Determination of Error for the Purdue Regional Objective Analysis of the Mesoscale (PROAM)

David R. Smith and Michelle E. Pumphry
Department of Geosciences, Purdue University
West Lafayette, Indiana 47907

Introduction

Objective analysis of data is the utilization of a mathematical algorithm to interpolate values of measured variables onto a uniform network of grid points. Such interpolation schemes have proven to be very useful tools for processing meteorological data, since meteorological observation sites are not uniformly distributed in space. The input for the objective analysis scheme is the routinely measured meteorological variables (temperature, dew point temperature, pressure, winds, etc.) from stations located nonuniformly throughout a region. The mathematical algorithm then generates as output these measured variables at regularly spaced grid points. The analysis process is greatly facilitated since computerized graphics routines can assist in the display of such data fields. Furthermore, derived quantities such as divergence and vorticity can also be computed. There are numerous examples in the literature where researchers make use of such objective analysis schemes to diagnose siginficant meteorological events.

One of the more commonly used objective analysis algorithms in meteorology is one developed by Barnes (1,2). This scheme applies a Gaussian weighting function

$$W_{i,j,k} = \exp\left(- \frac{R^2_{i,j,k}}{\alpha}\right) \quad (1)$$

to a set of observations (Q_k) to compute an interpolated value ($Q^{\star}_{i,j}$) at each grid point [$R_{i,j,k}$ is the distance between the grid point (i,j) and the station (k) involved in each interpolation.] Barnes has shown the response for such a weighting function in dependent upon a filter parameter (α) and the wavelength (λ) that is being resolved by the data.

A variation of the Barnes scheme, referred to as PROAM (Purdue Regional Objective Analysis of the Mesoscale) has been developed for analysis of surface data over a regional area. This version of PROAM is a modified two-pass Barnes (2) scheme, in which an initial analysis of the data is corrected by making a second pass with the algorithm in order to reduce deviations between observed and interpolated values at each station. PROAM has proven to be an effective tool for the diagnosis and nowcasting of severe convective events [Snow *et al.*, (5); Smith and McCauley, (4)]. While such schemes as PROAM can be very useful, there are some problems which must be addressed in order to adequately determine its capability as a reliable analysis tool.

One of the matters that must be considered is a determination of the error limits associated with the algorithm employed by PROAM. In particular, how effectively does the algorithm, which interpolates the data to the grid points, reproduce the meteorological phenomena resolved by the data? Since surface meteorological features are rather complex due to the interaction of several scales of atmospheric disturbances, it is somewhat difficult to quantitatively establish error limits. Smith and Leslie (3) have performed error analyses for PROAM using analytic distributions typical of atmospheric temperatures. They computed average grid point errors (the mean deviation between the known analytic solution and the interpolated values from the gridded field) for a variety of analysis parameters. They found that this Barnes-type scheme could produce analyses of surface data with average grid point errors on the order of 0.5 °C, which is consistent with errors of measurement for temperature. One of their findings, however, demonstrated that part of the error in the analysis could be due to the location of the stations which influenced the interpolation for the grid point value. For example, if all stations that significantly contribute to a grid point value are located south of this point, one would expect a bias toward warmer temperatures at that location. An examination of any surface weather map reveals that weather stations are scattered nonuniformly throughout space, with clusters of stations near large metropolitan areas and large data void regions in rural areas. Such a station distribution tends to complicate the analysis because it is difficult to discriminate the error due to the algorithm from that associated with irregular station spacing. This study addresses the problem of error determination due to station location.

Objective and Methodology

The objective of this investigation is to determine error limits associated with PROAM without the additional complication of irregular station spacing. Following the technique of Smith and Leslie (3), analytic functions (Figure 1) were selected to

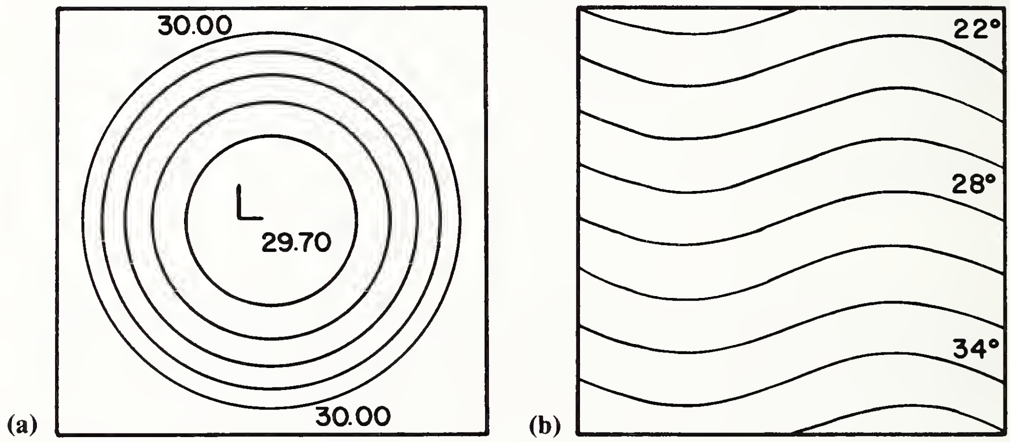

FIGURE 1.

a) Analytic function for pressure, with a contour interval of 0.05 *in. Hg.*

b) Analytic function for temperature, with a contour interval of 2 °C.

generate the data set used by the algorithm. However, rather than using actual surface station locations, the values were assigned to a uniformly distributed set of locations, (referred herein as "data sites"), then PROAM interpolated this data set onto a uniform network of grid points that is staggered from the station network. Root-mean square (RMS) and average grid point (AGP) errors were then computed for a variety of experiments. These errors can give guidance to the user, permitting a judicious choice of the analysis parameters which will minimize the errors associated with the scheme.

Several experiments were conducted to generate error (RMS and AGP) curves for the important analysis parameters used by PROAM. These parameters include a filter parameter (α), radius of influence (R_{inf}) and wavelengths (λ_T, λ_p) of the disturbances which are being resolved by the scheme. The filter parameter determines the degree of smoothing performed by the weighting function $W_{i,j,k}$. As α increases one expects greater smoothing of extreme data points in the analysis. The radius of influence (R_{inf}) determines the area surrounding each grid point from which data sites are selected to perform the interpolation. As R_{inf} increases, a greater number of data sites will contribute to the grid point value. The wavelengths (λ_T, λ_p) determine the scales of the disturbances to be resolved. The analytic temperature function (Figure 1) is a simple sinusoidal wave, while the pressure function is a circular pattern. By decreasing the wavelength, the frequency of occurrence of the waves is increased throughout the domain. The results presented show the effect of these parameters on the analysis, particularly the error distributions associated with variations of these parameters.

Results and Discussion

A series of experiments was conducted using PROAM in which one of the significant analysis parameters was allowed to vary through a range of values while holding the other parameters constant. The control conditions used in these experiments were $\alpha = 0.6\ \mathrm{cm}^2$, $R_{inf} = 1.0$ cm, $\lambda_T = 10$ cm, $\lambda_p = 20$ cm. (The units represent length as depicted on the map plane of the analysis. Since the maps are scaled by 10^{-7}, 10 cm corresponds to 1000 km, which is the actual distance across the domain of the analysis region.)

Experiment 1: Variation of α (0.1 - 4.8 cm^2)

The filter parameter (α) controls the degree of smoothing produced in the analysis of the data, whereby greater detail is generated for smaller values of α. One might expect the greater detail in the analysis would imply higher accuracy. The RMS error, which is actually a measure of the fit of the interpolated fields to the original data, is shown in Figure 2a. Note that RMS error increases monotonically over the range

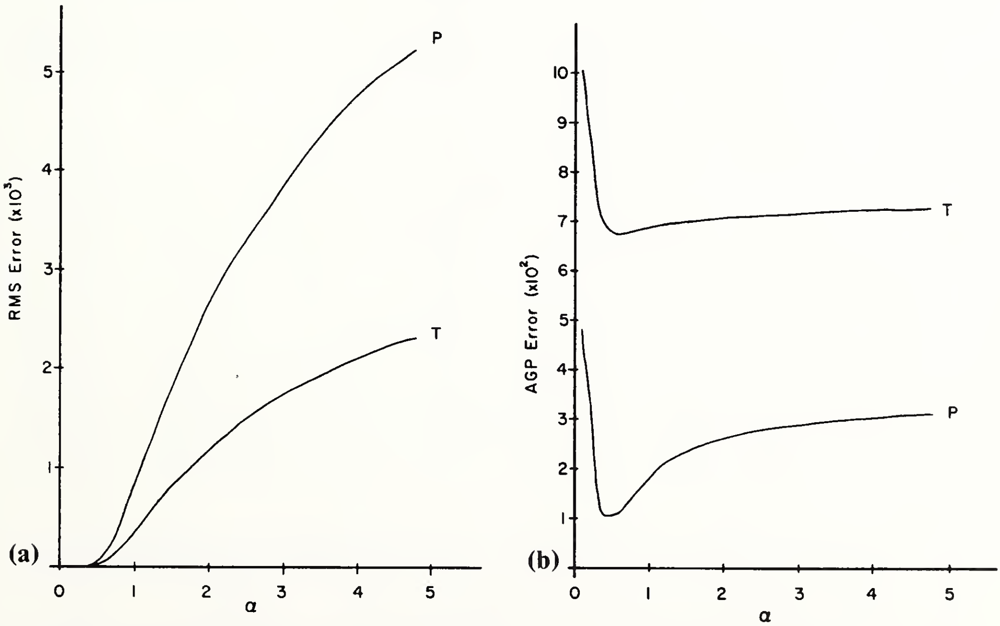

FIGURE 2.

a) Root mean square error of pressure (p) and temperature (T) functions with respect to the filter parameter (α).

b) Average grid point error of pressure (p) and temperature (T) functions with respect to α.

[*Error has been normalized by $A_x^{1/2} A_y^{1/2}$, where A_x is the amplitude of the wave in the X-direction (East-West) and A_y is the amplitude of the wave in the Y-direction (North-South).*] α is expressed in units of cm^2.

of α values. This indicates that additional smoothing, which reduces the effect of extreme values, produces an analysis that is less representative of the actual data. In other words extreme high and low observations have been modulated by values closer to the mean value of the field. The average grid point error (Figure 2b) however, shows that perhaps the analysis using smaller α values is questionable. Size less smoothing occurs, the contribution of shorter wave features is enhanced. Actually, these higher frequency waves are noise rather than significant features in the analysis, thereby producing larger errors in the analysis. A minimum in AGP error is apparent for $\alpha \sim 0.6$, with a slight increase in error for larger α values (effect of smoothing). This value of α corresponding to the minimum AGP error then suggests an optimum choice for α for the analysis. Determining an optimum α value thus provides the analyst with an effective selection criterion for the filter parameter.

Experiment 2: Variation of R_{inf} (0.5 - 4.0 cm)

The radius of influence controls the number of data sites influencing the interpolation at each grid point. For small R_{inf} values there is a higher probability of a single data site contributing to the interpolation, thereby introducing a bias in the analysis. This possible source of error is reduced for larger areas of influence, since multiple data sites contribute to the analysis. Figure 3a shows the RMS errors as a

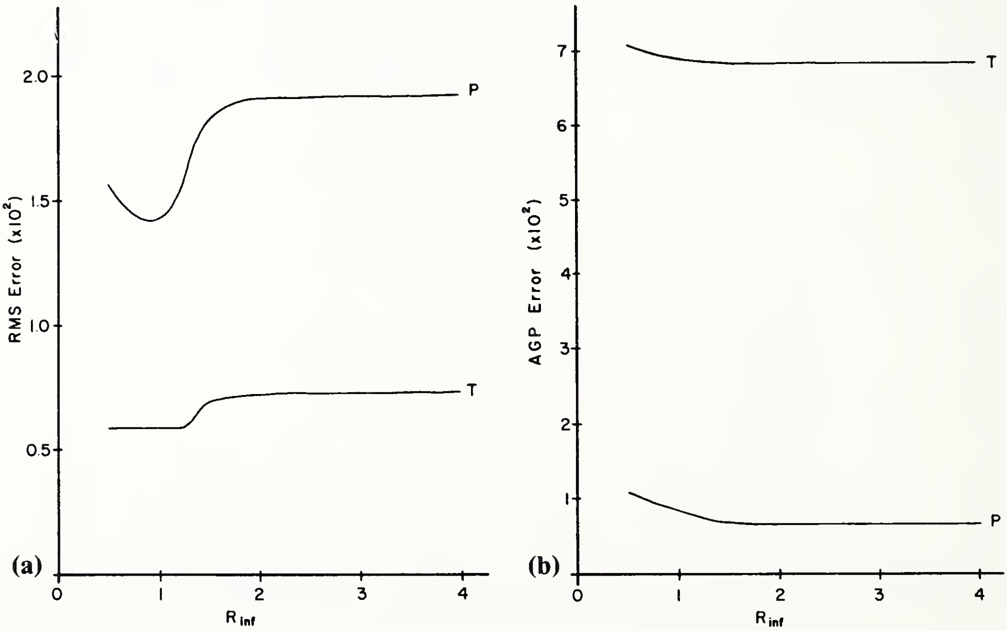

FIGURE 3.

a) Same as Figure 2a, except with respect to the radius of influence (R_{inf}).

b) Same as Figure 2b, except with respect to R_{inf}.

R_{inf} expressed in units of cm (distance as measured on the map plane).

function of R_{inf}. As expected, the P-curve shows an initial decrease in error as R_{inf} increases to 1. RMS error then increases for $1.0 \lesssim R_{inf} \lesssim 1.5$. This can be explained since the algorithm is selecting data sites further from the analyzed grid point to contribute to the interpolation. Since the analytic functions determine the observations for each data site, as we progress farther from the grid point, there is a greater deviation for these contributing data sites values from the value at the grid point, so error increases. Beyond a certain radius ($R_{inf} \gtrsim 2$) the contribution of these data sites is minimal (since the weighting function decreases exponentially with R^2), hence the RMS error levels off to a constant value. The average grid point error (Figure 3b) does show a gradual decrease in error as R_{inf} approaches 1.5. These error curves also show the same asymptotic behavior as RMS error at larger R_{inf}. Since larger R_{inf} increases the computational run time for PROAM, there is no real benefit, in terms of error reduction, to increasing R_{inf} beyond 1.5. Therefore, an optimum range of values for R_{inf} is 1.0 to 1.5 for this data set and data site distribution.

Experiment 3: Variation of λ (2-10 cm)

Since the response for the weighting function increases with λ, one would expect a better analysis for longer wavelength phenomena. Both RMS and AGP error analyses verify this (Figures 4a and b). In both sets of curves error decreases dramatically (for $\lambda < 5$) with an increase in λ, while at longer wavelengths there is little reduction in error. Since the response at $\lambda = 2$ cm (for $\alpha = 0.6$) is 80%, and it is nearly 99% for $\lambda > 5$ cm, one would expect to be able to resolve the smaller wavelengths with less accuracy than longer wavelengths. Furthermore, there is less significant improvement for $\lambda > 5$ cm.

Conclusions

This investigation has determined error limits for the PROAM scheme for analytic functions representative of surface temperature and pressure. This study used a uniform network of data sites in attempt to remove the effect of clustering and data void areas from the analysis. The major findings were:

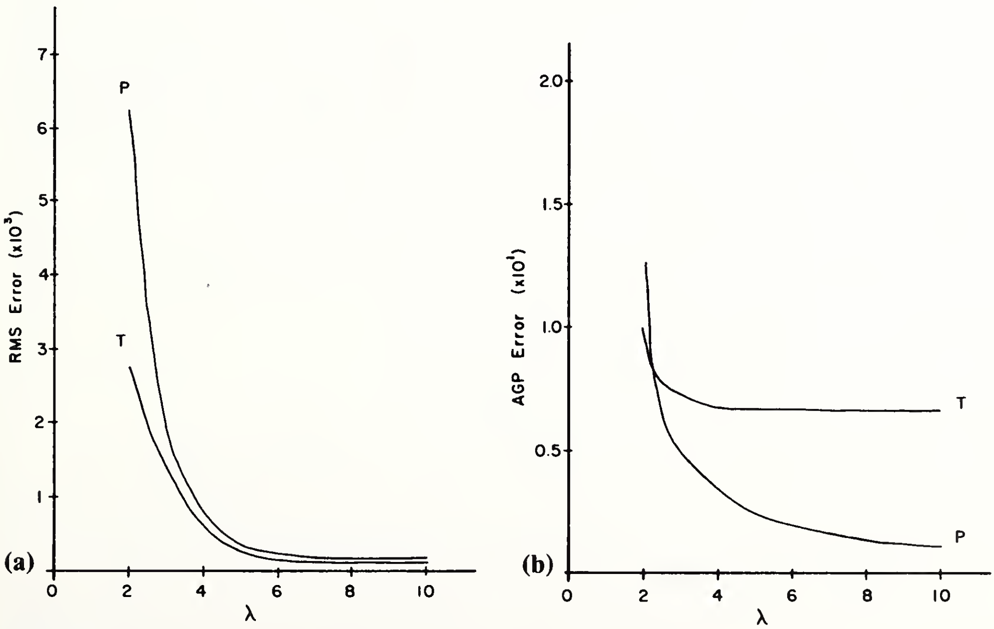

FIGURE 4.

a) Same as Figure 2a, except with respect to wavelength of the analytic functions (λ_p, λ_T).

b) Same as Figure 2b, except with respect to λ_p, λ_T.

λ *expressed in units of cm (distance as measured on the map plane).*

1) An increase in α produces larger RMS errors;
2) An optimum α value exists to minimize both RMS and AGP errors;
3) An optimum R_{inf} range exists to minimize errors;
4) Error reduction for R_{inf} does not justify additional computional run time; and
5) Error decreases as the wavelength of the disturbance increases.

Future work will involve a comparison of these results for a uniform data site distribution with that of a non-uniform station distribution typical of actual conditions.

Acknowledgments

The authors extend their appreciation to Dr. John T. Snow for his encouragement and review of this manuscript, to Ms. Claire Wooley and Barbara Chance for drafting the figures and to Ms. Helen Henry for typing the manuscript.

Literature Cited

1. Barnes, S. L., 1964. A technique for maximizing detail in numerical weather map analysis. *JAM, 3(4)*, pp. 396-409.
2. Barnes, S. L., 1973. *Mesoscale objective map analysis using weighted time-series observations*. NOAA Tech. Memo. ERL-NSSL-62, 60 pp.
3. Smith, D. R. and F. W. Leslie, 1982. *Evaluation of a Barnes-type objective analysis scheme for surface meteorological data*. NASA Tech. Memo. - 82509, 25 pp.
4. Smith, D. R. and S. D. McCauley, 1983. Mesoanalysis of the surface features associated with the Shelby County, KY tornado (20/21 March 1982), *Preprints of 13th Conf. on Severe Local Storms*, AMS Boston, MA, p. 308-311.
5. Snow, J. T., D. R. Smith, F. W. Leslie and R. H. Brady, 1983. Meso analysis of surface variables associated with the severe weather of 9-10 July 1980. *Nat. Wea. Dig., 8(1)*, p. 28-39.