - E(r)], \tag{3.11}$$

where $r + 1$ is the total cluster size; and suppose the residual $\Delta U - E\{\Delta U|r\}$ is otherwise correlated to extent ρ within a cluster. We find (see Bartlett, [2], p. 266)

$$\begin{aligned}&\nu(z|r) \\ &\quad = E\{\lambda_c[f_r(z) + 2f_{r-1}(z) + \cdots + rf_1(z)]\,[1 + \xi\xi^*\Delta U\Delta U' + (\xi + \xi^*)\Delta U]|r\},\end{aligned} \tag{3.12}$$

where ΔU, $\Delta U'$ are different ΔU in the same cluster, $f_r(z)$ is the rth convolution density of the interval between consecutive vehicles in a cluster, and λ_c is the average density of clusters. After substitution from (3.11) and

$$E\{\Delta U \Delta U' | r\} = \rho v_r + \beta^2[r - E(r)]^2, \tag{3.13}$$

where v_r is the variance of ΔU given r, we finally obtain

$$\nu(z) = E_r\{\lambda_c[f_r(z) + 2f_{r-1}(z) + \cdots + rf_1(z)] [1 + \xi\xi^*\{\rho v_r + \beta^2[r - E(r)]^2\} + \beta(\xi + \xi^*)[r - E(r)]\} \tag{3.14}$$

where E_r denotes averaging over r.

Two conclusions from this formula are

(i) if β in this model is zero, then $v_r = \sigma_u^2$, and the second term in (3.6) becomes $\rho\sigma_u^2\mu(z)$. In general, however, for $\beta \neq 0$, the relation of $\mu_u(z)$ to $\mu(z)$ is more complicated;

(ii) if $\beta = 0$, no cross-spectral density terms (coefficients of ξ and ξ^*) arise, but in general for $\beta \neq 0$ further information may be available from the cross-spectrum of $dN(x)$ and $\Delta U(x)dN(x)$. In particular, information on the sign of β is only available from the cross-spectrum.

4. Analysis of example (a)

The data for the first example consisted of the 50 lines shown in figure 1, with p from 0 to ∞ and θ from 0 to 2π coordinates given in table I. The latter

TABLE I

DATA FOR FIRST EXAMPLE

p	θ	p	θ	p	θ	p	θ	p	θ
2.45	2.461	48.32	4.683	12.40	0.913	5.60	0.217	28.76	5.931
11.85	5.771	19.39	2.244	29.15	1.154	28.74	3.630	36.19	2.799
39.31	4.584	21.17	1.534	64.52	1.856	66.72	3.074	41.70	3.879
59.69	4.893	7.97	1.637	60.26	0.392	12.99	4.820	7.69	5.280
23.48	0.017	27.03	4.223	62.76	0.484	10.63	0.582	43.59	3.436
12.86	0.557	47.14	0.000	30.92	3.261	21.31	1.443	54.86	4.962
16.56	3.737	48.12	0.883	44.03	4.573	26.77	0.671	37.16	0.173
26.76	5.110	45.39	4.009	39.85	5.085	69.96	5.808	26.91	0.375
40.04	0.983	5.64	1.540	12.40	1.346	67.00	5.945	3.10	2.922
11.37	5.142	19.08	3.038	8.73	0.116	11.92	1.307	56.70	5.731

values were obtained by calculating $\tan^{-1}(x/y)$ from a pair of independent normal variables x and y (*Tracts for Computers*, No. 25), and the former converted from uniformly distributed numbers in the range 0 to 100 (*Tracts for Computers*, No. 24) by dividing by $\sqrt{2}$, thus ensuring that the 50 lines intersected a circle of radius $50\sqrt{2}$, and hence most of them a square of side 50 (two did not, but were retained in the analysis).

The values of $I_s(\omega_p) = J_s(\omega_p)J_s^*(\omega_p)$ were computed for $\omega_p = 2\pi p/50$, for $p = 1, \cdots, 100$ and $s = -5$ to 5. Individual values are not reproduced, but the frequency tables for each s are summarized in table II, and the cumulative totals in steps of five at a time are given for each s in table III. The distributions

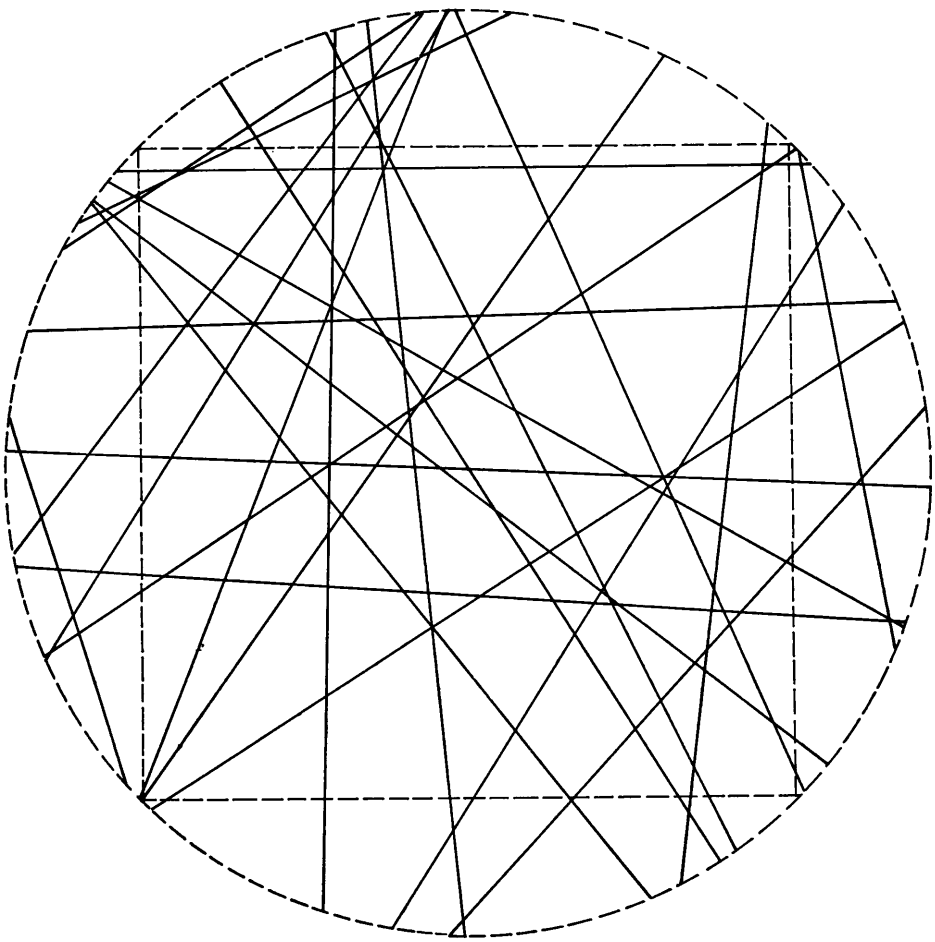

FIGURE 1

Fifty random straight lines, example (a).

in table II do not appear unreasonable, apart perhaps from rather more large values in the row for $s = +3$ than would be expected. However, the overall average of 2.10 is near to the theoretical average of 2, and the variation of the averages for the different rows gives a χ^2 of 17.86 with 10 d.f., which does not reach significance at the $P = 0.05$ level, namely, 18.31.

TABLE II

FREQUENCY TABLES FOR $I_s(\omega_p)$

s	0–	1–	2–	3–	4–	5–	6–	7–	8–	9–	10–	11–	12–	13–	Total	Avg.
−5	38	27	8	8	10	5	1	1	—	2					100	2.13
−4	34	34	9	9	6	4	1	1	1	—	—	1			100	2.05
−3	42	19	16	10	4	4	3	1	—	—	1				100	2.02
−2	48	20	16	4	6	2	2	1	1						100	1.75
−1	33	21	21	14	7	3	—	1							100	2.05
0	37	20	16	9	7	5	2	3	1						100	2.23
1	40	27	17	6	5	1	1	1	1	—	—	—	1		100	1.87
2	38	26	14	10	3	7	—	1	—	1					100	1.97
3	34	17	18	7	8	4	2	4	1	—	2	1	1	1	100	2.80
4	46	19	12	7	5	1	3	3	2	1	—	—	1		100	2.15
5	37	30	10	8	7	2	2	—	2	2					100	2.08
Total	427	260	157	92	68	38	17	17	9	6	3	2	3	1	1100	2.10

TABLE III

CUMULATIVE TOTALS FOR $I_s(\omega_p)$

p \ s	−5	−4	−3	−2	−1	0	1	2	3	4	5
5	10.53	8.35	5.54	7.04	11.40	11.70	20.10	14.30	4.57	5.44	14.71
10	14.20	20.19	24.81	13.84	23.31	24.86	29.30	19.81	15.27	11.25	19.06
15	25.80	26.75	30.20	29.00	28.64	32.86	34.02	32.30	31.66	17.04	37.33
20	37.17	36.55	47.99	34.21	36.64	45.00	38.62	43.73	52.75	23.92	49.70
25	51.41	43.72	57.98	46.34	51.30	53.43	44.76	54.83	64.60	43.57	60.55
30	61.44	54.34	64.56	53.86	62.18	74.06	50.47	60.11	82.60	54.89	72.88
35	74.47	70.34	77.05	69.67	71.09	93.65	63.37	74.19	94.60	62.39	83.56
40	79.80	78.26	86.95	79.10	77.66	113.73	82.24	81.65	101.56	69.54	89.82
45	91.53	85.55	102.11	91.29	92.41	119.56	90.28	92.65	115.84	82.60	95.16
50	98.60	97.96	106.78	101.22	104.06	127.50	97.43	100.77	128.70	92.38	102.75
55	114.21	109.46	114.71	106.56	117.71	133.22	101.91	117.09	150.60	102.36	108.14
60	132.66	118.35	123.11	110.21	124.75	137.89	108.77	128.00	161.34	113.60	116.44
65	149.30	124.26	131.74	118.23	133.14	152.02	116.66	132.39	177.77	123.27	131.33
70	161.41	136.98	134.87	129.59	143.25	165.56	125.30	141.16	186.68	141.00	140.25
75	169.66	141.40	142.71	133.22	149.72	178.63	139.83	147.72	196.05	144.57	147.57
80	175.98	151.98	148.83	147.80	154.85	187.68	151.08	158.99	205.04	151.73	163.17
85	179.24	158.56	158.28	154.95	162.52	193.53	159.82	171.73	221.66	163.22	174.84
90	198.19	173.82	179.20	161.56	171.00	208.09	168.45	185.72	232.34	169.44	179.83
95	205.22	185.45	191.41	168.87	179.08	212.70	174.06	194.17	263.28	193.32	185.15
100	212.99	203.37	197.22	175.80	196.08	216.69	183.64	198.01	272.95	211.30	203.38

5. Analysis of example (b)

The traffic data for the second example were kindly supplied to me by the National Road Research Institute, Stockholm, and consisted of the time instants in seconds of vehicles passing a fixed point in the northbound direction on a two lane road ($E4$) between Stockholm and Uppsala on September 16, 1961. Velocity

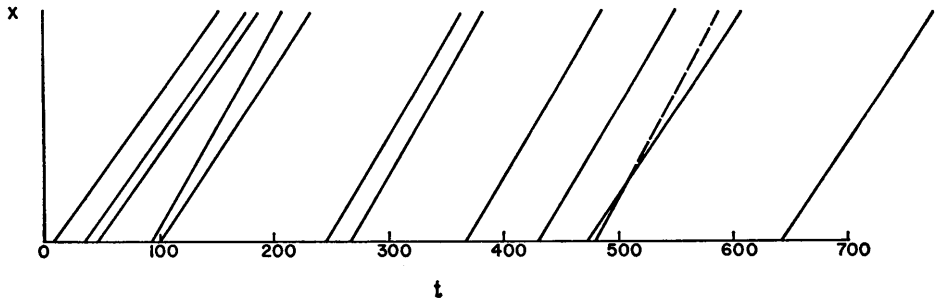

FIGURE 2

Times for first twelve vehicles, example (b),
with velocities depicted by the slopes of the lines (arbitrary scale).

measurements were only measured approximately in 10 km/hr group intervals. This may preclude a very accurate study of spacing-speed relations, but should be adequate for the type of spectral analysis described above. The entire series was quite extensive, consisting of 1215 observations, in which five velocities were missing. A series of 320 complete observations was chosen (the maximum available was 325). The data are not reproduced here, but a graph of the first twelve vehicle times and velocities is shown in figure 2. The results obtained from this set were checked from another set of 320 observations, containing only

TABLE IV

BLOCK TOTALS OF 16 FOR $H_p = \overline{U}^2 I(\omega_p)$ AND $H'_p = I_O^U(\omega_p)$

p	1st Series		2nd Series	
	H_p	H'_p	H_p	H'_p
1–16	14367	651.8	19777	532.3
17–32	14985	614.0	17370	480.7
33–48	14534	434.0	14540	416.5
49–64	14980	371.8	13278	311.2
65–80	12273	319.7	15317	229.9
81–96	16133	384.2	11158	223.6
97–112	9231	226.4	8217	332.2
113–128	12085	429.8	9639	190.5
129–144	6915	331.1	5019	250.0
145–160	10455	370.4	9147	226.3
161–176	11107	223.1	7400	184.7
177–192	8363	317.9	7856	188.4
193–208	8667	282.4	7109	184.6
209–224	6718	370.5	8660	212.7
225–240	7030	347.8	9818	217.4
241–256	4031	320.8	4094	288.9
257–272	8950	200.1	8659	218.7
273–289	5680	353.3	5484	190.5
289–304	5182	195.8	4823	146.1
305–320	6236	343.7	7160	153.4

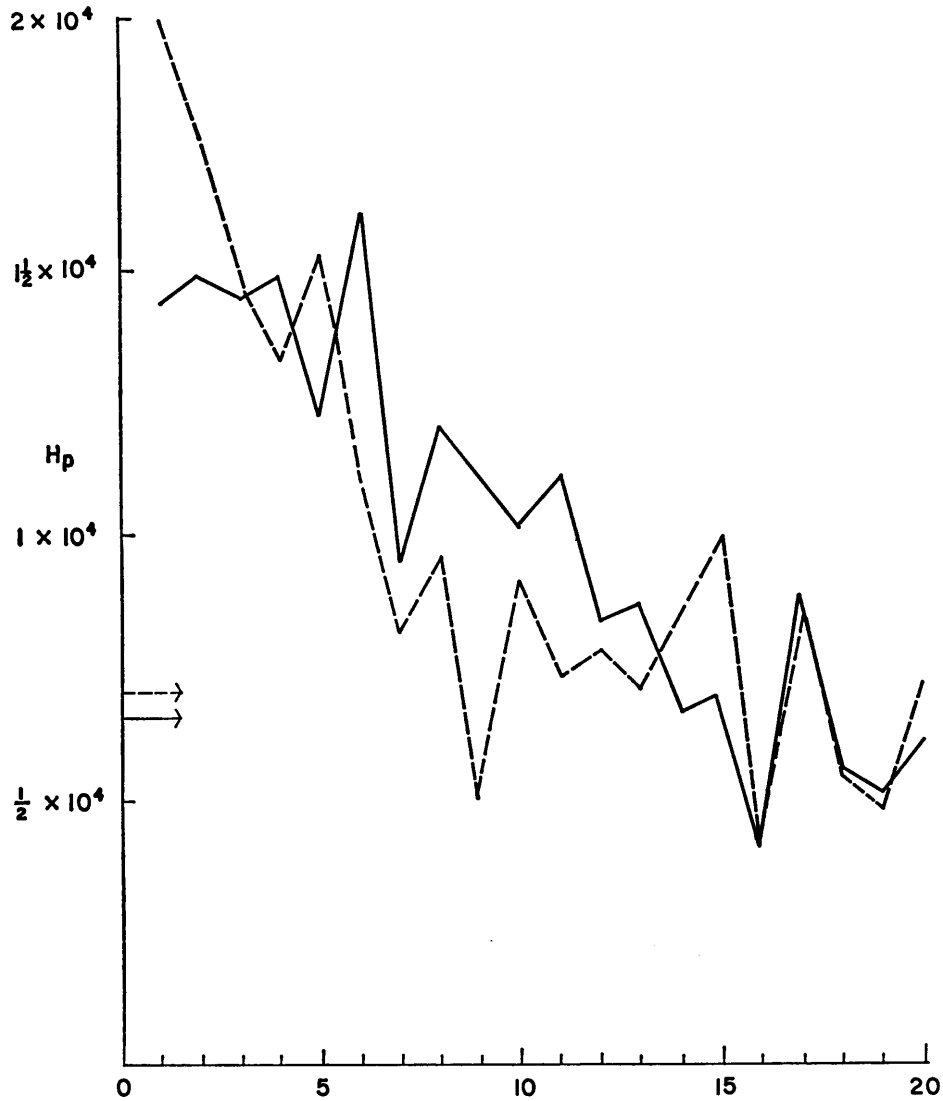

FIGURE 3

Values of $H_p = U^2I(\omega_p)$ summed over blocks of 16 (1st series: continuous line; 2nd series: dotted line). The expected ultimate values are indicated by arrows.

one missing velocity observation, for which the near average value of 75 km/hr was inserted.

In addition to the $J_U(\omega_p)$ of equation (3.2), a more standard point-spectrum analysis was made from $J(\omega_p)$, or rather from $\bar{U}J(\omega_p)$, so that in addition to $I_U(\omega_p)$ values were available of $\bar{U}^2I(\omega_p)$. The range of p taken was from 1 to 320,

and block totals of 16 were recorded. For the first series, the value of $\bar{U}$ is, in units of 5 km/hr, 14.77, so that the expected value of a block total of 16 in such units is $14.77^2 \times 2 \times 16 = 6661$ on the null hypothesis. The corresponding

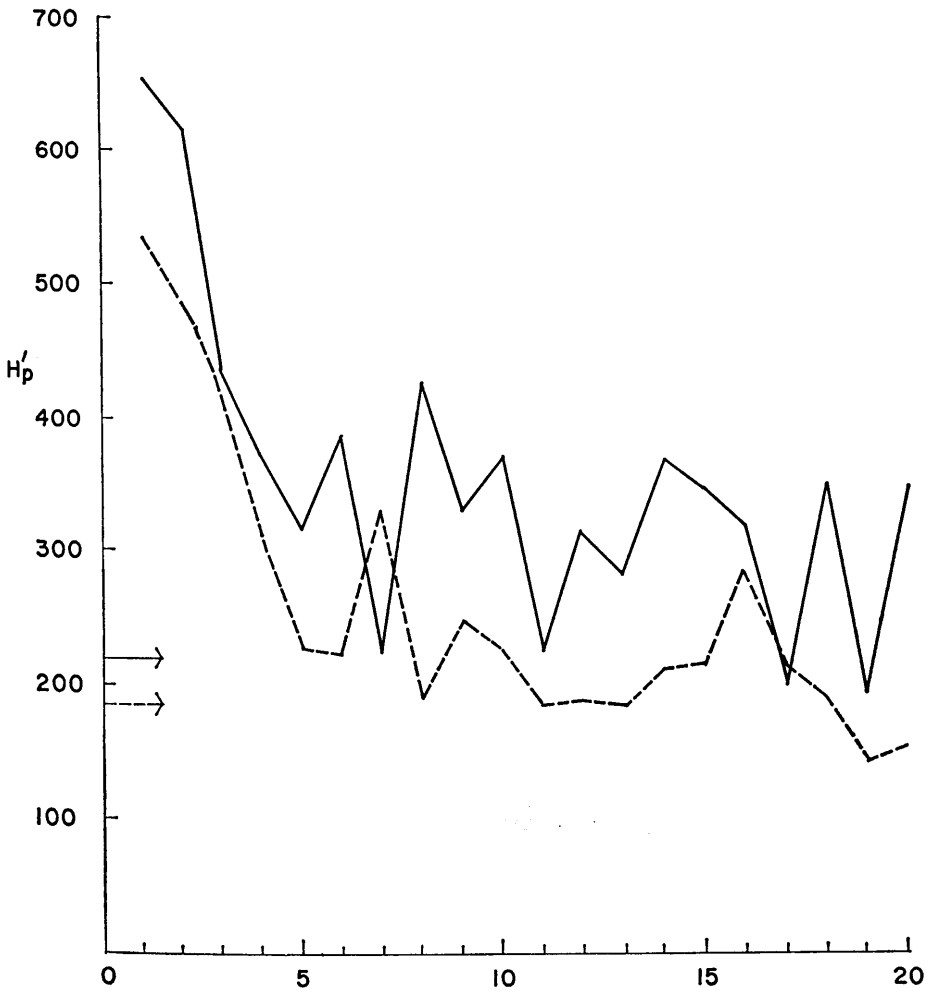

FIGURE 4

Values of $H_p' = I_U(\omega_p)$ summed over blocks of 16 (1st series: continuous line; 2nd series: dotted line). The expected ultimate values are indicated by arrows.

value on a random hypothesis for totals of $I_U(\omega_p)$ is $32\sigma_u^2$, estimated to be in the same units $32 \times 6.87 = 219.8$. The corresponding expected values for the second sum are $32 \times 14.84^2 = 7047$ and $32 \times 5.80 = 185.6$. The actual values obtained are given in table IV and figures 3 and 4. The significance of the rise

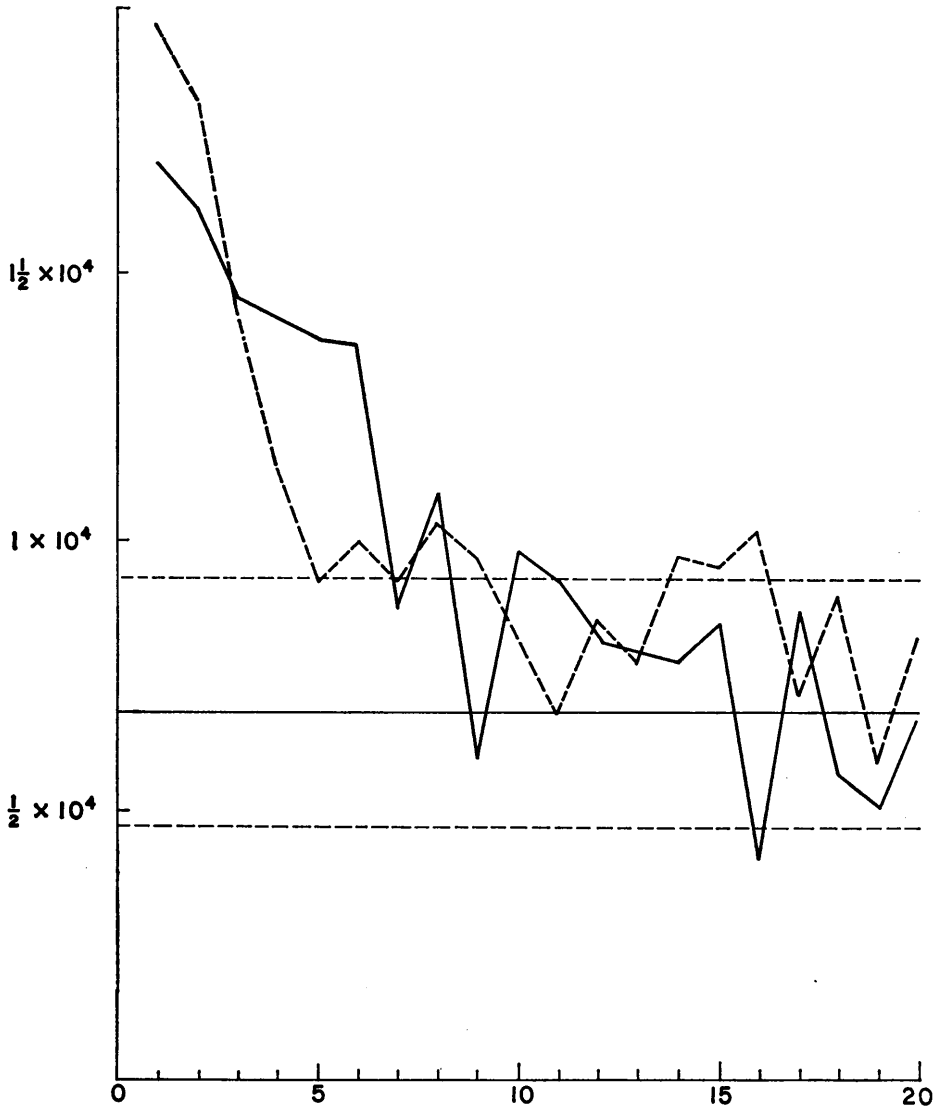

FIGURE 5

Average of H_p for both series (continuous line) with similar average for H'_p standardized to same ultimate level (dotted line). $P = 0.05$. Significance levels (two sides) for any point are indicated.

near the origin is clear from figure 5 which shows H_p averaged over the two series, with H'_p *standardized to the same ultimate level*, and $P = 0.05$ significance levels (two sides, for each separate point).

6. Further discussion of results for example (b)

The results for $I(\omega_p)$ were expected to show a spectrum similar to the one depicted for traffic data by Bartlett ([2], figure 1), and both series broadly agree in this. In fact, while the average intervals between vehicles is somewhat lower (12.35 secs for the first series and 10.63 secs for the second, compared with 15.81 secs in the earlier example), the density has been standardized to unity; the previous theoretical model, as specified in my 1963 paper [2], would appear reasonably compatible with the present results. It is recalled that it embodied a clustering process, with a modified geometric distribution for cluster size (excluding the leading vehicle)

$$p(r) = \begin{cases} 1 - c, & r = 0, \\ c\alpha^{r-1}(1 - d), & r = 1, 2, \cdots, \end{cases} \tag{6.1}$$

with $c = 1/9$, $\alpha = 2/3$. A dominant feature of the spectrum is the ratio of its value near $\omega = 0$ to its limiting value as ω increases, this being equal to

$$m + \frac{\sigma^2}{m} = \frac{(1 - \alpha)^2 + c(3 - \alpha)}{(1 - \alpha)(1 + c - \alpha)} \tag{6.2}$$

for the above model, where m and σ^2 are the mean and variance of $r + 1$. It will be noticed that rather indirect information is provided on c by formula (6.2).

The results for $I_U(\omega_p)$ are the more novel. The rise in $I_U(\omega_p)$ with $I(\omega_p)$, while somewhat more irregular, is present for both series, and is consistent with an anticipated correlation of velocities for vehicles in the same cluster. For the first series the values of $I_U(\omega_p)$ seem to remain a little high on average compared with the expected limit of 219.8 even for the larger values of ω. In general, the relation of $I_U(\omega_p)$ to $I(\omega_p)$ can be complicated (see formula (3.14)); but any apparent persistence of $I_U(\omega_p)$ above its ultimate value for large ω is not repeated for the second series; and it was decided to consider, at least provisionally, the simple clustering model where β is zero and velocity fluctuations within a cluster had constant correlation ρ. The individual differences of $I(\omega_p)$ or $I_U(\omega_p)$ from their ultimate values are of course subject to relatively large sampling error. However, the ratio $(H'_p/H_\infty - 1)/(H_p/H_\infty - 1)$ will be most accurate for large value of the denominator; and an overall estimate of ρ was made by weighting by the square of the denominator. The values H_p, H'_p were taken separately for the two series given in table IV, and the calculated values used for H_∞, H'_∞. The estimates of ρ so obtained are 0.76 and 0.78, respectively, suggesting rather a high correlation within clusters.

Such an effect should be demonstrable in other ways. The correlation ρ should

give rise to a detectable serial correlation between consecutive vehicle velocities, where

$$\rho' = \rho(m - 1)/m. \tag{6.3}$$

With $m = 4/3$, $\rho' = 0.19$ when $\rho = 0.76$, and 0.20 when $\rho = 0.78$. The actual serial correlations were computed to be 0.26 from the first series and 0.29 from the second. The agreement seems fair: though it could be somewhat improved either (i) by increasing m, or (ii) by increasing ρ, or (iii) supposing that additional heterogeneity in traffic density may contribute to the observed serial correlations.

With the apparent high correlation of velocities within clusters another rough check on the consistency of the model is possible. Suppose for simplicity we consider the correlation to be near unity. Runs of identical velocities will then be assumed to arise from two contingencies: (i) clusters; (ii) fortuitous runs. If the velocity distribution with discrete categories has probabilities $p_1, p_2, \cdots, p_k$, then runs of length s from a purely random series have probability

$$p_1^s q_1 + p_2^s q_2 + \cdots + p_k^s q_k. \tag{6.4}$$

From the observed velocity distributions (for each series of 320 observations separately), the probabilities in (6.4) yield the calculated distributions of table V,

TABLE V

DISTRIBUTION OF RUNS OF VEHICLES WITH SAME VELOCITY

s	1st Series		2nd Series	
	p_s	Observed	p_s	Observed
1	0.7541	126	0.7460	134
2	0.1763	44	0.1778	29
3	0.0488	11	0.0519	18
4	0.0146	4	0.0163	9
5	0.0044	1	0.0053	3
6	0.0012	3	0.0017	0
7	0.0004	0	0.0006	1
8	0.0001	2	0.0002	2
9+	0.0001	2	0.0002	0
Total	1.0000	193	1.0000	196
Mean	1.345	1.653	1.367	1.633

with the observed distributions shown for comparison. As the calculation is very rough, runs involving a single cluster of more than one for $r > 0$ are neglected (as well as the overlap of clusters). We then have the approximate equation for the first series, $1.345 + c/(1 - \alpha) = 1.653$, the second term on the left being the expected increase in length of run due to clusters of more than one. With $\alpha = 2/3$, this gives $c = 0.308/3 = 0.103$, a value compatible with the

value 1/9 previously assessed [2]. This estimate, while rather crude, is of some interest in view of the comparative paucity of information on c noted above. The corresponding figures for the second series are 1.367 (in place of 1.345), 1.633 (for 1.653), whence c is 0.266/3 (for the same α), that is, 0.089.

It might be noted that the mean value of the velocity for the larger runs ($\geqq 5$, say) is, in 5 km/hr units, 14.0 for the first series and 13.7 for the second, compared with an average over all vehicles of 14.8. This provides slight evidence of a $\beta < 0$ in (3.11), but hardly perhaps enough to justify fitting any more complicated model as represented by such formulae as (3.14). However, it was felt that calculation of the cross-spectrum would be of interest, and the results are described below. The relevant explicit evaluation of (3.14) for the clustering model is given in the appendix.

7. Calculation and discussion of the cross-spectrum

The cross-spectrum was conveniently computed by making use of the identity

$$\bar{U}^2 + (\delta U)^2 - U^2 = -2\bar{U}(\delta U), \tag{7.1}$$

where $\delta U = U - \bar{U}$. Thus, the spectrum of $UdN(x)$ was computed and hence, making use of (7.1), the cross-spectral function of $dN(x)$ and $\delta U dN x$. The results are given in table VI (and figure 6), which gives $-G_p = -\bar{U}I_{12}(\omega_p)$, where

$$I_{12}(\omega_p) = A_1(\omega_p)A_2(\omega_p) + B_1(\omega_p)B_2(\omega_p), \tag{7.2}$$

the subscript 1 referring to $dN(x)$ and 2 to $\delta U dN(x)$. (Notice that under the assumptions for our model the imaginary terms in the cross-spectral function do not appear, so that it is sufficient to calculate $I_{12}(\omega_p)$ above.)

TABLE VI

BLOCK TOTALS OF 16 FOR $-G_p$

1st Series				2nd Series			
1.	2313.9	11.	561.9	1.	1520.7	11.	−380.1
2.	2723.5	12.	1093.5	2.	1436.6	12.	−263.5
3.	387.5	13.	190.9	3.	1562.8	13.	−412.4
4.	2447.3	14.	670.4	4.	1695.2	14.	−8.5
5.	1521.8	15.	512.5	5.	67.8	15.	−407.9
6.	1777.5	16.	−84.0	6.	1278.7	16.	−629.6
7.	1765.8	17.	379.8	7.	588.7	17.	459.9
8.	345.9	18.	75.7	8.	575.0	18.	728.9
9.	−142.0	19.	−376.2	9.	635.1	19.	−92.3
10.	418.2	20.	287.0	10.	744.1	20.	42.9

The most significant feature of G_p is its negative value as $\omega \to 0$, implying the anticipated negative value for β. We shall confine our attention to the

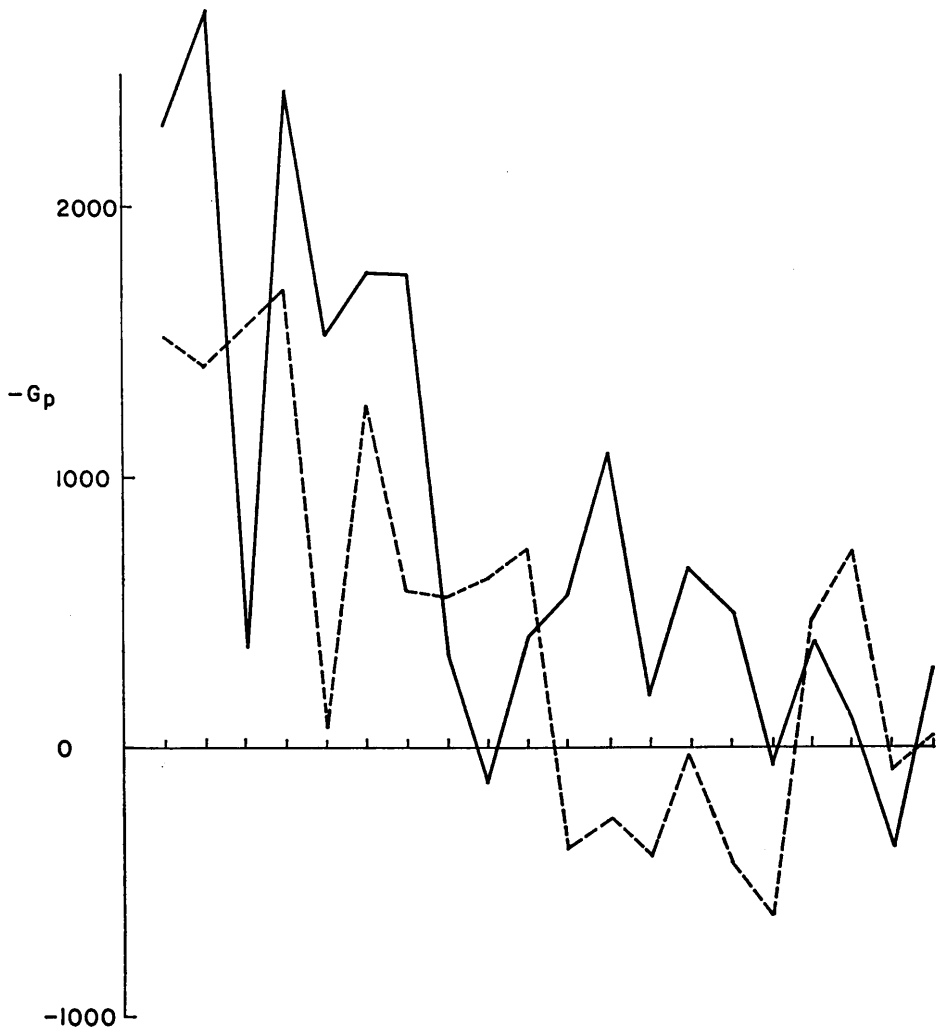

FIGURE 6

Values of $-G_p = -UI_{12}(\omega_p)$ summed over blocks of 16 (1st series: continuous line; 2nd series: dotted line).

values in the appendix at $\omega = 0+$. These values can only be appraised roughly from the graphs; but the following values were used:

1st series: $-G_0 = 2000, H_0 - H_\infty = 1\frac{1}{2} \times 10^4 - 6661, H_0' - H_\infty' = 600 - 220;$
2nd series: $-G_0 = 1500, H_0 - H_\infty = 1\frac{3}{4} \times 10^4 - 7047, H_0' - H_\infty' = 500 - 186.$

The estimate of β from the first series then yields -0.266, and from the second, -0.159, with a mean for the two series of -0.213. As a direct check on the order of magnitude and significance of this estimate, we may utilize the mean

values of U for the longer velocity runs noted at the end of the last section. These give (if we assume any such run all belongs to the same cluster) estimates of β of -0.141 ± 0.170 (1st series), -0.243 ± 0.210 (2nd series), or a (weighted) mean of -0.182 ± 0.132. However, the significance of this relation seems much more definite from the cross-spectrum (either from the overwhelming preponderance of negative values for G_p at the lower end of the frequency range, or from their individual significance if the covariance $I_{12}(\omega_p)$ is converted to a correlation).

With the estimate of β obtained from the cross-spectrum for each series, we may revise our estimates of the within-cluster velocity correlation. We now write this as

$$\rho_0 = \rho(1 - \rho_1^2) + \rho_1^2, \tag{7.3}$$

where ρ_1 is the correlation corresponding to β. Using the theoretical value of $(14/3)^{1/2}$ for σ_r^2 when $c = 1/9$, $\alpha = 2/3$, we have ρ estimated to be -0.127 (1st series) and -0.082 (2nd series). Making use of the expression for H_0 given in the appendix, we obtain estimates of ρ (with $v = \sigma_u^2(1 - \rho_1^2)$) of 0.815 (1st series) and 0.879 (2nd series), or finally of ρ_0 of 0.818 (1st series) and 0.880 (2nd series). These estimates are likely to have less bias, but to contain more error fluctuations than the previous estimates assuming $\beta = 0$, namely, 0.76 and 0.78. It is perhaps worth noting that with these somewhat higher correlations the expected serial correlations for the velocities are 0.20 and 0.22, a little nearer the observed values.

The interpretation of the above spectral analysis of traffic data in terms of a clustering model is not of course unique or exhaustive. An alternative (and not necessarily incompatible) interpretation in terms of flow density relations will be discussed elsewhere.

I am very much indebted to Stig Edholm, Head of the Traffic Department, National Road Research Institute, Stockholm, for sending me the traffic data for the second example. I am also much indebted to David Walley for his invaluable help in providing the computer programs and arranging the computations for these "extended" spectral analyses.

APPENDIX

Evaluation of the spectrum of $dM(x)$ for the clustering model. Equation (3.14) has the form

$$E_r\{(A + Br + Cr^2)(f_r + 2f_{r-1} + \cdots + rf_1)\}. \tag{A.1}$$

If we write $L(\psi)$ for the Laplace transform of f_1, we have for

$$\int_{-\infty}^{\infty} e^{-iz\omega_\nu}(z)\, dz \tag{A.2}$$

the expression

$$G(-i\omega) + G(i\omega), \tag{A.3}$$

where $G(\psi)$ is evaluated (if $v_r = v$, constant; otherwise the term in ρ is modified) as

$$\begin{aligned}(A.4)\quad & A\{LE(r) + L^2E'(r-1) + L^3E'(r-2) + \cdots\} \\ & + B\{LE(r^2) + L^2E'\{r(r-1)\} + L^3E'\{r(r-2)\} + \cdots\} \\ & + C\{LE(r^3) + L^2E'\{r^2(r-1)\} + L^3E'\{r^2(r-2)\} + \cdots\},\end{aligned}$$

E' denoting expectation over all nonnegative values. Now

$$\begin{aligned}E'\{r(r-s)\} &= E'\{(r-s)^2\} + sE'(r-s), \\ E'\{r^2(r-s)\} &= E'\{(r-s)^3\} + 2sE'\{(r-s)^2\} + s^2E'(r-s).\end{aligned} \tag{A.5}$$

Further, for the modified geometric distribution,

$$E'\{(r-s)^2\} = \alpha^s E'(r^2), \qquad E'\{(r-s)^3\} = \alpha^s E'\{r^3\},$$

$$E(r^2) = \frac{c(1+\alpha)}{(1-\alpha)^2}, \qquad E(r^3) = \frac{c(1+4\alpha+\alpha^2)}{(1-\alpha)^3}$$

$$\begin{aligned}(A.6)\quad G = {} & \frac{AcL}{(1-\alpha)(1-\alpha L)} + \frac{BcL(1+\alpha-2\alpha^2L)}{(1-\alpha)^2(1-\alpha L)^2} \\ & + \frac{CcL[(1+4\alpha+\alpha^2)(1-\alpha L)^2 + 2\alpha L(1-\alpha L)(1-\alpha^2) + \alpha L(1-\alpha)^2(1+\alpha L)]}{(1-\alpha)^3(1-\alpha L)^3},\end{aligned}$$

where further

$$\begin{aligned}A &= \lambda_C\left\{1 + \xi\xi^*\rho v + \frac{\beta^2c^2\xi\xi^*}{(1-\alpha)^2} - \frac{\beta c(\xi+\xi^*)}{(1-\alpha)^2}\right\}, \\ B &= -\lambda_C\left\{\frac{2\beta^2\xi\xi^*c}{1-\alpha} - \beta(\xi+\xi^*)\right\}, \\ C &= \lambda_C\beta^2\xi\xi^*.\end{aligned} \tag{A.7}$$

Rearranging terms, we may write this finally as

$$\begin{aligned}(A.8)\quad & \frac{\lambda_C(1+\xi\xi^*\rho v)cL}{(1-\alpha)(1-\alpha L)} \\ & + \frac{\lambda_C\beta^2\xi\xi^*cL}{(1-\alpha)^3(1-\alpha L)}\left\{c^2 - \frac{2c(1+\alpha-2\alpha^2L)}{1-\alpha L}\right. \\ & \quad \left. + \frac{(1+4\alpha+\alpha^2)(1-\alpha L)^2 + 2\alpha L(1-\alpha L)(1-\alpha^2) + \alpha L(1-\alpha)^2(1+\alpha L)}{(1-\alpha L)^2}\right\} \\ & + \frac{\beta\lambda_C(\xi+\xi^*)cL}{(1-\alpha)^2(1-\alpha L)}\left\{\frac{1+\alpha-2\alpha^2L}{1-\alpha L} - c\right\}.\end{aligned}$$

It is of interest to examine the relative values at $\omega = 0+$ $(L = 1)$ for the particular case $c = 1/9$, $\alpha = 2/3$. We obtain, with $\lambda_c = 3\lambda = 3/4$,

$$\frac{3}{4}(1 + \xi\xi^*\rho v) + 46\frac{5}{6}\beta^2\xi\xi^* + 5\beta(\xi + \xi^*). \tag{A.9}$$

REFERENCES

[1] M. S. BARTLETT, "Some problems associated with random velocity," *Publ. Inst. Statist. Univ. Paris*, Vol. 6 (1957), pp. 261–270.

[2] ———, "The spectral analysis of point processes," *J. Roy. Statist. Soc. Ser. B*, Vol. 25 (1963), pp. 264–296.

[3] ———, "The spectral analysis of two-dimensional point processes," *Biometrika*, Vol. 51 (1964), pp. 299–311.

[4] ———, *Introduction to Stochastic Processes*, Cambridge, Cambridge University Press, 1966 (2nd ed.).

[5] M. G. KENDALL and P. A. P. MORAN, *Geometrical Probability*, London, Griffin, 1963.