Ann. Geophys., 26, 305–314, 2008
www.ann-geophys.net/26/305/2008/
© European Geosciences Union 2008

Annales Geophysicae

A statistical approach for identifying the ionospheric footprint of magnetospheric boundaries from SuperDARN observations

G. Lointier[1], T. Dudok de Wit[1], C. Hanuise[1], X. Vallières[1], and J.-P. Villain[1]

[1]Laboratoire de Pysique et Chimie de l'Environnement, UMR 6115 CNRS/University of Orléans, 3A avenue de la recherche Scientifique, 45071 Orléans cedex 2, France

Received: 1 February 2007 – Revised: 10 September 2007 – Accepted: 10 September 2007 – Published: 26 February 2008

Abstract. Identifying and tracking the projection of magnetospheric regions on the high-latitude ionosphere is of primary importance for studying the Solar Wind-Magnetosphere-Ionosphere system and for space weather applications. By its unique spatial coverage and temporal resolution, the Super Dual Auroral Radar Network (SuperDARN) provides key parameters, such as the Doppler spectral width, which allows the monitoring of the ionospheric footprint of some magnetospheric boundaries in near real-time. In this study, we present the first results of a statistical approach for monitoring these magnetospheric boundaries. The singular value decomposition is used as a data reduction tool to describe the backscattered echoes with a small set of parameters. One of these is strongly correlated with the Doppler spectral width, and can thus be used as a proxy for it. Based on this, we propose a Bayesian classifier for identifying the spectral width boundary, which is classically associated with the Polar Cap boundary. The results are in good agreement with previous studies. Two advantages of the method are: the possibility to apply it in near real-time, and its capacity to select the appropriate threshold level for the boundary detection.

Keywords. Ionosphere (Auroral ionosphere) – Magnetospheric physics (Magnetosphere-ionosphere interactions)

1 Introduction

The interactions between the interplanetary medium and the Earth's magnetosphere have rapid consequences on the high-latitude ionosphere through reconfiguration of the magnetic field and through particle precipitation. Ground-based instruments located at high latitude offer a unique opportunity to monitor these phenomena in near real-time. One of the important parameters for space weather applications is the location of the ionospheric footprint of the various magnetospheric regions. We investigate here the possibility of tracking these footprints using the Super Dual Auroral Radar Network (SuperDARN) (Greenwald et al., 1995), which was originally designed to map ionospheric convection patterns at high latitude.

Many studies have revealed a connection between the statistical characteristics of the SuperDARN backscattered echoes and magnetic latitude or magnetic local time (Baker et al., 1995; Rodger et al., 1995; Pinnock et al., 1999; André et al., 2002; Villain et al., 2002). This has led to the idea that SuperDARN data could be used as a proxy for identifying the magnetospheric boundaries (Rodger, 2000). Most of the attention has focussed so far on the Doppler spectral width, which is derived by fitting a model to the autocorrelation function (ACF) of the backscattered signal. Although the interpretation of this spectral width is still elusive, Baker et al. (1995) and Rodger et al. (1995) have shown that it is enhanced at the ionospheric footprint of the cusp.

While a sharp latitudinal gradient in the Doppler spectral width (marking the latitudinal boundary between narrow and large spectral width) is frequently observed all around the polar ionosphere with SuperDARN, the relationship between the spectral width boundary (SWB) and physical boundaries is not obvious and seems to depend on magnetic local time (MLT). This relationship has been investigated in detail in several studies (Dudeney et al., 1998; Lester et al., 2001; Parkinson et al., 2002; Woodfield et al., 2002a,b; Chisham et al., 2004, 2005c,a). For the 18:00–02:00 and 08:00–12:00 MLT sector, the locations of the SWB and the open/closed field line boundary (OCB) seem to agree well. For the post-midnight and afternoon sectors, the SWB is located equatorward of the OCB, on the closed field lines that are co-located with the separation between the central plasma sheet and the plasma sheet boundary layer. The OCB is defined as the poleward border of the red line emission of the

Correspondence to: G. Lointier
(guillaume.lointier@cnrs-orleans.fr)

UV spectra (630 nm). In terms of electron precipitation, the OCB is equivalent to the boundary between the polar rain and precipitation from the boundary plasma sheet. Chisham and Freeman (2004) noticed that the nature of the transition (in terms of gradient and amplitude) is well-defined in a wide range of MLT, for which the comparison of the SWB with the OCB is precisely consistent.

Other studies have used the spectral width in order to deduce additional quantities such as the reconnection rate at the magnetopause (Baker et al., 1997; Pinnock et al., 1999), whose variations are of great interest for the understanding of the solar wind/magnetosphere interaction.

The spectral width is not sufficient for detecting boundaries, as its estimation suffers from some weaknesses. In particular, the models that are used to fit the backscattered spectrum (from which the spectral width value is computed) are not suitable for describing the variability in spectrum shapes. Indeed, André et al. (2002) have shown that the identification of boundaries can be improved by incorporating two more parameters (the quality of the regression of a model to the phase and to the modulus of the ACF). By applying ad-hoc thresholds, they defined three classes, each of them expressing one specific predominant physical process that affects the backscattered signal (such as microscopic plasma turbulence, Pc1 wave activity and large-scale vortices). From a statistical point of view, the authors revealed a good correlation between the projection of these classes onto a magnetic latitude and MLT frame and the ionospheric footprint of some magnetospheric regions. Their results indicate that the error made on the spectral width and on the drift velocity estimation offers a complementary description of the backscattering ionospheric regions. By performing a multivariate statistical analysis in which no classes were defined a priori, André and Dudok de Wit (2003) confirmed this result.

All studies involving SuperDARN data proceed by fitting a model to the complex ACF of the backscattered signal and interpreting parameters of that model, such as the Doppler spectral width. Having in mind a real-time monitoring of magnetospheric boundaries, we follow a different approach and focus instead on the statistical properties of the backscattered echoes. Using the singular value decomposition (SVD) as a data reduction technique, we show that these echoes can be described with three statistical parameters only. Using two of these parameters, we define a criterion for detecting the ionospheric footprint of magnetospheric boundaries.

The data selection used in this statistical study is discussed in Sect. 2. The data reduction method and its physical interpretation are respectively presented in Sects. 3 and 4. A case study is discussed in Sect. 5 and conclusions follow in Sect. 6.

2 SuperDARN data overview

2.1 Characterisation of the backscattered echoes

SuperDARN is an international network of HF coherent scatter radars that was designed for the study of ionospheric plasma convection over the auroral and polar regions (Greenwald et al., 1995). There are twelve and seven radars in operation in the Northern and Southern Hemisphere, respectively. Each radar consists of an array of 16 log-periodic antennas, from which 16 sequential beams are formed through electronic phase shifting. This provides an overall azimuth coverage of $\simeq 50°$. The transmission of a multi-pulse scheme, repeated through the integration time ($\simeq 7$ s in common sounding mode), gives access to the complex ACF of the backscattered signal by the field-aligned irregularities as a function of distance and azimuth. A typical cell size is $45 \times 100\,\mathrm{km}^2$, depending on distance from the radar. For each cell, three quantities are routinely estimated every 2 min by the FitACF algorithm (Greenwald et al., 1985; Villain et al., 1987; Hanuise et al., 1993; Baker et al., 1995):

- the Doppler spectral width, estimated from the decorrelation time of the modulus (or power) of the ACF: $|R(\tau)| = \sqrt{\Re[R(\tau)]^2 + \Im[R(\tau)]^2}$;
- the power of the backscattered signal: $|R(\tau = 0)|$;
- the line-of-sight Doppler velocity of the irregularities, estimated from the temporal drift of the ACF phase: $\varphi(\tau) = \arctan \Im[R(\tau)]/\Re[R(\tau)]$.

assuming a one-component spectrum, the Doppler spectral width is analytically estimated by fitting either a Lorentzian ($|R| \sim e^{-\lambda\tau}$) or a Gaussian function ($|R| \sim e^{-\sigma^2\tau^2}$) to the modulus of the ACF. Some ACFs are unambiguously Lorentzian or Gaussian, but most are in between. While this distinction between different models provides interesting insight into the microphysics of turbulence (Hanuise et al., 1993; Villain et al., 1996), it also questions the relevance of the Doppler spectral width as a proper parameter for characterising the backscattered signal. Because of this ambiguity, we bypass here the FitACF routine and choose to work directly on the raw ACF data. Our prime objective is to characterise the statistical properties of the ACFs without imposing a priori a physical model.

The phase and the modulus of the ACF embody all the available information and therefore are the starting point for our analysis. These two quantities, however, convey different information. The phase $\varphi(\tau)$ and its dependence on lag τ is associated with the bulk plasma velocity, and for a given plasma cell its value depends on the observing radar. The modulus $|R(\tau)|$ mainly bears the signature of the drift velocity distribution and the average lifetime of the irregularities inside the scattering volume (Hanuise et al., 1993; Villain

Table 1. Selection criteria for the our database of ACFs.

S/N ratio	$\geq$15 dB
Number of bad-lags	0
ErrPwr	$\leq$0.4
Spectral width	[60, 600] m s^{-1}
Distance from radar	[900, 3500] km

et al., 1996; Ponomarenko and Waters, 2003). The backscattered power ($|R(\tau=0)|$) is connected to the presence of field-aligned irregularities and to propagation effects, including the absorption of the radio waves.

Earlier studies have revealed the importance of the modulus for locating boundaries (Rodger, 2000; André et al., 2002; Hosokawa et al., 2002; André and Dudok de Wit, 2003; Chisham and Freeman, 2003). For that reason, we shall focus here on the modulus only rather than both the modulus and the phase. Unless specified otherwise, the term ACF will always stand for the modulus of the complex ACF. Future improvements of our statistical model will necessarily include a statistical description of the full complex ACF.

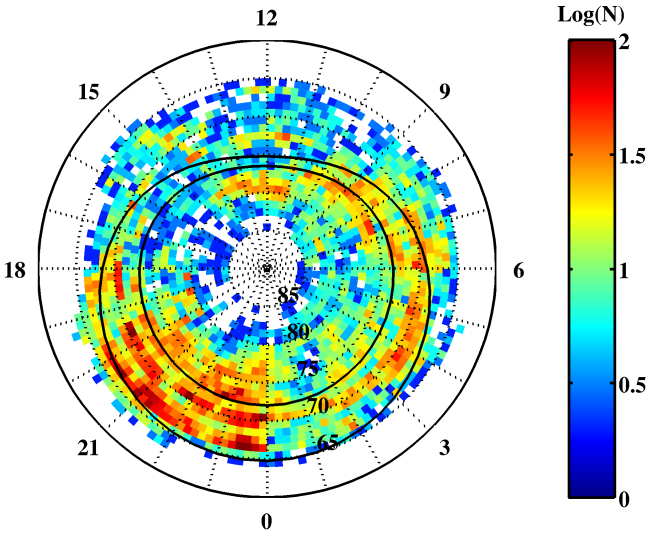


Fig. 1. Spatial distribution of the occurrence of radar echoes in our database, in MLT/Λ (Invariant Magnetic Latitude) AACGM coordinates. The occurrence is colour-coded on a logarithmic scale. The boundaries of the auroral oval, defined by Holzworth and Meng (1975) for a moderate geomagnetic activity ($3<K_p<4^-$) are overlaid.

2.2 Building the database

Before performing the statistical analysis, it is essential to compile a database that is both complete and representative of the different types of echoes. In order to do so, we consider all echoes recorded in the Northern Hemisphere during four typical days, one for each season: 12 February 2003, 24 May 2003, 1 September 2003 and 21 December 2003. These days have been chosen both for their high number of echoes and for their good spatio-temporal coverage of the ionospheric convection. The statistical model we obtain from these events does not significantly change if more days or radars with other field-of-views are included in the data set. This result is crucial, as it justifies the procedure whose description follows.

The solar wind conditions for these days reveal a moderate geomagnetic activity ($K_p<4$) with a regular northward to southward rotation of the Bz component. This could explain the high occurrence of the radar echoes with a regular injection of particles in the magnetosphere. No strong geomagnetic event (interplanetary shocks, geomagnetic substorms...) has been reported.

Our database must contain the largest possible number of radar echoes from different regions but several selection criteria also need to be applied in order to remove spurious data. Only data recorded during common operating times are selected in order to avoid any contamination from specific modes. Care is taken to select backscattered echoes originating from the F region only because at this altitude the plasma is driven by the magnetospheric activity. To eliminate echoes from the E region, we only select ACFs associated with ranges higher than 900 km from the radar site (André et al., 2002; Villain et al., 2002). The spectral width is extracted by least-squares fitting the ACFs with a Gaussian and a Lorentzian. The normalised residual error (ErrPwr) is indicative of the quality of the fit. We conserve data for which this normalised residual error is lower than 0.4. We also exclude ACFs whose spectral width is below 60 m s^{-1} to remove any contamination from ground scatter. The data are frequently corrupted by one or several spurious values (called "bad-lags") whose origin is generally instrumental. We discard ACFs in which the FitACF algorithm finds at least one bad-lag. Finally, we exclude ACFs with a signal-to-noise ratio below 15 dB and a spectral width greater than 600 m s^{-1}. The latter corresponds to narrow ACFs whose decorrelation time is too short to enable a correct estimation of the Doppler spectral width. These different constraints are summarised in Table 1.

After applying these selection criteria, our database contains 21 972 ACFs, 35% for 12 February 2003, 6% for 24 May 2003, 32% for 1 September 2003 and 27% for 21 December 2003. This represents 20% of the total number of echoes recorded in the Northern Hemisphere during these four days. This sample is large enough to allow statistical quantities to be properly estimated.

Figure 1 shows the spatial distribution of the radar echoes versus MLT and invariant magnetic latitude (Λ), based on the Altitude Corrected Geomagnetic (AACGM) coordinate system (Baker and Wing, 1989). We checked that this relatively small sample is representative of larger samples that were

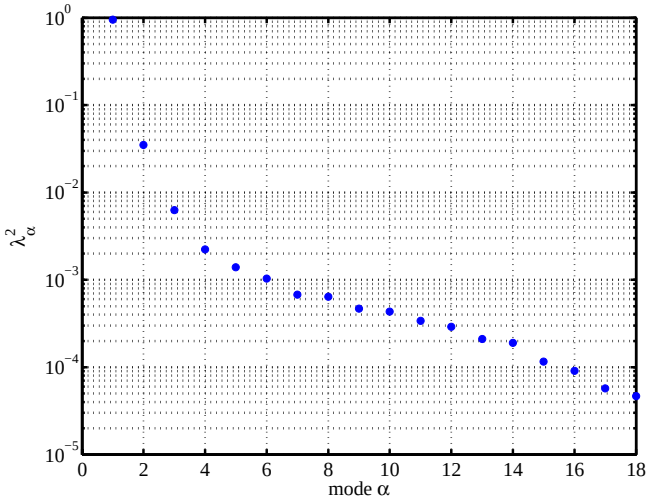


Fig. 2. Distribution of the variance of the 18 modes. Each weight is expressed as relative fraction of the total variance ($\sum_{k=1}^{n} {\lambda_k}^2$). Most of the variability of the data set is concentrated in a few low order modes.

used in earlier statistical studies (e.g. Ruohoniemi and Greenwald, 1997; Villain et al., 2002; André et al., 2002; André and Dudok de Wit, 2003). The non-uniform spatial distribution is a direct consequence of the coupling between the magnetosphere and the ionospheric backscattering regions. The high backscatter rate in the night-side sector corresponds to the auroral oval, as defined by Holzworth and Meng (1975) for a moderate geomagnetic activity ($3<K_p<4^-$).

3 Data analysis method

3.1 Principle of the method

All ACFs in our database are smooth, monotonically decaying functions, which suggests that a small set of parameters could describe their variability of shapes. Let us therefore perform a data reduction and define a set of basis functions or modes $g(\tau)$, so that each ACF can be expressed as a linear combination of few of them; τ is one of the 18 time lags at which the ACFs are estimated. We look for those modes that provide the best fit of the ACFs in a least-squares sense. The smallest set of modes is directly given by the SVD (Golub and Van Loan, 1993), which is closely related to principal component analysis (Chatfield and Collins, 1995). The SVD is a powerful data reduction technique that allows the replacement of a set of correlated variables with a smaller set of new uncorrelated variables. The same method has been used by André and Dudok de Wit (2003) with the SuperDARN data analysed by the FitACF algorithm.

3.2 Singular value decomposition of the ACFs

Let $R(t,\tau)$ be a $21\,972\times18$ rectangular matrix, in which each row is the modulus of an ACF recorded at a given time t; columns correspond to lags τ. Using the SVD, we decompose this data set into a unique ensemble of n=18 orthonormal modes

$$R(t,\tau) = \sum_{\alpha=1}^{n} \lambda_\alpha \cdot f_\alpha(t) \cdot g_\alpha(\tau) \ , \tag{1}$$

where each mode $g_\alpha(\tau)$ can be interpreted as a basis function of the ACFs and $f_\alpha(t)$ is the projection of the ACF at time t on mode $g_\alpha(\tau)$. Both functions are normalised

$$\langle f_\alpha(t),\, f_\beta(t)\rangle = \langle g_\alpha(\tau),\, g_\beta(\tau)\rangle = \begin{cases} 0 & \text{if} \quad \alpha \neq \beta \\ 1 & \text{if} \quad \alpha = \beta \ . \end{cases} \tag{2}$$

An additional coefficient $\lambda_\alpha \geq 0$ must be defined to express the weight of the α'th mode. From a geometrical point of view, the weight λ_α (or the variance ${\lambda_\alpha}^2$) expresses the dispersion of the data projected along mode α. These weights are conventionally sorted in decreasing order.

In this kind of statistical analysis, data pre-processing is important. We perform here the SVD on the ACFs without normalising them. This implies that the modes are more weighted by intense events and by ionospheric regions that have a high occurrence rate. One could also normalise each ACF by its power at lag 0, but this amplifies the contribution of weak echoes, whose signal-to-noise ratio is lower.

The distribution of the weights λ_α is the key to the interpretation of the SVD. This distribution is illustrated in Fig. 2, in which the variances have been normalised for easier reading: ${\lambda_\alpha}^2 \rightarrow \frac{{\lambda_\alpha}^2}{\sum_{k=1}^{n} {\lambda_k}^2}$. The key result is the presence of a few outstanding weights with large values, which suggests that the first few modes capture a dominant fraction of the variance of the data. If the ACFs had been random functions of event time t and lag τ, then a flat weight distribution would have been obtained. The steep weight distribution that we observe instead is a direct consequence of the redundancy of the ACFs. By projecting the data on the first mode alone, we can describe 95% of the total variance, on 2 modes 98.5%, and on 3 modes 99.2%. This means that we can reconstruct the ACFs with high accuracy using the truncated expansion:

$$\hat{R}(t,\tau) = \sum_{\alpha=1}^{m\leq n} \lambda_\alpha \cdot f_\alpha(t) \cdot g_\alpha(\tau) \ . \tag{3}$$

There is no clear criterion for determining what the number m of significant modes should be. The inflexion point in the weight distribution is often used (Chatfield and Collins, 1995). This criterion suggests that 3 modes should suffice here. Low order modes capture large-scale reproducible features of the ACFs whereas high order modes merely describe variations that occur locally as a function of t and τ.

The first four modes $g_\alpha(\tau)$ with α=[1, 2, 3, 4] are displayed in Fig. 3. Only the first three are smooth functions

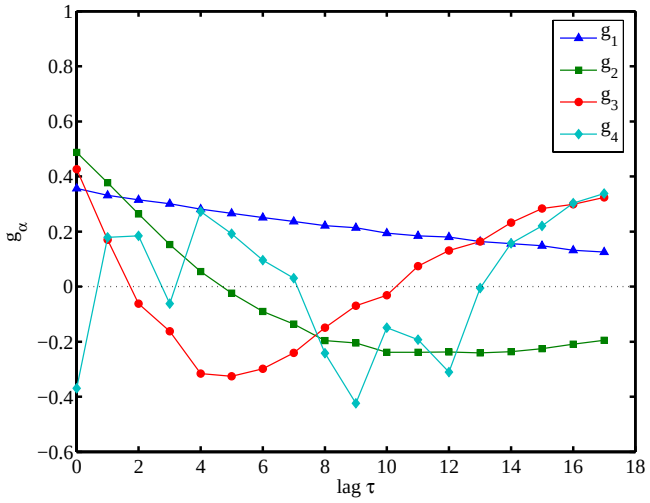


Fig. 3. The first four modes $g_\alpha(\tau)$, with α=[1, 2, 3, 4]. Only the first three modes have a regular decaying shape. Thus, these modes are used to reconstruct all ACFs.

of τ. The fourth and subsequent modes will be discarded in what follows since they are much weaker; their shape in addition is much more irregular and less reproducible. The first mode $g_1(\tau)$ is a mere average of all ACFs and looks like a decaying exponential. Modes $g_2(\tau)$ and $g_3(\tau)$ can be interpreted as higher order corrections that are needed to describe the range of possible shapes. The salient features of our data set can be reconstructed by a linear combination of these three modes only

$$\hat{R}(t,\tau) = \lambda_1 f_1(t) g_1(\tau) + \lambda_2 f_2(t) g_2(\tau) + \lambda_3 f_3(t) g_3(\tau) \ , \quad (4)$$

which suggests that each ACF can be parametrized just by three quantities $\{f_1(t), f_2(t), f_3(t)\}$. It should be stressed that this compact description of the ACFs issues from a statistical description that is less inductive than other methods in the sense that it does not impose any constraint on the actual shape of the ACFs. Figure 4 illustrates how a large variety of different ACFs can be fitted with these three modes only (letters a to f refer to different regions identified in Fig. 6).

The three modes are a direct consequence of the ionospheric activity in a general way. Further studies based on other days confirm the resilience of these modes $g_\alpha(\tau)$ and their weights λ_α. Our next task is to interpret the three projections $\{f_1(t), f_2(t), f_3(t)\}$ of each ACF in terms of physical quantities such as the spectral width. To do so, we now compare these projections to the outputs of the FitACF routine.

4 Classification and interpretation of the data reduction

The question now arises: how do these statistical parameters compare with physical parameters such as the Doppler

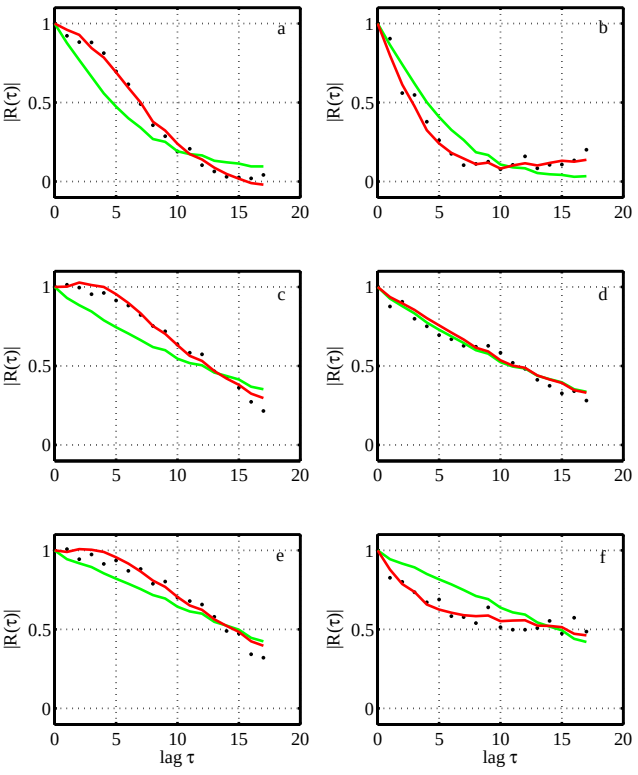


Fig. 4. Illustration of six typical ACFs reconstructed from SVD modes. Each ACF (black dots) has been reconstructed using either the first two modes (green curve) or the first three modes (red curve). Panels **(a)** to **(f)** refer to the location of these different ACFs in the parameter space displayed in Fig. 6.

spectral width and the total power? The latter is strongly correlated with the first parameter f_1, which can be roughly interpreted as an average of the ACF over all lags. The information about the shape of the ACFs (and hence the spectral width) should therefore appear in the values of f_2 and f_3 relative to f_1. The dimensionless ratio f_2/f_1 is plotted in Fig. 5 against the Doppler spectral width, as obtained by the FitACF routine (for each ACF, we have estimated the spectral width by fitting both a Lorentzian and a Gaussian).

Figure 5 convincingly shows a strong correlation between the Doppler Spectral Width (W in Hz) and the dimensionless f_2/f_1 ratio. Values beyond 35 Hz correspond to very peaked ACFs that are hard to characterise because only a few values at small lags only exceed the noise level. The spread increases with the spectral width (i.e. with the narrowness of the ACFs). Simulations carried out with analytical forms of the ACFs, incorporating a realistic noise level, confirm the difficulty in characterising the shape of peaked ACFs. Note also that the spread is exacerbated by the logarithmic scale.

Figure 5 shows two ridges, which suggests the presence of two distinct populations. A deeper investigation reveals that the upper and lower ridges actually consist of values

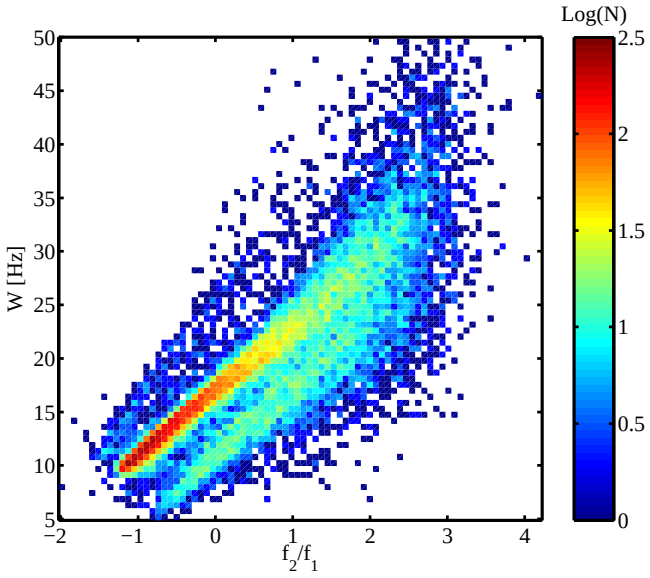


Fig. 5. Histogram of the number of radar echoes versus the f_2/f_1 ratio and the Doppler spectral width W (in Hz). The number of echoes is colour coded on a logarithmic scale. This histogram was built using 80^2 equispaced bins.

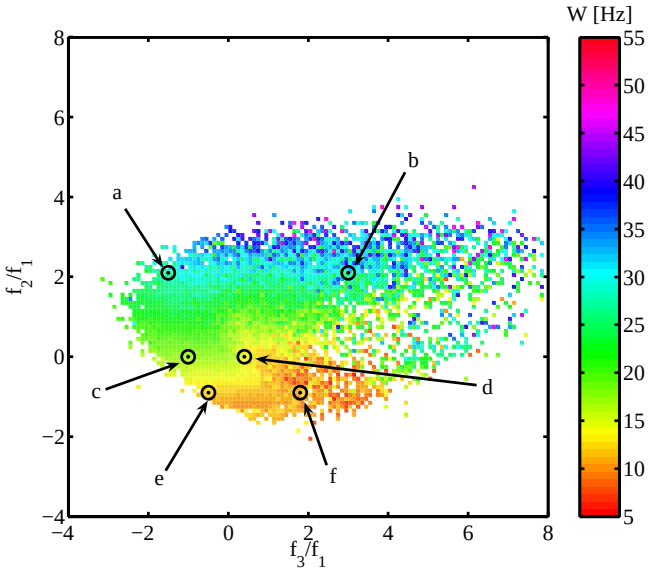


Fig. 6. Distribution of the Doppler spectral width W versus the dimensionless ratios f_2/f_1 and f_3/f_1. This distribution was built using 120^2 equispaced bins. Each letter defines a region which corresponds to a specific shape of ACF, see Fig. 4.

computed by the FitACF routine using a Gaussian and a Lorentzian model, respectively. In practice, the model that best fits the data is conserved. We see here that this choice leads to a systematic difference of about 5 Hz between the two estimates, which is far from negligible. This offset clearly illustrates how the choice of the underlying model introduces a bias that can affect the characterisation of the ACFs.

The role of f_3 is illustrated in Fig. 6, which shows the distribution of the Doppler spectral width versus the dimensionless f_2/f_1 and f_3/f_1 ratios. Although a third mode is definitely needed to fit the ACFs (see Fig. 4), there is no clear one-to-one relationship between f_3/f_1 and spectral width as with f_2/f_1; while f_2/f_1 characterizes the spectral width, f_3/f_1 appears merely as a first order correction. Each panel of the Fig. 4 shows a typical example of ACFs characterising a specific region of the $\{f_3/f_1, f_2/f_1\}$ space (see panels a to f in Fig. 6). Although the f_3/f_1 ratio is not important for our purposes, it does carry physical information and in particular points to the difference between Lorentzian- and Gaussian-shaped ACFs.

Interestingly, there is an outstanding population in Fig. 6 defined by $f_2/f_1 \in [-1, 0]$ and $f_3/f_1 \in [0, 4]$, for which the spectral width behaves differently of the rest of the sample. A closer examination reveals that these events mainly consist of backscattered echoes having a drift velocity smaller than 100 m s^{-1} and a high signal-to-noise ratio. These features are typical of E-region echoes, which can thus be identified that way. This example shows that the identification of E region echoes based upon their range is not a sufficient criterion. From now on, however, we focus on the spectral width and on the f_2/f_1 ratio, which is our proxy of the spectral width.

5 Boundary detection: case study

The Doppler spectral width of backscattered echoes has already been used in the past for identifying the polar cap boundary in the cusp region (noon sector). This boundary corresponds to the separatrix between the ionospheric projections of the open and closed geomagnetic field lines (OCB) (Baker et al., 1995; Chisham and Freeman, 2003). The boundary appears as a latitudinal gradient between narrow spectral widths below the equatorward edge of the cusp and large spectral widths within the cusp. These results suggest that the observed spectral width could be used as a criterion for separating the two classes of spectra and hence the two geophysical regions. The determination of these two classes is a delicate problem since the distribution of f_2/f_1 is broad, with no clear indication for a threshold value. This leads us to advocate a Bayesian decision criterion for discriminating the two classes of spectral widths, based on the value of the f_2/f_1 ratio. Bayesian approaches are widely used in decision theory and in pattern recognition (Duda et al., 2001). They are powerful, provided that the problem can be expressed in probabilistic terms. To illustrate our results, we focus on the 10 October 1999 event because it has already been analysed in detail by Hosokawa et al. (2003).

Let X be the space of observations. We consider a measure x, the estimate of the spectral width proxy f_2/f_1, to be a continuous random variable. Let $\omega{=}\omega_1$ and $\omega{=}\omega_2$ be the two possible classes of spectral width. The conditional probability of observing class ω_j given that x has been measured is given by:

$$P(\omega_j|x) = \frac{p(x|\omega_j)P(\omega_j)}{p(x)} \ , \tag{5}$$

where $P(\omega_j)$ is the a priori probability (or prior probability) that the measure X belongs to ω_j. Uppercase $P(\cdot)$ stands here for a probability and lowercase $p(\cdot)$ stands for a probability density.

The quantity of interest here is the posterior probability $P(\omega_j|x)$. Bayes' formula shows that by observing the value of x, we can convert the prior probability $P(\omega_j)$ to the posterior probability $P(\omega_j|x)$. The evidence $p(x)$ is the probability density function (pdf) of x, and is merely viewed as a normalisation factor that ensures $P(\omega_1|x){+}P(\omega_2|x){=}1$. The term $p(x|\omega_j)$ is the likelihood of ω_j with respect to x and expresses the conditional probability density of a measure x given the class ω_j.

To apply Bayes' theorem, let us define the likelihood of ω_1 and ω_2. The blue curve in the Fig. 7 is the probability density function of f_2/f_1, using data from October 10, 1999, restricted to the Northern Hemisphere. We apply selection criteria similar to those defined in Sect. 2.2 except that the constraint on the number of bad-lags is relaxed. This number can now be as large as 10. The advantage of this choice is the larger size of the statistical ensemble, which increases by an order of magnitude, now containing 366 168 ACFs. Without this, the size would be insufficient to detect the SWB. The bad-lags are at this stage detected by a subroutine of the FitACF algorithm, and then discarded before the estimation of the f_2/f_1 ratio. By performing statistical tests, we found that the f_2/f_1 ratio is strongly correlated with the spectral width, as long as the number of bad-lags does not exceed 10. Incidentally, by projecting the ACFs on the SVD modes, one can readily identify bad-lags, which stand out as outliers. A study is under way for using this property to automatically detect and interpolate bad-lags.

To avoid the bias introduced by multi-frequency scans between each radar, we also weight f_2/f_1 by the radar frequency.

Figure 7 (upper panel) suggests that the bimodal distribution of f_2/f_1 can be approximated by the sum of two distributions. We assume that each likelihood follows a normal law and that the two classes are equiprobable ($P(\omega_1){=}P(\omega_2){=}\frac{1}{2}$). The evidence is then given by:

$$p(x) = \sum_{j=1}^{2} p(x|\omega_j)P(\omega_j) \ . \tag{6}$$

Actually, no prior $P(\omega_j)$ is needed since we fit the $p(x|\omega_j)P(\omega_j)$ product directly to the data and not the likelihood $p(x|\omega_j)$. In making the two classes equiprobable, we minimise our prior on the discrimination.

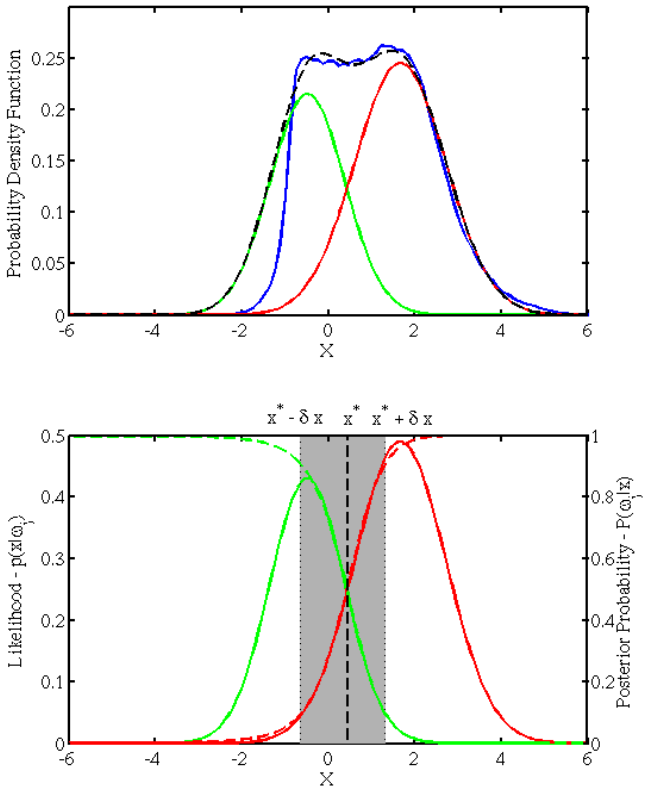


Fig. 7. Upper panel: probability density functions for October 10, 1999, over the Northern Hemisphere: probability density of f_2/f_1 (blue curve); evidence of ω_1 (green curve); evidence of ω_2 (red curve). The black dashed curve expresses the approximation of $p(x)$ Lower panel: the green and red solid and dashed lines are the respective likelihoods and posterior probabilities of ω_1 and ω_2. The superimposed grey color represents the confidence interval of the boundary defined between $P(\omega_2|x){\geq}10\%$ and $P(\omega_1|x){\geq}10\%$.

The green and red curves in the upper panel of the Fig. 7 show the approximations of the two terms of Eq. (6), respectively for ω_1 and ω_2. The two normal distributions were fitted by adjusting their mean, variance and amplitude. The resulting evidence $p(x)$ is shown by a dashed curve, which fits the observed pdf remarkably well. Note the truncated shape of $p(x)$ for $x{<}{-}1$, which can be explained by the lower limit applied on the spectral width value to discard ground echoes. A third class could be added to fit $p(x)$ more accurately for $x{<}{-}1$ but this does not bring a significant improvement to the SWB identification.

The lower panel in Fig. 7 shows the likelihoods $p(x|\omega_j)$ of classes ω_1 and ω_2 (green and red). The dashed curve with the same colour code express the posterior probability $P(\omega_j|x)$. The equality of the two posteriors $P(\omega_1|x^*){=}P(\omega_2|x^*){=}\frac{1}{2}$ sets the limit (or decision boundary) between both classes at

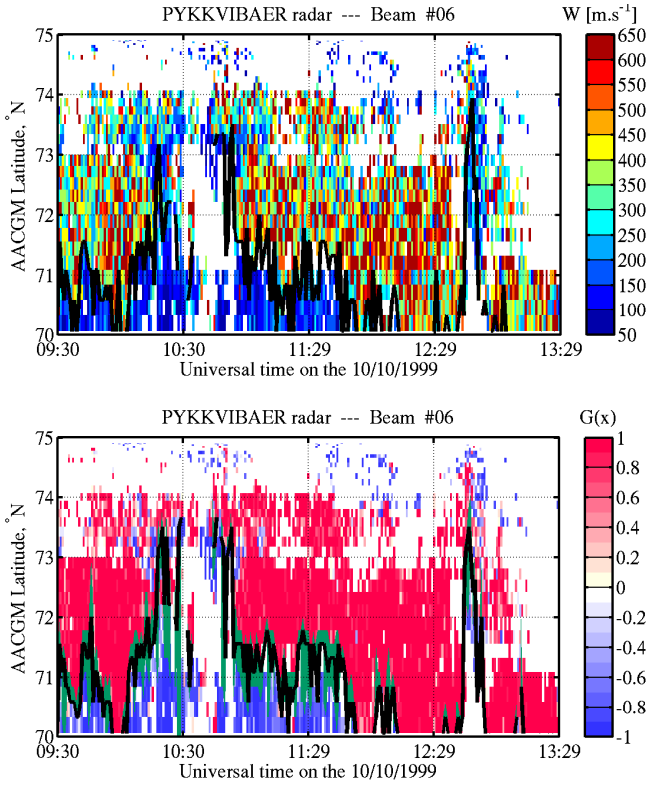


Fig. 8. SWB determination between 09h30-13h30 for the October 10, 1999 event, as observed by Þykkvibaer beam #06, using two methods. The upper panel shows the SWB (thick black line) obtained by applying a direct threshold to the spectral width ($W^*=200\,m\,s^{-1}$), as determined from the FitACF routine. The colour code refers to the spectral width. The lower panel shows the SWB identified from the discriminant function $G(x)$, with its confidence interval (green).

x^*=0.52. This will be the location of the boundary since:

- if $x<x^*$, $P(\omega_1|x)>P(\omega_2|x)$ and ω_1 is the prevailing class;
- if $x>x^*$, $P(\omega_1|x)<P(\omega_2|x)$ and ω_2 is the prevailing class.

Note that the boundary is estimated by maximising the uncertainty on the origin of each measure, where the two classes have a maximum overlap. One of the advantages of this approach is the possibility to define a confidence interval for the location of the boundary. This interval is defined by the range of values for which $P(\omega_2|x)\geq 10\%$ and $P(\omega_1|x)\geq 10\%$. The resulting range is shown by the grey interval in the lower panel of Fig. 7.

For the sake of convenience, we define a single discriminant function:

$$G(x) \equiv P(\omega_2|x) - P(\omega_1|x)\ , \qquad (7)$$

so that $G(x)$ is negative (resp. positive) when class ω_1 (ω_2) dominates. Figure 8 shows the spectral width (upper panel) and $G(x)$ (lower panel) as a function of invariant magnetic latitude and time between 09:30 and 13:30 UT of the 10 October 1999 event, for the beam #06 of the Iceland east radar (Þykkvibaer). This event is exactly the same as in Hosokawa et al. (2003). During this period, the radar field-of-view covers the mean latitude of the central plasma sheet and the low latitude boundary layer up to the polar cap across the cusp around the noon sector. In both panels, two regions stand out: a poleward one with high values of the spectral width and an equatorward one with small spectral widths.

In the upper panel, we employ the same strategy as in Hosokawa et al. (2003) to identify the SWB and set the boundary at a spectral width of $W^*=200\,m\,s^{-1}$. For each time, we search poleward for the first pair of successive gates that exceed the threshold value W^*. The first of these two gates gives the latitudinal position of the SWB, which is indicated in Fig. 8 by a thick black line. In the lower panel, we use our Bayesian criterion and set the boundary at $G(x)=0$. We associate $G(x)>0$ with large spectral widths and $G(x)<0$ with narrow spectral widths. The Bayesian boundary is identified with the same strategy that was used for locate the SWB. The green color indicates the latitudinal extent of the confidence interval as defined above.

This comparison shows that the two methods are in excellent agreement. The main advantages of our approach are that:

- the Bayesian boundary suffers less from latitudinal fluctuations than the SWB since the Bayesian criterion maximises the separation between the two classes of spectral width, and in this sense is better suited for tracking the boundary with sparse data. More importantly, our Bayesian approach bypasses the subjective task of defining a universal threshold, as already mentioned in Baker et al. (1997) and Chisham and Freeman (2003).
- the confidence interval is an important by-product, as it allows the validation of the results. This interval can be defined because of the gradual transition between regions of small and large spectral width.
- we detect the boundary using a proxy for the spectral width, which does not require the ACF to be modelled explicitly by a Lorentzian or a Gaussian. This choice between two models introduces an additional uncertainty.

We extended this comparison to other radars and always found an excellent agreement between the two approaches.

From a geophysical point of view, it is difficult to determine which boundary is the best since the association between the SWB and magnetospheric regions requires detailed comparisons with satellite data. Moreover, various effects (electric field, turbulence, homogeneity of the velocity

distribution, propagation ...) can affect the location of the SWB (Baker et al., 1995; Moen et al., 2001; André et al., 2000; Vallières et al., 2004). One of these is the orientation of the radar field-of-view. Off-meridonal beams can observe large spectral widths induced by a reversal of the flow inside the convection cell, which is unrelated to the OCB (Lester et al., 2001; Villain et al., 2002; Chisham et al., 2005b). Beam 6 displayed in Fig. 8 is an off-meridional beam. An investigation of the drift velocity clearly reveals that between 11:00 and 12:30 UT the SWB and the Bayesian boundary are co-located with the flow reversal boundary. Before 11:00 UT, the SWB and Bayesian boundaries are linked to the equartorward boundary of the cusp. This is a typical example where two distinct regions (OCB and flow reversal boundary) can lead to the same spectral width or f_2/f_1 ratio.

Several improvements are under way. Although the f_3/f_1 contributes weakly to the identification of the spectral width, it nevertheless provides an additional parameter that could bring additional leverage. This parameter is significant for narrow ACFs only (see Sect. 4) and so it could help better constrain the description of regions with large spectral widths.

6 Conclusion

In this study, we present a novel method for monitoring the ionospheric projection of magnetospheric regions from the SuperDARN HF coherent scatter radar network. Using a statistical approach based on the SVD, we show that the ACFs of the observed backscattered echoes can be reconstructed from a linear combination of three modes only; two of them give a good proxy for the spectral width. The application of a Bayesian criterion provides an objective method for identifying the boundary between magnetospheric regions, based on the value of the spectral width. We illustrate this method with the polar cap boundary in the cusp region.

Our approach is the first step towards an empirical model to identify boundaries in near real-time. For that purpose, it is necessary to stay closer to the raw autocorrelation data and bypass physical models, which may give rise to artefacts, e.g. Ponomarenko and Waters (2006). The problem of manually selecting thresholds is overcome by using a Bayesian criterion as a self-calibrating classifier, which is well suited for real-time analysis.

The method is still open to improvements. The description of the backscattered echoes remains incomplete without including the phase of the ACFs. This phase, however, depends on the location of the radars, and so its investigation is likely to be more complex and of little relevance for an automatic boundary detection scheme. To evaluate the accuracy of the method, we also plan to study the effect of the noise and bad-lags. Eliminating bad-lags is an important issue. Because our method starts by projecting the ACFs on three modes, such outliers can readily be identified. The determination of a detection scheme based on this property is in progress. For a validation purposes, the comparison of the identified boundaries with other diagnostics is foreseen.

Acknowledgements. This work was supported by the Institut National des Sciences de l'Univers (INSU) trough its Programme National Soleil-Terre (PNST).

Topical Editor M. Pinnock thanks G. Chisham, K. Hosokawa, and M. Lester for their help in evaluating this paper.

References

André, R. and Dudok de Wit, T.: Identification of the ionospheric footprint of magnetospheric boundaries using SuperDARN coherent HF radars, Planet. Space Sci., 51, 813–820, 2003.

André, R., Pinnock, M., and Rodger, A. S.: On the factor conditioning the Doppler spectral width determined from SuperDARN HF radars, Int. J. Geomagnetism Aeronomy, 2, 77–86, 2000.

André, R., Pinnock, M., Villain, J.-P., and Hanuise, C.: Influence of magnetospheric processes on winter radar spectra characteristics, Ann. Geophys., 20, 1783–1793, 2002, http://www.ann-geophys.net/20/1783/2002/.

Baker, K. B. and Wing, S.: A new magnetic coordinate system for conjugate studies at high latitudes, J. Geophys. Res., 94, 9139–9143, 1989.

Baker, K. B., Dudeney, J. R., Greenwald, R. A., Pinnock, M., Newell, P. T., Rodger, A. S., Mattin, N., and Meng, C.-I.: HF radar signatures of the cusp and low-latitude boundary layer, J. Geophys. Res., 100, 7671–7695, 1995.

Baker, K. B., Rodger, A. S., and Lu, G.: HF-radar observations of the dayside magnetic merging rate : A Geospace Environment Modeling boundary layer, J. Geophys. Res., 102, 9603–9617, 1997.

Chatfield, K. and Collins, A. J.: introduction to Multivariate Statistics., Chapman & Hall, London, 1995.

Chisham, G. and Freeman, M. P.: A technique for accurately determining the cusp-region polar cap boundary using SuperDARN HF radar measurements, Ann. Geophys., 21, 983–996, 2003, http://www.ann-geophys.net/21/983/2003/.

Chisham, G. and Freeman, M. P.: An investigation of latitudinal transitions in the SuperDARN Doppler spectral width parameter at different magnetic local time, Ann. Geophys., 22, 1187–1202, 2004, http://www.ann-geophys.net/22/1187/2004/.

Chisham, G., Freeman, M. P., and Sotirelis, T.: A statistical comparison of SuperDARN spectral width boundaries and DMSP particle precipitation boundaries in the nightside ionosphere, Geophys. Res. Lett., 31, L02 804, doi:10.1029/2003GL019074, 2004.

Chisham, G., Freeman, M. P., Lam, M. M., Abel, G. A., Sotirelis, T., Greenwald, R. A., and Lester, M.: A statistical comparison of SuperDARN spectral width boundaries and DMSP particle precipitation boundaries in the afternoon sector ionosphere, Ann. Geophys., 23, 3645–3654, 2005a.

Chisham, G., Freeman, M. P., Sotirelis, T., and Greenwald, R. A.: The accuracy of using the spectral width boundary measured in off-meridional SuperDARN HF radar beams as a proxy for the open-closed field line boundary, Ann. Geophys., 23, 2599–2604, 2005b.

Chisham, G., Freeman, M. P., Sotirelis, T., Greenwald, R. A., Lester, M., and Villain, J.-P.: A statistical comparison of SuperDARN spectral width boundaries and DMSP particle precipitation boundaries in the morning sector ionosphere, Ann. Geophys., 23, 733–743, 2005c.

Duda, R., Hart, P., and Stork, D.: Pattern Classification, A Wiley-Interscience Publication, 2 edn., 2001.

Dudeney, J. R., Rodger, A. S., Freeman, M. P., Picket, J., Scudder, J., Sofko, G., and Lester, M.: The nightside ionospheric response to IMF By changes, Geophys. Res. Lett., 25, 2601–2604, 1998.

Golub, G. and Van Loan, C.: Matrix Computations, The John Hopkins University Press, Baltimore, 4 edn., 1993.

Greenwald, R. A., Baker, K. B., and Hanuise, C.: An HF phased-array radar for studying small-scale structure in the high-latitude ionosphere, Radio Sci., 20, 63–79, 1985.

Greenwald, R. A., Baker, K. B., Dudeney, J. R., Pinnock, M., Jones, T. B., Thomas, E. C., Villain, J.-P., Cerisier, J.-C., Senior, C., Hanuise, C., Hunsucker, R. D., Sofko, G., Koehler, J., Nielsen, E., Pellinen, R., Walker, A. D. M., Sato, N., and Yamagishi, H.: DARN/SUPERDARN : A Global View of the Dynamic of High-Latitude Convection, Space Sci. Rev., 71, 761–796, 1995.

Hanuise, C., Villain, J.-P., Gresillon, D., Cabrit, B., and Baker, K. B.: Interpretation of HF radar ionopsheric Doppler spectra by collective wave scattering theory, Ann. Geophys., 11, 29–39, 1993,
http://www.ann-geophys.net/11/29/1993/.

Holzworth, R. and Meng, C.-I.: Mathematical representation of the auroral oval, Geophys. Res. Lett., 2, 377–380, 1975.

Hosokawa, K., Woodfield, E. E., Lester, M., Milan, S. E., Sato, N., Yukimatu, A. S., and Iyemori, T.: Statistical characteristics of Doppler spectral width as observed by the conjugate SuperDARN radars, Ann. Geophys., 20, 1213–1223, 2002,
http://www.ann-geophys.net/20/1213/2002/.

Hosokawa, K., Woodfield, E. E., Lester, M., Milan, S. E., Sato, N., Yukimatu, A. S., and Iyemori, T.: Interhemispheric comparison of spectral width boundary as observed by SuperDARN radars, Ann. Geophys., 21, 1553–1565, 2003,
http://www.ann-geophys.net/21/1553/2003/.

Lester, M., Milan, S. E., Besser, V., and Simth, R.: A case study of HF radar spectra and 630 nm auroral emission in the pre-midnight sector, Ann. Geophys., 19, 327–339, 2001,
http://www.ann-geophys.net/19/327/2001/.

Moen, J., Carlson, H. C., Milan, S. E., Shumilov, N., Lybekk, K., Sandholt, P. E., and Lester, M.: On the collocation between dayside auroral actvity and coherent HF radar backscatter, Ann. Geophys., 18, 1531–1549, 2001,
http://www.ann-geophys.net/18/1531/2001/.

Parkinson, M. L., Dyson, P. L., Pinnock, M., Devlin, J. C., Hairston, M. R., Yizengaw, E., and Wilkinson, P. J.: Signatures of the mid-night open-closed magnetic field line boundary during balanced dayside and nightside reconnection, Ann. Geophys., 20, 1616–1630, 2002,
http://www.ann-geophys.net/20/1616/2002/.

Pinnock, M., Rodger, A. S., Baker, K. B., Lu, G., and Hairston, M.: Conjugate observations of the day-side reconnection electric field: A GEM boundary layer campaign, Ann. Geophys., 17, 443–454, 1999,
http://www.ann-geophys.net/17/443/1999/.

Ponomarenko, P. V. and Waters, C. L.: The role of Pc1-2 waves activity in spectral broadening of SuperDARN echoes from high latitudes, Geophys. Res. Lett., 30, L1122, doi:10.1029/2002/GL016333, 2003.

Ponomarenko, P. V. and Waters, C. L.: Spectral width of SuperDARN echoes : measurement, use and physical interpretation, Ann. Geophys., 24, 115–128, 2006,
http://www.ann-geophys.net/24/115/2006/.

Rodger, A. S.: Ground-based imaging of magnetospheric boundaries, Adv. Space Res., 25, 1461–1470, 2000.

Rodger, A. S., Mende, S. B., Rosenberg, T. J., and Baker, K. B.: Simultaneous optical and HF radar observations of the ionospheric cusp, Geophys. Res. Lett., 22, 2045–2048, 1995.

Ruohoniemi, J.-M. and Greenwald, R. A.: Rates of scattering ocurrence in routine HF radar observations during solar cycle maximum, J. Geophys. Res., 32, 1051–1070, 1997.

Vallières, X., Villain, J.-P., Hanuise, C., and André, R.: Ionospheric propagation effects on spectral widths measured by SuperDARN HF radars, Ann. Geophys., 22, 2023–2031, 2004,
http://www.ann-geophys.net/22/2023/2004/.

Villain, J.-P., Greenwald, R. A., Baker, K. B., and Ruohoniemi, J.-M.: HF radar observations of E region plasma irregularities produced by oblique electron streaming, J. Geophys. Res., 92, 12 327–12 342, 1987.

Villain, J.-P., André, R., Hanuise, C., and Gresillon, D.: Observation of the high latitude ionosphere by HF radars: interpretation in terms of collective wave scattering and characterization of turbulence, J. Atmos. Terr. Phys., 58, 943–958, 1996.

Villain, J.-P., André, R., Pinnock, M., Greenwald, R. A., and Hanuise, C.: A Statistical study of the Doppler spectral width of high-latitude ionospheric F-region echeos recorded with SuperDARN coherent HF radars, Ann. Geophys., 20, 1769–1781, 2002,
http://www.ann-geophys.net/20/1769/2002/.

Woodfield, E. E., Davies, J. A., Eglitis, P., and Lester, M.: A case study of HF radar spectral width in the post midnight magnetic local time sector and its relationship to the polar cap boundary, Ann. Geophys., 20, 501–509, 2002a.

Woodfield, E. E., Davies, J. A., Lester, M., Yeoman, T. K., Eglitis, P., and Lockwood, M.: Nightside studies of coherent HF Radar spectral width behaviour, Ann. Geophys., 20, 1399–1413, 2002b.