A Variational Method for Filling in Missing Data in Doppler Velocity Fields

VINCENT T. WOOD,[a] ROBERT P. DAVIES-JONES,[a] AND ALAN SHAPIRO[b]

[a] *NOAA/OAR/National Severe Storms Laboratory, Norman, Oklahoma*
[b] *School of Meteorology, and Center for Analysis and Prediction of Storms, University of Oklahoma, Norman, Oklahoma*

ABSTRACT: Single-Doppler radar data are often missing in important regions of a severe storm due to low return power, low signal-to-noise ratio, ground clutter associated with normal and anomalous propagation, and missing radials associated with partial or total beam blockage. Missing data impact the ability of WSR-88D algorithms to detect severe weather. To aid the algorithms, we develop a variational technique that fills in Doppler velocity data voids smoothly by minimizing Doppler velocity gradients while not modifying good data. This method provides estimates of the analyzed variable in data voids without creating extrema. Actual single-Doppler radar data of four tornadoes are used to demonstrate the variational algorithm. In two cases, data are missing in the original data, and in the other two, data are voided artificially. The filled-in data match the voided data well in smoothly varying Doppler velocity fields. Near singularities such as tornadic vortex signatures, the match is poor as anticipated. The algorithm does not create any velocity peaks in the former data voids, thus preventing false triggering of tornado warnings. Doppler circulation is used herein as a far-field tornado detection and advance-warning parameter. In almost all cases, the measured circulation is quite insensitive to the data that have been voided and then filled. The tornado threat is still apparent.

KEYWORDS: Tornadogenesis; Radars/Radar observations; Optimization; Statistics

1. Introduction

Doppler radar data are often lost in significant regions of a severe storm because of low return power or low signal-to-noise ratio. Missing radar data or data voids are a pervasive problem for diagnostic data analysts who employ quantitative research methods. The sources of data voids are ground clutter associated with normal and anomalous propagation (black speckles, Figs. 1a–d, for example), missing radials from radar (often associated with partial or total beam blockage), (Figs. 1e,f), and other causes (Serafin and Wilson 2000). The blockage is typically caused by nearby obstacles (terrain, buildings, and wind farms). The analysts rely on a variety of ad hoc algorithms that attempt to "fix" the missing data by discarding incomplete cases or by filling in (or "imputing") the data voids. These attempts may be prone to bias that may affect the representativeness of the result. The use of a gap-filling procedure to provide more complete data coverage over a computational domain can benefit detection algorithms, rainfall maps, and wind field analyses.

Our main motivation in this paper is to improve the performance of radar-based tornado-warning algorithms. Missing data present difficulties for severe-weather algorithms. Although filling in null values smoothly will not vastly improve detection of vortices, a bad filling algorithm can increase false alarms by introducing spurious maxima of shear or decrease probability of detection by smoothing out singularities (Nuss and Titley 1994, p. 1615; Mahalik et al. 2019). Ideally an algorithm for filling in gaps in gridded data should preserve good data and not insert extrema that could mislead the algorithm.

There are numerous ways to interpolate data. Franke (1982) reviews and evaluates schemes for scattered data such as the multiquadric method (Hardy 1971; Nuss and Titley 1994). Franke defines a global method as one for which the interpolated value at a point is dependent on all data points and addition or deletion of a data point propagates throughout the domain of definition. Global methods are not well suited to radar applications where the domain is often a user-defined region of a storm. Many of the schemes assessed by Franke are global. Moreover, the schemes are not required to reproduce the data exactly at the scattered observation points. Some methods have at least one free parameter that has to be chosen judiciously. A bad choice of a parameter can lead to instability (Isom et al. 2009, p. 901). Other methods make restrictive statistical assumptions about the data (e.g., kriging). Typically, these methods assume that the data are stationary, isotropic, and/or fit a normal distribution. We conclude that schemes for scattered data are not really applicable to radar data, which is ordered on a grid with some points lacking data.

We now consider local methods for gridded data that are more appropriate for use with radar algorithms. In this context, Mahalik et al. (2019) used a least squares-plane method. The least squares fit is to data defined on local neighborhoods (3×3 or larger subgrids in their study). Their median filter modifies the data prior to their regression analysis. Interpolated values would not match gridded observation values even if the data were unfiltered. Methods such as bicubic spline interpolation are more advantageous as the interpolant passes through the good data points. The fact that the extrema of the interpolant may lie outside the range of the data is an undesirable feature for our specific application. We next discuss the properties that
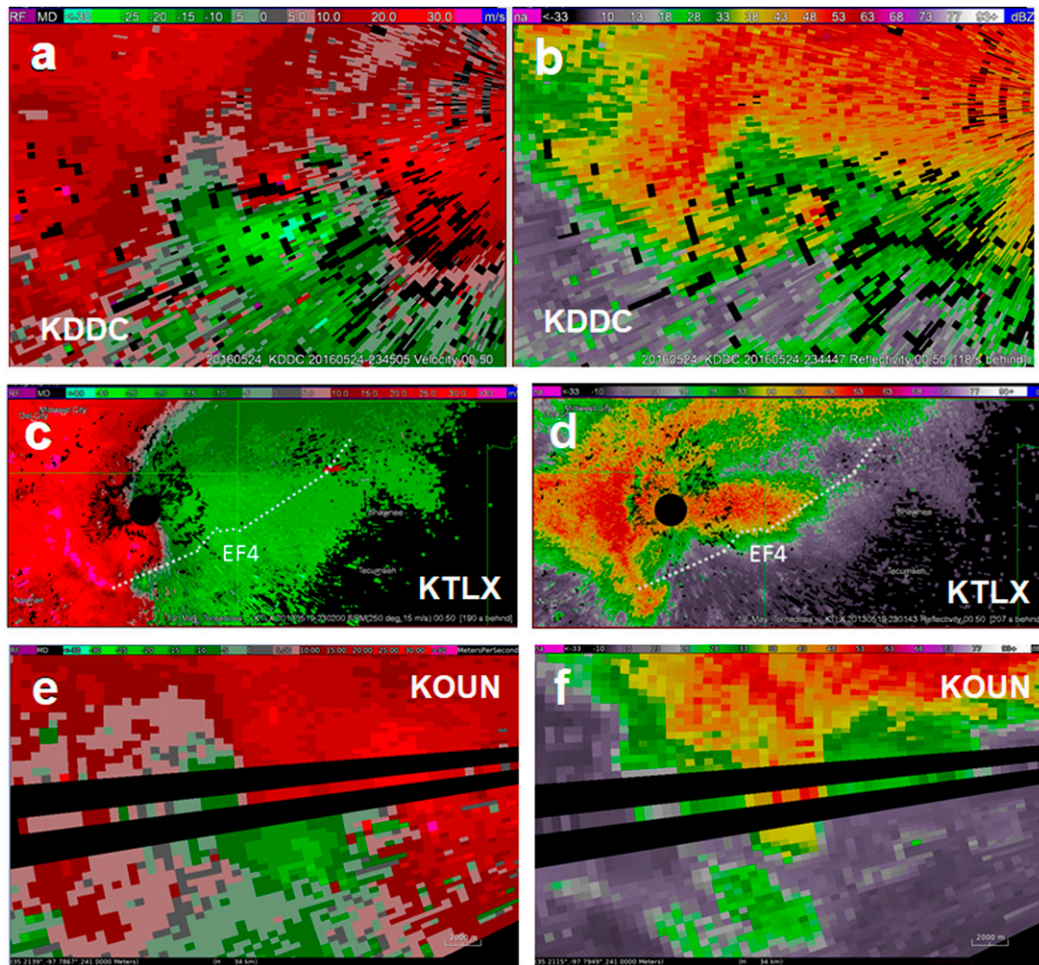
FIG. 1. DDC WSR-88D scans of (a) ground-relative, mean Doppler velocity (V_r, m s^{-1}) and (b) radar reflectivity (Z, dBZ) of the tornadic storm at the 0.5° launch angle at ~2345 UTC 24 May 2016. Oklahoma City (TLX) WSR-88D scans of (c) ground-relative, mean Doppler velocity and (d) radar reflectivity of a storm with an EF4 tornado as collected at 0.5° launch scan at 2302 UTC 19 May 2013. A dotted curve depicts damage path of the Lake Thunderbird–Bethel Acres–Shawnee tornado. Norman (OUN) WSR-88D scans of (e) ground-relative, mean Doppler velocity and (f) radar reflectivity of a significant mesocyclone signature in association with a hook echo at an approximate range of 36 km and azimuth of 262°, as collected at the 0.5° launch angle at 0027 UTC 6 May 2015. The superresolution data were collected with 0.25-km range and 0.5°-azimuth intervals. Data are missing along five radials from 260.22° to 262.22° and along five radials from 263.22° to 265.22°.

we believe a gap-filling technique should have for algorithms that detect tornadoes and mesocyclones.

Unlike image processing where outliers are regarded as bad pixels, the outliers in vortex detection are important information rather than unwanted noise. No method for filling in missing data, however simple or complex, will perform well near a singularity. A tornado is almost always unresolved by the grid and appears as a singularity. A tornadic vortex signature (TVS; Brown et al. 1978) can indicate the presence of a tornado. A large value of Doppler circulation (Davies-Jones et al. 2020, hereafter DJ20) reveals the possibility of a large violent tornado.

For our purposes we need a gap-filling algorithm that applies to gridded data of a scalar field (viz., Doppler velocity) on a specified 2D domain with some gaps in the data (i.e., data is missing at some assortments of data points). We wish to fill in the data by minimizing a cost function, namely, the mean squared magnitude of the gradient of the scalar, over each data *void*. Our other requirements are that (i) good data are not filtered or modified, (ii) maxima or minima are not introduced into the field, (iii) gaps are filled using only the nearest surrounding good data so the method is local, (iv) there are no free parameters that have to be chosen judiciously for each application, and (v) the algorithm is stable and works without fail. Undoubtedly other researchers with different needs will prefer different methods.

These requirements eliminate methods that (i) find best fits of mathematical functions to scattered data (Hardy 1971), (ii) obtain best least squares plane fits that do not pass through the data points (Mahalik et al. 2019), (iii) bicubic splines that fit

the data but with possible overshoots, (iv) median-filtering techniques used to fix bad pixels and produce pleasing images, (v) statistical methods that minimize some statistical measure while making assumptions about the statistical distribution of the data, and (vi) methods that have free parameters.

A variational method does meet all our requirements. Variational analysis provides the necessary Euler–Lagrange equation for minimizing the cost function in the data voids (along with the natural boundary conditions if the void has points on the domain boundary). The solution of the variational problem exists and is unique. The variational method is the one that minimizes gradients across data voids. There is no need to collect extensive statistics on numerous case studies and patterns of missing data to demonstrate that this is the preferred method for our specifications. Owing to the maximum principle (Weinberger 1965, 55–57), the algorithm will not place a local extremum in a data void that is fully within the interior of the domain. The method produces plots of Doppler velocity that vary smoothly across data voids without insertion of false spikes and peaks. More complex interpolation methods could introduce false extrema that a vortex-detection algorithm could latch on to.

The data-filling algorithm developed in the present study is similar to one of the procedures examined by Ellis (1997). However, while Ellis worked with numerical model data on a Cartesian grid, we consider actual radar data on a radar grid. Ellis tested gap-filling algorithms based on the Cartesian form of two-dimensional (2D) and three-dimensional (3D) Laplace and biharmonic equations so that Doppler radar data could be used to initialize regional scale models. Laplace's equation arose from the minimization of the (squared) first spatial derivatives of the field being void filled, while the biharmonic equation arose from the minimization of the (squared) second spatial derivatives. Ellis performed gap-filling experiments in an observing system simulation experiment (OSSE) framework using numerically simulated supercell data. To simulate radar data, the model output was assumed to be "missing data" if the rainwater mixing ratio value was below a specified threshold at a particular grid point. This was accomplished by flagging the grid point as one to be data filled. In tests with numerically simulated Cartesian velocity data, it was determined that the Laplace equation approach yielded better results than the biharmonic equation approach. Moreover, the 2D procedure was notably better than the 3D procedure.

This paper is organized as follows. Section 2 describes a variational method (Hildebrand 1965) to fill in data voids on surfaces of constant launch angle (Davies-Jones et al. 2019, hereafter DJ19) with values that minimize gradients of a scalar such as Doppler velocity. Several illustrative examples are provided in section 3. Section 4 uses actual single-Doppler data of four different tornadoes to explore the utility of the algorithm. Conclusions and future work follow in section 5.

2. Description of the variational methodology

Our variational method fills in voids in gridded data of a general scalar field with grid values that minimize the mean square magnitude of the gradient in each void. The method works for all configurations of missing data. It does not introduce local extrema into the field. To develop and test the performance of the method in a radar coordinate system, we followed the DJ19 approach. Under standard refraction conditions, the ray curvature is negligible, and we use the measurable radar coordinates (r, α, β) centered on the radar where r is the slant range (arclength along a stationary ray), α is the launch angle (elevation angle of the radar ray at the radar antenna), and β is the azimuth angle measured clockwise from due north. For straight rays, the surfaces of constant α are conical for $\alpha \neq 0$ and the surface $\alpha = 0$ is the plane tangent to the (assumed spherical) Earth at the radar. The gradient of a scalar field S on a particular surface of constant elevation angle α_o is

$$\nabla_\alpha S = \frac{\partial S}{\partial r}\hat{\mathbf{r}} + \frac{1}{r\cos\alpha_o}\frac{\partial S}{\partial \beta}\hat{\boldsymbol{\beta}}, \qquad (1)$$

where $\hat{\mathbf{r}}$ and $\hat{\boldsymbol{\beta}}$ are, respectively, unit vectors along the ray and perpendicular to the ray in the azimuthal direction. We seek the function $S(r, \beta)$ that provides a stationary value of the integral J

$$J \equiv \frac{1}{2}\iint_A \left[\left(\frac{\partial S}{\partial r}\right)^2 + \frac{1}{r^2\cos^2\alpha}\left(\frac{\partial S}{\partial \beta}\right)^2 \right] r\cos\alpha_o \, dr \, d\beta, \qquad (2)$$

where the quantity in square brackets is the magnitude squared of $\nabla_\alpha S$, A is an area in the α_o surface bounded by slant-range circles $r = r_1$ and $r = r_I$ and radials $\beta = \beta_1$ and $\beta = \beta_K$, and $r\cos\alpha_o dr d\beta$ is the element of area dA. The subscript I in r_I (K in β_K) represents the maximum index in the range (azimuth) direction. For a stationary value of the cost function J, we set $\delta J = 0$. By the calculus of variations (Hildebrand 1965), we have

$$\delta J = 0 = \iint_A \left(r\frac{\partial S}{\partial r}\delta\frac{\partial S}{\partial r} + \frac{1}{r\cos^2\alpha_o}\frac{\partial S}{\partial \beta}\delta\frac{\partial S}{\partial \beta} \right)\cos\alpha_o \, dr \, d\beta$$

$$= \iint_A \left(r\frac{\partial S}{\partial r}\frac{\partial\delta S}{\partial r} + \frac{1}{r\cos^2\alpha_o}\frac{\partial S}{\partial \beta}\frac{\partial\delta S}{\partial \beta} \right)\cos\alpha_o \, dr \, d\beta$$

$$= -\iint_A \left[\frac{\partial^2 S}{\partial r^2} + \frac{1}{r}\frac{\partial S}{\partial r} + \frac{1}{r^2\cos^2\alpha_o}\frac{\partial^2 S}{\partial \beta^2} \right]\delta S \, dA$$

$$+ \int_{\beta_1}^{\beta_2}\left[r\frac{\partial S}{\partial r}\delta S \right]_{r_1}^{r_2}\cos\alpha_o \, d\beta + \int_{r_1}^{r_2}\left[\frac{1}{r\cos\alpha_o}\frac{\partial S}{\partial \beta}\delta S \right]_{\beta_1}^{\beta_2} dr, \quad (3)$$

where we have used integration by parts. The variation δS is zero where the data are good and is arbitrary where the data are missing. In the data voids (where $\delta S \neq 0$), the minimizing function $S(r, \beta)$ must satisfy the Euler–Lagrange equation:

$$\frac{\partial^2 S}{\partial r^2} + \frac{1}{r}\frac{\partial S}{\partial r} + \frac{1}{r^2\cos^2\alpha_o}\frac{\partial^2 S}{\partial \beta^2} = 0, \qquad (4)$$

and the natural boundary conditions

$$\frac{\partial S}{\partial r} = 0 \quad \text{at} \quad r = r_1, r_I \quad \text{and} \quad \frac{\partial S}{\partial \beta} = 0 \quad \text{at} \quad \beta = \beta_1, \beta_K \qquad (5)$$
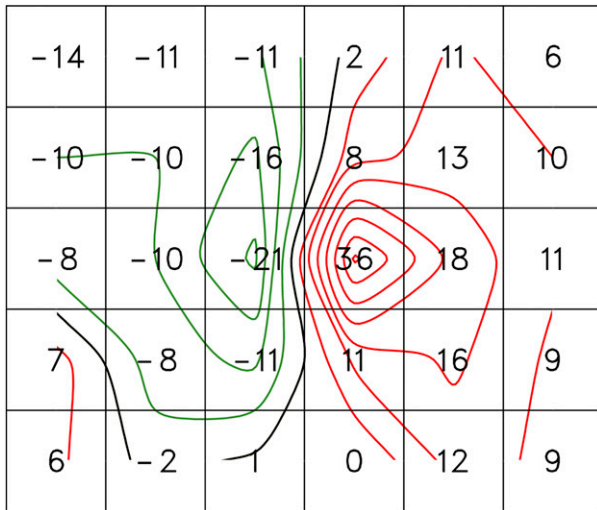
FIG. 2. Superimposition of Doppler velocity contours on the field of actual mean Doppler velocities (m s^{-1}) at 3.8° launch angle for the Union City tornado at 1546 CST 24 May 1973. Doppler velocity contours are drawn at intervals of 5 m s^{-1}. Green (red) contours represent flow toward (away from) the radar site. Black contours represent flow perpendicular to the radar viewing direction.
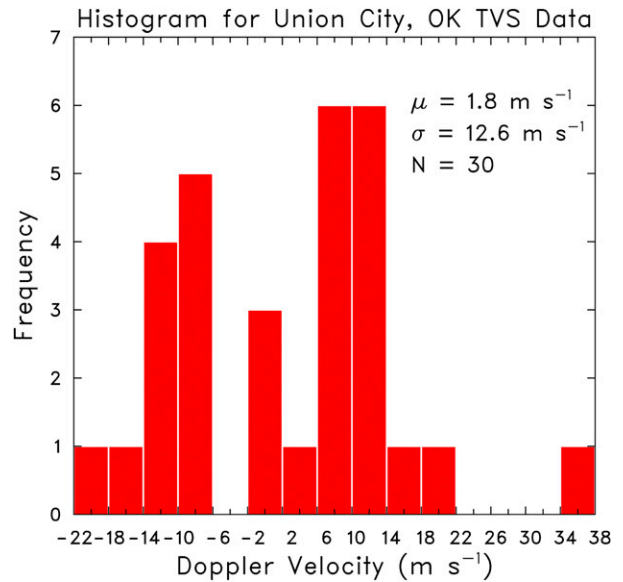

FIG. 3. Histogram of Doppler velocity data in Fig. 2. The values μ, σ, and N are the mean, standard deviation, and number of data points, respectively.

if S is missing at a domain boundary point. For the terms in (4) and (5) we use the following finite-difference expressions obtained from Taylor-series expansion to second order:

$$\left(\frac{1}{r}\frac{\partial S}{\partial r}\right)_{i,k} = \begin{cases} \dfrac{S_{i+1,k} - S_{i-1,k}}{2r_i\Delta r} & \text{for } 2 \le i \le I-1 \\ 0 & \text{for } i = 1, I \end{cases}, \quad (6)$$

$$\left(\frac{\partial^2 S}{\partial r^2}\right)_{i,k} = \begin{cases} \dfrac{S_{i+1,k} - 2S_{i,k} + S_{i-1,k}}{(\Delta r)^2} & \text{for } 2 \le i \le I-1 \\ \dfrac{2S_{i+1,k} - 2S_{i,k}}{(\Delta r)^2} & \text{for } i = 1, \text{ and} \\ \dfrac{2S_{i-1,k} - 2S_{i,k}}{(\Delta r)^2} & \text{for } i = I \end{cases} \quad (7)$$

$$\left(\frac{\partial^2 S}{\partial \beta^2}\right)_{i,k} = \begin{cases} \dfrac{(\beta_{k+1} - \beta_k)S_{i,k-1} - (\beta_{k+1} - \beta_{k-1})S_{i,k} + (\beta_k - \beta_{k-1})S_{i,k+1}}{0.5(\beta_{k+1} - \beta_{k-1})(\beta_{k+1} - \beta_k)(\beta_k - \beta_{k-1})} & \text{for } 2 \le k \le K-1 \\ \dfrac{2S_{i,2} - 2S_{i,1}}{(\beta_2 - \beta_1)^2} & \text{for } k = 1, \\ \dfrac{2S_{i,K-1} - 2S_{i,K}}{(\beta_K - \beta_{K-1})^2} & \text{for } k = K \end{cases} \quad (8)$$

where i is the index in the range direction and k is the index in the azimuth direction. We have used the natural boundary conditions in the formulas at boundary points. Note that the grid spacing in the radial direction $\Delta r = r_{i+1} - r_i$ is constant, but the spacing in the azimuthal direction may vary to accommodate phased-array radar. Inserting (6)–(8) into (4) results in a finite-difference version of the Euler–Lagrange equation, which relates the scalar at an interior grid point (i, k) to the scalar at the two adjacent points on the same radial $(i - 1, k)$ and $(i + 1, k)$, and the two adjacent points $(i, k - 1,)$ and $(i, k + 1)$ on the same slant-range circle. The same applies for boundary and corner

points except neighboring points that lie outside the domain are eliminated. The discrete Euler–Lagrange equation is

$$D_i S_{i+1,k} + 2\mu_i^2 E_k S_{i,k+1} + B_i S_{i-1,k} + 2\mu_i^2 A_k S_{i,k-1}$$
$$- 2(1 + \mu_i^2)C_k S_{i,k} = 0, \quad (9)$$

where $\lambda_i \equiv \Delta r/r_i$, $\mu_i \equiv \Delta r/(r_i \cos\alpha_o)$ and

$$B_i = \begin{cases} 0 & \text{for } i = 1 \\ 1 - \lambda_i/2 & \text{for } 1 < i < I, \\ 2 & \text{for } i = I \end{cases} \quad (10)$$

FIG. 4. Illustrative examples 1–4 are drawn from the example shown in Fig. 2. Data are missing only at the pink points. Values at the neighboring green points are used to fill in values at the pink points. Values at the white points surrounding green points have no effect on the filled-in values at the pink points.

$$D_i = \begin{cases} 2 & \text{for} \quad i = 1 \\ 1 + \lambda_i/2 & \text{for} \quad 1 < i < I \\ 0 & \text{for} \quad i = I \end{cases}, \tag{11}$$

$$C_k = \begin{cases} (\beta_{k+1} - \beta_k)^{-2} & \text{for} \quad k = 1 \\ (\beta_{k+1} - \beta_k)^{-1}(\beta_k - \beta_{k-1})^{-1} & \text{for} \quad 1 < k < K \\ (\beta_k - \beta_{k-1})^{-2} & \text{for} \quad k = K \end{cases}, \tag{12}$$

$$A_k = \begin{cases} 0 & \text{for} \quad k = 1 \\ (\beta_{k+1} - \beta_{k-1})^{-1}(\beta_k - \beta_{k-1})^{-1} & \text{for} \quad 1 < k < K \\ (\beta_k - \beta_{k-1})^{-2} & \text{for} \quad k = K \end{cases}, \tag{13}$$

and

$$E_k = \begin{cases} (\beta_{k+1} - \beta_k)^{-2} & \text{for} \quad k = 1 \\ (\beta_{k+1} - \beta_{k-1})^{-1}(\beta_{k+1} - \beta_k)^{-1} & \text{for} \quad 1 < k < K \\ 0 & \text{for} \quad k = K \end{cases}. \tag{14}$$

Prior to filling in data voids, we pick a domain defined by two range circles and two radials on a particular surface of constant launch angle. Since the algorithm minimizes the gradient of the scalar, the domain, however, should be chosen so that it encompasses a storm or mesocyclone with little wasted area outside the echo. The algorithm does not discriminate between missing values owing to poor data quality and those due to low reflectivity. WSR-88D algorithms such as tornado detection algorithm (Mitchell et al. 1998) and mesocyclone detection algorithm (Stumpf et al. 1998; Mahalik et al. 2019) can use the output from our algorithm.

The procedure for filling in data voids is as follows. We flag the grid points with missing data. At grid points with data, the scalar value is the observed one. For efficiency we use two preliminary nonessential steps. At step 1 we look for isolated missing data points (i.e., those points whose neighbors all have good data). At each of these points, the scalar is calculated from the discrete Euler–Lagrange equation in (9) and the point is then unflagged (i.e., its flag is removed) because the data void has been filled. At step 2 we look for two-point voids (data missing at just two contiguous points). The Euler–Lagrange equation provides two simultaneous equations, which are solved for the scalar at the two points, which are
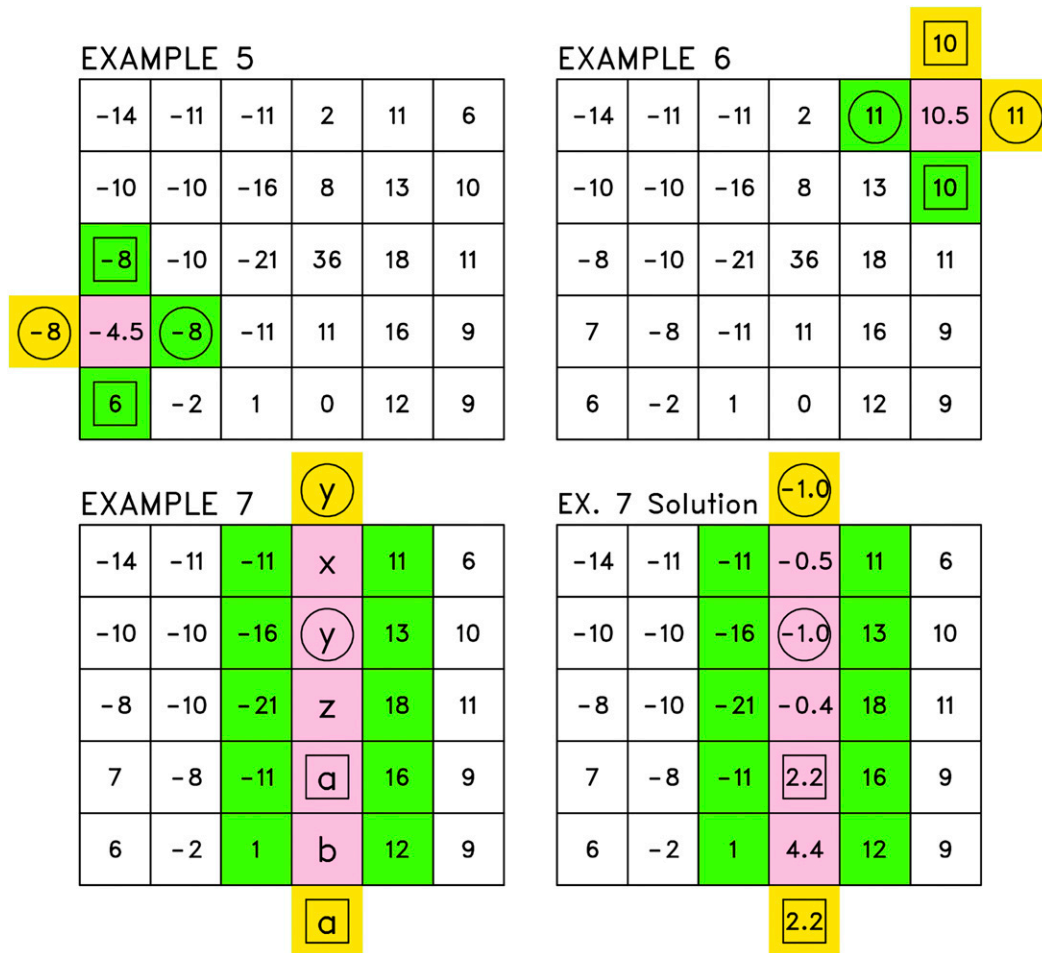
FIG. 5. As in Fig. 4, but illustrating how missing data are filled in at boundary points. Example 5 is for a boundary point that is not at a corner of the domain. Setting the yellow-circled value equal to the green-circled one ensures that the natural boundary condition is satisfied at the pink boundary point. Example 6 demonstrates how a missing datum is filled in at a corner point of the domain. The natural boundary condition is satisfied at the pink corner point when the circled values are equal and the boxed values are equal. In example 7, the data along the radial (shaded pink) through the peak value in the original field (36 in Fig. 2) are assumed missing. The natural boundary condition is satisfied at the pink boundary points when the circled values are equal and the boxed values are equal. The variables x, y, z, a, b, along the radial in example 7 are the fill values to be determined. The values obtained by the variational algorithm are shown in example 7 solution.

then unflagged. Finally, at step 3, all the larger data-void regions are filled simultaneously by solving a system of equations consisting of (9) at the remaining flagged points and $S_{i,k}$ equal to its observed or previously filled-in value at the unflagged points. At the points that are still flagged, S is set initially equal to $0\,\mathrm{m\,s^{-1}}$. The system is solved for the unknown values of S at the flagged points using a red-black relaxation method (e.g., Press et al. 1996). Each grid point is portrayed with a red or black dot according to whether $i + k$ is odd or even. The resulting grid is a checkerboard pattern with red points having only black neighbors and vice versa. We update the value at a red flagged point first by computing a new value that makes the residual there vanish. This increases the residuals at the black neighboring flagged points, but not

at any other red points. We can therefore perform updates at all the red flagged points simultaneously. We then use the updated values at the red points to update the values at all the black-flagged points. We repeat the process until all the residuals are below a specified tolerance. The relaxation method always converges, albeit slowly (Thompson 1961, p. 94; Haltiner and Williams 1980, p. 158). The method converges in fewer iterations if overrelaxation is used. Thus, the new values are computed using a larger correction than needed to make the residual vanish (Haltiner and Williams 1980, p. 159). The optimum overrelaxation parameter is greater than 1 by definition and just below 2 for data voids with a very large number of missing data points. We use an overrelaxation coefficient of 1.3 because the sizes of the data voids are usually small.

The algorithm is extremely stable. Even under the extreme conditions of missing data at all grid points but one, it will find a solution. This solution (a constant field equal to the one good data value) is obviously trivial, but the point is that the algorithm does not fail.

Note that the method is local rather than global. Even though the gaps are filled simultaneously, they are filled independently of one another. An advantage of the method is that the filling of an interior gap is independent of domain choice.

We justify our choice of cost function as follows. First consider the problem of finding the electrostatic potential S inside a closed curve when it is given on the curve and the interior is devoid of positive and negative electric charges. The potential is the solution of Laplace's equation subject to Dirichlet conditions (Farlow 1993, 245–263). (If there were a charge distribution in the interior, the equation would be a Poisson one instead.) The solution is the one that minimizes the squared magnitude of the potential gradient (the cost function) over the region. Conversely, minimization of this cost function results in an Euler–Lagrange equation that is Laplace's equation (Hildebrand 1965, 138–139). In the absence of an interior charge distribution, it follows from the maximum principle (Weinberger 1965, 55–57) that extreme values of the potential must occur on the boundary curve, not inside the curve.

Now consider the problem of minimizing the same cost function for a Doppler velocity field (the potential in this case) with a small data void that is not extremely close to the radar. We can neglect the divergence of the radials and the curvature of the surface of constant launch angle. Then the geometry of the region containing the data void is planar to a good approximation. The problem becomes analogous to the electrostatic one so the Euler–Lagrange equation for minimizing $|\nabla_\alpha S|^2$ over a planar data void is Laplace's equation $\nabla_a^2 S = 0$. Thus, the main advantages of our choice of cost function for our application are (i) the filled-in values are determined solely by values at the boundary of the void, and (ii) there are no interior sources and sinks (charges) of Doppler velocity so extreme values do not occur in the interior of the void.

3. Illustrative examples using a simplified version of the variational procedure

This section describes a simplified version of the variational algorithm to illustrate the basic ideas behind the method. The illustrative examples concern the Doppler radar signature of a tornado. Initially, we make a few oversimplifications to make calculations straightforward. In section 4, we outline how the algorithm works in radar applications.

We begin with the original Doppler velocity field $S = V_r$ (Fig. 2). This is actually a part of the Doppler velocity field of the Union City, Oklahoma, tornadic vortex signature (TVS) observed on the Norman Doppler radar at a height of about 3.5 km AGL (Brown et al. 1978). The deduced mean rotational velocity of the TVS is 28.5 m s^{-1} = [36 − (−21)]/2 m s^{-1}. This velocity is calculated as $\overline{V}_{rot} = \Delta V/2$, where $\Delta V = V_{ob} - V_{ib}$ is the velocity difference between the outbound (ob, positive)
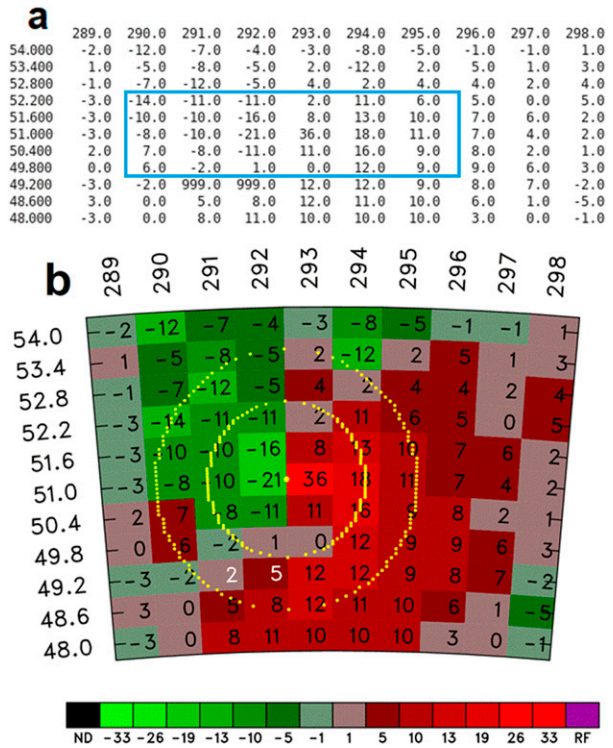

FIG. 6. (a) The B scan of the field of actual, storm-relative mean Doppler velocities (m s^{-1}) at 3.8° launch angle for the Union City TVS at 1546 CST 24 May 1973. The domain here has been expanded from the blue boxed domain shown in Fig. 2. The value of 999.0 represents a missing data parameter. Each missing datum is indicated with a flag. (b) The mean Doppler velocity field with two filled-in Doppler velocity values (white numbers). Black Doppler velocity values are unmodified. The Doppler circulations and mean convergences are calculated from the observed Doppler velocity field around the yellow dotted circles (1.5 and 2.5 km in radius) centered on the TVS axis (yellow heavy dot at 51 km range and 292.5° azimuth) in the 3.8° surface of constant launch angle. Ranges (km) on the left side and azimuths (°) on the top are from Norman Doppler radar. The horizontal color label bar provides the contour levels for Doppler velocity. Black "ND" represents no data (missing data); purple "RF" represents a range-folded value.

and inbound (ib, negative) Doppler velocity peaks in the characteristic velocity couplet (e.g., Wood and Brown 1997). To keep the calculations as simple as possible, we will assume that (i) the geometry is Cartesian, (ii) the grid spacing is the same in the radial and azimuthal directions, and (iii) the domain consists just of the grid points shown in Fig. 2. With these simplifications, the Euler–Lagrange equation reduces to Laplace's equation and the centered second ordered finite difference of the Laplacian becomes the formula for bilinear interpolation. For voids with points on the domain boundary, the natural boundary conditions are still needed for closure of the equations.

Figure 3 provides a histogram of Doppler velocity data for this case. The standard deviation σ is large and the bimodal frequency distribution is not a good fit to a normal
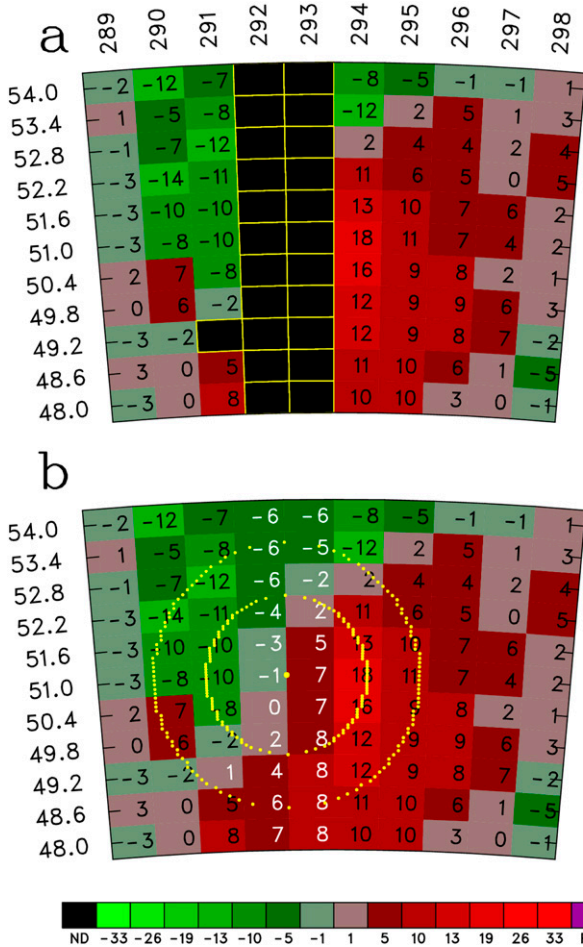
FIG. 7. As in Fig. 6, but (a) data along two radials passing through the mean Doppler velocity peaks at 292° and 293° azimuths have been voided (black pixels) and (b) the filled-in Doppler velocities (white numbers) along these two radials.

FIG. 8. As in Fig. 7, but the data along the 291° and 294° radials are voided instead.

distribution or any other standard probability distribution function. The most significant data for tornado warning purposes, the peak values, are two large standard deviations from the mean and are outliers arising not from chance but from the unresolved vortex. Hence, a statistical approach to data filling based on an assumed standard distribution is not useful. In contrast, our algorithm is independent of statistical assumption such as normality, stationarity, and isotropy.

We create data voids by deleting the original V_r data at some points. We then use the algorithm to fill in the missing data smoothly without altering any of the good data. To fit the definition of missing data, we should lose all memory of the deleted data because in real-life environments there is no way of validating filled-in data. However, in the current experimental framework, we compare the original prior-to-deleted data with filled-in fields to reveal the consequences of the variational procedure. With the Cartesian square grid as used in this section, the condition that $|\nabla V_r|$ is a minimum in each
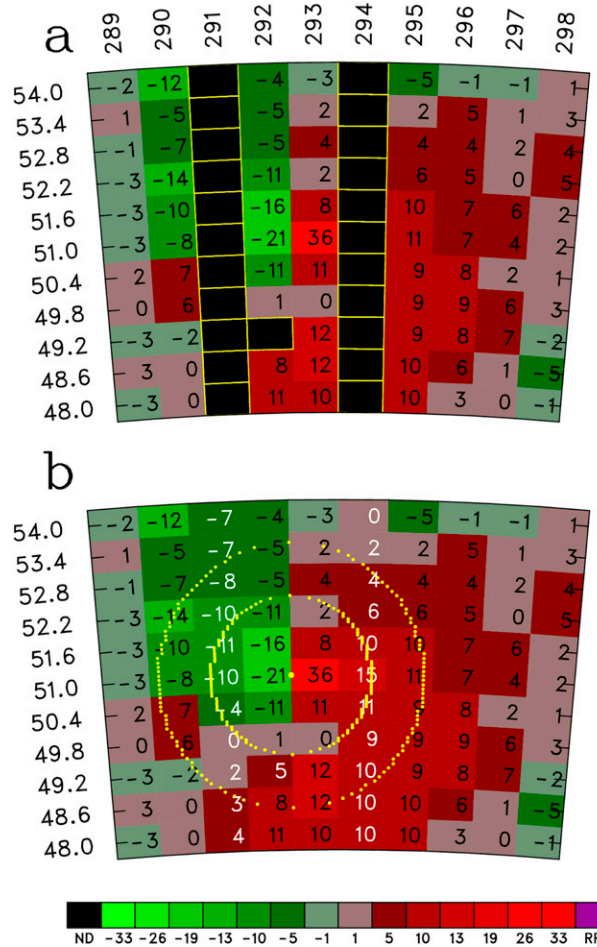
region devoid of data is that the Laplacian, $\nabla^2 V_r$, equals zero in each data void (Hildebrand 1965, p. 138) (i.e., at each grid point with missing data). The Laplacian is evaluated using a five-point stencil. It is given by

$$\nabla^2 V_r = \frac{(V_{i+1,k} + V_{i,k+1} + V_{i-1,k} + V_{i,k-1} - 4V_{i,k})}{\Delta^2}, \qquad (15)$$

where the subscript r in V_r is dropped for convenience, and Δ is the constant grid spacing. Thus, the condition that $\nabla^2 V_r = 0$ at the i, k grid point is satisfied if

$$4V_{i,k} \approx V_{i+1,k} + V_{i,k+1} + V_{i-1,k} + V_{i,k-1}. \qquad (16)$$

In other words, the filled-in value at missing data point ($V_{i,k}$) is equal to the average of values at the four nearest-neighbor points (i.e., $V_{i+1,k}$, $V_{i, k+1}$, $V_{i-1, k}$, and $V_{i,k-1}$). At boundary points with missing data (as will be shown subsequently), we need to satisfy the natural boundary condition that the normal gradient of V vanishes (i.e., $\partial V/\partial n = 0$). We do this through the device of adding adjacent exterior points. Note that by the maximum principle (Duff and Naylor 1966, p. 135), the
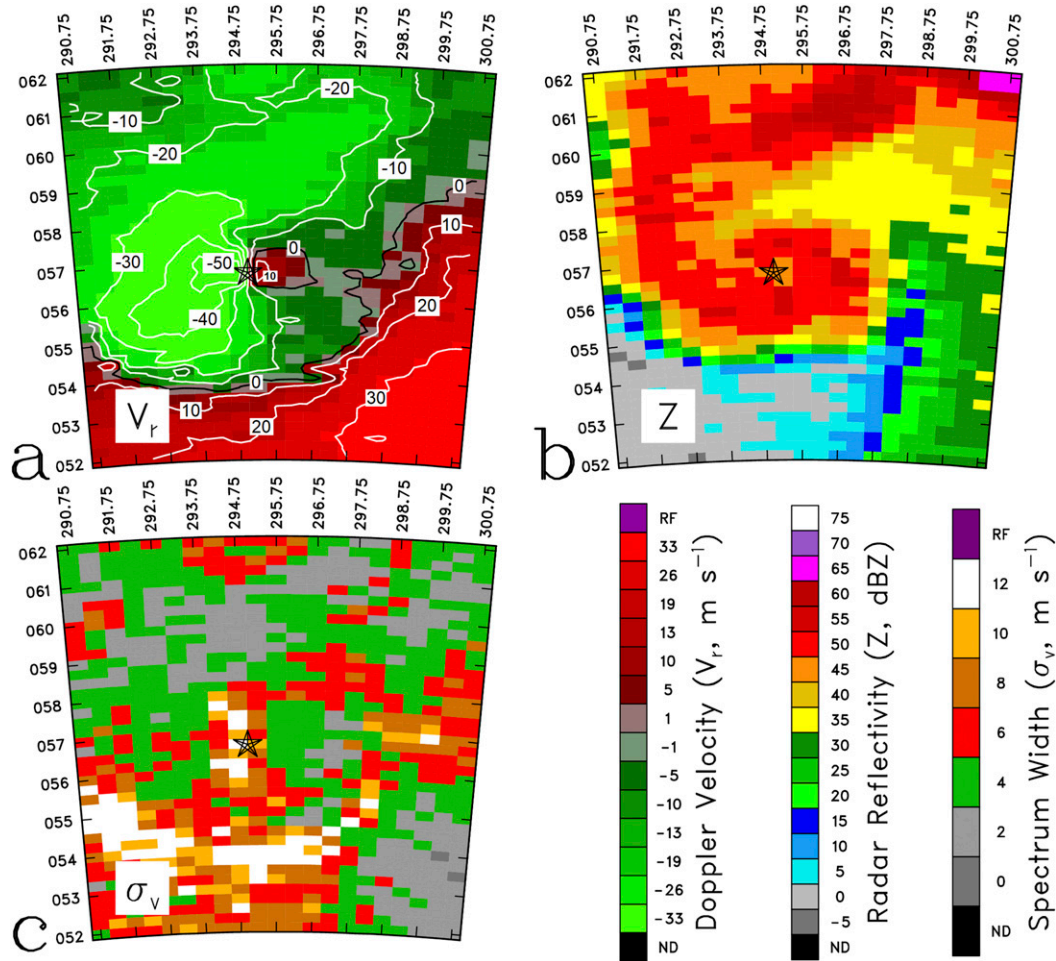
FIG. 9. KOUN scans of (a) ground-relative, mean Doppler velocity (V_r, m s^{-1}); (b) radar reflectivity (Z, dBZ); and (c) spectrum width (σ_v, m s^{-1}) of the El Reno tornado as collected at 0.97° launch angle at 2311 UTC 31 May 2013. The center of the significant TVS is located at 56.75 km and 295.0° (black star). The TVS's positive and negative Doppler velocity peaks, separated by about 0.8°, are, respectively, +14 and −58 m s^{-1}. In (a), the Doppler velocity is contoured at intervals of 10 m s^{-1}. The label bars indicate the V_r, σ_v, and Z contours with ND (black pixel) denoting no data (missing data) and RF (purple) range folding. The superresolution data have 0.25-km range and 0.5°-azimuth intervals. The labels on the left and top sides of the panels are ranges (km) and azimuths (°) from the KOUN radar.

maximum and minimum values in a data void are attained on its boundary. Therefore, the data filling does not introduce local maxima or minima into the field. In particular, the algorithm cannot reintroduce local maxima or minima that have been discarded from the original field and it cannot insert local maxima or minima that are not present in the original field. Other methods such as cubic spline interpolation can "overshoot" and amplify extrema. This is undesirable because we do not want interpolated (rather than actual) data to trigger, for example, the tornado detection algorithm (TDA; Mitchell et al. 1998) into false alarms. Too many false alarms lead to the public not taking tornado warnings seriously.

In our presentation of examples 1–7 (Figs. 4 and 5), we use the convention that locations with missing data are

shaded pink. The interior points with good data that are used to fill in missing data at the pink points are green points, the fake exterior points used to apply the natural boundary condition where necessary are yellow points, and the points that have no impact on the data filling are white points. We use seven illustrative examples (1–7) to show how the simple version of the variational algorithm works (Figs. 4 and 5).

a. Example 1

In the first example, we assume that the positive peak velocity in the TVS, the value 36 in Fig. 2, is missing (pink point in Fig. 4). The algorithm has no knowledge of the missing peak in the original field and so the algorithm cannot recreate it. The Laplacian is zero at the pink point so x, the filled-in value there, equals the average of the values at the four green
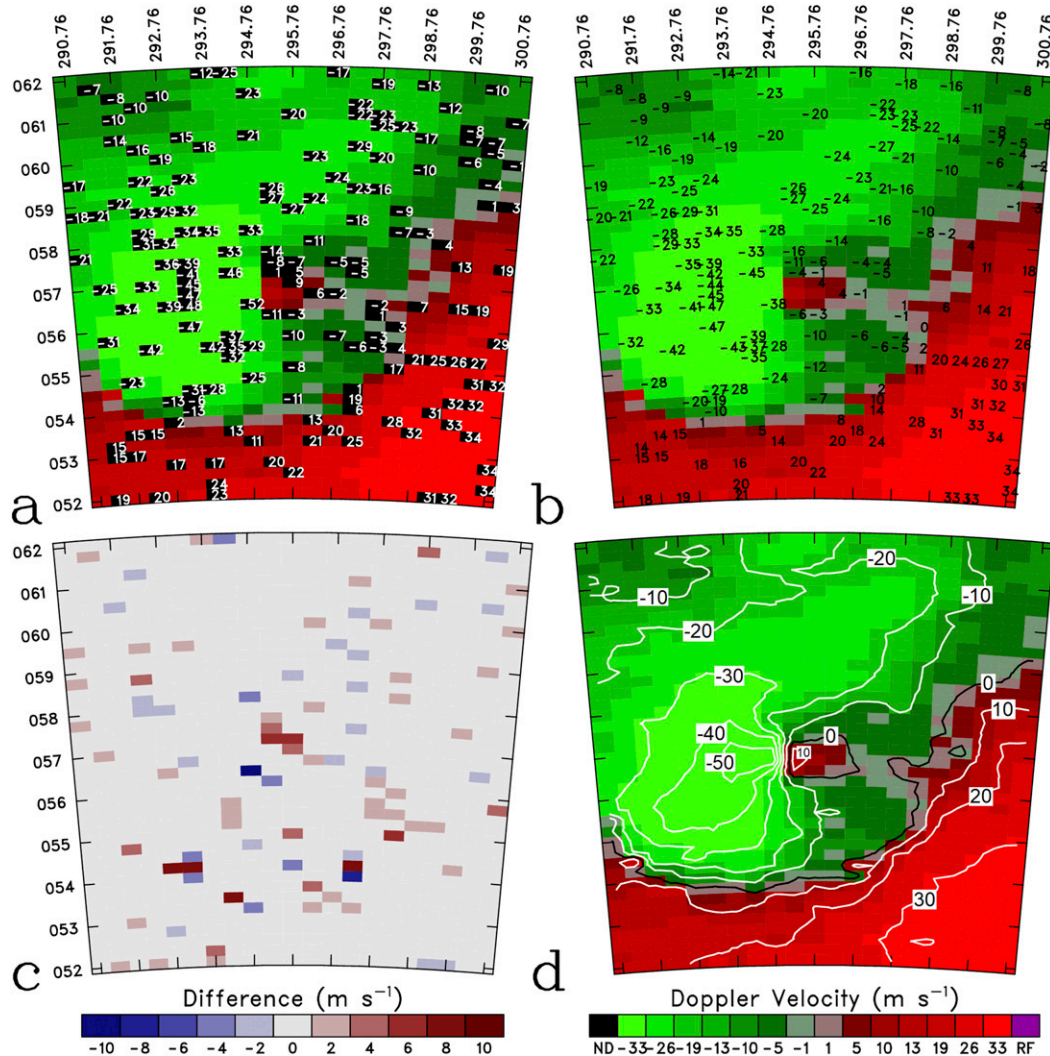
FIG. 10. (a) Missing Doppler velocity data created randomly by voiding data (superimposed by black pixels) based on the ϕ value of 0.2. (b) Filled-in Doppler velocity data provided by the algorithm. White numbers inside black pixels in (a) represent actual nonmissing data for comparison to filled-in Doppler velocity values (white numbers) in (b). (c) Doppler velocity differences [actual Doppler velocity values in (a) minus filled-in Doppler velocity values in (b)]. The difference scale is indicated at bottom of (c). (d) Contours of the filled-in Doppler velocity field. The V_r scale is indicated inside the label bar in (d).

nearest-neighbor points (Fig. 4) with values at the white points having no impact. Thus, x is equal to $(8 + 18 + 11 - 21)/4 = 4$. It is much different from the voided value 36, but that should be expected near a singularity. According to photogrammetric measurements (Golden and Purcell 1978), there is in real life a tornado with peak toward and away velocities of $80\,\mathrm{m\,s}^{-1}$ separated by roughly half the distance between grid points. A TVS is still evident in Fig. 2, but the deduced mean rotational velocity, $19\,\mathrm{m\,s}^{-1}$, is reduced and the peaks are further apart than in the original data.

b. Example 2

In example 2 (Fig. 4), we void the negative peak velocity of the TVS (i.e., the value -21 in Fig. 2) instead of the positive

one. In this case, the filled-in value is $-0.25 = (-16 + 36 - 11 - 10)/4$. In the filled-in field, there is still a TVS with the deduced mean rotational velocity of $23\,\mathrm{m\,s}^{-1}$.

c. Example 3

We now investigate the case where both the negative and positive peak velocities of the TVS are missing (the pink data void in example 3 of Fig. 4). Let x and y be the filled-in values for the left and right pink points, respectively. These depend mutually on each other and six surrounding green points. The conditions that $\nabla^2 V = 0$ at the pink points are $4x = -16 - 10 - 11 + y$ and $4y = 8 + 18 + 11 + x$. The solutions of these two simultaneous equations are $x = -7.4$ and $y = 7.4$. As expected, the velocity peaks are further apart than in the original data.
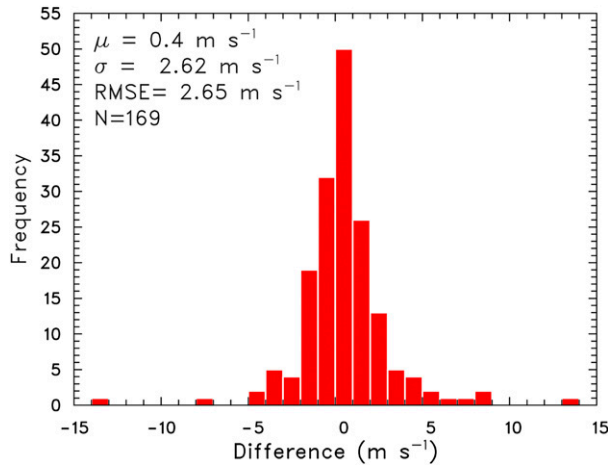
FIG. 11. Frequency histogram of the Doppler velocity differences (m s^{-1}) (actual Doppler velocity minus filled-in Doppler velocity) in 1 m s^{-1} velocity bins for the El Reno tornado as collected at the 0.97° launch angle at 2311 UTC 31 May 2013. The values μ, σ, RMSE, and N are the mean, standard deviation, root-mean-square error, and total number of data points, respectively.
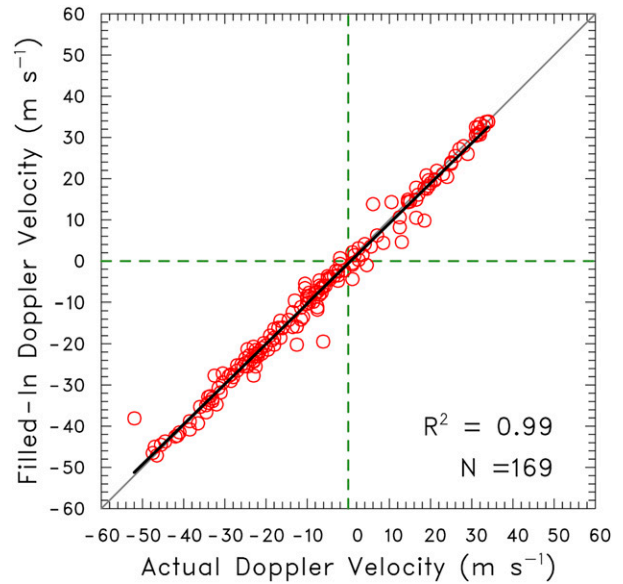
Thus, the filled-in field has a mesocyclone signature with a deduced mean rotational velocity of 14 m s^{-1} instead of a TVS. Since the good data outside the signature is unmodified, the large far-field circulation value, which indicates the possible or potential presence of a tornado (DJ20) is unaffected.

d. Example 4

In example 4 of Fig. 4 we void the data values −10 to the immediate left of the left peak and the value 18 to the immediate right of the right peak in Fig. 2. The data voids are separated far enough that the equations for the unknowns x and y are uncoupled and the voids can be filled in individually. Thus, $x = (−10 − 21 − 8 − 8)/4 = −11.75$ and $y = (13 + 11 + 16 + 36)/4 = 19$ depend on the leftmost and rightmost disjoint clusters of four green points, respectively. In this case filled-in values are close to the deleted values (Fig. 2). This is because the data voids are in velocity gradients that are varying slowly.

e. Example 5

Example 5 (Fig. 5) shows how to fill in missing data at an isolated pink point that is on a boundary but not at a corner. In this case, we have voided the value 7 on the left side of the domain (Fig. 2). We use the natural boundary condition by inserting the exterior yellow point as shown and assigning it a value of V such that the normal gradient of V vanishes at the pink point. Giving the unknown at the yellow point, the circled value at the interior green point across the boundary (viz., −8) makes $\partial V/\partial n = 0$. The filled-in value at the pink boundary point is thus $−4.5 = (−8 − 8 + 6 − 8)/4$. In this case, the filled-in value −4.5 is not close to the discarded value 7 because the normal gradient of V at the pink point is

FIG. 12. Scatterplots (red circles) of actual vs filled-in Doppler velocity data for the El Reno tornado as collected at the 0.97° launch angle at 2311 UTC 31 May 2013. R^2 is the correlation coefficient squared. The gray diagonal is the straight line through the origin of slope 1.0. The black line is the linear regression line.

significant (Fig. 2) and not zero as specified by the natural boundary condition.

f. Example 6

Example 6 (Fig. 5) illustrates how to fill in missing data at an isolated pink point at a corner point of the domain. Here, two exterior points are needed so that ∇V at the pink point vanishes in both the radial and azimuthal directions. The exterior points are assigned the values of the associated adjacent interior points as shown. Hence, the filled-in value at the pink corner point is $10.5 = (10 + 11 + 10 + 11)/4$.

g. Example 7

The final example demonstrates the filling-in of a missing radial, in particular, the pink radial that passes through the peak value (36) of the original field (Fig. 2). We need two exterior yellow points as shown in example 7 of Fig. 5. To satisfy the natural boundary condition ($\partial V/\partial n = 0$) at the endpoints of the radial, the boxed values must be the same and likewise for the circled values. We find the filled-in values shown at the pink points as the solutions of a system of five simultaneous equations (x, y, z, a, and b), namely, the finite-difference version of Laplace's equation at the five pink points. Applying this rule at pink points in succession starting at either top or bottom of the domain, the following five equations are given by
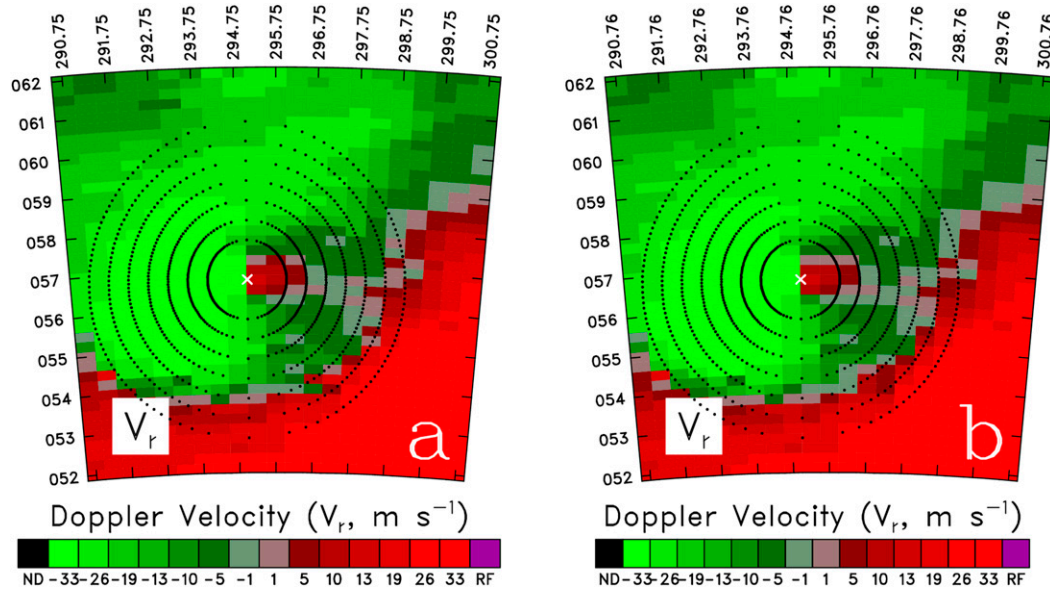
$$4x = y + 11 + y − 11,$$

$$4y = x + z + 13 − 16,$$

FIG. 13. KOUN WSR-88D scans of (a) actual and (b) filled-in, ground-relative, mean Doppler velocity V_r (m s^{-1}) at launch angle of 0.97° for the El Reno tornado at 2311 UTC 31 May 2013. Black dotted curves are concentric circles of radii $\rho = 1.0, 1.5, 2.0, 2.5, 3.0, 3.5,$ and 4.0 km centered on white ×, the estimated center of the TVS. The height of the center is 1.17 km AGL.

$$4z = y + 18 + a - 21,$$
$$4a = z + 16 + b - 11, \quad \text{and}$$
$$4b = a + 12 + a + 1.$$

As a result, the solutions of the five simultaneous equations, as shown in example 7 and its solution, are $x \approx -0.49$, $y \approx -0.98$, $z \approx -0.44$, $a \approx 2.23$, and $b \approx 4.37$. These equations depend on the good, unmodified data at the green points along the neighboring radials but not on the data at the white points. As in example 1, there is a TVS with a deduced mean rotational velocity of 19.5 m s^{-1} in the filled-in field.

We now indicate how the method is adapted to actual Doppler radar data on a surface of constant elevation angle α_o. The natural boundary condition (where needed) is still $\partial V / \partial n = 0$. Since the grid is not Cartesian and the grid spacing is different in the radial and azimuthal directions, the correct equation to apply in data voids is (4), not $\nabla^2 V = 0$. At pink points where the data are missing, the centered finite-difference version of (4) is used. At pink boundary points with missing data, we once more employ exterior yellow points to satisfy the natural boundary condition. At green and white points where the data are good, V is again set equal to its observed value. The maximum principle still applies.

Let N be the total number of interior and boundary points in the domain, M be the number of pink points at which data are missing, and E be the number of needed exterior yellow points. We now have a system of $N + E$ simultaneous equations in $M + E$ unknowns. Note that $N - M$ of these equations apply to green and white points with good data;

they simply set the value of V there to the observed value. A further M equations consist of the finite-difference form of (4) applied at the M pink points where data are missing and filled-in values need to be found. The remaining E equations equate V at exterior yellow points to V at adjacent interior points to satisfy the natural boundary condition (as in examples 5–7). We easily eliminate the E quantities at the exterior points, thereby reducing the system to N equations with $N - M$ known values and M unknowns.

The algorithm fills in all sized and shaped data voids in the domain. It performs two minor preliminary steps to speed up the algorithm. First, it searches for isolated pink points and fills in values there [as done similarly in examples 1, 2, 5, and 6 but using (4) instead of $\nabla^2 V = 0$]. Second, it does the same for isolated pairs of adjacent pink points (as done similarly in examples 3 and 4). After these steps, we have a modified system of N equations with less unknowns. The algorithm fills in the remaining data voids simultaneously by numerically solving the modified system (as done similarly in example 7).

4. Case studies

Our main interest in filling in data is for number crunching in a tornado-warning algorithm. Dealing with missing data (i.e., null values) is easiest done by filling in values in a smooth, neutral, uncomplicated way. Our aim is successful warnings, not best fits. In this regard we need to quantify tornado threat with and without missing data. This requires a measure of tornado threat that is quite robust to missing data and not reliant on the presence or absence

of a few data points. Scoring in the current TDA (Mitchell et al. 1998, p. 360) is based on maximum gate-to-gate velocity differences on surfaces of constant launch angle (as well as some signature height and extent criteria). Voiding a peak value on a surface has a large effect on the score. A more suitable metric than the one used by the TDA is Doppler circulation because it incorporates the more abundant far-field data (DJ20), which is partly missing at worst, instead of a localized TVS measure, for which the data may be entirely missing. Large circulation is the clue that there is an actual or potential tornado and circulation values are very high for large devastating tornadoes. Compared to circulation, gate-to-gate velocity differences are more range dependent, more dependent on tornado location within a grid cell, and more vulnerable to missing data. For our narrow requirements, a successful filling algorithm is one that retains most of the tornado threat (high circulation value) even when a lot of the data is missing.

A minor consideration is the filling of fields with "missing pixels" to produce display images that are pleasing to the eye. To this end, we have in a few examples randomly voided 20% of the data, filled in the values by the variational algorithm, and computed statistics for the differences between voided and filled values.

As examples of the "data filling" algorithm, we present analyzed actual Doppler velocity data of four tornadoes: (i) the Union City, Oklahoma, tornado of 24 May 1973; (ii) the El Reno, Oklahoma, tornado of 31 May 2013; (iii) the Binger, Oklahoma, tornado of 22 May 1981; and (iv) the Dodge City, Kansas, tornado of 24 May 2016. In some of these cases, data were missing originally. In others, we voided data either along radials or randomly.

As is the case with other algorithms (e.g., Mahalik et al. 2019), we assume that actual reflectivity data have been quality-controlled to eliminate the nonmeteorological radar echoes such as biological returns, anomalous propagation, instrument artifacts, and ground clutter (e.g., Lakshmanan et al. 2014) and that actual Doppler radial velocity data presented in this study have been carefully dealiased.

a. The Union City tornado of 24 May 1973

Doppler velocity data were collected by the Norman Doppler radar in the surface of 3.8° launch angle at 1546 central standard time (CST) during the mature stage of the violent tornado that struck Union City on 24 May 1973 (Lemon et al. 1978; Brown et al. 1978). This is presented in B scan (range–azimuth format) in Fig. 6a for a larger domain than in Fig. 2. The algorithm replaces two missing data (e.g., 999 in Fig. 6a) with filled-in Doppler velocity values (2 and 5 m s^{-1} in Fig. 6b). DJ20 computed Doppler circulation values of 56.96×10^3 and 53.04×10^3 m^2 s^{-1} and mean Doppler convergence values of 2.250×10^{-3} s^{-1} and 2.884×10^{-3} s^{-1} at the circle radii of 1.5 and 2.5 km, respectively (yellow dotted circles). After doubling to compensate for the unobserved wind components being set to zero, the circulation value (-1.1×10^5 m^2 s^{-1}) at
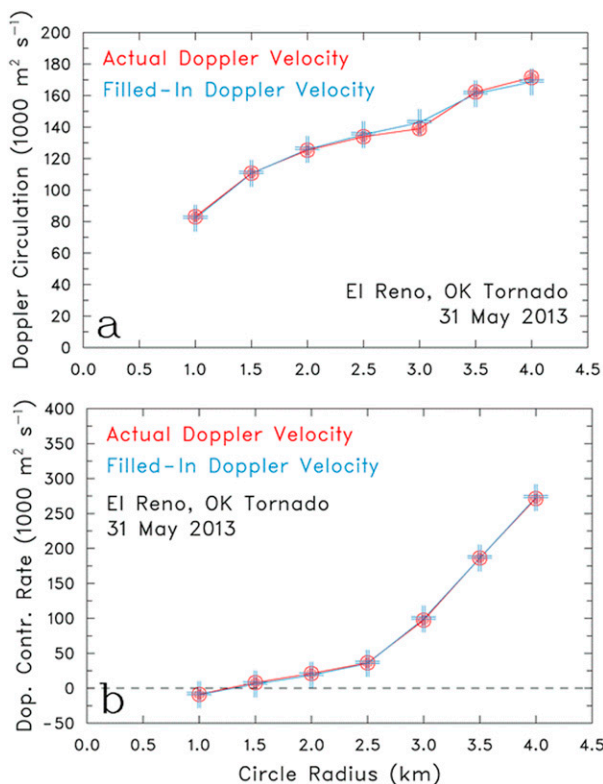

FIG. 14. (a) Doppler circulations and (b) Doppler areal contraction rates as a function of circle radius for the El Reno tornado of 31 May 2013. The circles are centered on the intersection of the TVS axis with the 0.97° surface of constant launch angle. The red dots (blue pluses) are the values calculated from the actual (filled in) Doppler velocity field depicted in Fig. 13a (Fig. 13b).

3.5 km AGL agreed closely with the photogrammetrically measured value near the ground. The results indicate that the mature tornado was embedded in a region of nearly uniform strong convergence ($\sim 5.5 \times 10^{-3}$ s^{-1} after doubling) that was about 6 km in diameter without a mesocyclone present at the time.

We artificially created missing data by voiding two contiguous radials of data that pass through the extreme positive and negative Doppler velocity values in the significant TVS region at 292° and 293° azimuths (Fig. 7a). This situation bears a resemblance to missing data extending radially from the KOUN WSR-88D radar shown in Fig. 1e. The deduced mean rotational velocity is now 16 m s^{-1} and the filled-in Doppler velocity field resembles that of a convergent mesocyclone signature instead of the TVS. Doppler circulation values at the circle radii of 1.5 and 2.5 km, respectively, are 55.43×10^3 and 51.98×10^3 m^2 s^{-1}; Doppler mean convergence values at the corresponding radii are 2.461×10^{-3} and 3.017×10^{-3} s^{-1}, respectively. Even though we have eliminated the TVS, high values of circulation and convergence still indicate a significant tornado potential.
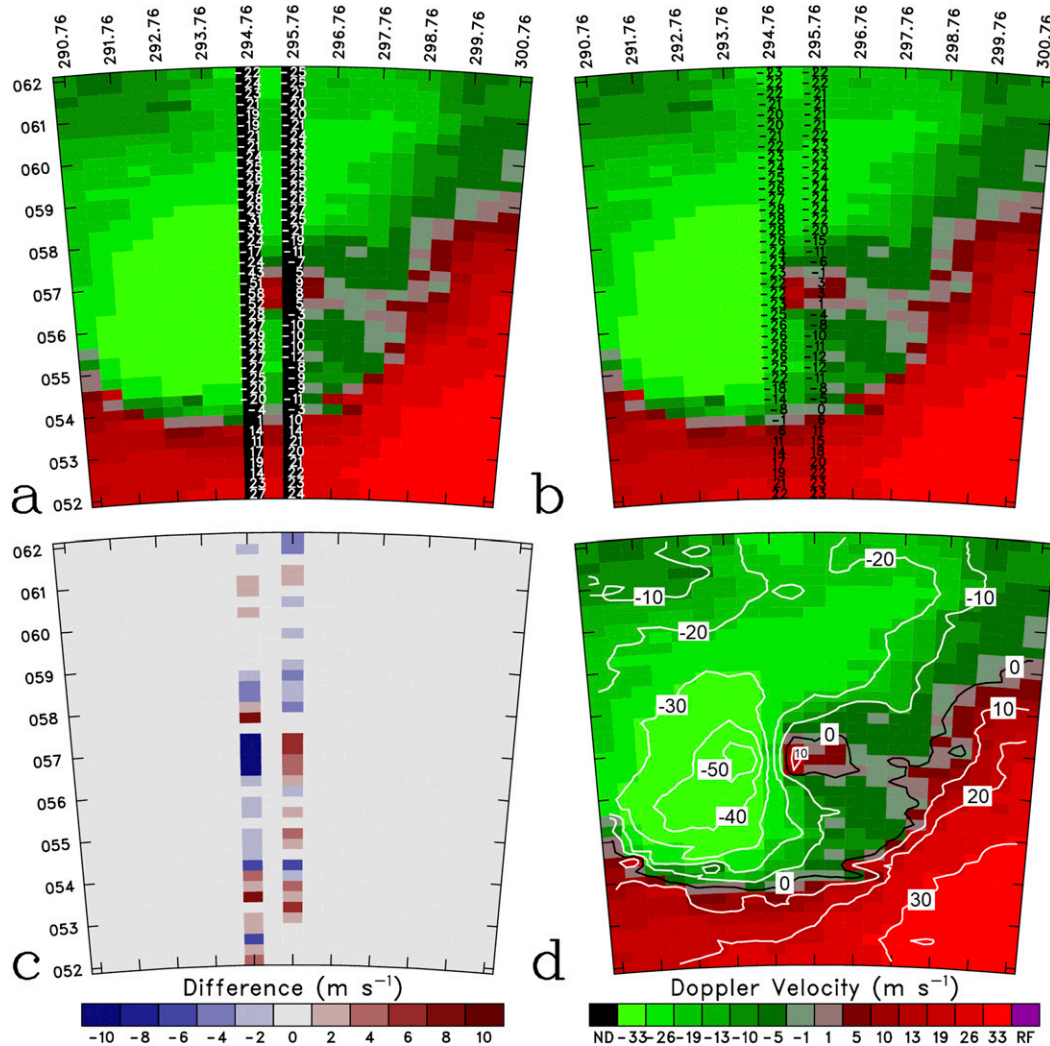
FIG. 15. As in Fig. 10, except that the black bins of voided data in (a) are along azimuths of 294.76°
and 295.76° that pass through the Doppler velocity peaks of the vortex signature. The filled-in values
of Doppler velocity are shown in (b) and (d). Doppler velocity contours with intervals of $10\,\mathrm{m\,s^{-1}}$ are
drawn in (d).

We artificially voided two separate radials of missing data
at azimuth 291° and 294° (Fig. 8). These radials resemble the
situation in Fig. 1e. The rotational velocity is $28.5\,\mathrm{m\,s^{-1}}$ and the
signature is a TVS as in the original, actual data (Fig. 6).
Doppler circulation values at the circle radii of 1.5 and 2.5 km,
respectively, are 45.61×10^3 and $53.88 \times 10^3\,\mathrm{m^2\,s^{-1}}$; Doppler
mean convergence values at the corresponding radii, re-
spectively, are 2.487×10^{-3} and $2.292 \times 10^{-3}\,\mathrm{s^{-1}}$ in good
agreement with the original values. The measures of a strong
convergent vortex remain high even when two radials are
missing.

Although the tornado threat is not masked when data
along two radials is filled variationally, this is no longer the
case when the number of adjacent, missing radials per gap is
large. As shown in Figs. 1e and 1f in another case, a non-
tornadic mesocyclone signature with an associated hook

echo, centered at an approximate 36-km range and 262°
azimuth, was heavily masked by 10 missing radials. The ra-
dials were caused by a processing error when the scanning
KOUN antenna "bumped" for a few azimuths (C. Kuster
2019, personal communication). These gaps are too large to
fill realistically.

b. The El Reno tornado of 31 May 2013

We now explore the statistical differences between origi-
nal and the filled-in fields utilizing a histogram and a lin-
ear regression model. Figure 9 presents KOUN WSR-88D
scans of ground-relative base Doppler velocity (V_r), base
reflectivity (Z) of and base spectrum width (σ_v) of the violent
El Reno tornado as collected with superresolution (0.5° azi-
muthal interval and 250-m range increment) at the 0.97°
launch angle at 2311:04 UTC 31 May 2013. Tornadogenesis,

tornado evolution, photogrammetric and polarimetric analyses, and aerial damage survey in the El Reno tornadic supercell have been extensively documented by Wakimoto et al. (2015, 2016) and Bluestein et al. (2018, 2019). We chose a large domain of original, good data in the α surface (launch angle) of 0.97° bounded by slant-range circles $r_1 =$ 52 km and $r_I =$ 62 km and azimuthal radials $\beta_1 =$ 290.75° and $\beta_K =$ 300.75°, with the subscripts $I =$ 41 and $K =$ 21 (861 grid points).

Associated with the large tornado is a strong TVS embedded in a strong convergent mesocyclone signature at low altitudes (Fig. 9a). This Doppler velocity pattern bears a resemblance to that at the low-altitude scan (0.5°) from the KTLX WSR-88D at 2314:52 UTC (see Fig. 2b of Wakimoto et al. 2015, for example). The extreme positive and negative Doppler velocity values of the TVS at a range of 56.75 km, respectively, are +14 and -58 m s^{-1}, resulting in a deduced mean rotational velocity of 36 m s^{-1}. The approximate center azimuth of the TVS is 294.9°—essentially the same as the center of the reflectivity minimum, which results from centrifuged radar targets (e.g., Dowell et al. 2005; Wood et al. 2009). The estimated center of the TVS at 2311:04 UTC was very close to the tornado track at 2310:00 and 2312:08 UTC (see Figs. 1 and 7 of Wakimoto et al. 2016).

At low altitudes, a significant gust front signature, as indicated by a curved, zero Doppler velocity band (Fig. 9a), is associated with a well-pronounced hook echo (Fig. 9b). Spectrum widths (Fig. 9c) are high along the gust front and around the TVS. These large values may indicate either strong turbulence or strong winds. The flow near the tornado can be smooth, owing to the rotation damping turbulence, but still have considerable variability within a sampling volume.

To create randomly black data voids (speckles) in the field of good Doppler velocity data (Fig. 10a), we used $\phi = 0.2$, which signifies that the data has been randomly removed at about 20% of the data points (169 out of 861). The resulting Doppler velocity field is shown in Fig. 10a. We then filled in the Doppler velocity field variationally using the natural boundary condition as needed. This filled-in field is depicted in Fig. 10b. Figure 10c presents actual minus filled-in Doppler velocities at the grid points where data have been filled in. Large (small) differences between actual and filled-in values highlight regions where the actual data has a lot of low-frequency (high frequency) variability. Figure 11 presents a frequency histogram of these differences. The mean value of the difference, 0.4 m s^{-1}, is insignificant. The standard deviation and RMSE values (2.62 and 2.65 m s^{-1}, respectively) indicate that the differences between the voided and filled-in data are generally small.

Linear regression (Fig. 12) has a R^2 value of 0.99. Despite the few outliers shown in Figs. 10 and 11, the algorithm typically fills in data gaps well when 20% of the data is randomly voided.

We calculated Doppler circulations and Doppler areal contraction rates (approximately one-half the real values) for the circles shown in Fig. 13 by the methods described in DJ20.
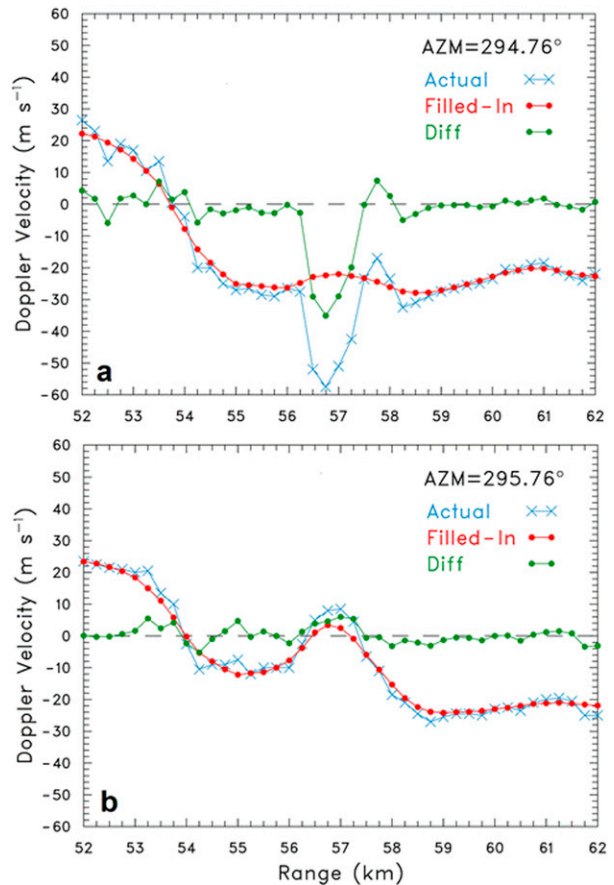

FIG. 16. Radials of actual Doppler velocity (blue curve with × markers), filled-in Doppler velocity (red dotted curve), and Doppler velocity difference (green dotted curve) at azimuths of (a) 294.76° and (b) 295.76° for the El Reno tornado of 31 May 2013.

These are shown in Figs. 14a and 14b. Randomly voiding 20% of the data and then filling in the missing data via the algorithm has very little effect on the Doppler circulation and the areal contraction rate.

Instead of randomly annulling data, we arbitrarily voided two radials of missing data that pass through the peak positive and negative Doppler velocity values of the TVS (Fig. 15a) as a worst-case scenario. The filled-in Doppler velocities, via the variational procedure, and the differences between the actual and filled-in Doppler velocity values are shown in Figs. 15b, 15c, 16, and 17. Large velocity differences occur near the inbound peak of the TVS along the 294.76° azimuth at around 57 km range. These differences are associated with a small-scale band of damaging winds on the rear side of the TVS. This local maximum in the actual inbound velocity is naturally missing in the filled-in field because the algorithm is designed to patch data voids smoothly without manufacturing small-scale peaks. Consequently, the patched data do not fit the actual data along the 294.76° azimuth very well (Figs. 16a and 18a) with the coefficient of determination R^2 being only 0.8. As
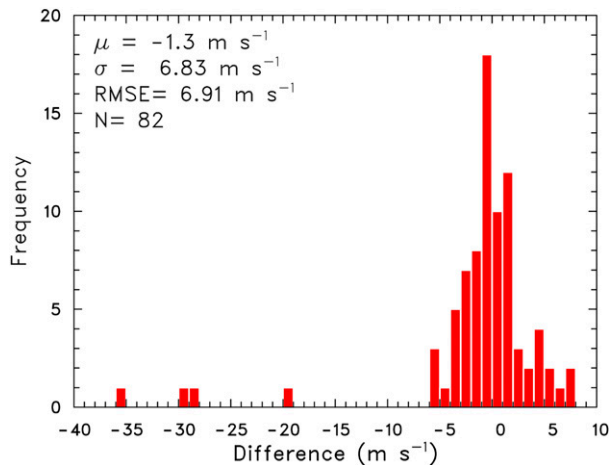
FIG. 17. As in Fig. 11, but for combined data along (a) 294.76° and (b) 295.76° azimuths shown in Fig. 16.

expected, the mean, standard deviation, and root-mean-square of the differences are quite high in this situation (Fig. 17). The variation of the Doppler velocity along the 295.76° azimuth through the outbound peak is smaller. Consequently, the difference between voided and filled-in values is smaller (Fig. 16b) and the correlation between the two sets of Doppler velocities is higher ($R^2 = 0.98$; Fig. 18b).

c. The Binger tornado of 22 May 1981

We now present an example where data are missing along the domain boundary as well as along a radial. Zrnić et al. (1985) investigated automatic detection by the Norman Doppler radar of strong azimuthal shears in association with large intense tornadoes. Figure 19a is almost identical to their Fig. 8, which displayed a B scan of mean Doppler velocities of the Binger tornado of 22 May 1981. As seen in Fig. 19a, there are many points with missing data on the domain boundary where the natural boundary condition (i.e., $\partial V_r / \partial n = 0$) is needed as in examples 5–7 (Fig. 5). These points include one corner boundary point as in example 6. However, Zrnić et al. (1985) did not show the radial of missing data at 281.20° azimuth between the 280.70° and 281.70° azimuths. At the closest point, the 281.20° azimuth is about 2 km to the rear of the TVS axis. Using the variational procedure, we filled in data along this azimuth and in some smaller data voids (Fig. 19b). Contours of the patched field look reasonable with no obvious discontinuity at 281.20°.

From the patched Doppler velocity field (Fig. 19b), we progressively computed Doppler circulations (about one-half the actual circulations) of 0.70×10^5 to 1.48×10^5 m^2 s^{-1} around circles (Fig. 20) of 0.5 through 2.0 km radius, respectively, centered on the TVS axis at a height of 1.3 km AGL. These large circulations are consistent with the TVS's strong, measured mean rotational velocity (48 m s^{-1}). The Doppler circulation decreases with decreasing circle radius, indicating an intense mesocyclone, whose presence is not
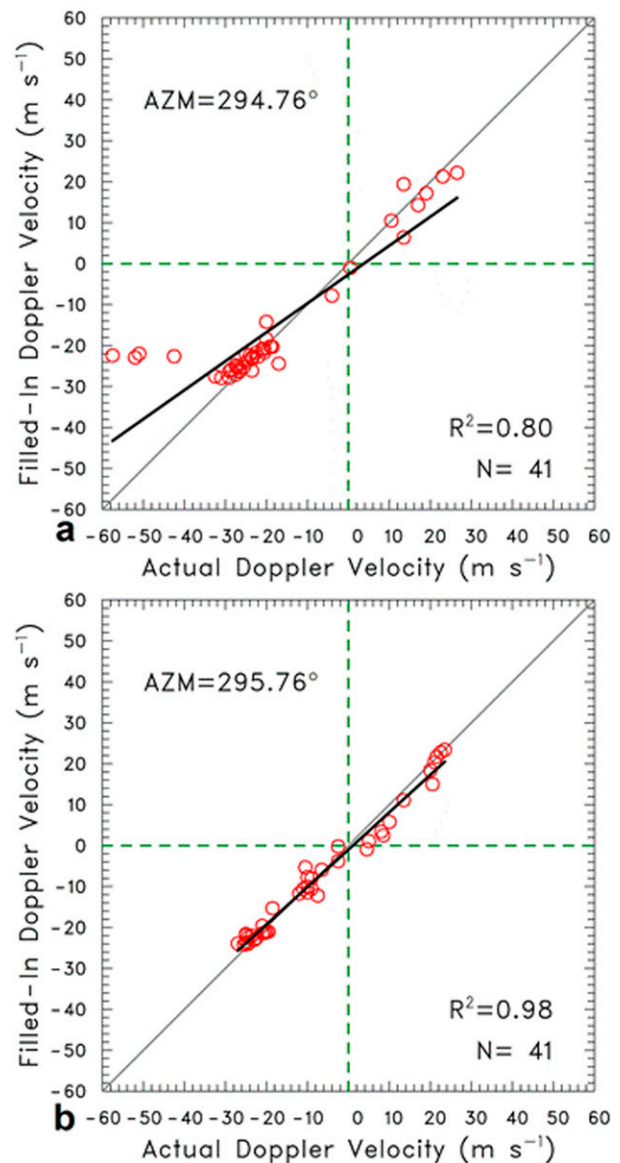

FIG. 18. As in Fig. 12, but for data along (a) 294.76° and (b) 295.76° azimuths in the case of the El Reno tornado at 2311 UTC 31 May 2013.

masked by the filled in data. The corresponding measured Doppler areal contraction rates (DJ20), -0.27×10^4 and 1.55×10^4 m^2 s^{-1} around of circles of 0.5 and 2.0 km radius, are probably small as a result of a great deal of debris centrifuging masking convergent airflow (e.g., Dowell et al. 2005; Wood et al. 2009).

d. The Dodge City tornado 8 of 24 May 2016

As a final example, we used Dodge City (DDC) WSR-88D scans of ground-relative base Doppler velocity (V_r) and base reflectivity (Z) of the Dodge City tornadic storm on 24 May 2016 (Wakimoto et al. 2018). The data were collected at the 0.5° launch angle at 2345 UTC (Figs. 1a,b). At this
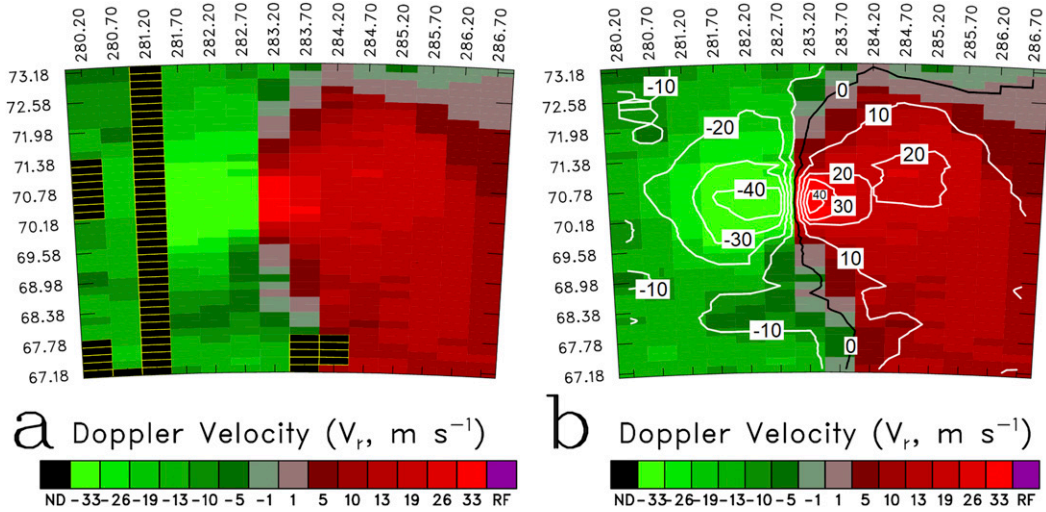
FIG. 19. (a) Actual, mean Doppler velocities (V_r, m s^{-1}) with black data voids and (b) filled-in Doppler velocities (V_r, m s^{-1}) in the Norman Doppler radar data coverage at 0.8° launch angle of the Binger tornado at 1909–1910 CST 22 May 1981. In (a), black rectangles represent missing Doppler velocity values that the algorithm has filled in (b). The significant TVS center is at 70.8 km and 283° from the radar and at the height of 1.31 km. (b) Contours of the filled-in Doppler velocity field with a contour interval of 10 m s^{-1}. Range and azimuth increments, respectively, are 150 m and 0.5°. Doppler velocity scale is indicated at bottom. ND stands for no data (missing data); RF stands for range folding.

time there was a dissipating EF0 tornado (#8 in Fig. 2 of Wakimoto et al. 2018). Within the mesocyclone region, there is a lot of missing data (Figs. 21a,b). The data could result in mesocyclone detection failure. Forecasters would appreciate any missing data being filled in, especially if the mesocyclone was moving close to the radar in a populated area (C. Kuster 2020, personal communication). Such data quality issues cause considerable problems for forecasters making tornado warning decision. The variational procedure produces realistic-looking fields of filled-in Doppler velocity and radar reflectivity, respectively (Figs. 21c,d).

Calculated Doppler circulations and Doppler areal expansion rates for different sized circles in the 0.5° constant-launch-angle surface are presented in Fig. 22. The filled-in data supplied by the algorithm produce useful smooth radial profiles of Doppler circulations and Doppler areal expansion rates. These profiles indicated a strong mesocyclone in a wide updraft.

5. Conclusions and recommendations

We have presented a variational, gap-filling algorithm, which minimizes the gradients of Doppler velocity. It fills in
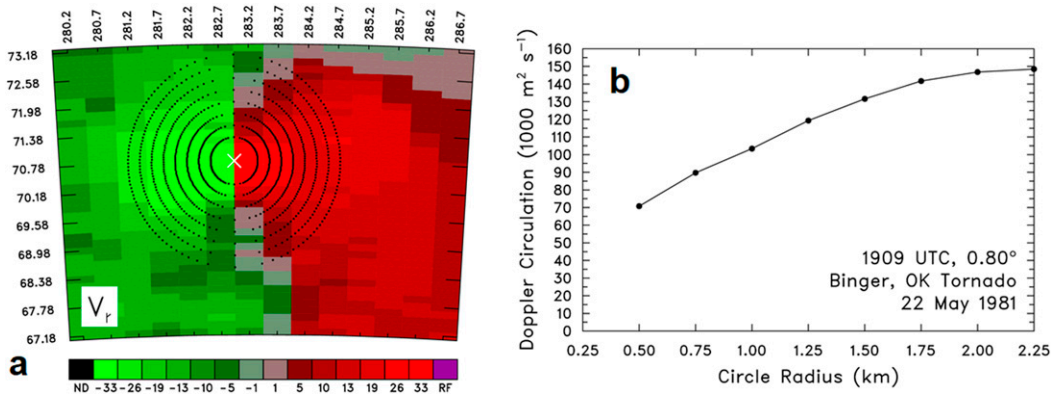

FIG. 20. (a) Norman Doppler radar scan of filled-in, ground-relative, mean Doppler velocity (V_r, m s^{-1}) at launch angle of 0.8° at 1909 CST (as in Fig. 19b). Black dotted curves are concentric circles of radii ρ = 0.5, 0.75, 1.0, 1.25, 1.5, 1.75, 2.0, and 2.25 km centered on white $\times$, the estimated center of Doppler velocity signatures. The height of the center is 1.31 km AGL. (b) Doppler circulations as a function of circle radius for the Binger tornado of 22 May 1981 centered on the mesocyclone axis.
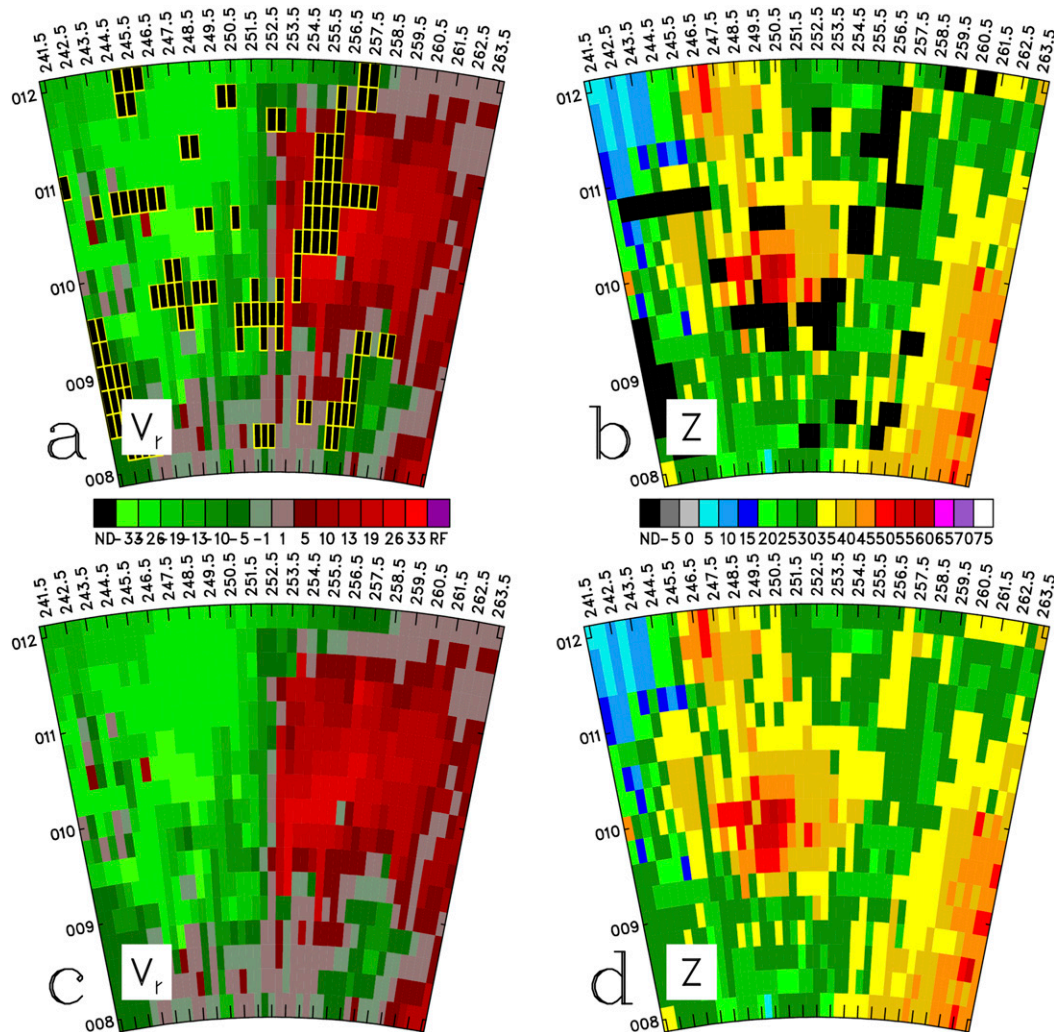
Fig. 21. DDC WSR-88D scans at 0.5° launch angle and at ~2345 UTC of (a) actual, ground-relative, mean Doppler velocity (V_r, m s^{-1}) field with black data voids and (b) actual radar reflectivity (Z, dBZ) field with black data voids of the Dodge City tornadic storm on 24 May 2016. In (a) and (b), each black rectangle represents a range gate of missing value that has been filled in via the algorithm in (c) and (d).

missing Doppler velocity data in data voids smoothly while leaving good data unmodified. We used single-Doppler data of four tornadoes to explore and assess the utility of the variational procedure. In tests in which Doppler velocity data are artificially voided, the algorithm performs well in regions where the field varies smoothly, and, as expected for any method, achieves poorly near artificially eliminated singularities such as TVSs. The gap-filling algorithm is not a data panacea (e.g., Trapp and Doswell 2000) because it is unable to put back small-scale peaks (such as a TVS). On the other hand, it is incapable of introducing a peak into the data and so will not create a false TVS. Data voids, such as two missing radials or 20% randomly distributed missing data points, slightly affect far-field tornado warning parameters such as Doppler circulation without concealing the threat.

Since our algorithm is stable and always converges to a unique solution rapidly in actual time, in principle it can be automated and incorporated in the NEXRAD system. An automated method for depicting mesocyclone track positions can be implemented in the MDA and TDA. Based on the past track positions, a future track position at the beginning of WSR-88D's next volume scan can be estimated. The automated algorithm can define an estimated center position before defining beginning and ending ranges and azimuths of a domain (say, 10 km × 10 km). The algorithm then calls to "search" any missing data parameter entirely within the domain. If missing data are detected, then the automated algorithm calls the variational method to fill in data. If there are no missing data within the domain, the variational algorithm gets bypassed, and the automated MDA and TDA continue.
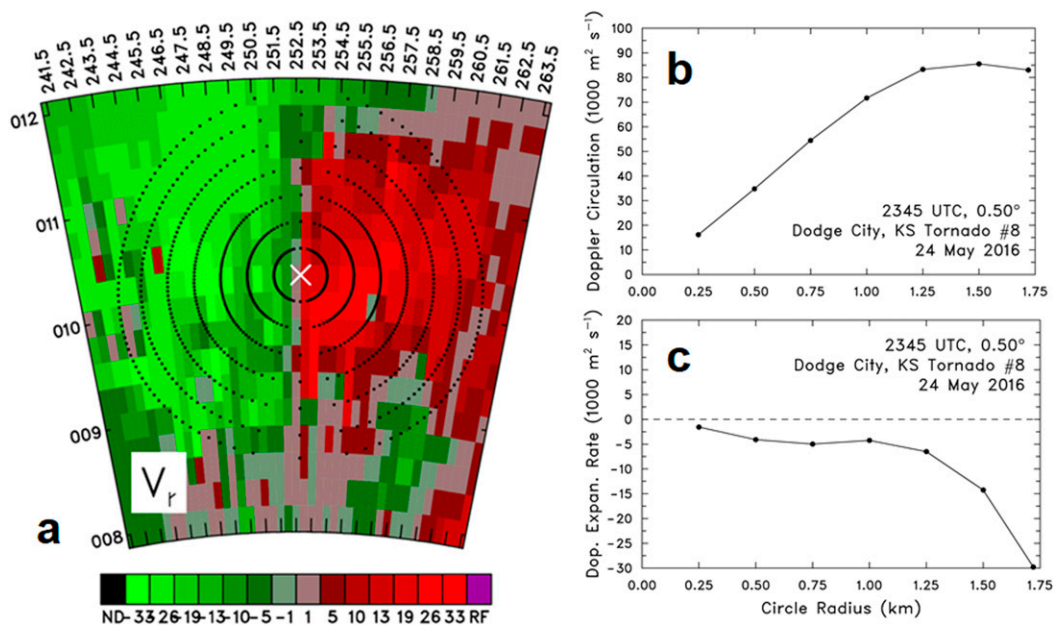
FIG. 22. (a) DDC WSR-88D scan of filled-in, ground-relative, mean Doppler velocity V_r (m s^{-1}) at launch angle of 0.5° at 2345 UTC (as in Fig. 21c). Black dotted curves are concentric circles of radii ρ = 0.25, 0.5, 0.75, 1.0, 1.25, 1.5, and 1.72 km centered on white ×, which is the estimated center of Doppler velocity signatures. The height of the center is 97 m AGL. (b) Doppler circulations (m^2 s^{-1}) and (c) Doppler areal expansion rates (m^2 s^{-1}) as a function of circle radius for the Dodge City tornadic storm on 24 May 2016.

REFERENCES

Bluestein, H. B., K. J. Thiem, J. C. Snyder, and J. B. Houser, 2018: The multiple-vortex structure of the El Reno, Oklahoma, tornado on 31 May 2013. *Mon. Wea. Rev.*, **146**, 2483–2502, https://doi.org/10.1175/MWR-D-18-0073.1.

——, ——, ——, and ——, 2019: Tornadogenesis and early tornado evolution in the El Reno, Oklahoma, supercell on 31 May 2013. *Mon. Wea. Rev.*, **147**, 2045–2066, https://doi.org/10.1175/MWR-D-18-0338.1.

Brown, R. A., L. R. Lemon, and D. W. Burgess, 1978: Tornado detection by pulsed Doppler radar. *Mon. Wea. Rev.*, **106**, 29–38, https://doi.org/10.1175/1520-0493(1978)106<0029:TDBPDR>2.0.CO;2.

Davies-Jones, R., V. T. Wood, and M. A. Askelson, 2019: Ray curvature on a flat earth for computing virtual WSR-88D signatures of simulated supercell storms. *Mon. Wea. Rev.*, **147**, 1065–1075, https://doi.org/10.1175/MWR-D-18-0356.1.

——, ——, and E. N. Rasmussen, 2020: Doppler circulation as a fairly range-insensitive far-field tornado detection and precursor parameter. *J. Atmos. Oceanic Technol.*, **37**, 1117–1133, https://doi.org/10.1175/JTECH-D-19-0116.1.

Dowell, D. C., C. R. Alexander, J. M. Wurman, and L. J. Wicker, 2005: Centrifuging of hydrometeors and debris in tornadoes: Radar-reflectivity patterns and wind-measurements errors. *Mon. Wea. Rev.*, **133**, 1501–1524, https://doi.org/10.1175/MWR2934.1.

Duff, G. F. D., and D. Naylor, 1966: *Differential Equations of Applied Mathematics.* Wiley, 423 pp.

Ellis, S., 1997: Hole-filling data voids in meteorological fields. M.S. thesis, School of Meteorology, University of Oklahoma, 198 pp. [Available from National Weather Center Library, 120 David L. Boren, Blvd., Room 4300, Norman, OK 73072.]

Farlow, S. J., 1993: *Partial Differential Equations for Scientists and Engineers.* Dover, 414 pp.

Franke, R., 1982: Scattered data interpolation: Tests of some methods. *Math. Comput.*, **38**, 181–200, https://doi.org/10.1090/S0025-5718-1982-0637296-4.

Golden, J. H., and D. Purcell, 1978: Airflow characteristics around the Union City tornado. *Mon. Wea. Rev.*, **106**, 22–28, https://doi.org/10.1175/1520-0493(1978)106<0022:ACATUC>2.0.CO;2.

Haltiner, G. J., and R. T. Williams, 1980: *Numerical Prediction and Dynamic Meteorology.* 2nd ed. Wiley, 477 pp.

Hardy, R. L., 1971: Multiquadric equations of topography and other irregular surfaces. *J. Geophys. Res.*, **76**, 1905–1915, https://doi.org/10.1029/JB076i008p01905.

Hildebrand, F. B., 1965: *Methods of Applied Mathematics.* 2nd ed. Prentice Hall, 362 pp.

Isom, B. M., and Coauthors, 2009: Detailed observations of wind turbine clutter with scanning weather radars. *J. Atmos. Oceanic Technol.*, **26**, 894–910, https://doi.org/10.1175/2008JTECHA1136.1.

Lakshmanan, V., C. Karstens, J. Krause, and L. Tang, 2014: Quality control of weather radar data using polarimetric variables. *J. Atmos. Oceanic Technol.*, **31**, 1234–1249, https://doi.org/10.1175/JTECH-D-13-00073.1.

Lemon, L. R., D. W. Burgess, and R. A. Brown, 1978: Tornadic storm airflow and morphology derived from single-Doppler radar measurements. *Mon. Wea. Rev.*, **106**, 48–61, https://doi.org/10.1175/1520-0493(1978)106<0048:TSAAMD>2.0.CO;2.

Mahalik, M. C., B. R. Smith, K. L. Elmore, D. M. Kingfield, K. L. Ortega, and T. M. Smith, 2019: Estimates of gradients in radar moments using a linear least squares derivative algorithm. *Wea. Forecasting*, **34**, 415–434, https://doi.org/10.1175/WAF-D-18-0095.1.

Mitchell, E. D., S. V. Vasiloff, G. J. Stumpf, A. Witt, M. D. Eilts, J. T. Johnson, and K. W. Thomas, 1998: The National Severe Storms Laboratory tornado detection algorithm. *Wea. Forecasting*, **13**, 352–366, https://doi.org/10.1175/1520-0434(1998)013<0352:TNSSLT>2.0.CO;2.

Nuss, W. A., and D. W. Titley, 1994: Use of multiquadric interpolation for meteorological objective analysis. *Mon. Wea. Rev.*, **122**, 1611–1631, https://doi.org/10.1175/1520-0493(1994)122<1611:UOMIFM>2.0.CO;2.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, 1996: *Numerical Recipes in Fortran 90: The Art of Parallel Scientific Computing.* Vol. 2, *Fortran Numerical Recipes*, 2nd ed. Cambridge University Press, 1486 pp.

Serafin, R. J., and J. W. Wilson, 2000: Operational weather radar in the United States: Progress and opportunity. *Bull. Amer. Meteor. Soc.*, **81**, 501–518, https://doi.org/10.1175/1520-0477(2000)081<0501:OWRITU>2.3.CO;2.

Stumpf, G. J., A. Witt, E. D. Mitchell, P. L. Spencer, J. T. Johnson, M. D. Eilts, K. W. Thomas, and D. W. Burgess, 1998: The National Severe Storms Laboratory mesocyclone detection algorithm for the WSR-88D. *Wea. Forecasting*, **13**, 304–326, https://doi.org/10.1175/1520-0434(1998)013<0304:TNSSLM>2.0.CO;2.

Thompson, P. D., 1961: *Numerical Weather Analysis and Prediction.* Macmillan, 170 pp.

Trapp, R. J., and C. A. Doswell III, 2000: Radar data objective analysis. *J. Atmos. Oceanic Technol.*, **17**, 105–120, https://doi.org/10.1175/1520-0426(2000)017<0105:RDOA>2.0.CO;2.

Wakimoto, R. M., N. T. Atkins, K. M. Butler, H. B. Bluestein, K. Thiem, J. Snyder, and J. Houser, 2015: Photogrammetric analysis of the 2013 El Reno tornado combined with mobile X-band polarimetric radar data. *Mon. Wea. Rev.*, **143**, 2657–2683, https://doi.org/10.1175/MWR-D-15-0034.1.

——, and Coauthors, 2016: Aerial damage survey of the 2013 El Reno tornado combined with mobile radar data. *Mon. Wea. Rev.*, **144**, 1749–1776, https://doi.org/10.1175/MWR-D-15-0367.1.

——, Z. Wienhoff, H. B. Bluestein, and D. Reif, 2018: The Dodge City tornadoes on 24 May 2016: Damage survey, photogrammetric analysis combined with mobile polarimetric radar data. *Mon. Wea. Rev.*, **146**, 3735–3771, https://doi.org/10.1175/MWR-D-18-0125.1.

Weinberger, H. F., 1965: *A First Course in Partial Differential Equations with Complex Variables and Transform Methods.* Blaisdell, 446 pp.

Wood, V. T., and R. A. Brown, 1997: Effects of radar sampling on single-Doppler velocity signatures of mesocyclones and tornadoes. *Wea. Forecasting*, **12**, 928–938, https://doi.org/10.1175/1520-0434(1997)012<0928:EORSOS>2.0.CO;2.

——, ——, and D. C. Dowell, 2009: Simulated WSR-88D velocity and reflectivity signatures of numerically modeled tornadoes. *J. Atmos. Oceanic Technol.*, **26**, 876–893, https://doi.org/10.1175/2008JTECHA1181.1.

Zrnić, D. S., D. W. Burgess, and L. D. Hennington, 1985: Automatic detection of mesocyclonic shear with Doppler radar. *J. Atmos. Oceanic Technol.*, **2**, 425–438, https://doi.org/10.1175/1520-0426(1985)002<0425:ADOMSW>2.0.CO;2.