

Radio Science

RESEARCH ARTICLE

10.1029/2019RS006892

Key Points:

- Radar antenna beam sidelobe contaminations cause dual-peaked Doppler spectra to be detected in Jicamarca radar MST probing
- Sorting the mean Doppler shifts and related power of detected spectral peaks into mainlobe and sidelobe sets is required in wind estimation
- We use weighted least squares fits of dual-peaked spectra followed by sorting and classification steps to obtain bias free wind estimates

Supporting Information:

- Figure S1
- Figure S2
- Figure S3
- Figure S4
- Figure S5
- Figure S6
- Figure S7
- Figure S8
- Figure S9
- Figure S10
- Figure S11
- Figure S12
- Figure S13

Correspondence to:

E. Kudeki,
erhan@illinois.edu

Citation:

Lee, K., Kudeki, E., Reyes, P. M., Lehman, G. A., & Milla, M. (2019). Mesospheric wind estimation with the Jicamarca MST radar using spectral main lobe identification. *Radio Science*, 54, 1222–1239. <https://doi.org/10.1029/2019RS006892>

Received 27 MAY 2019

Accepted 21 OCT 2019

Accepted article online 31 OCT 2019

Published online 10 DEC 2019

©2019. American Geophysical Union.
All Rights Reserved.

Mesospheric Wind Estimation With the Jicamarca MST Radar Using Spectral Mainlobe Identification

Kiwook Lee¹ , Erhan Kudeki¹ , Pablo M. Reyes^{1,2}, Gerald A. Lehman³ ,
and Marco Milla⁴ 

¹Department of Electrical and Computer Engineering, University of Illinois, Urbana-Champaign, Urbana, IL, USA,

²Now at Center for Geospace Studies, SRI International, Menlo Park, CA, USA, ³Department of Physics and Astronomy,

Clemson University, Clemson, SC, USA, ⁴Radio Observatorio de Jicamarca, Instituto Geofísico del Perú, Lima, Peru

Abstract MST (mesosphere, stratosphere, troposphere) radar observations at Jicamarca use four antenna beams, one vertical, others tilted to the east, west, and south, to detect the scattered pulse returns from mesospheric heights (~55–85 km). Doppler shifts of scattered returns, estimated by fitting the observed signal spectra by generalized Gaussian-shaped models, are used to estimate mesospheric wind vectors. At some heights two spectral peaks are seen in which case a dual-peaked model is fitted the spectrum. Dual peaks are more common for returns from the east and west tilted beams with stronger sidelobes. When sidelobe-caused peaks are dominant and are mistaken for mainlobe peaks, wind errors occur since the estimation algorithm uses the pointing angle of the mainbeam. To avoid such errors we implemented a clustering-based machine learning procedure to identify and use only the mainbeam components of dual-peaked spectra. Wind estimates made before and after the procedure will be presented to assess the improvements achieved by this new method to be used routinely in Jicamarca mesospheric wind measurements and applied to past MST data.

1. Introduction

In MST radar (mesosphere, stratosphere, troposphere remote probing technique using large VHF radars with “MW-acre” class power-aperture products) observations of the mesosphere at the Jicamarca Radio Observatory (JRO), four antenna beams, three of which are tilted off-zenith in east, west, and south directions, are used to detect the scattered pulse returns from mesospheric turbulence layers. In this paper we will describe a new analysis procedure developed for Jicamarca MST radar data collected with this quad-beam observation mode to obtain reliable vector wind velocity estimates free of errors and ambiguities that can result from sidelobe contaminations. The procedure is based on a maximum-likelihood fitting of the measured MST signal spectra by dual-peaked generalized Gaussian models, which is an extension of the least squares fitting techniques described in Sheth et al. (2006) and Smith (2014), but it also includes a novel and machine learning-based spectral main lobe identification feature to eliminate antenna sidelobe-caused ambiguities.

For the essentials of the MST radar technique (and how it differs from the incoherent scatter radar (ISR) technique) we refer the reader to Sheth et al. (2006), and references therein. An important point underlined in Sheth et al. (2006), that is crucial to what we are describing here, is that for MST radars the shape of the backscattered signal spectrum is largely determined by the antenna beam shape of the radar as a consequence of a “beam broadening” mechanism (as opposed to the thermodynamics of the scattering plasma which is the dominant factor for ISR technique). When the beam broadening mechanism (e.g., Kudeki et al., 1990) dominates, the MST signal spectrum becomes a rescaled version of the antenna “two-way” angular beam function (the product of the radar antenna transmission gain and receiving area functions in 2-D) collapsed to 1-D in the direction of the prevailing mesospheric wind advecting the electron density irregularities that cause the backscattering of the transmitted radar pulse. The modeling details of this mapping (including “aspect sensitivity” caused modifications that may be important) can be found in MST radar literature—see Kudeki et al. (1990) for a comprehensive description—but it is sufficient for our descriptions in this paper to understand that MST radar returns would be expected to display single peaked Doppler spectra if the two-way radar beam were free of sidelobes, and, conversely, display multi-peaked shapes if/when beam sidelobes were nonnegligible. Since practical MST radar probing beams used in Jicamarca

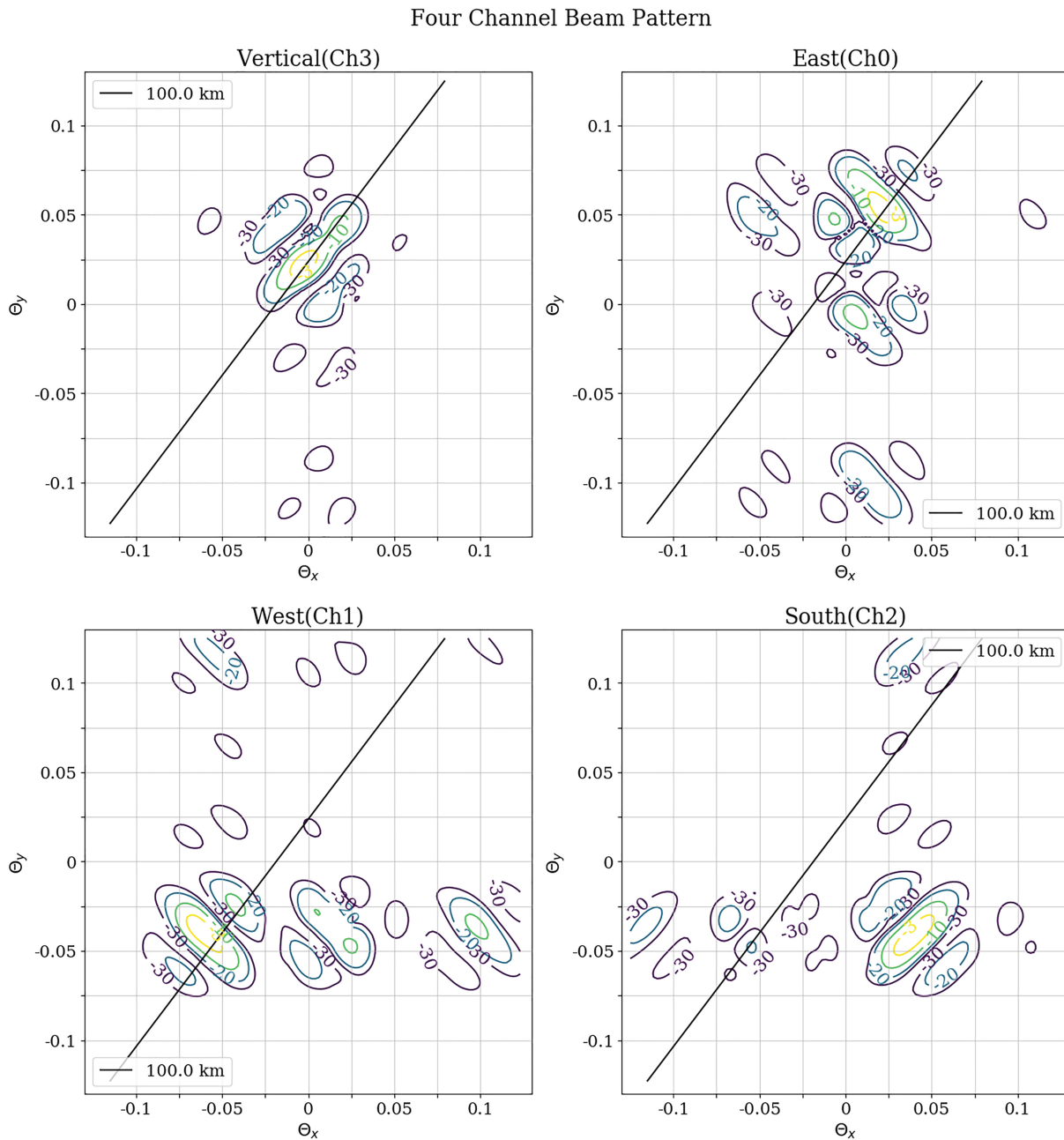


Figure 1. Antenna beam pattern designed for 2015 and 2016 where yellow contour line indicates main lobe beam for each channel, and black line indicates the loci of perpendicularity to the geomagnetic field at 100 km of altitude.

measurements have a variable but nonnegligible sensitivity in angular directions outside the intended main beams (as shown in Figure 1), the MST Doppler spectra recorded at Jicamarca are frequently observed to exhibit dual-peaked shapes.

The improved MST wind estimation method for Jicamarca to be described here works briefly as follows: Least squares fitting provides single or dual Doppler velocity shift estimates depending on whether a single- or dual-peaked spectral model best fits the observed MST spectrum. The fitted Doppler spectral shifts are also accompanied by spectral width and “shape index” estimates as well as uncertainty estimates accompanying each of the fit parameters. Signal power is simply the area under the measured spectrum minus the noise level fit estimate, but it can be partitioned into “main beam” and “sidelobe” power components by calculating separately the “areas” under the fitted dual Gaussians.

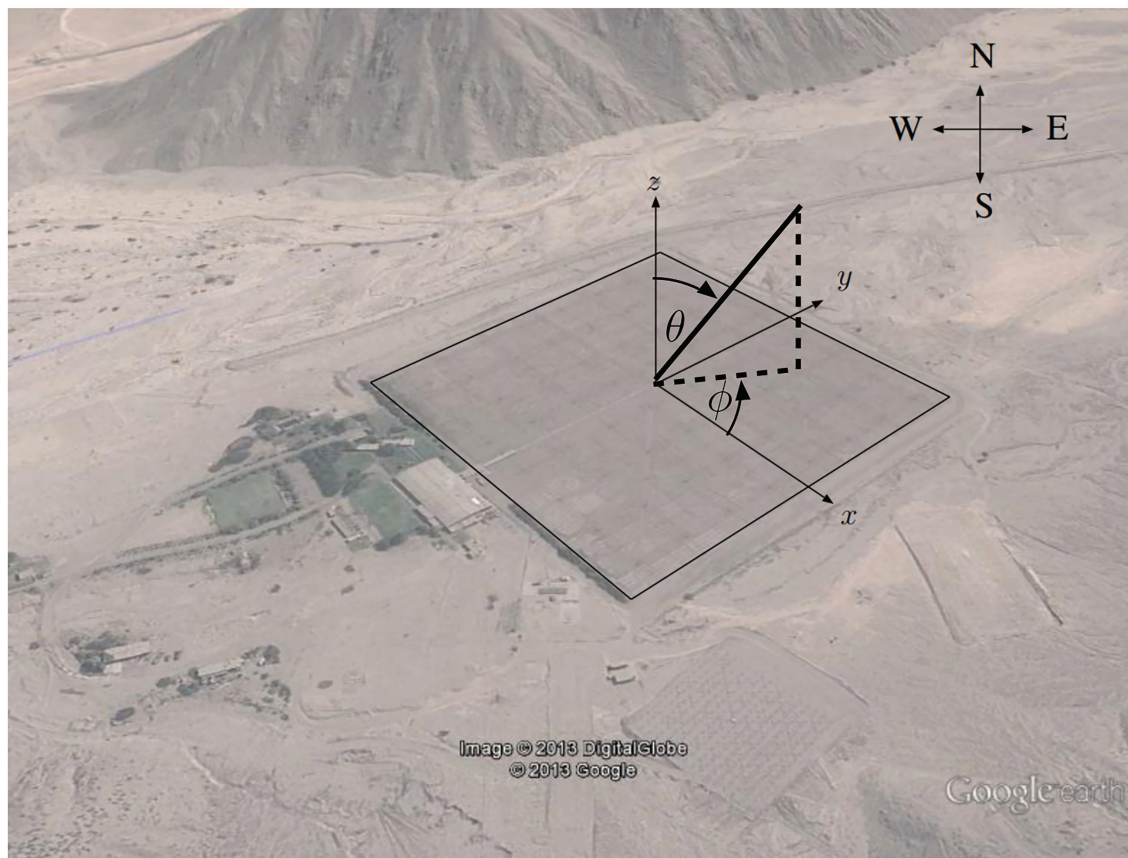


Figure 2. Jicamarca MST radar geometry.

The sorting of the dual peak fitting results (and associated powers) is accomplished with a “clustering step” of mixed data sets containing all the fitted Doppler shifts obtained for each radar range gate, followed by a procedure in which the clusters from all of the range gates are classified into separate main beam and side-lobe sets. Vector wind velocity estimates are then computed using standard least squares methods with the line-of-sight Doppler velocity data picked from the “main beam sets” belonging to “equal heights” (making a careful distinction between “height” and “range gate” on individual antenna beams). Velocity maps organized by observation times and altitudes are finally produced for the geophysically relevant u , v , w wind components to accompany the scattered power maps (RTI plots) produced separately from the main lobe and sidelobe sets—these power maps are in effect the “reliability measures” of our wind estimates derived from the main beam sets.

This paper is organized as follows: Section 2 provides a general overview of the Jicamarca MST radar MST-ISR mode (and its MST-ISR-EEJ extension) and the spectral estimation method used with the raw voltage data collected in this mode. Section 3 describes the dual-peaked least squares fitting of the quad-beam spectral estimates obtained in MST-ISR experiments. Clustering and classification of the fit results into main beam and sidelobe sets are described in section 4. Section 5 presents examples of mesospheric velocity and power maps extracted from April 2015 MST-ISR-EEJ data and a discussion of our new wind estimation method. Conclusions of this study and our plans for disseminating the derived wind estimates via the Madrigal database are presented in section 6.

2. MST Radar Observations at Jicamarca

The MST radar measurements discussed in this paper make use of the MST-ISR observational mode established at Jicamarca in late 2004 (Akiray, 2007; Lehmacher et al., 2007) and continued beyond 2011 in its MST-ISR-EEJ extension (e.g., Lehmacher et al., 2019). Both of these modes make use of three distinct antenna beams tilted toward the east, west, and south of the antenna array unit normal (called *on-axis*

Table 1

List of Main Beam and Major Sidelobe (-10 dB Strength) Pointing Directions in JRO MST-ISR Experiments Conducted in 2015 and 2016—Direction Cosines Θ_x and Θ_y (see text) and Line-of-Sight (LOS) Vector Projection Components of Main Beams and Sidelobes in Geodetic (E, N, U) Directions (East, North, Up)

Beam	Θ_x	Θ_y	E	N	U
East main lobe	0.0204	0.0545	0.03535	0.0029	0.9994
West main lobe	−0.0603	−0.0406	−0.0893	0.0042	0.9960
South main lobe	0.0397	−0.0412	−0.0259	−0.0732	0.9970
Vertical main lobe	−0.0003	0.0226	−0.0039	−0.0002	0.99992
East side lobe	0.0059	−0.0072	−0.0214	−0.0255	0.9994
West side lobe	0.0977	−0.0365	0.0148	−0.1147	0.9933

direction from hereon) with the two-way beam patterns depicted in Figure 1 as contour plots, plus a fourth antenna beam pointing in the vertical direction slightly to the north-east of the on-axis direction. These two-way radar beam intensities are shown in Figure 1 as a function of *direction cosines* $\Theta_x \equiv \sin \theta \cos \phi$ and $\Theta_y \equiv \sin \theta \sin \phi$, where θ and ϕ are zenith and azimuth angles, respectively, in the local coordinate system depicted in Figure 2, where the z axis coincides with the on-axis direction of the JRO antenna array. The patterns shown in Figure 1 display substantial sidelobe structures particularly for the beams with east and west tilted main beams or “main lobes”. The direction cosines of the four main lobes shown in Figure 1 as well as the strongest sidelobes accompanying the east and west main lobes are listed in Table 1—these values correspond to the 2-D moments of the two-way beam pattern functions (gain $\times$ effective area products) computed over the main lobes and sidelobes, respectively.

In MST-ISR observational mode the radar transmitter cycles between distinct pulse sets designed for mesospheric (MST) and ionospheric (ISR) probing in a tightly interleaved schedule—ISR samples are collected to enable absolute calibration of the MST radar returns as described in Akgiray (2007), Guo et al. (2007), and Lehman et al. (2007). MST probing in this mode uses 64-baud complementary coded pulse pairs with 49.92 MHz carrier frequencies and 1 μ s baud lengths and an interpulse period of 202.5 μ s (using the usual conversion factor of 150 km/ms). A group of 20 such pulses are transmitted before ISR pulsing mode is entered to return back to the next MST pulse set after a duration of 20,250 μ s (135 ms).

The received MST pulse returns are decoded and *coherently* integrated over 20 IPPs, and subsequently 64-point time series consisting of $N_c = 64$ coherently integrated samples (one complex voltage “ $I + jQ$ ” per MST pulse set per sampled radar range) are fast Fourier transformed to produce periodograms that are then *incoherently* integrated (e.g., Farley, 1985) over a 60.48 s interval to obtain the MST *power spectral density* (PSD) or *power spectrum* estimates like those shown in Figure 3 in black (dots and dashes). These spectra have a Nyquist Doppler bandwidth of $v_{\max} = \lambda_B f_{\max}$ (11.11 m/s), where λ_B is the Bragg wavelength (3 m at 50 MHz) and $f_{\max}$ is the Nyquist bandwidth (3.7 Hz), the inverse of twice the sampling interval of $T_s = 135$ ms. Doppler velocity resolution (spacing between dots in the figure) of these spectra is $\Delta v = 2v_{\max}/N_c$ (0.347 m/s). The blue and red dashed curves shown in Figure 3 correspond to general Gaussian fits that we will be describing in section 3.

Assuming that the fitting procedure provides the estimates of the so-called “radial Doppler velocity” of the scattering targets (Bragg scale refractive index fluctuations in the probed volume) contained within the main beam (as opposed to side lobes), four radial velocity estimates V_b obtained from four MST-ISR beams $b \in E, W, S, V$ can be combined to estimate the (v_x, v_y, v_z) components (defined in the coordinate system of Figure 2) of the vector wind velocity in the probed region, to be translated to the geophysically relevant (u, v, w) components as described in section 4. Wind analysis of data taken in the MST-ISR-EEJ mode also works similarly—in this extension of MST-ISR, a third burst of short pulses is included for probing the electrojet altitudes without saturating the radar, but MST probing details remain the same as in MST-ISR observations (see Lehman et al., 2019).

3. Spectral Fits of Jicamarca MST Returns

The first step in Jicamarca MST wind estimation is to least squares fit the *measured* Doppler spectra with appropriately chosen spectral models (assumed to include convolutional distortions of the intrinsic

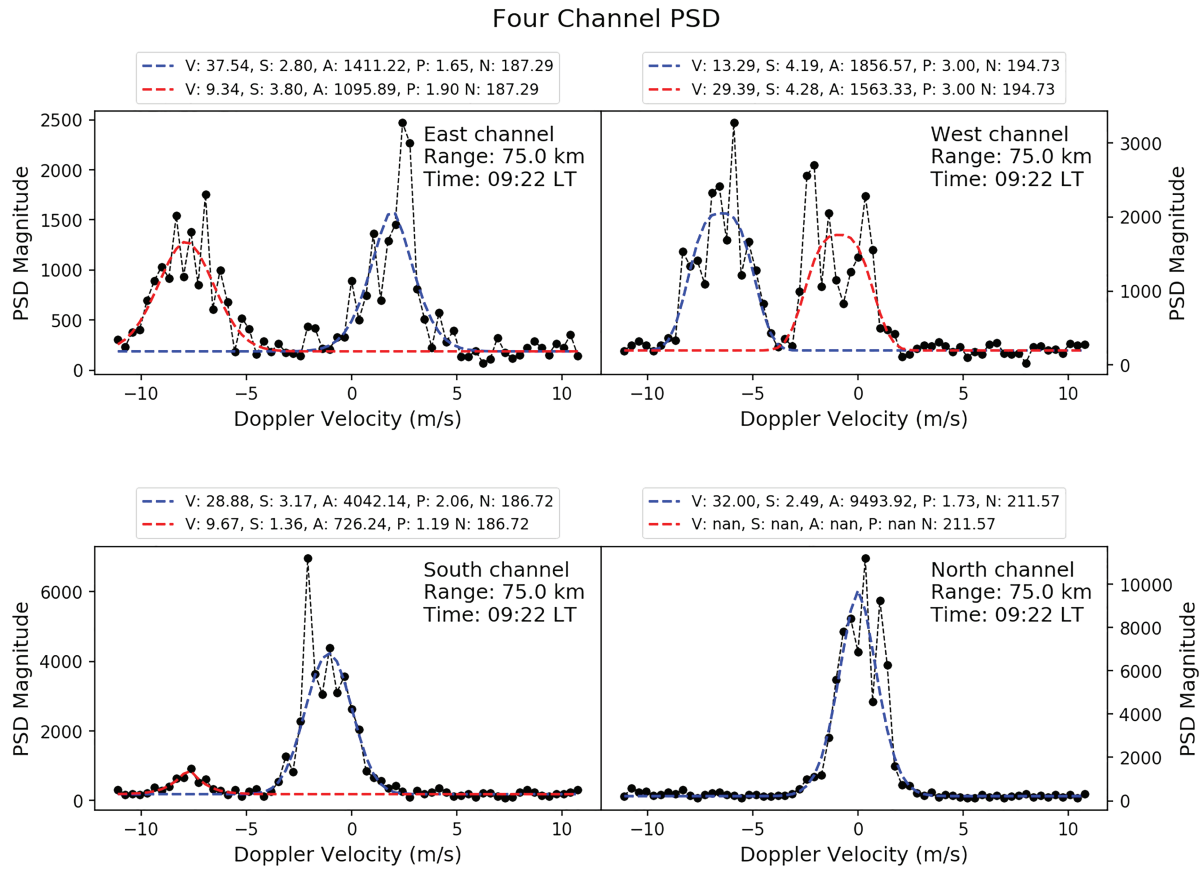


Figure 3. Four-channel dual Doppler spectral data fitted on 16 April 2015. A blue dashed line indicates our spectral peak from the mainbeam, a red dashed line indicates our spectral peak from the sidelobe, and a black dash-dotted line indicates our raw PSD data.

spectrum due to sampling and time windowing effects). Figure 3 shows examples of measured Doppler spectra fitted by a model consisting of the superposition of two generalized Gaussian functions (e.g., Sheth et al., 2006). If the measured spectrum is single peaked the “best fit” of our dual Gaussian model will yield a very small amplitude for one of the fitted Gaussians as illustrated in some of the additional spectrum examples shown in Figure 4. In these spectra the independent variable is discrete samples of Doppler velocity v , corresponding to $-\lambda_B f$, expressed in terms of Doppler frequency variable f , the independent variable of fast Fourier transform outputs. In terms of v , our single-peaked generalized Gaussian spectral model is

$$G_1(v) \equiv A \exp \left(-\frac{1}{2} \left| \frac{v - V}{S} \right|^p \right), \quad (1)$$

where A is the amplitude of the generalized Gaussian, V is the mean Doppler velocity shift (positive for motion away from the radar), S is Doppler spectral width, and p is the shape parameter of the generalized Gaussian. A dual-peaked extension of this is

$$G_2(v) \equiv A_1 \exp \left(-\frac{1}{2} \left| \frac{v - V_1}{S_1} \right|^{p_1} \right) + A_2 \exp \left(-\frac{1}{2} \left| \frac{v - V_2}{S_2} \right|^{p_2} \right), \quad (2)$$

with parameters A_i , V_i , S_i , and p_i describing the two separate Gaussians. An extension of (2) that accounts for possible occurrences of frequency aliasing as well as a noise floor N is

$$G(v) \equiv G_2(v - 2v_{\max}) + G_2(v) + G_2(v + 2v_{\max}) + N, \quad (3)$$

where $v_{\max}$ is the Nyquist Doppler bandwidth. The dual-peaked fitting model used in this work is actually based on (3) utilized for $|v| \leq v_{\max}$.

In practice, the natural logarithm of the measured 64-point spectra are fitted to discrete samples of the natural logarithm of model (3) (see Sheth et al., 2006) using least squares fitting modules from the Python

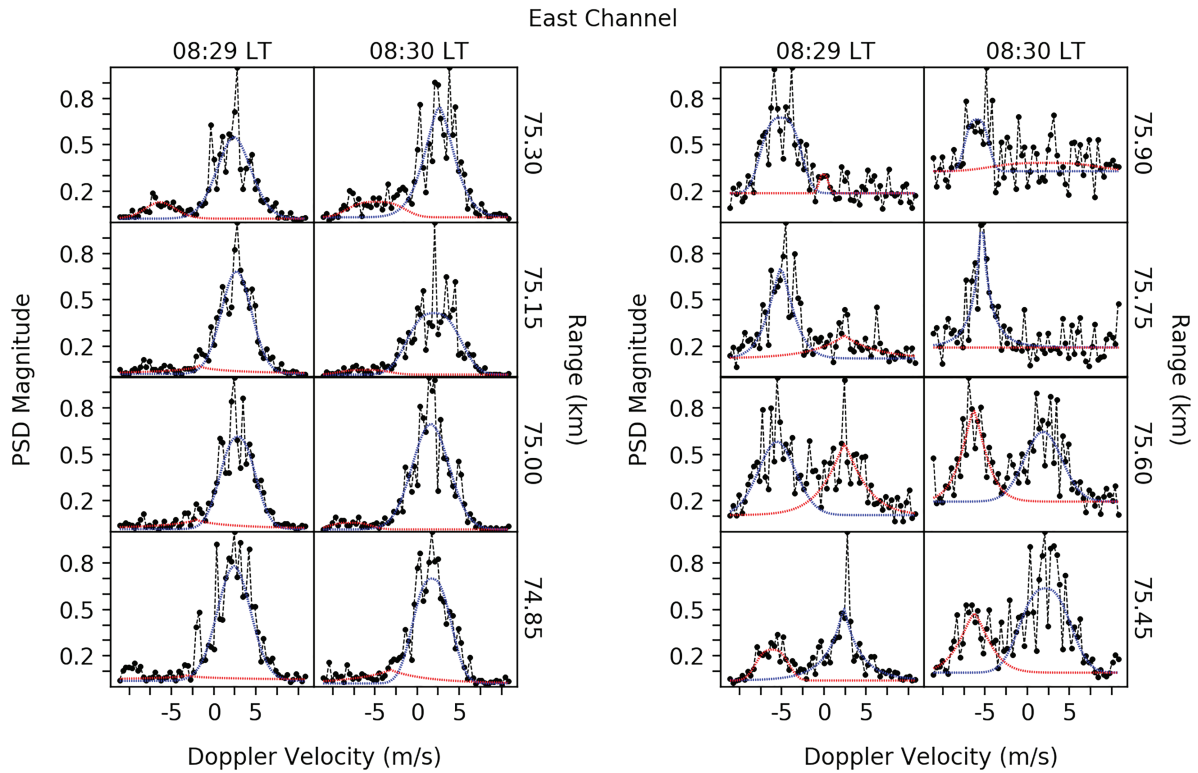


Figure 4. Fitted PSD plots from Jicamarca MST data on 16 April 2015. Blue dashed line indicates our preferred generalized-Gaussian (GG) peak, and red dashed line indicates secondary GG peak, which we assume to have come from sidelobe.

package `scipy.optimize` (Jones et al., 2001)—more specifically, `fmin_l_bfgs_b` function, which uses L-BFGS-B algorithm of Zhu et al. (1997), was chosen among the available nonlinear optimization methods. This algorithm allows constraints to be imposed for the model fit parameters—we constrain the noise variable N as

$$0.9 \cdot \text{noiseHS} \leq N \leq 1.1 \cdot \text{noiseHS}, \quad (4)$$

where `noiseHS` is a spectral noise floor estimate derived from the measured spectrum using a technique described in Hildebrand and Sekhon (1974); amplitude A is constrained as

$$0 \leq A \leq \infty, \quad (5)$$

Doppler shift V as

$$-v_{\max} \leq V \leq v_{\max}, \quad (6)$$

Doppler width S as

$$\Delta v \leq S \leq \text{powerwidth } \Delta v, \quad (7)$$

where `powerwidth` is the number of Doppler velocity bins in the measured signal spectrum between 8% and 92% of cumulative signal power and finally the shape parameter p as

$$0.5 \leq p \leq 3. \quad (8)$$

The fitting routine calculates, among other parameters, for example, the amplitudes A_1 and A_2 , the best fit values of Doppler velocity shifts V_1 and V_2 , and one of these is selected as the “radial velocity” estimate V_b for the antenna beam $b \in E, W, S, V$ in use, to be utilized subsequently in vector velocity estimation as

described below. Initially V_b is selected as the shift variable V_1 or V_2 matching the larger one of the fitted amplitudes A_1 and A_2 , but other strategies were developed for better results as discussed in section 4.

Assuming that V_b is the radial Doppler velocity of radar targets detected within the main antenna beam with known direction cosines Θ_{xb} and Θ_{yb} , we have a simple projection (i.e., dot product) model

$$V_b = \Theta_{xb}v_x + \Theta_{yb}v_y + \Theta_{zb}v_z, \quad (9)$$

involving vectors $(\Theta_{xb}, \Theta_{yb}, \Theta_{zb})$ and (v_x, v_y, v_z) , where $\Theta_{zb} \equiv \sqrt{1 - \Theta_{xb}^2 - \Theta_{yb}^2}$ and $v_{x,y,z}$ are the scalar wind velocity components for the probed altitude expressed in the local coordinate system depicted in Figure 2. We estimate the $v_{x,y,z}$ wind components from V_b radial Doppler data using a straightforward pseudo-inverse operation applied on (9), $b \in E, W, S, V$.

In order to convert $v_{x,y,z}$ into geophysically relevant u, v, w wind components we simply dot multiply the vector wind expression

$$\vec{v} = v_x\hat{x} + v_y\hat{y} + v_z\hat{z}, \quad (10)$$

where $\hat{x}, \hat{y}, \hat{z}$ are the unit vectors of the JRO coordinate system shown in Figure 2, with unit vectors $\hat{e}, \hat{n}, \hat{u}$ directed from the location of JRO toward geodetic east, north, and zenith, respectively—this leads to

$$u = v_x\hat{x} \cdot \hat{e} + v_y\hat{y} \cdot \hat{e} + v_z\hat{z} \cdot \hat{e}, \quad (11)$$

$$v = v_x\hat{x} \cdot \hat{n} + v_y\hat{y} \cdot \hat{n} + v_z\hat{z} \cdot \hat{n}, \quad (12)$$

$$w = v_x\hat{x} \cdot \hat{u} + v_y\hat{y} \cdot \hat{u} + v_z\hat{z} \cdot \hat{u}. \quad (13)$$

In (11)–(13)

$$\hat{e} = (-\sin\phi, \cos\phi, 0) = (0.9818, 0.1897, -0.0088) \quad (14)$$

$$\hat{n} = (-\sin\theta\cos\phi, -\sin\theta\sin\phi, \cos\theta) = (0.0536, -0.2327, 0.9711) \quad (15)$$

$$\hat{u} = (\cos\theta\cos\phi, \cos\theta\sin\phi, \sin\theta) = (0.2222, -0.9528, -0.2070) \quad (16)$$

where ϕ and θ are the East longitude and geodetic latitude of JRO with -76.8723° and -11.95° values, respectively, while the unit-normal (on-axis direction) for JRO is

$$\hat{z} = (\cos\lambda\cos(\gamma/4 + \phi), \cos\lambda\sin(\gamma/4 + \phi), \sin\lambda) = (0.2022, -0.9536, -0.2229), \quad (17)$$

where λ and γ denote the declination and hour angle of the on-axis direction with -12.88° and -4.6167° values, respectively. Also,

$$\hat{x} = \frac{\hat{u} \times \hat{z}}{|\hat{u} \times \hat{z}|} \quad (18)$$

(because the JRO antenna array plane is slightly tilted upward parallel to its x -axis edge—as runners around the array know very well), whereas $\hat{y} = \hat{z} \times \hat{x}$. Three components of each unit vector given in (14)–(18) are the axial projections in an Earth-centered coordinate system with the first and second axes directed to the equator on the prime meridian (Greenwich) and 90° east meridian, respectively, and the third axis toward the North Pole.

In the above description we glossed over one fine detail: Since the four antenna beams have their main lobes pointing in a nonsymmetric way distributed about the local zenith direction $\hat{u}$, equal ranges along different beams $b \in E, W, S, V$ do not correspond to equal heights above the local geoid passing through the center of the Jicamarca antenna array. Consequently, when we utilize four radial velocity estimates V_b , $b \in E, W, S, V$, in pseudo-inverting (9) for $v_{x,y,z}$ prior to u, v, w estimation, care must be taken to select samples of V_E, V_W, V_S , and V_V taken from different radar ranges corresponding to equal radar heights on the four radar beams (see section 4). This practice protects our u, v, w estimates from shear related systematic errors and only horizontal wind inhomogeneities may impact the results to introduce zero-mean fluctuating errors (see Kudeki et al., 1993).

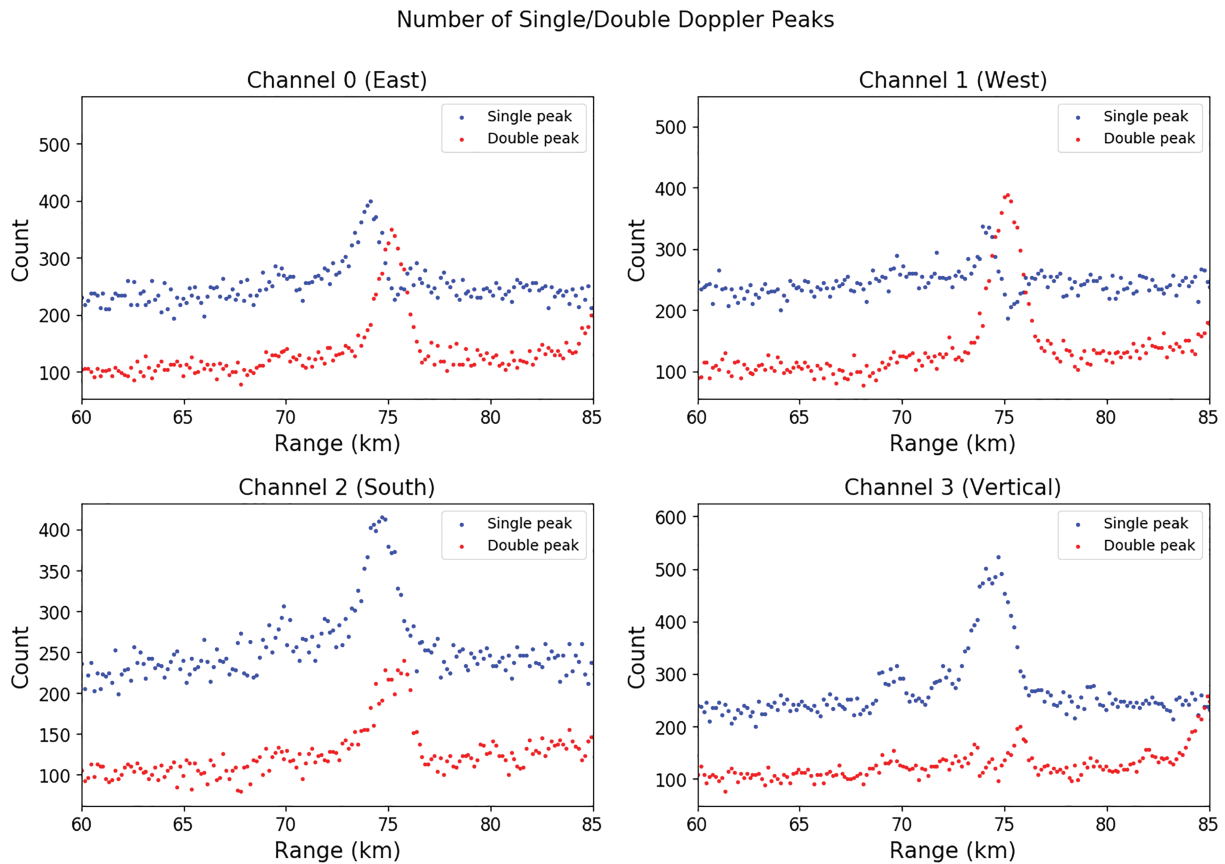


Figure 5. For each of four radar beams/channels used during the April 16, 2015 MST-ISR experiment the count of single peaked (blue dots) and double peaked (red dots) spectral observations as a function of range gate.

4. Classification of Main Beam and Sidelobe MST Radar Returns

In the previous section we have seen that Jicamarca MST wind estimation hinges on correctly identifying the radial Doppler velocity V_b derived from each of the four main antenna beams, $b \in E, W, S, V$, utilized in the observations. This identification occasionally fails when the measured Doppler spectra are double peaked, and the peak associated with the main beam or mainlobe is weaker than the peak associated with the sidelobe. This can happen, for example, when larger zenith angle sidelobes contain stronger turbulence at a given radar range compared to a smaller zenith angle mainlobe at the same radar range—a common occurrence at heights near the vicinity of the tops of narrow scattering layers frequently encountered in the mesosphere. Figure 5 shows the full counts of single- and double-peaked spectra detected in all range gates and all four beams of an MST-ISR experiment conducted at Jicamarca on 16 April 2015, over the entire observation interval. According to this figure, in close to one half of the observations at some ranges in E and W beams, there will be some chance of confusing the fitted Doppler shifts of the antenna sidelobes with the antenna mainlobe Doppler shift and thereby contaminating the final u, v, w estimates derived from the observations.

We will next describe our newly developed strategy to address this difficulty in MST-ISR wind analysis. The “trick” is to sort out and identify the mainbeam Doppler velocity shifts within the full set of mean Doppler shift results obtained from our fits.

We use a global strategy (developed as a part of the M.S. thesis work of the first author) where collections of all the fitted Doppler shifts of single- and double-peaked spectral models over all range gates are inspected jointly and partitioned into two or more sets based on clustering principles. One of these clusters is then identified with the mainbeam and utilized as described in the previous section in vector wind velocity estimation.

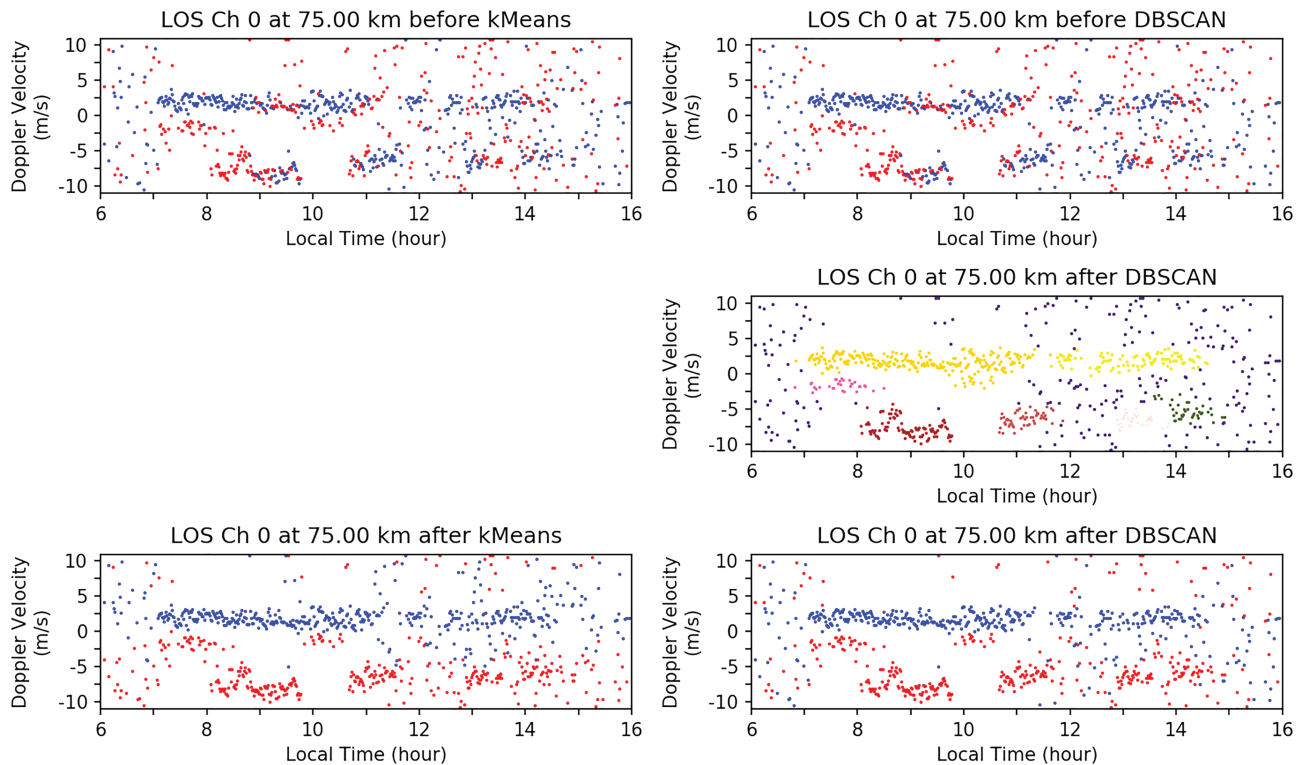


Figure 6. Identical scatter plots shown in both columns of the first row represent line-of-sight velocities with respect to time before a clustering algorithm is applied. Blue and red dots are the Doppler shifts of the spectral fits at a given time with blue dots having more spectral power than red dots. The second row shows an intermediate step for DBSCAN algorithm. The algorithm clusters tuple points (t_i, V_i) into multiple groups based on their closeness (seven groups in this case). Finally, the third row illustrates the reclustered Doppler shifts by kMeans algorithm (left) and DBSCAN algorithm (right).

The major steps of our clustering strategy applied to each of the four antenna beams (with mainlobe and sidelobe components) are as follows:

1. For each beam and range gate produce a set of fitted Doppler velocity V_i and observation time t_i tuples (t_i, V_i) spanning the duration of the experiment (typically all daytime hours)—these sets will include two different tuples for the same t_i when the spectrum for time t_i is double peaked.
2. Process each tuple set to produce two (or more) distinct clusters, one of which is to be identified as the mainbeam cluster in the next step.
3. For each beam process all the tuple sets jointly for identifying the mainbeam cluster of each range gate (starting with the lowest range gate) using a procedure explained below.

Step (1) amounts to a straightforward sorting and labeling of the Doppler velocity shift outputs of the fitting procedure described in the previous section. An example of such a sorted Doppler velocity set for 75.0 km range is shown in the top row of Figure 6 on both columns—each plot contains either single or double Doppler velocity shift data V_i at a given time t_i —a red dot or a red/blue pair. Notice that red and blue dots are not well separated in the top panels, they frequently appear in a mixed fashion among close neighbors.

In Step (2) clustering is performed on sets formed in Step (1). Clustering can be performed using a variety of techniques including the kMeans (e.g., Lloyd, 1982) and DBSCAN (e.g., Ester et al., 1996) algorithms provided by the Python machine learning package *Scikit-learn* (e.g., Pedregosa et al., 2011).

In kMeans, k initial clusters of V_i data are “reshuffled” until at the end, where k clusters are left to be optimally congregated about their own mean V_i values. Consider the case of $k = 2$ clusters initially consisting of the red and blue dots shown in the top row of Figure 6. Associated with these clusters are the mean values V_{red} and V_{blue} . Transfer any V_i in the blue cluster into the red cluster if V_i is found to be closer to V_{red} than to V_{blue} , and, conversely, any V_i in the red cluster into the blue cluster if V_i is closer to V_{blue} . Recalculate V_{red}

and V_{blue} after all the required exchanges are done, and repeat the process until there is no longer any need for any further transfers between the clusters. We implemented kMeans using the

$$\text{sklearn.cluster.kMeans}(n_cluster=2).fit([V_i]) \quad (19)$$

function call—this call can be made with an initial guess of the cluster means (e.g., V_{red} than V_{blue}) in order to speed up the convergence of the algorithm (Arthur & Vassilvitskii, 2007); otherwise, the k means are randomly initiated within sklearn.

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm (e.g., Ester et al., 1996), on the other hand, is called with user-specified parameters of eps and $min_neighbors$ for classification. The algorithm begins by searching a “core tuple” (t_i, V_i) that contains at least $min_neighbors$ neighboring tuples within an Euclidean distance of eps —in Euclidian distance computations squared distances are taken as the sum of squared time variable t_i entered in minutes and a squared Doppler shift index related to V_i ranging from 0 to 630 (as it maps to -11.11 to 10.77 m/s values of V_i after $fftshift$). Once found, all tuples that include the core tuple within its radius eps will be classified as the same cluster as the core tuple. Multiple core tuples can naturally exist in a single cluster formed this way. This process repeats until all the core tuples are found along with their corresponding neighbor tuples. At the end DBSCAN returns a variable number of clusters identified in this manner as well as an additional “outlier cluster” consisting of all the tuples having fewer neighbors within an eps radius than $min_neighbors$. We have used DBSCAN with varying input parameters on different days, for instance on 16 April 2015, as

$$\text{sklearn.cluster.DBSCAN}(eps=25, min_neighbors=20).fit([t_i, V_i]). \quad (20)$$

Sample results illustrating the use of kMeans and DBSCAN are shown in Figure 6 on the left and right columns, respectively. We observed DBSCAN to be a superior clustering algorithm for our purposes for several reasons: First, given the complicated side lobe structures of the MST-ISR antenna patterns, there is the possibility of having several distinct clusters of V_i data instead of just two clusters. Therefore, given the variable number of expected clusters—some very minor or noise-like— k of kMeans is uncertain and DBSCAN better handles this situation by depending on the Euclidean distance between all possible tuple pairs. Second, DBSCAN algorithm can find a cluster of any shape (e.g., elongated in 2-D) as long as its tuples are within eps radius from the core tuples of the cluster, while kMeans is only applicable if the clusters are separated by straight line borders on the (t_i, V_i) plane, not a universally valid happening.

Finally, step (3) of the procedure is for determining the correct cluster at each range gate associated with the main radar beam by referring to the “previous” range gate. The average velocity V_{avg} of each cluster in a range gate (there may be many such clusters when DBSCAN is used as opposed to kMeans) is compared to the average velocity of the mainlobe cluster decided for the previous range gate over the same set of t_i 's as the current cluster, and the cluster closest to the mainbeam cluster of the previous range gate is marked as the mainlobe cluster. This closest distance is required to be less than a user-defined $clust_dist$ variable—if the closest distance exceeds $clust_dist$, then the current cluster is marked as a sidelobe cluster and no mainbeam cluster may be assigned for that range gate if all the clusters spanning different t_i 's end up being marked as sidelobe clusters. Also, the mainbeam cluster of the previous range gate is required to be at least as large as $noise_thresh$ times the present cluster size over its own set of t_i 's. Notice that vector velocity estimation will only be possible at range gates with nonempty mainbeam clusters on at least three of the four radar beam directions—avoiding faulty/biased wind estimation is more important than having some estimate at all heights and times (which is not possible in the MST technique in any case because of the layered nature of mesospheric scattering layers). This algorithm works because at the lowest range gates of each scattering layer the Doppler spectra are all single peaked and therefore belong to mainlobe sets by default.

The clustering and sorting method based on DBSCAN is further illustrated in Figure 7. On the left column we notice that default choices of the preferred and secondary Doppler peaks are significantly mixed because of the sidelobe corruption. However, after applying DBSCAN algorithm two distinct clusters can be observed as shown on the right column.

Once all four beam Doppler data are sorted and classified we are ready for vector wind estimation except for one final detail—conversion of range gates of each channel to appropriate “height gates.”

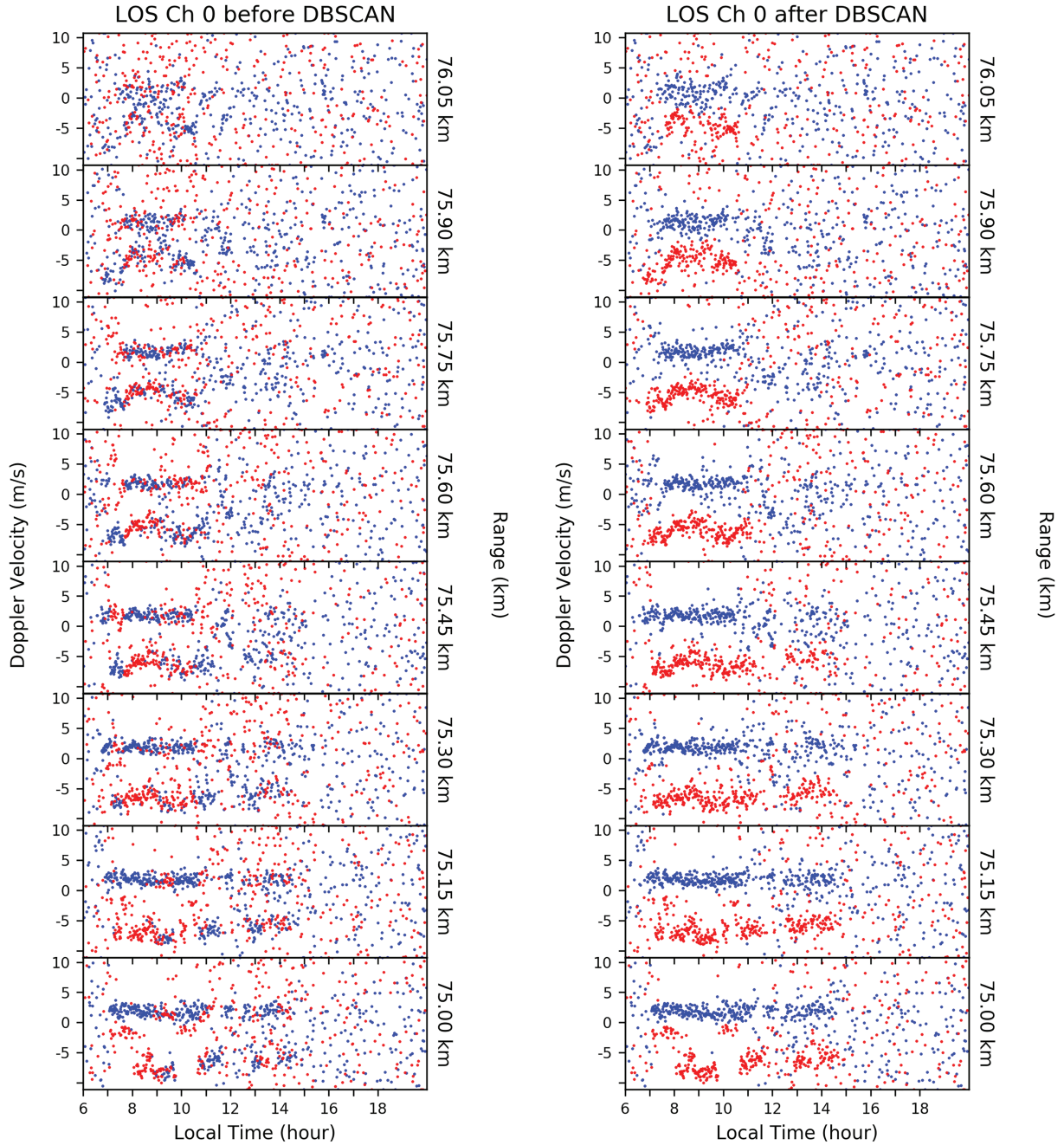


Figure 7. Sequence of line-of-sight velocities plots with respect to time index on 16 April 2015 before (left) and after DBSCAN (right). Blue mark indicates preferred peak and red mark indicates secondary peak.

On each antenna beam, we can associate with each range gate $r_n = n\Delta r$, with $\Delta r = 0.15$ km, the corresponding height above the local geoid given by

$$h_n = (\Theta_x, \Theta_y, \Theta_z) \cdot \hat{u} r_n, \quad (21)$$

where Θ_x and Θ_y are the direction cosines of the main beam directions specified in Table 2. Each range r_n (on each beam) is then assigned a “height index” m as the closest integer to the floating number $h_n/0.15$. Final wind calculations are subsequently completed using mainlobe $V_{b,m \leftarrow n}$ data selected from all four beams $b \in E, W, S, V$ for a set of discrete heights $0.15m$ km spanning the probed mesospheric heights.

Table 2

Direction Cosines Θ_x and Θ_y of Antenna Main Beams Used During Different Epochs of MST-ISR Experiments Since 2004

Date	Ch0 (Θ_x, Θ_y)	Ch1 (Θ_x, Θ_y)	Ch2 (Θ_x, Θ_y)	Ch3 (Θ_x, Θ_y)
Jan–Dec 2016	0.0204, 0.0545	−0.0603, −0.0406	0.0397, −0.0412	−0.0027, 0.0226
Jan–Dec 2015	0.0204, 0.0545	−0.0603, −0.0406	0.0397, −0.0412	−0.0027, 0.0226
Feb–Dec 2014	0.0171, 0.0556	−0.0603, −0.0406	0.0397, −0.0412	−0.0032, 0.0276
Jan 2014	0.0126, 0.0561	−0.0591, −0.0275	0.0404, −0.0408	−0.0034, 0.0276
Jan–Dec 2009	0.0126, 0.0561	−0.0580, −0.0284	0.0404, −0.0408	−0.0331, 0.0556
Jan–Dec 2007	0.0305, 0.0563	−0.0290, −0.0066	0.0289, −0.0066	−0.0304, 0.0563
Jan–Dec 2006	0.0305, 0.0563	−0.0290, −0.0066	0.0289, −0.0066	−0.0304, 0.0563
Jan–Dec 2005	0.0305, 0.0563	−0.0271, −0.0084	0.0289, −0.0066	−0.0284, 0.0584

Note. The $\Theta_z = \sqrt{1 - \Theta_x^2 - \Theta_y^2}$.

5. Results and Discussion

In previous sections we have provided a detailed description of the newly developed mesospheric vector wind velocity estimation procedure to be used with Jicamarca MST-ISR mode raw data sets (existing since 2005). In this section we will present sample power and wind velocity maps—height-time-intensity plots where the vertical axis indicates heights 0.15m rather than ranges 0.15n km (as explained in the last section)—illustrating the improved new results and capabilities using the MST-ISR data collected in April 2015.

In Figure 8 we present a selection of signal-to-noise ratio S/N maps in dB computed for 16 April 2015 day-time observations for all four MST-ISR beams—all the maps in the figure are dominated by an enhanced signal-to-noise ratio feature corresponding to a turbulent layer centered about 75 km height. The second and third column maps in the figure are based on the mainbeam and sidelobe echo power components identified with the classification method described in this paper. The fourth column combines these power components into a single map but with distinct and recognizable color signatures to affirm that the combination is compatible with the unsorted total power maps shown in the first column. Channels 2 and 3 have

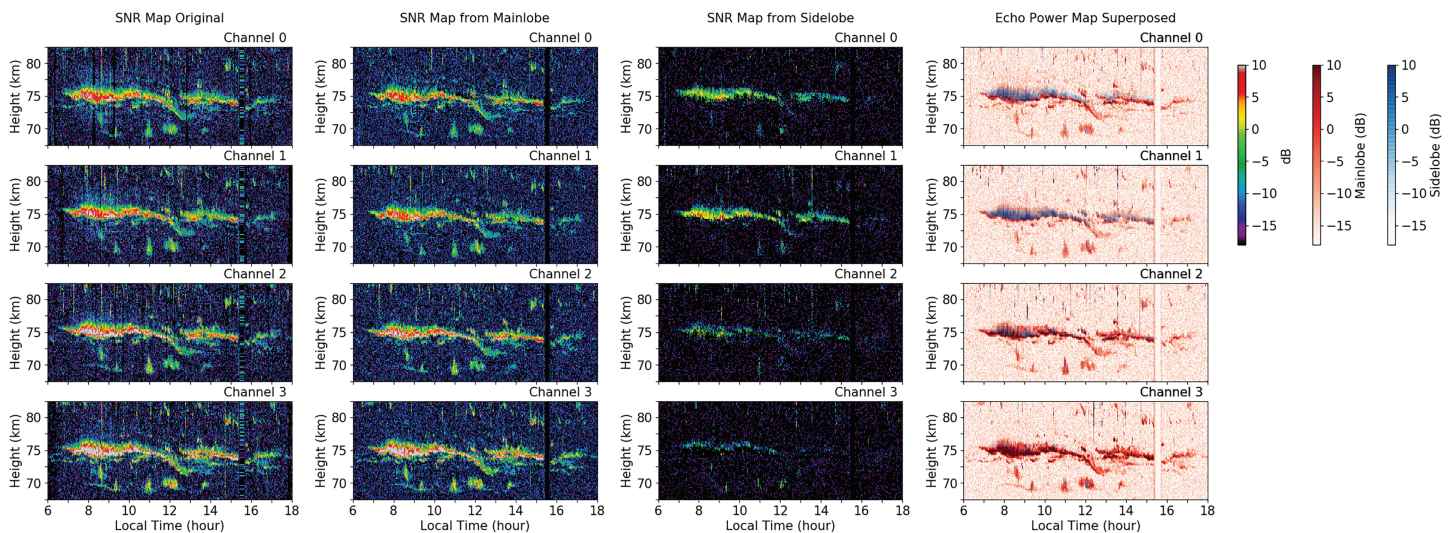


Figure 8. RTI maps representing the signal to noise ratio S/N in dB based on MST data from 16 April 2015—left column is based on total power measurements in the four MST-ISR beams, columns 2 and 3 account for mainbeam and sidelobe power calculations using the clustering and classification method described in the paper. The last column on the right represents the mainbeam (red tones) and sidelobe (blue tones) semitransparent maps superposed together. Clearly, sidelobe echoes are only prominent in channels 0 and 1, and in channel 0 the sidelobe echoes dominate over mainbeam echoes at larger ranges while in channel 1 mainbeam echoes and relatively weaker sidelobe echoes roughly coincide in range; this behavior is consistent with the fact that for channel 0 most of the sidelobes are more oblique than the mainbeam while for channel 1 the mainbeam is more oblique than most of the sidelobes as well as the true vertical direction.

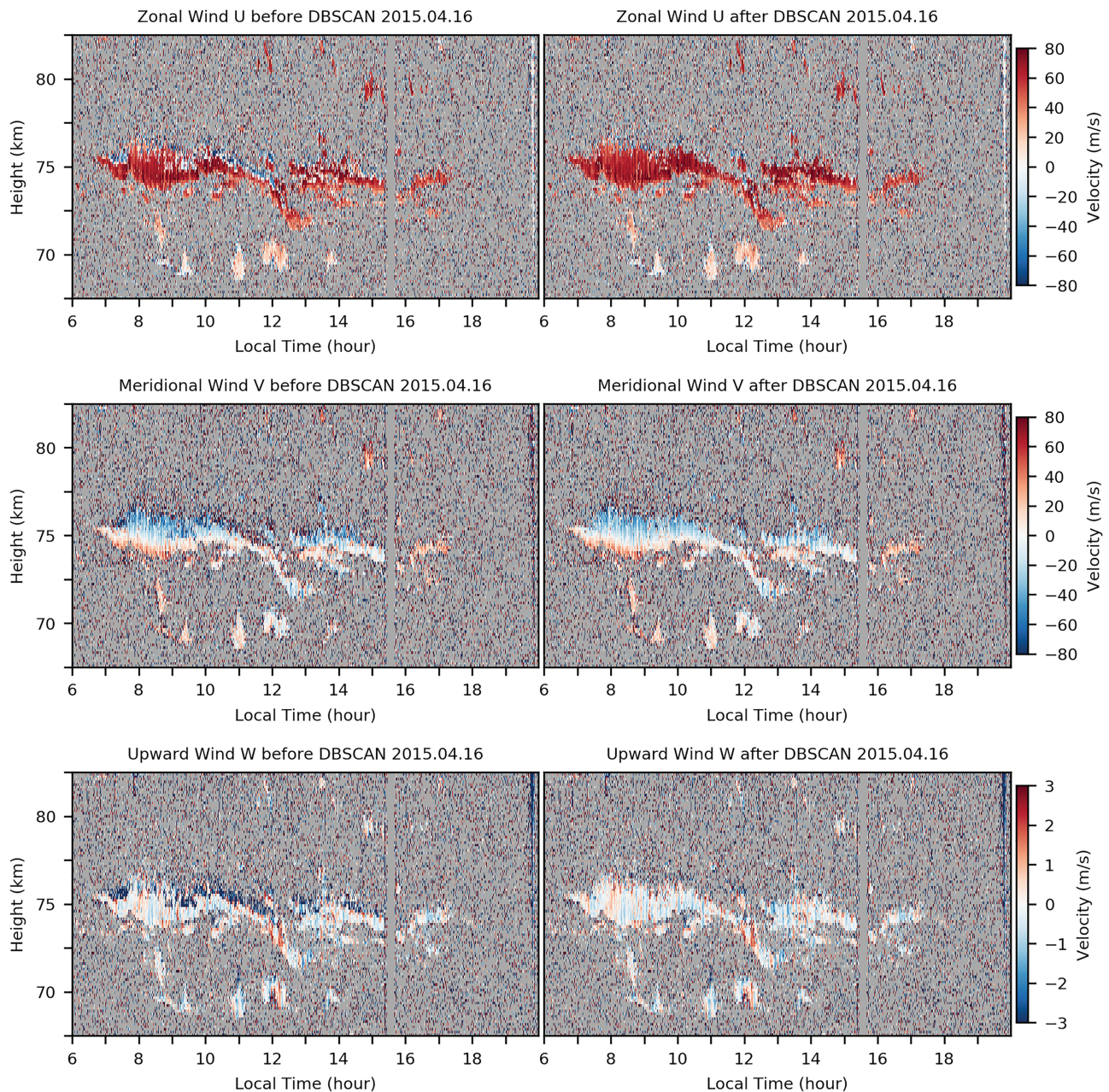


Figure 9. Eastward windmap (top), westward windmap (middle), and upward windmap (bottom) before and after DBSCAN clustering algorithm ($\text{eps}=25$, $\text{min_neighbors}=10$, $\text{clust_dist}=14$, and $\text{noise_thresh}=0.9$ for east, south, and vertical channel, and $\text{eps}=10$, $\text{min_neighbors}=7$, $\text{clust_dist}=12$, and $\text{noise_thresh}=0.9$ for west channel) based on MST data from 16 April 2015.

very small levels of sidelobe returns, while the strongest sidelobe returns seen in channel 0 appear at higher ranges than the mainbeam returns because of the more oblique direction of most of the sidelobes in channel 0 in comparison to the mainbeam.

The top row of Figure 9 shows the eastward wind map—the u -map—constructed for the same day as in Figure 8 before (left) and after (right) applying the clustering/classification method. The windmap on the right was obtained with DBSCAN-based clustering and is clearly much improved compared to the original windmap shown on the left—a discontinuity in wind velocity at upper turbulence layer of zonal windmap was greatly reduced after applying clustering algorithm. The second and third rows in Figure 9 show the meridional (northward) and vertical (upward) windmaps in a similar format, and it can be seen once more that the maps on the right obtained after the application of clustering and classification algorithm are much

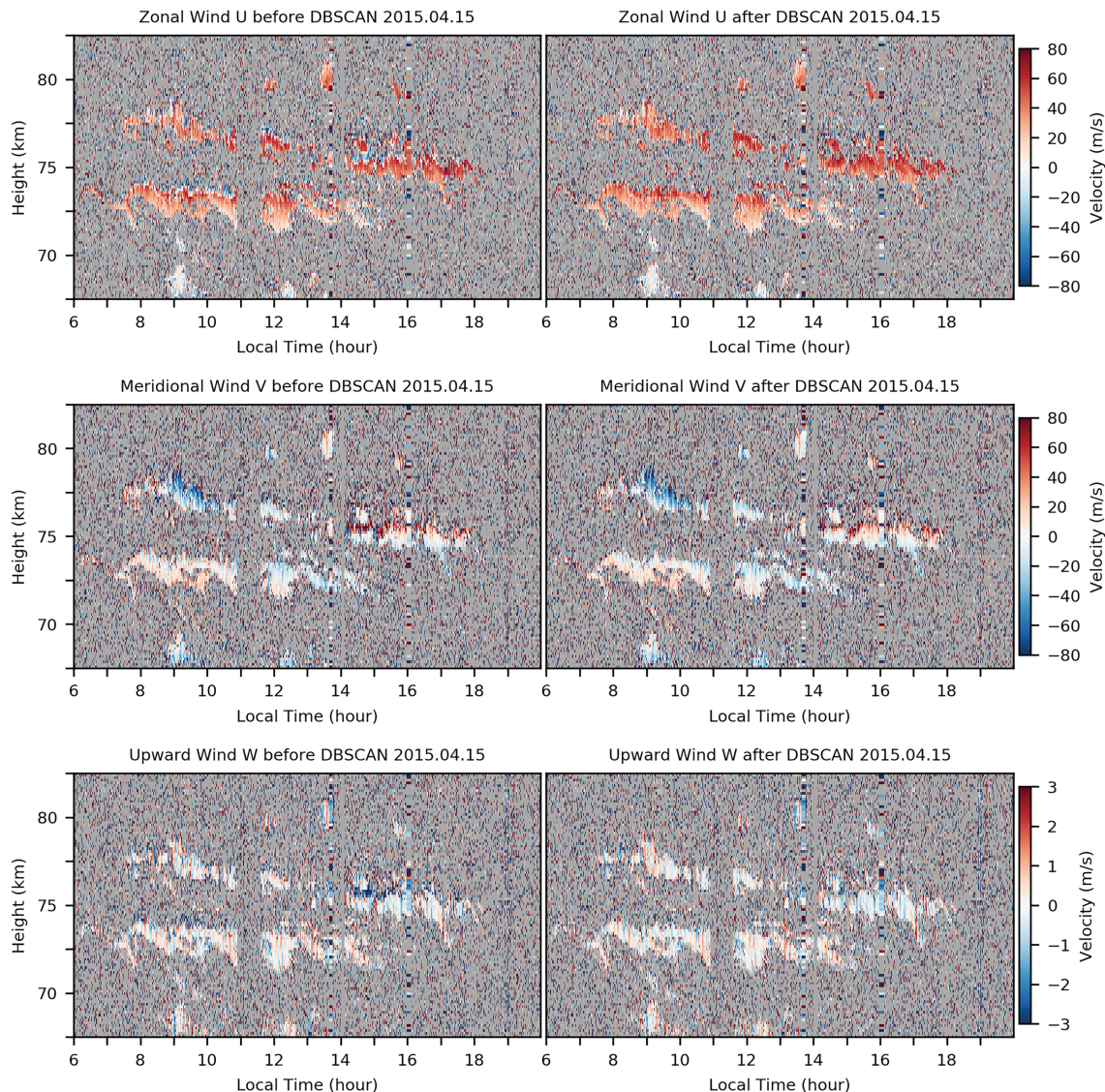


Figure 10. Eastward windmap (top), westward windmap (middle), and upward windmap (bottom) before and after DBSCAN clustering algorithm ($\text{eps}=25$, $\text{min_neighbors}=10$, $\text{clust_dist}=10$, and $\text{noise_thresh}=0.6$ for all channels) based on MST data from 15 April 2015.

improved. Also notice that the meridional windmap indicates the presence of a very strong vertical shear within the scattering layer, which is very typical since mesospheric scattering layers are frequently formed as a consequence of shear-driven turbulence. Other windmaps for 15 and 17 April, shown in Figures 10 and 11, respectively, also exhibit evidence for substantial improvements wind estimation after the application of DBSCAN based classification (as well as sheared flows).

A complete set of powermaps and windmaps from the entire MST-ISR mode observation campaigns conducted since 2005 can be accessed in our remote server (http://remote1.ece.illinois.edu/MST_ISR/uvvw.py).

On this web page results of each observation day can be accessed by clicking on the corresponding windmap icon. While inside the display page, tabs above the plot panels provide different displays of the data, including Wind_gg Map and Wind_gg+ Map, which are windmaps before and after DBSCAN algorithm, respectively—see an instruction page button, located above the tabs, to access further information about how to use the site. One of the features of the site is a channel selectable scatter plot display of entire sets of Doppler shift data similar to Figure 7. These scatter plots play a crucial role in fine tuning the wind estimation method described in this paper.

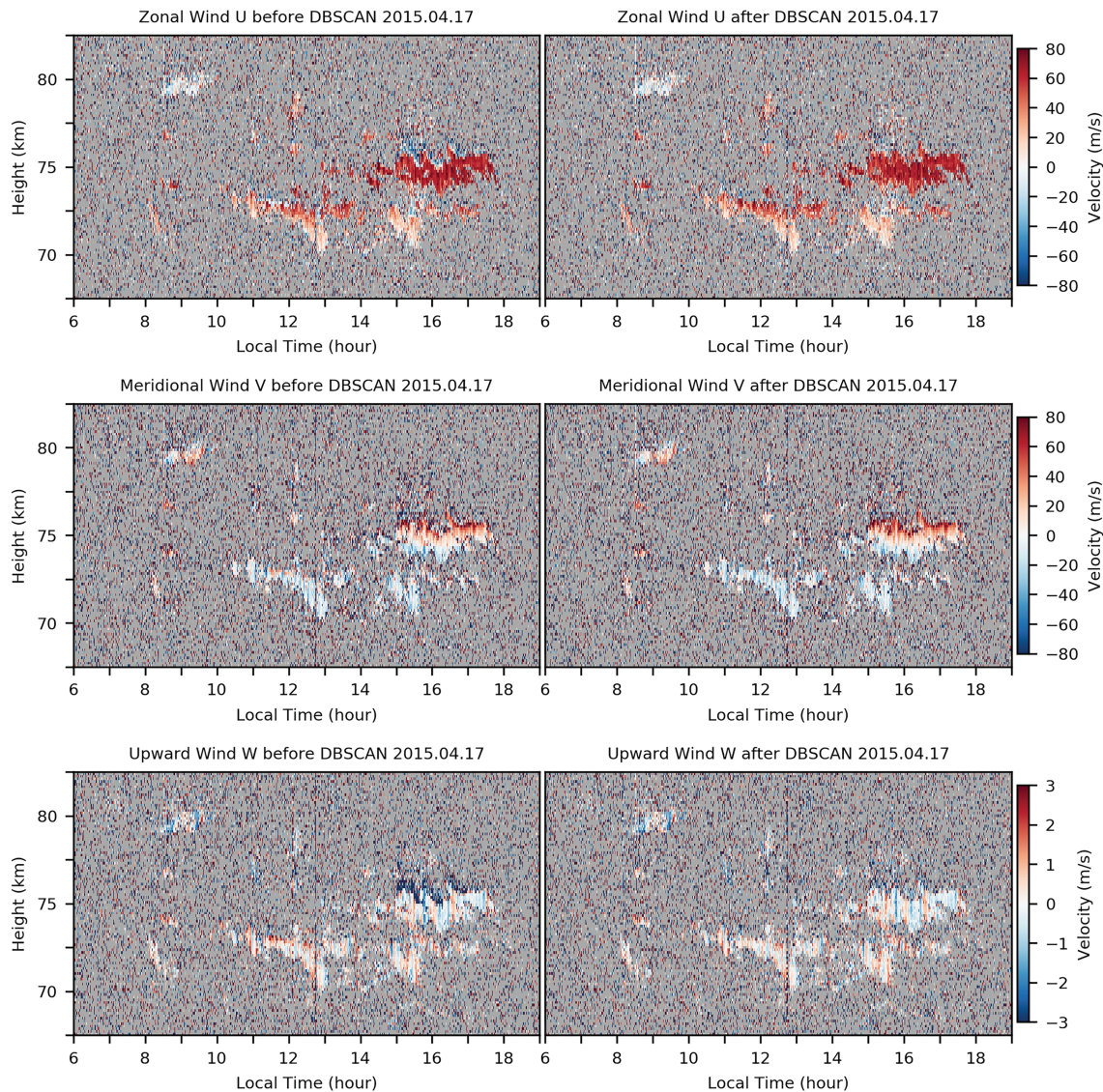


Figure 11. Eastward windmap (top), westward windmap (middle), and upward windmap (bottom) before and after DBSCAN clustering algorithm ($\text{eps}=25$, $\text{min_neighbors}=10$, $\text{clust_dist}=16$, and $\text{noise_thresh}=0.6$ for all channels) based on MST data from 17 April 2015.

In analyzing the MST-ISR data set (dating back to 2005), a fully automatized machine implementation of the clustering and classification procedure outlined earlier in the paper is difficult to achieve—there are significant day-to-day variations in mesospheric scattering scenarios that necessitates, in practice, to implement a “user-supervised” or “human-in-the-loop” version of what was described above. The first pass application of DBSCAN with a set of reasonable input parameters is automatized. However, the resulting wind maps—similar to Figure 9—are visually inspected to look for any residual signatures of sidelobe-caused artifacts. If such artifacts are noticed scatter plots shown in `MST_ISR/uvw.py` site (such as Figure 7) are inspected to understand the nature of clustering errors and DBSCAN is rerun with different input parameters until clustering and classification errors are eliminated—this human-in-the-loop step is quick and efficient (up to about 30 min of corrective action per observation day) and superior to both a fully manual clustering approach that would be impossible in practice or attempts to engineer a fully automatized approach to perfection (which has not been attained).

Once the vector wind velocities are estimated with some accuracy over a sufficiently long observation interval, it becomes possible to make, with the sorted and classified mainbeam Doppler velocity $V_{b,m \leftarrow n}$ data

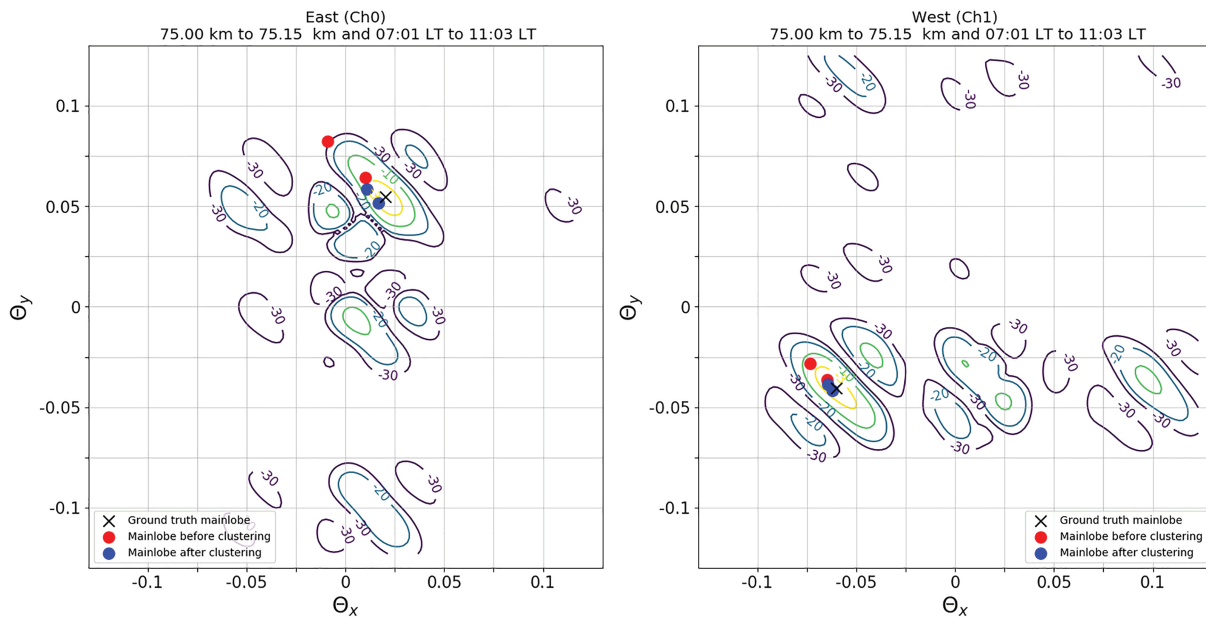


Figure 12. Beam pattern and computed mainlobe direction cosines of channel 0 (left) and channel 1 (right) using data from 75.00 and 75.15 km altitude and 07:01 to 11:03 LT on 16 April 2015; x marks the nominal direction cosines while red and blue dots the inversion results prior to and after DBSCAN-based side lobe identification step.

covering the entire interval, highly accurate “recalculations” of the direction cosines of the antenna mainlobes. Such calculations are useful to provide “reality checks” for the overall inversion process we have developed.

Figure 12 shows the recalculated mainlobe beam directions (direction cosines) for Channels 0 and 1 with east and west directed main beams shown in the figure as contour plots duplicated from Figure 1. The dots in the figure show the direction cosine calculation results obtained with the mainbeam Doppler velocity set prior to (red) and after (blue) the clustering-based mainbeam identification step using the data from 75.00 and 75.15 km height gates and 07:01 to 11:03 LT time interval. Note that while u, v, w wind estimates are made every minute using $V_{b,m \leftarrow n}$ data, the present inversion for direction cosines Θ_x and Θ_y is based on about 4 hr of beam Doppler data and wind estimates. To understand the meaning of this calculation and plot, consider what would happen—in the inversion of the set of equations relating the beam directions Θ_x and Θ_y and wind components u, v, w to line of sight $V_{b,m \leftarrow n}$ data taken from some height with index m —if and when $V_{b,m \leftarrow n}$ data were erroneous as a consequence of imperfect mainlobe identification. With some of $V_{b,m \leftarrow n}$ data coming from the main beams but the rest from other directions, the inversion process would fail to provide the correct values of Θ_x and Θ_y (indicated by x in the figure) as is the case with the red dots shown in the figure computed using $V_{b,m \leftarrow n}$ sets prior to mainlobe identification. After mainlobe identification, computed Θ_x and Θ_y values shown in blue, are where they belong, indicating for us a measure of quality of our mainbeam identification procedure. Two altitudes selected for this figure were close to the top of a scattering layer where the observed spectra are frequently double peaked and sidelobe peaks are at times stronger than the mainbeam peaks, making the mainlobe identification procedure (described in this paper) particularly important for these altitudes.

Finally, Figure 13 illustrates a larger set of results spanning 74.25 to 75.90 km altitudes (containing 11 height gates) and 06:00 to 19:59 LT on 16 April 2015, with red and blue markers once again indicating results prior to and after mainlobe identification, respectively. This is a large data set including many cases of dual-peaked spectra but relatively small instances of mainlobe/sidelobe ambiguities. Mainlobe directions are indicated by dots, computed sidelobe direction by crosses, and vector wind estimates by arrows.

In both channels, the calculated mainlobe directions are correctly placed (evidently classification step is not so important as blue and red dots overlap), while inversion with the “sidelobe data sets” provides locations

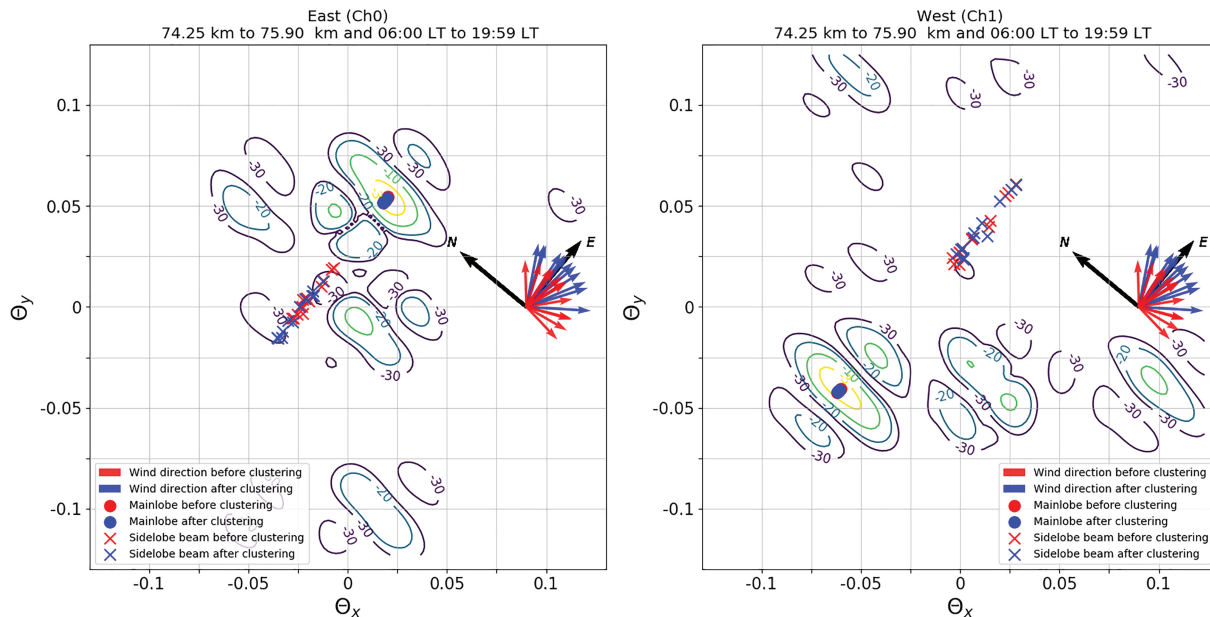


Figure 13. Beam pattern with computed mainlobe and sidelobe direction cosines of channel 0 (left) and channel 1 (right) with MST data from heights 74.25 to 75.90 km and from time 06:00 to 19:59 LT on 16 April 2015 using DBSCAN algorithm. Note that blue dots are masking red dots nearly colocated.

(indicated by crosses) near the zenith displaced from the main beam location in the direction of the computed wind vector, consistent with the beam broadening mechanism (invoked earlier on) combined with mesospheric aspect sensitive scattering favoring the zenith direction.

It should be mentioned that “mainlobe” based range to height mapping used for all the computations makes “sidelobe”-related inversions shown in the figure less important/meaningful than “mainlobe”-related reality checks already discussed above.

6. Conclusions

We have presented in this paper the state-of-the-art reached in mesospheric wind velocity estimation efforts at the Jicamarca Radio Observatory. This most advanced MST-ISR wind estimation method developed for Jicamarca makes use of a quad-beam radar configuration using the full transmission power capacity of the system while maintaining a 150 m range resolution. Doppler spectral estimates obtained with the quad-beam data streams exhibit single- or dual-peaked shapes which are least squares fitted to models consisting of a single or dual general Gaussians with adjustable amplitudes, widths, shape factors, and Doppler velocity shifts. Fitted Doppler velocity data collected in this fashion have been shown to be associated with the quad-beam mainlobe and/or sidelobe look directions. A new clustering and classification procedure introduced in this paper identifies the mainlobe data from the mixed initial data sets (of fitting products) and the Doppler velocity data associated with the known mainlobe beam directions can be subsequently combined to obtain high-resolution vector velocity estimates that should be bias free except for possible wind-inhomogeneity-caused fluctuating zero mean errors discussed in Kudeki et al. (1993).

The procedure just summarized has been applied to MST-ISR mode of Jicamarca observations conducted since 2005. Examples of results obtained with April 2015 MST-ISR observations have been presented in this paper. Vector wind estimates obtained in JRO MST-ISR campaigns since 2005 have been made public in the Madrigal data base.

References

- Akgriray, A. H. (2007). Calibration of Jicamarca radar using *F* region incoherent scatter for measurements of *D* region backscatter RCS (M.S. thesis), University of Illinois at Urbana-Champaign, Urbana, Illinois. Retrieved from <http://vufind.carli.illinois.edu/vf-uiu/Record/uiu5431543>

Acknowledgments

We thank the staff and engineers of the Jicamarca Radio Observatory for their assistance with the observations. Jicamarca Observatory is operated by the Instituto Geofísico del Perú, with support from the National Science Foundation. This work was supported by the Aeronomy Program, Division of Atmospheric and Geospace Sciences of the NSF, through Grants AGS 0422661 and 1143523. Results from the entire set of MST-ISR and MST-ISR-EEJ observations made at Jicamarca since 2005 are on display at this site (http://remote1.ece.illinois.edu/MST_ISR/uvw.py).

- Arthur, D., & Vassilvitskii, S. (2007). K-means++: The advantages of careful seeding. *Proc. 2. ACM-SIAM Symp. Discr. Algorithms 18th* (pp. 1027–1035). New Orleans, LA: Society for Industrial and Applied Mathematics.
- Ester, M., Kriegel, H., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering 3. clusters in large spatial databases with noise. *Proc. Int. Conf. KDD 2nd*, (226–231) Portland, OR: AAAI.
- Farley, D. T. (1985). On-line data processing techniques for MST radars. *Radio Science*, 20(6), 1177–1184.
- Guo, L., Lehmacher, G. A., Kudeki, E., Akgiray, A., Sheth, R., & Chau, J. (2007). Turbulent kinetic energy dissipation rates and eddy diffusivities in the tropical mesosphere using Jicamarca radar data. *Advances in Space Research*, 40, 744–750. <https://doi.org/10.1016/j.asr.2007.05.068>
- Hildebrand, P. H., & Sekhon, R. S. (1974). Objective determination of the noise level in Doppler spectra. *Journal of Applied Meteorology*, 13(7), 808–811.
- Jones, E., Oliphant, T., & Peterson, P. (2001). Scipy: Open source scientific tools for Python.
- Kudeki, E., Rastogi, P. K., & Sürücü, F. (1993). Systematic errors in radar wind estimation: Implications for comparative measurements. *Radio Science*, 28(2), 169–179. <https://doi.org/10.1029/92RS01931>
- Kudeki, E., Sürücü, F., & Woodman, R. (1990). Mesospheric wind and aspect sensitivity measurements at Jicamarca using radar interferometry and poststatistics steering techniques. *Radio Science*, 25(4), 595–612. <https://doi.org/10.1029/RS025i004p00595>
- Lehmacher, G. A., Guo, L., Kudeki, E., & Chau, J. (2007). High-resolution observations of mesospheric layers with the Jicamarca VHF radar. *Advances in Space Research*, 40, 734–743. <https://doi.org/10.1016/j.asr.2007.05.059>
- Lehmacher, G. A., Kudeki, E., Reyes, P. M., Lee, K., Heale, C. J., & Snively, J. B. (2019). Gravity wave ducting observed in the mesosphere over Jicamarca, Peru. *Journal Geophysical Research: Atmospheres*, 124, 5166–5177. <https://doi.org/10.1029/2019JD030264>
- Lloyd, S. P. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28, 129–137.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Sheth, R., Kudeki, E., Lehmacher, G. A., Sarango, M. F., Woodman, R. F., Chau, J. L., et al. (2006). A high-resolution study of mesospheric fine structure with the Jicamarca MST radar. *Annales Geophysicae*, 24(5), 1281–1293. <https://doi.org/10.5194/angeo-24-1281-2006>
- Smith, J. A. (2014). Spectral processing and wind estimation with Jicamarca mesospheric radar data (M.S. thesis), University of Illinois at Urbana-Champaign, Urbana, Illinois. Retrieved from <http://hdl.handle.net/2142/50631>
- Zhu, C., Byrd, R. H., Lu, P., & Nocedal, J. (1997). Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization. *ACM Transactions on Mathematical Software*, 23(4), 550–560.