JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 89, NO. B9, PAGES 7962–7968, SEPTEMBER 10, 1984

Gravity Empirical Covariance Values for the Continental United States

CLYDE C. GOAD

National Geodetic Survey, Charting and Geodetic Services, National Ocean Service, National Oceanic and Atmospheric Administration, Rockville, Maryland

C. CHRISTIAN TSCHERNING

Danish Geodetic Institute, Charlottenlund

M. MIRANDA CHIN

National Geodetic Survey, Charting and Geodetic Services, National Ocean Service, National Oceanic and Atmospheric Administration, Rockville, Maryland

Variances and correlation distances have been determined for terrain corrected Bouguer anomalies in all 30 × 30 arc min blocks covering the continental United States. Also linear trend values were subtracted from the data. In order to make the determination independent of the frequently irregular local data distribution, gravity values predicted in a regular grid were used. The prediction was made using least squares collocation, constructing local surfaces representing the gravity anomalies for each 30 × 30 arc min block. For this purpose 544,000 terrain-corrected Bouguer anomalies were available in the U.S. National Geodetic Survey gravity data base. Therefore in order to limit the computational effort, a sequence of surfaces was determined by selecting in each block an increasing number of data points until a limit of the absolute value of the error equal to 4 mGal was reached. This procedure reduced the number of gravity observations used per block to a very few (in areas of little gravity variation) to maximally 188. The correlation distance was found to be smallest in areas of high topography. Large variances are associated with the midcontinental gravity high and tectonic features on the Pacific coast.

1. Introduction

The method of least squares collocation [*Moritz*, 1980] is widely used for local interpolation and prediction of gravity field related quantites such as gravity anomalies and deflections of the vertical. The optimal procedure is to remove the long-wavelength field (by subtracting out the contribution from a high degree and order spherical harmonic expansion) and the effect of the topography. The residuals are then used as data in least squares collocation. In addition, to the residual anomalies, this method requires knowledge of the covariance function.

The most important parameters describing a covariance function are variance and correlation distance. We have determined these quantities for residual Bouguer anomalies in North America for 30 × 30 arc min areas. Since a spherical harmonic expansion representing wavelengths down to 30 arc min was not available, a simple linear trend was removed.

Our results show that the parameters are quite similar in some areas (i.e., the covariance functions are nearly stationary) and vary greatly in other areas. These differences appear to correspond well with the changing topographic variation and with well-known tectonic features.

Before giving further computational details, we would like to describe the background for this investigation.

Recently, at the U.S. National Geodetic Survey, the method of least-squares collocation was implemented to provide interpolated gravity anomalies for the evaluation of Stoke's and Vening-Meinesz integrals [*Schwarz*, 1978]. At first, gravity anomalies were determined using a common covariance function for all of the continental United States. Having realistic estimates for the uncertainties of the predictions (which can be obtained as a part of the least squares collocation procedure) allows applications to be enhanced. But this requires that the actual covariance function be available.

One of the most obvious uses of these statistics is the determination of where additional observations should be made [see *Tscherning*, 1975, 1980]. Another is in estimating errors in inertial navigation or surveying resulting from failure to take gravity disturbances into account [*Chatfield et al.*, 1975; *Forsberg and Tscherning*, 1981].

Covariance functions may be determined directly from observational data, but the often erratic data distribution may cause poor results. We then chose to use the implemented least squares collocation sytem to interpolate data in a regular grid for use in the covariance estimation.

The implemented algorithm is described in section 2. Section 3 contains a discussion of how the effect of local topography was removed. The selection of data to be used for the prediction is described in section 4, and in section 5 the estimation of the covariance function parameters are described. A geophysical interpretation of the varying values of these parameters is attempted in section 6, and in section 7 we derive simple expressions relating the prediction error and the parameters.

2. The Gravity Prediction Algorithm

A detailed discussion of the least squares collocation algorithm used here can be found by *Moritz* [1980]. Since the algorithm requires the solution of a system of linear equations with as many unknowns as the number of data points, local

Paper number 4B0740.

solutions are constructed from data in local, overlapping areas. Global effects are removed by subtracting the effect of a high-order spherical harmonic approximation or of a local trend surface as explained in the following. Also, the number of data points needed in order to be within a given upper limit for the error of prediction will decrease considerably if the local field is smoothed by subtracting out local topographic effects. Details concerning the data selection algorithm used are given in section 4, and the removal of topographic effects is described in section 3.

For the United States it was decided to determine local solutions for 30 × 30 arc min as a compromise for as small an area as possible but still economizing on computer storage and the number of blocks to handle and maintain. Since a spherical harmonic expansion representing wavelengths down to the size of this area was not available, another solution had to be found.

Figure 1 is a typical plot of the distribution of gravity data over a 30 × 30 arc min sector. A visual examination reveals that a first-order surface fits the global character of the data values. This can be expressed mathematically by

$$\Delta g = x_0 + x_1\phi + x_2\lambda + s + \varepsilon \qquad (1)$$

where Δg is the gravity anomaly value, ϕ, λ are the geodetic coordinates of the observations, s is the unexplained or unmodeled signal part of the measurement, ε is measurement noise, and x_0, x_1, and x_2 are the systematic surface parameters.

Equation (1) is linear in x_0, x_1, and x_2 and thus allows one to directly estimate them using the method of least squares collocation.

Let

$$X = \begin{bmatrix} x_0 \\ x_1 \\ x_1 \end{bmatrix}$$

and A be the matrix of partial derivatives,

$$A = \begin{bmatrix} 1 & \phi_1 & \lambda_1 \\ 1 & \phi_2 & \lambda_2 \\ \vdots & \vdots & \vdots \\ 1 & \phi_n & \lambda_n \end{bmatrix}$$

Then, (1) is rewritten in matrix notation as

$$[\Delta g] = AX + S + \varepsilon \qquad (2)$$

The least squares solution for the vector X is given by

$$\hat{X} = [A^T(C_{ss} + R)^{-1}A]^{-1}A^T(C_{ss} + R)^{-1}[\Delta g] \qquad (3)$$

where R is the measurement noise matrix, and C_{ss} is the signal covariance matrix. Note that R is usually a diagonal (white) noise covariance matrix and C_{ss} is a full signal covariance matrix.

The (i, j) element of the C_{ss} matrix contains the signal covariance between the ith and jth measurements. Since the data were terrain-corrected anomalies, a homogeneous and isotropic covariance model was used in this study, given by

$$E(s_i s_j) = C(\psi_{ij}) = \sum_{n=360}^{\infty} c_n P_n(\cos \psi_{ij}) \qquad (4)$$

where ψ_{ij} is the angular distance between location i and location j. *Kaula*'s [1966, p. 98] rule was chosen for the evaluations of the c_n in (4). The summation first starts at $n = 360$ to represent all unmodeled variations with wavelengths shorter than 30 arc min. The upper limit of infinity in the summation was approximated by choosing $n_{max} = 2000$. Figure 2 is a plot of the signal covariance function used.

This covariance function was used initially to fit the interpolating gravity anomaly surfaces to the U.S. gravity data. Experience has shown that the form of the covariance function does not materially affect the quality of the interpolation [*Tscherning*, 1975], however, it does impact the quality of the error estimates [*Tscherning*, 1980].

The estimates of the signal contributions are given by

$$\hat{S}_p = C_{s_p s}[C_{ss} + R]^{-1}\{[\Delta g] - A\hat{X}\} \qquad (5)$$

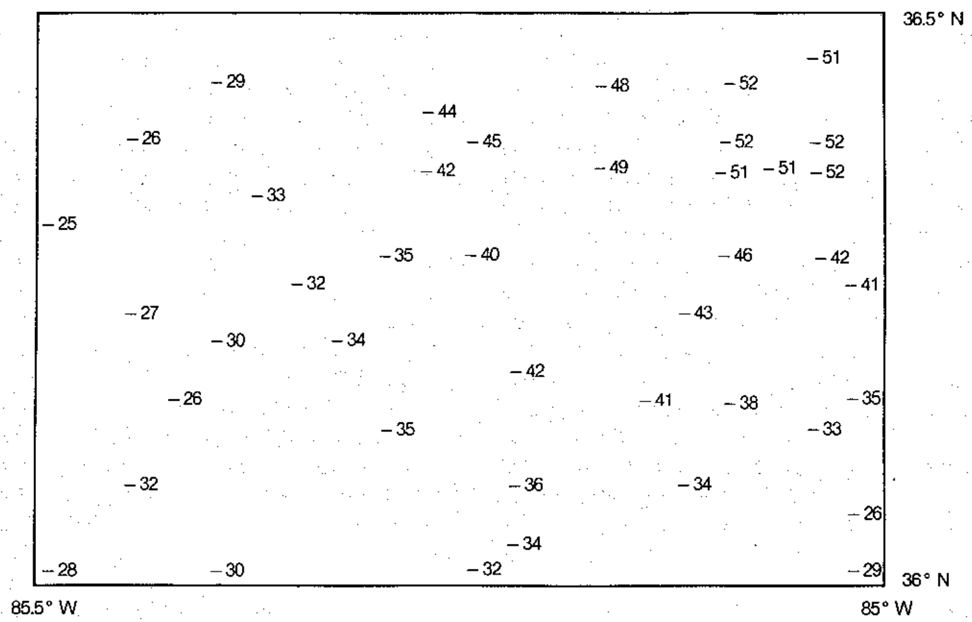

Fig. 1. Typical gravity distribution over a 30 × 30 arc min area.

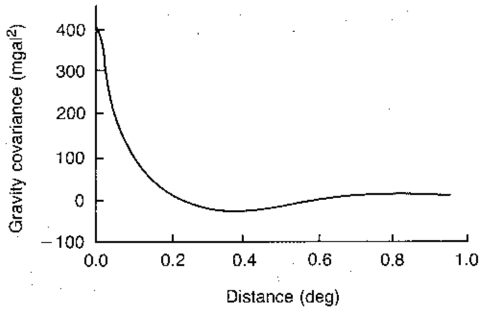

Fig. 2. Preliminary gravity covariance function.

Both equations (3) and (5) were solved using a Cholesky decomposition algorithm. At this stage, no matrix inverse actually needs to be computed; only a system of linear equations needs to be solved. The estimate of a gravity anomaly at a point becomes the combination of (3) and (5):

$$\Delta g_p = A_p\hat{X} + \hat{S}_p \qquad (6)$$

where again the subscript p denotes the predicted point, with coordinates implied in A_p.

The systematic coefficients $\hat{X}$ and interpolation vector $[C_{ss} + R]^{-1} \{[\Delta g] - A\hat{X}\}$ were determined for 3637 30 × 30 arc min areas (within the continental United States) and stored within the computer for quick access.

3. Terrain Correction Algorithm

By far the most important correction to free air gravity anomaly data is the Bouguer plate reduction. Also the removal of nearby topographic mass effects is needed to permit the use of the azimuth-independent covariance functions. Now that digitized terrain data are available for the United States on a 30 × 30 arc sec grid interval, terrain corrections can be considered in preprocessing to permit modeling of gravity at the required 5-mGal level. One can then use consistent terrain-corrected gravity anomalies with similarly corrected deflections of the vertical in a combined adjustment. *Forsberg and Tscherning* [1981] chose to model the topographic height variations (called residual terrain models) using parallelepipeds. Since the terrain data are available on square grids of latitude and longitude, no special height interpolation schemes need to be used when the parallelepiped mass model is chosen. *MacMillan* [1958] has given the equations for modeling the potential and its derivatives using parallelepipeds. *Forsberg and Tscherning* [1981] used the MacMillan formulation with a dense grid of parallelepipeds near the computation point and a coarser grid for distant areas to decrease the computation burden. Here the MacMillan formulation is used for the nearby contributions, but a simpler formulation was chosen for the distant masses. The same small grid interval is used throughout. The distant masses are still modeled as parallelepipeds (or prisms) but with a less rigorous formulation. First-order Taylor expansions of the differential potential with respect to the center of the prisms are analytically integrated over the volume. A density of 2.67 g/cm³ was used for all calculations. The grid heights are given to the nearest 10 m. Differentiation with respect to the center coordinates yields components of force, which are used to smooth gravity anomalies and also deflection of the vertical observations (horizontal force components). Little difference is found between our numerical smoothing algorithm and that of *Forsberg and Tscherning* [1981]. Considering all the approximations made, we expect to have removed all topographic effects to within an error of ±2 mGal. This error will not play a role in the estimation of covariance function parameters as described in section 5.

The most dramatic evidence we have found of the success of terrain corrections for data smoothing is in the analysis of deflection of the vertical measurements in the White Sands, New Mexico, area, Here the north-south components show variations about the mean of 2 arc sec before removal of local mass effects and 1 arc sec after removal. East-west variations are 7 arc sec before and only 2 arc sec after removal of local mass contributions (out to a distance of 30 arc min). In practice, once the smoothed systematic surfaces and signal interpolation coefficients are available, predictions are made using (6), and then the topographic mass effects are restored.

It should also be pointed out that identifying errors in gravity data in mountainous areas at the 10-mGal level is almost impossible without the use of terrain corrections to obtain smoothly varying gravity anomalies. The contributions of the terrain effects can easily reach 15 mGal with 25- to 30-mGal effects occurring occasionally.

4. Data Selection

The accuracy of the empirical covariance values depends very much on the data distribution. The determination would be optimum if a regularly distributed set of anomalies were available. In practice, the situation is as described below.

Figure 3 shows the distribution of continental U.S. gravity data. Some 30 × 30 arc min areas have several thousand gravity observations, while others have only a few tens or less. Some areas have so many observations, it is impossible from a computational standpoint to use them all. Thus a data selection algorithm was employed to reduce the computational burden while still providing adequate coverage in the areas of less dense data. First, an initial pass was made through all data in a 30 × 30 arc min area, selecting data points closest to the points of some chosen grid (normally a 6 × 6 arc min grid, but sometimes a 3 × 3 arc min grid in mountainous regions). After the covariance values are determined from a fit to these selected data, predictions are made at the locations of all the observations not selected. Data points where large discrepancies are identified are added to the original list of considered observations. A second fit is then made, and as before, predictions at all remaining, nonchosen locations are performed. Points where large discrepancies occur are again identified, and observations with large residuals are included, and so on. This process is continued until all residuals are less than some chosen cutoff, or the number of iterations reaches a preassigned maximum. The fit tolerance used in this study was 4.0 mGals.

Figure 4 gives the distribution of the number of observations actually used per 30 × 30 arc min quad. One can easily see that the number of observations required to model the gravity surface was usually less than 50. Areas of highly variable topography normally require additional data to model higher frequency effects. A fit in an area of smooth topography that requires a large number of observations is usually traceable to improperly calibrated data, bad data, or data that were tagged with incorrect elevation information. Figure 5 shows the ratio of number of observations used to the number of observations available for the 30 × 30 arc min quads.

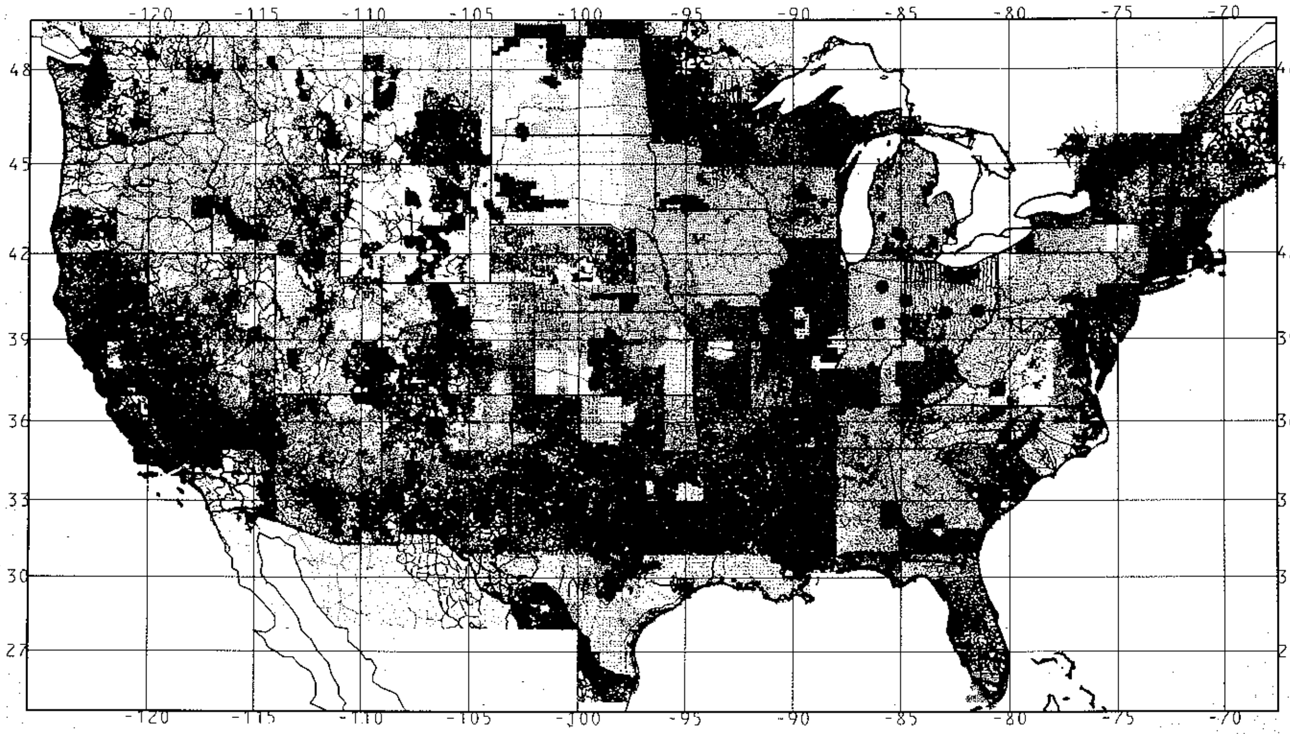

Fig. 3. U.S. National Geodetic Survey gravity data.

5. COVARIANCE ESTIMATION

The basic definition of the global empirical gravity anomaly covariance function on a sphere is given by *Moritz* [1980, section 34] as the expected value over the sphere of the product of all pairs of gravity values located a fixed distance apart. This covariance function may be represented by a Legendre series as given in (4) and may be used as the basis for deriving covariance functions for other gravimetric quantities [for example, *Tscherning and Rapp*, 1974].

A local covariance function is obtained by considering only products of values associated with points within a specific local area. Legendre series may still be used to represent analytically the empirically determined values. However, generally they cannot be used as a basis for the derivation of covariance functions of other quantities. Various attempts have been made to give a more operational definition of a local covariance function since these functions are needed when implementing the method of collocation.

Here we will try to give a definition that explains and justifies what is done in practice when a local covariance function is estimated [for example, see *Tscherning*, 1974, section 4; *Schwarz and Lachapelle*, 1980]. In practice, data outside the areas are disregarded. It is also presupposed that wavelengths larger than the maximal extent of the area have been removed (e.g., by subtracting out the effect of a spherical harmonic expansion or by removing a low degree polynomial trend as described in section 2).

We then arrive at the following definition of a local covariance function:

> A local covariance function is a special case of a global covariance function where the information content of wavelengths longer than the extent of the local area has been removed, and the information outside, but nearby, the area is assumed to vary in a manner similar to the information within the area.

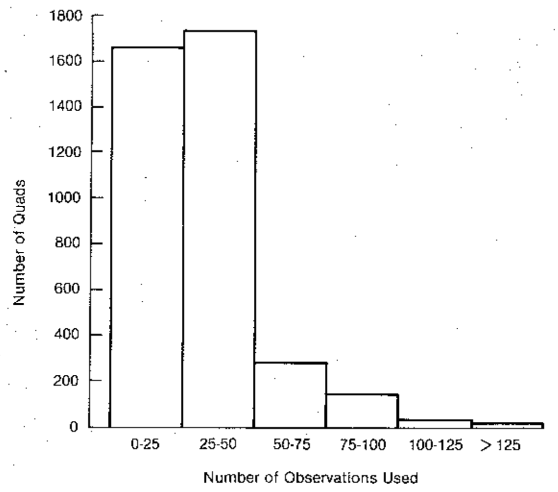

Fig. 4. Historgram of number of quads versus number of observations used.

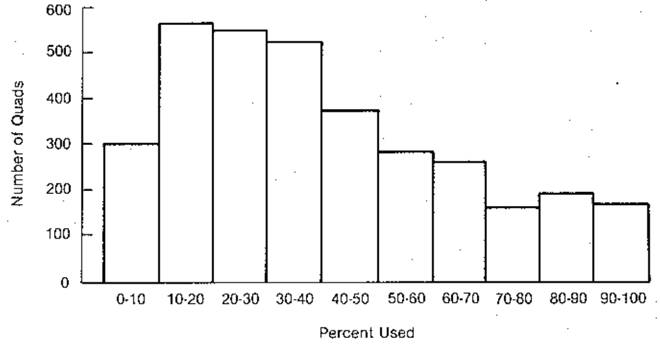

Fig. 5. Ratio of observations used to observations available per quad.

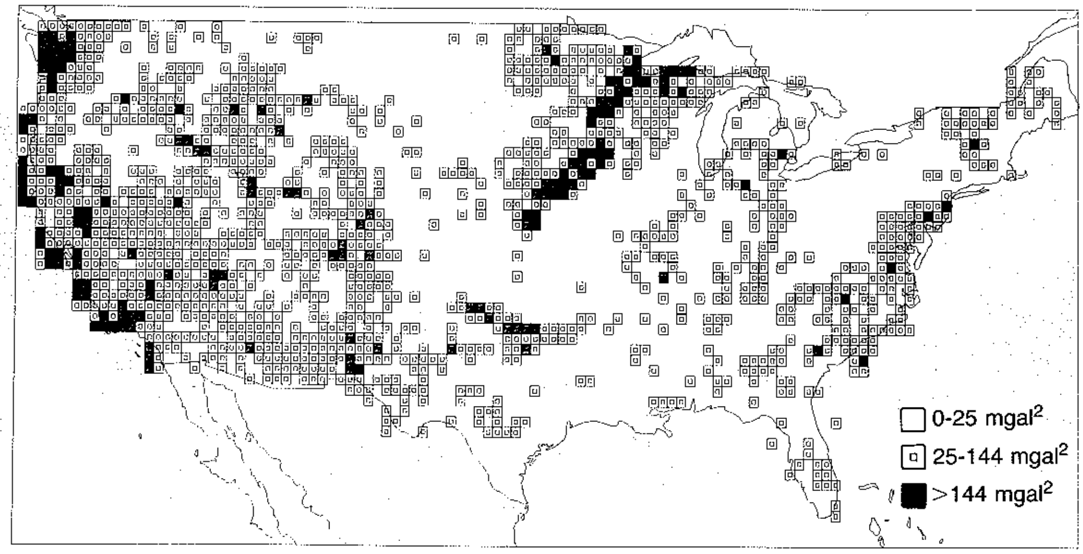

Fig. 6. Distribution of variances for gravity for the continental United States (over 30 × 30 arc min areas).

If the conditions are not fulfilled to a good approximation, then the error estimates of the predictions at or near the edge of the local areas may not be realistic. This also holds for covariance functions of other quantities derived using the Legendre series or equivalent global covariance function models.

In order to calculate the covariances for an area, we must compute the integral of the products of all pairs of anomaly values within the area after subtracting the long-wavelength systematic trend within the area. For this integration, we use the predictions of the signal variations, $\hat{s}$. The integration is approximated by a finite sum, whereby we get the usual estimate from (7) below.

In order to find the needed number of products, we used a varying number of equally distant points, 9, 25, 36, 49, 64, 81, etc., within a 30 × 30 arc min area. It was found that 64 points were sufficient, even for areas with a strongly varying gravity field.

Sixteen sets (or bins) denoted by $l_j, j = 1, 16$ were constructed based on the distance between predicted points. Each set was composed of all pairs of points with distances in the range $[2(j-1), 2j]$ min. Then sample covariances for each set were calculated as follows:

$$C_{l_j} = \sum_{l_j} [\hat{s}_i \hat{s}_k]/N_j \qquad (7)$$

where the summation is over all pairs $(\hat{s}_i, \hat{s}_k)$ in the set l_j and N_j is the number of pairs in the set l_j.

We felt that these small areas and limited data would only support the estimates of variance and correlation distance, C_0 and ψ_0. More dense data over larger areas with similar geophysical properties might enable one to estimate additional parameters, such as curvature [*Moritz*, 1980].

The estimate of C_0 is given by

$$C_0 = \frac{1}{64} \sum_{i=1}^{64} \hat{s}_i^{\,2} \qquad (8)$$

For the estimation of the correlation distance, a model of the covariance behavior must be used. Here, for this purpose we use the following approximation to the covariance function:

$$C(\psi) = C_0[1 - \tfrac{1}{2}(\psi/\psi_0)^2] \qquad (9)$$

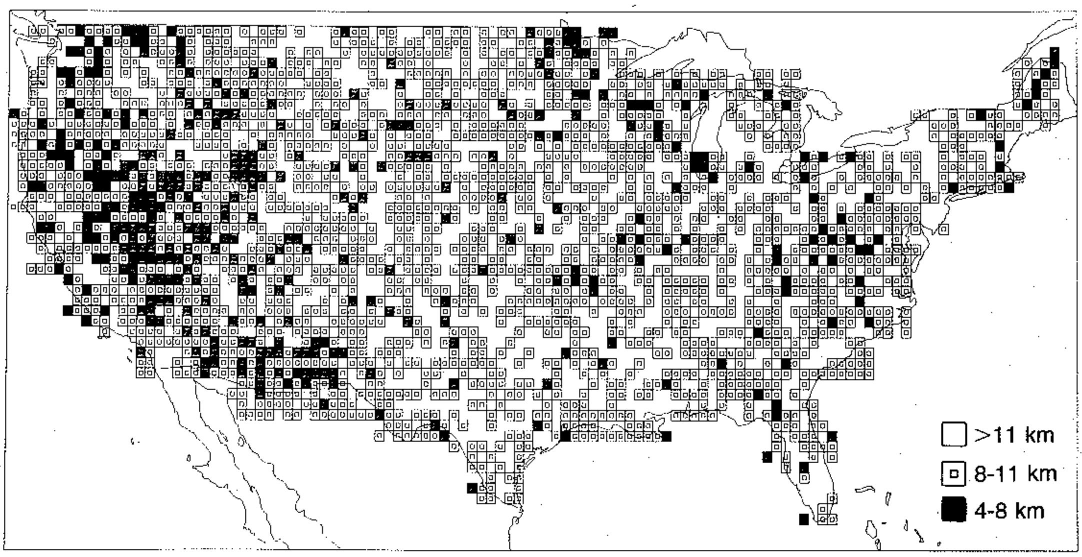

Fig. 7. Distribution of correlation distances for gravity for the continental United States (over 30 × 30 arc min areas).

where ψ is limited to a distance of 21 km or 12 arc min; ψ_0 is the value of ψ for which the covariance is equal to 50% of the variance and is called the correlation distance [*Moritz*, 1980]. It is interesting to note that for small ψ, (9) is equivalent to the third-order Markov undulation model given by *Jordan* [1972]:

$$\phi_{gg} = \sigma_g{}^2(1 + \gamma/D - \gamma^2/2D^2)e^{-\gamma/D}$$

where ϕ_{gg} is $C(\psi)$, $\sigma_g{}^2$ is C_0, γ is ψ, and $D = 1.4\ \psi_0$.

Now using (9) and the sample covariances accumulated for each set, C_{l_j}, with distances between points less than 12 arc min, the correlation distance is estimated as follows:

$$\psi_0 = \left\{\left[\sum_{l_j=l_2}^{l_6} \left(1 - \frac{C_{l_j}}{C_0}\right) \frac{2}{D_{l_j}{}^2}\right] \Big/ N\right\}^{-1/2} \quad (10)$$

where C_0 is available from (8), C_{l_j} is the covariance from set l_j, D_{l_j} is the average distance of pair products, and N is the number of sets used.

In practice, a modified version of (4) is then used which incorporates the C_0 and ψ_0 values:

$$C(\psi) = C_0\left[\sum_{n=360}^{\infty} c_n{}'\rho^n P_n(\cos\psi)\right] \quad (11)$$

where the ρ is chosen to provide the proper correlation distance (ψ_0) and the prime denotes that all c_n have been scaled by a common factor so the summation inside the bracket is equal to unity for $\psi = 0$.

6. Geophysical Interpretation

Figures 6 and 7 summarize the results of the covariance function study for the continental United States. Figure 6 is probably the most striking of the two, since it reveals a clear coincidence of large variance values in the vicinity of the well-known midcontinent gravity high. This geophysically interesting area has been studied extensively by several investigations. For example, *Ocala and Meyer* [1973] have shown that all data collectively (gravity and seismic profiles) can be used to argue for a rift system. They quote seismic velocities from 3.5 to 6.9 km/s, which is qualitatively in agreement with our large variances. But none of their local planar crustal models appear to show spatial variations shorter than 20 km, also in agreement with the rather long correlation distances found in this area.

Another area with large variances and long correlation distances apparent in Figures 6 and 7 is the Wichita-Arbuckle structural trend in the Texas-Oklahoma area. Groups of large variance values are also found on the Pacific Coast. Medium-sized variance values seem to be correlated with topography. Isostatic compensation should exhibit long wavelength behavior, generally speaking. However, short-wavelength contributions from isostatic compensation would contribute to our derived variance values. This also would contribute to the correlation found between large variances and topography. The maximum value found was 1136 mGal^2. Variance values may have been underestimated in some areas due to a lack of adequate data coverage. This would be most likely in the mountainous areas of the northwest, which can be seen by an examination of the gravity data distribution in Figure 3. On the average, the signal correlation distances for the United States are in the 8- to 11-km range. Short distances are associated mainly with areas of high topography.

7. A Simple Error Expression

If no systematic contributions were present in equation (2) (no AX term), then the uncertainty of a prediction can be approximated by

$$\sigma_p{}^2 \approx C_0 - C(\psi_p)C_0{}^{-1}C(\psi_p) \quad (12)$$

where ψ_p is the distance to the closest observed gravity value.

Substituting for $C(\psi)$ from (9) into (12) yields

$$\sigma_p{}^2 \approx C_0 \frac{\psi_p{}^2}{\psi_0{}^2}\left[1 - \frac{\psi_p{}^2}{\psi_0{}^2}\right] \quad (13)$$

This approximation can be used with local values of C_0 and ψ_0 to determine the maximum spacing for a given accuracy or, alternatively, the approximate uncertainty for a given distance ψ_p to the closest measured value of gravity (assuming, of course, that the noise contribution is small). Instead of using the closest point, the "mean distance" between the points in an area may be used, e.g., equal to the square root of the area divided by the number of points. The corresponding equations for this case can be found by *Forsberg and Tscherning* [1981]. This was, in fact, the way the model equation (9) was found, because it explained the number of observations needed to obtain a predefined mean prediction error [see *Forsberg and Tscherning*, 1981, Figure 1].

8. Conclusions

Variances and correlation distances for 3637 30 × 30 arc min areas using over 544,000 terrain-corrected Bouguer anomalies have been determined for the continental United States. Short correlation distances are associated with the midcontinental gravity high and areas near the Pacific Coast.

The parameters are quite similar in some areas but vary greatly in other areas. In general, the covariance functions must be considered nonstationary.

However, the values can be used to properly compute uncertainties in predicted gravity when using the systematic surfaces and interpolation coefficients from which the statistical information was obtained. The validity of the error estimates for points at the boundaries of the block will naturally have to be judged, considering the change of the parameters from the actual block to the neighboring block.

Also areas requiring additional observations in order to obtain better surface fits can be easily identified.

Acknowledgment. This paper is a contribution to a project supported by NATO grant 320/82.

References

Chatfield, A. B., M. M. Bennett, and T. Chen, Effect of gravity model inaccuracy on navigation performance, *AAIA J.*, *13*(11), 1494–1501, 1975.

Forsberg, R., and C. C. Tscherning, Deflection and gravity anomaly prediction for inertial surveying using collocation, Proceedings of the Second International Symposium on Inertial Technology for Surveying and Geodesy, pp. 89–95, Can. Inst. of Surv., Ottawa, 1981.

Jordan, S. K., Self-consistent statistical models for the gravity anomaly, vertical deflections, and undulation of the geoid, *J. Geophys. Res.*, *77*(20), 3660–3670, 1972.

Kaula, W. M., *Theory of Satellite Geodesy*, Blaisdell Publishing, Waltham, Mass., 1966.

MacMillan, W. D., *The Theory of the Potential*, Dover, New York, 1958.

Moritz, H., *Advanced Physical Geodesy*, H. Wichmann Verlag, Karlsruhe, Germany, 1980.

Ocala, L. C., and R. P. Meyer, Central North American rift system, 1, Structure of the axial zone from seismic and gravimetric data, *J. Geophys. Res.*, *78*(23), 5173–5199, 1973.

Schwarz, C. R., Deflection computations for network adjustment in the United States, in *Proceedings of the Second International Symposium on Problems Related to the Redefinition of North American Geodetic Networks*, pp. 91–102, NOAA/NGS, Rockville Md., 1978.

Schwarz, K. P., and G. Lachappelle, Local characteristics of the gravity anomaly covariance function, *Bull. Geod.*, *54*, 21–36, 1980.

Tscherning, C. C., A FORTRAN IV program for the determination of the anomalous potential using stepwise least squares collocation, *Rep. 212*, Dep. of Geod. Sci., Ohio State Univ., Columbus, 1974.

Tscherning, C. C., Application of collocation for the planning of gravity surveys, *Bull. Geod.*, *116*, 183–198, 1975.

Tscherning, C. C., The role and computation of gravity for the processing of levelling data, Proceedings of the Second International Symposium on Problems Related to the Redefinition of North American Vertical Networks, pp. 505–524, Can. Inst. of Surv., Ottawa, 1980.

Tscherning, C. C., and R. H. Rapp, Closed covariance expressions for gravity anomalies, geoid undulations, and deflections of the vertical implied by anomaly degree-variance models, *Rep. 208*, Dep. of Geod. Sci., Ohio State Univ., Columbus, 1974.

M. M. Chin and C. C. Goad, National Geodetic Survey, Charting and Geodetic Services, National Ocean Service, National Oceanic and Atmospheric Administration, Rockville, MD 20852.

C. C. Tscherning, Geodetic Institute, Gamlehave Alle 22, DK-2920 Charlottenlund, Denmark.

(Received August 15, 1983;
revised April 11, 1984;
accepted May 15, 1984.)