Bulletin of the Seismological Society of America, Vol 71, No. 4, pp. 1105–1130, August 1981

TELESEISMIC LOCATION TECHNIQUES AND THEIR APPLICATION TO EARTHQUAKE CLUSTERS IN THE SOUTH-CENTRAL PACIFIC

BY THOMAS H. JORDAN AND KEITH A. SVERDRUP

ABSTRACT

Improved methods for single- and multiple-event hypocenter determinations are developed and applied to the problem of locating earthquake clusters in the South-Central Pacific Ocean. Bayesian statistical methods are used to incorporate *a priori* information about arrival-time variance into the derivation of hypocenter confidence ellipsoids, permitting a more realistic calculation of critical parameters in the case where the number of stations is small. The diagonal elements of certain projection operators, called "data importances" by Minster *et al.* (1974), are used to evaluate network balance. The *hypocentroid* of an event cluster is defined to be the average location of events within the cluster, and the deviations of individual hypocenters from the hypocentroid are called *cluster vectors*. The problem of estimating the cluster vectors can be decoupled from the problem of estimating the hypocentroid by a simple but fundamental mathematical result, here termed the hypocentroidal decomposition theorem. The algorithm based on this analysis appears to have many advantages over other published methods for multiple-event location, both in its efficient use of available information and its computational speed. The application of this method to three clusters of shallow intraplate seismicity in the South-Central Pacific, designated Regions A, B, and C, demonstrates that the seismicity within each cluster is very localized; the rms lengths of the cluster vectors for each group of epicenters are estimated to be only 9, 6, and 12 km, respectively. Estimates of the epicentroids are

Region A	**(4 events)**	**7.40 ± 0.023°S**	**148.21 ± 0.026°W**
Region B	**(3 events)**	**18.40 ± 0.018°S**	**132.87 ± 0.018°W**
Region C	**(7 events)**	**20.76 ± 0.015°S**	**126.95 ± 0.019°W.**

INTRODUCTION

The intraplate seismicity of the South-Central Pacific Ocean has been mapped by Talandier and Kuster (1976) and Okal *et al.* (1980) using the 15-station French Polynesian Network and global station arrays. During the recording period 1965–1979 low-magnitude activity was diffusely distributed throughout the area comprised by these studies, but most of the seismicity, including 13 of the 17 earthquakes with $m_b \geqq 4.9$, was concentrated at three discrete localities, designated Regions A, B, and C (Figure 1). Over the past 15 yr these localities have been the most intense centers of seismicity wthin the Pacific plate interior, excluding Hawaii, and together have accounted for more than 90 percent of the seismic energy release in the South-Central Pacific. Focal mechanisms obtained in Regions A, B, and C have nearly horizontal, NW-trending compressional axes oriented approximately parallel to the direction of Pacific plate motion, an observation which led Okal *et al.* (1980) to conclude that the mechanisms are indicative of a regional tectonic stress field, rather than locally disturbed stress patterns.

A regional stress field with small lateral variation does not account for the tendency of earthquakes in the South-Central Pacific to be spatially localized, however. As Okal *et al.* (1980) have noted, this conspicuous aspect of the seismicity suggests the existence of inhomogeneities within the oceanic plate that act to

concentrate stress or to weaken the lithosphere. The characterization of these inhomogeneities remains an important problem of oceanic intraplate tectonics.

Accurate location of the seismicity is essential for this purpose, especially to permit the correlation of epicenters with potentially diagnostic bathymetric features. Based on the times of water multiples (*pnwP* phases) from events in Regions A, B, and C, Okal *et al.* (1980) concluded that the hypocenters are very shallow, probably lying within the oceanic crust. Using this information they relocated some of the larger events by standard location procedures.

This paper reexamines various aspects of the location problem, with an emphasis on techniques for constraining the "epicentroids" of seismic clusters as well as the relative displacements of events within a cluster. Several improvements to standard location algorithms and hypothesis-testing procedures are devised and implemented in this context.

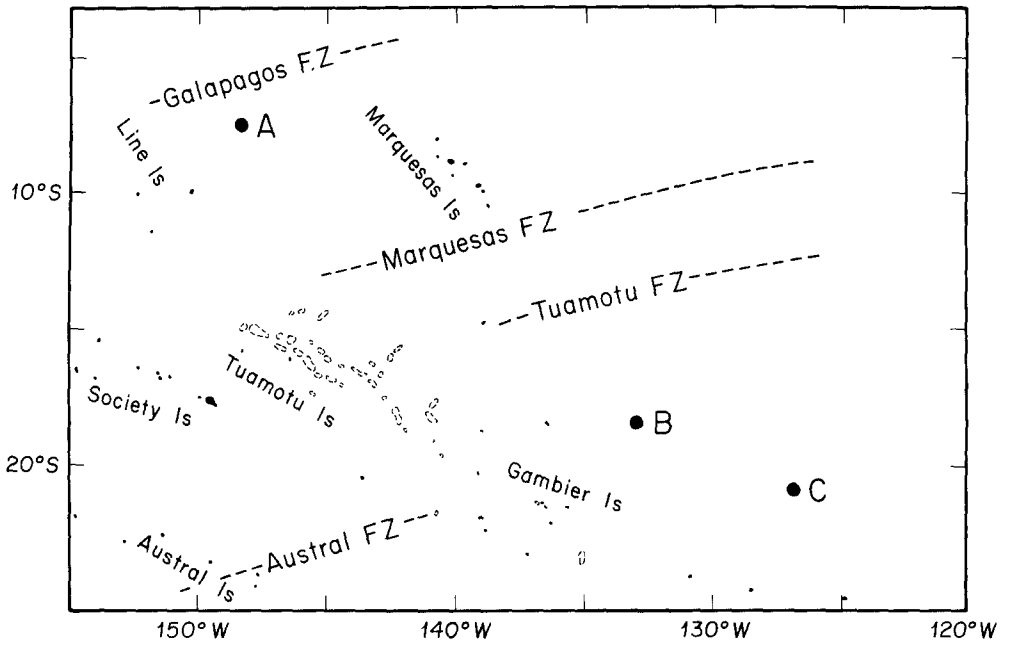

FIG 1 Map of the South-Central Pacific Ocean showing the locations of Regions A, B, and C.

ARRIVAL-TIME DATA

The set of 14 events selected for study comprised 4 earthquakes from Region A, 3 from B, and 7 from C (Table 1). Two of the Region C events (C3 and C4) do not appear in the catalog of Okal *et al.* (1980), but have been recently identified by the International Seismological Centre (ISC) during their routine searches of reported times (R. D. Adams, private communication, 1980). Each event was large enough ($m_b \geqq 4.7$) to be recorded by 20 or more stations at teleseismic distances. *P*-wave arrival times were read from short-period vertical-component seismograms by the authors or obtained from ISC or United States Geological Survey (USGS) bulletins. Although good P_n phases were recorded for most events on stations of the French Polynesian Network, these first arrivals were excluded from the analysis because their concentration at certain azimuths and their anisotropic propagation speeds

TABLE 1

SINGLE-EVENT LOCATIONS OF EARTHQUAKES USED IN THIS STUDY*

Region	Event	m_b (ISC)	No of Data	Date (y/m/d)	Origin Time (h m s)	s d † (sec)	Latitude (°N)	s d † (°)	Longitude (°E)	s d † (°)	ζ
A	1	4 9	42	68/07/29	02:45.45 6	0.25	−7 47	0.048	−148 16	0 058	1.10
	2	5.0	42	69/08/06	17:15 40.4	0.21	−7.40	0.047	−148 19	0.048	0.82
	3	4 9	37	73/01/19	07 36:34 0	0 18	−7 36	0.041	−148 24	0 044	0 85
	4	4.8	27	73/01/19	15.26:26.9	0.26	−7 31	0.055	−148.27	0.062	0 95
B	1	5 4	73	65/03/06	11:10 52.2	0.17	−18.40	0 032	−132.85	0.039	0.86
	2	5.0	40	66/09/18	06 40·35 8	0 19	−18.40	0 031	−132.86	0 060	0 79
	3	4 9	43	75/05/25	14.16:32 8	0 17	−18.40	0.037	−132.92	0 055	0 80
C	1	5.0	45	77/10/31	08·19 13.8	0 24	−20 72	0 040	−126 95	0 053	0 83
	2	5 5	112	78/01/05	03·23:25.3	0 12	−20 80	0 023	−126 94	0.026	0 96
	3	4 7	20	78/02/19	23 57 59.6	0 34	−20 83	0.062	−126.70	0 168	0.73
	4	4 8	36	78/03/11	03 32 10 1	0 29	−20 89	0.049	−126.96	0 061	0 86
	5	5 1‡	36	78/07/13	18 04·18 2	0 38	−20 78	0 070	−127 09	0 071	0.93
	6	5.2‡	55	78/07/25	07 54·07.8	0 23	−20.76	0 041	−126.97	0.057	0.95
	7	5 0‡	32	79/02/26	06·31·53.2	0 53	−20 94	0.104	−126.74	0 132	0.93

* Depths fixed at 10 km

† Standard deviations of marginal distributions computed with $K = 8$

‡ Magnitudes from National Earthquake Information Service

(Talandier and Bouchon, 1979) could bias the locations. *PKP* phases and arrivals with very large residuals (>5 sec in magnitude) were also excluded.

The resulting distributions of arrival-time data are displayed in Figure 2. Although the stations recording the earthquakes tend to be concentrated in the northeastern (North America) and southwestern (Australia) quadrants, the azimuthal control is reasonably good. Relatively few stations recorded all events in each cluster, however.

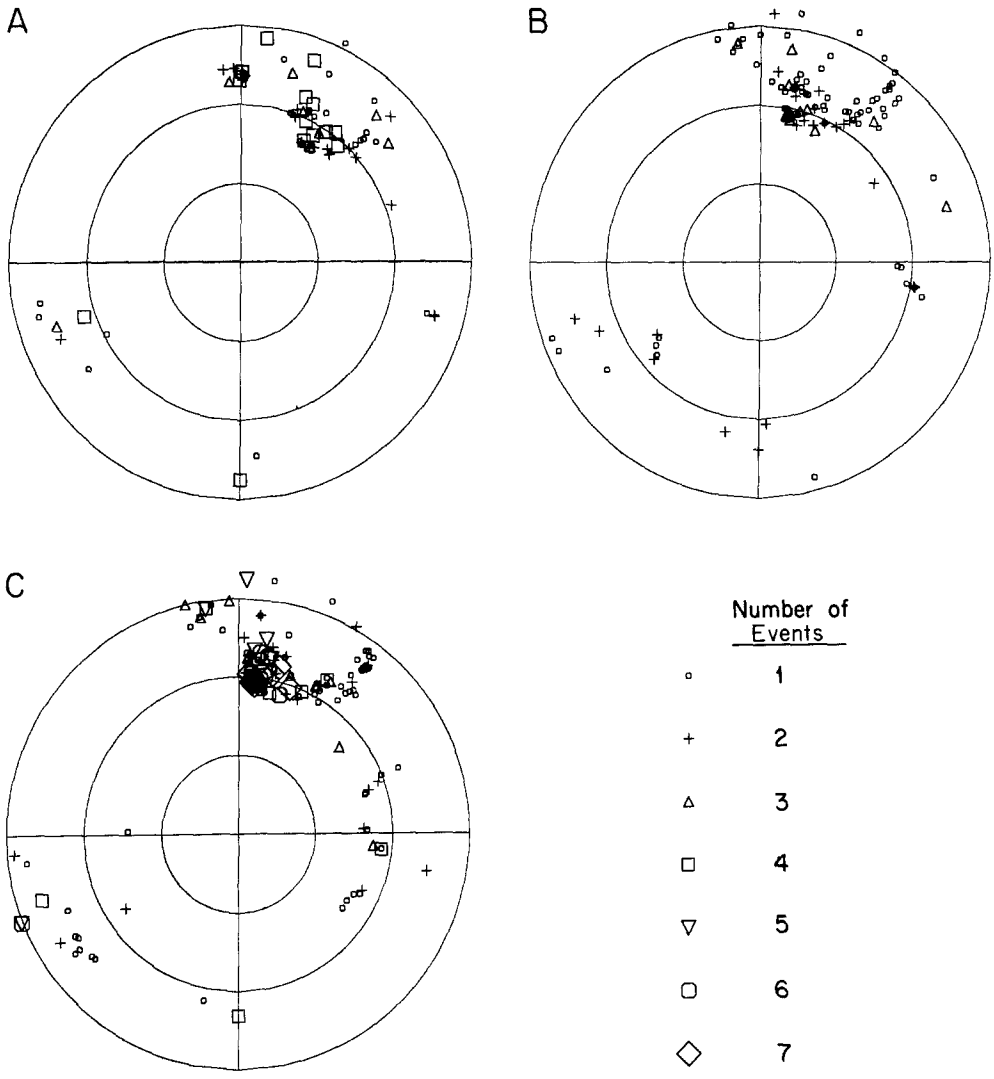

FIG 2 Distributions of arrival-time data for Regions A, B, and C plotted on azimuthal equidistant projections centered on the epicentroids; north is up, and circles are drawn at intervals of 30 geocentric degrees.

A priori estimates of the observational uncertainty were assigned to individual readings. High-quality picks made by the authors, which included about 50 per cent of the data for A1 to A4, 20 per cent for C2, and 10 per cent for C6, were generally assigned standard errors of ±0.7 sec, whereas low-quality readings and bulletin times were assigned errors of ±1.0 sec. These values are consistent with our previous

experience and the estimates of other authors (e.g., Freedman, 1967; Evernden, 1969a). An assessment of the uncertainty in these *a priori* values of observational error and its impact on location confidence ellipsoids is a subject of our investigation.

Single-Event Location

Geiger's (1910) iterative, least-squares method and its various extensions and modern improvements (e.g., Flinn, 1965; Buland, 1976) are the basis of nearly all numerical algorithms for earthquake location, including those discussed here (see, however, Lomnitz, 1977). A point seismic source is specified by a space-time vector $\boldsymbol{x}$ of dimension $M \leqq 4$, the inequality pertaining whenever certain components, such as depth, are fixed throughout the calculation. This location vector is to be constrained by an N-component vector of observed arrival times $\boldsymbol{t}^O$. Formally, $\boldsymbol{x} \in E^M$ and $\boldsymbol{t}^O \in E^N$, where E^N is the Euclidian space of dimension N. Since the number of observation points N is usually large compared to M, the problem is over-determined, although it may still be ill-conditioned. From an initial estimate $\boldsymbol{x}_0$, an arrival-time vector $\bar{\boldsymbol{t}}(\boldsymbol{x}_0)$ is calculated by tracing rays through an average radial earth structure; the overbar indicates the assumption of spherical symmetry and in this paper always refers to the theoretical times of Herrin (1968). By expanding $\bar{\boldsymbol{t}}$ about $\boldsymbol{x}_0$ we can write

$$\overline{\Delta \boldsymbol{t}} \equiv \bar{\boldsymbol{t}}(\boldsymbol{x}) - \bar{\boldsymbol{t}}(\boldsymbol{x}_0) = \boldsymbol{A} \cdot \Delta \boldsymbol{x} + \mathcal{O}[|\Delta \boldsymbol{x}|^2] \tag{1}$$

where $\Delta \boldsymbol{x} \equiv \boldsymbol{x} - \boldsymbol{x}_0$ and $\boldsymbol{A}$ is an $N \times M$ matrix of partial derivatives evaluated at $\boldsymbol{x}_0$. If rank $[\boldsymbol{A}] = M$, as shall be supposed, a refined estimate of $\boldsymbol{x}$ can be uniquely obtained from the residual vector $\Delta \boldsymbol{t}^O \equiv \boldsymbol{t}^O - \bar{\boldsymbol{t}}(\boldsymbol{x}_0)$ by minimizing a squared norm of the error vector

$$\boldsymbol{e}^O = \Delta \boldsymbol{t}^O - \boldsymbol{A} \cdot \Delta \boldsymbol{x}. \tag{2}$$

To account for the curvature of the $\bar{\boldsymbol{t}}(\boldsymbol{x})$ manifold, the estimate of $\Delta \boldsymbol{x}$ is added to $\boldsymbol{x}_0$ and the procedure is iterated to convergence, which is usually rapid in the teleseismic case.

Nonlinear effects have been discussed by Flinn (1965), Buland (1976), and Spence (1980), among others. In certain situations, especially for hypocenters determined by sparse local networks, the nonlinear aspects of the problem can be important. For teleseismic location refinements of the sort discussed here, however, where the hypocentral displacements and confidence-ellipse dimensions do not exceed 100 km, a linear analysis completely suffices (see, e.g., Spence, 1980). Therefore, terms of higher order than the first in equation (1) are ignored throughout our analysis.

Estimation procedure. The residual vector $\Delta \boldsymbol{t}^O$ is considered to be a sample of a stochastic process $\Delta \boldsymbol{t}$ expressible as the sum of two terms

$$\Delta \boldsymbol{t} = \overline{\Delta \boldsymbol{t}} + \boldsymbol{n}. \tag{3}$$

The first term is defined by the equation

$$\boldsymbol{A} \cdot \Delta \boldsymbol{x} = \overline{\Delta \boldsymbol{t}} \tag{4}$$

and the second is a noise process. The noise process is presumed to be normally

distributed with mean $\boldsymbol{r}$ and a variance matrix $\boldsymbol{V}_n$; i.e.,

$$\langle \boldsymbol{n} \rangle = \boldsymbol{r} \tag{5}$$

$$\langle (\boldsymbol{n} - \boldsymbol{r})(\boldsymbol{n} - \boldsymbol{r})^T \rangle = \boldsymbol{V}_n. \tag{6}$$

Here, as elsewhere, the angle brackets denote the expected value. The vector $\boldsymbol{r}$ includes the bias introduced into $\Delta \boldsymbol{t}^O$ by ellipticity, station elevation, instrumental group delays, station anomalies, etc., whereas $\boldsymbol{V}_n$ represents the dispersion caused by reading errors, timing errors, and other sources of random or quasi-random noise.

Estimates of $\boldsymbol{r}$ and $\boldsymbol{V}_n$ may be available, but neither is precisely known. A goal of the analysis is to account for the imprecision of this information in the estimation and hypothesis-testing procedures. Of course, statistical assumptions that are unjustified, or justified only by the lack of contradictory data, must be made in order to permit a feasible solution.

Let $\tilde{\boldsymbol{r}}$ be an estimate of the bias vector that includes an observed set of mean station anomalies. For simplicity $\tilde{\boldsymbol{r}}$ is taken to be a sample from a normally distributed random process with mean $\boldsymbol{r}$ and a diagonal variance matrix $\boldsymbol{V}_r$ whose nonzero elements are set equal to the sample variances of the station anomalies. To the extent that these conditions are correct, making the substitutions

$$\Delta \boldsymbol{t}^O - \tilde{\boldsymbol{r}} \rightarrow \Delta \boldsymbol{t}^O \tag{7}$$

$$\boldsymbol{V}_n + \boldsymbol{V}_r \rightarrow \boldsymbol{V}_n \tag{8}$$

redefines the noise process $\boldsymbol{n}$ to have zero mean. In so doing, the noise variance is increased in proportion to the variance assigned to the bias. The increase can be substantial for station anomalies computed from a small number of events. The station anomalies used here, however, are those of Poupinet (1979), derived from a very extensive set of ISC P-wave residuals; their formal standard errors never exceed 0.1 sec and therefore contribute insignificantly to the travel-time variances, although, to the extent that station anomalies do not completely account for the path anomalies appropriate to these particular events, some bias will remain.

Nevertheless, it is supposed for the moment that this procedure removes all bias from $\Delta \boldsymbol{t}^O$, so that

$$\langle \Delta \boldsymbol{t} \rangle = \overline{\Delta \boldsymbol{t}} \tag{9}$$

$$\langle (\Delta \boldsymbol{t} - \overline{\Delta \boldsymbol{t}})(\Delta \boldsymbol{t} - \overline{\Delta \boldsymbol{t}})^T \rangle = \boldsymbol{V}_n. \tag{10}$$

Furthermore, the variance matrix $\boldsymbol{V}_n$ is taken to be diagonal and, within a constant factor, to be known exactly; in other words, a diagonal matrix $\hat{\boldsymbol{V}}_n$ is specified such that

$$\boldsymbol{V}_n = \nu^2 \hat{\boldsymbol{V}}_n \tag{11}$$

for some (possibly unknown or poorly known) value of ν^2. This choice assumes the errors in the elements of $\Delta \boldsymbol{t}$ to be uncorrelated, and it fixes the ratios of their variances.

The classical solution to this inverse problem is to require the estimate to maximize a likelihood function proportional to $\exp[-\frac{1}{2}\boldsymbol{e}^O \cdot \boldsymbol{V}_n^{-1} \cdot \boldsymbol{e}^O]$. Under the

statistical assumptions stated here, the maximum-likelihood estimate can be obtained by minimizing with respect to $\Delta \boldsymbol{x}$ the squared norm of a scaled error process

$$\hat{\boldsymbol{e}} = \Delta \hat{\boldsymbol{t}} - \hat{\boldsymbol{A}} \cdot \Delta \boldsymbol{x} \tag{12}$$

for any element of its ensemble, where the carat denotes normalization by the "standard-deviation" matrix $\hat{\boldsymbol{V}}_n^{1/2}$

$$\Delta \hat{\boldsymbol{t}} \equiv \hat{\boldsymbol{V}}_n^{-1/2} \cdot \Delta \boldsymbol{t} \tag{13}$$

$$\hat{\boldsymbol{A}} \equiv \hat{\boldsymbol{V}}_n^{-1/2} \cdot \boldsymbol{A}. \tag{14}$$

This yields

$$\Delta \boldsymbol{x} = \hat{\boldsymbol{A}}^{\dagger} \cdot \Delta \hat{\boldsymbol{t}}. \tag{15}$$

The matrix operator $\hat{\boldsymbol{A}}^{\dagger}$ is the "generalized inverse" of $\hat{\boldsymbol{A}}$ (Penrose, 1955), which we compute by singular-value decomposition (SVD); a description of the SVD algorithm and its application in this context is given by Golub and Reinsch (1971).

The generalized inverse has the properties that $\hat{\boldsymbol{A}}^{\dagger} \cdot \hat{\boldsymbol{A}} = \boldsymbol{I}^{(M)}$ is the identity operator on E^M, and

$$\boldsymbol{P} = \hat{\boldsymbol{A}} \cdot \hat{\boldsymbol{A}}^{\dagger} \tag{16}$$

is the orthogonal projection operator that maps E^N onto the subspace spanned by the columns of $\hat{\boldsymbol{A}}$. The subspace is M dimensional if, as we assume, $\hat{\boldsymbol{A}}$ has full rank. Hence, the matrix

$$\boldsymbol{Q} = \boldsymbol{I}^{(N)} - \boldsymbol{P} \tag{17}$$

is the projection operator onto the $(N - M)$ dimensional null space of $\hat{\boldsymbol{A}}$. Using (15) in equation (12), we find

$$\boldsymbol{e} = \boldsymbol{Q} \cdot \Delta \hat{\boldsymbol{t}}. \tag{18}$$

Minster *et al.* (1974) call the diagonal elements of $\boldsymbol{P}$ *data importances*, because these quantities describe the weighting given to particular data in the computation of the least-squares estimate. Each data importance is a scalar lying between 0 and 1, and they are additive: the sum of all data importances equals the rank of $\hat{\boldsymbol{A}}$, in this case M. Other properties and examples of their applications are discussed by Minster *et al.* (1974) and Minster *et al.* (1977). As we shall see, data importances are useful in assessing earthquake location estimates in the South-Central Pacific.

The particular estimate of $\Delta \boldsymbol{x}$ available to the observer is $\Delta \tilde{\boldsymbol{x}}^O$, obtained by substituting $\Delta \hat{\boldsymbol{t}}^O$ into equation (15); under our statistical assumptions, it represents one realization of a normally distributed stochastic process $\Delta \tilde{\boldsymbol{x}}$, whose mean and variance matrix are easily shown to be

$$\langle \Delta \tilde{\boldsymbol{x}} \rangle = \Delta \boldsymbol{x} \tag{19}$$

$$\boldsymbol{V}_x \equiv \langle (\Delta \tilde{\boldsymbol{x}} - \Delta \boldsymbol{x})(\Delta \tilde{\boldsymbol{x}} - \Delta \boldsymbol{x})^T \rangle = \nu^2 \hat{\boldsymbol{V}}_x \tag{20}$$

where

$$\hat{\boldsymbol{V}}_x = (\hat{\boldsymbol{A}}^T \cdot \hat{\boldsymbol{A}})^{-1}. \tag{21}$$

Similarly, the moments of the normalized error vector are

$$\langle \hat{\boldsymbol{e}} \rangle = \boldsymbol{Q} \cdot \langle \Delta \hat{\boldsymbol{t}} \rangle = \boldsymbol{Q} \cdot \hat{\boldsymbol{A}} \cdot \Delta \boldsymbol{x} = 0 \tag{22}$$

$$\langle \hat{\boldsymbol{e}} \hat{\boldsymbol{e}}^T \rangle = \boldsymbol{Q} \cdot \langle \Delta \hat{\boldsymbol{t}} \Delta \hat{\boldsymbol{t}}^T \rangle \cdot \boldsymbol{Q} = \nu^2 \boldsymbol{Q}. \tag{23}$$

Its squared length $\hat{e}^2 = \hat{\boldsymbol{e}} \cdot \hat{\boldsymbol{e}}$ is χ^2 distributed and has the expected value

$$\langle e^2 \rangle = \nu^2 \operatorname{tr} \boldsymbol{Q} = \nu^2 (N - M). \tag{24}$$

Thus, the random variable $\hat{e}^2/\nu^2$ is χ^2 distributed with $N - M$ degrees of freedom. These results are, of course, just those of classical least-squares analysis.

Confidence ellipsoids. In constructing confidence ellipsoids about $\Delta \tilde{\boldsymbol{x}}^O$, the random variable of interest is

$$k^2 = (\Delta \boldsymbol{x} - \Delta \tilde{\boldsymbol{x}})^T \cdot \hat{\boldsymbol{V}}_x^{-1} \cdot (\Delta \boldsymbol{x} - \Delta \tilde{\boldsymbol{x}}). \tag{25}$$

A critical value κ_α^2 is sought for some $\alpha \in (0, 1)$ such that the probability of k^2 exceeding κ_α^2 is α. The set of all vectors $\Delta \boldsymbol{x}'$ that satisfy

$$(\Delta \boldsymbol{x}' - \Delta \tilde{\boldsymbol{x}}^O)^T \cdot \hat{\boldsymbol{V}}_x^{-1} \cdot (\Delta \boldsymbol{x}' - \Delta \tilde{\boldsymbol{x}}^O) \leqq \kappa_\alpha^2 \tag{26}$$

is called the $100(1 - \alpha)$ per cent confidence ellipsoid of $\Delta \tilde{\boldsymbol{x}}^O$, and, for any vector $\Delta \boldsymbol{x}''$ that does not obey (26), the hypothesis that $\Delta \boldsymbol{x}'' = \Delta \boldsymbol{x}$ can be rejected at the $100(1 - \alpha)$ per cent confidence level.

Equations (19) and (20) yield

$$\langle k^2 \rangle = \hat{\boldsymbol{V}}_x^{-1} : \boldsymbol{V}_x = \nu^2 M \tag{27}$$

from which it is seen that k^2/ν^2 is χ^2 distributed with M degrees of freedom. Suppose it is known *a priori* that $\nu^2 = 1$; then, $\boldsymbol{V}_n = \hat{\boldsymbol{V}}_n$ and

$$\kappa_\alpha^2 = \chi_\alpha^2[M] \tag{28}$$

where $\chi_\alpha^2[M]$ is the critical value of χ^2 for M degrees of freedom. In standard location algorithms, however, the assumption that $\nu^2 = 1$ is not made, but rather the normalized sample variance

$$\hat{s}^2 = |\hat{\boldsymbol{e}}^O|^2/(N - M) \tag{29}$$

is used to estimate ν^2 *a posteriori* (e.g., Flinn, 1965). The ratio $(N - M)k^2/M\hat{e}^2$ is F-distributed with M and $N - M$ degrees of freedom, so, letting $F_\alpha[M, N - M]$ be the corresponding critical value, we obtain

$$\kappa_\alpha^2 = M\hat{s}^2 F_\alpha[M, N - M]. \tag{30}$$

This is the usual expression for regression estimates.

Evernden (1969a) has criticized the application of equation (30) to the teleseismic location problem on the grounds that, as N becomes small, the confidence ellipsoids become unrealistically large. Instead, he advocates the use of equation (28). Evernden's criticism is certainly justified; the critical value given by (30) diverges as $N \rightarrow M$ because no prior information is assumed to exist regarding ν^2, which is untrue. His remedy is not justified, however; the value given by (28) assumes $\boldsymbol{V}_n$ to be perfectly known, which is also untrue.

A more satisfactory solution to this long-standing problem can be obtained by adopting the Bayesian technique of imposing a prior distribution on ν^2 to describe how well the noise variance matrix is thought to be known before any estimate of $\boldsymbol{\Delta x}$ is made. Our procedure shall be outlined without formal derivation; general descriptions of the Bayesian methodology on which the analysis is based are given by Jeffreys (1961) and Cox and Hinkley (1974). The results are simple, however, and can be heuristically motivated by supposing that ν^2 has been previously estimated by a normalized sample variance $\hat{s}_0^2$ from an experiment with K degrees of freedom; then, an estimate of ν^2 "updated" with the new information contained in $\boldsymbol{\Delta t}^O$ is

$$\frac{K\hat{s}_0^2 + |\hat{\boldsymbol{e}}^O|^2}{K + N - M}. \tag{31}$$

In such a case, it would seem reasonable to replace $\hat{s}^2$ by (31) and $N - M$ by $K + N - M$ in equation (30).

The Bayesian arguments for this procedure begin by supposing that $\hat{\boldsymbol{V}}_n$ is the best estimate of $\boldsymbol{V}_n$ constructed from previously available information, so that, *a priori*, $\nu^{-2}\boldsymbol{V}_n$ can be considered a stochastic process with an expected value of $\hat{\boldsymbol{V}}_n$. A natural choice for the prior distribution is to take K/ν^2 to be a χ^2 random variable with K degrees of freedom; the parameter K then controls the variance of ν^{-2} about unity

$$\langle (\nu^{-2} - 1)^2 \rangle = 2/K. \tag{32}$$

The prior density function imposed on the reciprocal ν^2/K is the so-called inverse-γ distribution (Cox and Hinkley, 1974, p. 371) with K degrees of freedom and an expected value equal to $(K-2)^{-1}$. For the problem at hand, the inverse-γ distribution is conjugate to the normal distribution assumed for $\boldsymbol{\Delta t}$; therefore, the posterior distribution of ν^2 is itself an inverse-γ distribution. Using Bayes's theorem it can be easily shown that, conditional on the event $\boldsymbol{\Delta t} = \boldsymbol{\Delta t}^O$, the variable $\nu^2/(K + |\hat{\boldsymbol{e}}^O|^2)$ is inverse-γ distributed with $K + N - M$ degrees of freedom and an expected value of $(K + N - M - 2)^{-1}$. Hence, its reciprocal is χ^2 distributed with $K + N - M$ degrees of freedom, and we obtain

$$\kappa_\alpha^2 = M\hat{s}^2 F_\alpha[M, K + N - M] \tag{33}$$

where

$$\hat{s}^2 = \frac{K + |\hat{\boldsymbol{e}}^O|^2}{K + N - M}. \tag{34}$$

In these expressions the parameter K is allowed to vary continuously on the

interval $[0, \infty)$. $K = 0$ corresponds to the assumption of *no* prior information about ν^2, in which case the expressions for $\hat{s}^2$ and ${\kappa_\alpha}^2$ reduce to equations (29) and (30), respectively. As $K \to \infty$, $s \to 1$, and ${\kappa_\alpha}^2$ reduces to (28), corresponding to the assumption of *exact* prior information; *viz.*, $\nu^2 = 1$. Thus, the procedures of Flinn (1965) and Evernden (1969a) appear as limiting cases.

In theory, if the noise processes represented by $\boldsymbol{n}$ depended only on station characteristics or other stable parameters and were stationary with respect to source location and origin time, then K could be specified as the total number of degrees of freedom in all previous experiments used to derive $\hat{\boldsymbol{V}}_n$, which, after some time, could be very large. In practice, however, allowances must be made for the various types of nonstationarity that cause $\boldsymbol{V}_n$ to vary from experiment to experiment: lateral heterogeneity, source magnitude, seismic noise modulation, and so forth.

We shall therefore adopt a "subjectivist's" strategy and specify K according to an assumed uncertainty in the standard deviation variable $\xi \equiv \nu^{-1}$, whose square has the χ^2 prior distribution given by (32). Using formulae listed by Zelen and Severo (1972, p. 943), we find

$$\langle \xi \rangle = \left(1 - \frac{1}{2K}\right)^{1/2} \left[1 + \frac{1}{16K(K-1)}\right] + \mathbb{O}(K^{-4}) \tag{35}$$

$$\langle (\xi - \langle \xi \rangle)^2 \rangle = \frac{1}{2K} - \frac{1}{8K^2} - \frac{1}{16K^3} + \mathbb{O}(K^{-4}). \tag{36}$$

The standard deviation of ξ,

$$\text{s.d.}[\xi] = \langle (\xi - \langle \xi \rangle)^2 \rangle^{1/2} \tag{37}$$

is a convenient measure of the uncertainty in $\hat{\boldsymbol{V}}_n$; K is plotted as a function of s.d.$[\xi]$ in Figure 3.

The s.d.$[\xi]$ could itself be formally estimated in conjunction with $\hat{\boldsymbol{V}}_n$ from an ensemble of location experiments, but to our knowledge no results have been published. However, based on our experience and that of other authors (e.g., Freedman, 1967; Evernden, 1969a), we are willing to assert that 10 per cent $<$ s.d.$[\xi]<$ 40 per cent. This subjective assessment implies $50 > K > 2.8$. For the purposes of calculation we have adopted (and recommend for standard usage in teleseismic location) the intermediate value of s.d.$[\xi] = 25$ per cent, yielding $K = 8$.

Table 2 illustrates the effect of this choice on the average size of 95 per cent confidence ellipsoids. The parameter ${\kappa_\alpha}^2/\hat{s}^2$ (which scales as cross-sectional area) is listed for $M = 3$ and various values N for three cases: $K = 8$, and the two limiting cases $K = 0$ (completely *a posteriori*) and $K = \infty$ (completely *a priori*). Clearly, incorporation of prior information about the probable range of ξ, even the relatively low-grade information implied by $K = 8$, significantly reduces the expected size of the confidence ellipsoids relative to the standard regression method when the number of stations is small ($N \leqq K + M$), thus overcoming the objections of Evernden (1969a) to the use of F-statistics in the teleseismic location problem. Of course, when the number of stations is large ($N \gg K + M$), the information contained in $\Delta \boldsymbol{t}^O$ dominates over the prior information, and the difference between the Bayesian and regression procedures is very small.

Application to Regions A, B, and C. In the application of these procedures to earthquakes in the South-Central Pacific, the hypocentral depths were fixed at 10

km, in accordance with Okal *et al.*'s (1980) analysis of *pnwP* phases; hence, $M = 3$. The locations derived from equation (15) are listed in Table 1, and the epicenters are plotted on Figure 4. The standard deviations in Table 1 and the 95 per cent confidence ellipses on the figures correspond to marginal distributions with $M' = 1$ and $M' = 2$, respectively, for which the variance matrix V_x' is the appropriate $M' \times M'$ submatrix of V_x and the critical parameter is

$$\kappa_\alpha'^2 = M'\hat{s}^2 F_\alpha(M', K + N - M). \tag{38}$$

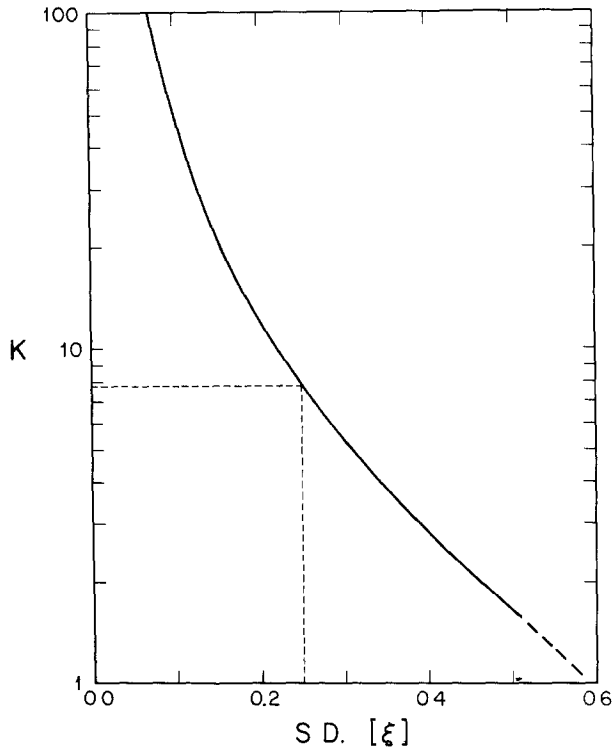

FIG 3. The Bayesian parameter K plotted against the standard deviation of ξ, calculated from equation (36) The choice of s.d.[ξ] = 25 per cent implies $K \simeq 8$, which is the value used in the single-event locations

TABLE 2

VALUES OF $\kappa_\alpha^2/\hat{s}^2 = MF_\alpha(M, K + N - M)$ FOR $\alpha = 0.05$ AND $M = 3$

N	$K = 0$	$K = 8^*$	$K = \infty$
5	57 48	11 13	7.80
10	13 05	9 87	7.80
15	10 47	9 30	7 80
20	9 60	8 97	7 80
30	8 88	8.64	7 80
50	8 44	8 34	7.80
100	8 13	8 10	7 80
∞	7 80	7 80	7.80

* Recommended value for teleseismic location

(A discussion of marginal distributions in the context of the location problem is given by Flinn, 1965.) The value of $K = 8$ was assumed, but, because N equals 20 or greater for every event, the differences in the critical parameters for $K = 0$ and $K = 8$ are slight (Table 2).

The fit to the arrival-time data, as measured by the parameter $\hat{s}$, conforms to our *a priori* expectations: the individual values of $\hat{s}$ range from 0.73 to 1.10, with a mean of 0.88 (Table 1), commensurate with the value $\xi \simeq 1 \pm 0.25$ implied by $K = 8$. Only four residuals had magnitudes exceeding 3 sec out of a total data set of 640, and in all cases these were associated with very low data importances ($\leqq 0.04$).

Data importances are useful not only in assessing the impact of possibly spurious readings, but also in monitoring "network balance." If too much importance is concentrated at a particular azimuth, the location is particularly susceptible to bias introduced by uncorrected path anomalies. This is a possible danger for Region A

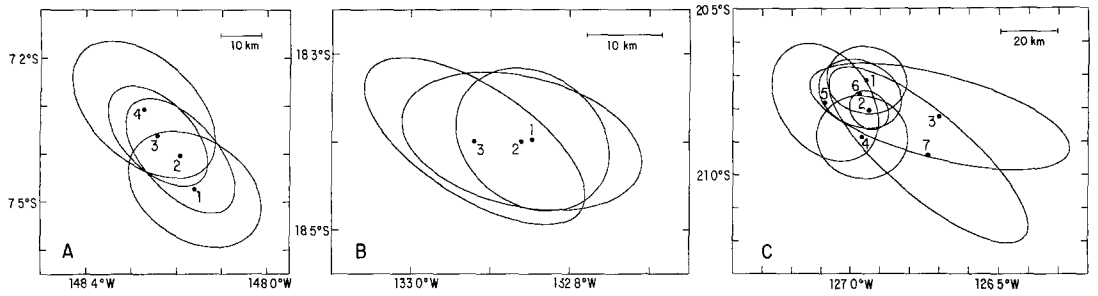

FIG 4 Single-event epicenters and 95 per cent confidence ellipsoids for earthquakes in Regions A, B, and C ($h = 10$ km); event numbers correspond to Table 1.

TABLE 3

DISTRIBUTION OF STATIONS AND CUMULATIVE DATA IMPORTANCES FOR SINGLE-EVENT LOCATIONS IN REGION A BY QUADRANTS

Event	Number of Stations				Cumulative Data Importance			
	NE	SE	SW	NW	NE	SE	SW	NW
A1	34	4	3	1	1.38	0 72	0.85	0.05
A2	36	0	4	2	1 62	0	1.26	0.12
A3	25	2	6	4	1 32	0 31	1.14	0.23
A4	19	0	4	4	1 50	0	1.23	0.27

events, for example, since all but a small fraction of the reporting stations are within a narrow range of azimuths in the northeastern quadrant (Figure 2). Fortunately, the azimuthal distribution of data importance is not nearly so lopsided as the station distribution (Table 3). The SE and NW quadrants are not well represented in either distribution, but there is a reasonably good balance in the partitioning of data importance between the NE and SW quadrants.

The network balance for events in Regions B and C is generally as good as or better than that for Region A, with the exception of event C7 for which 28 of 32 stations are concentrated in North America. For this event the cumulative data importance in the NE quadrant is very high (1.97). The poor station distribution, in particular the lack of control in the SE and NW quadrants, is also reflected in the large 95 per cent confidence ellipse associated with the C7 epicenter and in the NW-

SE trend of its major axis. All of the other Region C locations except C5 lie within this confidence ellipse, so the displacement of C7 to the SE side of the cluster may not be significant.

In fact, there is generally substantial overlap among the 95 per cent confidence ellipses within all three clusters, suggesting that the source regions for these events may be very small. Further quantification of this hypothesis is the primary goal of the next section.

Multiple-Event Locations

Multiple-event location procedures are founded on the observation that the noise contaminating travel times from a set of nearby earthquakes or explosions tends to be strongly correlated; in particular, the error introduced by incorrect assumptions regarding earth structure (e.g., spherical symmetry) has a nearly constant value for times measured at the same station. Since path-correlated noise of this type dominates the sample standard deviations computed for single-event locations (Freedman, 1967), the relative locations of events within a seismic cluster can be improved by taking these correlations into account. Various algorithms have been devised for this purpose (Freedman, 1967; Douglas, 1967; Evernden, 1969b; Dewey, 1972; Ansell and Smith, 1975; Veith, 1975; Spence, 1980). Behind all such algorithms is a simple yet fundamental mathematical result, termed here the hypocentroidal decomposition theorem. Although some of its implications are obvious and have been clearly recognized by previous authors, the theorem itself, to our knowledge, has never been precisely stated or formally proved. We do this below. We then formulate a multiple-event location algorithm based on this analysis that is efficient in both its use of available information and its computational speed, and we apply it to events in Regions A, B, and C. The algorithm appears to have many advantages over other published methods for multiple-event location.

In setting up the problem it will be worthwhile to develop a compact notation. Let us suppose there are P closely grouped events to be analyzed whose (unknown) location vectors relative to a reference point $\boldsymbol{x}_0$ are $\{\Delta \boldsymbol{x}_p \in E^M; p = 1, 2, \cdots, P\}$, and let each be constrained by N_p arrival-time observations contained in a residual vector $\Delta \boldsymbol{t}_p{}^O \in E^{N_p}$. Because a common space-time reference point has been chosen for all P events, the components of $\Delta \boldsymbol{t}_p{}^O$ and the origin-time component of $\Delta \boldsymbol{x}_p$ may share a large additive constant; this is of no consequence in the analysis, since the relationship between origin time and residual is strictly linear. The total number of unknown parameters is $M_T = MP \leqq 4P$, and the total number of data is $N_T = \Sigma \mid_{p=1}^{P} N_p$. As in the case of single-event location, $\Delta \boldsymbol{t}_p{}^O$ is considered to be a sample of stochastic process $\Delta \boldsymbol{t}_p = \overline{\Delta \boldsymbol{t}}_p + \boldsymbol{n}_p$, where $\overline{\Delta \boldsymbol{t}}_p$ is the arrival-time difference computed from a spherically symmetric earth structure and $\boldsymbol{n}_p$ is a Gaussian process with a mean $\boldsymbol{r}_p$ and a variance matrix $\boldsymbol{V}_{n_p}$. Again it is assumed that the spatial components of $\Delta \boldsymbol{x}_p$ are small enough so that the linear relationship

$$\boldsymbol{A}_p \cdot \Delta \boldsymbol{x}_p = \overline{\Delta \boldsymbol{t}}_p \tag{39}$$

is an adequate approximation, where the partial derivatives in the $N_p \times M$ matrices $\boldsymbol{A}_p$ are all evaluated at the common reference point $\boldsymbol{x}_0$.

Suppose the set of vectors $\{\Delta \boldsymbol{t}_p{}^O\}$ consists of residuals observed at a set of Q stations. Then, $N_p \leqq Q$ for all $p = 1, 2, \cdots, P$, and each matrix $\boldsymbol{A}_p$ consists of a subset of row vectors from a $Q \times M$ matrix $\boldsymbol{A}_0$ defined such that $\boldsymbol{A}_0 \cdot \Delta \boldsymbol{x}_p$ is the Q-component vector of theoretical residuals at all Q stations and contains $\overline{\Delta \boldsymbol{t}}_p$ as a

subset; in other words, $\boldsymbol{A}_0$ comprises all of the partial derivatives needed at $\boldsymbol{x}_0$. It will be convenient to write

$$\boldsymbol{A}_p = \boldsymbol{B}_p \cdot \boldsymbol{A}_0 \tag{40}$$

for some matrix $\boldsymbol{B}_p$. By definition, $\boldsymbol{B}_p$ is just the $N_p \times Q$ matrix whose (j, q)th element equals unity if the jth residual $\Delta \boldsymbol{t}_p{}^O$ is from the qth station and is zero otherwise.

If $\boldsymbol{s} \in E^Q$ is the vector of path anomalies appropriate to this particular set of events, then the assumption that the bias is path-correlated can be expressed by the equation

$$\boldsymbol{r}_p = \boldsymbol{B}_p \cdot \boldsymbol{s}. \tag{41}$$

The problem of specifying the set of vectors $\{\boldsymbol{r}_p\}$ is thus reduced to the problem of specifying $\boldsymbol{s}$.

To compose the various vectors described above into a single system of equations, a simple notation is employed: $\Delta \boldsymbol{X} \in E^{M_T}$, $\Delta \boldsymbol{T}^O \in E^{N_T}$, and $\boldsymbol{R} \in E^{N_T}$ are defined to be the partitioned column vectors with subvectors $\{\Delta \boldsymbol{x}_p\}$, $\{\Delta \boldsymbol{t}_p{}^O\}$, and $\{\boldsymbol{r}_p\}$, respectively. For example,

$$\Delta \boldsymbol{X} = \begin{bmatrix} \Delta \boldsymbol{x}_1 \\ \Delta \boldsymbol{x}_2 \\ \vdots \\ \Delta \boldsymbol{x}_P \end{bmatrix}. \tag{42}$$

Similarly, $\mathscr{A}$ and $\mathscr{V}_n$ are defined to be the $N_T \times M_T$ and $N_T \times N_T$ partitioned matrices whose P diagonal submatrices are $\{\boldsymbol{A}_p\}$ and $\{\boldsymbol{V}_{n_p}\}$, respectively, and whose $P(P-1)$ off-diagonal submatrices are zero; e.g.,

$$\mathscr{A} = \begin{bmatrix} \boldsymbol{A}_1 & \boldsymbol{O} & \cdots & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{A}_2 & \cdots & \boldsymbol{O} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{O} & \boldsymbol{O} & \cdots & \boldsymbol{A}_P \end{bmatrix}. \tag{43}$$

Hence, the vector $\Delta \boldsymbol{T}^O$ is a sample of a Gaussian process

$$\Delta \boldsymbol{T} = \overline{\Delta \boldsymbol{T}} + \boldsymbol{N} \tag{44}$$

specified by the following equations

$$\mathscr{A} \cdot \Delta \boldsymbol{X} = \overline{\Delta \boldsymbol{T}} \tag{45}$$

$$\langle \boldsymbol{N} \rangle = \boldsymbol{R} \tag{46}$$

$$\langle (\boldsymbol{N} - \boldsymbol{R})(\boldsymbol{N} - \boldsymbol{R})^T \rangle = \mathscr{V}_n = \nu^2 \hat{\mathscr{V}}_n. \tag{47}$$

To incorporate the assumption of path-correlated bias, we define the $N_T \times Q$

partitioned matrix

$$\mathscr{B} = \begin{bmatrix} \underline{\boldsymbol{B}_1} \\ \underline{\boldsymbol{B}_2} \\ \vdots \\ \underline{\boldsymbol{B}_P} \end{bmatrix} \tag{48}$$

so that the bias vector can be written as

$$\boldsymbol{R} = \mathscr{B} \cdot \boldsymbol{s} \tag{49}$$

for any given set of path anomalies $\boldsymbol{s}$.

The unbiased, minimum-variance estimate of $\Delta \boldsymbol{X}$ is obtained by minimizing the squared norm of the error process.

$$\hat{\boldsymbol{E}} = \Delta \hat{\boldsymbol{T}} - \hat{\mathscr{B}} \cdot \boldsymbol{s} - \hat{\mathscr{A}} \cdot \Delta \boldsymbol{X}. \tag{50}$$

Here, in analogy with the single-event case, the carat denotes normalization by the standard-deviation matrix $\hat{\mathscr{V}}_n^{1/2}$; e.g. $\Delta \hat{\boldsymbol{T}} = \hat{\mathscr{V}}_n^{-1/2} \cdot \Delta \boldsymbol{T}$. The solution to this problem is again found by applying the generalized inverse to the bias-corrected residual vector

$$\Delta \tilde{\boldsymbol{X}} = \hat{\mathscr{A}}^{\dagger} \cdot (\Delta \hat{\boldsymbol{T}} - \hat{\mathscr{B}} \cdot \boldsymbol{s}). \tag{51}$$

In fact, if the same set of path anomalies is used in (7) and (51), the estimates given by (15) and (51) are identical.

Previous work. Of course, the problem with applying equation (51) is that the path anomaly vector $\boldsymbol{s}$ is generally not known *a priori*. One set of techniques for dealing with this lack of information is based on the use of only arrival-time differences at particular stations to constrain relative locations (e.g., Ansell and Smith, 1975); such differences are clearly independent of $\boldsymbol{s}$. These so-called *homogeneous-station methods* are simple to implement with codes written for single-event location, but they make very inefficient use of the available data. For example, of the 68 stations reporting times for the Region A earthquakes listed in Table 1, 42 recorded 2 or more events, but only 15 recorded all 4 events. Thus, 27 stations with information about the relative locations of Region A events would have to be deleted in order to apply a homogeneous-station method to all 4 events. For Regions B and C, respectively, 24 of 36 stations and 59 of 68 stations would have to be deleted to insure homogeneity.

Another set of methods for multiple-event location assumes *a priori* knowledge of $\Delta \boldsymbol{x}_p$ for some p, say, $p = 1$. The *master-event technique* (e.g., Evernden, 1969b) is to choose the elements of $\boldsymbol{s}$ to equal their corresponding elements in the error vector $\boldsymbol{e}_1{}^O = \Delta \boldsymbol{t}_1{}^O - \boldsymbol{A}_1 \cdot \Delta \boldsymbol{x}_1$. Obviously, the method requires that the "master event" ($p = 1$) be recorded at all stations used in the multiple-event location scheme. Like the homogeneous-station methods, the master-event technique is easy to implement, but it also requires the deletion of useful data. In the case of Region A, only 32 stations recorded the best master event (A1) and at least one other earthquake; hence, 10 stations with information about relative locations would have to be

discarded, including several at azimuths not well represented by the reduced station set.

A less restrictive technique is the method of *joint hypocenter determination* (JHD), developed by Douglas (1967) and Dewey (1971, 1972) and applied by Blamey and Gibbs (1968), Fairhead and Girdler (1969), Billington and Isacks (1975), and Pascal *et al.* (1978), among others. In the JHD method the error criterion $|\hat{\boldsymbol{E}}|^2$ derived from (50) is minimized with respect to variations in both $\Delta \boldsymbol{X}$ and $\boldsymbol{s}$; estimates of $\Delta \boldsymbol{X}$ and $\boldsymbol{s}$ are thus derived by simultaneous inversion. The linear system resulting from this minimization is ill-conditioned (in our formulation, it is singular), so side constraints are usually imposed to stabilize its inversion, often by fixing the location of a master event (Douglas, 1967; Dewey, 1971) or the path anomalies to a set of "calibration stations" (Dewey, 1978). (Stabilization could equally well be accomplished by truncating the singular values of small magnitude in the SVD computation of the generalized inverse, but evidently this has not been attempted.) Besides the assumption of *ad hoc* side conditions, the JHD method has the distinct disadvantage of requiring the solution to a system of equations that increases in size approximately as Q^2. More precisely, the matrix system has dimension $N_T' \times [M_T + Q' - M]$, where N_T' is the number of elements of $\Delta \boldsymbol{T}^O$ corresponding to the Q' stations with two or more arrival times. In the case of Region C, for example, this system has dimension 230×89.

Given the large capacity and rapid speed of available computers, the solution of such a system is quite feasible. As shall be seen, however, it is not at all necessary. A procedure for multiple-event location has been developed that not only eliminates the need for fixing a master event but also requires at most the solution of an $N_T' \times M_T$ linear system. The theoretical foundation for the algorithm rests on a particular decomposition of the error vector $\hat{\boldsymbol{E}}$.

Hypocentral decomposition. The set of location perturbations $\{\Delta \boldsymbol{x}_p;\ p = 1, 2, \cdots, P\}$ has the decomposition

$$\Delta \boldsymbol{x}_p = \Delta \boldsymbol{x}_0 + \delta \boldsymbol{x}_p \tag{52}$$

where

$$\Delta \boldsymbol{x}_0 \equiv P^{-1} \sum_{p=1}^{P} \Delta \boldsymbol{x}_p \tag{53}$$

$$\sum_{p=1}^{P} \delta \boldsymbol{x}_p = 0. \tag{54}$$

We shall call $\boldsymbol{x}_0 + \Delta \boldsymbol{x}_0$ the *hypocentroid* and $\{\delta \boldsymbol{x}_p\}$ the *cluster vectors* of the event group.

Let $\delta \boldsymbol{X}$ be the M_T-component column vector with subvectors $\{\delta \boldsymbol{x}_p;\ p = 1, 2, \cdots, P\}$, and let $\Delta \boldsymbol{X}_0$ be the M_T-component column vector whose P subvectors all equal $\Delta \boldsymbol{x}_0$; then, equation (52) can be written

$$\Delta \boldsymbol{X} = \Delta \boldsymbol{X}_0 + \delta \boldsymbol{X}. \tag{55}$$

An $M_T \times M$ matrix $\mathscr{H}$ maps $\Delta \boldsymbol{x}_0$ into $\Delta \boldsymbol{X}_0$

$$\Delta \boldsymbol{X}_0 = \mathscr{H} \cdot \Delta \boldsymbol{x}_0. \tag{56}$$

In terms of the identity on E^M,

$$\mathscr{H} = \left[\begin{array}{c} \mathbf{I}^{(M)} \\ \hline \mathbf{I}^{(M)} \\ \hline \vdots \\ \hline \mathbf{I}^{(M)} \end{array}\right] \quad P \text{ partitions.} \tag{57}$$

Clearly, $\mathscr{H}^T \cdot \mathscr{H} = P\boldsymbol{I}^{(M)}$, and the generalized inverse of $\mathscr{H}$ is $\mathscr{H}^\dagger = P^{-1}\mathscr{H}^T$; hence, the operator that projects vectors in E^{M_T} onto the range space of $\mathscr{H}$ is $\mathscr{P}_H = P^{-1}\mathscr{H} \cdot \mathscr{H}^T$, and its complement, projecting vectors in E^{M_T} onto the null space of $\mathscr{H}^\dagger$, is $\mathscr{Q}_H = \boldsymbol{I}^{(M_T)} - \mathscr{P}_H$. The utility of these operators derives from the fact that, for any $\boldsymbol{\Delta X} \in E^{M_T}$,

$$\boldsymbol{\Delta X}_0 = \mathscr{P}_H \cdot \boldsymbol{\Delta X} \tag{58}$$

$$\boldsymbol{\delta X} = \mathscr{Q}_H \cdot \boldsymbol{\Delta X}. \tag{59}$$

These matrices thus specify the decomposition of $\boldsymbol{\Delta X}$ into a unique *hypocentroid vector* $\boldsymbol{\Delta X}_0$ and a unique *cluster vector* $\boldsymbol{\delta X}$.

Consider now the decomposition of the error vector (50) provided by the complementary projection operators $\mathscr{P}_{\hat{B}} = \hat{\mathscr{B}} \cdot \hat{\mathscr{B}}^\dagger$ and $\mathscr{Q}_{\hat{B}} = \boldsymbol{I}^{(N_T)} - \mathscr{P}_{\hat{B}}$

$$\hat{\boldsymbol{E}}_H = \mathscr{P}_{\hat{B}} \cdot \hat{\boldsymbol{E}} \tag{60}$$

$$\hat{\boldsymbol{E}}_C = \mathscr{Q}_{\hat{B}} \cdot \hat{\boldsymbol{E}}. \tag{61}$$

By definition, $\hat{\boldsymbol{E}}_H \cdot \hat{\boldsymbol{E}}_C = 0$, so the squared norm of $\hat{\boldsymbol{E}}$ is the sum of two terms

$$|\hat{\boldsymbol{E}}|^2 = |\hat{\boldsymbol{E}}_H|^2 + |\hat{\boldsymbol{E}}_C|^2. \tag{62}$$

From equations (40), (43), (56), and (57) it is easily seen that

$$\hat{\mathscr{A}} \cdot \boldsymbol{\Delta X}_0 = \hat{\mathscr{B}} \cdot \boldsymbol{A}_0 \cdot \boldsymbol{\Delta x}_0. \tag{63}$$

Therefore, $\mathscr{P}_{\hat{B}} \cdot \hat{\mathscr{A}} \cdot \boldsymbol{\Delta X}_0 = \hat{\mathscr{B}} \cdot \boldsymbol{A}_0 \cdot \boldsymbol{\Delta x}_0$, $\mathscr{Q}_{\hat{B}} \cdot \hat{\mathscr{A}} \cdot \boldsymbol{\Delta X}_0 = 0$, and the expressions for $\hat{\boldsymbol{E}}_H$ and $\hat{\boldsymbol{E}}_C$ reduce to

$$\hat{\boldsymbol{E}}_H = \mathscr{P}_{\hat{B}} \cdot \Delta\hat{\boldsymbol{T}} - \hat{\mathscr{B}} \cdot \boldsymbol{s} - \hat{\mathscr{B}} \cdot \boldsymbol{A}_0 \cdot \boldsymbol{\Delta x}_0 - \mathscr{P}_{\hat{B}} \cdot \hat{\mathscr{A}} \cdot \boldsymbol{\delta X} \tag{64}$$

$$\hat{\boldsymbol{E}}_C = \mathscr{Q}_{\hat{B}} \cdot \Delta\hat{\boldsymbol{T}} - \mathscr{Q}_{\hat{B}} \cdot \hat{\mathscr{A}} \cdot \boldsymbol{\delta X}. \tag{65}$$

These equations constitute the hypocentroidal decomposition theorem. Their remarkable form immediately suggests an algorithm for estimating the cluster vector and the hypocentroid. Since the path anomaly vector $\boldsymbol{s}$ does not appear in equation (65), an estimate of $\boldsymbol{\delta X}$ independent of $\boldsymbol{s}$ can be first obtained by minimizing $|\hat{\boldsymbol{E}}_C|^2$. Then, using this estimate in equation (64), an estimate of $\boldsymbol{\Delta x}_0$ can be found by minimizing $|\hat{\boldsymbol{E}}_H|^2$.

Before investigating this procedure in detail we give an easily computed formula

for the projection operator $\mathscr{P}_{\hat{B}}$ based on the expression

$$\mathscr{P}_{\hat{B}} = \hat{\mathscr{B}} \cdot \hat{\mathscr{B}}^{\dagger} = \hat{\mathscr{B}} \cdot (\hat{\mathscr{B}}^{T} \cdot \hat{\mathscr{B}})^{-1} \cdot \hat{\mathscr{B}}^{T}. \tag{66}$$

The column vectors of $\hat{\mathscr{B}}$ are mutually orthogonal, so $\hat{\mathscr{B}}^{T} \cdot \hat{\mathscr{B}}$ is a $Q \times Q$ positive-definite diagonal matrix expressible as the square of a positive-definite diagonal matrix $\boldsymbol{W}$

$$\hat{\mathscr{B}}^{T} \cdot \hat{\mathscr{B}} = \boldsymbol{W}^2 = \text{diag}(w_1^2, w_2^2, \cdots, w_Q^2). \tag{67}$$

The qth diagonal element of $\boldsymbol{W}^2$ is

$$w_q^2 = \sum_{j=1}^{N_T} B_{jq} \hat{\sigma}_j^{-2} \tag{68}$$

where B_{jq} is the (j, q)th element of $\mathscr{B}$ (a one or a zero) and $\hat{\sigma}_j^2$ is the jth diagonal element of $\hat{\mathscr{V}}_n$. Let $q(j) \in \{1, 2, \cdots, Q\}$ be the station index corresponding to the jth residual (i.e., $B_{jq(j)} = 1$), and define the $N_T \times N_T$ matrix

$$\Delta_{ij} = \begin{cases} 1, & \text{if the } i\text{th and } j\text{th residuals are from} \\ & \text{the same station} \\ 0, & \text{otherwise.} \end{cases} \tag{69}$$

Then, it is clear from equation (66) that the (i, j)th element of $\mathscr{P}_{\hat{B}}$ is

$$[\mathscr{P}_{\hat{B}}]_{ij} = w_{q(j)}^{-2} \hat{\sigma}_i^{-1} \hat{\sigma}_j^{-1} \Delta_{ij}. \tag{70}$$

In terms of the Kronecker delta δ_{ij}, the (i, j)th element of $\mathscr{Q}_{\hat{B}}$ is

$$[\mathscr{Q}_{\hat{B}}]_{ij} = \delta_{ij} - w_{q(j)}^{-2} \hat{\sigma}_i^{-1} \hat{\sigma}_j^{-1} \Delta_{ij}. \tag{71}$$

With these formulae projecting vectors in E^{N_T} using either $\mathscr{P}_{\hat{B}}$ or $\mathscr{Q}_{\hat{B}}$ can be coded very efficiently for machine calculation.

Estimation of the cluster vector. One estimate of the cluster vector minimizing $|\hat{\boldsymbol{E}}_C|^2$ is the generalized inverse solution

$$\delta\tilde{\boldsymbol{X}} = (\mathscr{Q}_{\hat{B}} \cdot \hat{\mathscr{A}})^{\dagger} \cdot \mathscr{Q}_{\hat{B}} \cdot \Delta\hat{\boldsymbol{T}}. \tag{72}$$

The minimum is clearly not unique, since adding any vector of the form $\mathscr{H} \cdot \Delta\boldsymbol{x}$ to (72) does not change $\hat{\boldsymbol{E}}_C$. However, equation (72) is the only solution to the minimization problem with the property that $\mathscr{P}_H \cdot \delta\tilde{\boldsymbol{X}} = 0$ and, therefore, the only one yielding an unbiased estimate of the true cluster vector $\delta\boldsymbol{X}$

$$\begin{aligned} \langle \delta\tilde{\boldsymbol{X}} \rangle &= (\mathscr{Q}_{\hat{B}} \cdot \hat{\mathscr{A}})^{\dagger} \cdot \mathscr{Q}_{\hat{B}} \cdot \langle \Delta\hat{\boldsymbol{T}} \rangle \\ &= (\mathscr{Q}_{\hat{B}} \cdot \hat{\mathscr{A}})^{\dagger} \cdot \mathscr{Q}_{\hat{B}} \cdot (\hat{\mathscr{A}} \cdot \Delta\boldsymbol{X} + \hat{\mathscr{B}} \cdot \boldsymbol{s}) \\ &= (\mathscr{Q}_{\hat{B}} \cdot \hat{\mathscr{A}})^{\dagger} \cdot \mathscr{Q}_{\hat{B}} \cdot \hat{\mathscr{A}} \cdot \Delta\boldsymbol{X} \\ &= \mathscr{Q}_H \cdot \Delta\boldsymbol{X} \\ &= \delta\boldsymbol{X}. \end{aligned} \tag{73}$$

The next-to-last equality asserted in (73), though nontrivial, is true as long as $\hat{\mathscr{A}}$ has full rank.

Equation (73) implies that the error process obtained by substituting (72) into (65) has zero mean: $\langle \hat{\boldsymbol{E}}_C \rangle = 0$. Solving for the second moment yields

$$\langle \hat{\boldsymbol{E}}_C \hat{\boldsymbol{E}}_C{}^T \rangle = \nu^2 (\mathscr{Q}_{\hat{B}} - \mathscr{P}_{Q\hat{A}}) \tag{74}$$

$$\langle |\hat{\boldsymbol{E}}_C|^2 \rangle = \nu^2 \operatorname{tr}(\mathscr{Q}_{\hat{B}} - \mathscr{P}_{Q\hat{A}}) \tag{75}$$

where $\mathscr{P}_{Q\hat{A}}$ is the projection operator whose diagonal elements are the "data importances" for this problem

$$\mathscr{P}_{Q\hat{A}} = \mathscr{Q}_{\hat{B}} \cdot \hat{\mathscr{A}} \cdot (\mathscr{Q}_{\hat{B}} \cdot \hat{\mathscr{A}})^{\dagger} = \mathscr{Q}_{\hat{B}} \cdot \hat{\mathscr{A}} \cdot (\hat{\mathscr{A}}^T \cdot \mathscr{Q}_{\hat{B}} \cdot \hat{\mathscr{A}})^{\dagger} \cdot \hat{\mathscr{A}}^T \cdot \mathscr{Q}_{\hat{B}}. \tag{76}$$

The trace of $\mathscr{P}_{Q\hat{A}}$, and thus the sum of the data importances, is $M_T - M = (P - 1)M$, the rank of $\mathscr{Q}_{\hat{B}} \cdot \hat{\mathscr{A}}$ and the trace of $\mathscr{Q}_H$. From equation (71) it is easily seen that $\operatorname{tr} \mathscr{Q}_{\hat{B}} = N_T - Q$. Therefore, the number of degrees of freedom is

$$\langle |\hat{\boldsymbol{E}}_C|^2 \rangle / \nu^2 = N_T - Q - (P - 1)M. \tag{77}$$

The variance matrix of the estimation process can be written as

$$\begin{aligned} \mathscr{V}_C &\equiv \langle (\boldsymbol{\delta}\tilde{\boldsymbol{X}} - \boldsymbol{\delta X})(\boldsymbol{\delta}\tilde{\boldsymbol{X}} - \boldsymbol{\delta X})^T \rangle \\ &\equiv \nu^2 \hat{\mathscr{V}}_C = \nu^2 (\hat{\mathscr{A}}^T \cdot \mathscr{Q}_{\hat{B}} \cdot \hat{\mathscr{A}})^{\dagger}. \end{aligned} \tag{78}$$

Hence, the random variable analogous to that defined by equation (25) is

$$k_C{}^2 = (\boldsymbol{\delta X} - \boldsymbol{\delta}\tilde{\boldsymbol{X}})^T \cdot \hat{\mathscr{V}}_C^{\dagger} \cdot (\boldsymbol{\delta X} - \boldsymbol{\delta}\tilde{\boldsymbol{X}}). \tag{79}$$

It has the expected value $\langle k_C{}^2 \rangle = \hat{\mathscr{V}}_C^{\dagger} : \hat{\mathscr{V}}_C = \nu^2 (P - 1)M$, so if no *a priori* information about ν^2 is assumed, the critical value at the $100(1 - \alpha)$ per cent confidence level takes the form

$$(\kappa_C)_{\alpha}{}^2 = (P - 1)M \hat{s}_C{}^2 F_{\alpha}[(P - 1)M, N_T - Q - (P - 1)M] \tag{80}$$

where the normalized sample variance is

$$\hat{s}_C{}^2 = |\hat{\boldsymbol{E}}_C{}^O|^2 / [N_T - Q - (P - 1)M]. \tag{81}$$

These expressions could be modified to accommodate a prior distribution on ν^2 using Bayesian arguments identical to those given for single-event locations, but the number of degrees of freedom is usually large enough to make this unnecessary.

As it stands, equation (72) requires the SVD of the $N_T \times M_T$ matrix $\mathscr{Q}_{\hat{B}} \cdot \hat{\mathscr{A}}$. It is clear from equations (71) and (76), however, that the arrival-time residual from any station recording only one event will have a data importance exactly equal to zero: such stations contribute no information about the relative locations of events within a cluster. Hence, stations with only one arrival can be deleted from the system without changing the estimate $\boldsymbol{\delta}\tilde{\boldsymbol{X}}$, thereby reducing the dimension of the system to $N_T' \times M_T$, where, as before, N_T' is the number of elements of $\boldsymbol{\Delta T}^O$ corresponding to

the Q' stations with two or more arrival times. Since $N_T' - Q' = N_T - Q$, reducing the system of equations in this manner does not alter the number of degrees of freedom and leaves the values of $(\kappa_C)_\alpha^2$ and $\hat{s}_C^2$ unaffected.

Estimation of the hypocentroid. Equation (64) makes explicit the tradeoff between path anomalies and hypocentroidal travel times and, hence, the need to constrain $\boldsymbol{s}$ in order to estimate $\Delta\boldsymbol{x}_0$. One could impose a prior normal distribution on $\boldsymbol{s}$, for example, and solve for both $\boldsymbol{s}$ and $\Delta\boldsymbol{x}_0$ by maximizing the appropriate likelihood function; prior information about $\boldsymbol{s}$ could be obtained, say, from an earthquake or explosion whose location was independently constrained. Instead, we shall simply adopt an estimate $\tilde{\boldsymbol{s}}$ that includes ellipticity, elevation, and station corrections, as we did in the case of single-event location, and minimize the squared length of (64) with respect to $\Delta\boldsymbol{x}_0$, employing (72) as the estimate of $\delta\boldsymbol{X}$. The solution to this minimization problem can be written as the difference of two terms

$$\Delta\tilde{\boldsymbol{x}}_0 = \Delta\tilde{\boldsymbol{x}}_0' - \Delta\tilde{\boldsymbol{x}}_0'' \tag{82}$$

$$\Delta\tilde{\boldsymbol{x}}_0' = (\hat{\mathscr{B}}\cdot\boldsymbol{A}_0)^\dagger\cdot(\Delta\hat{\boldsymbol{T}} - \hat{\mathscr{B}}\cdot\tilde{\boldsymbol{s}}) \tag{83}$$

$$\Delta\tilde{\boldsymbol{x}}_0'' = (\hat{\mathscr{B}}\cdot\boldsymbol{A}_0)^\dagger\cdot\hat{\mathscr{A}}\cdot\delta\tilde{\boldsymbol{X}}. \tag{84}$$

The fact that $(\hat{\mathscr{B}}\cdot\boldsymbol{A}_0)^\dagger\cdot\mathscr{P}_{\hat{B}} = (\hat{\mathscr{B}}\cdot\boldsymbol{A}_0)^\dagger$ has been used to derive (83) and (84). $\Delta\tilde{\boldsymbol{x}}_0'$ is just the estimate of $\Delta\boldsymbol{x}_0$ obtained by fitting a single hypocenter to the entire set of corrected residuals. This estimate is biased, however. Some manipulation yields

$$\langle\Delta\boldsymbol{x}_0'\rangle = \Delta\boldsymbol{x}_0 + (\hat{\mathscr{B}}\cdot\boldsymbol{A}_0)^\dagger\cdot\hat{\mathscr{B}}\cdot(\boldsymbol{s} - \tilde{\boldsymbol{s}}) + (\hat{\mathscr{B}}\cdot\boldsymbol{A}_0)^\dagger\cdot\hat{\mathscr{A}}\cdot\delta\boldsymbol{X}. \tag{85}$$

The second term on the right-hand side of (85) is the bias caused by incorrect path-anomaly corrections. The third term is introduced because, in the calculation of $\Delta\tilde{\boldsymbol{x}}_0'$, the data for each event in the cluster are not equally weighted; rather, $\Delta\tilde{\boldsymbol{x}}_0'$ is shifted toward the locations of events with more and/or better data. subtracting $\Delta\tilde{\boldsymbol{x}}_0''$ corrects for this latter type of bias, which can be seen by taking its expected value

$$\langle\Delta\tilde{\boldsymbol{x}}_0''\rangle = (\hat{\mathscr{B}}\cdot\boldsymbol{A}_0)^\dagger\cdot\hat{\mathscr{A}}\cdot\delta\boldsymbol{X}. \tag{86}$$

Therefore, the only bias in $\Delta\tilde{\boldsymbol{x}}_0$ is that ascribable to uncorrected path anomalies. We shall henceforth assume that the elements of the difference vector $\boldsymbol{s} - \tilde{\boldsymbol{s}}$ vary randomly enough as functions of distance and azimuth so that this bias is small enough to be ignored. Of course, the existence of uncorrected path anomalies will introduce scatter into the arrival-time residuals that cannot be fit by moving the hypocentroid; this scatter is treated as random noise, just as in the case of single-event locations. In other words, $\boldsymbol{s} - \tilde{\boldsymbol{s}}$ is considered to be a sample of a random process with zero mean, and the variance this process induces in the arrival times is incorporated into $\mathscr{V}_n$.

The variance matrix for the hypocentroid,

$$\boldsymbol{V}_H \equiv \langle(\Delta\tilde{\boldsymbol{x}}_0 - \Delta\boldsymbol{x}_0)(\Delta\tilde{\boldsymbol{x}}_0 - \Delta\boldsymbol{x}_0)^T\rangle \tag{87}$$

can be formulated by noting that the covariance between $\Delta\tilde{\boldsymbol{x}}_0'$ and $\Delta\tilde{\boldsymbol{x}}_0''$ is zero

$$\langle(\Delta\boldsymbol{x}_0' - \langle\Delta\boldsymbol{x}_0'\rangle)(\Delta\boldsymbol{x}_0'' - \langle\Delta\boldsymbol{x}_0''\rangle)^T\rangle$$
$$= \nu^2(\hat{\mathscr{B}}\cdot\boldsymbol{A}_0)^\dagger\cdot\mathscr{P}_{\hat{B}}\cdot\mathscr{Q}_{\hat{B}}\cdot(\hat{\mathscr{B}}\cdot\boldsymbol{A}_0)^{\dagger T} = 0. \tag{88}$$

Thus, the variance matrix has the decomposition

$$V_H = V_H' + V_H'' \tag{89}$$

where the two terms on the right-hand side are the variance matrices of $\Delta\tilde{x}_0'$ and $\Delta\tilde{x}_0''$, respectively. These can be reduced to the following expressions

$$V_H' = \nu^2 V_H' = \nu^2 (A_0{}^T \cdot W^2 \cdot A_0)^{-1} \tag{90}$$

$$V_H'' = (\hat{\mathscr{B}} \cdot A_0)^{\dagger} \cdot \hat{\mathscr{A}} \cdot \mathscr{V}_C \cdot \hat{\mathscr{A}}^T \cdot (\hat{\mathscr{B}} \cdot A_0)^{\dagger T}. \tag{91}$$

Consider now the statistics of the error process obtained by substituting the estimate (82) into equation (64)

$$\hat{E}_H = \mathscr{P}_{\hat{B}} \cdot \Delta\hat{T} - \hat{\mathscr{B}} \cdot s - \hat{\mathscr{B}} \cdot A_0 \cdot \Delta\tilde{x}_0' - [\hat{\mathscr{B}} \cdot A_0 \cdot \Delta\tilde{x}_0'' - \mathscr{P}_{\hat{B}} \cdot \hat{\mathscr{A}} \cdot \delta X]. \tag{92}$$

Using equations (72) and (84) it is possible to show that the term in square brackets is identically equal to zero (the proof is omitted to conserve space). Taking the first and second moments of (92), we find

$$\langle \hat{E}_H \rangle = 0 \tag{93}$$

$$\langle \hat{E}_H \hat{E}_H{}^T \rangle = \nu^2 [\mathscr{P}_{\hat{B}} - \hat{\mathscr{B}} \cdot A_0 \cdot (A_0{}^T \cdot W^2 \cdot A_0)^{-1} \cdot A_0{}^T \cdot \hat{\mathscr{B}}^T] \tag{94}$$

$$\langle |\hat{E}_H|^2 \rangle = \operatorname{tr} \langle \hat{E}_H \hat{E}_H{}^T \rangle = \nu^2 (Q - M). \tag{95}$$

For this problem the normalized sample variance is, therefore,

$$\hat{s}_H{}^2 = |\hat{E}_H{}^O|^2/(Q - M). \tag{96}$$

Generally, $\hat{s}_H{}^2$ will exceed $\hat{s}_C{}^2$, because the former includes the dispersion caused by uncorrected path anomalies, whereas the latter is insensitive to such anomalies.

Confidence ellipsoids for either $\Delta x_0'$ or $\Delta x_0''$ can be constructed in the usual manner. For example, the random variable

$$(k_H')^2 = (\Delta\tilde{x}_0' - \Delta x_0')^T \cdot (\hat{V}_H')^{-1} \cdot (\Delta\tilde{x}_0' - \Delta x_0') \tag{97}$$

has an expected value $\nu^2 M$ and a critical value proportional to $\hat{s}_H{}^2$

$$(\kappa_H')_\alpha{}^2 = M\hat{s}_H{}^2 F_\alpha[M, Q - M]. \tag{98}$$

An analogous variable $(k_H'')^2$ can be defined in terms of the scaled variance matrix $\hat{V}_H''$, obtained by substituting $\hat{\mathscr{V}}_C$ for $\mathscr{V}_C$ in equation (91); its expected value is also $\nu^2 M$, but its critical value is proportional to $\hat{s}_C{}^2$

$$(\kappa_H'')_\alpha{}^2 = M\hat{s}_C{}^2 F_\alpha[M, N_T - Q - (P - 1)M]. \tag{99}$$

Using the additive properties of χ^2 distributions an expression for the confidence ellipsoid of the hypocentroid estimate itself—the difference between $\Delta\tilde{x}_0'$ and $\Delta\tilde{x}_0''$—can be derived; the $100(1 - \alpha)$ per cent confidence ellipsoid for the hypocentroid

comprises all vectors $\Delta \boldsymbol{x}$ that satisfy the inequality

$$(\Delta \boldsymbol{x} - \Delta \tilde{\boldsymbol{x}}_0{}^O)^T \cdot [\hat{s}_H{}^2 \hat{\boldsymbol{V}}_{H'} + \hat{s}_C{}^2 \hat{\boldsymbol{V}}_{H''}]^{-1} \cdot (\Delta \boldsymbol{x} - \Delta \tilde{\boldsymbol{x}}_0{}^O) \leqq MF_\alpha[M, N_T - M_T]. \quad (100)$$

It is clear from this expression and from equation (89) that the bias-corrected estimate $\Delta \tilde{\boldsymbol{x}}_0$ will always have a larger uncertainty than $\Delta \tilde{\boldsymbol{x}}_0'$; the bias in the latter can be reduced only by increasing its variance.

To compute $\Delta \tilde{\boldsymbol{x}}_0'$ and $\Delta \tilde{\boldsymbol{x}}_0''$ directly from equations (83) and (84) requires the calculation of the generalized inverse of $\hat{\mathscr{B}} \cdot \boldsymbol{A}_0$, an $N_T \times M$ matrix. However, the dimension of this system can be reduced by rewriting these equations using the expression $(\hat{\mathscr{B}} \cdot \boldsymbol{A}_0)^\dagger = (\boldsymbol{A}_0{}^T \cdot \boldsymbol{W}^2 \cdot \boldsymbol{A}_0)^{-1} \cdot \boldsymbol{A}_0{}^T \cdot \hat{\mathscr{B}}^T$; for example,

$$\begin{aligned} \Delta \tilde{\boldsymbol{x}}_0' &= (\boldsymbol{A}_0{}^T \cdot \boldsymbol{W}^2 \cdot \boldsymbol{A}_0)^{-1} \cdot \boldsymbol{A}_0 \cdot (\hat{\mathscr{B}}^T \cdot \Delta \hat{\boldsymbol{T}} - \boldsymbol{W}^2 \cdot \tilde{\boldsymbol{s}}) \\ &= (\boldsymbol{A}_0{}^T \cdot \boldsymbol{W}^2 \cdot \boldsymbol{A}_0)^{-1} \cdot \boldsymbol{A}_0 \cdot \boldsymbol{W} \cdot [\boldsymbol{W}^{-1} \cdot \hat{\mathscr{B}}^T \cdot \Delta \hat{\boldsymbol{T}} - \boldsymbol{W} \cdot \boldsymbol{s}] \\ &= (\boldsymbol{W} \cdot \boldsymbol{A}_0)^\dagger \cdot [\boldsymbol{W}^{-1} \cdot \hat{\mathscr{B}}^T \cdot \Delta \hat{\boldsymbol{T}} - \boldsymbol{W} \cdot \tilde{\boldsymbol{s}}] \end{aligned} \quad (101)$$

The computation of $\Delta \tilde{\boldsymbol{x}}_0'$ thus requires only the SVD of the $Q \times M$ matrix $\boldsymbol{W} \cdot \boldsymbol{A}_0$

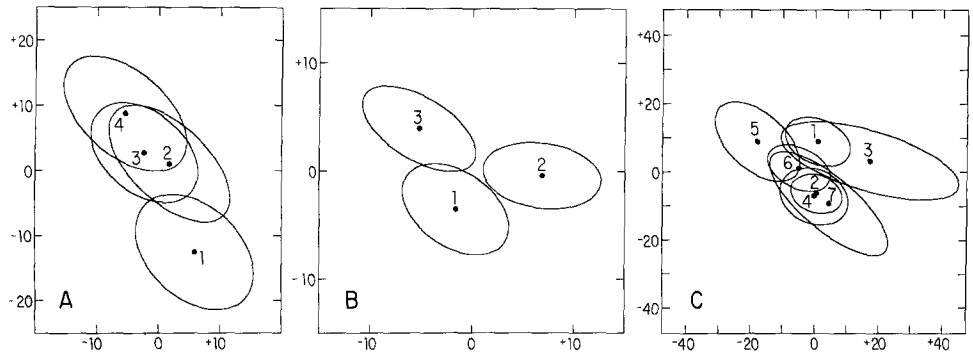

FIG. 5 Cluster vectors for earthquakes in Regions A, B, and C calculated from equation (72), ellipses delineate 95 per cent confidence regions for individual epicenters. North is up, units are kilometers, and event numbers refer to Table 1

and the multiplication of its generalized inverse into the station-averaged, bias-corrected residual vector enclosed by the square brackets. Moreover, the diagonal elements of the $Q \times Q$ projection operator

$$\boldsymbol{P}_{WA_0} = \boldsymbol{W} \cdot \boldsymbol{A}_0 \cdot (\boldsymbol{W} \cdot \boldsymbol{A}_0)^\dagger \quad (102)$$

are the data importances—or, here, perhaps more aptly termed "station importances"—for this problem.

Application to Regions A, B, and C. The location procedures based on hypocentroidal decomposition have been applied to the earthquakes of Table 1. The cluster vectors derived from equation (72) are displayed in Figure 5, where again the hypocentral depths have been fixed at 10 km.

Deserving attention are several comparisons of these results with those of the single-event procedure. First, in the two methods the information regarding relative locations, as indicated by the data importances, is distributed differently across the data sets. Whereas stations with only one arrival per cluster contribute to the

relative locations in the single-event calculation, such stations are assigned zero importance in the multiple-event method, because this information could be biased by uncorrected path anomalies. There is a corresponding increase in the importance of stations with more than one arrival; the cumulative importance of the 15 stations recording all four Region A events increases from 43 per cent of the total to 63 per cent, for example.

Second, the sizes of the confidence ellipsoids for individual epicenters are substantially reduced by joint location. The reduction of errors in relative location is attributable, of course, to the removal of path-correlated errors from the arrival-time data; the values of the normalized standard deviations given by equation (81) are $\hat{s}_C = 0.58$, 0.28, and 0.39 for Regions A, B, and C, respectively, a factor of 2 or so smaller than the normalized standard deviations typical of single-event locations (Table 1). The average magnitude in each cluster ($\bar{m}_b = 4.9$, 5.1, and 5.0, respectively) correlates inversely with $\hat{s}_C$, which is consistent with the hypothesis that these smaller standard deviations are dominated by reading errors and, hence, reflect the average signal-to-noise ratio. Indeed, on the seismograms examined by the authors, arrivals from the smaller Region A events are generally more difficult to pick than those from the larger events in Region B.

Finally, a comparison of Figure 5 with Figure 4 shows that the multiple-event calculation reduces the overall scatter of Region C epicenters, but not that of Region A or B; in fact, Region B epicenters are somewhat more dispersed when jointly determined. The rms lengths of the cluster vectors in Regions A, B, and C are 8.9, 5.9, and 12.0 km, respectively. The marginal distributions for individual events within each cluster yield confidence ellipsoids that are generally smaller than these rms distances, suggesting the scatter is not entirely caused by reading errors. However, a formal test of this hypothesis cannot be based on the marginal distributions alone; it must also account for the event-to-event error cross-correlations implied by the other off-diagonal elements of $\mathcal{V}_C$.

Let the subscript S denote the purely spatial parts of the cluster vector and its error matrix, and consider the null hypothesis that the epicenters in each region are confined to a single spatial point; i.e., the hypothesis that $\boldsymbol{\delta X}_S = 0$. If $(k^O_{CS})^2 \equiv \boldsymbol{\delta \tilde{X}}_S{}^{OT} \cdot \hat{\mathcal{V}}_{CS}{}^{\dagger} \cdot \boldsymbol{\delta \tilde{X}}_S{}^{O}$ exceeds some critical value $(\kappa_{CS})_\alpha{}^2$, then this null hypothesis can be rejected at the $100(1 - \alpha)$ per cent confidence level. Calculations involving the appropriate marginal distributions yield the following table

Region	$(k^O_{CS})^2$	$(\kappa_{CS})_\alpha{}^2$	
		$\alpha = 0.05$	$\alpha = 0.01$
A	4.8	4.5	6.2
B	1.0	0.8	1.2
C	10.6	3.3	4.2.

In all three cases, the hypothesis that the epicenters are identical can be rejected at the 95 per cent confidence level, but only for Region C does the hypothesis fail at the 99 per cent level.

Since the arrival-time data resolve the finite sizes of the source regions at a reasonable confidence level, some of the features of the distributions in Figure 5 may have tectonic significance. In particular, there are alignments of epicenters in both Regions A and C along NW-trending axes. The Region-A alignment is nearly

parallel to small-scale faults identified within the epicentral zone during a 1979 oceanographic survey (Sverdrup and Jordan, 1979). These faults, which were presumably generated at or near the spreading center, are perhaps being reactivated by intraplate stresses.

"Weighted" estimates of the epicentroids calculated from equation (83) are listed in Table 4 and are compared with the single-event locations in Figure 6. As pointed out in the theoretical discussion, this formula provides the minimum-variance estimate of Δx_0, but its value is biased toward the larger, better-recorded events. In the three cases considered here, however, the bias-correction term given by (84) is not significantly different from zero at the 95 per cent confidence level, so, to avoid increasing their variances, this correction has not been made to any of the epicentroid estimates.

TABLE 4

WEIGHTED EPICENTROIDS CALCULATED FROM EQUATION (83)*

Region	Latitude (°N)	s d (°)	Longitude (°E)	s d (°)
A	−7.40	0.023	−148.21	0.026
B	−18.40	0 018	−132 87	0 018
C	−20.79	0 015	−126 95	0 019

* Depths fixed at 10 km

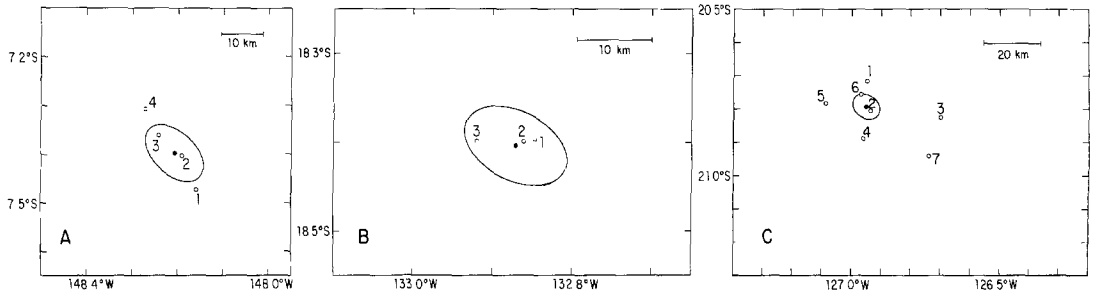

FIG 6. Weighted epicentroids for Regions A, B, and C and their 95 per cent confidence ellipsoids Open circles are single-event locations numbered according to Table 1

DISCUSSION

The epicentroid locations of Table 4 thus constitute our best estimates of where to look for bathymetric expressions of the seismicity (e.g., faulting or disruption of sediments) or for other features associated with seismic activity (e.g., active seamounts). The 95 per cent confidence ellipsoids for the epicentroids have characteristic dimensions on the order of 10 km (Figure 6), whereas the clusters themselves appear to be confined to areas with characteristic horizontal dimensions not more than about twice this value (Figure 5). If the bias due to uncorrected path anomalies is this small or smaller, as we suspect, then detailed bathymetric and seismic surveys of Regions A, B, and C using standard marine geophysical techniques are feasible, requiring the expenditure of only several days of ship time at each site. A preliminary survey of Region A has already been accomplished (Sverdrup and Jordan, 1979), and the interpretation of the oceanographic data in conjunction with the earthquake location results obtained here will be the subject of a separate publication.

The theoretical techniques developed in this paper should have wide application

to other teleseismic location problems. Although the method of hypocentroidal decomposition appears to be new, the basic idea behind the method—the use of projection operators to separate a linear system of equations into two or more smaller systems—has been employed previously in seismological inverse problems. In particular, three groups of investigators (Pavlis and Booker, 1980; Rodi *et al.*, 1980; Spencer and Gubbins, 1980; see also Gubbins, 1980) have developed and applied, independently of one another, the projection technique and its generalizations to the joint determination of hypocenters and velocity structure in regions of lateral velocity variations. As in the case of hypocentroidal decomposition, these techniques lead to algorithms requiring much less work than previous formulations. The use of projection operators in setting up "hierarchies" of inverse problems has been discussed in the context of earth-structure determination by Jordan (1980), who describes a general class of results he calls "averaging theorems". The hypocentroidal decomposition theorem is a member of this class.

Acknowledgments

The authors thank R. D. Adams of the International Seismological Centre for providing the data and preliminary locations of events C3 and C4 before their publication, and they are grateful to E. A. Okal and W. L. Rodi for helpful discussions, and to J. W. Dewey for critical comments on the typescript. This research was completed during the tenure of T. H. J. as an Alfred P. Sloan Research Fellow and was supported by the Office of Naval Research under Grant N00014-75-C-0152.

References

Ansell, J. H. and E. G. C. Smith (1975). Detailed structure of a mantle seismic zone using the homogeneous station method, *Nature* **253**, 518–520.

Billington, S. and B. Isacks (1975). Identification of fault planes associated with deep earthquakes, *Geophys. Res. Letters* **2**, 63–66.

Blamey, C. and P. G. Gibbs (1968). The epicenters and origin times of some large explosions, *Geophys. J.* **16**, 1–7.

Buland, R. (1976). The mechanics of locating earthquakes, *Bull. Seism. Soc. Am.* **66**, 173–187.

Cox, D. R. and D. V. Hinkley (1974). *Theoretical Statistics*, Chapman and Hall, London, 511 pp.

Dewey, J. W. (1971). Seismicity studies with the method of joint hypocenter determination, *Ph.D. Dissertation*, University of California, Berkeley.

Dewey, J. W. (1972). Seismicity and tectonics of western Venezuela, *Bull. Seism. Soc. Am.* **62**, 1711–1751.

Dewey, J. W. (1978). Joint hypocenter determination using calibration stations: applications to eastern North America (abstract), *Earthquake Notes* **49**, 35.

Douglas, A. (1967). Joint epicentre determination, *Nature* **215**, 47–48.

Evernden, J. F. (1969a). Precision of epicenters obtained by small numbers of world-wide stations, *Bull. Seism. Soc. Am.* **59**, 1365–1398.

Evernden, J. F. (1969b). Identification of earthquakes and explosions by use of teleseismic data, *J. Geophys. Res.* **74**, 3828–3856.

Fairhead, J. D. and R. W. Girdler (1969). How far does the rift system extend through Africa? *Nature* **221**, 1018–1020.

Flinn, E. A. (1965). Confidence regions and error determinations for seismic event location, *Rev. Geophys.* **3**, 157–185.

Freedman, H. W. (1967). A statistical discussion of *P* residuals from explosions. Part II, *Bull. Seism. Soc. Am.* **57**, 545–561.

Geiger, L. (1910). Herdbestimmung bei Erdbeben aus den Ankunftzeiten, *K. Gesell. Wiss. Goett.* **4**, 331–349.

Golub, G. H. and C. Reinsch (1971). Singular value decomposition and least squares solutions, in *Linear Algebra*, J. H. Wilkinson, and C. Reinsch, Editors, Springer-Verlag, New York, 134–151.

Gubbins, D. (1980). Source location in laterally varying media (preprint).

Herrin, E. (Chairman) (1968). Seismological tables for *P* phases, *Bull. Seism. Soc. Am.* **58**, 1193–1241.

Jeffreys, H. (1961). *Theory of Probability*, Oxford, 3rd ed., 439 pp.

Jordan, T. H. (1980). Earth structure from seismological observations, in *Physics of the Earth's Interior*

Proceedings of the International School of Physics "Enrico Fermi," Course LXXVIII, A M. Dziewonski and E Boschi, Editors, North Holland, Amsterdam, 1–40

Lomnitz, C. (1977) A fast epicenter location program, *Bull Seism Soc. Am.* **67,** 425–431.

Minster, J B , T. H Jordan, P Molnar, and E Haines (1974). Numerical modelling of instantaneous plate tectonics, *Geophys. J.* **36,** 541–576.

Minster, J. F , J B. Minster, M Treuil, and C. J Allegre (1977) Systematic use of trace elements in igneous processes, II, Inverse problem of the fractional crystallization process, *Contrib Mineral Petrol* **61,** 49–77

Okal, E. A , J. Talandier, K. A Sverdrup, and T H Jordan (1980) Seismicity and tectonic stress in the South-Central Pacific, *J Geophys Res* **85,** 6479–6495.

Pascal, G., B L Isacks, M Barazangi, and J Dubois (1978). Precise relocations of earthquakes and seismotectonics of the New Hebrides island arc, *J Geophys Res* **83,** 4957–4973

Pavlis, G. L and J R. Booker (1980). The mixed discrete-continuous inverse problem: application to the simultaneous determination of earthquake hypocenters and velocity structure, *J Geophys Res.* **85,** 4801–4810.

Penrose, R A. (1955) A generalized inverse for matrices, *Proc Camb. Phil Soc* **51,** 406–413

Poupinet, G (1979). On the relation between *P*-wave travel time residuals and the age of continental plates, *Earth Planet Sci Letters* **43,** 149–161.

Rodi, W L , T H. Jordan, J. F. Masso, and J. M Savino (1980) Determination of the three-dimensional structure of eastern Washington from the joint inversion of gravity and earthquake travel-time data, *Report SSS-R-80-4516*, Systems, Science and Software, La Jolla, 143 pp

Spence, W. (1980) Relative epicenter determination using *P*-wave arrival-time differences, *Bull. Seism Soc Am* **70,** 171–183

Spencer, C and D. Gubbins (1980) Travel time inversion for simultaneous earthquake location and velocity structure determination in laterally varying media, *Geophys. J.* **63,** 95–116.

Sverdrup, K and T. H. Jordan (1979). Bathymetric survey of Seismic Region A, South Central Pacific Ocean (abstract), *EOS, Trans Am. Geophys Union* **60,** 957.

Talandier, J. and M. Bouchon (1979) Propagation of high-frequency Pn waves at great distances in the South Pacific and its implication for the structure of the lower lithosphere, *J. Geophys Res* **84,** 5613–5619.

Talandier, J and G. T Kuster (1976). Seismicity and submarine volcanic activity in French Polynesia, *J Geophys Res.* **81,** 936–948.

Veith, K F. (1975). Refined hypocenters and accurate reliability estimates, *Bull Seism Soc. Am.* **65,** 1199–1222

Zelen, M and C. Severo (1972) Probability functions, in *Handbook of Mathematical Functions*, M. Abramowitz, and I. A Stegun, Editors, Dover, New York, 925–995

Geological Research Division
Scripps Institution of Oceanography
La Jolla, California 92093

Manuscript received October 30, 1980