

# Optimization of the Cross-Correlation Algorithm for Two-Component Wind Field Estimation from Single Aerosol Lidar Data and Comparison with Doppler Lidar

MASAKI HAMADA, PIERRE DÉRIAN, CHRISTOPHER F. MAUZEY, AND SHANE D. MAYOR

*California State University, Chico, Chico, California*

(Manuscript received 1 January 2015, in final form 19 October 2015)

## ABSTRACT

Numerical and field experiments were conducted to test an optimized cross-correlation algorithm (CCA) for the remote sensing of two-component wind vectors from horizontally scanning elastic backscatter lidar data. Each vector is the result of applying the algorithm to a square and contiguous subset of pixels (an *interrogation window*) in the lidar scan area. Synthetic aerosol distributions and flow fields were used to investigate the accuracy and precision of the technique. Results indicate that in neutral static stability, when the mean flow direction over the interrogation window is relatively uniform, the random error of the estimates increases as the mean wind speed and turbulence intensity increases. In convective conditions, larger errors may occur as a result of the cellular nature of convection and the dramatic changes in wind direction that may span the interrogation window. Synthetic fields were also used to determine the significance of various image processing and numerical steps used in the CCA. Results show that an iterative approach that dynamically reduces the block size provides the largest performance gains. Finally, data from a field experiment conducted in 2013 in Chico, California, are presented. Comparisons with Doppler lidar data indicate excellent agreement for the 10-min mean wind velocity computed over a set of 150 h: the root-mean-square deviations (and slopes) for the  $u$  and  $v$  components are  $0.36 \text{ m s}^{-1}$  (0.974) and  $0.37 \text{ m s}^{-1}$  (0.991), respectively, with correlation coefficients  $R^2 > 0.99$ .

## 1. Introduction

The cross-correlation algorithm (CCA) is a mainstay in the field of motion estimation. It is used to compute the apparent motion of objects and fluid flows in the fields of robotics, navigation, medical imaging, and geosciences (Murray et al. 2009; Emery et al. 2003; Avants et al. 2008; Schubert et al. 2013; Adrian and Westerweel 2011; Cheng et al. 2005; Antoine et al. 2013). In the atmospheric sciences, the cross-correlation algorithm has been applied to satellite imagery (Leese et al. 1971; García-Pereda and Borde 2014), radar data (Rinehart and Garvey 1978), and lidar data (Eloranta et al. 1975; Shimizu et al. 1981; Kolev et al. 1988; Schols and Eloranta 1992; Piironen and Eloranta 1995). Under the assumption that macroscopic aerosol features are advected by the wind, the wind velocity can be estimated remotely from the apparent motion of the features in lidar image sequences.

The CCA has been the primary numerical method for determining fluid motion in particle image velocimetry (PIV) experiments.<sup>1</sup> In most PIV experiments, the fluid is deliberately seeded with very small particles that serve as robust tracers of the local flow. The particles are typically illuminated with a laser light plane and individual particles are discernible in the rapidly collected sequence of images taken by a camera. The particles do not change appreciably in shape nor in brightness as they move. For low and moderate particle density experiments, in which the fluid between the particles appears dark and does not contribute any information, the motion of the small particles is solely relied upon for determining the motion field. As a result, the cross-correlation functions (CCFs) for PIV experiments contain sharp peaks. Furthermore, few particles straddle the edges of the interrogation window. Particles that appear or disappear within the time between two frames, either

---

Corresponding author address: Shane D. Mayor, California State University, Chico, 400 W. First St., Chico, CA 95929.  
E-mail: sdmayor@csuchico.edu

---

<sup>1</sup> Recently, optical flow methods have become an alternative. For example, Corpetti et al. (2006) applies an optical flow scheme to study the plane turbulent mixing layer, the wake of a circular cylinder, and vorticity measurement.

by moving into or out of the interrogation window, or into or out of the illumination plane, only contribute to incoherent variance confined to the zero lag of the CCF. The application of the CCA to derive air motion from atmospheric aerosol backscatter lidar images raises a set of issues that are not present in most PIV experiments (at least low-density experiments). These issues result from the time required for the lidar to collect a scan, the inability to discern individual particles, and the continuous range-dependent image intensity that is a proxy to particle concentration (Held et al. 2012).

At the heart of the CCA is the computation of a CCF from a pair of images. The location of the peak of the CCF relative to its origin indicates the displacement of dominant image features. A consequence of selecting a single peak is the loss of information that describes the true velocity field that causes the motion of features. As reflected in the title of the paper by Schols and Eloranta (1992), some have regarded the velocity estimate to be the *area average* of the wind field in the interrogation window. However, Schols and Eloranta (1992) do not elaborate on their use of area average. The wind is a vector and the average of a set of vectors may be computed by *vector averaging* (summing the components first and then computing the mean vector) or *scalar averaging* (averaging the speeds and directions that result from each vector in a set). In the former, a mean displacement would result. In the later, the mean speed may be different from the magnitude of the mean displacement, and of more interest to the wind energy community, for example. Moreover, it is possible that, because of the organization of the wind and the aerosol fields within the interrogation window, CCA estimates do not equal an area average at all. To investigate these issues, we generated synthetic aerosol images and velocity fields and applied the CCA to the synthetic aerosol backscatter images. This enabled us to compare the results of the CCA with the mean of the synthetic velocity field that was used to advect the synthetic aerosol structures.

In addition, the synthetic data are used to confirm the necessary sequence of steps within the CCA. Mayor et al. (2012) compared the results of a rudimentary CCA applied to scanning elastic backscatter aerosol lidar data to tower-mounted sonic anemometer wind measurements between 10 and 30 m AGL. The study showed that the use of the CCA for interrogation windows that were 500 m  $\times$  500 m and smaller produced plausible but noisy results. However, a large and important part of the spectrum of atmospheric motions exists at these microscales. Thus, a higher spatial resolution of the wind velocity fields is desired. The algorithm used by Mayor et al. (2012) was functional but not optimal. It did not include several steps tested in this paper that may

deliver better results. They are 1) zero padding to account for nonperiodic images, 2) use of a window function to reduce the undesirable effects of sharp edges and aerosol features entering and leaving the block area, 3) multipass interrogation for iterative refinement of the motion estimation, 4) multigrid interrogation to improve the spatial resolution of the resulting flow fields, and 5) deformation of images to correct for the mean advection of aerosol features during the time required to complete one scan as described by Sasano et al. (1982).

Finally, because the synthetic images and velocity fields are only an approximation to reality, we also conducted a field experiment to validate the results of the optimized CCA applied to real lidar data. The elastic backscatter aerosol lidar system used was the Raman-shifted Eye-safe Aerosol Lidar (REAL) (Mayor and Spuler 2004; Spuler and Mayor 2005; Mayor et al. 2007). A single compact Doppler lidar was used to validate the wind velocity fields resulting from the optimized CCA applied to the REAL aerosol backscatter images.

This paper is organized as follows: Section 2 introduces the CCA and describes the options that improve the wind velocity estimation. In section 3, the performance of the CCA with these options is evaluated by using synthetic lidar backscatter images and wind velocity fields. In section 4, the performance of the optimized CCA is validated by field experiments using the Doppler lidar (DL) data as a reference.

## 2. Application of the cross-correlation algorithm to elastic lidar backscatter images

Scanning lidar data are collected in a spherical coordinate system with coordinates of azimuth, elevation, and range. The data are processed and interpolated to a Cartesian grid before the CCA is applied. The “pre-processing” that is applied to each range-dependent lidar backscatter array includes the calculation and subtraction of the raw background signal, multiplication of the waveform by the range squared, conversion to decibels, and the application of a high-pass median filter (Schols and Eloranta 1992). The application of the high-pass median filter is critical. It removes large-scale trends in the backscatter arrays caused by attenuation (or real atmospheric structure) and leaves the smaller-scale backscatter perturbations that are more likely to be advected by the local wind. For the REAL data, we typically use a 333-point filter window (500 m in range) resulting in the removal of features larger than one-half the filter length (250 m). The preprocessed data are then interpolated to a Cartesian grid, which results in a 2D array referred to as the *image*. In this work, the image is constructed by interpolating to a Cartesian grid with a spacing of 10 m in

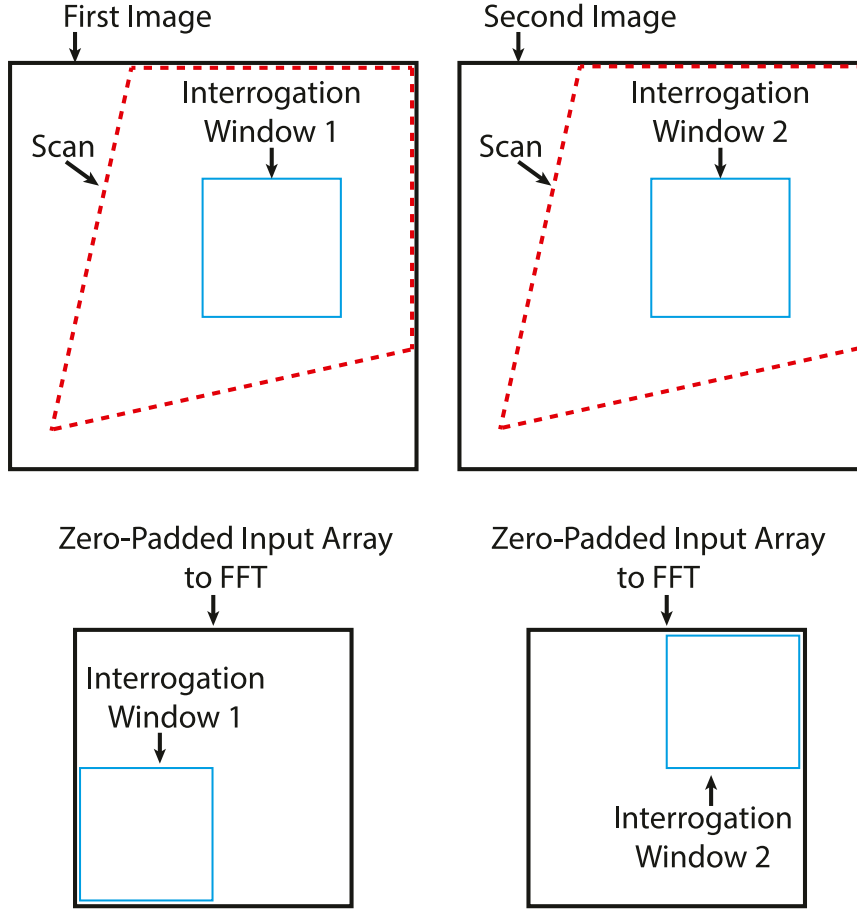


FIG. 1. Diagram defining terminology used and relative locations of each object.

both the east–west and north–south directions. The CCA can then be applied to any square subset of the image. This subset is referred to as the *interrogation window* or the *block* (see Fig. 1). In practice we often start with a 1 km<sup>2</sup> block containing 100 × 100 data points. The CCA is applied to as many different blocks as desired to compute a vector flow field.

*a. The cross-correlation function*

The CCF is a description of the correlation of two arrays, as a function of delay (in time) or lag (in space). When applied to blocks from a pair of consecutive images, the peak location of the CCF indicates the displacement of the dominant image features within the time interval  $\delta t$  between the images. Let  $S(x, y, t)$  stand for the preprocessed lidar data in a 2D image at discrete times and let  $L$  stand for the length of one side of the interrogation window. Then, the interrogated signal,  $S_{pq}(x, y, t)$ , at the window center  $(x_p, y_q)$  is expressed as

$$S_{pq}(x, y, t) = S(x, y, t; x_p, y_q), \quad (1)$$

whose spatial domain is  $x \in [x_p - (L/2), x_p + (L/2)]$  and  $y \in [y_q - (L/2), y_q + (L/2)]$ . Next, let  $t_1$  and  $t_2 = t_1 + \delta t$  for reference times for two consecutive images. Then, the corresponding interrogated signals  $S_{pq,1}(x, y)$  and  $S_{pq,2}(x, y)$  are defined as

$$S_{pq,1}(x, y) = S_{pq}(x, y, t_1) \quad (2)$$

and

$$S_{pq,2}(x, y) = S_{pq}(x, y, t_2). \quad (3)$$

The normalized 2D CCF at the interrogation window,  $r_{pq}$ , for two series,  $S_{pq,1}(x, y)$  and  $S_{pq,2}(x, y)$ , is defined as

$$r_{pq} = \frac{\text{Cov}[S_{pq,1}(x, y), S_{pq,2}(x, y)]}{\sigma_1 \sigma_2}, \quad (4)$$

where  $\text{Cov}[S_{pq,1}(x, y), S_{pq,2}(x, y)]$  is the covariance of the overlapped portions of  $S_{pq,1}(x, y)$  and  $S_{pq,2}(x, y)$ ,  $\sigma_1$  is the standard deviation of  $S_{pq,1}(x, y)$ , and  $\sigma_2$  is the standard deviation of  $S_{pq,2}(x, y)$  (Davis and Sampson 2002). For computational efficiency, the fast Fourier

transform (FFT) is widely used instead of Eq. (4). Let  $S_{pq,j}(x_m, y_n)$ ,  $j = 1, 2$  be discrete signals in the interrogation window,  $N_x$  be the number of data points in the  $x$  direction,  $N_y$  be the number of points in the  $y$  direction,

$k_x$  be the wavenumber corresponding to the  $x$  coordinate, and  $k_y$  be the wavenumber corresponding to the  $y$  coordinate. The FFT of the given signals,  $\text{FFT}_{pq,j}$ , can be expressed as

$$\text{FFT}_{pq,j} = \frac{\sum_{m=1}^{N_x} \sum_{n=1}^{N_y} S_{pq,j}(x_m, y_n) e^{-i2\pi[(k_x x_m)/(N_x) + (k_y y_n)/(N_y)]}}{N_x N_y}, \quad j = 1, 2, \quad (5)$$

and the cross-correlation function at the interrogation window  $r_{pq}$  can be expressed as

$$r_{pq} = \frac{\text{FFT}^{-1}(\text{FFT}_{pq,1} \text{FFT}_{pq,2}^*)}{\sigma_1 \sigma_2}, \quad (6)$$

where  $\text{FFT}_{pq,1}$  is the FFT of  $f_{pq,1}(x, y)$ ,  $\text{FFT}_{pq,2}^*$  is the complex conjugate of  $\text{FFT}_{pq,2}$ , and  $\text{FFT}^{-1}$  represents the inverse fast Fourier transform.

#### b. Zero padding

The FFT is designed for periodic signals, but real atmospheric lidar data are not periodic. Thus, it is important to circumvent the assumption of periodicity by using “zero padded” arrays (Adrian and Westerweel 2011). Each dimension is padded by zeros over a domain that is twice the size of the original signal, so three-quarters of the arrays for input to the FFT are filled with zeros (Bastiaans 2000). The first input array contains the data from the first interrogation window in the lower-left corner and zeros elsewhere. The second input array contains the data from the second interrogation window in the upper-right corner and zeros elsewhere. Figure 1 shows this arrangement. This is done because the complex conjugate of the FFT of the second input array is used to compute the CCF. According to Hamada (2014), the performance of a CCA using 100 pairs of synthetic backscatter images and synthetic wind velocity fields are slightly increased by using zero-padded images.

#### c. Histogram equalization

Histogram equalization enhances image contrast by adjusting the histogram distribution of pixel intensity. The histogram function is used to obtain the density distribution of the interrogated signal. Let  $S_j(x_m, y_n)$  stand for the discrete signal in the interrogation window. Then, the histogram function  $H_j$  is defined as

$$H_j = \sum_{m=1}^{N_x} \sum_{n=1}^{N_y} P[S_j(x_m, y_n), i], \quad i = 1, 2, \dots, \frac{S_{\max} - S_{\min}}{w}, \quad (7)$$

where  $N_x$  is the number of data points in the  $x$  direction,  $N_y$  is the number of points in the  $y$  direction,  $S_{\max}$  is the maximum value of  $S_j(x_m, y_n)$ ,  $S_{\min}$  is the minimum value of  $S_j(x_m, y_n)$ ,  $w$  is the bin width, and

$$P[S_j(x_m, y_n), i] = \begin{cases} 1: & i \leq \frac{S_j(x_m, y_n) - S_{\min}}{w} \leq i + 1, \\ 0: & \text{otherwise.} \end{cases}$$

Term  $H_j$  is integrated to obtain the cumulative density probability function and it is transformed to a histogram-equalized byte array. After processing, the values of the discrete signals range from 0 to 255.

Histogram equalization has been used previously for computing horizontal wind vectors from lidar backscatter images (Schols and Eloranta 1992). Without histogram equalization, the motion of small areas of bright features may dominate the CCF. In that case, the cross correlation may be biased toward the motion of the small and bright features. On the other hand, with histogram equalization, other dimmer features in the image are able to influence the CCF. However, according to Hamada (2014), histogram equalization tends to broaden the peak of the CCF thereby increasing uncertainty of the peak location. Since our experiment to test histogram equalization involved only one relatively strong wind case (mean wind speed of approximately  $10 \text{ m s}^{-1}$ ), the effects of histogram equalization should be investigated for other scenarios in the future to determine whether it is broadly beneficial.

#### d. Window function

The performance of the CCA decreases if bright features straddle the edges of the interrogation window. This tends to distort the shape of the CCF, shifts the location of its peak, and leads to an underestimation of the wind velocity vector. Window functions, such as the Tukey window, may be applied to taper the backscatter intensity near the image block edges. Let  $N$  be the  $x$  dimension of a 1D array. Then, the 1D Tukey window  $w(x)$  is defined as

$$w(x) = \begin{cases} \frac{1}{2} \left( 1 + \cos \left\{ \pi \left[ \frac{2x}{\alpha(N-1)} - 1 \right] \right\} \right) : & 0 \leq x \leq \frac{\alpha(N-1)}{2}, \\ 1 : & \frac{\alpha(N-1)}{2} \leq x \leq 1(N-1) \left( 1 - \frac{\alpha}{2} \right), \\ \frac{1}{2} \left( 1 + \cos \left\{ \pi \left[ \frac{2x}{\alpha(N-1)} - \frac{2}{\alpha} + 1 \right] \right\} \right) : & (N-1) \left( 1 - \frac{\alpha}{2} \right) \leq x \leq (N-1), \end{cases}$$

where  $\alpha$  is a constant that determines the width of the cosine lobe of the window (Tukey 1967). It is set to 0.2 for this study. The Tukey window function  $w(x)$  can be extended to 2D by multiplying by  $w(y)$  in the  $y$  direction. The 2D Tukey window is expressed as  $w(x, y) = w(x)w(y)$ . The Tukey window effectively decreases the intensity of aerosol features near the image block edges, thus reducing undesirable effects such as Gibbs phenomenon. Other window functions (such as triangular, Parzen, Hann, Blackman, etc.) may also be effective.

*e. Multipass interrogation*

The CCF for two consecutive lidar backscatter blocks gathers contributions from aerosol features that appear in both blocks. Because aerosol features are advected by the wind, some aerosol features in the first image block may move out of the interrogation window in the time interval between consecutive scans. During that time, aerosol features initially outside the first block may appear in the second block. In this case, these features do not contribute to the CCF, and the wind velocity may be underestimated. For example, if the wind velocity field is nonuniform within an interrogation window, then the aerosol features with lower velocity tend to remain within two consecutive image blocks, while those moving faster tend to disappear. In this case, the CCA is biased and underestimates the wind velocity field in the interrogation window. A multipass interrogation (Raffel et al. 2007) can minimize such effects. This approach iterates two steps: (i) compute a displacement vector from two image blocks by the cross correlation and then (ii) displace the center of the second block according to this vector. Each vector is an incremental refinement of the solution, and the process repeats until the magnitude of the incremental vector falls below 1 pixel—typically, after two or three iterations. The subpixel location of the peak of the CCF is then estimated by curve fitting following Piironen and Eloranta (1995). Finally, the solution vector is given by the sum of the incremental estimations and the subpixel location. Following this process, the displaced second block contains a similar number of the features that also appeared in the first image block. The CCF is therefore

better defined, and the accuracy of the motion estimation increases.

*f. Multigrid interrogation*

The CCA provides one wind velocity vector per interrogation window. The typical size of a large interrogation window, for elastic lidar backscatter images, is about 1 km × 1 km, which is larger than the size of most turbulent coherent structures. With a large block size, most microscale structures cannot be resolved although they are important meteorological phenomena. In this case, multigrid interrogation can be used to increase the spatial resolution of the wind velocity vector field. Multigrid interrogation is similar to multipass interrogation except that the dimensions of the blocks are reduced after each pass (Adrian and Westerweel 2011). In general, the number of features that appeared in both blocks decreases as the block size is reduced—typically, the size reduction factor is 50%. However, displacing the second block increases the similarity of these blocks and makes it possible to resolve the wind velocity vector for a relatively small region (less than 500 m × 500 m). Mayor and Eloranta (2001) applied multipass interrogation and multigrid interrogation but were not able to validate the resulting flow fields.

*g. Image deformation*

Lidar scans do not represent an instantaneous distribution of aerosol features as in a “snapshot” because of the time required to complete a scan; that is, during a scan some aerosol features are observed before others, while all features are advected by the wind. The result is a distorted image relative to the ideal snapshot. Because of this, apparent displacements are inaccurate and the motion estimation is biased. This problem was first discussed by Sasano et al. (1982), who proposed an iterative correction of image deformation. In this process, aerosol features that are observed before and after a reference time within a given scan are translated forward and backward, respectively, according to the mean estimated wind velocity in the entire region of the scan. The process is repeated until the mean velocity is changed by less than 1%. After the image correction, the deformed image approximates a snapshot of the true

lidar backscatter distribution at the reference time, here corresponding to the time associated with the backscatter arrays collected in the center of the sector. While processing and comparing the time series data presented in section 4, we noticed a clear bias in the velocity estimates when the image deformation was turned off. The magnitude of the bias increased with the mean wind speed and was eliminated with the image deformation turned on.

#### h. Quality control

In general, the signal-to-noise ratio (SNR) of elastic backscatter data decays as one over the range squared. At long ranges, the noise amplitude may dominate the backscatter from coherent aerosol features. Applying the CCA to such data can result in areas of spurious, random vectors. In other circumstances, the aerosol features can lead to a peak of the CCF at a location that does not represent the actual motion. Such situations are more likely to occur close to the scan edges, or when high wind speeds are involved. They result in isolated spurious vectors, known as outliers. These two distinct sets of erroneous estimates are detected by two different mechanisms inspired from the PIV expertise.

A first step consists of discarding vectors for which the value of the peak of the CCF is below a certain threshold. This test is efficient at detecting erroneous vectors resulting from low SNR backscatter data, typically removing patches of vectors in the far range. The second step is handled once the whole vector field has been estimated. It is the *normalized median test*, as described in Adrian and Westerweel (2011) and Westerweel and Scarano (2005). It assumes a local, spatial coherence of the vector field and therefore is able to detect isolated outliers that were missed by the previous test.

For a displacement vector  $\mathbf{d}$ , these tests can be written as

$$\left. \begin{array}{l} \text{if } r_{pq,\max} < \tau_r (\text{noisy data}) \\ \text{or } \frac{|\mathbf{d} - \mathbf{d}_m|}{\sigma_m + \sigma_\epsilon} < \tau_m (\text{isolated outlier}) \end{array} \right\} \Rightarrow \mathbf{d} \text{ discarded,} \quad (8)$$

where  $r_{pq,\max}$  is the value of the peak of the CCF,  $\tau_r$  is a threshold value of the peak of the CCF,  $\mathbf{d}_m$  is the median of the eight vectors  $\mathbf{d}_i$  neighboring  $\mathbf{d}$ ,  $\sigma_m$  is the median of the neighboring residuals  $\{|\mathbf{d}_i - \mathbf{d}_m|, i = 1, \dots, 8\}$ ,  $\sigma_\epsilon$  is the acceptable fluctuation level due to the CCF, and  $\tau_m$  is a threshold value for the normalized median test. In this work, the threshold values are  $\tau_r = 0.2$  and  $\tau_m = 2$ . The typical root-mean-square noise level of the PIV data is about 0.1 pixels (Westerweel 2000), so  $\sigma_\epsilon = 0.1$  was chosen for this study. Since the normalized CCFs

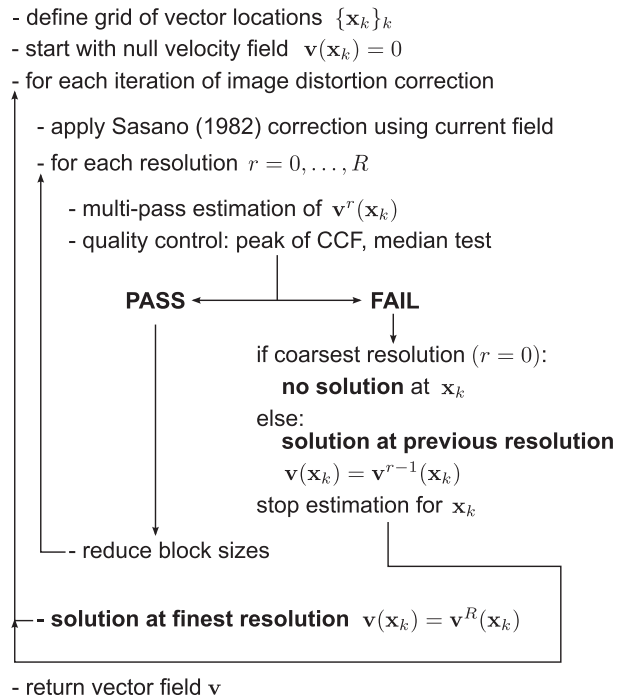


FIG. 2. Simplified diagram of the CCA.

were used, the same threshold value  $\tau_r = 0.2$  can be used for different size of blocks. In addition, the universality of the normalized median test makes it suitable for iterative PIV interrogation schemes (Raffel et al. 2007). Thus, both of these tests do not depend on the size of the blocks and can be integrated to a multigrid interrogation process. They are applied after each step of the multigrid interrogation. Vectors flagged as spurious are replaced by their value at the previous step of the estimation, if available, and the estimation process stops. As such, the algorithm is adaptive: the estimation proceeds to smaller blocks (finer motion scales) only when the quality of data is locally good enough to support it.

#### i. Implementation

A simplified diagram of the operational algorithm is presented in Fig. 2. It combines iterations of the distortion correction (section 2g), the multigrid interrogation (section 2f), and the multipass estimation (section 2e). For each vector, there can be up to 27 iterations total (three for each of the distortion correction, multigrid, and multipass), with as many evaluations of the CCF. To complete the execution of the motion estimation within the time between two scans of the REAL, the core pieces of the method (CCF, histogram equalization, interpolation for the distortion correction) are written in the Compute Unified Device Architecture

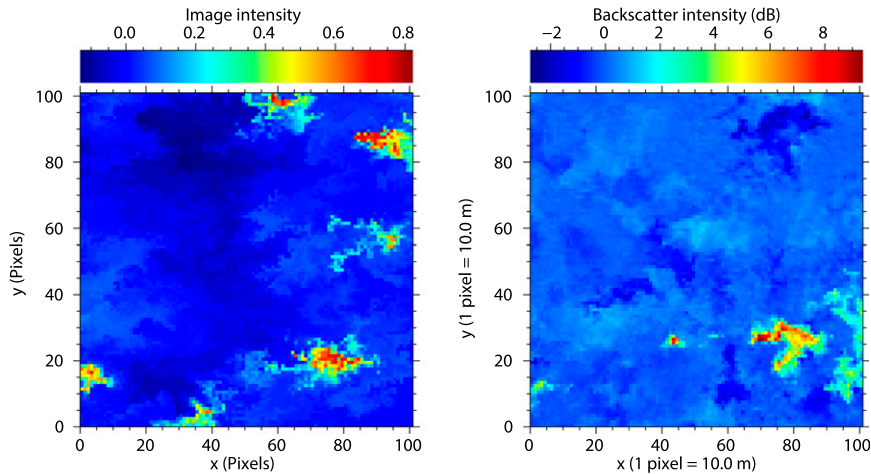


FIG. 3. Comparison of (left) a synthetic backscatter image and (right) a REAL backscatter image. The REAL backscatter image was collected at the University Farm at California State University, Chico, on 17 Oct 2013.

(CUDA) language (Mauzey et al. 2012). These functions are executed in a massively parallel fashion on specific graphic processing units that are designed for scientific computation, thus enabling real-time execution.

### 3. Tests using synthetic backscatter images and wind velocity fields

Unlike other wind measurement techniques that sample relatively small volumes of the atmosphere, the CCA relies upon spatial data over a large area to make a velocity estimate that is assigned to a single point (the location of the center of the interrogation window). The spatial data are the aerosol backscatter field in the interrogation window. If the actual wind field is uniform throughout the interrogation window, then the spatial distribution of aerosol features that contribute to the CCF does not matter. However, the wind velocity in the real world is spatially variable. Therefore, it is reasonable to question whether the peak of the CCF represents a central tendency of the actual spatial wind field within the interrogation window. This is because the aerosol features may not be uniformly distributed across the interrogation window, or they may be correlated with velocity structures.

To study this problem, we developed synthetic velocity fields and synthetic aerosol backscatter images. The synthetic velocity fields, although not continuously defined, are high in spectral integrity, of much higher resolution than can be obtained from present observing techniques, and enable us to calculate the spatial mean velocity. The synthetic backscatter field enables us to confirm the previously suggested notion that the CCA results in an area average of the wind field in the interrogation window.

The structure of turbulence depends on the static stability (Moeng and Sullivan 1994; Khanna and Brasseur 1998). Turbulence may be generated by shear or convection or some combination of both. During periods of neutral static stability, turbulence is generated only by shear. Coherent wind and aerosol structures tend to be elongated and aligned with the mean wind direction (streaks). During periods of pure convection in the lower boundary layer, cellular flow structures exist. The cells are characterized by broad areas of subsidence and divergence and share borders of strong convergence and updrafts (Schmidt and Schumann 1989). Often, vortices form near the intersections of the convergence lines (Kanak 2005).

Because these two stability regimes produce different turbulence coherent structures, we created two types of synthetic flow fields for testing. For pure shear-induced turbulent flows, we added small-scale turbulent perturbations to a uniform velocity field. For pure convection, the four fundamental kinematic components (divergence, rotation, stretching deformation, and shearing deformation) (Holton and Hakim 2013, 23–25) of fluid flow fields were generated.

To generate synthetic backscatter and velocity fields, we first defined a Cartesian grid. The grid spacing ( $\delta x = \delta y = 10$  m) and the time between two consecutive images ( $\delta t = 10$  s) were chosen so that the motion of one unit (10 m) during a time step (10 s) represents a velocity of  $1 \text{ m s}^{-1}$ . A synthetic backscatter image is created initially by assigning random numbers to the grid points. Then, a  $25 \times 25$  pixel rectangular smooth is applied to the random numbers to approximate a field of coherent structures with the characteristic length scale of 250 m. Next, small Gaussian features were randomly added to simulate local sources of aerosol features that are

routinely observed in REAL data. Then, a synthetic turbulent velocity perturbation field, as generated by the model of Mann (1994, 1998) was used to diffuse both the sum of the Gaussian features and background. The Mann model is a spectral tensor turbulence model based on rapid distortion theory (Durbin and Reif 2010) that provides idealized three-component spatial velocity fields with realistic turbulence structure for neutral atmospheric surface-layer flows. The model is widely used in the field of wind engineering. To run the model, we specified three-dimensional domains containing  $512 \times 512 \times 32$  wind vectors spaced at 10-m intervals in all dimensions. The turbulence intensity was controlled by the roughness length and the mean velocity. Multiple independent velocity perturbation fields were used to diffuse the aerosol field. Figure 3 shows a synthetic backscatter image resulting from this procedure and a REAL backscatter image. The REAL backscatter image was extracted from a lidar scan collected at the University Farm at California State University, Chico, on 17 October 2013. The REAL backscatter data was also interpolated to a  $10\text{ m} \times 10\text{ m}$  grid. From Fig. 3, similar spatial gradients and backscatter intensity ranges can be observed in both images although the exact distributions of aerosol features are different.

Next, the synthetic backscatter field must be displaced to generate a synthetic backscatter field at a subsequent time. A second image corresponding to  $t_2$  was generated by displacing the synthetic backscatter image according to a synthetic velocity field as described below. Values on the Cartesian grid were obtained from the bicubic interpolation of the displaced image data using homogeneous Neumann boundary conditions. The synthetic velocity field used for the displacement of image synthetic aerosol features from one frame to the next was created by adding new turbulent perturbations to analytic functions. Let  $u(x, y)$  and  $v(x, y)$  be the east-west and the north-south components of the wind velocity, and  $u'(x, y)$  and  $v'(x, y)$  be the corresponding turbulent perturbations generated by the Mann model (Fig. 4). Then, the wind velocity field can be expressed as

$$\begin{bmatrix} u(x, y) \\ v(x, y) \end{bmatrix} = \begin{bmatrix} u_a x + u_b y + u_c + u'(x, y) \\ v_a x + v_b y + v_c + v'(x, y) \end{bmatrix}, \quad (9)$$

where  $u_a, u_b, u_c, v_a, v_b,$  and  $v_c$  are constants. For shear-driven turbulent flow (neutral static stability) we set  $u_a = u_b = v_a = v_b = 0$ , so the velocity fields can be expressed as

$$\begin{bmatrix} u(x, y) \\ v(x, y) \end{bmatrix} = \begin{bmatrix} u_c + u'(x, y) \\ v_c + v'(x, y) \end{bmatrix}. \quad (10)$$

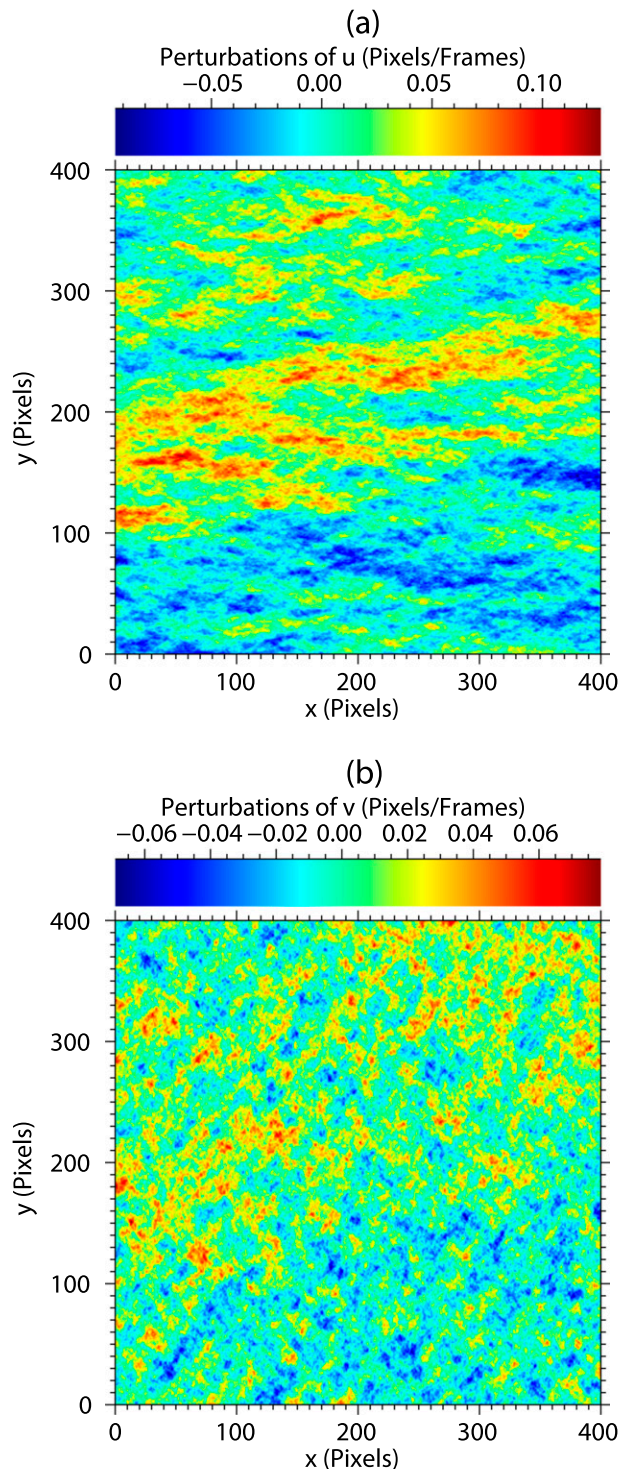


FIG. 4. The (a)  $u$  and (b)  $v$  components of shear-induced turbulent wind perturbations resulting from the Mann model.

Since the Mann model is designed for neutral conditions, we chose not to use it to generate turbulent perturbations for convective conditions. Therefore, to simulate flow structures that we routinely observe with

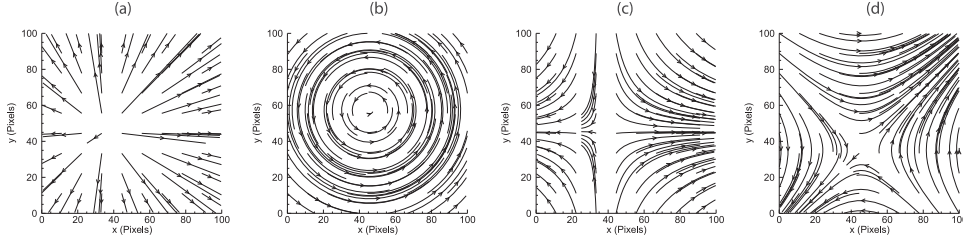


FIG. 5. Streamlined flow fields for pure (a) divergence, (b) rotation, (c) stretching deformation, and (d) shearing deformation. The centers of the functions were deliberately offset because the interrogation window is unlikely to be perfectly centered on such flow structures when applied to real data.

the lidar in the lower convective boundary layer, we set  $u'(x, y) = v'(x, y) = 0$  resulting in

$$\begin{bmatrix} u(x, y) \\ v(x, y) \end{bmatrix} = \begin{bmatrix} u_a x + u_b y + u_c \\ v_a x + v_b y + v_c \end{bmatrix}. \quad (11)$$

To specify these remaining terms, we used analytic functions for four fundamental kinematic flow structures: divergent flow [ $u_a = 0.1 \text{ s}^{-1}$ ,  $v_b = 0.1 \text{ s}^{-1}$ ,  $u_c = -u_a x_0$ , and  $v_c = -v_b y_0$ , where  $(x_0, y_0)$  is the center of divergence], rotational flow ( $u_b = -0.1 \text{ s}^{-1}$ ,  $v_a = 0.1 \text{ s}^{-1}$ ,  $u_c = -u_b y_0$ , and  $v_c = -v_a x_0$ ), stretching deformation ( $u_a = 0.1 \text{ s}^{-1}$ ,  $v_b = -0.1 \text{ s}^{-1}$ ,  $u_c = -u_a x_0$ , and  $v_c = -v_b y_0$ ), and shearing deformation ( $u_b = 0.1 \text{ s}^{-1}$ ,  $v_a = 0.1 \text{ s}^{-1}$ ,  $u_c = -u_b y_0$ , and  $v_c = -v_a x_0$ ). Figure 5 shows streamlines of these flow fields. While the finescale turbulent perturbations used in the neutral shear cases are absent, we assert that the analytic functions represent the most important horizontal velocity transport features, resulting from buoyancy-driven flows at altitudes in the lower boundary layer. These are highly idealized changes in wind speed and direction associated with convection cells that occur on length scales similar to the dimensions of interrogation windows and are responsible for the advection and distortion of aerosol features on the time scales required to repeat lidar sector scans.

The CCA was applied to the synthetic pairs of images and velocity vectors were computed. The precise peak location of the CCF is estimated by fitting a 2D surface (polynomial of degree 2) to the peak ( $5 \times 5$  pixels). Without this subpixel interpolation, the theoretical error on the estimated motion is  $\pm 0.5\delta x/\delta t$  on each velocity component. Each experiment was repeated on 100 different pairs of synthetic backscatter images, and the mean and the standard deviation of the velocity vectors were compared with the mean velocity of the given synthetic wind velocity field.

#### a. Significance of optional steps

The experiments with synthetic data were used to elucidate two aspects of the CCA. First, they were used

to study the significance of the impact of the optional numerical procedures within the CCA (described in section 2 and Table 1). Results are shown in Table 2. The image distortion correction, as described by Sasano et al. (1982), was not included since there is no image distortion for the synthetic lidar backscatter images. The five options of the CCA are the multipass interrogation (MP), the multigrid interrogation (MG), the zero padding (ZP), the Tukey window (TW), and the histogram equalization (HE). None of these options are included for test 1, but all five options are included for test 6. A block size of  $25 \times 25$  pixels, which corresponds to  $250 \text{ m} \times 250 \text{ m}$  regions of REAL backscatter images, was chosen to evaluate the performance of the CCA in the smallest of the block sizes used in the real-time operational version of our algorithm.

The multipass and multigrid interrogations contribute the most to the performance of the CCA, while the other options bring relatively small improvements. The histogram equalization and the Tukey window tend to slightly underestimate the  $u$  component but provide better estimation of the  $v$  component, making a correction of the wind direction estimation. On the other hand, the zero padding tends to reduce the underestimation of the  $u$  component and improve the estimation of the wind speed. Overall, the results suggest that the performance of the CCA for the relatively small region is optimized by applying all five options.

TABLE 1. The tests of the CCA using synthetic backscatter images and the velocity fields. The five options of the CCA are MP, MG, ZP, TW, and HE. Symbols  $\surd$  and  $\times$  represent that these options are turned on and off, respectively.

Options	MP	MG	HE	ZP	TW
Test 1	$\times$	$\times$	$\times$	$\times$	$\times$
Test 2	$\surd$	$\times$	$\times$	$\times$	$\times$
Test 3	$\surd$	$\surd$	$\times$	$\times$	$\times$
Test 4	$\surd$	$\surd$	$\surd$	$\times$	$\times$
Test 5	$\surd$	$\surd$	$\surd$	$\surd$	$\times$
Test 6	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$

TABLE 2. The results of the tests of the CCA using synthetic backscatter images for light (top two rows), moderate (middle two rows), and strong (bottom two rows) wind cases. The 2D synthetic velocity field and the test results are expressed in meters per second. The first and the second rows show the mean velocity and the standard deviation (SD), respectively, obtained from 100 estimations.

	Velocity field ( $\text{m s}^{-1}$ )	Test 1 ( $\text{m s}^{-1}$ )	Test 2 ( $\text{m s}^{-1}$ )	Test 3 ( $\text{m s}^{-1}$ )	Test 4 ( $\text{m s}^{-1}$ )	Test 5 ( $\text{m s}^{-1}$ )	Test 6 ( $\text{m s}^{-1}$ )
Mean	(1.027, 0.002)	(0.509, 0.008)	(0.563, 0.008)	(1.018, -0.001)	(0.988, 0.001)	(1.006, 0.001)	(1.008, 0.002)
SD		(0.282, 0.063)	(0.345, 0.060)	(0.192, 0.011)	(0.187, 0.007)	(0.004, 0.002)	(0.014, 0.011)
Mean	(5.811, 0.088)	(3.538, -0.005)	(4.282, -0.154)	(5.676, -0.036)	(5.648, 0.156)	(5.718, 0.154)	(5.608, 0.142)
SD		(1.891, 1.155)	(1.901, 2.202)	(0.394, 0.493)	(0.383, 0.165)	(0.309, 0.153)	(0.452, 0.191)
Mean	(11.79, 0.194)	(5.305, -0.134)	(6.575, 0.748)	(11.70, -0.672)	(11.64, 0.612)	(11.68, 0.766)	(11.32, 0.586)
SD		(6.654, 4.579)	(8.139, 17.39)	(0.798, 1.173)	(0.795, 0.658)	(0.752, 0.810)	(0.498, 0.749)

### b. Accuracy and precision of the wind estimates

The use of a Cartesian grid and a discrete time interval between scans tends to quantize the CCA velocity estimates. Fitting a 2D polynomial function to the  $5 \times 5$  domain around the peak of the CCFs, following Piironen and Eloranta (1995), ameliorates quantization somewhat, but it does not eliminate it. In our use of synthetic data with  $\delta x = \delta y = 10 \text{ m}$  and  $\delta t = 10 \text{ s}$ , we observed that the accuracy and precision of the CCA speed estimates oscillates as a function of the true mean speed, and it becomes increasingly pronounced as the flow field increases in uniformity. The period of the oscillations is  $1 \text{ m s}^{-1}$ , reflecting our choice of  $\delta x$ ,  $\delta y$  and  $\delta t$ . The experiments showed that the CCA tends to underestimate the velocity when the mean speed is

below the halfway point between intervals of  $\delta x/\delta t$ , and it tends to overestimate the velocity when the mean speed is above the halfway point between intervals of  $\delta x/\delta t$ . For wind speeds between 1 and  $2 \text{ m s}^{-1}$ , this can be as much as  $\pm 25\%$  relative error. Furthermore, for mean wind speeds near the halfway point between integer multiples of  $\delta x/\delta t$  ( $1.5, 2.5, 3.5 \text{ m s}^{-1}$ , etc.), the standard deviation of velocity estimates reaches local maxima as large as  $0.3 \text{ m s}^{-1}$ —indicating a loss of precision at those mean speeds. It appears that these errors, which oscillate with speed, are common as described by Chen and Katz (2005) and are highlighted in Fig. 1 in their paper. Chen and Katz (2005) offer an alternative method to achieve subpixel velocity resolution. Unfortunately, however, implementing their algorithm was not possible in our project.

By including turbulence (i.e., including spatial variability of the wind field) and gradually increasing the mean wind speed for neutral conditions, we observed that the oscillations in systematic errors become secondary to random errors resulting from the effects of a

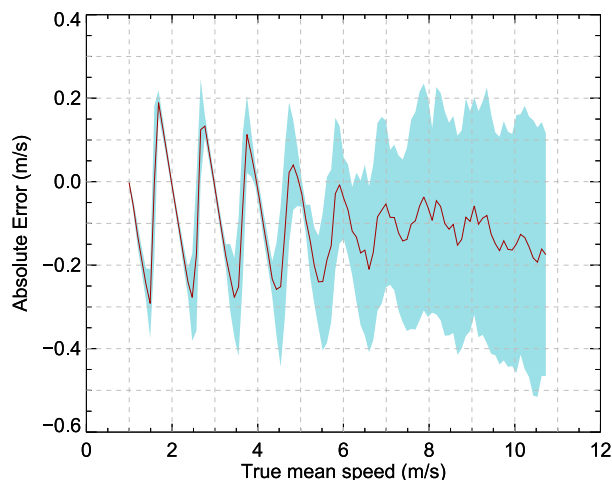


FIG. 6. Absolute error of the optimized CCA as a function of mean wind speed based on experiments with synthetic flow fields (neutral static stability) and synthetic aerosol distributions. The red line represents accuracy, or the systematic error associated with the mean of 100 trials for a given true mean speed. The light blue shaded region corresponds to one standard deviation and reflects the precision. The oscillations at slow true mean speeds are the result of the less-than-ideal performance of the polynomial fit used to obtain subpixel velocity resolution. The increasing standard deviation of the estimates with speed is the result of increasing turbulence intensity.

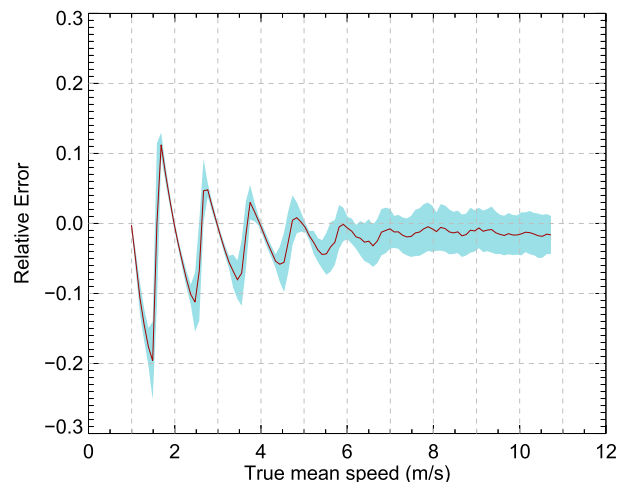


FIG. 7. Relative errors of the optimized CCA as a function of mean wind speed based on experiments with synthetic flow fields (neutral static stability) and synthetic aerosol distributions. These are the same data as plotted in Fig. 6, except they have been normalized by the corresponding true mean speed and are therefore unitless.

TABLE 3. The results of the tests of the optimized CCA using synthetic backscatter images for divergent flow, rotational flow, stretching deformation, and shearing deformation wind cases. The 2D synthetic velocity field and the test results are expressed in meters per second. The first and the second rows show the mean velocity and the SD, respectively, obtained from 100 estimations.

		Velocity field ( $\text{m s}^{-1}$ )	Optimized CCF result ( $\text{m s}^{-1}$ )
Divergent flow	Mean	(1.000, 0.000)	(0.967, -0.026)
	SD		(0.718, 0.454)
Rotational flow	Mean	(1.000, 0.000)	(0.816, 0.0795)
	SD		(0.733, 0.653)
Stretching deformation	Mean	(1.000, 0.000)	(0.875, 0.062)
	SD		(0.654, 0.498)
Shearing deformation	Mean	(1.000, 0.000)	(0.652, 0.0453)
	SD		(0.629, 0.510)

nonuniform wind field and the resulting deformation of aerosol features. The result is shown in Fig. 6. The red line was computed from  $(\bar{\mathbf{V}}_e - \|\bar{\mathbf{V}}_e\|)/\bar{\mathbf{V}}_t$ , where  $\|\bar{\mathbf{V}}_e\|$  represents the mean speed of 100 CCA estimates and  $\|\bar{\mathbf{V}}_t\|$  represents the true spatial mean speed in the  $25 \times 25$  pixel interrogation window. The light blue shaded region represents one standard deviation of the 100 CCA estimates. Figure 7 shows the same results scaled according to the true mean wind speed. It indicates that the systematic error is approximately 2% and the random error is approximately 6% at  $10 \text{ m s}^{-1}$  true mean speed.

The optimized CCA (with all five options) was also applied to the four flow structures that are typically observed in the lower *convective* boundary layer (divergent, rotational, stretching deformation, and shearing deformation). Table 3 lists the results of the tests for the four flows. By comparing test 6 of Table 2 and Table 3, one can see that the performance of the optimized CCA for convective flows was lower than all three cases of shear-induced turbulent flow. We attribute this to the diversity of wind directions within the interrogation window, sometimes displacing aerosol structures in opposing directions. In the unlikely case where a CCF contains two peaks of equal value, the algorithm will report only one. Therefore, when applied to regions of strong shear (i.e., an interrogation window with half easterly flow and half westerly flow as a worst case), the CCA will fail to provide the vector average of the true velocity field.

While the use of synthetic images and flow fields are powerful tools for testing the algorithm, they also have severe limitations and miss relevant physics that occur in the real world. Foremost, our 2D synthetic images and flow fields lack the realism of the 3D nature of aerosol and wind. In the real world, aerosol features may pass through the scan plane, resulting in false apparent motions. Also, in the real world, air parcels could move circuitously and not at constant velocity during the time between scans.

#### 4. Comparison with Doppler lidar wind measurements

As described by Mayor et al. (2012), a rudimentary CCA was applied to REAL data collected during the Canopy Horizontal Array Turbulence Study (CHATS; Patton et al. 2011), from March to June of 2007 near Dixon, California. The REAL was located 1.61 km north of the National Center for Atmospheric Research (NCAR) Integrated Surface Flux Facility (ISFF) 30-m vertical tower. The tower was surrounded by an orchard ( $800 \text{ m} \times 800 \text{ m}$ ) of walnut trees approximately 10 m tall. Five Campbell Scientific CSAT3 3D sonic anemometers were located on the tower at 12.5, 14.0, 18.0, 23.0,

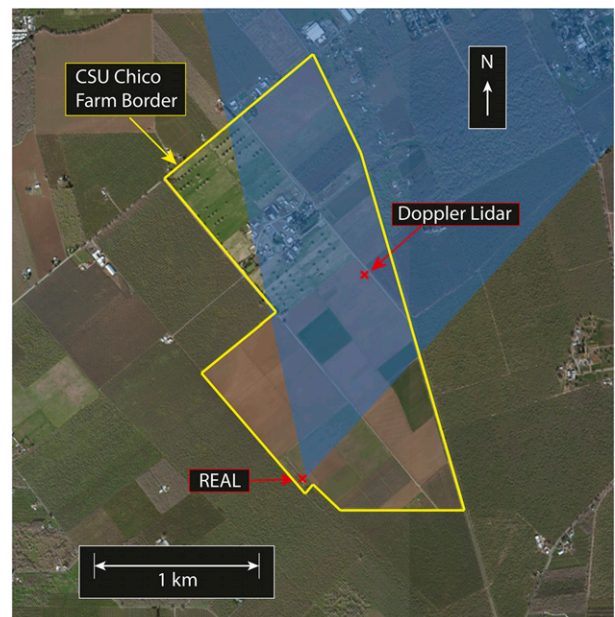


FIG. 8. A map showing the experimental setup at the University Farm at California State University, Chico, in 2013. The yellow lines represent the University Farm border. The two red symbols (×) represent the locations of the REAL system and the DL, respectively. The blue region represents the PPI scans collected by the REAL system for the experiment.

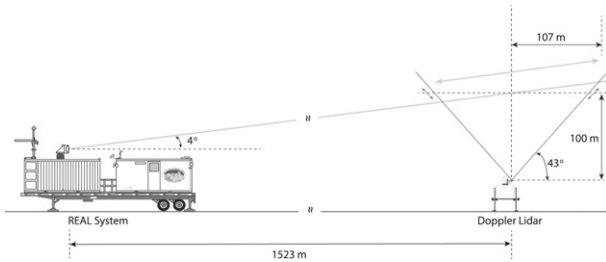


FIG. 9. Vertical cross-section diagram for the 2013 Chico field experiment. The REAL scans the atmosphere at  $4^\circ$  elevation. The DL is located 1523 m from the REAL and operated in vertical profile mode. With an elevation angle of  $43^\circ$ , the DL samples at 100 m AGL were 107 m from the center location.

and 29.0 m AGL to measure the wind velocity. The REAL scanned the atmosphere nearly horizontally over the orchard, and the rudimentary CCA was applied to estimate the wind velocity in a series of different sized blocks centered on the tower.

The REAL dataset from CHATS was groundbreaking but it had some deficiencies. In particular, the ISFF tower and nearby trees caused hard-target reflections, creating intermittent bright pixels and large shadows in the aerosol backscatter data that prevented optimal retrieval of wind fields by cross correlation. Moreover, the REAL platform settled into the soil during the experiment and precise measurements of the pitch and roll of the instrument were not available. This resulted in uncertainty of the altitude of the lidar scan at the location of the tower. Given the strong speed shear with increasing altitude in the roughness sublayer just above the top of the canopy, Mayor et al. (2012) chose not to calculate and compare temporal mean wind velocity data and instead focused on the instantaneous wind vectors resulting from the CCA.

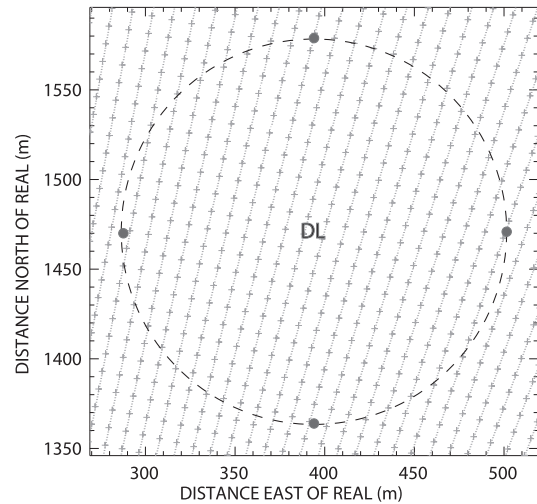


FIG. 10. Diagram of lidar data density in a  $250\text{ m} \times 250\text{ m}$  area at 100 m above the DL location. REAL aerosol backscatter (+) and DL radial velocity measurement ( $\bullet$ ).

To move forward, a new field experiment was conducted in Chico, California, from May of 2013 through January of 2014. Chico is 130 km north of Dixon and has less variable relative humidity than Dixon due to its distance from the Sacramento–San Joaquin River delta. However, nearby agriculture activities and boundary layer convection offer good conditions for testing. In the Chico experiment, a Halo Photonics Streamline Doppler lidar (serial number 0811–35, manufactured in November 2011) was employed as the reference system in order to avoid hard target reflections associated with mast-mounted in situ sensors. The DL operated at a wavelength of  $1.5\ \mu\text{m}$ , a pulse energy of  $20\ \mu\text{J}$ ,

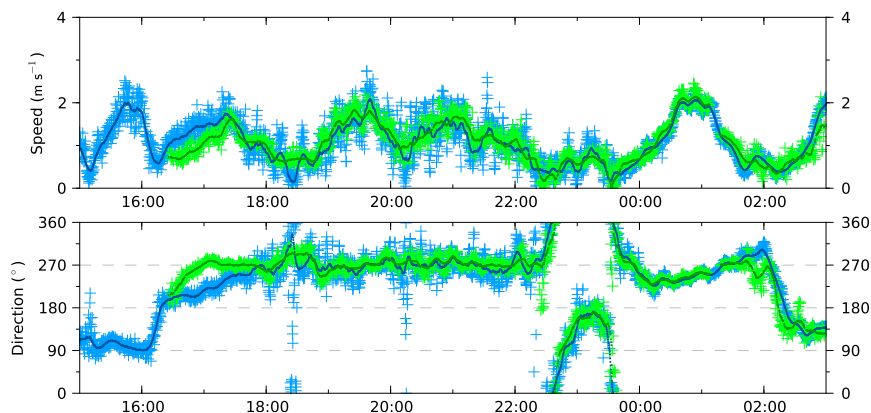


FIG. 11. Time series of wind speed and direction, as estimated by the DL (blue) and by the optimized CCA (green) from REAL backscatter images for a light wind case starting at 1500 UTC 23 Oct 2013. Each plus (+) represents an individual estimation or measurement, separated by approximately 17 s. Solid lines represent 10-min rolling averages.



TABLE 4. RMS deviation of differences, linear regression variables (slope, offset),  $R^2$ , number of points, and recovery percentage w.r.t. the DL reference for the 10-min averaged wind component  $u$  (west–east), for the three specific cases, and for the 15 days considered for the temporal validation.

Case	RMS ( $\text{m s}^{-1}$ )	Slope	Offset ( $\text{m s}^{-1}$ )	$R^2$	No. points	Recovery (%)
Light	0.17	0.984	0.079	0.950	64	88.9
Moderate	0.34	0.937	−0.06	0.971	72	100
Strong	0.51	0.947	0.16	0.961	72	100
15 days	0.36	0.974	−0.05	0.993	891	99.0

within a  $250\text{ m} \times 250\text{ m}$  block at the altitude of 100 m AGL. The distortion correction, as discussed by [Sasano et al. \(1982\)](#), was applied to all REAL backscatter images before the wind velocity estimation.

The quality of the aerosol backscatter data depends upon the performance of the instrument and the state of the atmosphere. In the following, for sake of brevity, only three cases featuring different wind conditions are presented as examples: low, moderate, and strong wind speeds. Then, a statistical analysis of 15 days of data that occurred in September and October of 2013 are considered. These 15 days present the best-quality aerosol backscatter data at the DL location during daytime while covering a broad range of wind speeds ( $0\text{--}16\text{ m s}^{-1}$ ) and constant to variable wind directions. Therefore, these days constitute the best dataset to analyze the performance of the algorithm while minimizing any weaknesses due to the instrument or nonoptimal atmospheric conditions. They were selected by analyzing the “image SNR” as detailed in [Dérian et al. \(2015\)](#).

#### 1) LIGHT WIND CASE

[Figure 11](#) shows time series of wind speeds and directions, as estimated by the DL (blue) and by the optimized CCA (green) from REAL backscatter images for a 12-h period starting from 1500 UTC 23 October 2013. Skies were clear with a maximum air temperature at the REAL of  $29.2^\circ\text{C}$  at 1632 Pacific daylight time (PDT) (2332 UTC). Range–height indicator (RHI) scans made by the REAL reveal a shallow convective boundary layer that did not exceed approximately 300-m depth during the day. Plus symbols represent the individual estimates and the line a 10-min rolling average.

This is an example of a light wind case where the wind speed ranges between 0 and  $2\text{ m s}^{-1}$  and the wind direction is variable. The results show that the time series of both methods are very similar except for a 2-h period between 1500 and 1700 UTC, where no coherent aerosol features were present. The correlation coefficients  $R^2$  for the 10-min averaged wind speed and direction between the optimized CCA and the DL are 0.829 and 0.658, respectively.

#### 2) MODERATE WIND CASE

[Figure 12](#) shows time series of wind speeds and directions, as estimated by the DL (blue) and by the optimized CCA (green) from REAL backscatter images for a 12-h period starting from 1500 UTC 17 September 2013. Skies were clear with a maximum air temperature at the REAL of  $27.3^\circ\text{C}$  at 1659 PDT (2359 UTC). RHI scans by the REAL show a mixed layer reaching approximately 1100 m deep in the afternoon. This is an example of a moderate wind case where the wind speed ranges between 0 and  $8\text{ m s}^{-1}$  and the wind direction is approximately constant for the first half of the period but varies after 2000 UTC. The time series of both methods are remarkably similar for both wind speed and direction. Correlation coefficients  $R^2$  for the 10-min averaged wind speed and direction between the optimized CCA and the DL are 0.973 and 0.938, respectively.

#### 3) STRONG WIND CASE

[Figure 13](#) shows the time series of wind speeds and directions, as estimated by the DL (blue) and by the optimized CCA (green) from REAL backscatter images for a 12-h period starting from 1500 UTC 9 October

TABLE 5. RMS deviation of differences, linear regression variables (slope, offset),  $R^2$ , number of points, and recovery percentage w.r.t. the DL reference for the 10-min averaged wind component  $v$  (south–north), for the three specific cases and the 15 days considered for the temporal validation.

Case	RMS ( $\text{m s}^{-1}$ )	Slope	Offset ( $\text{m s}^{-1}$ )	$R^2$	No. points	Recovery (%)
Light	0.35	0.504	0.020	0.458	64	88.9
Moderate	0.28	0.984	0.01	0.987	72	100
Strong	0.44	0.885	−0.76	0.879	72	100
15 days	0.37	0.991	0.06	0.995	891	99.0

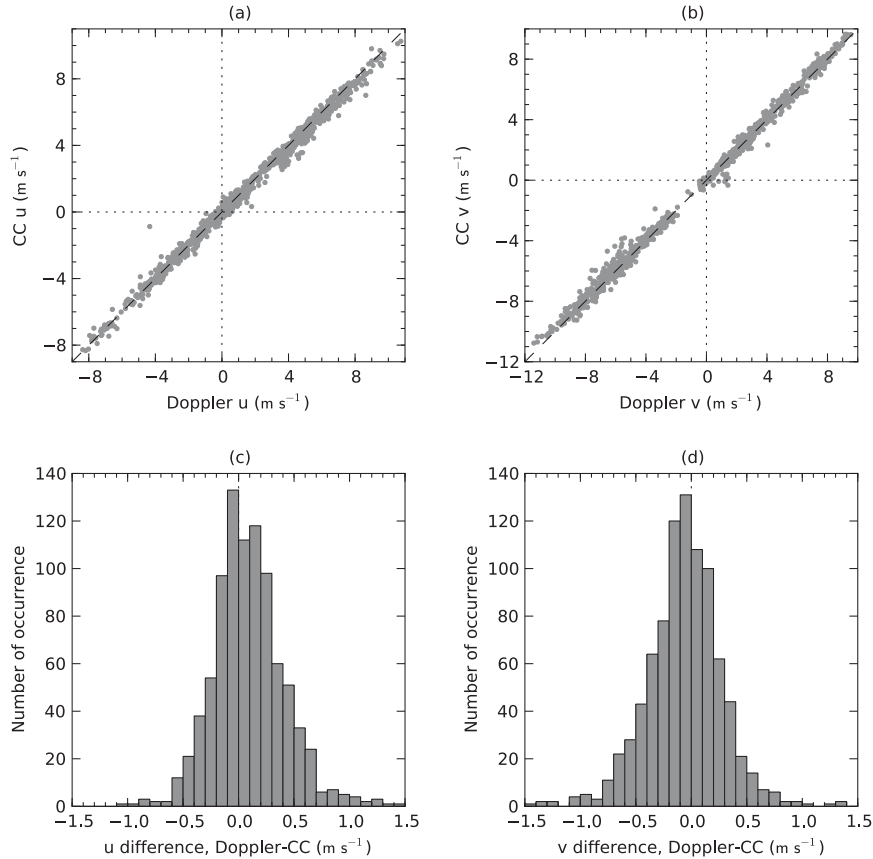


FIG. 14. (top) Scatterplots of 10-min averaged  $u$  and  $v$  components of the wind velocity estimated by the optimized CCA (vertical axis) vs that estimated by the DL at 100 m AGL (horizontal axis), for 15 days, during daytime (891 intervals). The histogram distributions of differences for the same dataset are shown in the bottom panels.

2013. Skies were clear with a maximum air temperature at the REAL of 24.2°C at 1537 PDT (2237 UTC). RHI scans made by the REAL show a mixed layer approximately 1 km deep. This is an example of a strong wind case where the average wind speed is about 10 m s<sup>-1</sup> (with gusts to 14 m s<sup>-1</sup>) and the wind direction is approximately constant (from the northwest direction). As in the moderate wind case, the time series of both methods are strikingly similar for both wind speed and direction. Correlation coefficients  $R^2$  for the 10-min averaged wind speed and direction between the optimized CCA and the DL are 0.929 and 0.968, respectively.

#### 4) STATISTICAL ANALYSIS OF TEMPORAL VALIDATION

Statistical comparisons between optimized CCA and the DL are summarized in Tables 4 and 5 for all three cases (weak, moderate, and strong winds). Figure 14 presents scatterplots for the 10-min means collected over 15 days (891 intervals), and the corresponding statistical results are available in Tables 4 and 5. The root-mean-square (RMS)

deviations for components  $u$  and  $v$  over 15 days are 0.36 m s<sup>-1</sup> and 0.37 m s<sup>-1</sup>, respectively. The scatterplots show excellent agreement between the cross-correlation motion estimates and the DL measurements. This is confirmed by the slopes of the best-fit lines (0.974 for  $u$  and 0.991 for  $v$ ) and  $R^2$  coefficients (0.993 for  $u$  and 0.995 for  $v$ ).

The time series of Figs. 11–13 reveal that the wind velocities obtained from the cross correlation have less variability than the DL measurements. Figure 15 shows scatterplots of the turbulent kinetic energy (TKE) measured by the DL and the CCA over the 891 ten-minute intervals. Three sets of results are presented for the CCA, corresponding to three levels of the multigrid estimation: 1000 m × 1000 m, 500 m × 500 m, and 250 m × 250 m. In all three sets, the cross-correlation method underestimated the TKE. However, as the block size is reduced, more TKE is recovered: from ≈25% with the largest blocks to ≈39% with 250 m × 250 m blocks. These results support the notion that smaller block sizes are able to capture smaller-scale velocity perturbations. As stated in Mayor et al. (2012), estimating the wind

directly from small block sizes leads to noisier results. However, with the addition of multipass, multigrid, and quality control, the algorithm is now able to increase the resolution of the motion field.

### b. Wind velocity fields

Two-component velocity *fields* can be retrieved from the REAL backscatter images to a range of several kilometers via application of the optimized CCA. Each vector in the field is the result of placing the interrogation window on the region of the preprocessed backscatter intensity image surrounding the desired vector location. Figure 16 shows an example of a strong wind case on 3 October 2013 when the wind direction is relatively uniform. RHI scans near this time indicate mixing to altitudes above 1 km AGL. Mean wind speed at 100 m AGL at the location of the DL was about  $13 \text{ m s}^{-1}$ . Skies were clear and the high temperature on this day was  $24^\circ\text{C}$ . This is an example of a case where there are likely significant contributions from both shear and buoyancy. Figure 17 shows an example of a microscale vortex observed from the light wind case (23 October 2013) discussed in a previous section. The optimized CCA greatly improves the resolution of such velocity structures.

### c. Spatial validation

For this phase of the experiment, the DL was located on the roof of the REAL system and operated in fixed-beam mode pointing at  $45^\circ$  azimuth and  $2^\circ$  elevation to estimate the radial component of the wind velocity field at the center of the scan sector swept by the REAL. The corresponding radial component of the wind velocity was retrieved from the horizontal components of the wind velocity vector estimated by the optimized cross-correlation algorithm applied to REAL backscatter images.

The radial components of the wind velocity vectors at  $45^\circ$  azimuth and  $2^\circ$  elevation were retrieved by the algorithm with all options (MP, MG, HE, ZP, and TW) applied to image blocks ( $250 \text{ m} \times 250 \text{ m}$ ) in the range between 0.5 and 3 km from the REAL system. The REAL system scanned the atmosphere every 17 s (between  $15^\circ$  and  $75^\circ$  azimuth at  $2^\circ$  elevation). The radial component of the wind vectors as a function of time and range were compared. This experiment was conducted in December 2013 and January 2014.

As an example, Fig. 18 shows the radial component of the wind velocity as measured by the DL and estimated by the optimized cross-correlation algorithm, for an 8-h period starting at 1700 UTC 8 January 2014. The intensity of solar radiation on this day was weak due to partly sunny skies and the low sun angle associated with the winter season. The high temperature was about  $13^\circ\text{C}$ .

Wind speeds at 100 m did not exceed approximately  $4 \text{ m s}^{-1}$ . The two radial velocity fields shown in the figure bear strong resemblance.

Statistics on the 10-min mean radial velocities, computed at different ranges, are presented in Fig. 19. These results were computed over 8-h periods (from 1700 to 0100 UTC) for 8 days selected in December 2013 and January 2014. It shows that  $R^2$  (Fig. 19d) remains above 0.97 until the 1.2-km range and then it decreases with the range due to the decaying SNR for both instruments. The scatterplot of radial velocities (Fig. 19a) indicates that the cross correlations gradually overestimate the radial velocities as the range increases. This is confirmed by the histogram of velocity differences (Fig. 19b), biased toward negative values, as well as the slopes of linear regressions (Fig. 19c). A similar trend was found for the same dataset using a different motion estimation method (Dérian et al. 2015). This likely indicates a slight misalignment of both instrument beams. A mismatch in the elevation angle would result in a difference in altitude of the beams that increases with the range, thus explaining the lower velocities measured by the DL.

## 5. Conclusions

This paper describes the results of a research program that utilized two very different approaches to characterize and improve the performance of the CCA as applied to elastic lidar data for remote wind estimation. The first approach, described in detail in Hamada (2014), used synthetic aerosol backscatter fields and synthetic turbulent velocity fields to conduct highly controlled numerical experiments. By using synthetic images and velocity fields, we were able to evaluate the significance of the various procedures included in the algorithm on the resulting motion vectors. An operational real-time version of the optimized CCA is now available for use with the REAL. Moreover, the synthetic test results enabled us to estimate the accuracy and precision of the technique. Results suggest that when the turbulence is shear-generated under neutral conditions, systematic errors dominate at slow mean speeds ( $\leq 5 \text{ m s}^{-1}$ ) and random errors dominate at high mean speeds ( $\geq 5 \text{ m s}^{-1}$ ). The systematic errors are due to the less-than-ideal performance of the polynomial fit that is intended to eliminate velocity quantization. The random errors are due to the turbulence and the resulting distortion of aerosol features. When the boundary layer is driven by convection, the flow structures become cellular. To understand the possible errors that occur under these conditions, we used the flow fields of the four elemental kinematic fluid properties. In these cases, the flow direction changes dramatically in the interrogation









- single aerosol lidar data. *J. Atmos. Oceanic Technol.*, **32**, 1759–1778, doi:10.1175/JTECH-D-15-0010.1.
- Durbin, P. A., and B. A. P. Reif, 2010: Rapid distortion theory. *Statistical Theory and Modeling for Turbulent Flows*, 2nd ed. John Wiley & Sons, Ltd., 281–301, doi:10.1002/9780470972076.ch11.
- Eloranta, E. W., J. M. King, and J. A. Weinman, 1975: The determination of wind speeds in the boundary layer by monostatic lidar. *J. Appl. Meteor.*, **14**, 1485–1489, doi:10.1175/1520-0450(1975)014<1485:TDOWSI>2.0.CO;2.
- Emery, W. J., D. Baldwin, and D. Matthews, 2003: Maximum cross correlation automatic satellite image navigation and attitude corrections for open-ocean image navigation. *IEEE Trans. Geosci. Remote Sens.*, **41**, 33–42, doi:10.1109/TGRS.2002.808061.
- García-Pereda, J., and R. Borde, 2014: The impact of the tracer size and the temporal gap between images in the extraction of atmospheric motion vectors. *J. Atmos. Oceanic Technol.*, **31**, 1761–1770, doi:10.1175/JTECH-D-13-00235.1.
- Hamada, M., 2014: Evaluations of the performance of a cross-correlation algorithm for wind velocity estimation using synthetic backscatter lidar images and velocity fields. M.S. thesis, Dept. of Geological and Environmental Sciences, California State University, Chico, 202 pp.
- Held, A., T. Seith, I. M. Brooks, S. J. Norris, and S. D. Mayor, 2012: Intercomparison of lidar aerosol backscatter and in-situ size distribution measurements. *European Aerosol Conf.*, Granada, Spain, European Aerosol Assembly, B-WG01S2P05.
- Holton, J. R., and G. J. Hakim, 2013: *An Introduction to Dynamic Meteorology*. 5th ed. Academic Press, 532 pp.
- Kanak, K. M., 2005: Numerical simulation of dust devil-scale vortices. *Quart. J. Roy. Meteor. Soc.*, **131**, 1271–1292, doi:10.1256/qj.03.172.
- Khanna, S., and J. G. Brasseur, 1998: Three-dimensional buoyancy- and shear-induced local structure of the atmospheric boundary layer. *J. Atmos. Sci.*, **55**, 710–743, doi:10.1175/1520-0469(1998)055<0710:TDBASI>2.0.CO;2.
- Kolev, I., O. Parvanov, and B. Kaprielov, 1988: Lidar determination of winds by aerosol inhomogeneities: Motion velocity in the planetary boundary layer. *Appl. Opt.*, **27**, 2524–2531, doi:10.1364/AO.27.002524.
- Leese, J. A., C. S. Novak, and B. B. Clark, 1971: An automated technique for obtaining cloud motion from geosynchronous satellite data using cross correlation. *J. Appl. Meteor.*, **10**, 118–132, doi:10.1175/1520-0450(1971)010<0118:AATFOC>2.0.CO;2.
- Mann, J., 1994: The spatial structure of neutral atmospheric surface-layer turbulence. *J. Fluid Mech.*, **273**, 141–168, doi:10.1017/S0022112094001886.
- , 1998: Wind field simulation. *Probab. Eng. Mech.*, **13**, 269–282, doi:10.1016/S0266-8920(97)00036-2.
- Mauzey, C. F., J. P. Lowe, and S. D. Mayor, 2012: Real-time wind velocity estimation from aerosol lidar data using graphics hardware. *GPU Technology Conf.*, San Diego, CA, NVIDIA, AV10. [Available online at [http://on-demand.gputechconf.com/gtc/2012/posters/P0501\\_cmauzey\\_jlowe\\_sdmayor\\_gtc\\_2012\\_poster\\_draft\\_final.pdf](http://on-demand.gputechconf.com/gtc/2012/posters/P0501_cmauzey_jlowe_sdmayor_gtc_2012_poster_draft_final.pdf).]
- Mayor, S. D., and E. W. Eloranta, 2001: Two-dimensional vector wind fields from volume imaging lidar data. *J. Appl. Meteor.*, **40**, 1331–1346, doi:10.1175/1520-0450(2001)040<1331:TDVWFF>2.0.CO;2.
- , and S. M. Spuler, 2004: Raman-shifted eye-safe aerosol lidar. *Appl. Opt.*, **43**, 3915–3924, doi:10.1364/AO.43.003915.
- , —, B. M. Morley, and E. Loew, 2007: Polarization lidar at 1.54- $\mu\text{m}$  and observations of plumes from aerosol generators. *Opt. Eng.*, **46**, 096201, doi:10.1117/1.2786406.
- , J. P. Lowe, and C. F. Mauzey, 2012: Two-component horizontal aerosol motion vectors in the atmospheric surface layer from a cross-correlation algorithm applied to scanning elastic backscatter lidar data. *J. Atmos. Oceanic Technol.*, **29**, 1585–1602, doi:10.1175/JTECH-D-11-00225.1.
- Moeng, C.-H., and P. P. Sullivan, 1994: A comparison of shear- and buoyancy-driven planetary boundary layer flows. *J. Atmos. Sci.*, **51**, 999–1022, doi:10.1175/1520-0469(1994)051<0999:ACOSAB>2.0.CO;2.
- Murray, J. C., H. R. Erwin, and S. Wermter, 2009: Robotic sound-source localisation architecture using cross-correlation and recurrent neural networks. *Neural Networks*, **22**, 173–189, doi:10.1016/j.neunet.2009.01.013.
- Patton, E. G., and Coauthors, 2011: The Canopy Horizontal Array Turbulence Study. *Bull. Amer. Meteor. Soc.*, **92**, 593–611., doi:10.1175/2010BAMS2614.1.
- Piironen, A. K., and E. W. Eloranta, 1995: Accuracy analysis of wind profiles calculated from volume imaging lidar data. *J. Geophys. Res.*, **100**, 25 559–25 567, doi:10.1029/94JD02605.
- Raffel, M., C. E. Willert, S. Wereley, and J. Kompenhans, 2007: *Particle Image Velocimetry: A Practical Guide*. 2nd ed. Springer, 448 pp.
- Rinehart, R. E., and E. T. Garvey, 1978: Three-dimensional storm motion detection by conventional weather radar. *Nature*, **273**, 287–289, doi:10.1038/273287a0.
- Sasano, Y., H. Hirohara, T. Yamasaki, H. Shimizu, N. Takeuchi, and T. Kawamura, 1982: Horizontal wind vector determination from the displacement of aerosol distribution patterns observed by a scanning lidar. *J. Appl. Meteor.*, **21**, 1516–1523, doi:10.1175/1520-0450(1982)021<1516:HWVDFT>2.0.CO;2.
- Sathe, A., and J. Mann, 2013: A review of turbulence measurements using ground-based wind lidars. *Atmos. Meas. Tech.*, **6**, 6815–6871, doi:10.5194/amtd-6-6815-2013.
- , —, J. Gottschall, and M. S. Courtney, 2011: Can wind lidars measure turbulence? *J. Atmos. Oceanic Technol.*, **28**, 853–868, doi:10.1175/JTECH-D-10-05004.1.
- Schmidt, H., and U. Schumann, 1989: Coherent structures of the convective boundary layer derived from large-eddy simulations. *J. Fluid Mech.*, **200**, 511–562, doi:10.1017/S0022112089000753.
- Schols, J. L., and E. W. Eloranta, 1992: The calculation of area-averaged vertical profiles of the horizontal wind velocity from volume imaging lidar data. *J. Geophys. Res.*, **97**, 18 395–18 407, doi:10.1029/92JD01051.
- Schubert, A., A. Faes, A. Kääb, and E. Meier, 2013: Glacier surface velocity estimation using repeat TerraSAR-X images: Wavelet-vs. correlation-based image matching. *ISPRS J. Photogramm. Remote Sens.*, **82**, 49–62, doi:10.1016/j.isprsjprs.2013.04.010.
- Shimizu, H., Y. Sasano, Y. Yasuoka, H. Ueda, N. Takeuchi, and M. Okuda, 1981: The remote measurement of wind direction and velocity by a laser radar using the spatial correlation technique. *Oyobutsuri*, **50**, 616–620.
- Spuler, S. M., and S. D. Mayor, 2005: Scanning eye-safe elastic backscatter lidar at 1.54 microns. *J. Atmos. Ocean. Technol.*, **22**, 696–703, doi:10.1175/JTECH1755.1.
- Tukey, J. W., 1967: An introduction to the calculations of numerical spectrum analysis. *Spectral Analysis of Time Series*, B. Harris, Ed., John Wiley and Sons, 25–46.
- Westerweel, J., 2000: Theoretical analysis of the measurement precision in particle image velocimetry. *Exp. Fluids*, **29**, S003–S012, doi:10.1007/s003480070002.
- , and F. Scarano, 2005: Universal outlier detection for PIV data. *Exp. Fluids*, **39**, 1096–1100, doi:10.1007/s00348-005-0016-6.