

A Comparison of Interpolation Methods for Sparse Data: Application to Wind and Concentration Fields

WILLIAM R. GOODIN,¹ GREGORY J. McRAE AND JOHN H. SEINFELD

Environmental Quality Laboratory, California Institute of Technology, Pasadena 91125

(Manuscript received 1 December 1978, in final form 23 February 1979)

ABSTRACT

In order to produce gridded fields of pollutant concentration data and surface wind data for use in an air quality model, a number of techniques for interpolating sparse data values are compared. The techniques are compared using three data sets. One is an idealized concentration distribution to which the exact solution is known, the second is a potential flow field, while the third consists of surface ozone concentrations measured in the Los Angeles Basin on a particular day. The results of the study indicate that fitting a second-degree polynomial to each subregion (triangle) in the plane with each data point weighted according to its distance from the subregion provides a good compromise between accuracy and computational cost.

1. Introduction

A problem common to many disciplines is the development of continuous fields from discrete data sets. For example, in meteorology, wind fields are often generated using a two-step procedure. The first element is the interpolation of the raw station data to a finer mesh. Objective analysis procedures are then employed to adjust the wind vectors at each grid point so that an applied physical constraint, such as minimum field divergence, is satisfied. Much of the literature is devoted to the second step; what is frequently neglected is that the final form of the field is often critically dependent on the results of the initial interpolation. Formally, the objective of this paper is to address the problem: given a bounded region of r -space containing n , error-free data values C_i , at locations $\mathbf{x}^i = [x_1^i, \dots, x_r^i]$, $i = 1, 2, \dots, n$, develop a function, $f(\mathbf{x})$, which will assign a value of C at any arbitrary location \mathbf{x} . While simply stated, there is, in general, no unique solution to the interpolation problem. As a result, when alternative techniques are applied to the same discrete data set, different fields are generated. This study was undertaken to identify and test computationally efficient methods for interpolating sparse data measurements onto a regular mesh.

2. A survey of methods for interpolation of sparse data

a. Weighted interpolation methods

A common approach to interpolation of sparse data onto a regular grid is to assume that the grid value

is some weighted average of the surrounding data values, i.e.,

$$C_{ij} = \frac{\sum_{k=1}^n C_k W_k(r)}{\sum_{k=1}^n W_k(r)}, \quad (1)$$

where C_k is the measured value at the k th measuring station, $W_k(r)$ the weighting function, and r the distance from the grid point to the station.

In an early study, Cressman (1959) reported on a procedure for use in pressure-surface height analysis which used the weighting factor

$$W(r) = \frac{R^2 - r^2}{R^2 + r^2}, \quad (2)$$

where R is the distance at which the weighting factor goes to zero, i.e., the "radius of influence." This weighting technique aided the interpolation procedure in areas of sparse data. Decreasing values of R were used on successive scans to analyze a spectrum of scales. The values obtained from each scan were then averaged to produce the final field.

Endlich and Mancuso (1968) combined both polynomial fitting and distance weighting in their interpolation technique. A least-squares fit to a first-order polynomial was performed using five of the nearest station values, according to

$$W(r) = \frac{a}{(r + r^*)^2 + a}, \quad (3)$$

where a is a constant, r the distance to the station and r^* a distance factor ($0 \leq r^* \leq r$) that indicates whether the observation is in an upwind-downwind ($r^* = r$) or crosswind ($r^* = 0$) direction from the grid point.

¹ Present affiliation: Advanced Technology Group, Dames & Moore, Los Angeles, CA 90024.

Shepard (1968) discussed an interpolation technique in which a direction factor was also included which accounted for shadowing of the influence of one data point by a nearer one in the same direction. The method also included the effect of barriers. If a "detour" of length $b(r)$, perpendicular to the line between the point (i, j) and the k th measuring station, was required to travel around the barrier between the two points, then $b(r)$ was considered to be the length of the barrier. An effective distance r' was defined by

$$r' = [r^2 + b(r)^2]^{\frac{1}{2}}. \quad (4)$$

If no barrier separated the two points, then $b(r) = 0$.

Shenfeld and Boyer (1974) presented a technique for interpolation of a velocity field similar to that proposed by Endlich and Mancuso. The velocity was computed as in Eq. (1). For example, for the x component of the velocity at grid point (i, j) ,

$$u_{ij} = \frac{\sum_{k=1}^n u_k W_k(r)}{\sum_{k=1}^n W_k(r)}, \quad (5)$$

with the weighting function defined by

$$W_k(r) = \exp \left[- \left(\frac{x_k^2}{S_{xk}} + \frac{y_k^2}{S_{yk}} \right) \right], \quad (6)$$

where $x_k^2 + y_k^2 = r^2$. The coordinate system was oriented in the direction of the observed wind with the origin at the k th station. S_{xk} was defined by

$$S_{xk} = S_y \left(1 + \frac{V_k}{V_s} \right), \quad (7)$$

where V_k was the magnitude of the velocity vector at the k th station and V_s a scaling velocity. S_y was assigned a value between 40 and 70, depending on the density of wind stations. The larger the number of stations, the lower the value of S_y . For low values of S_y , the computed wind velocity at any grid point was more dependent on nearby stations. Also since S_{xk} was always greater than or equal to S_y , the computed wind velocity was more dependent on downwind distance (x_k) than on crosswind distance (y_k).

For interpolation in regions of sparse data, Fritsch (1971) used a cubic spline technique. He first fitted spherical surfaces to the data to obtain an initial field, and then iteratively adjusted these values using the splines until convergence was obtained. He compared his technique with that of Cressman using an idealized data set with a known solution, and the mean error ($\sim 3\%$) was approximately half that of Cressman's.

MacCracken and Sauter (1975) used a Gaussian weighting scheme to eliminate complete dominance of a measuring station near a grid point, i.e.,

$$W(r) = \exp(-0.1r^2). \quad (8)$$

Hovland *et al.* (1977) computed wind and temperature fields using data from the Environmental Protection Agency's Regional Air Pollution Study (RAPS) conducted in St. Louis. An iterative scan procedure was used in which the radius of influence was decreased and the number of stations increased empirically on successive iterations. The advantage to this strategy is that small-scale motions which are only detected in an area of dense station coverage are not transmitted to outlying areas. Moreover, during the initial iterations this procedure places significant weight on outlying stations which may be less reliable than those in the center of the region. The weighting function used was

$$W(r) = \begin{cases} \left(\frac{R^2 - r^2}{R^2 + r^2} \right)^4, & r \leq R \\ 0, & r > R. \end{cases} \quad (9)$$

This function decreases rapidly with increasing distance r from a maximum of 1 at $W(0)$.

Recently, Boone and Samuelson (1977) described the application of a distance and directional weighting technique to the display of air pollution data. The weighting factor used in Eq. (1) was

$$W_k(r) = S_k^2 (1 + T_k), \quad (10)$$

where, based on the work of Shepard (1968), the weighting factor was defined by

$$S_k = \begin{cases} 1/r_k, & 0 < r_k \leq R/3 \\ \frac{27}{4R} \left(\frac{r_k}{R} - 1 \right)^2, & R/3 < r_k \leq R \end{cases} \quad (11)$$

where r_k is the distance from the k th station to the (i, j) grid point. The directional weighting factor is computed from

$$T_k = \frac{\sum_{l=1}^m S_l (1 - \cos A)}{\sum_{l=1}^m S_l}, \quad (12)$$

where m is the number of points within the radius R and angle A is defined by the segments (x_k, y_k) , (i, j) and (i, j) , (x_l, y_l) .

b. Least-squares polynomial interpolation

The second class of methods for producing a continuous surface over a grid is a least-squares fit of a polynomial to the data points. The technique requires minimization of χ^2 , the goodness of fit to the data. In a second-degree polynomial, for example,

$$\begin{aligned} \chi^2 &\equiv \sum_{k=1}^n (\Delta C_k)^2 \\ &= \sum_{k=1}^n (C_k - a_1 - a_2 x_k - a_3 y_k - a_4 x_k y_k - a_5 x_k^2 - a_6 y_k^2)^2 \end{aligned} \quad (13)$$

must be a minimum, where C_k is the measured concentration (or wind speed) at point (x_k, y_k) . The minimum value of χ^2 can be determined by setting the derivatives of χ^2 with respect to each of the coefficients a_i equal to zero. For a second-degree polynomial, six simultaneous equations must be solved for the optimum coefficients. The concentration $C(x, y)$ at any grid location (x, y) can then be computed from

$$C(x, y) = a_1 + a_2x + a_3y + a_4xy + a_5x^2 + a_6y^2. \quad (14)$$

In an early paper, Panofsky (1949) used third-degree polynomials to fit wind and pressure fields for use in weather map construction. The technique was later modified to handle areas with sparse data by Gilchrist and Cressman (1954). These polynomial-fitting procedures were typically applied to the whole grid.

An alternative to applying the polynomial interpolation to the entire grid is to perform the interpolation over areas within the influence of individual stations. Thiessen (1911) defined space-filling polygons over particular areas by assuming that each station measurement is associated with the local region of the area nearer to that station than to any other. Formally, the resultant planar divisions are defined as Dirichlet tessellations although they are sometimes called Voroni or Thiessen polygons. While the recent algorithm of Green and Sibson (1978) simplifies the task of defining the polygons, the attainment of slope continuity in C from one region to another is a difficult problem.

A simpler technique is to triangulate the region using the station locations for the vertex positions. Lawson (1977) described a number of algorithms which perform this task. Given n_b points on the convex boundary of the region, the number of non-overlapping triangles $n_t = 2n - n_b - 2$ is unique even though the triangulations may be different. Once the triangle vertices have been established, a variety of interpolation schemes can be used. For example, $C(x, y)$ within each triangle can be determined from the equation of a plane oriented to pass through the three vertex data points.

A natural extension of this idea is to use higher order polynomials to achieve slope continuity between adjacent triangles. Lawson (1977) employed a cubic polynomial based on the finite element method of Clough and Tocher (1965) to obtain

$$C(x, y) = \sum_{i=0}^M \sum_{j=0}^{M-i} a_{ij} x^i y^j, \quad (15)$$

where $M=3$. Powell and Sabin (1977) used piecewise quadratic approximation to obtain an interpolation function with continuous first derivatives.

Akima (1975) developed a method of bivariate interpolation and smooth surface fitting for irregularly spaced data based on a fifth-degree polynomial ($M=5$) in x and y defined in each triangular cell. For each

polynomial 29 coefficients must be determined. In addition to the values of the function at the data locations, the first- and second-order partial derivatives are also required. Partial derivatives of the function differentiated in the direction perpendicular to each side of the triangle are considered in order that the resulting polynomials intersect smoothly at the triangle edge. Use of higher order polynomials requires more coefficients and accordingly more computer time to solve for the coefficient matrix a_{ij} .

McLain (1974, 1976) has discussed a different polynomial fitting approach, in which the domain is divided into triangular subregions by connecting the data points. A second-degree polynomial is then fitted to each triangle using all data points with each value weighted according to its distance from the given triangle. The weighting scheme r^{-2} was used in the present formulation with the radius of influence effectively set to the dimension of the region. At each edge of the triangle, $C(x, y)$ is required to be the same as that of adjacent polynomials in order that position continuity of the resulting surface is assured.

In each triangle, the final value $C(x, y)$ is found as the weighted average of the three functions [Eq. (14)] corresponding to the vertices, i.e.,

$$C(x, y) = W_1C_1 + W_2C_2 + W_3C_3. \quad (16)$$

To ensure smooth transition from one triangle to the next, each weight W_i and its leading derivatives must be identically zero along the side of the triangle opposite to the i th vertex. This can be achieved by making W_i proportional to the third power of the distance from that side. The distance d_i from the point (x, y) to the side is a linear function of x and y , i.e.,

$$d_i = l_i x + m_i y + n_i, \quad (17)$$

where l_i, m_i, n_i are the coefficients defining a line through triangle edge opposite vertex i , scaled such that $d_i=1$ at vertex i . The weight W_i , within the triangle, is then

$$W_i = d_i^3 / (d_1^3 + d_2^3 + d_3^3). \quad (18)$$

c. Optimum interpolation

A third general technique, first presented by Gandin (1963), known as optimum interpolation, has proved useful for interpolation of synoptic-scale meteorological data (wind and temperature) for initializing global or synoptic circulation models (Dartt, 1972; Schlatter, 1975; Julian and Thiebaux, 1975; Thiebaux, 1975). The technique produces an interpolated field from data points irregularly spaced in both space and time. The interpolation function can be formulated directly in terms of statistical properties of the wind (temperature) field, i.e., past behavior determines the function's form. This implies, however, that a historical record of data values must be available in

TABLE 1. Summary of interpolation methods and their attributes.

Method	Attributes	Applications
Weighted interpolation	1. Easy to implement 2. May be modified if directional influence is important 3. More features of the data are smoothed as n decreases 4. Influence of a station becomes very localized as n (or a) increases 5. Radius of influence R may be fixed or variable	Cressman (1959) Endlich & Mancuso (1968) Shepard (1968) Shenfeld & Boyer (1974) MacCracken & Sauter (1975) Hovland <i>et al.</i> (1977) Boone & Samuelson (1977)
(a) $W(r) = \left(\frac{R^2 - r^2}{R^2 + r^2}\right)^n$ (b) $W(r) = \exp(-ar^n)$ (c) $W(r) = r^{-n}$		
Least-squares polynomial interpolation	1. Complex to implement 2. Resulting field depends strongly on distribution of data points when using (a) 3. Resulting field is smoothed as n decreases when using (a) 4. (b) fits data almost exactly 5. Execution time increases with n	Panofsky (1949) Akima (1975) McLain (1974, 1976) Lawson (1977)
(a) Polynomial of degree n fitted to full grid (b) Polynomials of degree n fitted to subregions of grid		
Optimum interpolation	1. Complex to implement 2. Much historical data may be required 3. Statistics of the data accounted for	Gandin (1963) Dartt (1972) Schlatter (1975)

order to calculate these statistical properties (covariances). Moreover, the variance from the long-term mean of a wind measurement taken at a particular hour may not be useful for a mesoscale flow since the local stability and terrain influences on a particular day may be critically important in determining the flow.

d. Summary

Table 1 summarizes the methods that have been discussed in Section 2.

3. Optimum radius of influence

Theoretically, the choice of weight function and radius of influence for use in a distance-weighted interpolation procedure depends on the field statistics. Since insufficient measured data are available in most practical applications, the choice of an optimum radius must be based on the average station separation. For a two-dimensional domain of area A with N stations randomly distributed over the plane, the average station density ρ and separation d are given by $\rho = N/A$ and $d = (A/N)^{1/2}$. Stephens and Stitt (1970) have shown empirically that the optimum search radius R for large signal-to-noise ratios is $R/d \approx 1.6$. A further finding of their work is that R should be overestimated rather than underestimated. In practice this is important since nonrandom station distributions can lead to situations in which R is commensurate with the dimension of the data separation. In a separate study, Barnes (1964, 1973) came to similar conclusions and was also able to relate R/d to L/d , where L is the half wavelength of a disturbance. As an example, the parameters d and R/d can be calculated for the network of approximately 50 air monitoring stations in the South Coast Air Basin of Southern

California (in 1974). Distributed over an area of about 19 400 km², the average station separation is calculated to be ~ 20 km. This implies that a radius of influence between 20 and 40 km would be optimum. In order to calculate air quality distributions over the ocean and desert areas, which have few measurement stations, a larger radius of influence of 30–50 km must sometimes be used.

Careful consideration must be given to the choice of R . If the objective is to produce a set of gridded values, then each point must be within the radius of influence of at least one measurement station. While increasing R reduces the field variance, it does so at an increase of computational cost. For randomly distributed stations the interpolation cost per grid value increases quadratically since the number of data points within each radius of influence is proportional to R^2 . Thus, knowledge of an "optimum" radius of influence is important both from the point of view of accuracy of the final interpolated field as well as cost of the calculation. The simple technique discussed in this section for computing an optimum radius of influence, while developed for a random distribution of points, provides a useful range of values even when the measurement points are not randomly located.

4. Comparison of interpolation procedures for scalar fields

The results of interpolating test data sets for which the exact solution is known provide a means to assess the performance of interpolation routines.

A test data set consisting of a hemispherically shaped surface was constructed to compare the various interpolation methods. The hemisphere was arbitrarily centered above the center of a 100×50 grid system overlaying the South Coast Air Basin. Each grid cell was 3.2 km×3.2 km. The height of the surface above

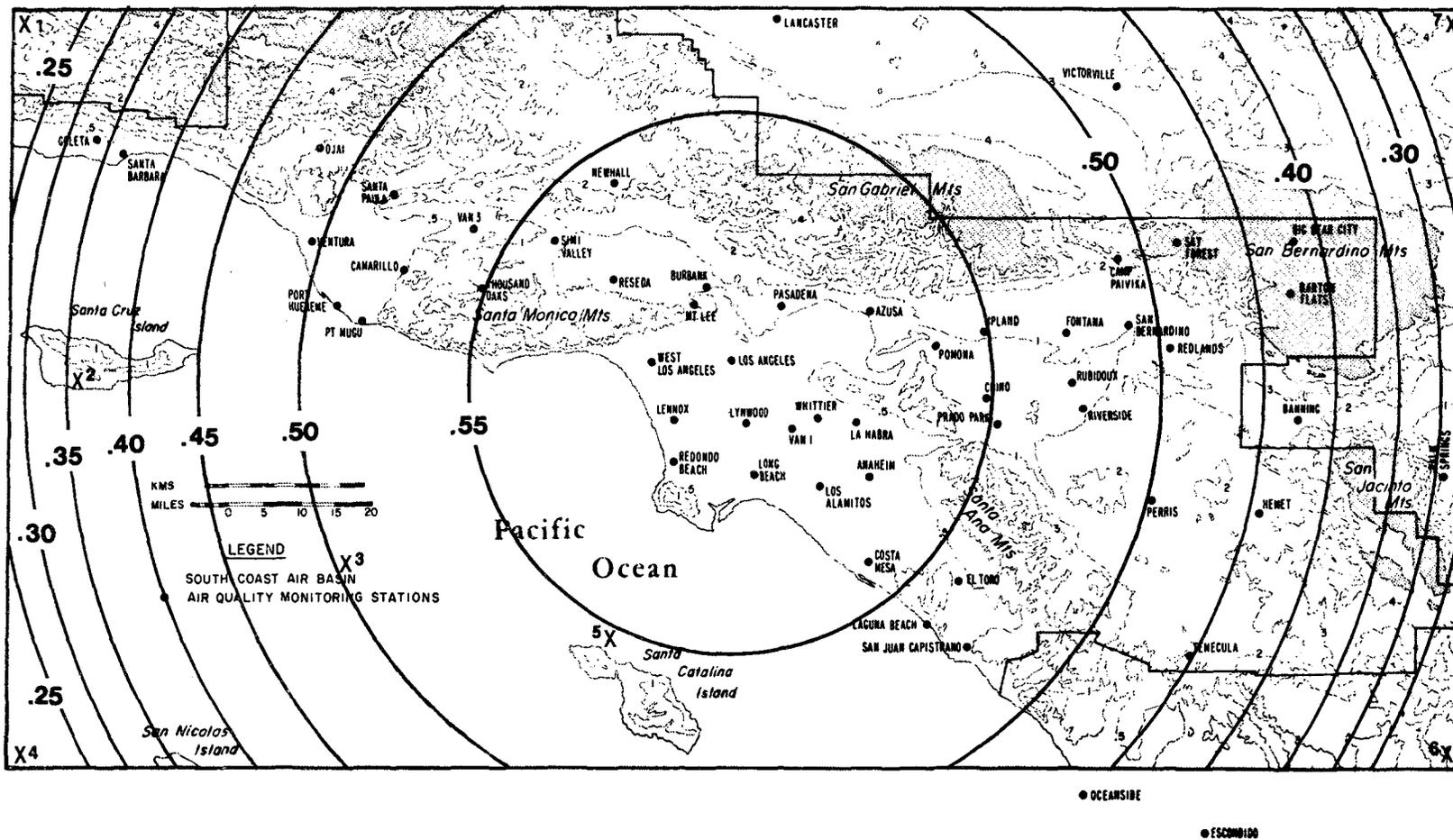


FIG. 1. Distribution of measurement stations and hemispherical test surface.

each grid point corresponded to the field value at that point. The height of the constructed surface was evaluated at each measuring station location (56 points) and these values were used as raw data in the test. Fig. 1 shows the field isopleths and the measuring station locations. The test consisted of attempting to reconstruct the hemispherical surface from these data points (no flow barriers were used).

The methods compared are the second-degree polynomial with r^{-2} weighting and the simple distance-weighting scheme with various weights, the results of which are shown in Table 2. Statistics are given for the station locations as well as for all grid points; in each case the percentage relative residual error E_r is computed from

$$E_r = 100 \left[\frac{C_{\text{comp}}}{C_{\text{act}}} - 1 \right], \quad (19)$$

where C_{comp} is the computed surface height and C_{act} its actual height. When an interpolation scheme is applied to a set of field data, the residual error at each station is the only measure of the technique's performance. However, for this test data set, the residual error can be computed over the whole grid, indicating the level of accuracy that can be expected away from measuring stations. Such a procedure cannot be expected to resolve sharp gradients which are not reflected in the data.

Contour plots of the reconstructed fields are shown in Fig. 2. Generally, in regions where the number of measuring stations is relatively large, the error is smallest (i.e., toward the center of the region). However, near the boundaries where the data network is less dense the errors can be much larger.

The results indicate that among the simple distance-weighting schemes, the $(R^2 - r^2)/(R^2 + r^2)$ weighting produces the most hemisphere-like surface. However, the departure from measured values is largest with

this scheme. More accuracy is obtained near the data points with the r^{-2} and $\exp(-br^2)$ schemes, but at the expense of accuracy away from measuring stations. The second-degree polynomial fitting procedure provides a compromise. The hemispherical shape is preserved and the accuracy at the measuring station locations is also acceptable. The execution times do not include the square-to-station distance (or formulation of the triangles) calculation times since they need only be calculated once for a given set of stations.

5. Application of selected interpolation procedures to air quality data

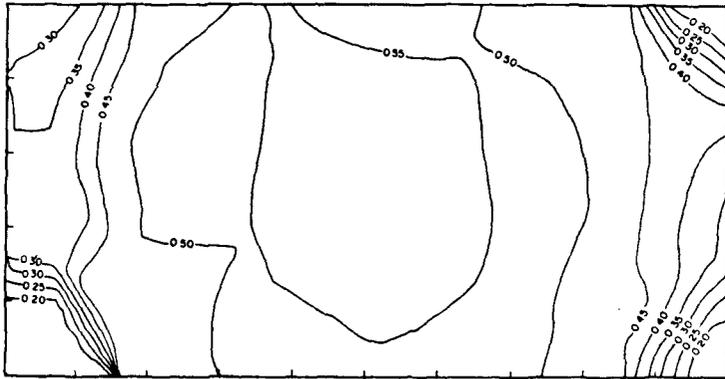
On the basis of the simple problem described above as well as a variety of other test cases, three techniques were selected for further examination using actual measured data. These were distance-weighting schemes r^{-2} and $\exp(-0.1r^2)$ and second-degree polynomial fit with r^{-2} weighting. The distance-weighting schemes were chosen based on a compromise between station accuracy, accuracy over the whole grid, and cost. The actual data chosen for testing purposes were ozone measurements on 26 June 1974 taken within the South Coast Air Basin. All data have been corrected for measurement as well as interference errors. The distance-weighting techniques were tested using a fixed radius of influence of 48 km as well as a variable radius of influence. The variable radius of influence was specified to include at least two data points.

The results of the ozone data interpolation tests are displayed in Fig. 3 and Table 3. The r^{-2} weighting scheme produces the smallest residuals in the vicinity of the measuring stations whether a fixed or variable radius of influence is used. While minimizing the station residuals, the overall field variance for the r^{-2} weighting can be much higher than the polynomial fitting procedure as demonstrated in the test problem. Fig. 4 shows a three-dimensional perspective plot of

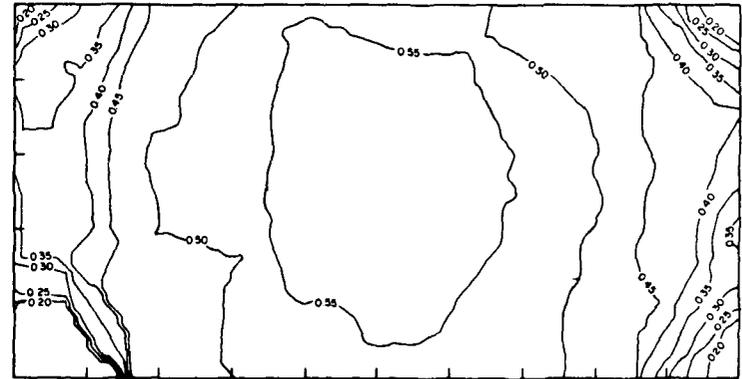
TABLE 2. Percentage relative residual error tabulated for hemispherical surface, for various weighting functions using 56 stations and a radius of influence (R) of 48 km.

Weighting function* $W(r)$	Residual error E_r (%) at measuring stations				Residual error E_r (%) at all grid points				Relative computation time
	Mean	Minimum	Maximum	Std. Dev.	Mean	Minimum	Maximum	Std. Dev.	
r^{-1}	0.06	-1.27	2.31	0.57	0.52	-52.47	52.12	8.23	2.1
r^{-2}	<0.01	-0.14	0.29	0.06	0.03	-52.47	52.10	8.59	1.4
r^{-3}	<0.01	-0.04	0.11	0.02	-0.38	-52.47	52.08	9.19	3.4
$\frac{R^2 - r^2}{R^2 + r^2}$	0.93	-9.29	36.64	5.51	-0.06	-52.47	52.08	9.47	1.7
$e^{-0.1r^2}$	0.02	-3.28	3.58	0.72	-1.07	-55.88	54.77	12.42	2.3
$e^{-0.5r^2}$	<0.01	-0.66	0.72	0.16	-1.14	-58.0	58.27	13.15	2.5
Second degree polynomial (r^{-2} weighting)	0.09	-0.75	0.87	0.32	-1.78	-18.92	10.64	4.0	1.0

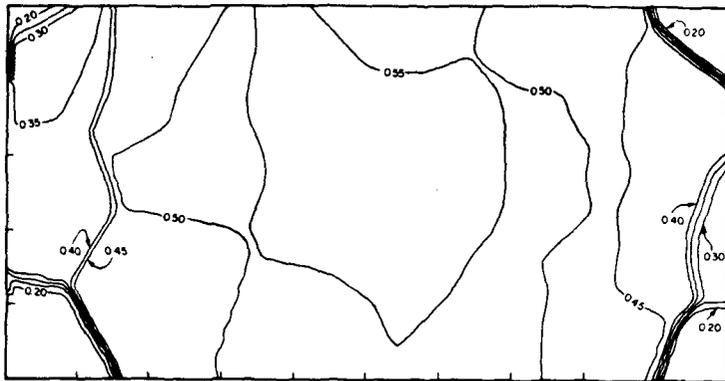
* r is the distance from grid point to measuring station; for $r > R$, $W(r)$ is set to zero.



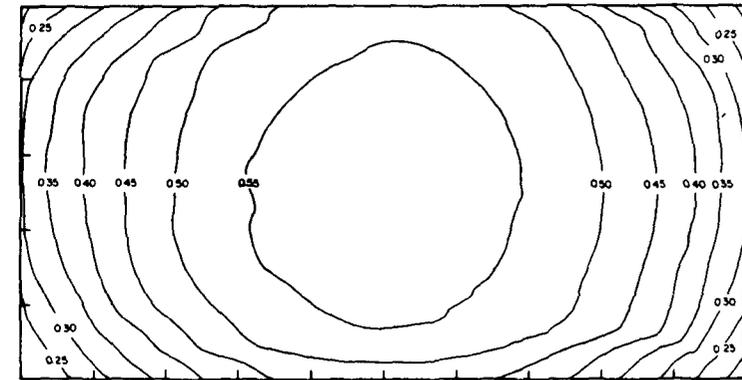
(a)



(b)

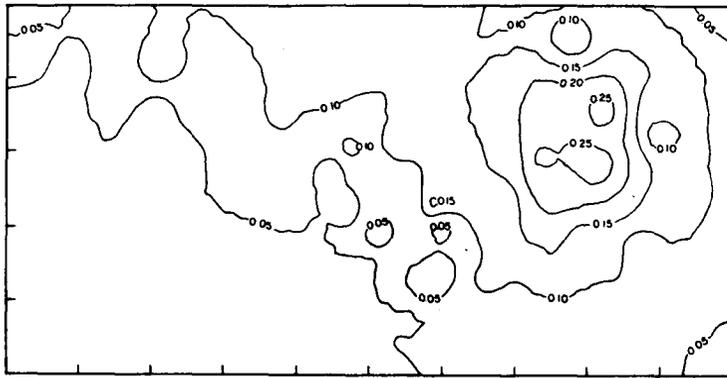


(c)

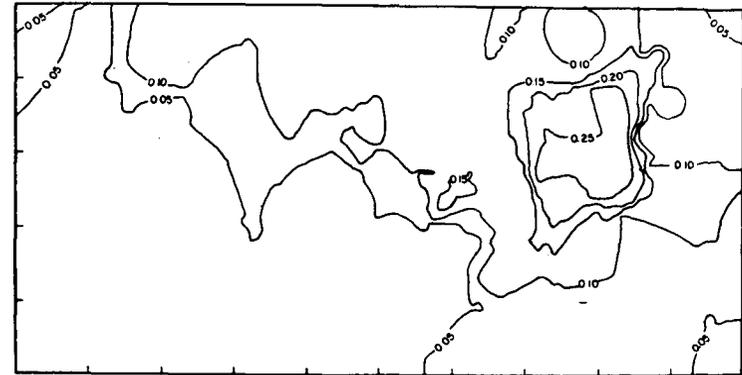


(d)

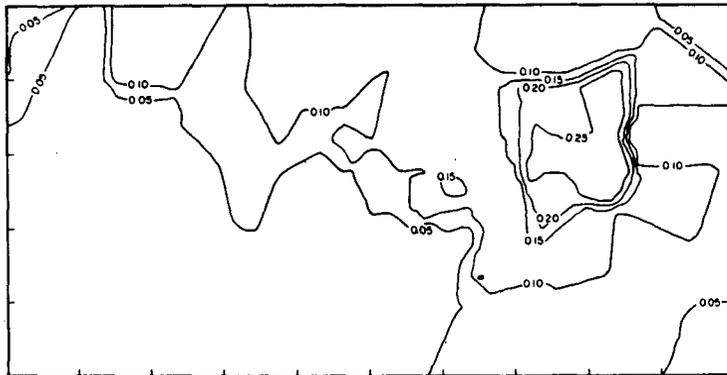
FIG. 2. Surfaces resulting from interpolation of hemispherical data using (a) $(R^2 - r^2)/(R^2 + r^2)$ weighting of data values, (b) r^{-2} weighting of data values (similar visual results were obtained for the r^{-1} and r^{-3} cases), (c) $\exp(-0.1r^2)$ weighting of data values (similar visual results were obtained for $\exp(-0.5r^2)$), and (d) polynomial fitting procedure.



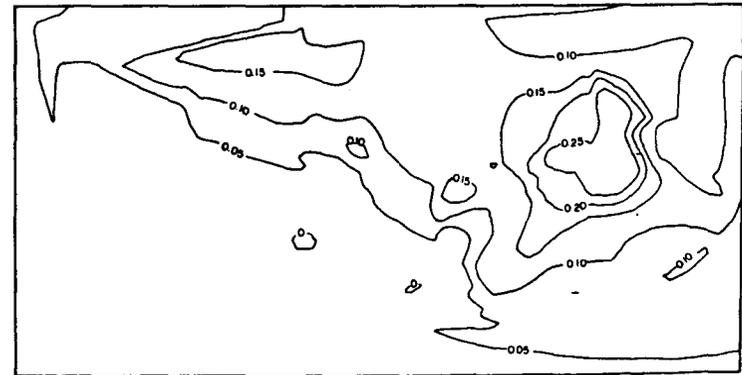
(a)



(b)



(c)



(d)

FIG. 3. Surfaces resulting from interpolation of measured ozone values from 1600 PST 26 June 1974 using (a) r^{-2} weighting of data values with a fixed radius of influence R , (b) r^{-2} weighting of data values with a variable R , (c) $\exp(-0.1r^2)$ weighting of data values with a variable R (similar visual results were obtained with a fixed R), and (d) polynomial fitting procedure.

TABLE 3. Percentage relative residual error E_r , tabulated from ozone data analysis for various weighting functions using 56 stations.

Weighting function* $W(r)$	Residual error E_r (%) at stations with radius of influence $R=48$ km				Residual error E_r (%) at stations with variable radius of influence R			
	Mean	Minimum	Maximum	Std. Dev.	Mean	Minimum	Maximum	Std. Dev.
r^{-2}	0.67	-3.08	10.42	2.04	0.14	-3.22	3.79	1.00
$e^{-0.1r^2}$	5.56	-28.75	85.48	19.62	3.01	-21.59	71.72	15.69
Second degree polynomial (r^{-2} weighting)	4.82	-13.02	49.18	12.60				

* r is the distance from grid point to measuring station, for $r > R$, $W(r)$ is set to zero.

the ozone surface generated by the polynomial fitting procedure.

6. Application of interpolation procedures to vector fields

The procedures outlined in Section 2 can easily be applied to vector as well as scalar fields, for example, by treating the u and v components separately. Based on the results of the hemisphere test data set, the same three techniques used in Section 5 were selected for further detailed examination here. These were distance-weighting schemes r^{-2} and $\exp(-0.1r^2)$ and second-degree polynomial fit with r^{-2} weighting.

The test wind field data set, potential flow over a flat plate inclined at angle α to the flow, free stream velocity v_0 , was chosen to illustrate the inclusion of a barrier to flow. The exact solution to the problem can be calculated from potential flow theory. The velocity potential ϕ and streamfunction ψ in an x - y Cartesian

coordinate system are given by

$$\phi(x,y) = v_0 \left(x + \frac{x \cos 2\alpha - y \sin 2\alpha}{x^2 + y^2} \right), \tag{20}$$

$$\psi(x,y) = v_0 \left(y - \frac{x \sin 2\alpha + y \cos 2\alpha}{x^2 + y^2} \right). \tag{21}$$

The velocity components are computed from the velocity potential

$$u(x) = \frac{\partial \phi}{\partial x} = v_0 \left(1 + \frac{\cos 2\alpha (y^2 - x^2) + 2xy \sin 2\alpha}{(x^2 + y^2)^2} \right), \tag{22}$$

$$v(y) = \frac{\partial \phi}{\partial y} = v_0 \left(\frac{\sin 2\alpha (y^2 - x^2) - 2xy \cos 2\alpha}{(x^2 - y^2)^2} \right). \tag{23}$$

For the present analysis, the angle of inclination α was chosen to be $\pi/4$. Indicated on Fig. 5 are the

TABLE 4. Percentage relative residual error statistics for interpolation of a potential flow problem.

Weighting function* $W(r)$	Velocity component	Residual error E_r (%) at measuring stations				Residual error E_r (%) at all grid points			
		Mean	Minimum	Maximum	Std. Dev.	Mean	Minimum	Maximum	Std. Dev.
$e^{-0.1r^2}$ $R=8$ squares	u	16.36	-57.23	231.43	49.86	3.49	-80.71	281.16	19.87
	v	12.43	-81.38	185.68	54.04	18.20	-787.53	544.39	81.04
r^{-2} $R=8$ squares	u	0.21	-1.08	2.46	0.63	7.53	-81.26	165.14	14.25
	v	-0.10	-1.21	2.86	0.71	-66.27	-989.70	966.45	151.65
Polynomial	u	9.72	-40.85	99.13	24.98	1.62	-76.93	170.49	14.17
	v	-4.85	-55.76	48.25	24.06	-99.93	-192.21	137.29	19.61
$e^{-0.1r^2}$ R variable (includes 2 data points)	u	10.13	-27.64	202.69	39.12	0.70	-81.82	202.69	15.16
	v	2.88	-158.82	151.86	62.89	9.43	-819.73	555.60	87.15
r^{-2} R variable (includes 2 data points)	u	0.10	-0.31	2.24	0.43	0.23	-81.82	198.00	12.76
	v	0.01	-0.59	0.76	0.31	12.47	-702.78	577.71	74.22

* r is the distance from grid point to measuring station; for $r > R$, $W(r)$ is set to zero.

Note: In calculating the error statistics, points on the plate and one grid cell away from the plate were ignored.

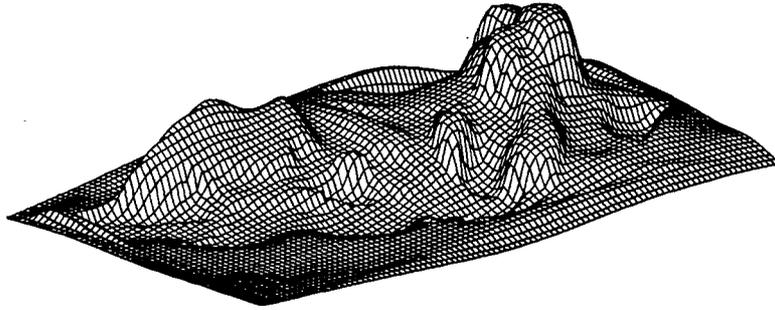


FIG. 4. Perspective plot of ozone distribution generated by polynomial interpolation scheme (a view from the southwest).

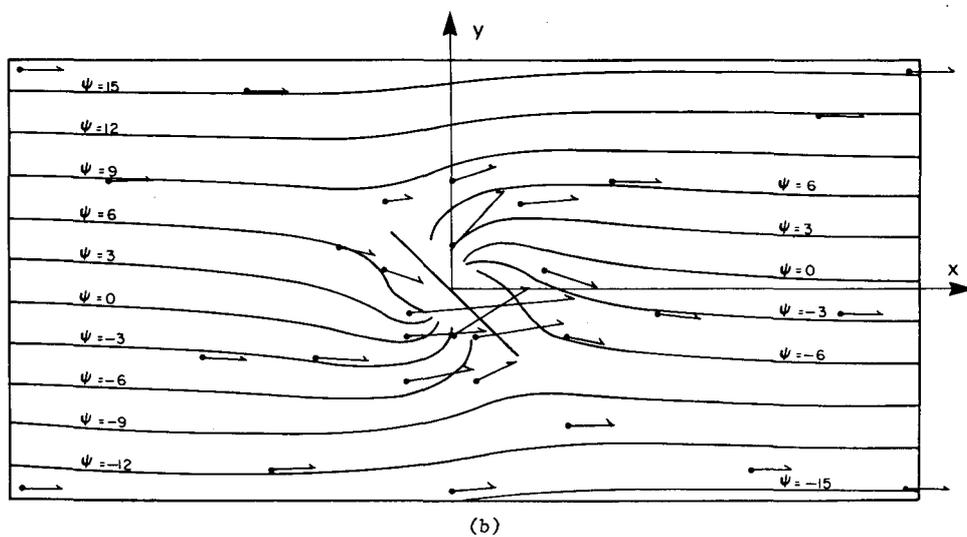
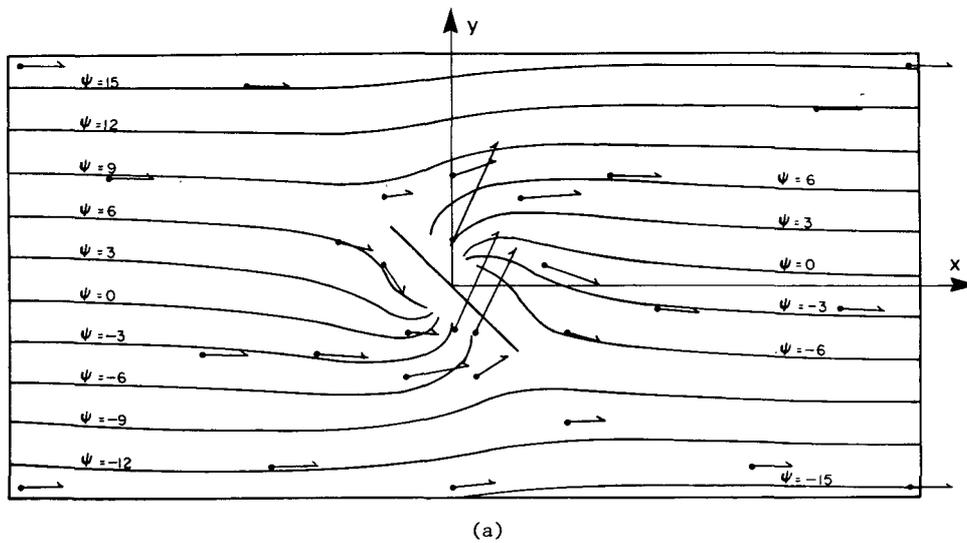


FIG. 5. Streamfunctions and location of velocity data points (solid circles) for potential flow over a flat plate. Reconstructed velocity field for potential flow over a flat plate for (a) r^{-2} with variable R and (b) polynomial fitting scheme.

streamfunction and velocity data points together with the calculated velocity vectors developed using the interpolation procedures over a 40×20 grid.

In the distance-weighting procedures, a check is made to see whether a line connecting a measuring station to a grid point intersects the barrier. If so, that station's value is not included in the weighting procedure at that grid point. In the polynomial fitting procedure a barrier is treated by requiring that it be an edge of two adjacent triangles. The results of this vector test problem are shown in Table 4 and Fig. 5.

7. Conclusions

Since the interpolation of a set of sparse data does not have a unique solution, it is important that the calculation be done carefully, producing physically realistic resulting surfaces. The interpolated field is critical to calculations such as wind field divergence reduction, contouring of data values, and initialization of pollutant transport calculations. A comparison of a number of procedures for interpolating sparse data indicates that the second-degree polynomial fitting procedure with an r^{-2} distance-weighting scheme provides a good compromise between computational costs and the accuracy of the final surface adjacent to and further away from measurement stations.

Acknowledgments. Portions of this work were supported by the California Air Resources Board under Contract A5-046-87, and by the Department of Energy under Institutional Grant EY-76-G-03-1305.

REFERENCES

- Akima, H., 1975: A method of bivariate interpolation and smooth surface fitting for values given at irregularly distributed points. Dept. of Commerce Rep. 75-70, 51 pp. [NTIS COM-75-10812].
- Barnes, S. L., 1964: A technique for maximizing details in numerical weather map analysis. *J. Appl. Meteor.*, **3**, 396-409.
- , 1973: Mesoscale objective analysis using weighted time-series analysis. NOAA Tech. Memo. ERL NSSL-62, 60 pp. [NTIS COM-73-10781].
- Boone, D. R., and G. S. Samuelson, 1977: Computer mapping of air quality. *J. Environ. Eng. Div. Proc. ASCE*, **103**, EE6, 969-979.
- Clough, R. W., and J. L. Tocher, 1965: Finite element stiffness matrices for the analysis of plates in bending. *Proc. Conf. Matrix Methods in Structural Mechanics*, Air Force Inst. and Tech., Wright Patterson AFB, Ohio.
- Cressman, G. P., 1959: An operational objective analysis system. *Mon. Wea. Rev.*, **87**, 367-374.
- Dartt, D. G., 1972: Automated streamline analysis utilizing optimum interpolation. *J. Appl. Meteor.*, **11**, 901-908.
- Endlich, R. M., and R. L. Mancuso, 1968: Objective analysis of environmental conditions associated with severe thunderstorms and tornadoes. *Mon. Wea. Rev.*, **96**, 342-350.
- Fritsch, J. M., 1971: Objective analysis of a two-dimensional data field by the cubic spline technique. *Mon. Wea. Rev.*, **99**, 379-386.
- Gandin, L. S., 1965: *Objective Analysis of Meteorological Fields*. Israel Program for Scientific Translations, Jerusalem, 242 pp.
- Gilchrist, B., and G. P. Cressman, 1954: An experiment in objective analysis. *Tellus*, **6**, 309-318.
- Green, P. J., and R. Sibson, 1978: Computing Dirichlet tessellations in the plane. *Comput. J.*, **21**, 168-173.
- Hovland, D., D. Dartt and K. Gage, 1977: An objective analysis technique for the regional air pollution study, Part I. EPA, Research Triangle Park, N.C., 44 pp. [NTIS PB 266255].
- Julian, P. R., and H. J. Thiebaut, 1975: On some properties of correlation functions used in optimum interpolation schemes. *Mon. Wea. Rev.*, **103**, 605-616.
- Lawson, C. L., 1977: Software for C' surface interpolation. *Mathematical Software*, Vol. 3, J. R. Rice, Ed., Academic Press, 161-194.
- MacCracken, M. C., and G. D. Sauter, Eds., 1975: *Development of an Air Pollution Model for the San Francisco Bay Area*, Vol. 2, Appendices. Lawrence Livermore Laboratory, UCRL-51920, 229-230 [NTIS UCRL-51920].
- McLain, D. H., 1974: Drawing contours from arbitrary data points. *Comput. J.*, **17**, 318-324.
- , 1976: Two dimensional interpolation from random data. *Comput. J.*, **19**, 178-181.
- , 1976: Errata. *Comput. J.*, **19**, 384.
- Panofsky, H. A., 1949: Objective weather-map analysis. *J. Meteor.*, **6**, 386-392.
- Powell, M. J. D., and M. A. Sabin, 1977: Piecewise quadratic approximation on triangles. *Mathematical Software*, Vol. 3, J. R. Rice, Ed., Academic Press, 316-325.
- Schlatter, T. W., 1975: Some experiments with a multivariate statistical objective analysis scheme. *Mon. Wea. Rev.*, **103**, 246-257.
- Shenfeld, L., and A. E. Boyer, 1974: The utilization of an urban air pollution model in air management. Paper presented at Fifth Meeting NATO/CCMS Expert Panel on Air Pollution Modeling, Roskilde, Denmark.
- Shepard, D., 1968: A two-dimensional interpolation function for irregularly spaced data. *Proc. 23rd ACM Nat. Conf.*, Las Vegas, 517-524.
- Stephens, J. J., and J. M. Stitt, 1970: Optimum influence radii for interpolation with the method of successive corrections. *Mon. Wea. Rev.*, **98**, 680-687.
- Thiebaut, H. J., 1975: Experiments with correlation representatives for objective analysis. *Mon. Wea. Rev.*, **103**, 617-627.
- Thiessen, A. H., 1911: Precipitation averages for large areas. *Mon. Wea. Rev.*, **39**, 1082-1084.