A Method for Improved Analyses of Scalars and Their Derivatives

PHILLIP L. SPENCER
Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, Norman, Oklahoma

DAVID J. STENSRUD
NOAA/National Severe Storms Laboratory, Norman, Oklahoma

J. MICHAEL FRITSCH
Department of Meteorology, The Pennsylvania State University, University Park, Pennsylvania

(Manuscript received 16 December 2002, in final form 1 May 2003)

ABSTRACT

Analytic observations are used to compare the traditional and triangle methods for the objective analysis of scalar variables. The traditional method for objective analysis assigns gridpoint values based on the distance from the grid points to each member of the set of observations. Subsequently, spatial derivatives are derived by applying a finite-differencing scheme to the field of gridded observations. The triangle method for objective analysis calculates the spatial first-order derivatives directly from each set of nonoverlapping triangles that are formed by the observations, and the derivatives are assigned to the triangle centroids. By calculating the first-order derivatives directly from the observations, the triangle method bypasses the need for finite differencing. Triangle centroid estimates of the scalar field itself are simply arithmetic averages of the three observations comprising each triangle. The centroid estimates of the scalar variable and its spatial derivatives are then treated as observations and mapped to a uniform grid via the traditional method.

Results indicate that the traditional method for the analysis of a scalar variable is superior to the triangle method for scalar analysis because the simple averaging involved in creating the triangle centroid estimates of the scalar exposes the triangle analysis to the potential for significant damping of the input field. Indeed, although the patterns of the scalar analyses from the two methods are comparable, the analysis from the triangle method does not reproduce the amplitude of the scalar field as well as the analysis from the traditional method. Gradient and Laplacian fields computed from the triangle method, however, are generally superior to those derived by the traditional method, which tends to force all of the gradient information into the gaps between observing stations.

To overcome the deficiency of the triangle method’s ability to produce an acceptable scalar analysis and the deficiency of the traditional method’s ability to produce an acceptable derivative analysis, a variational objective analysis scheme is developed that combines the best aspect of the triangle method with that of the traditional method. Analyses of the scalar and its spatial derivatives from the variational analysis scheme are generally superior to analyses from both the traditional and triangle methods.

1. Introduction

The purpose of objective analysis is to provide gridded estimates of variables from observations that are irregularly distributed in space and/or time. There are several reasons why gridded fields may be desired. One reason is for the simple purpose of displaying and contouring the data; most contouring packages require gridded input. In addition, some of the simpler analysis schemes used today (e.g., Cressman 1959; Barnes 1964, 1973) have predictable and controllable response characteristics that allow unwanted scales to be suppressed or virtually eliminated from the analysis. A second reason is that gridded observations are required by widely used and popular techniques used today for diagnostic calculations that involve derivative estimates. The equations for moisture convergence, frontogenesis, vertical vorticity tendency, and kinematically derived vertical motion, for example, are replete with terms involving spatial derivatives. These derivatives typically are estimated by applying a finite-differencing scheme to the gridded observations. A third reason for objective analysis is to provide the initial conditions to a numerical
forecast model. Although more sophisticated techniques are popular for today’s suite of operational and research models (e.g., optimal interpolation and variational methods), simple objective analysis schemes still are used on occasion in the research environment because of their simplicity and efficiency.

Lorenc (1986) rightly states that no method for obtaining gridded estimates of a meteorological variable is ideal in all respects. For some purposes, the sophistication of optimal interpolation or variational schemes may be desired; in others, the simplicity of traditional successive correction schemes may be preferred. For example, the use of sophisticated schemes that involve balance constraints may be more appropriate than simpler schemes when the analysis acts as the initial conditions for a numerical forecast model. Unless a suitable balance exists between the initial mass and motion fields, the generation of high-frequency gravity-inertia waves within the model may render the forecast useless. On the other hand, simple successive correction schemes may be quite satisfactory—or even preferable—for diagnostic studies (Lorenc 1986). The following quote from Parsons and Dudhia (1997) says it well:

There are still classes of problems in the atmospheric sciences where it is scientifically and/or philosophically preferable to derive an estimate of the state of the atmosphere that is independent of any numerical model through undertaking objective analysis of measurements alone. One example of this class of problems is special field projects designed to address poorly understood and poorly forecasted processes.

Therefore, when the observations are most important, with little weight given to prior relationships, then methods such as traditional successive corrections are appropriate (Lorenc 1986). Caracena (1987) supports this view by noting that the diagnostician’s main concern may be to evaluate the state of an observed portion of the atmosphere with the greatest fidelity possible. He notes that the analyst may not want to use constraints in the analysis that would be counter to this purpose, such as using a geostrophic constraint when studying ageostrophic effects. The easy availability of model analyses that are created by sophisticated operational analysis systems admittedly makes them very tempting for use in diagnostic studies. However, another compelling reason for applying a simple successive correction scheme in some instances is for the analysis of observations that are not incorporated into the operational analysis systems. For example, the high spatial and temporal resolution Weather Surveillance Radar–1988 Doppler (WSR-88D) level-II data (Crum et al. 1993) from the Next Generation Doppler radar (NEXRAD) network currently are not incorporated into operational numerical weather prediction (NWP) analysis systems. Therefore, in order to obtain gridded analyses of these data, analysts often resort to traditional successive correction techniques involving distance-dependent weighting (e.g., Trapp and Doswell 2000; Askelson et al. 2000). Also, successive correction schemes may be the best available option for data originating from special field programs, since these data also may be excluded from operational analysis systems.

Having thus established the benefits of simple successive correction objective analysis schemes for meteorological studies, we note that two schemes have remained popular for decades among diagnosticians: the Cressman (1959) and Barnes (1964, 1973) analysis schemes. These schemes have been used to analyze all sorts of meteorological data, including surface (e.g., Koch and Saleeby 2001), rawinsonde (e.g., Ogura and Chen 1977), satellite (e.g., Koch et al. 1983), profiler (e.g., Spencer et al. 1999), radar (e.g., Askelson et al. 2000), rainfall (e.g., Bussières and Hogg 1989), and aircraft (e.g., Fankhauser et al. 1985) data. Both the Cressman and Barnes schemes assign gridpoint values based on the distance from the grid points to each member of the set of observations; nearby observations affect a gridpoint value more than distant observations. These schemes typically are iterated until the analysis converges—to some extent—to the observations. Barnes (1994a), however, showed that the goodness of fit to the observations is a poor measure of the quality of an analysis. In fact, an analysis that fits the observations very closely may be judged to be very poor if derivative information is important. All of the gradient information is forced into the data voids between stations for an analysis that closely fits the observations (Barnes 1994a). Therefore, the quality of spatial derivative estimates from such an analysis obtained by finite differencing is questionable, at best. Since derivative terms permeate most of the equations used by diagnosticians, derivative estimation from a set of observations is an important issue.

Bellamy (1949) skirted the issue of finite differencing by developing a procedure for calculating spatial derivatives of the wind field (namely, divergence and vorticity) directly from the observations. His method computes divergence, for example, by calculating the fractional rate of change of a triangular area due to the wind field observed by three noncolinear stations. Ceselski and Sapp (1975) and Schaefer and Doswell (1979) used line integrals around triangles of observations to estimate the divergence and vorticity at triangle centroids. Zamora et al. (1987) and Doswell and Caracena (1988) applied the linear vector point function (LVPF) method for the same purpose. Each of these techniques for diagnosing the kinematic quantities assumes a linear variation of the wind field along each leg of the triangle. Because of this common assumption, the techniques are essentially equivalent (Davies-Jones 1993).

Schaefer and Doswell (1979) showed that the method of first estimating the derivatives directly from the observations and then applying an analysis scheme to create a gridded derivative field is superior to the traditional method of applying a finite-differencing scheme to a set
of gridded observations. Doswell and Caracena (1988) explored the theoretical foundation of the superiority of the former method and ascribe its superiority to the failure of the traditional method to use all the information in the data. Spencer and Doswell (2001) confirmed the superiority of triangle techniques over traditional methods and went so far as to say that if the goal of doing an objective analysis is to diagnose derivatives of the wind field (divergence, vorticity, and deformations) with as much accuracy as possible, then there is little alternative to doing the derivative estimates via some form of the triangle method (Bellamy, line integral, or LVPF).

These methods for calculating derivatives directly from the observations apply only to a vector field—namely, the wind field. The techniques, as developed, provide no means for calculating the spatial derivatives of scalar quantities such as temperature, moisture, or pressure. Since the results of Schaef er and Doswell (1979), Doswell and Caracena (1988), and Spencer and Doswell (2001) all suggest that derivatives of the wind field are best calculated directly from the observations and then mapped to a uniform grid as opposed to applying a finite-differencing scheme to a set of gridded wind components, then it seems reasonable to suggest that spatial derivatives of scalar quantities also should be calculated directly from the observations. Unfortunately, this methodology appears less popular than the triangle method for calculating kinematic quantities of the wind field. Popular software today, such as GEM-PAK (general meteorological data assimilation, analysis, and display software package; Koch et al. 1983), continue to use the traditional method for derivative estimation from both scalar and vector fields, despite mounting evidence of a superior alternative. Endlich and Clark (1963) developed such a method for calculating spatial derivatives of any scalar quantity directly from observations using a triangle technique that assumes a linear variation of the scalar between reporting stations. However, they make no mention of using an objective analysis routine to grid the triangle centroid derivative estimates; in fact, isolines were drawn subjectively. Endlich and Clark (1963) also gave little attention to the analysis of the scalar variable itself. They did mention, however, that the triangle technique provides a centroid estimate of the scalar that is simply the arithmetic average of the three observations comprising the triangle. This type of averaging exposes the triangle analysis to a possible reduction in the amplitude of the scalar field. Barr et al. (1971) modified the method of Endlich and Clark (1963) in two important ways. First, they installed an objective routine—based on the Cressman (1959) scheme—to provide gridded fields from the triangle centroid estimates. Second, Barr et al. (1971) included both the centroid estimates of the scalar variable and the actual observations themselves when creating a gridded scalar field. They found that by weighting the actual observations by an amount 10 times greater than the centroid estimates of the scalar, the scalar analysis did not suffer the severe reduction in amplitude that would characterize a scalar analysis based solely on the centroid estimates of the scalar. Although Barr et al. (1971) claim that the chosen weighting factor of ten is “reasonable,” they also admit that it is “somewhat arbitrary.”

In this paper, a somewhat different method for the objective analysis of a scalar variable is developed. The technique described herein uses the triangle approach of Endlich and Clark (1963) for calculating spatial derivatives of an analytic scalar variable directly from the observations. These derivative fields are subsequently mapped to a uniform grid by a traditional Barnes objective analysis scheme. The gridded derivative fields then are combined with a Barnes analysis of the analytic scalar variable itself through a variational procedure to produce gridded scalar and derivative fields that are mutually consistent.

2. Creating analytic data and artificial observing networks

a. Analytic observations

Observations that are created by sampling an analytic scalar field are useful for objective analysis studies because the true field and its derivatives are known at all points in the domain. The analytic observations used in this study are determined from samples of the two-dimensional scalar variable \( S_a \):

\[
S_a(x, y) = 10 \cos \left( \frac{2\pi x}{L} \right) \cos \left( \frac{2\pi y}{L} \right),
\]

where the wavelength \( L \) is an adjustable parameter. The gradient of \( S_a \) is given by

\[
\nabla S_a = -\left( \frac{20\pi}{L} \right) \left[ \sin \left( \frac{2\pi}{L} x \right) \cos \left( \frac{2\pi}{L} y \right) \hat{i} + \cos \left( \frac{2\pi}{L} x \right) \sin \left( \frac{2\pi}{L} y \right) \hat{j} \right],
\]

where \( \hat{i} \) and \( \hat{j} \) are unit vectors in the positive \( x \) and \( y \) directions, respectively. The Laplacian of \( S_a \) is given by

\[
\nabla^2 S_a = -\left( \frac{8\pi^2}{L^2} \right) S_a.
\]

Each of these fields represents a checkerboard pattern. Examples for \( L = 6\Delta \) and \( L = 20\Delta \) are shown in Fig. 1.\(^1\) where \( \Delta \) represents the average data spacing (section 2b).

\(^1\) Note that for all \( L = 6\Delta \) contour plots shown in this paper, only the interior one-half of the domain is presented. Also note that for all contour plots of the gradient, it is the magnitude of the gradient that is actually contoured.
b. Artificial observing networks

The method described by Doswell and Lasher-Trapp (1997) for creating an artificial array of observing sites is used in this study. This procedure allows us to compare analyses for data distributions ranging from perfectly regular to those associated with high degrees of clustering and relatively large data gaps, such as the current U.S. surface network.

The method for creating an artificial observing network begins with a $21 \times 21$ triangular array of observing sites (Fig. 2a). Each site within the interior of the domain has six nearest neighbors, each of which is

Fig. 2. Array of artificial observing sites (a) before displacement occurs, (b) after displacement using SC = 0.5; (c) after the superob algorithm has been applied to the stations in (b); and (d) showing the Delauney triangulation applied to the stations shown in (c). (d) Triangles with a minimum angle less than 15° are not drawn and are removed from further consideration; lightly shaded areas indicate interior regions where triangles have been removed.

Fig. 1. Contour plots of the analytic field of the scalar for (a) $L = 6\Delta$, (b) $L = 20\Delta$; magnitude of the scalar gradient for (c) $L = 6\Delta$, (d) $L = 20\Delta$; scalar Laplacian for (e) $L = 6\Delta$, and (f) $L = 20\Delta$. Dashed contour lines indicate negative values. Note that for the $L = 6\Delta$ plots, only the interior one-half of the analysis domain is presented.
spaced 20 km from the central site. This average data spacing is hereafter referred to as \( \Delta \) (i.e., \( \Delta = 20 \) km). This regularly distributed array of observing sites represents a perfectly uniform observing network.

To create an observing network that is irregularly distributed, each observing site in Fig. 2a is randomly displaced in each direction by a distance that is no greater than some fraction of \( \Delta \). This fraction is referred to as the scatter constant (SC) and has values \( 0.0 \leq SC \leq 1.0 \). Using SC = 0.0 results in an observing network that is regularly distributed (Fig. 2a), whereas using SC = 1.0 generates an observing network that is highly irregularly distributed. If an observing site is displaced far enough such that it falls outside any boundary of the domain, then it is merely reflected inside the opposite boundary that same distance in order to keep it inside the domain. A data distribution generated using SC = 0.5 is shown in Fig. 2b.

For moderate to large values of the scatter constant, one or more stations may become clustered with other stations. Since clustering may have deleterious effects on the analyses, the simple algorithm developed by Spencer (2002) to generate a superobservation \(^1\) (superob) from clustered stations is applied. This algorithm is described in the appendix. For the present study, two observing stations are considered to be clustered if the distance between them is less than the subjectively chosen value of 0.25\( \Delta \). Figure 2c shows the locations of the observations that result after the superob algorithm is applied to the observations shown in Fig. 2b.

c. Delauney triangulation

Once the superob algorithm has been applied and we have arrived at the final set of observations that are to be used by the objective analysis routines, a triangular tessellation of the data network is performed. As we will see, this tessellation is required by the triangle method of objective analysis (described in section 3b).

A set of nonoverlapping triangles is defined from the observational network according to the Delauney triangulation method (Ripley 1981). This procedure selects the set of triangles containing the largest minimum angle from among the two competing triangular tesselations that comprise each quadrilateral formed by four nearby observing stations. Another criterion of the method is that any point within any triangle should have as its closest data point one of the vertices of that triangle. A more thorough description of the Delauney triangulation is provided by Ripley (1981). The triangular tessellation of the stations in Fig. 2c is presented in Fig. 2d.

3. Objective analysis schemes

In this section, we describe the two objective analysis schemes that form the basis of our initial comparisons. The prototype of the traditional method, which applies a finite-differencing scheme to a gridded set of observations to calculate spatial derivatives, is based on a three-pass modification of the Barnes (1973) two-pass analysis scheme. The three-pass analysis scheme is discussed at length by Achtemeier (1987, 1989) and Barnes (1994b). The triangle method, which calculates spatial derivatives directly from the observations, is based on the work of Endlich and Clark (1963).

a. Traditional method

The traditional method for objective analysis applies the following analysis equation to the observations:

\[
S_g = S_b + \frac{\sum_{k=1}^{N} w_k (S_k - S_{bg})}{\sum_{k=1}^{N} w_k},
\]

where \( S_g \) is the gridded scalar field, \( S_b \) is the background (first-guess) estimate, \( N \) is the number of observations, \( S_k \) is the scalar observation at the \( k \)th data point, \( S_{bg} \) is the scalar estimate at the \( k \)th data point obtained by bilinear interpolation of the background field, and \( w_k \) is the weight assigned to the \( k \)th observation. A Barnes-type exponential weighting function is used in (4) and is given by

\[
w_k = e^{-R_k^2/\lambda^2},
\]

where \( R_k \) is the distance between the \( k \)th observation and the grid point in question and \( \lambda \) is the smoothing parameter that controls the response characteristics of the analysis; that is, \( \lambda \) determines the smoothness of the analysis.\(^4\) The weights for observations farther than a distance \( 5 \Delta \) from a particular gridpoint are set to zero.\(^5\)

In this study, a three-pass successive corrections version of (4) and (5) is applied to the observations. Achtemeier (1987) showed that improvements in the analysis, particularly in the short but resolvable wavelengths, are possible if a three-pass scheme is used instead of the popular two-pass scheme proposed by Barnes (1973). Since no background field is used in this study, \( S_b \) is zero during the initial pass. Also, following the recommendation of Achtemeier (1989), the first-pass analysis is very smooth, retaining significant amplitude.

---

\(^1\) Doswell and Lasher-Trapp (1997) showed that little or no increase in the irregularity of the station distribution is obtained as SC increases beyond 1.0.

\(^2\) A superobservation represents the average location and value derived from a set of clustered stations.

\(^4\) The analytic observations and the data distributions presented herein are well-suited for the isotropic weighting function given by Eq. (5). Anisotropy in the observations and/or data distributions may be accounted for by means of an anisotropic weighting function (e.g., Benjamin and Seaman 1985; Spencer et al. 1999; Askelson et al. 2000).

\(^5\) The distance at which the weights are set to zero often is referred to as the “cutoff radius.”
only at the well-resolved wavelengths. Specifically, for
the initial, smooth first-pass analysis, we choose \( \lambda = 1.75\Delta \). The response curve for this choice of \( \lambda \) is the dashed line in Fig. 3.

For the second pass of the analysis scheme (first correction pass), the background field \( S_b \) is nonzero and is simply the first-pass analysis. Similarly, for the third pass of the analysis scheme (second correction pass), the background field is simply the second-pass analysis. For each of the two correction passes, we use \( \lambda = 1.0\Delta \); that is, the smoothing parameter is held constant, following the procedure of Achtemeier (1989). The final response is the solid curve in Fig. 3. Clearly, the final response curve suggests a sharp cutoff between wavelengths unresolved by an observing network (L < 2\( \Delta \)) and those that are considered marginally well sampled (6\( \Delta \) = L = 12\( \Delta \); Doswell and Caracena 1988).

The analysis grid contains 169 \( \times \) 146 grid points in the \( x \) and \( y \) directions, respectively, where the grid spacing \( \Delta = \Delta/8 \). This fine-resolution grid reduces the effect of truncation error when the finite-differencing scheme is applied to calculate spatial derivatives. Fourth-order centered finite differencing for estimating derivatives is used wherever enough grid points are available. Elsewhere (near the boundaries), second-order centered or one-way finite-differencing schemes are used to estimate spatial derivatives.

Barnes (1994a) suggests that the analysis standard is not how closely the analysis replicates the observations, but how accurately the analyzed gridpoint values reconstruct the scalar field and its derivatives over the area of interest. For this reason, both \( \nabla S \) and \( \nabla^2 S \) are computed from the gridded scalar field, where the spatial derivatives are calculated using a fourth-order differencing scheme.

b. Triangle method

The triangle method for the analysis of a scalar variable \( S \) begins with the assumption that \( S \) varies linearly within each triangle created by the Delauney triangulation. Therefore, \( S \) may be written as

\[
S(x, y) = S_c + \frac{\partial S_c}{\partial x}(x - x_c) + \frac{\partial S_c}{\partial y}(y - y_c),
\]

where \( S_c \) is the scalar estimate at the triangle centroid, \( \partial S_c/\partial x \) and \( \partial S_c/\partial y \) are the horizontal gradient components of \( S \) at the centroid, and \( (x_c, y_c) \) is the location of the centroid. If the linear expansion (6) is written for each of the three stations comprising a triangle, then we have

\[
S_1 = S_c + \frac{\partial S_c}{\partial x}(x_1 - x_c) + \frac{\partial S_c}{\partial y}(y_1 - y_c) \quad (7a)
\]
\[
S_2 = S_c + \frac{\partial S_c}{\partial x}(x_2 - x_c) + \frac{\partial S_c}{\partial y}(y_2 - y_c) \quad (7b)
\]
\[
S_3 = S_c + \frac{\partial S_c}{\partial x}(x_3 - x_c) + \frac{\partial S_c}{\partial y}(y_3 - y_c) \quad (7c)
\]

where \( S_1, S_2, \) and \( S_3 \) represent the scalar observations and \( (x_1, y_1) \) are their locations. According to Endlich and Clark (1963), the triangle centroid estimate of the scalar variable is simply the arithmetic average of the three values comprising the triangle; that is,

\[
S_c = \frac{S_1 + S_2 + S_3}{3}. \quad (8)
\]

Given (8), any two of the Eqs. (7) may be combined to solve for the components of the scalar gradient. For example, combining (8) with (7a) and (7b) yields

\[
\frac{\partial S_c}{\partial x} = \frac{\Delta S_1}{\Delta x_1} + \left( \frac{\Delta S_1}{\Delta y_1} - \frac{\Delta S_1}{\Delta y_2} \right) \frac{\Delta y_2}{\Delta x_1} \quad \text{and} \quad (9)
\]
\[
\frac{\partial S_c}{\partial y} = \frac{\Delta S_1}{\Delta y_1} - \frac{\Delta S_1}{\Delta y_2} \frac{\Delta x_1}{\Delta x_2} \quad (10)
\]

where \( \Delta S_1 = S_1 - S_c, \Delta x_1 = x_1 - x_c, \) and \( \Delta y_1 = y_1 - y_c \). Therefore, (8), (9), and (10) provide estimates of the scalar and its horizontal gradient \( \nabla S \) at the centroid of each triangle. Note that the gradient is estimated without a prior mapping of the scalar field to a uniform grid and resorting to a finite-differencing scheme as is required by the traditional method; the triangle method estimates spatial derivatives directly from the observations. To obtain gridded estimates of the scalar and its gradient for direct comparison with the traditional method, the three-pass analysis scheme (4) and (5) is applied separately to the triangle centroid estimates of...
the scalar and its gradient, where all analysis parameters are equal to those used for the traditional method analysis.

The Laplacian of the scalar field $S$ is calculated according to

$$\nabla^2 S = \nabla \cdot (\nabla S),$$

where $\nabla S$ is the gridded scalar gradient described in the previous paragraph.\(^6\)

### 4. Comparison of analysis methods

#### a. Measures of analysis error

The root mean square error (rmse) between each scalar analysis ($S_a$) and the analytic field ($S_g$) is calculated according to

$$\text{rmse} = \sqrt{\frac{\sum_{i,j} (S_{a,i,j} - S_{g,i,j})^2}{N_g}},$$

where $N_g$ is the number of grid points. The calculation is performed over the interior $110 \times 87$ grid points to prevent boundary errors from contaminating the statistic. For the gradient—a vector quantity—the root mean square vector error (rmsve) is calculated according to

$$\text{rmsve} = \sqrt{\frac{\sum_{i,j} \left[(S_{x,i,j} - \frac{\partial S}{\partial x})^2 + (S_{y,i,j} - \frac{\partial S}{\partial y})^2\right]}{N_g}},$$

where $S_x$ and $S_y$ represent the gridded gradient components from the analyses and the analytic gradient components are given by (2). The rmse of the Laplacian analyses are given by (12), except that $S_a$ is replaced by $\nabla^2 S_a$ [(given by Eq. (3)] and $S_g$ is replaced by $(S_{xg} + S_{yg})$, where $S_{xg}$ and $S_{yg}$ represent gridded second-derivative estimates of the scalar in the $x$ and $y$ directions, respectively.

A second measure of analysis error, the Pearson correlation coefficient, is calculated according to

$$\text{correlation coefficient} = \frac{\sum_{i,j} (A_{i,j} - \bar{A})(T_{i,j} - \bar{T})}{\sqrt{\sum_{i,j} (A_{i,j} - \bar{A})^2 \sum_{i,j} (T_{i,j} - \bar{T})^2}}^{1/2},$$

where $A$ represents a gridpoint analysis, $\bar{A}$ represents the average value of the analysis, $T$ represents the associated analytic field ("truth") and $\bar{T}$ represents the average value of the analytic field (Wilks 1995). The correlation coefficient\(^7\) provides a better measure of correctness of the pattern of an analysis than does the rmse, whereas the rmse provides a better measure of correctness of the amplitude of an analysis. The correlation coefficients are calculated over the same interior portion of the analysis domain as are the rmse.\(^8\)

#### b. Results

To demonstrate differences in the analysis techniques, consider Figs. 4 and 5, which present analyses of waves whose wavelengths are $L = 6\Delta$ and $L = 20\Delta$, respectively. Wavelength $L = 6\Delta$ represents the smallest of those waves that are considered marginally well sampled (Dowell and Caracena 1988) and whose spatial derivatives are considered crudely approximated, whereas $L = 20\Delta$ represents a well-sampled wave. For these figures, we use SC = 0.5 (Figs. 2b,c), which creates an artificial observing network whose degree of spatial irregularity mimics the irregularity of the station distribution of a declustered U.S. surface network (Spencer 2002).

Figures 4a,b and 5a,b suggest that for scalar objective analysis, the traditional method is superior. The checkerboard patterns appear to be represented quite well by both methods (as evidenced by the high correlation coefficients), but the amplitudes of the waves are reproduced better by the traditional method.\(^9\) As noted by Endlich and Clark (1963), each triangle centroid value of the scalar variable is simply the average of the observations comprising the triangle, thus exposing triangle scalar analyses to the possibility of significant damping.

The real value of the triangle method arises from its ability to analyze derivative fields more accurately than the traditional method (Figs. 4c–f and 5c–f). Significant distortions in the traditional method gradient and Laplacian analyses are quite evident, whereas only minor

\(^6\) Note that the finite differencing scheme required by (11) is applied to the gridded gradient field rather than to the gridded scalar field.

\(^7\) The value of the correlation coefficient lies from $-1$ to $+1$.

\(^8\) For simplicity, the term rmse often is used to encompass the rmsve of the gradient, as well.

\(^9\) The rmse and correlation coefficient for each analysis are listed in the upper-right corner of each plot. See the caption of Fig. 4 for details.

---

Fig. 4. Analyses of the $L = 6\Delta$ wave for the following analysis methods and variables: (a) traditional method and (b) triangle method for the scalar; (c) traditional method and (d) triangle method for the magnitude of the scalar gradient; (e) traditional method and (f) triangle method for the scalar Laplacian. For these analyses, the scatter constant (SC) is 0.5. Dashed contour lines indicate negative values. The numbers in each upper-right corner are rmse and correlation coefficient, respectively. The gradient rmse values have been multiplied by $10^7$ and the Laplacian rmse values have been multiplied by $10^9$. 

Unauthenticated | Downloaded 06/05/24 01:41 PM UTC
FIG. 5. Same as in Fig. 4 except for $L = 20\Delta$. 
distortions are generated by the triangle method. Therefore, these results support those that have already been established by Schaefer and Doswell (1979), Doswell and Caracena (1988), and Spencer and Doswell (2001)—namely, when the fidelity of spatial derivatives is of primary concern, estimating them directly from the observations is preferable to applying a finite differencing scheme to a set of gridded observations. Doswell and Caracena (1988) present a theoretical explanation for the superiority of the triangle method, whereas Spencer (2002) provides a purely intuitive explanation. The reader is referred to these papers for details.

The circumstances for which one analysis technique may be preferred over the other is summarized with the aid of Fig. 6, which shows rmse and correlation coefficient differences across the scatter constant and wavelength spectrum for analyses of the scalar, scalar gradient, and scalar Laplacian. Figure 6a indicates that throughout the spectrum, the traditional method provides scalar analyses that are superior to those created by the triangle method. This is consistent with the figures presented earlier (Figs. 4 and 5) and with the previous discussion regarding the averaging involved in creating triangle centroid estimates of the scalar variable. Figure 6a also suggests that as the wavelength decreases, the superiority of the traditional method over the triangle method increases. This is not surprising, since the basic assumption of the triangle method is that the scalar variable varies linearly across the triangle. Therefore, for a nonlinear field sampled by a given observational network, smaller waves violate the linearity assumption more than larger waves. Figure 6b indicates that the quality of the patterns of the scalar analyses from the two analysis methods—as measured by the correlation coefficient—is virtually indistinguishable across the spectrum.

Figures 6c and 6d compare the rmse and correlation coefficients of the traditional method scalar gradient analyses to those of the triangle method. Figure 6c suggests that the triangle method provides superior gradient analyses over a large portion of the scatter constant and wavelength spectrum. In fact, for data distributions typical of current observing networks (SC ≥ 0.25), the triangle method is superior to the traditional method for all but the smallest of marginally well-sampled waves (L ≤ 9Δ). Figure 6d indicates that the triangle method produces gradient analyses that have higher correlation coefficients than those produced by the traditional method across the entire scatter constant and wavelength spectrum (especially for large scatter constants and small wavelengths), suggesting that the patterns of the analyses from the triangle method are superior to those of the traditional method. For example, Fig. 6c indicates that for L = 6Δ and SC = 0.5, the traditional method provides the superior gradient analysis. However, the difference in the correlation coefficient clearly suggests that the triangle method provides the superior analysis (Fig. 6d). These seemingly contradictory conclusions are supported by Figs. 4c and 4d, which show that the amplitude of the traditional gradient analysis is superior to that of the triangle method, whereas the pattern of the triangle gradient analysis obviously is superior to that of the traditional method. Clearly, neither the rmse nor the correlation coefficient provides a complete picture of analysis error.

Figures 6e and 6f compare rmse and correlation coefficients of the traditional method scalar Laplacian analyses to those of the triangle method. Figure 6e suggests that the triangle method provides Laplacian analyses that are superior to those of the traditional analyses over the entire spectrum of data distributions associated with current U.S. observing networks. The greatest improvement is for small waves and highly irregularly distributed observing stations. Figure 6f indicates that the triangle method produces Laplacian analyses that have higher correlation coefficients over the entire spectrum, again suggesting that the patterns of the triangle analyses are superior.

To summarize, if an accurate analysis of the scalar variable itself is the only concern of the analyst, then the traditional method should suffice. However, if information about the spatial distribution of the scalar (i.e., its derivatives) is important, then the results presented in this section indicate that the triangle method generally should be used instead of the traditional method. In the following section, we develop a variational objective analysis method that combines the best aspect of each method.

5. A variational objective analysis scheme

a. The cost function and Euler–Lagrange equation

In order to create a scalar analysis resembling that of the traditional method and whose derivatives resemble those of the triangle method, we propose a variational formulation whereby the variance between the scalar analysis and the traditional scalar analysis is minimized subject to the weak constraint that the gradient computed from the scalar analysis equals the gradient analysis from the triangle method; that is, we propose minimizing the following cost function J:

\[
J = \int \int \left[ (S_w - S_p)^2 + \gamma \left( \frac{\partial S_w}{\partial x} - \frac{\partial S_p}{\partial x} \right)^2 + \gamma \left( \frac{\partial S_w}{\partial y} - \frac{\partial S_p}{\partial y} \right)^2 \right] dx \, dy, \tag{15}
\]
where $S_y$ represents the desired gridded scalar field (the variational analysis), $S_y$ is the traditional method scalar analysis, $\partial S^2/\partial x$ and $\partial S^2/\partial y$ are the gridded gradient components from the triangle method, and $\gamma$ is a user-defined weight factor that determines the relative strength of the constraint. The Euler–Lagrange (EL) equation associated with (15) is

$$\left( S_y - S_y \right) + \gamma \left[ \frac{\partial}{\partial x} \left( \frac{\partial S^2}{\partial x} \right) + \frac{\partial}{\partial y} \left( \frac{\partial S^2}{\partial y} \right) - \nabla^2 S_y \right] = 0.$$  

(16)

The complete derivation of the EL equation appears in Spencer (2002). Equation (16) suggests that if $\gamma = 0$, then the variational scalar analysis matches the traditional analysis. Rewriting (16) as

$$\nabla^2 S_y - \frac{S_y}{\gamma} = \nabla \cdot (\nabla S^2) - \frac{S_y}{\gamma},$$  

(17)

where $\nabla S^2 = (\partial S^2/\partial x)\hat{x} + (\partial S^2/\partial y)\hat{y}$, it is obvious that as $\gamma$ gets very large, the Laplacian of the variational scalar analysis ($\nabla^2 S_y$) approaches the Laplacian analysis derived from the triangle method ($\nabla \cdot \nabla S^2$). Equation (17) is an elliptic, Helmholtz-type partial differential equation with a forcing function. This equation is solved numerically for $S_y$ using relaxation (Haltiner and Williams 1980, p. 157 ff.).

b. Choosing an appropriate weight factor $\gamma$

To determine what value of the weight factor $\gamma$ is appropriate when the accuracy of the derivative estimates is as important as the accuracy of the scalar field itself, consider Figs. 7 and 8, which plot rmse and correlation coefficients, respectively, of the variational analyses as a function of $\gamma$ for each of three fields ($S_y$, $\nabla S_y$, $\nabla^2 S_y$) and two wavelengths ($L = 6\Delta$ and $L = 20\Delta$). The figures indicate that for $\gamma \approx 10^6$, the variational scalar analyses are equivalent to the traditional analyses (Figs. 7a and 8a). Similarly, Figs. 7c and 8c suggest that for $\gamma > 10^{10}$, the variational Laplacian analyses are equivalent to the triangle Laplacian analyses. Therefore, when $\gamma < 10^6$, the first term on the right-hand side of (15) dominates the cost function $J$.

and when $\gamma > 10^{10}$, the second and third terms dominate.

Clearly, the transition occurs for values $10^6 \leq \gamma \leq 10^{10}$. Curiously, minima in the rmse for $L = 6\Delta$ occur at $\gamma = 10^8$ for the derivative fields (Figs. 7b and 7c). Similarly, a maximum in the correlation coefficient for $L = 6\Delta$ occurs at $\gamma = 10^9$ for the scalar field (Fig. 8a).

Therefore, while the results shown in Figs. 7b and 7c argue for choosing $\gamma = 10^8$, the results in Fig. 8a argue for choosing $\gamma = 10^9$. The remaining results (Figs. 7a, 8b, and 8c) argue for choosing either $\gamma < 10^8$ or $\gamma \geq 10^9$. As a compromise, we choose $\gamma = 5 \times 10^8$ for the variational objective analysis scheme.

c. Results

Variational analyses for $L = 6\Delta$ and $L = 20\Delta$ are shown in Fig. 9. Comparing these to Figs. 4 and 5, it is clear that the use of the variational analysis procedure satisfies our goal — namely, the variational scalar analyses resemble those of the traditional method while at the same time the variational derivative analyses resemble those of the triangle method. A significant portion of the scalar amplitude lost through the triangle technique has been recovered by the variational analysis method. Fortunately, this has not come at the expense of the derivative fields. Only small differences are apparent visually between the variational derivative analyses and those produced by the triangle method. In fact, for these particular plots, the rmse and correlation coefficients of the variational derivative analyses generally suggest that the variational method for derivative analysis is slightly superior to the triangle method.

Plots of rmse and correlation coefficient differences between the traditional and variational analyses across the scatter constant and wavelength spectrum are presented in Fig. 10. Comparing these plots with those presented in Fig. 6, the most obvious difference is the improvement in the scalar analysis across the spectrum that the variational method provides (cf. Figs. 6a and 10a). The variational method scalar rmse are actually

---

13 The units of $\gamma$ are length squared.

14 The results from each of the 10 runs that comprise the average show this same structure.

15 Figure 11, which shows rmse and correlation coefficient differences between the triangle and variational analyses, corroborates the conclusions made in this paragraph and is presented for completeness. Note that Fig. 11 is created simply by subtracting the results of Fig. 6 from those of Fig. 10.
lower than those of the traditional method over much of the spectrum, most notably for well-sampled waves and moderate to large scatter constants (Fig. 10a). Slight improvements in the correlation coefficients provided by the variational method are seen by comparing Figs. 6b and 10b (see also Fig. 11b). As mentioned earlier, improvements in the variational scalar analyses have not occurred at the expense of the derivative fields. In fact, a comparison of Figs. 6c with 10c and Figs. 6e with 10e suggests that the variational method for derivative analysis provides greater improvements over the traditional method than does the triangle method. The improvements are most notable in the marginally well-sampled portion of the spectrum (Figs. 11c,e), which is consistent with Fig. 7. Similarly, a comparison of Figs. 6d and 10d indicates that the variational method provides slightly greater improvements in the correlation coefficient of the gradient analyses than those produced by the triangle method, most notably in the marginally well-sampled portion of the spectrum (see also Fig. 11d). A comparison of Figs. 6f and 10f suggests that differences in the patterns of the triangle and variational Laplacian analyses are inconsequential (see also Fig. 11f).

d. Origin of the improvement of the variational analysis method

The goal in developing the variational method for objective analysis is to provide scalar analyses that re-
semble those produced by the traditional method and to provide derivative analyses that resemble those produced by the triangle method. In the previous section, we discovered a curious result: the variational method may, in fact, provide scalar analyses that are superior to those of the traditional method and at the same time provide derivative analyses that are superior to those of the triangle method.

To help understand this behavior, consider Fig. 12, which shows a small portion of the scalar and gradient analyses for $L = 20\Delta$ that were presented earlier. The letter S in these plots represents locations of the scalar observations and the letter C represents locations of the triangle centroids. As mentioned earlier, the traditional method is susceptible to generating a poor derivative analysis within data voids. For example, the outlined area in Figs. 12a,b encloses an elongated data void region that is characterized by a poor derivative analysis (Fig. 12b). Although the triangle method produces an inferior scalar analysis whose amplitude is somewhat damped (Fig. 12c), it is able to produce a superior derivative analysis (Fig. 12d) by virtue of calculating derivatives directly from the observations. Note that the observations required to create gridded analyses from the traditional method are located at positions marked S (e.g., Fig. 12a), whereas the “obser-
Fig. 9. Same as Fig. 1, except for the variational method analyses. The numbers in each upper-right corner are rmse and correlation coefficient, respectively. The gradient rmsses have been multiplied by $10^5$ and the Laplacian rmsses have been multiplied by $10^9$. 
Fig. 10. Same as in Fig. 6 except that the differences are for the traditional method minus the variational method. (a), (c), (e) Solid lines indicate regions where the variational method rmse is lower; (b), (d), (f) dashed contour lines indicate regions where the variational method correlation coefficients are higher.
Fig. 11. Same as in Fig. 6 except that the differences are for the triangle method minus the variational method.
Fig. 12. Analyses of the $L = 20\Delta$ wave for the following analysis methods and variables: traditional method for the (a) scalar and (b) magnitude of the scalar gradient; triangle method for the (c) scalar and (d) magnitude of the scalar gradient, variational method for the (e) scalar and (f) magnitude of the scalar gradient. These contour plots are simply zoomed-in versions of various plots from Figs. 5 and 9. The letter S represents station locations and the letter C represents triangle centroid locations. The enclosed area highlights a significant data-void region.

required to create gridded analyses from the triangle method are located at positions marked C (e.g., Fig. 12c), which are necessarily offset from the actual scalar observations S. The variational analysis method uses analyses from both the traditional and triangle methods [see Eq. (15)] and therefore uses the information that is available at the positions marked S as well as those marked C (Figs. 12e,f); that is, the var-
ational method incorporates the traditional scalar analysis derived from the observations located at S in Fig. 12a in combination with the triangle derivative analysis derived from the “observations” located at C in Fig. 12d. Therefore, by using both sets of information, the variational method incorporates derivative information within the regions where the traditional analysis is most poor—between observations (see the enclosed area in Figs. 12e,f). By doing this, the variational method is able to produce superior analyses.

6. Summary and discussion

A comparison of different methods for the analysis of scalar variables and their derivatives has been presented. The traditional method for scalar analysis, which maps observations to a grid via a distance-dependent weighted averaging scheme, is shown to be superior to the triangle method, which maps triangle centroid scalar estimates to a grid using the same weighted averaging scheme. The triangle method for derivative estimation, however, is generally superior to the traditional method, which applies a finite-differencing scheme to the set of gridded observations. Slight distortions in the traditional scalar analysis manifest themselves when finite differencing is used to estimate derivatives. The triangle method, on the other hand, estimates derivatives directly from the observations and is not prone to the problems associated with finite differencing.

The superiority of the traditional method for scalar analysis and the superiority of the triangle method for derivative analysis has motivated the development of an objective analysis scheme based on a variational method. This variational objective analysis scheme combines the best aspect of the traditional method for objective analysis with the best aspect of the triangle method; namely, the scalar analysis from the variational method resembles that produced by the traditional method, whereas the derivative analyses from the variational method resemble those of the triangle method. In fact, the results using analytic observations indicate that with a judicious choice of the weight factor \( g \) associated with the variational formulation, the rmse of the variational scalar analyses actually may be lower than those associated with the traditional analyses, especially for long wavelengths and moderately to highly irregular station distributions. When the correlation coefficients are compared, the variational scalar analyses are slightly superior for all combinations of the data distribution and wavelength. Similarly, for most combinations of the wavelength and data distribution, the variational derivative analysis rmse is superior to those of both the traditional and triangle methods. The correlation coefficients of the variational derivative analyses are higher than those of the traditional derivative analyses for all combinations of the wavelength and data distribution (increasingly so as the data become more irregularly distributed). Differences in the correlation coefficients of the derivative analyses between the variational and triangle methods are small.

The goal in developing the variational method was not to outperform the traditional and triangle methods, per se, but simply to combine the best aspects of each method in a logical manner to create scalar and derivative analyses that are mutually consistent. The superiority of the variational method is seen to be a consequence of its better use of available information—the variational method not only uses the scalar observations themselves, but derivative information at points between observations is used, as well. The use of both sets of information allows the variational method to produce better analyses than if only one set of information were used.

The use of derivative information for the analysis of a scalar variable is somewhat reminiscent of a multivariate analysis technique for the analysis of geopotential heights. In such a scheme, the analysis of the height field is accomplished with the aid of wind observations, which represent the derivative of the heights through the geostrophic relationship. Similarly, our variational scheme uses derivative information to constrain the scalar analysis, although this is not done through a dynamic constraint, but rather through the arbitrary choice of the weight factor \( g \). We also note that multivariate techniques, by definition, necessarily involve more than one variable, whereas our technique is strictly univariate.

We hope that the results of this study inspire analysts to reconsider using popular methodologies for producing gridded estimates of scalar quantities and their derivatives. For example, the use of traditional successive correction schemes such as those of Cressman (1959) and Barnes (1964, 1973) for creating gridded scalar fields and the finite differencing, Bellamy (1949), line integral, and LVPF methods for creating gridded derivative fields should be reevaluated in light of our findings. We do not proclaim that these other methods are necessarily inappropriate when they are applied intelligently; we simply wish to emphasize that superior analyses may be available through a variational objective analysis scheme. Because terms involving spatial derivatives of scalar quantities permeate many of the important equations used by diagnosticists, these derivatives should be estimated with great care. The variational objective analysis scheme developed herein provides a means for accurately estimating these derivative terms while simultaneously providing scalar analyses of the variables that are consistent with the derivatives.

An avenue for future research is to incorporate the philosophy of our variational analysis method into numerical modeling systems that use variational methods that do not drive the analysis toward derivative “observations.” Perhaps improved forecasts of precipitation, for example, may be obtained if consideration is made for fitting the analysis to derivatives of the observed wind field (especially divergence) and to derivatives of the observed scalar variables (see Daley 1985).
Our variational objective analysis scheme was tested using analytic observations that form checkerboard patterns of various wavelengths. Although these simple patterns provide a good initial test of the analysis scheme, they do not replicate complex atmospheric patterns. To further test the robustness of the superiority of the variational scheme for objective analysis, future work will compare analyses derived from more realistic and complex patterns. For example, a set of gridpoint data randomly selected from a model analysis may be treated as observations by the various analysis methods and the resulting analyses may be compared to the complete model analysis. The random grid points can be selected such that any degree of irregularity in the spatial distribution of the pseudo-observations may be produced, ranging from highly regular to highly irregular. This flexibility will allow us to determine how well the various analysis techniques perform as a function of the spatial irregularity of the data, much like what was done through the use of the scatter constant parameter for the analytic data studies. In fact, the distribution of the randomly selected grid points can be forced to replicate the distribution of stations within actual observing networks in order to provide comparisons when the analysis schemes are applied to data from such networks.

The use of randomly chosen model gridpoint data as input for the variational analysis scheme also should provide useful information regarding an appropriate choice for the user-defined weight factor $\gamma$. The tests using analytic observations indicate that reasonable choices for $\gamma$ may span a range of an order of magnitude or more (Figs. 7 and 8). Additional tests, especially those involving datasets containing more realistic atmospheric structure, need to be performed in order to determine appropriate values of $\gamma$ for diagnostic studies of the atmosphere.

Acknowledgments. This work could never have been completed without the encouragement, support, and guidance of many individuals. This research summarizes the Ph.D. research of the first author (PLS) and he is grateful to his coauthors for their availability and advice. The first author is also thankful to Mr. Mike Eilts, Dr. Jeff Kimpel, and the Cooperative Institute for Mesoscale Meteorological Studies (CIMMS), for encouraging me and allowing me to pursue the degree at The Pennsylvania State University after working at the National Severe Storms Laboratory (NSSL) for several years. We also wish to thank the other committee members, Drs. Paul Markowski, David Stauffer, and Mark Levi, for their comments on an earlier draft of the manuscript. Thanks also to Dr. Chuck Doswell for his continuing interest and comments regarding this work. We also wish to acknowledge Doswell and Mr. Dave Watson for providing us the Delauney triangulation code, which makes this work possible. Finally, we would like to express our appreciation to the two anonymous reviewers for their suggestions and comments. Major funding for this research was provided under NOAA-OU Cooperative Agreement #NA17RJ1227.

APPENDIX

Superob Technique

The superob technique used for this study is illustrated with the aid of Fig. A1, which depicts four clustered observing sites, indicated by the circled numbers. The algorithm first averages the locations (and data values) of stations 1 and 2, thus generating a superob at location A. Stations 1 and 2 are then no longer considered to be part of the dataset. Clustered stations A and 3 are then combined to produce superob B. Notice that because two observations influence the location and value at A, superob A receives twice the weight when it is combined with station 3. Once superob A and station 3 are weighted accordingly and combined to create superob B, superob A and station 3 are removed from the dataset. Finally, following the same procedure, clustered stations B and 4 are weighted accordingly and combined to create superob C. Superob B and station 4 are then removed from the dataset. Therefore, in this illustration, clustered stations 1–4 are combined to produce one superob located at C.

REFERENCES


——, 1973: Mesoscale objective analysis using weighted time-series
obs. NOAA Tech. Memo. ERL NSSL-62, National Severe Storms Laboratory, Norman, OK, 41 pp. [NTIS COM-73-10781.]


