AGU PUBLICATIONS

Radio Science

RESEARCH ARTICLE
10.1002/2016RS005989

Key Points:
- Frequently whistlers occur amidst multiple dispersions
- Multiple dispersions complicate whistler identification
- We have developed a new automated identification algorithm for this

Correspondence to:
A. R. Jacobson,
abramj@u.washington.edu

Citation:
Jacobson, A. R., R. H. Holzworth, R. Pfaff, and R. Heelis (2016), Automated identification of discrete, lightning-generated, multiple-dispersed whistler waves in C/NOFS-VEFI very low frequency observations, *Radio Sci.*, *51*, 1547–1569, doi:10.1002/2016RS005989.

Received 23 FEB 2016
Accepted 31 AUG 2016
Accepted article online 8 SEP 2016
Published online 24 SEP 2016

©2016. American Geophysical Union. All Rights Reserved.

Automated identification of discrete, lightning-generated, multiple-dispersed whistler waves in C/NOFS-VEFI very low frequency observations

Abram R. Jacobson[1], Robert H. Holzworth[1], Robert Pfaff[2], and Roderick Heelis[3]

[1]Earth and Space Sciences Department, University of Washington, Seattle, Washington, USA, [2]NASA Goddard Spaceflight Center, Greenbelt, Maryland, USA, [3]Center for Space Sciences, University of Texas at Dallas, Richardson, Texas, USA

Abstract Automated wave feature detection is required to efficiently analyze large archives of very low frequency broadband recordings for discrete whistler identification and feature extraction. We describe a new method to do this, even in the presence of simultaneous, multiple whistler phase dispersions. Previous techniques of whistler identification were unable to deal with simultaneous, multiple phase dispersions. We demonstrate the new method with data from the Vector Electric Field Investigation (VEFI) payload on the Communication/Navigation Outage Forecast System (C/NOFS) satellite, from the mission years 2008–2014.

1. Introduction

Since the early 1960s, electric and magnetic field wave instruments on research satellites have recorded lightning-generated discrete whistler waves in the magnetosphere [*Gurnett et al.*, 1965]. These whistlers have been useful in the study of broadband wave propagation in the magnetoplasma. Lightning-generated discrete whistlers have been considered as potentially modifying the radiation belt structure, although this has been contentious [*Draganov et al.*, 1992; *Green et al.*, 2005; *Meredith et al.*, 2006, 2007]. What is less controversial is that lightning-generated whistlers are often implicated in precipitation of magnetospheric electrons (see review in *Bortnik et al.* [2003] and *Bortnik et al.* [2006a, 2006b]).

Improved modern digitizers and mass memory on low-Earth-orbit research satellites like Detection of Electro-Magnetic Emissions Transmitted from Earthquake Regions (DEMETER) and Communication/Navigation Outage Forecast System (C/NOFS) in the past ~15 years have allowed the amassing of large quantities of broadband radio recordings containing numerous discrete, lightning-generated whistler disturbances [*Berthelier et al.*, 2006; *Burkholder et al.*, 2013; *Chum et al.*, 2006, 2009; *Fiser et al.*, 2010; *Jacobson et al.*, 2014; *Jacobson et al.*, 2011; *Santolik et al.*, 2009]. The large number of wave features available for later study provides a priceless opportunity for research on atmosphere-to-magnetosphere radio coupling, research on wave propagation within the geoplasma, and testing of whether lightning can affect inner-magnetosphere structure and particle populations. Due to the staggering size of the wave archives, it has been necessary to develop automated algorithms for identifying discrete whistler waves [*Fiser et al.*, 2010; *Jacobson et al.*, 2011] embedded amidst the data.

These automated approaches to wave identification have assumed that at any time, a sample of the record will be affected by just a single whistler wave dispersion. The simultaneous presence of more than one whistler wave dispersion defeats those first approaches to wave identification. Since multiple simultaneous dispersions are quite common, we have now developed a suitable algorithm for identifying discrete whistlers and determining the dispersion of these whistler waves, *in the presence of multiple phase dispersions simultaneously*. The ability to deal with multidispersion waves simultaneously is a new development. We describe this new "multidispersion" approach in section 3. Then, in section 4, we show how the multiple values of dispersion correspond to different propagation parameters. Finally, in section 5, we examine statistics of the Vector Electric Field Investigation (VEFI) peaks detected with the new method, compare these with lightning source ground truth, and use these statistics to validate the VEFI peaks.

2. C/NOFS Satellite and Payloads Used in This Study

The Communication/Navigation Outage Forecast system (C/NOFS) [*de La Beaujardiere*, 2004] was a satellite launched by the U.S. Air Force in 2008 to study ionospheric irregularities and dynamics at low magnetic

Table 1. C/NOFS Details

Name of satellite	Communication/Navigation Outage Forecast System (C/NOFS)
Mission developer	Air Force Research Laboratory (U.S.)
Mission scientist	Dr. O. de la Beaujardiere
Year of mission start	2008
Year of mission end	2015
Orbital inclination	approximately 13°
Perigee	401 km (decayed to ~390 km during 2014)
Apogee	867 km
Spacecraft attitude	Nonspinning; nadir stabilized
Electric field payload	Vector Electric Field Investigation (VEFI)
Electric field PI	Dr. R. Pfaff, NASA Goddard Space Flight Center
Ions, neutrals payload	Coupled Ion-Neutral Dynamics Investigation (CINDI)
Ions, neutrals PI	Dr. R. Heelis, Hanson Center for Space Science/University of Texas at Dallas
VEFI antennas	Three orthogonal boom pairs, each approximately 20 m tip to tip
VEFI recording	Records of three-axis electric fields, digitized at ~32 kilosamples/s; record duration typically ~12 s

latitudes. The useful epoch for this study was middle 2008 through middle 2014. The satellite mission ended in 2015. C/NOFS was in a slightly elliptical orbit with inclination 13°. The nominal altitude was 401 km at perigee and 867 km at apogee. Key to this study is that the satellite was a nonspinner and was nadir stabilized. For more details see Tables 1 and 2.

The Vector Electric Field Investigation (VEFI) payload [*Pfaff et al.*, 2010] measured three orthogonal electric field components, with a suite of different digitization rates from quasi-DC to 32 kilosamples/s. The antennas were balls at the ends of three pairs of opposing booms, each 9.5 m long. Two antenna pairs lay in the orbital plane, inclined ±45° with respect to the ram direction. The third antenna pair was normal to orbital plane. Although the main research challenge for VEFI was extremely accurate measurement of the low-latitude quasi-DC electric field, the payload also performed preprogrammed burst-mode digitizations of up to 32 kilosamples/s. We use the burst-mode, 32 kilosamples/s recordings in this study to analyze very low frequency (VLF; 3–30 kHz) whistler waves attributable to lightning. The passband actually useful to our measurements is limited below by 2 or 3 kHz and above by 15 kHz. The lower limit is imposed by the cutoffs inherent in the Earth-ionosphere waveguide modes, while the upper limit is due to the onboard antialiasing filter in the VEFI payload. Thus, we do not cover the upper half (15–30 kHz) of the VLF passband, so only with this proviso may we continue to characterize this study as being pertinent to the VLF. The recording times are usually preprogrammed, or triggered by ionospheric irregularities, and are not intentionally planned to coincide with nearby lightning.

The C/NOFS payloads also provided a very useful Planar Langmuir Probe [*Pfaff et al.*, 2010] time series of total ion density. This Langmuir probe was ideal for indicating the level of plasma irregularities along the orbit. We relied upon this data to exclude periods of significant ion-density irregularities.

Table 2. VEFI Data Set for This Study

Required sampling rate	"Burst mode", with ~32 kilosamples/s
Required data: *E* field	Three-axis, full vector electric field
Required data: electron density	Time series of relative ion density from C/NOFS planar Langmuir probe (PLP)
Required data: ion composition	Contemporary tabular data of ion composition from CINDI
Data quality requirement 1	E_{wave} perpendicular to $\mathbf{B}_0$: $<\mathbf{E}_{parallel}{}^2>/<\mathbf{E}_{perp}{}^2> < 0.015$
Data quality requirement 2	Relative electron-density fluctuations small: root-mean-square $\delta n/n < 0.005$
Data quality requirement 3	No fewer than one peak per record identified with new method dechirp
Number of qualifying burst records meeting all requirements above	6,905
Number of peaks detected with new (old) method dechirp	1,494,827 (341,840)
Number of WWLLN located strokes during the burst records	412,494
Number of peaks detected with new (old) method dechirp having width < 15 samples	385,055 (101,155)
Number of new (old) method dechirp peaks with width < 15 samples and correlated with WWLLN strokes	64,807 (18,415)

We performed a "sanity check" on VLF wave data to validate the satellite attitude control, the antenna boom locations, and the amplifier chain tracking between the three axes. For waves at frequencies significantly below the electron-cyclotron frequency, any known wave in the magnetized plasma should have negligible wave electric field parallel to the DC background magnetic field. We checked for orthogonality of recorded $\mathbf{E}_{\mathrm{wave}}$ with respect to $\mathbf{B}_0$ taken from the International Geomagnetic Reference Field (IGRF). To satisfy the sanity check, we required that the data satisfied

$$< E_{\mathrm{parallel}}{}^2 > / < E_{\mathrm{perp}}{}^2 > \quad < 0.005$$

during a record. For many records this criterion was failed so that the record could not be included in this study. Failing this criterion was almost entirely due to the observed, and obvious, presence of lower hybrid waves caused by the scattering of whistler waves from electron-density fluctuations [*Bell and Ngo,* 1988, 1990], whose presence was indicated by the planar Langmuir probe. Because the lower hybrid waves tend to be decameter scale, *the differential measurement using 20 m boom-pair sensor cannot measure the point electric field of such waves*. We could identify no instances where the sanity check failed in the absence of ion-density fluctuations. This implies that the satellite attitude control and the stability of the antenna configuration are generally within a couple of degrees of nominal.

Another C/NOFS payload also supported this study: The Coupled Ion-Neutral Dynamics Investigation (CINDI) [*Heelis et al*., 2009]. In order to infer wave vector from three-axis wave electric field [*Jacobson et al*., 2014], we need the ion composition (which determines the index of refraction). For more details on payloads, see Tables 1 and 2.

3. Automated Whistler Detection in Recorded VLF Satellite Data

3.1. Prior Methods

Lightning strokes in the Earth's lower atmosphere radiate copiously into the very low frequency (VLF; 3–30 kHz) radio band. The penetration of some of the lightning VLF wave energy into the ionosphere/magnetosphere can generate an oblique whistler wave [see, e.g., *Helliwell*, 2006, chap. 3]. The lightning-generated whistlers observed from satellites within the plasmasphere, when far from the plasmapause or any other ducting structure, are obliquely propagating [*Gurnett et al*., 1965], rather than ducted, whistlers. Here we shall be concerned with "zero-hop" oblique whistlers, corresponding to propagation upward from the atmosphere toward a low-Earth-orbit satellite.

A typical lightning return-stroke-signal VLF pulse width at the source is less than 1 ms [*Cummer*, 2000], but the zero-hop whistler pulse width observed at a low-Earth-orbit satellite is generally 1 to 2 orders of magnitude longer. This is due to the frequency dispersion of impulsive, lightning-generated whistler waves, which both broadens their pulse width and lowers their peak amplitude. Both of these effects make the waves less detectable among a noisy background and complicate the separation of two (or more) whistlers that are close together, even if each is well above the noise. Moreover, it is often necessary to correct the arrival time of the whistler by compensating for the dispersive delay. This step is extremely useful in determining to which individual lightning stroke the whistler may be attributed [*Chum et al*., 2006; *Fiser et al*., 2010; *Jacobson et al*., 2011; *Santolik et al*., 2009].

There have been at least three prior attempts to deal with frequency dispersion in whistler detection using satellite recordings:

1. *Eyeballing the data*. The human eye is a superb pattern-recognition device, and if the data set is small, visual examination of frequency/time spectrograms can be used to identify whistler occurrences, as was done in an early study of DEMETER and Magion-5 satellite data [*Chum et al*., 2006].
2. *Correlation with a fixed dispersion template*. In a crude approximation, dispersion during daytime does not vary or during nighttime does not vary. To the extent that this approximation applies, some pulse compression can be obtained by correlating the data time/frequency spectrogram with a template spectrogram: one template for a fixed daytime dispersion and another template for a fixed nighttime dispersion. Applying this to DEMETER [*Fiser et al*., 2010], it was possible to identify several tens of thousands of discrete zero-hop whistlers. This number of identifications exceeded what could be done reasonably by an "eyeball" approach.

3. *Automated fit of each brief data tranche to its own spectral dispersion*. Under variable ionospheric conditions, a fixed dispersion template is not effective. We find, for example, that even during satellite recordings at identical magnetic latitude, altitude, and local time, there can be 100% variability of the preferred wave dispersion. Part of this is not due to the ionospheric variability: The whistler dispersion relation depends on the polar angle θ between the wave vector **k** and the background magnetic field $\mathbf{B}_0$. Since θ is unconstrained by the ionosphere, even a standard ionosphere could support variable whistler dispersion. But the real ionosphere is not "standard." A robust feature of the low-latitude ionosphere is that there are large-scale day-to-day variabilities in ionospheric profile and filling [*Heelis et al*., 2009; *Klenzing et al*., 2011]. These day-to-day variabilities are not captured by standard models, e.g., International Reference Ionosphere [*Bilitza et al*., 2012], particularly during the recent (2008–2009) solar minimum, and during the extended recovery from the minimum. The standard models, such as the International Reference Ionosphere, are only statistically accurate, and in any particular instance, there can be significant (several tens of percent) discrepancies between the actual ion compositions and densities, and those predicted by the model. The C/NOFS [*de La Beaujardiere*, 2004] satellite, using the VEFI (Vector Electric Field Instrument) [*Pfaff et al*., 2010], did over half of its burst-mode whistler observations [*Burkholder et al*., 2013; *Jacobson et al*., 2011] precisely during that period of low solar activity. Therefore, in dealing with the C/NOFS period, it is best to retrieve, from the whistler data itself, the contemporaneous spectral dispersion, without imposing any assumption, e.g., of fixed daytime dispersion or of fixed nighttime dispersion. This approach was developed into an automated algorithm, described in the first report on VEFI burst-mode VLF recordings [*Jacobson et al*., 2011], henceforth identified as the "old method" dechirp.

3.2. Outline of the Old Method Automated Dechirp

We briefly summarize the old method of automated dechirp here. A typical VEFI burst record lasts ~12 s. The sampling rate is ~ 32 kilosamples/s, for a total of almost 0.4 million samples. We used a single antenna's signal to search for peaks in the old method of dechirp, although of course all three axes' components are needed to retrieve the wave polarization. We moved a ~0.5 s analysis window (16,384 samples), advanced by 50% (8192 samples). Each analysis window's digitized time domain signal was then Fourier transformed into the frequency domain, to yield a complex Fourier coefficient $e(f)$. In the quasi-longitudinal approximation, for a pure-electron whistler, for frequencies f well below both the electron gyrofrequency f_{ce} and the electron plasma frequency f_{pe}, we expect the phase to vary with frequency according to [see *Helliwell*, 2006, section 3.3]

$$\begin{aligned}\varphi &= \left(-\frac{2\pi}{c}\right)\sqrt{f}\ \int\left[f_{pe}/\sqrt{f_{ce}\cos(\theta)}\right]\,ds\cdot k_{\text{unit}}\\ &= \text{phase_dispersion} * \sqrt{f}\end{aligned} \tag{1}$$

where φ is phase (radians) and θ is the angle between the magnetic field **B**, the wave vector **k**, and k_{unit} is the unit vector along **k**. The integration is performed along the group path whose differential element is ds. This takes the form that phase is proportional to $f^{1/2}$ if we further approximate that neither the group path nor θ varies with frequency over the signal bandpass. For the latter in hertz, the constant of proportionality (henceforth, "phase_dispersion") has units of $s^{1/2}$. We emphasize that the phase_dispersion in the second line of equation (1) is defined as the ratio of phase (in radians) to the factor $f^{1/2}$. Our phase_dispersion is based on phase, not group delay, and *does not coincide with the oft-used quantity "D" based on group delay* [see *Helliwell*, 2006, equation (3.21)]. To relate the two, note that the group delay τ_g is the derivative of phase with respect to angular frequency, viz

$$\begin{aligned}\tau_g &= \frac{\partial\varnothing}{\partial\omega}\\ &= \frac{\partial}{\partial\omega}\left[\text{phase_dispersion}X\sqrt{f}\right] = \frac{1}{4\pi}X\,\text{phase_dispersion}X\frac{1}{\sqrt{f}}\end{aligned} \tag{2}$$

Thus, the phase-based phase_dispersion is related to the familiar group delay-based D by

$$\text{phase_}dispersion = 4\pi D \tag{3}$$

where D is defined in terms of group delay by [see *Helliwell*, 2006, equation (3.21)]

$$D = \tau_g X\sqrt{f} \tag{4}$$

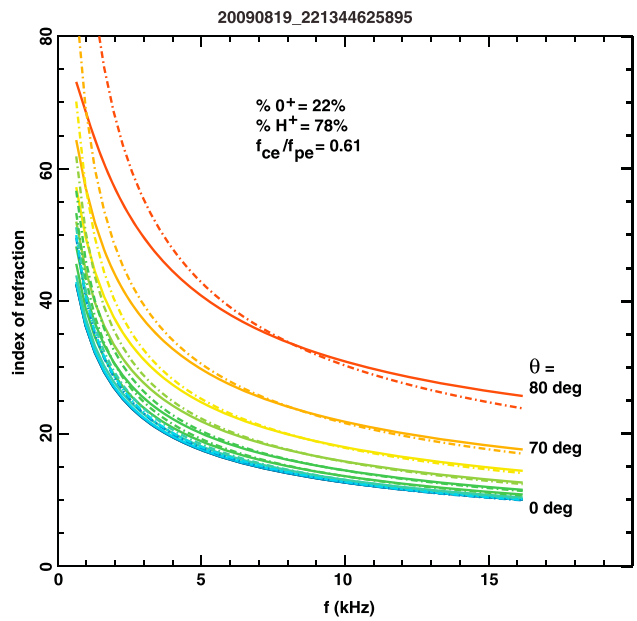

Figure 1. The index of refraction at the satellite versus frequency, parametrized by the polar angle θ between the wave vector k and the geomagnetic field B, in 10° steps from $\theta = 0°$ to $\theta = 80°$. The solid curves are the full, nonclosed-form solution including ion effects. The dashed curves are the closed-form electron whistler expression. We approximate the dispersion as if polar angle, θ (see equation (1)), does not vary versus frequency (see text).

In the frequency domain, the phase dispersion can be removed by multiplying the Fourier coefficient by the phasor $\exp(-i\varphi)$. We stepped the phase dispersion through a regular grid, running from zero to 200 $s^{1/2}$. At each step n (=1, 2, …, N) of the grid, we evaluated the corrected ("dechirped") Fourier transform e(f) $\exp(-i\varphi_n)$ and then reverse-Fourier transformed e(f) exp $(-i\varphi)$ to the time domain to get a dechirped time domain signal $E_n(t)$. Thus, we performed the reverse Fourier transform N times (but the forward transform only once.) We then competed the N dispersions against each other by testing for the value of n giving the highest and narrowest peaks in $E_n(t)$. The competition was done in a vector "figure of merit," which was a vector with N elements, one for each gridpoint of phase dispersion. The figure of merit consisted of the time-integrated eighth power of $E_n(t)$, integrated across the 0.5 s analysis window. There is no magic about using the eighth power and not the sixth power for example. *Use of any even power above the second power yields the largest time integral at the solution which creates the sharpest and highest peaks.* The second power is proportional to the power, and its integral is proportional to the energy during the time window. Since the energy is invariant to this phase transformation, the second power's time integral is invariant to the choice of phase dispersion.

We emphasize that our method, whether the old method just described or the "new method" described in section 3.4 below, works entirely in the frequency domain, where Fourier coefficient phases are adjusted according to the predicted whistler behavior. The phase manipulations in the frequency domain are particularly simple, fast, and amenable to automation, compared to the traditional method based on matching a dispersed spectrogram to a template based on a certain group dispersion D.

We further emphasize that both the old method and the new method approximate the dispersion as if polar angle θ (see equation (1)) does not vary versus frequency.

In this manner the automated algorithm based on the old method dechirp was satisfactory during the 2011 study. Each 0.5 s analysis window provides a dechirped waveform, containing zero, one, or multiple dechirped whistlers, as well as an optimum dispersion estimate for that analysis window. More details are available in the original report [*Jacobson et al.*, 2011]. However, the disadvantage of the old method dechirp is that it is confounded by the presence of more than one dispersion within the same analysis window. We now illustrate this confusion with an example from data recorded on 19 August 2009, which is the same data used to illustrate the new method in section 3.4 below.

3.3. Example of Old Method Dechirp in Presence of Multiple Dispersions

As stated earlier, the dechirp algorithm is based on equation (1), both for the old and new method dechirps. The automated dechirping requires that the closed-form electron whistler dispersion relation (see equation (1) above) at least approximates the full solution including ions [*Gurnett et al.*, 1965; *Stix*, 1962]. The latter is not closed form and hence is not convenient to use for automated dispersion retrieval, so we now check the validity of the electron-only approximation needed for the dechirp. The onboard CINDI instrument [*Heelis et al.*, 2009] furnishes the complete ion composition at the satellite. CINDI has time resolution as short as a

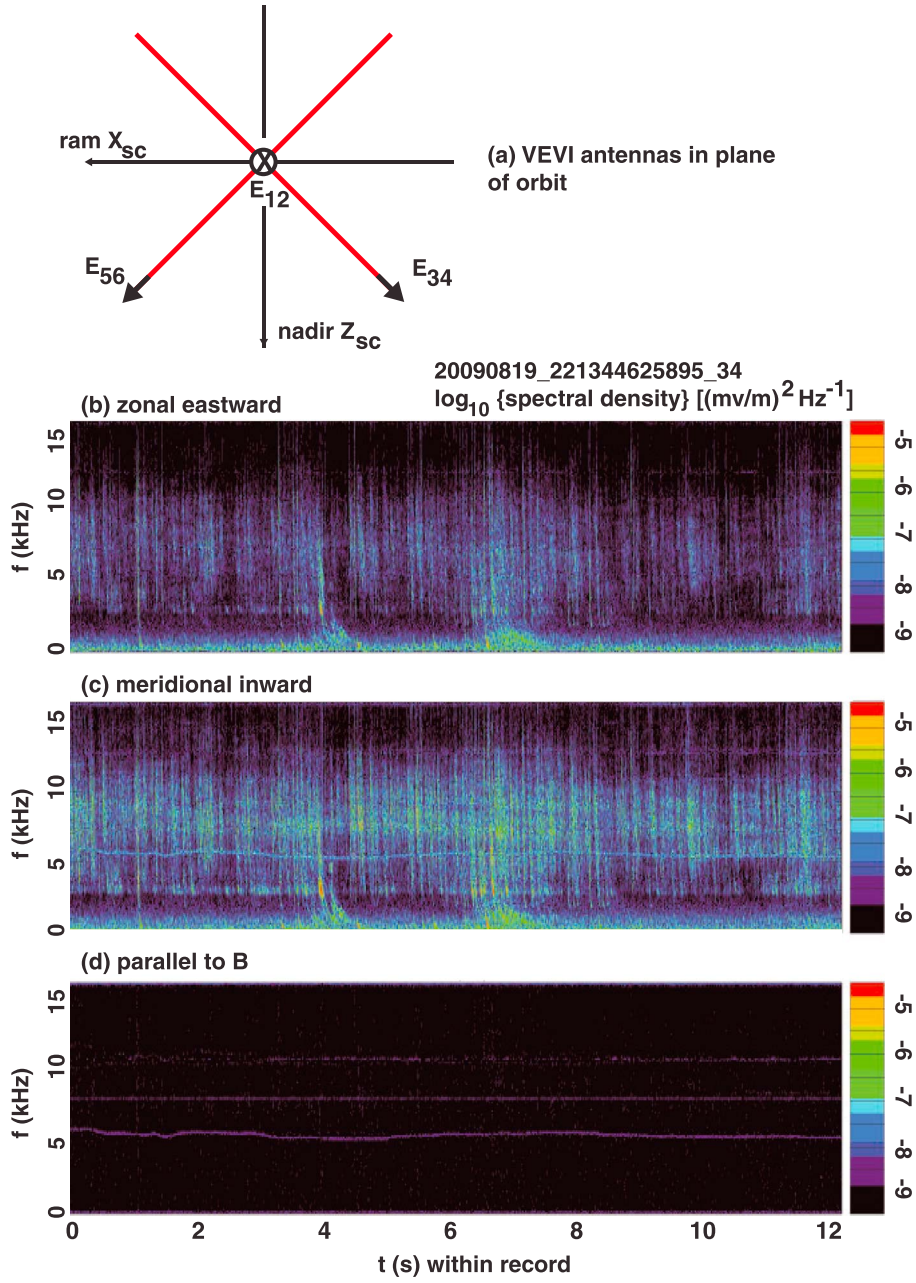

Figure 2. (a) Orientation of VEFI antennas in the orbital plane. (Figures 2b–2d) VEFI burst recording's power spectra, after transforming the antenna-based signals into magnetic coordinates: (b) Magnetic zonal eastward, (c) magnetic meridional inward, and (d) parallel to the geomagnetic field. Color scale is logarithmic. Each Fourier pane is 256 samples (8 ms) wide and is advanced by steps of 16 samples (0.5 ms). Note that the stepping (0.5 ms) is intentionally smaller than the temporal resolution set by the window width.

second [*Heelis et al.*, 2009]; we found that 15 s CINDI reporting intervals was adequate to describe the 12 s recording durations used for VEFI data. CINDI provides the total ion density as well as the percentages in each significant ion species, including those most important for our purposes: O^+, H^+, and He^+ ions. During the example recording to follow, CINDI determined that the ion compositions at C/NOFS were O^+ 22% and H^+ 78%, and (invoking the IGRF magnetic field) the ratio of electron gyrofrequency to electron plasma frequency was 0.61. (The latter captures the total ion density from CINDI.) Figure 1 shows the calculated index of refraction for the CINDI-specified ion composition. The solid curves include the ion effects [*Gurnett et al.*, 1965; *Stix*, 1962]. By contrast, the dashed curves ignore ion effects and use only the electron refractivity [*Helliwell*, 2006].

The curves are parameterized by the polar angle θ. Nearing perpendicularity, the electron-only approximation progressively fails to replicate the full dispersion relation including ions. However, our wave vector solutions (see section 4 below) usually have $\theta < 80°$. Thus, Figure 1 provides some confidence for our use of the electron-only dispersion relation (equation (1)) as a closed-form dechirp tool in this case.

The cartoon in Figure 2a shows the layout of the antenna axes. The diagram's plane is the orbital plane. Two antennas, "34" and "56," lie in the orbital plane. The third antenna, "12," is normal to the orbital plane. The ram direction is leftward, and nadir is downward. Figures 2b–2d shows the digitized electric field's moving-window spectrograms for the entire record (~12 s duration), transformed into geomagnetic coordinates (b) magnetic zonal eastward, (c) magnetic meridional inward, and (d) parallel to $\mathbf{B}_0$. Each Fourier pane is 256 samples (8 ms) wide and is advanced by steps of 16 samples (0.5 ms). Note that the stepping (0.5 ms) is intentionally smaller than the temporal resolution set by the window width. In Figure 2d, we see that the parallel component of the wave electric field is acceptably small. The residual carriers in the parallel component are probably aliased from out-of-band radio signals. The record in Figures 2b and 2c is dominated by multiple, discrete lightning-generated whistlers, as well as some long-duration, diffuse enhancements (see 5–10 kHz in Figure 2c) that we will ignore in what follows. Most, but not all, of the discrete whistlers' power is naturally high passed by the first couple of cutoffs of the Earth-ionosphere waveguide [*Hayakawa et al.*, 1994; *Yamashita*, 1978]. Apart from a few aliased out-of-band carriers, the wave electric field parallel to $\mathbf{B}_0$ is insensibly small (Figure 2d).

It should be noted that the moving-window spectrograms with short (256-sample) fast Fourier transform window, shown here, are mainly for illustrative purposes, to illustrate the signal's characteristics. However, we emphasize that *the dechirping methods discussed in this paper do not use the spectrograms.* Rather, the dechirping methods use the entire time domain signal within rather large (16,384-sample or 0.5 s) analysis domains.

For the old method dechirp, we use just the electric field along antenna axis 34. The old method dechirp is illustrated in Figure 3, applied to the first 16,384 samples (~0.5 s) of the record. This 16,384 samples is the "analysis window." The figure of merit, based on the eighth power of E_{34}, is shown in Figure 3a. The phase dispersion is swept in discrete steps between 0 and 200 $s^{1/2}$. At each step, the electric field's Fourier transform is phase corrected according to that phase dispersion, then back-Fourier transformed to the time domain. The figure of merit is the time-integrated, eighth power of this transformed electric field within the first analysis window. The normalization is by the maximum of that time integral, so the ordinate runs from 0 to 1. There are two peaks in the figure of merit: the stronger at ~100 $s^{1/2}$ and the weaker in the range 17–21 $s^{1/2}$. The old method algorithm chooses the highest point as the unique solution, which in this case means 100 $s^{1/2}$. The time history of E^2 is shown in Figure 3b, without any correction for phase dispersion. The corresponding time history of E^2 for a phase correction of 100 $s^{1/2}$ is shown in Figure 3c. Clearly, the phase correction has rearranged the time domain signal to have several intense, narrow peaks.

Figure 4 shows the meridional-inward electric field component's spectrogram in the first 16,384 samples of the first 32,768-sample analysis window. Each Fourier pane is 256 samples (8 ms) wide and is advanced by steps of 16 samples (0.5 ms). Note that the stepping (0.5 ms) is intentionally smaller than the temporal resolution set by the window width. An old method dechirp, retrieving a phase dispersion 100 $s^{1/2}$, has been performed by way of example, to illustrate the tendency of the most powerful disturbances to determine the fitted, single-dispersion dechirp. Two examples of these dominant features are indicated with horizontal arrows. The old method dechirp causes there to be artifacts which have been incorrectly dechirped, such as the two back-dechirped artifacts marked by the diagonal red arrows. Brown dashed lines serve as guides to estimate the amount of excessive dechirp associated with one of the back-chirped artifacts. We can make an eyeball estimate of the amount of excess dechirp in the back-chirped artifacts using the brown dashed lines as guides. The first of these artifacts passes through $f = f_2 = 15$ kHz at $t = t_2 \sim 0.12$ s and through $f = f_1 = 2$ kHz at $t = t_1 \sim 0.03$ s. The horizontal dashed lines mark these two frequencies, and the vertical dashed lines mark the two times. From equation (4), we can estimate the *apparent* group dispersion parameter "D_{app}" of the back-chirped artifacts, with frequencies in hertz:

$$D_{app} = \frac{(t_1 - t_2)}{\left(\frac{1}{\sqrt{f_1}} - \frac{1}{\sqrt{f_2}}\right)} \cong -6.4\ s^{1/2} \tag{5}$$

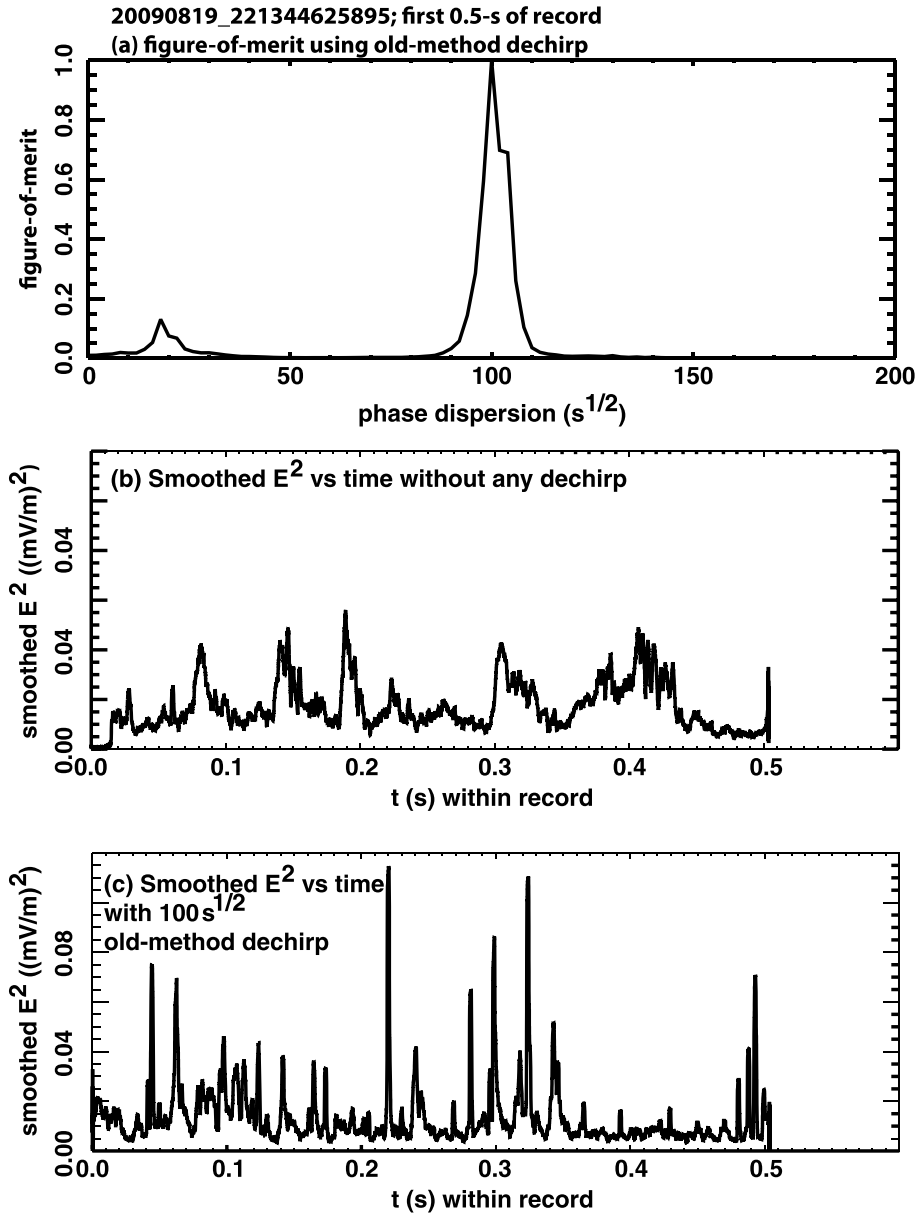

Figure 3. (a) Figure of merit versus phase dispersion, for the old method dechirp applied to the first ~0.5 s of record. (b) Smoothed E^2 versus time, for the first ~0.5 s of record, without dechirp. (c) Same but with optimum phase dispersion dechirp of 100 $s^{1/2}$.

From equation (3), this implies that the back-chirped artifacts have an *apparent* phase dispersion of $-80\,s^{1/2}$ *relative to the* $100\,s^{1/2}$ *dechirp already applied*. That is, the true phase dispersion of these features is $(100\,s^{1/2} - 80\,s^{1/2})$ or $20\,s^{1/2}$. This is consistent with the secondary peak in the figure of merit plot (Figure 3 above).

3.4. A New Method Allowing Simultaneous, Diverse Dispersions

The dispersion-fitting method described above suffices when the whistler waves in a given analysis window all display the same whistler dispersion. This corresponds to a single propagation path from the atmosphere (below the ionosphere) to the satellite. However, the VEFI data recordings often indicate multiple, simultaneous waves, leading to multiple, simultaneous-recorded phase dispersions. The phase-dispersion-fitting method described above has difficulty when there are multiple phase dispersions and in practice only locks onto the dispersion with the most spectral power. Thus, that algorithm misses the other dispersion(s) and at

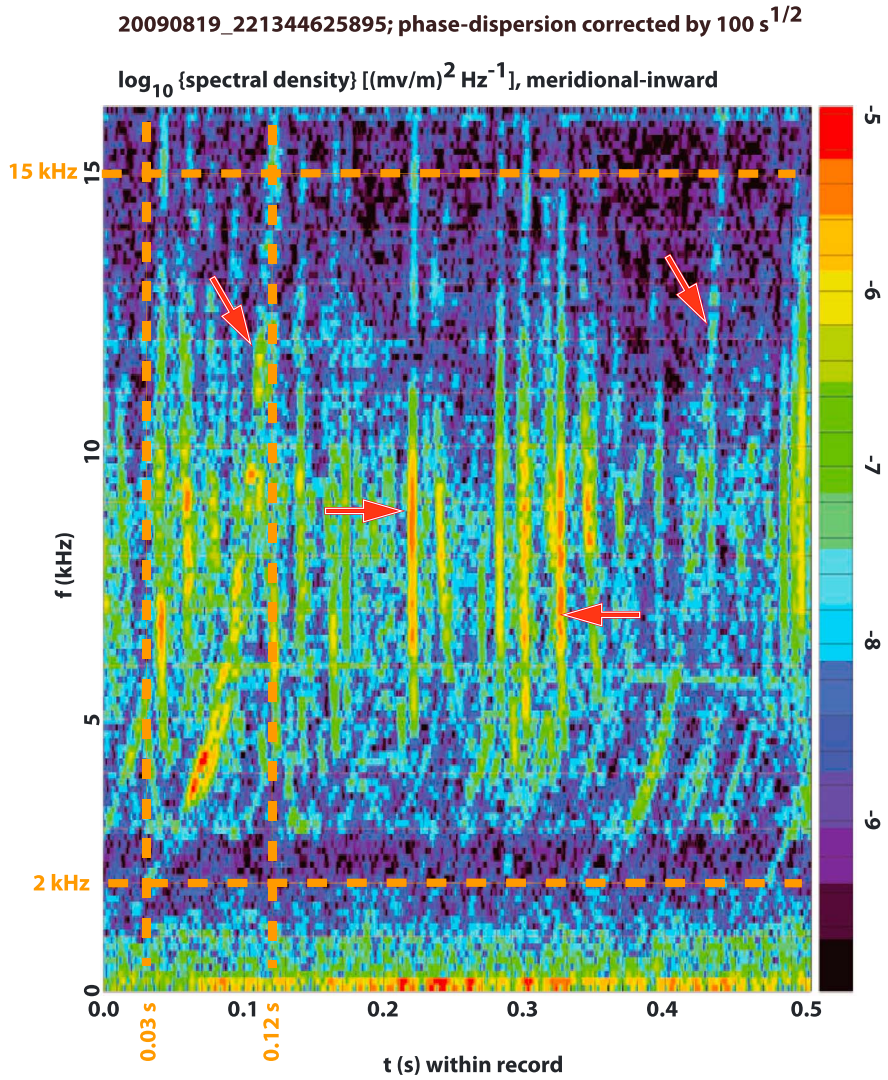

Figure 4. The meridional inward electric field component spectrogram in the first 16,384 samples of the first 32,768 sample analysis window. Each Fourier pane is 256 samples (8 ms) wide and is advanced by steps of 16 samples (0.5 ms). Note that the stepping (0.5 ms) is intentionally smaller than the temporal resolution set by the window width. An old method dechirp, retrieving a phase-dispersion 100 $s^{1/2}$, has been performed by way of example to illustrate the tendency of the most powerful disturbances to determine the fitted, single-dispersion dechirp. Two examples of these dominant features are indicated with horizontal arrows. The old method dechirp causes back-chirped artifacts which have been incorrectly dechirped, such as the two marked by the diagonal red arrows. Brown dashed lines serve as guides to estimate the amount of excessive dechirp associated with one of the back-chirped artifacts (see text).

the same time biases the dispersion estimate of the strongest waves due to competition from the nonzero power in the weaker waves(s). We have developed a more agile method to perform pulse compression, peak detection and phase-dispersion estimation in the presence of simultaneous, diverse dispersions. The method is illustrated as follows.

We divide the time domain into standard analysis windows, each 32,768 samples (1 s) long, and advance the window by discrete steps 8192 samples (1/4 s). Note that this analysis window is twice as long as in our original work [*Jacobson et al.*, 2011].

However, at the conclusion of the analysis of the 32,678 samples, we will zero the final 16,384 samples to prevent early times aliasing into late times. Thus, we are protected from signal rotating out the left boundary and back in through the right boundary due to the phase transformation during the phase transformation.

We now define a 2-D figure of merit matrix. To do this, we dechirp the 32,768-sample waveform *for each of 400 regularly spaced phase-dispersion values*, spaced by 0.75 $s^{1/2}$. Each dechirp provides a vector, containing the time domain signal dechirped with that particular value of phase dispersion. We then stack the vectors. This forms a matrix of $E_{n,\mathrm{dechirp}}(t)$ ($n = 1, 2, \ldots, 400$). Time runs horizontal, and dispersion runs vertical. We then raise that matrix to the fourth power and finally smooth in time (horizontal) by ± 0.5 ms.

Before searching the matrix for a candidate focus, we preexcavate triangular zones in the lower left and upper right corners of the plane so that no focus will be considered in those zones. This is to ensure that the bundle of rays passing through a candidate focus is not too truncated by nearby boundaries, as that would interfere with the determination of optimum phase dispersion. See the description of peak-quality-automated criteria below.

This provides a 2-D figure of merit matrix, shown in Figure 5a for the first half (16,384) samples of the 327,689-sample window. Time varies along the horizontal, while phase dispersion varies along the vertical. Note that *at each point in this matrix, all frequencies contribute*. The triangular excavations at the upper right and lower left corners are apparent.

One can see several conical foci in Figure 5a. The cone angle of each focus is proportional to that signal feature's bandwidth. Each conical focus is a candidate peak solution, already dechirped without that dechirp being "pulled" by competing peaks. The focus, or narrowest waist of the cone, occurs for the optimum phase-dispersion correction (project leftward to vertical axis) and the optimum time (project downward to horizontal axis). Diverse phase dispersions within the window may be detected. For example, one notes that although most of the conical foci in Figure 5a are near phase dispersion ~ 100 $s^{1/2}$, there is a more weakly dispersed peak at phase dispersion ~ 17 $s^{1/2}$, marked with a red arrow. Detection of both 100 $s^{1/2}$ and 17 $s^{1/2}$, dispersions simultaneously is straightforward with this algorithm.

We then use the figure of merit plane to find candidate peaks, in an iterative loop that finds the brightest remaining focus in the plane. In Figure 5a, the first candidate focus is marked by a red square. This candidate was chosen by virtue of its having the highest intensity in the plane. The red lines converging through that square indicate a conical zone within which the focus is further located (in phase dispersion.) We consider different levels of phase dispersion, up and down from the candidate focus, and at each level tabulate the maximum value of the matrix within the red-line-bound fiducial cone. We then choose the value of the phase dispersion giving the maximum figure of merit value. The maximum figure of merit versus dispersion, considered as a maximum within a row segment between the tilted conical boundaries, is graphed in Figure 5b versus phase dispersion. There is clearly a peak at 100 $s^{1/2}$, without interference from a secondary peak at 17 $s^{1/2}$.

We then vet this candidate peak by requiring that it meet several peak-quality-automated criteria [*Jacobson et al.*, 2011] adapted from the old method dechirp:

1. Reject if too close to peaks already found. We draw a rectangle around the candidate peak, ± 8 ms wide horizontally and ± 7.5 $s^{1/2}$ vertically. If there is any already-excavated pixel in that rectangle, then we ignore the new candidate peak as being too close to solutions found earlier.
2. Reject if candidate peak is closer than 4 ms to the temporal boundaries.
3. Reject if candidate peak's phase dispersion $> 292.5\ s^{1/2}$ or $< 7.5\ s^{1/2}$.
4. Reject if figure of merit peak (see Figure 5b) is too wide. We calculate the dispersion-width ratio as the ratio of the figure of merit integrated with respect to the phase-dispersion variable within the domain ± 37.5 $s^{1/2}$ about the peak over the peak figure of merit. The ratio has units of phase dispersion ($s^{1/2}$). We reject any candidate peak if this dispersion-width ratio $> 15\ s^{1/2}$. For reference, the peak shown in Figure 5a has a dispersion-width ratio of 12 $s^{1/2}$.

At the end of this cycle of the iterative loop, we zero the same rectangle around the candidate peak, ± 8 ms wide horizontally and ± 7.5 $s^{1/2}$ vertically, *whether or not the peak has been accepted or rejected by the peak quality criteria*, and begin the next iteration by finding the brightest surviving focus. This is shown in Figure 6. The first peak has been "excavated" from the data, and the second peak is further to the left (earlier in time). The algorithm will then test this second candidate peak, decide whether it meets peak quality criteria, and then excavate another small rectangular zone. In this manner the figure of merit plane becomes increasingly excavated, the remaining candidates become more likely to fail, and the loop proceeds to a threshold of percentage excavation of the plane. At that final iteration we collect our accepted peaks

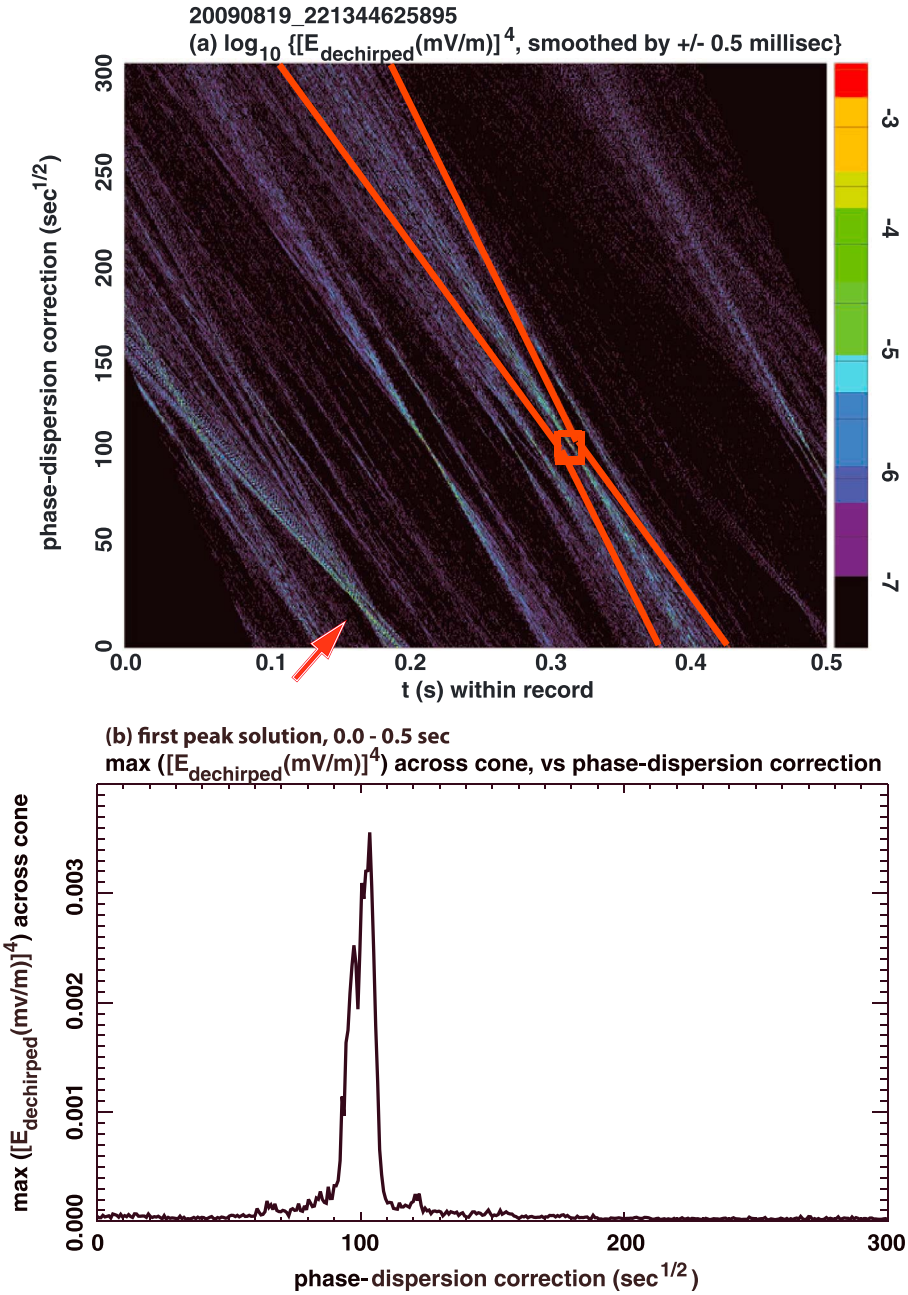

Figure 5. (a) Two-dimensional figure of merit matrix, for the first half (16,384) samples of the first 32,7689 sample window. Time runs horizontal, and dispersion runs vertical. The first candidate peak is marked by a red square. The red lines mark a conical fiducial zone within which the focus is further located (in phase dispersion). We consider different levels of dispersion, up and down from the candidate focus, and at each level tabulate the maximum value of the matrix within the left and right boundaries of the red-line-bound fiducial cone. A red arrow marks another peak at lower phase dispersion. (b) Maximum figure of merit, considered as a maximum within a row segment between the tilted conical boundaries, versus the phase dispersion of the rows.

from this window, move the analysis window 8192 samples rightward, and repeat the whole process with new data.

At the overall end of the process, when the analysis window has been stepped as far rightward has it can be without protruding past the end of the record, we accumulate the accepted peaks and display them as a scatter plot versus time (horizontal) and phase dispersion (vertical). This is shown in Figure 7a. This includes solutions from

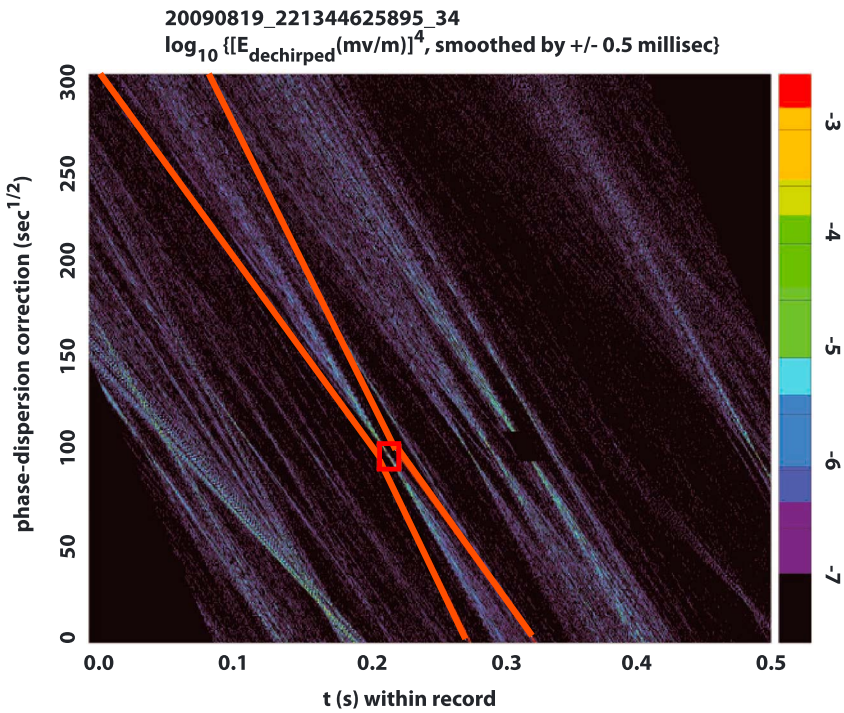

Figure 6. Similar to Figure 5a but for the second iteration of the new method dechirp example. A rectangular box around the first candidate peak has been blackened (excavated) to prevent that peak being counted again.

the entire record. We have detected three dispersion clusters, near (in descending order of population) ~100 $s^{1/2}$, ~17 $s^{1/2}$, and ~126 $s^{1/2}$. By comparison, Figure 7b shows the peaks gathered in the original, old method algorithm. In each analysis window in the old method algorithm, there are now peaks at only a single phase dispersion. For most of the first 8 s of the record, the winning competitor on the phase-dispersion axis tends to be ~100 $s^{1/2}$, interspersed by window placements during which the top competitors are at ~17 $s^{1/2}$. Following $t = 8$ s, the favored dispersion jumps to ~17 $s^{1/2}$. There are no third dispersion peaks near ~126 $s^{1/2}$ anywhere in this old method treatment. The gross number of accepted peaks is much larger with the 2-D, new method (466 peaks; Figure 7a) than with the 1-D, old method dechirp method (111 peaks; Figure 7b).

Approximately 40 (or 9%) of the new method solution peaks in Figure 7a are not part of the three clusters and appear instead to be in more solitary locations on the plane. We have little confidence in those outliers' significance. One way to exclude them would be to include only those peaks which are within a cluster with other peers, but there is an even easier way. It turns out empirically that we can exclude the outliers by a simpler means. When we archive the solution peaks, we include an "energy width" giving the E^2-weighted peak width in time. *It turns out that the outlier peaks have systematically wider temporal widths than the dispersion-clustered peaks do*, and we find that a selection of suitably narrow peak width < 15 samples (0.5 ms) is satisfactory to eliminate most of the outliers.

4. Inferred Whistler Propagation in the Case Example

We now illustrate how the multiple values of phase dispersion in the example record correspond to multiple paths of propagation from the base of the ionosphere to the satellite. There is nothing in the equations of ray tracing dictating that there will be a unique, sole path from the base of the ionosphere to the satellite [*Kimura*, 1966] nor does any rule dictate that there can be only one unique profile of the polar angle θ along a given path.

C/NOFS burst-mode digitization was done only for the VEFI electric field signals, without simultaneous recording of magnetic field wave signals. Ordinarily, this lack of at least five electromagnetic wave components prevents determination of the wave vector at the satellite, but a recently developed technique can use the polarization state (from the three components of wave ***E***) to retrieve the wave vector [*Jacobson et al.*, 2014]. We now apply this wave vector retrieval for the example of new method dechirping seen in section 3.4.

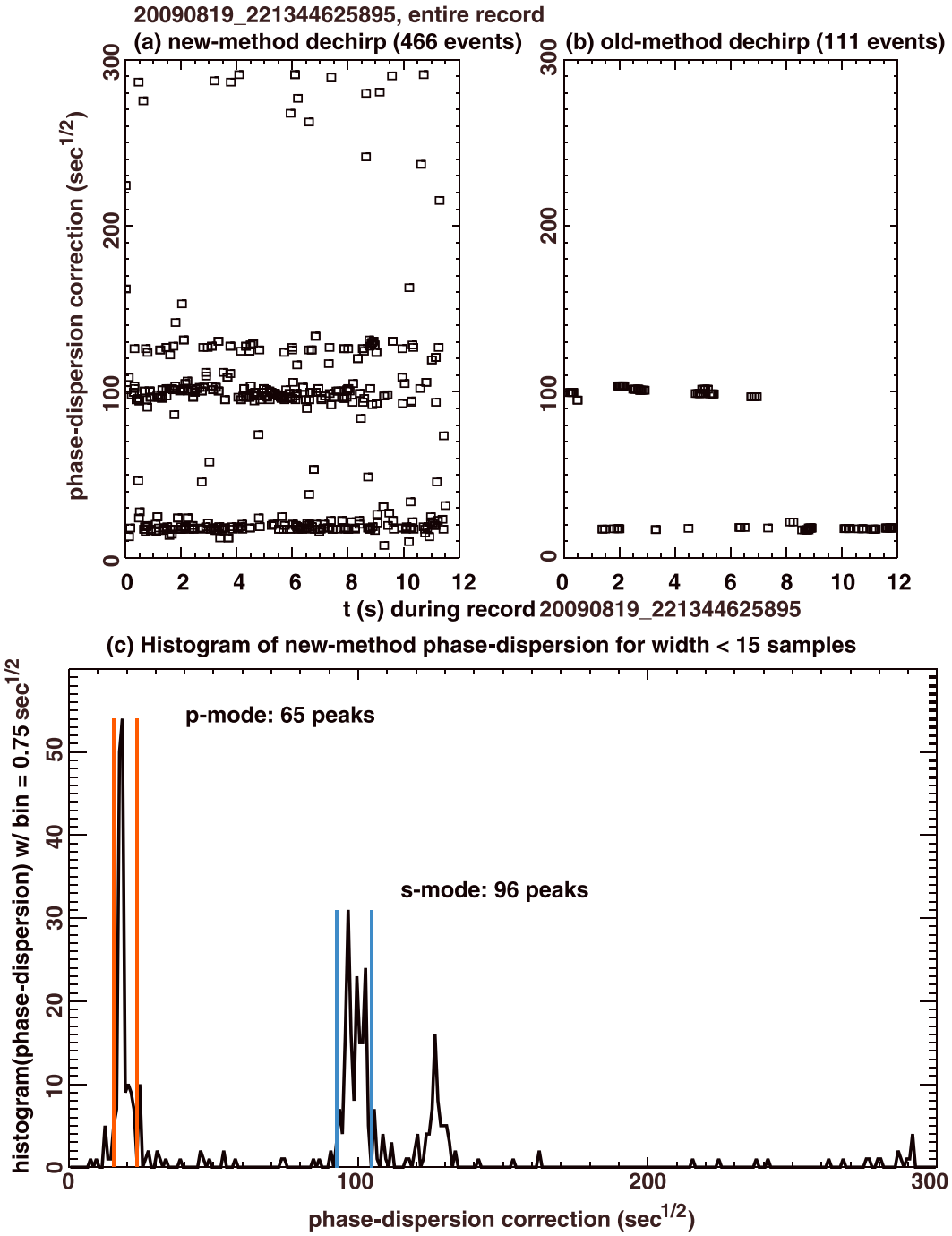

Figure 7. All detected and accepted peaks' phase dispersion versus time during the entire record, for (a) the new method algorithm (466 total peaks) and (b) in the original old method algorithm (111 total peaks). (c) The dispersion distribution for the narrow peaks (<15 samples energy width) for the entire, 12 s duration example record.

Figure 7c shows the phase-dispersion distribution for the narrow VEFI pulses (<15 samples energy peak width) for that example. There are two principal clusters ("p" and "s") which we shall use to illustrate the automated propagation approach. Figure 8 shows a moving-window spectrogram of the wave intensity (I_{stokes}) summed over (a) 65 p-cluster whistlers and (b) 96 s-cluster whistlers. The Fourier window is 32 samples (~1 ms) long and is advanced by two samples. The s cluster in Figure 8b has a narrower bandwidth than

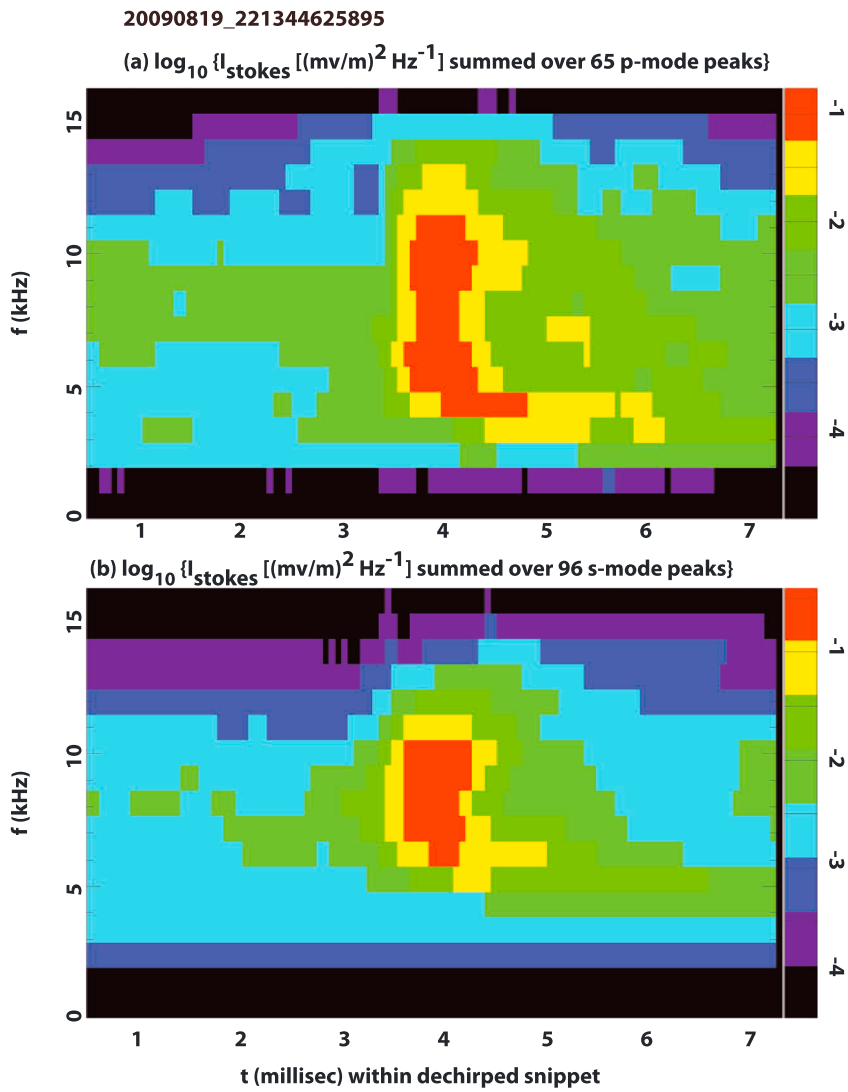

Figure 8. The wave intensity (Istokes) after dechirping, summed over (a) all 65 p cluster whistlers and (b) all 96 s cluster whistlers, versus time (horizontal) within 8 ms centered on the pulse, and versus frequency (vertical).

does the p cluster in Figure 8a. Both clusters exhibit near-complete high-pass filtering by the cutoffs in the Earth-ionosphere waveguide, and the coda at low frequencies is the "tweek" phenomenon [*Yamashita*, 1978] related to cutoff of the Earth-ionosphere waveguide. The color is for logarithmic scaling of intensity.

The two angles defining the direction of the wave vector $\boldsymbol{k}$ are θ, the polar angle of $\boldsymbol{k}$ with respect to the geomagnetic field **B**, and τ, the azimuthal angle of $\boldsymbol{k}$ around **B**. The azimuth convention is that $\tau = 0$ at magnetic zonal east. Our wave vector-retrieval method [*Jacobson et al.*, 2014] uses the measured polarization in the plane normal to **B**, plus the full dispersion relation (including ions), to infer the polarization in the plane normal to $\boldsymbol{k}$, which in turn allows us to infer $\boldsymbol{k}$. In Figure 9, we have regraphed intensity pixels summed within a propagation cluster (p or s), according to the polarimetry-derived determination of θ, τ. The event summing is for (a) the p dispersion and (b) the s dispersion. The summation is over peaks, over time, and over frequency. The color is for logarithmic scaling of intensity. Figure 9 reveals that there is an important difference in the wave vector of the two dispersions. The wave vector has polar angle near 55° for the p-cluster swarm of events but near 80° for the s-cluster swarm of events. Qualitatively, this is consistent with the direction of difference between the observed phase dispersions, 17 $s^{1/2}$ for the p cluster and 100 $s^{1/2}$ for the s cluster.

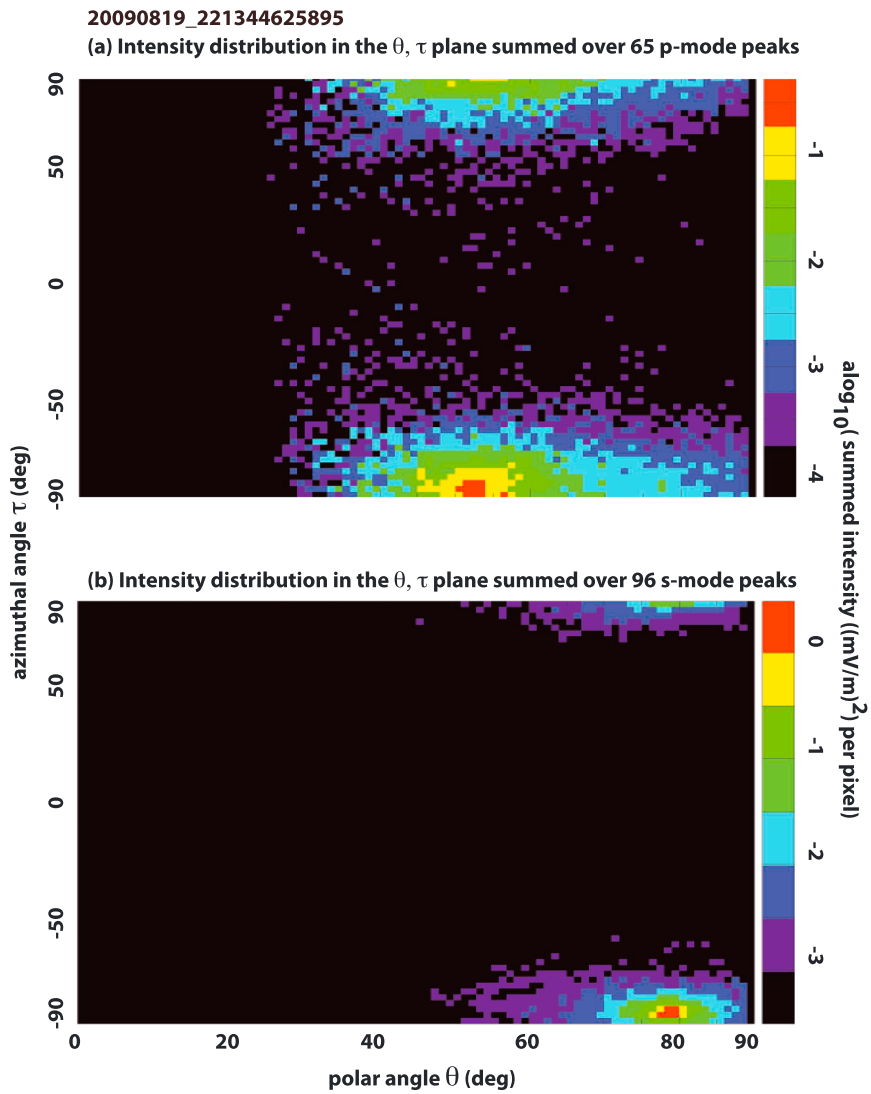

Figure 9. Intensity for each pixel, summed within a propagation cluster (p or s), according to the polarimetry-derived determination of θ, τ for that pixel. The event summing is for (a) the p dispersion cluster and (b) the s dispersion cluster. The color is for logarithmic scaling of intensity.

5. Performance Characteristics of the New Method Dechirping Algorithm

5.1. Temporal Correlation of VEFI Peaks With WWLLN Strokes

We have processed all suitable VEFI burst recordings with both the new method algorithm, and the original, old method algorithm. The suitable VEFI recordings are determined by requiring that they meet each of four additional criteria: First, they must have simultaneous recording of all three electric field antennas so that the polarimetric inference of wave vector can be done [*Jacobson et al.*, 2014]. Second, the record-averaged wave E spectral power parallel to $\mathbf{B}_0$ must be suitably small compared to the spectral power perpendicular to $\mathbf{B}_0$. Specifically, we require $\mathrm{avg}(E^2_{\mathrm{par}})/\mathrm{avg}(E^2_{\mathrm{perp}}) < 0.015$. This suffices to ensure adequately accurate polarimetry [*Jacobson et al.*, 2014]. Third, there must be simultaneous CINDI [*Heelis et al.*, 2009] ion composition data. Fourth, we require that root-mean-square electron-density fluctuations be relatively small: < 0.005. This is because stronger electron-density fluctuations tend to scatter the whistler waves into electrostatic lower hybrid waves [*Bell and Ngo,* 1988, 1990]. This restrictive set of burst recordings contains 6905 bursts, typically 12 s duration, spanning the date range from 28 August 2008 to 28 March 2014. The data set characteristics are summarized in Table 2.

During the entirety of these 6905 records, the new method dechirp algorithm identified 1.49 million dechirped peaks, compared to 0.34 million with the old method dechirp algorithm. Restricting to temporally narrow events (energy width $<$ 15 samples or ~ 0.5 ms), the peak counts fall to 0.39 million in the new, multi-dispersion algorithm and 0.10 million in the original, single-dispersion algorithm.

The World-Wide Lightning Location Network (WWLLN) [*Hutchins et al*., 2012, 2013] archive of global lightning stroke locations and times provides an indispensable ground truth for judging the validity of VEFI peaks identified by the new method dechirp algorithm. WWLLN has in turn been "calibrated" against continent-scale and regional-scale systems having very high stroke-detection efficiency (SDE), e.g., National Lightning Detection Network in the continental United States, known as NLDN [*Cummins and Murphy*, 2009]. Furthermore, a geographical *relative* calibration of WWLLN's SDE has been derived [*Hutchins et al*., 2012, 2013]. These efforts indicate that WWLLN achieves ~50% SDE in its best covered locales, e.g., the continental United States, and as low as $<$ 20% SDE in its most challenged areas, e.g., the Congo Basin, with a global mean on the order of 30%.

WWLLN has been demonstrated to have spatial location absolute accuracy within 15 km or less (depending on geographic region) [*Jacobson et al*., 2006]. The timing accuracy has been shown to be 30 μs or smaller. These spatial and temporal accuracies are sufficient to ensure that WWLLN may be regarded as a "ground-truth" stroke locater in comparisons with C/NOFS detections. Now consider the logical steps needed before we state that a given VEFI peak is likely "confirmed" by WWLLN ground truth:

1. Calculate the time differences between WWLLN stroke times, corrected for signal-propagation delay to the satellite, and the times of VEFI solutions. These are "corrected time differences."
2. Histogram these corrected time differences.
3. If there is *no* correlation peak in the histogram, then the only fair conclusion would be that there is no deterministic correspondence between the WWLLN and VEFI events.
4. If there *is* a correlation peak, recognizably higher than the "accidental" random floor, then one can demarcate the flanks of this peak and assert that WWLLN/VEFI pairs within these demarcations are possible true correlations.
5. The "accidental" random floor in the histogram, off to either sides of the correlation peak, shows the level of false coincidences. Thus, the chance that a WWLLN/VEFI pair within the correlation peak is *not* a true correlation will be approximately the ratio of the random floor level to the correlation peak level. Put another way, the signal-to-noise ratio (SNR) of the correlation peak is the ratio of the correlation peak level to the random floor level.

We demonstrate this process in Figure 10a, which presents a histogram of WWLLN-VEFI-corrected time differences. The histogram bin size is 1 ms. The histogram includes pairings between ~ 0.39 million candidate VEFI events with temporal peak width $<$ 15 samples and all ~ 0.41 million WWLLN events cumulatively occurring within the record epochs. There is a central correlation peak, approximately in the range -4 to +6 ms. On either sides of the peak is a dip to a minimum level, outside of which the baseline then recovers to a noise floor. Both the peak and the noise floor are indicated in Figure 10a by horizontal dashed lines. The dips below the noise floor, on either sides of the central correlation peak, are an artifact of the new method algorithm's inability to fully resolve VEFI events closer than ~ 20–30 ms of each other; this will be discussed in section 5.3 below. The peak-to-floor ratio, or SNR, in Figure 10a is 8:1, suggesting that even at the correlation peak, perhaps 12% of the inferred correlations, are likely false correlations, due to random time syncronicity.

5.2. Timing Skill of the VEFI Peak Solutions

The usual way of increasing the SNR of a time difference correlation peak is to use narrower histogram bins and to ensure that both data sources have time uncertainties that are as small as possible, no wider than the histogram bin. This is no problem with WWLLN, whose timing uncertainty for remotely located strokes is generally not worse than 10 μs [*Jacobson et al*., 2006]. Instead, the $>$ 5 ms width of the correlation peak in Figure 10a is entirely due to the way we identify peaks in the VEFI record. Referring to Figure 5 above, the solution for a peak's optimum time and optimum phase dispersion is a joint search in a 2-D plane. It is a matter of empirical observation that the 2-D search yields timing uncertainties of several milliseconds and phase-dispersion uncertainties of several $s^{1/2}$. Unfortunately, the two variables sought are not perfectly decorrelated from each other, and the solution search occurs in the presence of both noise and a failure of

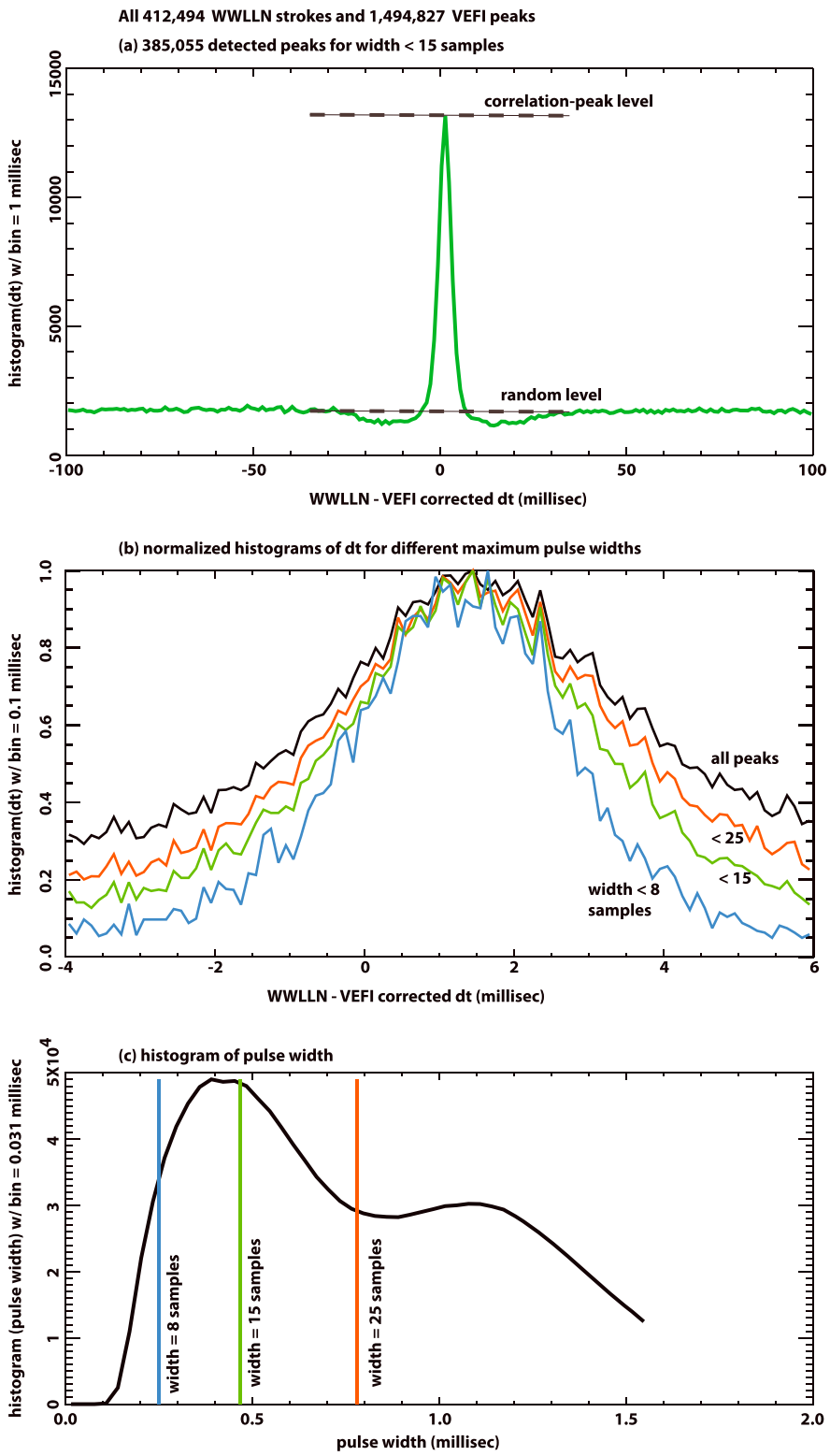

Figure 10. (a) Histogram of corrected WWLLN-VEFI time differences, with a 1 ms wide bin, using only those VEFI events with pulse width < 15 samples. The correlation peak value and the random noise floor are marked by horizontal black lines. (b) Histogram of corrected WWLLN-VEFI time differences, with a 0.1 ms wide bin, within the correlation zone −4 to +6 ms. Color marks the VEFI pulse width selection: Black: no selection; red: only < 25 samples; green: only < 15 samples; and blue: only < 8 samples. (c) Histogram of VEFI temporal pulse width. The colored lines mark specific width thresholds, with same color coding as in (Figure 10b).

the real signal to obey all of our approximations. The result is that our time/phase-dispersion solutions are uncertain to > 5 ms in timing and several $s^{1/2}$ in phase dispersion.

The intrinsic uncertainties of the time/phase-dispersion solutions in our method also prevent our realizing much further narrowing of the time difference histogram by using a more refined propagation model. The simple model used here (in "correcting" the WWLLN times for propagation to the satellite) does the following: (1) Assumes the signal propagates along a Great Circle path from the stroke location to the subsatellite point, at the speed of light; (2) adds a "vacuum" time delay for the signal to propagate from the subsatellite point vertically to the satellite, as if the plasma along the path did not exist; and (3) relies on the dechirp to correctly compensate for the contribution of the plasma to the time delay.

A more ambitious model for the propagation would include two refinements, which we have tested and found to be beyond the skill of the data:

1. Include small corrections in the group-propagation speed of the VLF energy in the Earth-ionosphere waveguide. However, this yields improvements of only tens of microseconds or less.
2. Ray-trace an actual path from the base of the ionosphere to the satellite. In practice, we reverse-ray-trace from the satellite to the base of the ionosphere, using the wave vector inferred from the wave polarization [*Abram R. Jacobson et al.*, 2014]. Then propagate the signal along a Great-Circle path from the stroke, *not to the sub-satellite point, but to the foot of the computed ray trace*. From the base of the ionosphere, add a "vacuum delay" as if the signal propagated along the actual raypath at the speed of light. *Due to the steepness of the ray trace paths we have studied, this refinement seldom makes a difference exceeding ~ 1 millisec*. Again, this is small compared to the timing skill of the VEFI peak solutions (~ 5 millisec).
3. Rely on the dechirp to correctly compensate for the contribution of the plasma to the time delay; this continues to work adequately.

The net conclusion is that the SNR in Figure 10a is probably not amenable to marked improvement by a more refined propagation model.

5.3. Signal Width as an Indicator of Signal Timing and Signal Quality

There is one last way of narrowing the time difference correlation peak, but it carries a cost. We have mentioned (section 3.4 above) that the width of the VEFI signal peak (after optimum dechirping) is related statistically to signal quality. The VEFI data contributing to Figure 10a have been selected to have width < 15 samples (~0.5 ms). Now we examine in more detail the effect of signal peak width on the timing skill of the dechirp method. Figure 10b shows relative histograms (each separately normalized to its maximum) of corrected time differences from WWLLN-VEFI. Unlike Figure 10a, the bin size is now finer (0.1 ms) and the total domain is only −4 to +6 ms. Each curve's color indicates the degree of width limitation: black for all widths, red for width < 25 samples, green for width < 15 samples (coinciding with Figure 10a), and blue for width < 8 samples. It is seen that as selection requires increasingly narrow pulse widths, the histogram becomes narrower too and the histogram correlation peak's SNR increases. There are two reasons why this happens:

1. We mentioned earlier (section 3.4 above) that requiring width < 15 samples seems to eliminate most of the possibly spurious peak solutions which are outside of clusters in the phase-dispersion distribution. Elimination of spurious events, that is, of events that are artifacts and can have no preferred temporal correspondence to WWLLN ground-truth strokes, can only improve the histogram correlation peak, making it narrower and higher relative to the background random noise in the histogram.
2. Peaks with more compact, that is, narrow, energy distributions are more likely to allow accurate time tagging. Therefore, it is not surprising that narrower pulse widths leads to narrower, and higher-SNR, time difference correlation peaks, as in Figure 10b.

The improvement of time-difference-correlation SNR is not, however, without cost. Within the "correlated" domain −4 to +6 ms, there are WWLLN/VEFI pairs involving 134,785 detected VEFI peaks with no width restrictions; 103,511 detected peaks having width < 25 samples; 66,368 detected peaks having width < 15 samples; and 16,659 detected peaks having width < 8 samples. Thus, not surprisingly, *there is a trade-off between quality of the peaks versus total number of the selected peaks*. Our particular compromise choice is to require width < 15 samples, but Figure 10b shows that depending on what is needed, one could restrict the width even further.

The overall distribution of event pulse width is shown in Figure 10c, with the abscissa in milliseconds. Vertical lines mark various widths in samples, with color coding the same as in Figure 10b. The 2-D search for events accepts peaks with widths only below 50 samples. Parenthetically, the distribution maxima in Figure 10c occur at +1.0 to +1.5 ms and not at 0. We are not able to rule out that this offset is traceable to the onboard timing on C/NOFS.

5.4. Inhibition of Neighbor Peak Detection

We mentioned earlier (sections 3.2,3.3 above) that the old method dechirp could resolve only one phase-dispersion value in each analysis window. Since the old method analysis window was advanced by 8192 samples, or roughly 1/4 s, this means that the old method dechirp was incapable of seeing more than one phase-dispersion value in a ~ 1/4 s period. Put another way, the old method could not see secondary solutions that differed in phase dispersion from the dominant peak but which occurred within roughly a ~1/4 s duration centered on the dominant peak. In this sense the detection of secondary peaks was inhibited if they were too close in time to a dominant peak. The new method dechirp largely avoids this problem. There is, however, a small mutual inhibition effect between solutions, whose limited scope we describe now, along with a comparison to the old method. To do this, we consider each event as a point at the origin (0, 0) of a plane. The event at the origin is called the "key" event. The horizontal axis in this plane is the neighbor-event minus key-event time difference. The vertical axis is the neighbor event minus key event phase-dispersion difference. This is a superposed epoch distribution in two dimensions. The distribution is summed over all the VEFI events with width < 15 samples. Each of those key events has neighbors in a 2-D plane centered at the key event. These pixelated superposed epoch distributions are shown in Figure 11. Figures 11a and 11c is for the new method dechirp. Figures 11b and 11d is for the old method dechirp. Color of each pixel codes the logarithm (base 10) of the neighbor population in that pixel, summed over all key events. Each pixel is 0.01 s wide in time and 0.75 $s^{1/2}$ tall in phase dispersion. The color scales for all four panels are exactly the same. Figures 11a and 11b shows the superposed epoch distribution for a time range of ± 1 s. Figures 11c and 11d shows the innermost ± 0.2 s. The color scale covers five decades. The normalization is by the brightest pixel on the page, which occurs on the left side.

The intense horizontal feature at zero relative phase dispersion in Figure 11 (all panels) corresponds to the tendency of detected VEFI peaks to occur within clusters of similar phase dispersion. The more diffuse and less intense regions above and below the bright band correspond to occasional occurrences of multiple phase dispersion. These occurrences are more populated in the new method (Figures 11a and 11c) then in the old- method (Figures 11b and 11d) dechirp.

The upward arrow in Figure 11b shows the region (vertical black band) of inhibition exerted against any neighbors with different phase dispersion and within about ~1/4 s relative to the key event in the old method dechirp. Figures 11b and 11d reveal that the region of inhibition in the old method dechirp has ragged and uneven borders in time but extends uniformly over the entire phase-dispersion range.

The new method dechirp (Figures 11a and 11c) also has inhibition effects, but they are more limited than in the old method dechirp:

1. First, note the white rectangle at the origin, barely visible in Figure 11a but more visible in the expanded view; Figure 11c. This is the excavated rectangle which is applied to any newly accepted peak solution. The iterative method of peak finding in the 2-D plane involves a form of decimation, in this case complete excavation of the zone around the newly found peak. The excavation mitigates against that peak being present in subsequent iterations; the excavation's finite extent also mitigates against remaining flanks of that peak appearing as solutions in subsequent iterations. The excavation is wide enough in time and high enough in phase dispersion, to fully remove the flanks in either direction.
2. Second, note the faint depletion along a diagonally trending cone in Figure 11c. This is an inevitable consequence of the way that the new method peak-finder works. In Figures 5a and 6, each candidate event occupies a diagonally trending cone. The energy of the candidate event is most concentrated at pixels near the apex of the cone and is more spread out as one goes up or down (in phase dispersion) from the candidate peak. Even this spread-out energy can cover and obscure another peak happening to possess an apex within the cone. That leads to a statistically significant depletion of neighbor events within the search cone in the superposed epoch distribution of Figure 11c. Comparison of

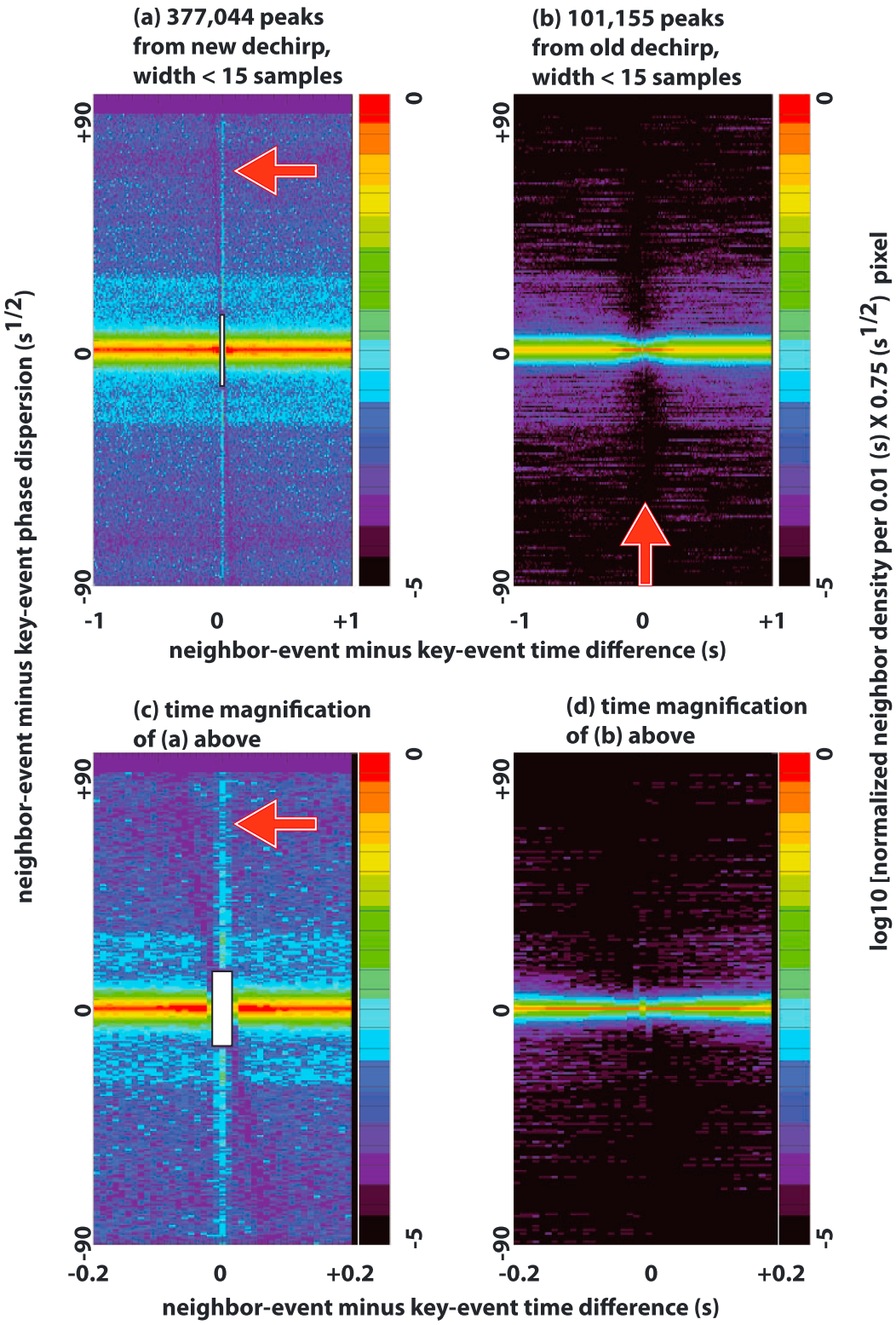

Figure 11. Two-dimensional superposed epoch distribution of neighbor events, in the plane with horizontal = time difference, and vertical = phase-dispersion difference. The origin (0, 0) is the key event, and neighbors in time/dispersion are counted as pixel population of the surrounding pixels. The population is shown logarithmically on the color scale. (a) VEFI peaks from new method dechirp, with pulse width < 15 samples. (b) VEFI peaks from old method dechirp, with pulse width < 15 samples. (c) Time magnification of VEFI peaks from new method dechirp, with pulse width < 15 samples. (d) Time magnification of VEFI peaks from old method dechirp, with pulse width < 15 samples. See text.

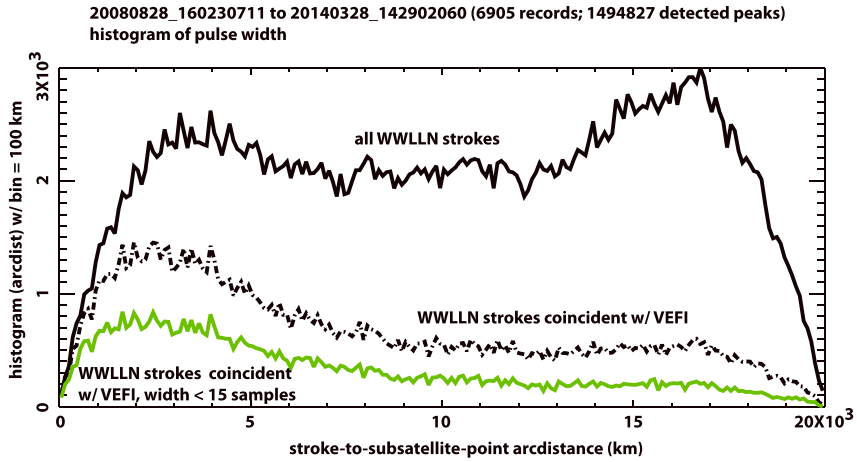

Figure 12. Histograms of Great-Circle distance from subsatellite point to WWLLN stroke locations. Solid black curve: All WWLLN strokes occurring during the cumulative 6905 recording epochs. Dashed black curve: Only WWLLN strokes temporally coincident with VEFI new method detected pulses. Green curve: Only WWLLN strokes temporally coincident with VEFI new method detected pulses having width $<$ 15 samples.

the left side, versus the right side of Figure 11 reveals that this new method inhibition effect is very small compared to that of the old method.

Most important for our application (study of upward wave propagation to the satellite) is that the new method inhibition zone allows detection of the simplest form of multiple dispersion: *simultaneous signal arrival along two (or more) multiple paths*. Simultaneous multipath neighbor events are completely destroyed in the old method, but in the new method dechirp, they appear as a vertical enhancement in the pixel population above and below the key event. See the red arrow in Figures 11a and 11c.

5.5. Distance Distribution

The distance distribution of strokes on the planet, relative to C/NOFS and simultaneous with VEFI burst recordings, is highly biased and is not even remotely a uniform sampling of all areas on the planet. The biases arise from two general sources: First, lightning has a biased distribution in local time and in location. Second, the C/NOFS orbit and burst-recording program imposed additional biases.

Most of the lightning on Earth occurs over land, not water [*Williams and Stanfill*, 2002], and within three longitudinal zones [*Christian et al.*, 1999]. These three zones cross the geographical Equator near Colombia, the Congo Basin, and the Maritime Continent, respectively. Within each zone, the annual average lightning maximizes in the tropics but shows secondary hot spots into midlatitudes. High-latitude lightning is less common on a statistical basis, although some occurs. Finally, terrestrial lightning is local-time selective; the superposition of the local-time maxima in the three favored longitudinal zones is responsible for the well-known "Carnegie curve" of global lightning occurrence versus universal time [*Williams*, 2009].

The C/NOFS orbit at 13° inclination and the satellite's low altitude cause it to "see" the global lightning distribution as quasi-two-dimensional, with the most lightning signals arriving zonally within the Earth-ionosphere waveguide and less arriving meridionally. One of several reasons for this is that the Antarctic ice's lossy electrical properties provide a near-complete barrier to transpolar VLF signal propagation in the Earth-ionosphere waveguide in the Southern Hemisphere [*Said et al.*, 2010]. Another reason is that both the Greenland ice sheet and the Arctic sea ice to some extent attenuate transpolar VLF signal propagation over the Northern Hemisphere as well [*Hutchins et al.*, 2012, 2013].

The VEFI burst-mode recording program heavily selected for local nighttime. Additionally, the 6905 records selected for this study are chosen to avoid periods when the satellite was transiting ionospheric irregularities; this selection imposes additional local-time bias in the data set [*de La Beaujardiere*, 2004; *Pfaff et al.*, 2010]. Putting all these various biases together, it is not surprising that we see the distance distributions shown in Figure 12. The upper black, solid curve is the distribution of Great Circle distance *to all the WWLLN strokes that*

occur during the accumulated 6905 records. Were the recording-contemporaneous lightning incidence independent of longitude and latitude, then the curve would peak at the center (10,000 km). However, the various biases discussed above produce almost the opposite: a near-proximal peak (3000–5000 km) and a near-antipodal peak (16,000–17,000 km). Between the peaks is a trough, precisely where the uniform-lightning peak would have been. The only feature of the overall WWLLN-stroke distance distribution that comports with a uniform-lightning model is the presence of a steep roll-off at either ends. That is a simple consequence of the area element's steep descent, either approaching zero or approaching 20,000 km (the antipode). Put more formally, the abscissa in Figure 12 is arcdistance. The area element dA of an annular ring on the Earth's surface varies versus arcdistance as the circumference of a geographic Small Circle at that arcdistance:

$$\frac{dA}{d(arcdistance)} = \text{circumference} = 2\pi * R_E * \sin\left\{\frac{\pi * \text{arcdistance}}{20,000\ km}\right\} \tag{6}$$

where R_E is the Earth's radius. This explains why the distributions in Figure 12 tend steeply to zero both proximally (arcdistance approaching 0) and antipodally (arcdistance approaching 20,000 km).

The dashed black curve in Figure 12 is the distance distribution of WWLLN strokes that are coincident with VEFI detected events using the new method dechirp. By "coincident" we mean having corrected WWLLN-VEFI time differences in the range −4 to +6 ms. Clearly, the dashed black curve has a different shape than does the solid black curve. That tends to validate the notion that the coincidences are, mostly, not accidental. Were the coincidences completely accidental, then the dashed curve would necessarily have the same shape as the solid curve. For distance $<$ 1000 km, most of the WWLLN strokes have corresponding VEFI detected peaks. At the proximal peak of the solid curve (3000–5000 km), VEFI still detects events coincident with $>$ 50% of the WWLLN strokes. At the near-antipodal peak (16,000–17,000 km), VEFI detects events coincident with $>$ 20% of the WWLLN strokes. The green curve is similar to the dashed black curve, except that the VEFI events are only the narrow events (peak width $<$ 15 samples). The narrow VEFI events coincide with ~ 25% of the WWLLN strokes at the proximal peak and then descend to ~ 10% approaching the antipode.

These percentage coincidences inferred from Figure 12 pertain to the percentage of WWLLN strokes that are coincidentally detected by VEFI and not the other way around.

6. Conclusions

We have established an archive of three-axis VEFI burst recordings of wave **E**, having suitably small projection of wave **E** onto the geomagnetic field $\mathbf{B}_o$, occurring during periods of quiet, undisturbed ionospheric conditions and having simultaneous CINDI ion composition data. This selected archive contains 6905 separate bursts, typically of 12 s duration each burst.

We have developed a new whistler dechirping, whistler identification, and phase-dispersion retrieval algorithm that has no trouble in the presence of multiple dispersions, as from multiple propagation paths from the whistler source to the C/NOFS satellite. This new algorithm differs from the old method algorithm, which could lock onto only one dispersion within an analysis window. The total number of detected discrete whistlers (accumulated from all 6905 bursts) is 1.49 million in the new method algorithm compared to 0.38 million in the original, old method algorithm.

We find that the narrowest peaks (energy width $<$ 15 samples or 0.5 ms) have the best reliability and rate of coincidence with WWLLN ground truth. Using narrow peaks only, we find that there are 64,807 in the new algorithm that are coincident with WWLLN and thus can be attributed with source location.

The old and new method algorithms impose different inhibitions on neighboring events. The old method algorithm cannot detect simultaneous signal arrivals at multiple phase dispersions, while the new method algorithm can.

Acknowledgments
This work was partially supported by NSF grant 1443011, "Wave-vector-resolved Study of Lightning Whistler Propagation and Energetics in the Low-latitude Plasmasphere." Readers wishing to examine the original data are invited to contact the corresponding author.

References

Bell, T. F., and H. D. Ngo (1988), Electrostatic waves stimulated by coherent VLF signals propagating in and near the inner radiation belt, *J. Geophys. Res.*, *93*, 2599–2618, doi:10.1029/JA093iA04p02599.

Bell, T. F., and H. D. Ngo (1990), Electrostatic lower hybrid waves excited by electromagnetic whistler mode waves scattering from planar magnetic-field-aligned plasma density irregularities, *J. Geophys. Res.*, *95*, 149–172, doi:10.1029/JA095iA01p00149.

Berthelier, J. J., et al. (2006), ICE, the electric field experiment on DEMETER, *Planet. Space Sci.*, *54*, 456–471, doi:10.1016/j.pss.2005.10.016.

Bilitza, D., S. A. Brown, M. Y. Wang, J. R. Souza, and P. A. Roddy (2012), Measurements and IRI model predictions during the recent solar minimum, *J. Atmos. Sol. Terr. Phys.*, *86*, 99–106, doi:10.1016/j.jastp.2012.06.010.

Bortnik, J., U. S. Inan, and T. F. Bell (2003), Frequency-time spectra of magnetospherically reflecting whistlers in the plasmasphere, *J. Geophys. Res.*, *108*(A1), 1030, doi:10.1029/2002JA009387.

Bortnik, J., U. S. Inan, and T. F. Bell (2006a), Temporal signatures of radiation belt electron precipitation induced by lightning-generated MR whistler waves: 1. Methodology, *J. Geophys. Res.*, *111*, A02204, doi:10.1029/2005JA011182.

Bortnik, J., U. S. Inan, and T. F. Bell (2006b), Temporal signatures of radiation belt electron precipitation induced by lightning-generated MR whistler waves: 2. Global signatures, *J. Geophys. Res.*, *111*, A02205, doi:10.1029/2005JA011398.

Burkholder, B. S., M. L. Hutchins, M. P. McCarthy, and R. H. Holzworth (2013), Attenuation of lightning-produced sferics in the Earth-ionosphere waveguide and low-altitude ionosphere, *J. Geophys. Res. Space Physics*, *118*, 3692–3699, doi:10.1002/jgra.50351.

Christian, H. J., et al. (1999), Global frequency and distribution of lightning as observed by the Optical Transient Detector (OTD), paper presented at 11th International Conference on Atmospheric Electricity, National Aeronautics and Space Administration (U. S.), Huntsville, Ala.

Chum, J., F. Jiricek, O. Santolik, M. Parrot, G. Diendorfer, and J. Fiser (2006), Assigning the causative lightning to the whistlers observed on satellites, *Ann. Geophys.*, *24*, 2921–2926.

Chum, J., O. Santolik, and M. Parrot (2009), Analysis of subprotonospheric whistlers observed by DEMETER: A case study, *J. Geophys. Res.*, *114*, A02307, doi:10.1029/2008JA013585.

Cummer, S. A. (2000), Modeling electromagnetic propagation in the Earth-ionosphere waveguide, *IEEE Trans. Antennas Propagations*, *48*(9), 1420–1429.

Cummins, K. L., and M. J. Murphy (2009), An overview of lightning location systems: History, techniques, and data uses, with an in-depth look at the U.S. NLDN, *IEEE Trans. Electromagn. Compat.*, *51*(3), 499–518.

de La Beaujardiere, O. (2004), C/NOFS: A mission to forecast scintillations, *J. Atmos. Sol. Terr. Phys.*, *66*, 1537–1591, doi:10.1016/j.jastp.2004.07.030.

Draganov, A. B., U. S. Inan, V. S. Sonwalkar, and T. F. Bell (1992), Magnetospherically reflected whistlers as a source of plasmaspheric hiss, *Geophys. Res. Lett.*, *19*, 233–236, doi:10.1029/91GL03167.

Fiser, J., J. Chum, G. Diendorfer, M. Parrot, and O. Santolik (2010), Whistler intensities above thunderstorms, *Ann. Geophys.*, *28*, 37–46.

Green, J. L., S. Boardsen, L. Garcia, W. W. L. Taylor, S. F. Fung, and B. W. Reinisch (2005), On the origin of whistler mode radiation in the plasmasphere, *J. Geophys. Res.*, *110*, A03201, doi:10.1029/2004JA010495.

Gurnett, D. A., S. D. Shawhan, N. M. Brice, and R. L. Smith (1965), Ion cyclotron whistlers, *J. Geophys. Res.*, *70*, 1665–1668, doi:10.1029/JZ070i007p01665.

Hayakawa, M., K. Ohta, and K. Baba (1994), Wave characteristics of tweek atmospherics deduced from the direction-finding measurement and theoretical interpretation, *J. Geophys. Res.*, *99*, 10,733–10,744, doi:10.1029/93JD02555.

Heelis, R. A., W. R. Coley, A. G. Burrell, M. R. Hairston, G. D. Earle, M. D. Perdue, R. A. Power, L. L. Harmon, B. J. Holt, and C. R. Lippincott (2009), Behavior of the O+/H+ transition height during the extreme solar minimum of 2008, *Geophys. Res. Lett.*, *36*, L00C03, doi:10.1029/2009GL038652.

Helliwell, R. A. (2006), *Whistlers and Related Ionospheric Phenomena*, 349 pp., Dover Publications, Mineola, N. Y.

Hutchins, M. L., R. H. Holtzworth, C. J. Rodger, and J. B. Brundell (2012), Far-field power of lightning strokes as measured by the World Wide Lightning Location Network, *J. Atmos. Oceanic Technol.*, *29*, 1102–1110, doi:10.1175/JTECH-D-1111-00174.

Hutchins, M. L., A. R. Jacobson, R. H. Holzworth, and J. B. Brundell (2013), Azimuthal dependence of VLF propagation, *J. Geophys. Res. Space Physics*, *118*, 1–5, doi:10.1002/jgra.50533.

Jacobson, A. R., R. Holzworth, J. Harlin, R. Dowden, and E. Lay (2006), Performance assessment of the World Wide Lightning Location Network (WWLLN), using the Los Alamos Sferic Array (LASA) array as ground-truth, *J. Atmos. Oceanic Technol.*, *23*, 1082–1092.

Jacobson, A. R., R. H. Holzworth, R. F. Pfaff, and M. P. McCarthy (2011), Study of oblique whistlers in the low-latitude ionosphere, jointly with the C/NOFS satellite and the World-Wide Lightning Location Network, *Ann. Geophys.*, *29*, 851–863, doi:10.5194/angeo-29-851-2011.

Jacobson, A. R., R. H. Holzworth, R. F. Pfaff, R. Heelis, and P. Colestock (2014), A method to estimate whistler wavevector from polarization using 3-component satellite E-field data, *Radio Sci.*, *49*, 131–145, doi:10.1002/2013RS005335.

Kimura, I. (1966), Effects of ions on whistler-mode ray tracing, *Radio Sci.*, *1*, 269–284, doi:10.1002/rds196613269.

Klenzing, J., F. Simoes, S. Ivanov, R. A. Heelis, D. Bilitza, R. Pfaff, and D. Rowland (2011), Topside equatorial ionospheric density and composition during and after extreme solar minimum, *J. Geophys. Res.*, *116*, A12330, doi:10.1029/2011JA017213.

Meredith, N. P., R. B. Horne, M. A. Clilverd, D. Horsfall, R. M. Thorne, and R. R. Anderson (2006), Origins of plasmaspheric hiss, *J. Geophys. Res.*, *111*, A09217, doi:10.1029/2006JA011707.

Meredith, N. P., R. B. Horne, S. A. Glauert, and R. R. Anderson (2007), Slot region electron loss timescales due to plasmaspheric hiss and lightning-generated whistlers, *J. Geophys. Res.*, *112*, A08214, doi:10.1029/2007JA012413.

Pfaff, R., et al. (2010), Observations of DC electric fields in the low-latitude ionosphere and their variations with local time, longitude, and plasma density during extreme solar minimum, *J. Geophys. Res.*, *115*, A12324, doi:10.1029/2010JA016023.

Said, R. K., U. S. Inan, and K. L. Cummins (2010), Long-range lightning geolocation using a VLF radio atmospheric waveform bank, *J. Geophys. Res.*, *115*, D23108, doi:10.1029/2010JD013863.

Santolik, O., M. Parrot, U. S. Inan, D. Buresova, D. A. Gurnett, and J. Chum (2009), Propagation of unducted whistlers from their source lightning: A case study, *J. Geophys. Res.*, *114*, A03212, doi:10.1029/2008JA0113776.

Stix, T. H. (1962), *The Theory of Plasma Waves*, McGraw Hill Book Company, New York.

Williams, E., and S. Stanfill (2002), The physical origin of the land-ocean contrast in lightning activity, *Comptes Rendus Phys.*, *3*, 1277–1292.

Williams, E. R. (2009), The global electric circuit: A review, *Atmos. Res.*, *91*, 140–152, doi:10.1016/j.atmosres.2008.05.018.

Yamashita, M. (1978), Propagation of tweek atmospherics, *J. Atmos. Sol. Terr. Phys.*, *40*, 151–156.