Correspondence to:
B. Bezděková,
baja@etranslator.biz

Using Principal Component Analysis to Characterize the Variability of VLF Wave Intensities Measured by a Low-Altitude Spacecraft and Caused by Interplanetary Shocks

B. Bezděková[1] [ID], **F. Němec[1]** [ID], **M. Parrot[2]** [ID], **O. Kruparova[3,4,5]** [ID], and **V. Krupar[3,4]** [ID]

[1]Faculty of Mathematics and Physics, Charles University, Prague, Czech Republic, [2]LPC2E/CNRS, Orléans, France, [3]Goddard Planetary Heliophysics Institute, University of Maryland, Baltimore, MD, USA, [4]Heliophysics Division, Heliospheric Physics Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD, USA, [5]Department of Space Physics, Institute of Atmospheric Physics, Czech Academy of Sciences, Prague, Czech Republic

Abstract Very low frequency wave intensity measurements provided by the French low-altitude DEMETER spacecraft are studied using the principal component analysis (PCA). We focus on both the physical interpretation of the first two principal components and their application to real physical problems. Variations of the first principal component (PC1) coefficients due to the geomagnetic activity and seasonal/longitudinal changes are studied. It is shown that their distribution corresponds to the wave intensity dependences obtained in previous studies. Moreover, the variations of PC1 coefficients around interplanetary shock arrivals are analyzed. The study is performed for fast forward (FF), fast reverse, slow forward, and slow reverse shocks separately. It shows that the most significant effect on the wave intensity is displayed in the FF case. Furthermore, it turns out that the wave intensity variations depend on the wave intensity detected before the shock arrival. Finally, the shock strength and interplanetary magnetic field orientation are also important. The performed analysis shows that PCA can be successfully applied to characterize large data sets of spacecraft measurements by limited sets of numbers—principal component coefficients (typically first one or two are enough), which still maintain a sufficient amount of information.

1. Introduction

Variations of the wave intensity in the very low frequency (VLF) range measured in the Earth's magnetosphere can be caused by various factors and are permanently under a vivid discussion, (e.g., Colman & Starks, 2013; Marshall et al., 2008; Němec et al., 2009; Park & Miller, 1979; Parrot, 1990; Píša et al., 2013; Šulić et al., 2016). The wave intensity can evolve due to, for example, background plasma density fluctuations (Yue et al., 2020), solar wind dynamic pressure variations (Liu et al., 2018), or along with substorm occurrence (Tsurutani et al., 2018, and references therein). A case when a substorm generated plasmaspheric hiss on the nightside was reported by Su et al. (2018).

Both the generation and properties of whistler mode electromagnetic waves are ultimately affected by the parameters of the solar wind. Among these, effects related to interplanetary (IP) shock occurrence have a significant and step-like influence. It is known that the characteristics of the shock arrivals affect both substorm (Heppner, 1955) and storm (Gonzalez et al., 1994) occurrence, auroral brightening and motion (Meurant et al., 2004; Su et al., 2011; Zhou & Tsurutani, 1999), and energetic particle dynamics (Blake et al., 1992; Hudson et al., 1997; Zong et al., 2009).

In few recent studies, a possible direct connection between the whistler mode waves and IP shocks was investigated. Su et al. (2015) reported a case study of the disappearance of plasmaspheric hiss (at frequencies between about 50 Hz and 1 kHz) after detecting an IP shock and a strong substorm (maximum AE ≈ 1400 nT) for about five hours. Similarly, an event analyzed by Liu et al. (2017) exhibited an instant disappearance of hiss waves after the arrival of an IP shock in the frequency range from 0.1 to 1.0 kHz, but also the appearance of weak waves above 1 kHz. This effect was simultaneously measured by several satellites along with a prolonged substorm activity (maximum AE ≈ 1100 nT). About one hour after the shock detection, recovery of plasmaspheric hiss waves was observed, and the increase of the chorus wave intensity at frequencies above 1 kHz was detected. During this recovery period, hiss wave intensities were about 2–5

times larger than before the shock. On the contrary, Fu et al. (2012) investigated chorus intensification in response to the IP shock arrival. They showed that the electron temperature anisotropy is significantly enhanced due to the magnetospheric compression, resulting in up to about 50% increase of the growth rate.

Zhima et al. (2014) investigated the evolution of VLF wave intensity measured by DEMETER (Detection of Electro-Magnetic Emissions Transmitted from Earthquake Regions) during geomagnetic storms driven by coronal mass ejections (CME). The analysis was performed as both case and statistical studies. It is shown that the wave intensity increases immediately after the compression of geomagnetic field by CME and a strong wave excitation happens predominantly at the end of the main phase and beginning of the recovery phase. Moreover, the wave intensity response was found to be frequency dependent. While waves below 3 kHz intensified substantially during the whole storm duration, higher frequency waves were enhanced rather during the later main and early recovery phases. If the wave frequency was too large (above 15 kHz), their intensity was amplified only in a very limited time slot between the end of the main phase and the beginning of the recovery phase.

Since the arrangement of instruments allowing to directly detect a magnetospheric wave response to IP shock arrival is not common, not many studies are dealing with the problem statistically. Yue et al. (2017) presented a statistical analysis of the effect of 86 fast forward IP shocks detected between 2010 and 2016 on the whistler mode waves. Using observations of the Van Allen Probes and THEMIS (Time History of Events and Macroscale Interactions during Substorms) spacecraft, they showed that chorus and plasmaspheric hiss waves respond to the IP shock occurrence differently. While chorus waves are usually intensified after a shock arrival, plasmaspheric hiss can be either attenuated or enhanced after a shock. It is stressed that also solar wind parameter variations, mainly solar wind dynamic pressure, can cause plasmaspheric hiss disappearance on the dayside.

Principal component analysis (PCA) is a statistical tool which is used to reduce a dimensionality of a given data set (see below). Due to its apparent practicality it has been already used in several geophysical applications. Holappa et al. (2014) showed that the first two principal components are enough to describe an overall geomagnetic activity. While the first principal component describes global average geomagnetic activity, the second principal component corresponds to the deviations from the global average caused by high-speed streams (HSS). The obtained principal components allowed to infer CME and HSS occurrence and also the long-term evolution of the interplanetary magnetic field (IMF) strength and the solar wind speed. Milan et al. (2015) also focused on searching for different aspects of geomagnetic activity by PCA. They applied PCA to the Birkenland current data obtained by the AMPERE experiment and studied patterns of different geomagnetic activity aspects. PCA was also used for detecting ULF geomagnetic effects associated with earthquakes (e.g., Gotoh et al., 2002; Hattori et al., 2006; Hayakawa et al., 2007; Li et al., 2011; Serita et al., 2005). Last but not least, PCA has been adapted for works dealing with ionospheric problems (e.g., Kim et al., 2012; Lin, 2013; Natali & Meza, 2010).

We aim to use the PCA method to systematically analyze variations of the VLF wave intensity in the Earth's inner magnetosphere as a function of the geomagnetic activity, season, longitude, and IP shock arrival. Low-altitude spacecraft measurements are particularly useful for this analysis as their relatively short orbital periods allow to sample a vast range of space during a comparatively short time. Data of the French microsatellite DEMETER were used in the present study. More details about this spacecraft are mentioned in Section 2. The PCA method used for data analysis is described in Section 3. The obtained results are presented in Section 4, discussed in Section 5, and summarized in Section 6.

2. Data Sets

Data used in the present analysis were obtained by the French microsatellite DEMETER. It operated between June 2004 and December 2010. Its original altitude was about 710 km, but since December 2005 it was decreased to about 660 km. The satellite was operating on an almost Sun-synchronous circular orbit which led to basically only two possible local time (LT) measurement intervals—either about 10:30 LT (daytime/dayside measurements) or 22:30 LT (nighttime/nightside measurements). The spacecraft performed about 14 orbits per day.

Survey mode wave measurements used in the present paper were continuously performed at geomagnetic latitudes lower than 65°. One electric and one magnetic field components have been obtained. Unfortunately, due to substantial interferences in the magnetic field data in the VLF range, only the electric field measurements are used. The electric field measurements were performed by the Instrument Champ Electrique (ICE) instrument (Berthelier et al., 2006). The VLF data consist from on board calculated power spectral densities of electric field fluctuations in the frequency range up to 20 kHz divided into 1,024 linearly spaced frequency channels with a time resolution of about 2 s and a frequency resolution of about 19.53 Hz.

The dayside and nightside measurements were analyzed separately in this study. Altogether, 29,335 daytime and 28,239 nighttime half orbits with VLF data available (corresponding to the entire DEMETER mission) were analyzed. However, in some cases, the measurements were not performed for the full duration of the spacecraft half orbit. Such spectrograms were excluded from the analysis (see below).

In the second part of the present study, the effect of IP shocks on the DEMETER wave intensity measurements was investigated. IP shocks which occurred during the DEMETER mission were identified in the Wind spacecraft data (close to the L1 point) by an automatic procedure and subsequent manual verification as described by Krupařová et al. (2013) and Bezděková et al. (2015). Altogether, 225 IP shocks were detected between June 2004 and December 2010. They were compartmentalized according to the classification described in Burlaga (1971). Shocks moving in the solar wind frame away from the Sun are forward shocks, whereas shocks moving in the solar wind frame toward the Sun are called reverse shocks. While both the density and magnetic field magnitude increase behind fast shocks, the density increases and magnetic field magnitude decreases behind slow shocks. Out of the given data set, 87 IP shocks were classified as fast forward (FF), 31 as fast reverse (FR), 59 as slow forward (SF), and 48 as slow reverse (SR) shocks. Effects of these shock types on the wave intensity measured by DEMETER on the dayside were analyzed separately. In order to capture the full time range of effects related to the IP shock arrivals, wave intensity measured by the spacecraft between 5 h before and 24 h after the shock arrival was investigated for every single shock and sorted according to its type. Statistics of the wave measurements obtained for given shock type was then performed. Remark that due to a comparatively large distance between the shock detection location and the magnetosphere, the times of the shock arrivals were recalculated taking into account the respective shock propagation times.

3. Method Description

PCA is a statistical tool for reducing a dimensionality of large data sets while conserving as much information as possible. The output of this routine is a new basis of components. These so-called principal components are linear functions of original vectors, they are orthogonal, and defined in such a way that they maximize the respective variances of the original projections. Moreover, they are sorted according to their variances, which correspond to the amount of the information from the original data set carried by the new variables. A comprehensive review of this technique and possible applications was given by Jolliffe and Cadima (2016).

In our case, the problem of finding principal components was slightly more complicated since the original data set consists of frequency-time spectrograms measured by DEMETER. However, considering that on the time scales of a single half orbit most of the observed intensity variations are spatial, that is, due to the spacecraft movement, the time dependence was recalculated to geomagnetic latitude dependence and the analysis was in fact performed using frequency-latitude spectrograms. Moreover, each spectrogram exhibits a latitudinal symmetry, as the spacecraft regularly crossed the equator. The data obtained within each DEMETER half orbit were thus divided into two frequency-geomagnetic latitude spectrograms around the geomagnetic equator (northern and southern parts) and the dependence on the geomagnetic latitude was analyzed. Consequently, two spectrograms were obtained from each of the analyzed half orbits. Note that in all plots shown hereinafter we choose the southern, that is, negative, latitudes to be plotted on the abscissa. Due to the generally different intensity variations on the dayside and nightside, the dayside and nightside spectrograms were investigated separately. Only spectrograms covering the entire duration of a given half (north or south) of spacecraft half orbits were considered. Altogether, the PCA of 56,940 dayside quarter orbits and 56,584 nightside quarter orbits was performed.
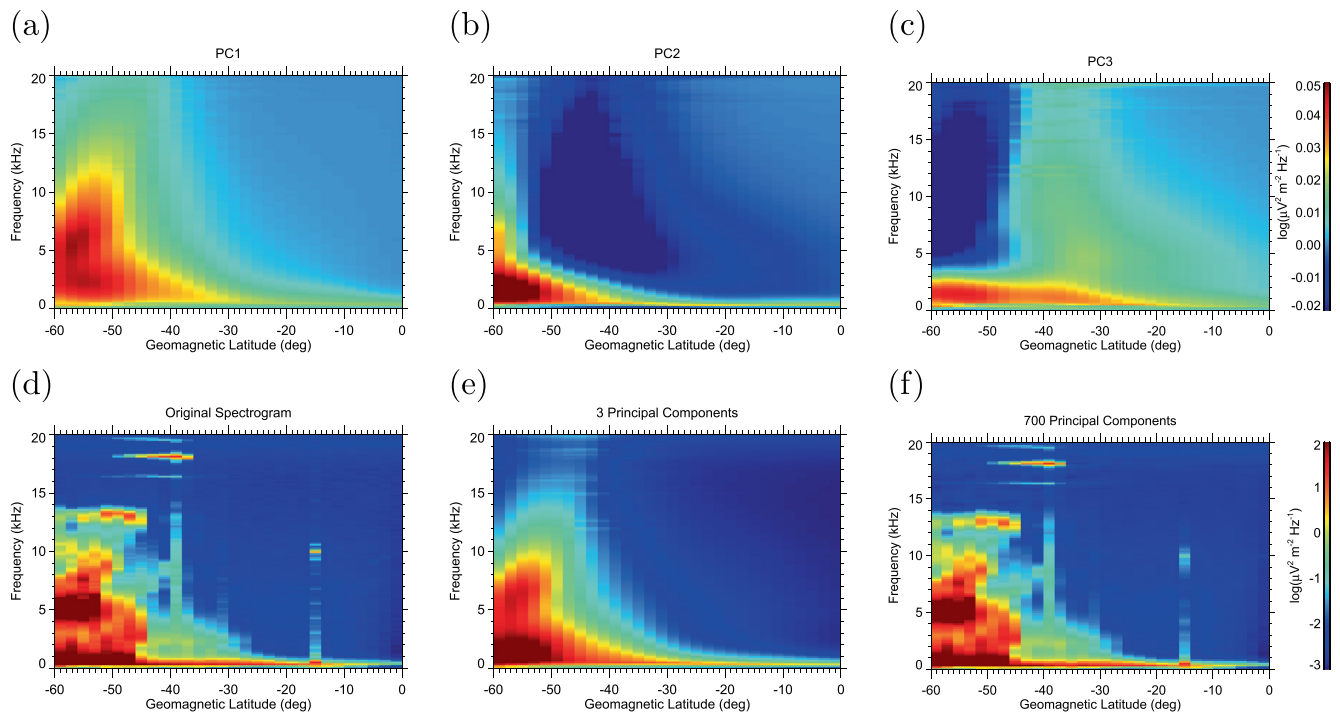
Figure 1. Application of the principal component analysis on the DEMETER data set. (a)–(c) The first three principal components (basis vectors) corresponding to the dayside DEMETER half orbit data set. The analysis was performed for the frequency-geomagnetic latitude plots (two for each DEMETER half orbit—northern, southern parts) and the results were then plotted with negative geomagnetic latitude signs (corresponding to the southern parts). (d)–(f) Example of using the calculated principal components for a backward reconstruction of a randomly chosen frequency-latitude spectrogram. (d) Original frequency-latitude spectrogram measured by the DEMETER spacecraft which was averaged as described in the text. (e) Reconstruction of the spectrogram depicted in (d) obtained using the first three principal components. (f) Reconstruction of the spectrogram depicted in (d) performed by using the first 700 principal components.

The computation of the principal components of such a large data set represents a time consuming task. To spare some computational time, the frequency resolution of the analyzed spectrograms was partially reduced and the spectrograms were hence averaged in the frequency domain. Finally, frequency-latitude spectrograms with the frequency resolution 156.25 Hz and latitudinal resolution 2° spanning the frequency range 0–20 kHz and latitudinal range 0°−60° were used as the initial data set. After the PCA calculation, a new set of variables—principal components—was obtained. Altogether, 3,840 principal components (i.e., frequency-geomagnetic latitude spectrograms forming a new principal component basis) were calculated for the dayside and nightside spectrogram data sets. The first three principal components obtained for the dayside spectrograms are shown in Figures 1a–1c. They carry as much as about 60% of the original information.

Having obtained the PCA basis, every single daytime frequency-geomagnetic latitude spectrogram measured by DEMETER can be expressed as a linear combination of the principal components. The principal components can be thus regarded as base vectors of the space of the analyzed spectrograms. The amount of information carried by individual principal components can be demonstrated by the backward reconstruction of a randomly chosen frequency-geomagnetic latitude spectrogram as shown in Figures 1d–1f. Figure 1d shows the original frequency-geomagnetic latitude spectrogram measured by DEMETER. The reconstruction of this spectrogram using the first three principal components (shown in Figures 1a–1c) is plotted in Figure 1e, while the reconstruction using the first 700 principal components is shown in Figure 1f. It is evident that already the first three principal components are enough to capture basic characteristics of the original spectrogram. Naturally, more components contain more information and the usage of a higher number of principal components thus leads to a more precise spectrogram reconstruction. Figure 1f demonstrates that already the first 700 principal components (out of 3,840) are enough to precisely capture almost all specific features of the original spectrogram.

The main advantage of the PCA is the robust reduction of the set dimensionality. Since the majority of the information is included in the first few components, only a handful of them are needed to characterize the original data with a sufficient accuracy. In our case, this means that a given original DEMETER spectrogram can be characterized by a set of coefficients calculated from its decomposition into the principal components. Moreover, the backward reconstruction shown in Figure 1e demonstrates that to characterize main aspects of a particular spectrogram, it is enough to investigate only the first two or three components.

A crucial task when applying PCA to problems in Physics is to interpret the physical meaning of the obtained principal components. It is generally assumed that several first principal components are enough to maintain most of the information included in the original data set. To get a better idea about the meaning of the first two principal components, Figure 2 was obtained. Figure 2a shows a scatter plot of the first two principal component coefficients (PC1, PC2) calculated for the dayside measurements. Since they describe more than 50% of the total variances, they readily characterize basic features of the overall wave intensity measurements performed in a given quarter orbit. Note that higher principal components eventually describe more specific features in the spectrograms, such as intensity increases due to ground-based transmitters.

To better illustrate a possible range of intensity variations given by the first two several principal components, Figures 2b–2e show spectrograms obtained by multiplying the first two principal components (shown in Figures 1a and 1b) by coefficients (−20, 15), (50, 30), (−20, 15), and (50, 30), respectively. The obtained spectrograms correspond to extreme intensity profiles measured by DEMETER on the dayside. The effect of coefficient values is apparent. While the first principal component determines the overall wave intensity, the second principal component gives the basic intensity profile. Remark that a scatter of the obtained PC coefficients in the top part of Figure 2a, where PC2 coefficients are positive, is much lower than in the bottom part of the plot. The variance is thus larger for spectrograms with large positive PC1 coefficients and small negative PC2 coefficients. Moreover, the intensity profiles obtained for the positive PC2 qualitatively correspond to mean intensity profile of the dayside DEMETER intensity measurements (not shown), while intensity profiles given for the negative PC2 are different. Note that the same analysis applied on the nighttime measurements yields similar results, just with a different principal component basis. Note also that there is a sign ambiguity depending on a choice of the PCA basis sign. If the opposite sign is chosen when constructing the basis, all the respective principal component coefficients would have a different sign as well. In the further analysis, only PC1 coefficients are studied. Although a similar analysis could be performed also for higher PC coefficients, the exact interpretation of their meaning is comparatively more difficult.

The fact that PC1 corresponds to an overall wave intensity is further demonstrated in Figure 3. It shows a dependence of the mean wave intensities in individual analyzed frequency-latitude spectrograms on the respective PC1 coefficients. It is clearly seen that the spectrograms with larger PC1 coefficients tend to have larger average wave intensities. Pearson correlation coefficient of these two quantities is about 0.94. Note that the correlation is not perfect, as the PC1 coefficients actually correspond to the weighted average wave intensities, with the weights determined in such a way that they emphasize the intervals where the intensity variability is most significant.

4. Results

4.1. Geomagnetic Activity and Seasonal Variations of the First Principal Component Coefficient

Having determined that the PC1 coefficient gives the information about the overall wave intensity, it can be used to characterize the intensity measured during each individual quarter orbit. The northern and southern quarter orbits are analyzed together since there is a clear wave intensity profile symmetry around the equator in the DEMETER spectrograms. Figures 4a and 4b show a dependence of the first principal component coefficient on Kp index obtained for dayside and nightside measurements, respectively. Black points correspond to individual measured frequency-latitude spectrograms and red curves denote median values of the PC1 coefficient in given Kp index intervals.

Figure 4 demonstrates that while the values of PC1 obtained for dayside measurements significantly increase during geomagnetically disturbed periods, PC1 coefficients calculated for nightside orbits are
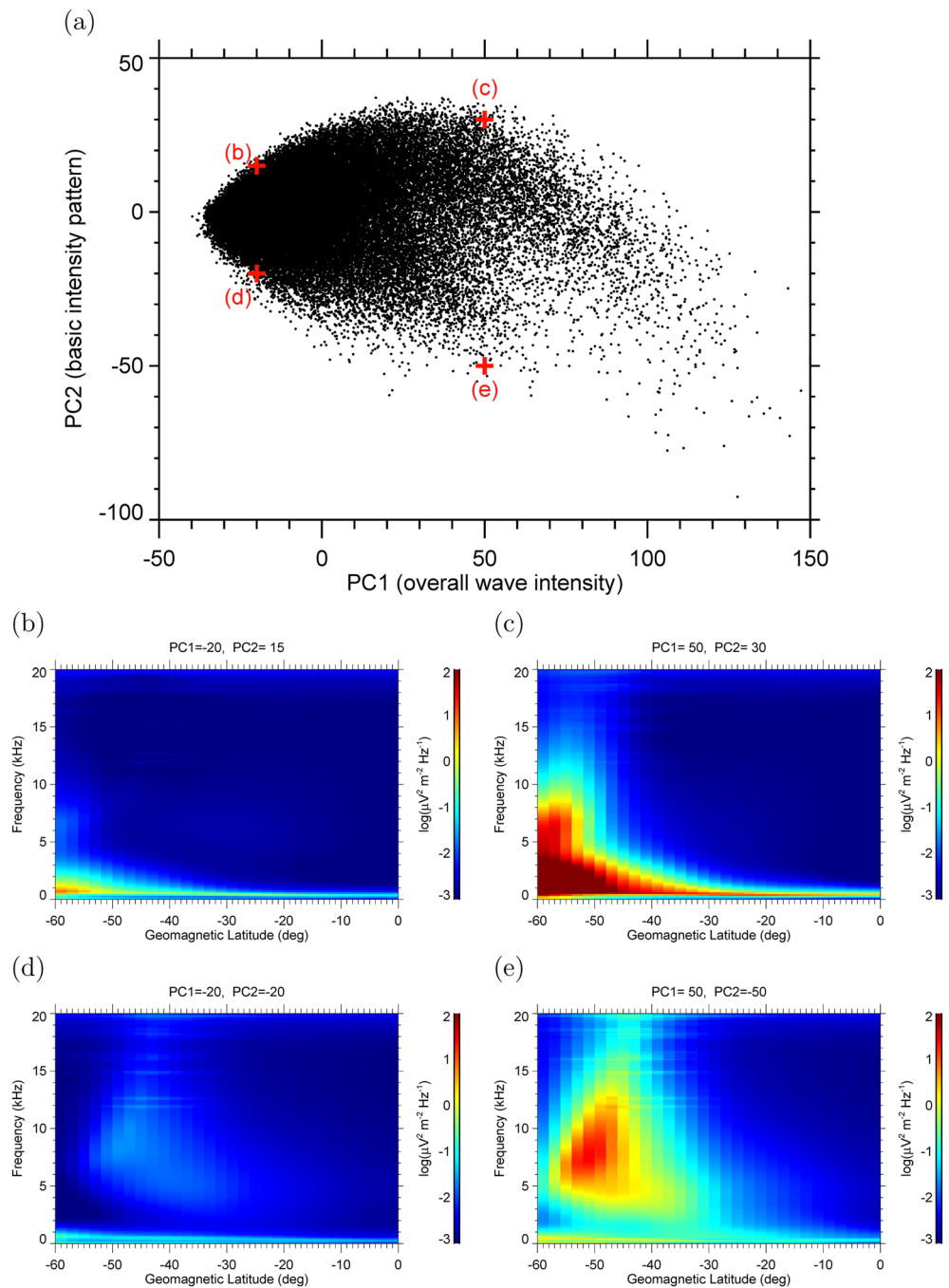
Figure 2. Effect of the first two principal components on the appearance of individual frequency-latitude spectrograms. (a) Scatter plot of individual principal component coefficients PC1 and PC2 obtained for all dayside spectrograms. (b)–(e) Frequency-latitude spectrograms obtained by using the first two principal component coefficients. Their coordinates are plotted in panel (a) by red crosses and correspond to (b) [−20, 15], (c) [50, 30], (d) [−20, −20], (e) [50, −50].

basically independent of the *Kp* index. Note that the observed variations of the median values for very large *Kp* indices are likely due to only few spectrograms detected under so high *Kp* index values.

Another possibility of how to check the behavior of PC1 coefficients is to investigate their seasonal and longitudinal variations. The obtained results are depicted in Figure 5. It shows the mean values of the first
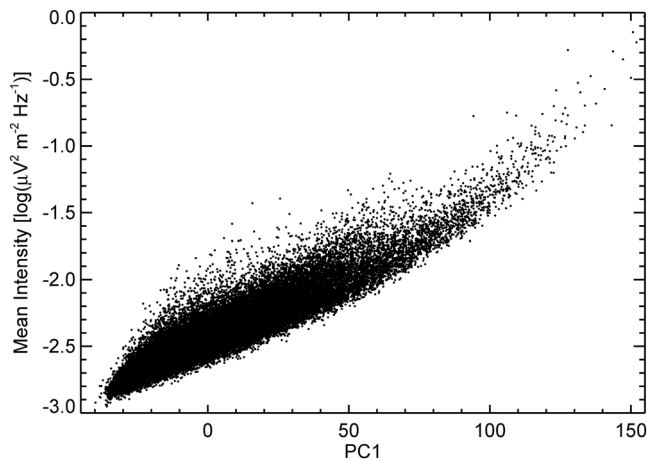
Figure 3. Dependence of the mean wave intensities in individual frequency-latitude spectrograms on the respective PC1 coefficients.

principal coefficient in given geomagnetic longitude-month bins. A longitudinal resolution of 10° was chosen. Results obtained for dayside and nighttime measurements are shown in Figures 5a and 5b, respectively.

At first sight, Figures 5a and 5b might be considered to be qualitatively similar. However, quantitatively they are not. The scale used for the mean PC1 values for dayside measurements has much lower range than for the nightside measurements. While mean daytime PC1 coefficient values span between about −15 and 34, nighttime coefficient values expand between about −50 and 66. Moreover, while the maximal values obtained for the dayside measurements occur almost exclusively during northern summer (between June and August), significant mean PC1 coefficient values for nighttime measurements occur also in the northern winter (in January, February, and December). Geomagnetic longitudes of maximal PC1 coefficients roughly correspond to locations of continents as discussed, for example, by Němec et al. (2010). Maximal PC1 coefficients are found at geomagnetic longitudes between about 300° and 360° both during the daytime and nighttime, which approximately corresponds to the longitudes of North America.

Note that due to the splitting of individual DEMETER frequency-latitude spectrograms to the northern and southern parts, we may plot the same dependences as in Figure 5 for the data measured in the northern and southern hemispheres separately (not shown). On the nightside, the PC1 coefficient values were found to be roughly the same in the northern and southern hemispheres for all the geomagnetic latitudes and seasons. This is in a good agreement with the results presented by Němec et al. (2010) and can be likely understood in terms of lightning generated whistlers propagating approximately along the magnetic field lines between the respective hemispheres. However, the difference of PC1 coefficient values between the northern and southern hemispheres for the daytime measurements is more substantial. Specifically, the PC1 coefficient values in the northern hemisphere have comparatively low values at geomagnetic longitudes of about 50° during all seasons. We believe that this can be explained by significant natural emissions at higher L-shells which are not measured by DEMETER at these particular longitudes in the northern hemisphere due to the latitudinal limit of DEMETER measurements and the non-dipole aspect of the geomagnetic field (Němec et al., 2020). However, for the purpose of the present study, the overall PC1 coefficient variations depicted in Figure 5 are sufficient to show that the PC1 coefficient values correspond to the overall wave intensity
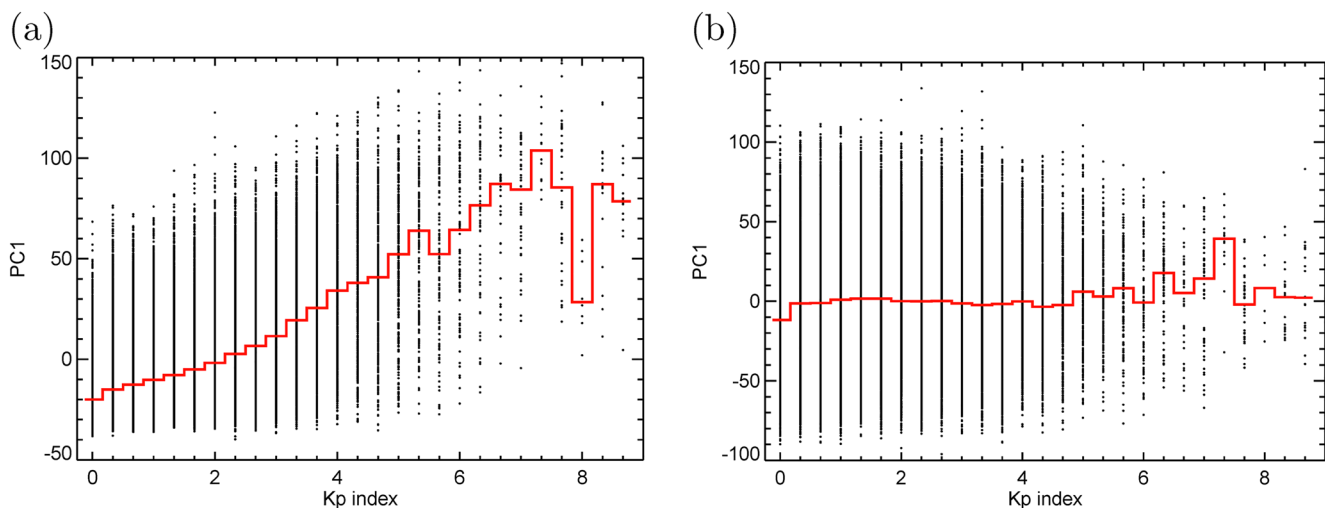

Figure 4. Dependence of the first principal component (PC1) coefficient on *Kp* index obtained for the (a) dayside and (b) nightside measurements. Black points correspond to individual frequency-latitude spectrograms measured by detection of electro-magnetic emissions transmitted from earthquake regions (DEMETER), while red curves show median values of the PC1 coefficient for given *Kp* index values.
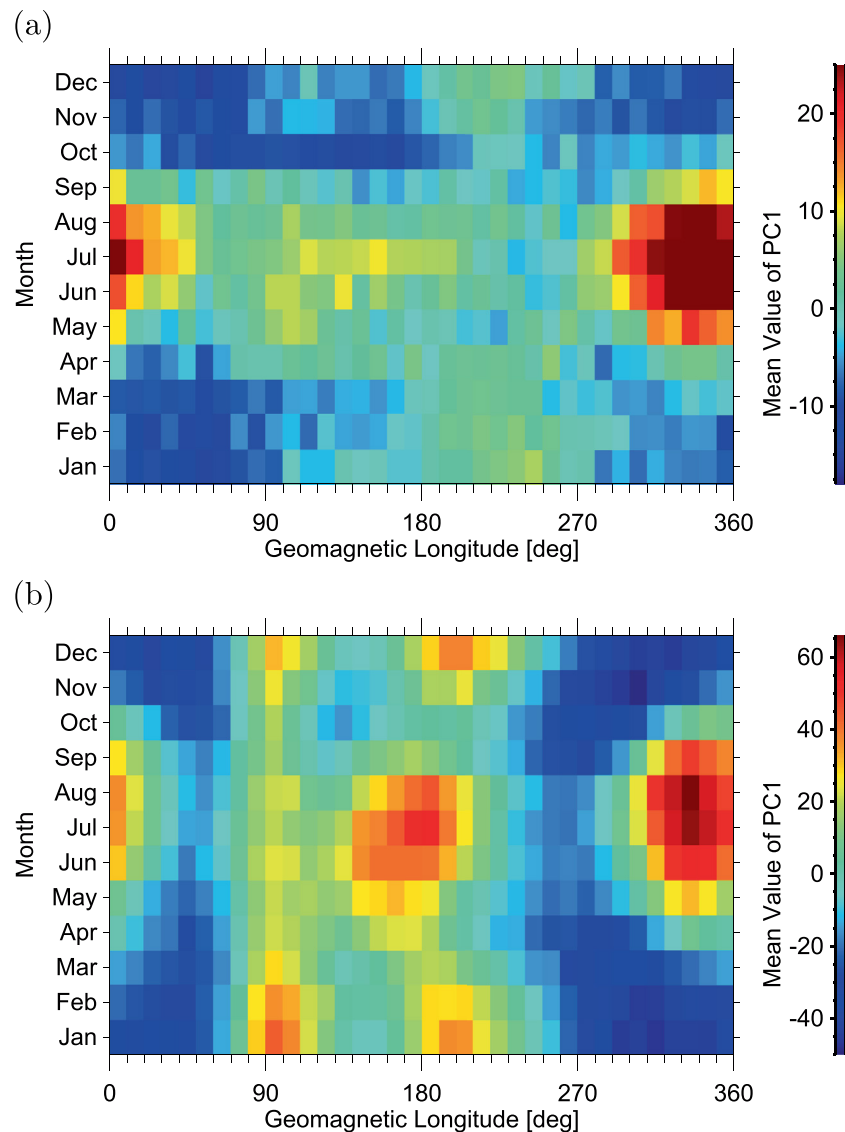
(a)

(b)

Figure 5. Mean values of the PC1 coefficients, roughly corresponding to overall wave intensities, as a function of the geomagnetic longitude (*x*-axis) and month (*y*-axis) obtained for the (a) dayside and (b) nightside detection of electromagnetic emissions transmitted from earthquake regions (DEMETER) measurements.

measured by DEMETER. The difference between the measurements in the northern and southern hemispheres is thus not further considered.

4.2. Variations of Principal Component Coefficients after Interplanetary Shock Arrivals

Having investigated the overall seasonal, longitudinal, and geomagnetic activity dependence of PC1 coefficients, it is possible to check the variations caused by the IP shock arrivals. As the nighttime wave intensities are generally quite unaffected by the geomagnetic activity changes and/or shock arrivals, only daytime (i.e., of the same local time) measurements and principal components are used for this purpose. Figures 6a–6d show variations of the daytime first principal component coefficient around the FF, FR, SF, and SR shock arrival times, respectively. The investigated time interval spans from 5 h before to 24 h after a shock arrival. PC1 coefficients corresponding to every spectrogram measured around the shock arrival time are plotted by black points in Figure 6 at a central time of the spectrogram. Median values of PC1 coefficients in every hour are drawn by red curves.
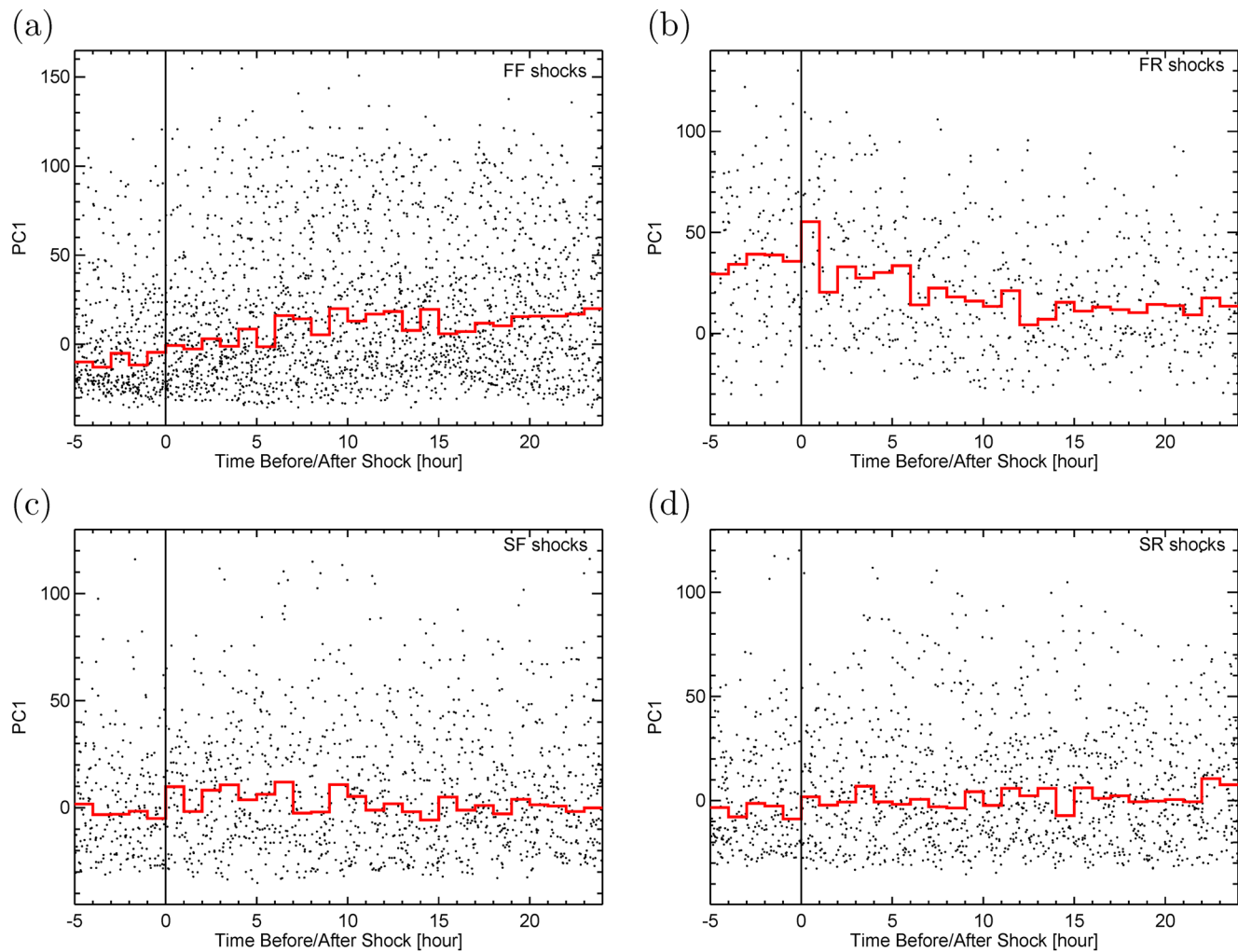
Figure 6. Variations of the daytime first principal component (PC1) coefficients with respect to the (a) FF, (b) FR, (c) SF, and (d) SR shock arrival times. Each point corresponds to a single spectrogram. Times of the shock arrivals are depicted by black vertical lines. Red curves show median values of PC1 coefficients calculated for each hour of the investigated time interval (from 5 h before to 24 h after the shocks).

Figure 6 shows that the dependence of the PC1 coefficient significantly changes according to the type of the shock. While there is a substantial evolution for fast shocks, no apparent trend for the PC1 coefficients before or after shock arrival is visible in case of slow shocks. Thus, these shocks are not investigated further. In the case of fast shocks, there is a clear opposite trend for forward and reverse shocks. While for FF shocks the PC1 coefficients tend to increase after the shock arrival, peaking about ten hours after the shock, the PC1 coefficients in the case of FR shocks rather systematically decrease after the shock arrival. However, its statistical significance is rather questionable due to the overall amount of found FR shocks (31), which is the lowest number of all analyzed shock types. Consequently, only PC1 coefficient variations around the times of FF shock arrivals are investigated hereinafter.

The evolution of PC1 coefficients obtained in Figure 6 shows that especially the FF shock arrival can cause a significant change of the PC1 coefficients. However, it is not clear how much the coefficients change for particular shock cases. As shocks can vary by their type, strength or orientation, also their effects on the wave intensity can be different. Figure 7 shows how the PC1 coefficients evolve for particular FF shocks. The format of the plots is the same as in Figure 6. The individual shock cases are distinguished by different colors. Mean values of the PC1 coefficients during 5 h before the shock arrivals were calculated and used to code individual shock cases on the scale from blue (low mean PC1 coefficient over the 5 h before the shock) to red (large mean PC1 coefficient over the 5 h before the shock).
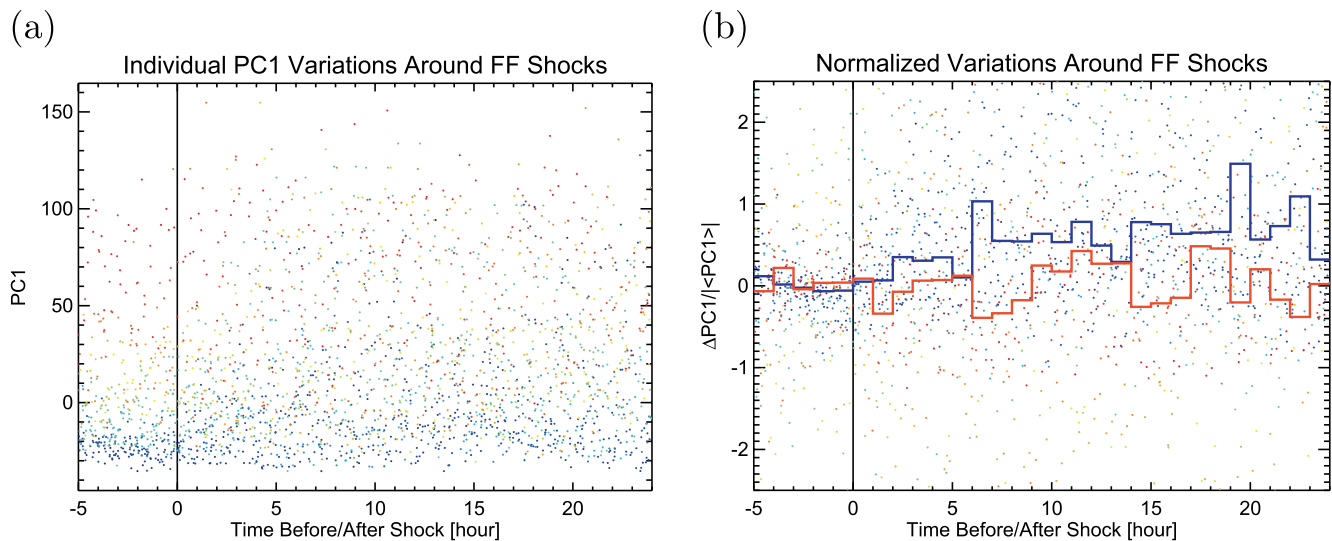
Figure 7. Detailed analysis of the daytime first principal component coefficient variations around fast forward (FF) shock arrivals. Individual cases are drawn by different colors spanning from blue to red. Particular shocks were sorted according to the mean values of the PC1 coefficients calculated within 5 h before the shock arrivals. Black vertical lines denote time of the shock arrival. (a) Color-coded variations of PC1 coefficients around FF shock arrival. (b) Relative change of PC1 variations around the shock arrival. Blue lines denote median values of PC1 for cases with mean values below −20, while red lines denote median values of PC1 with PC1 mean values above 30.

Figure 7a demonstrates how the PC1 coefficients evolve during individual shock arrivals. Figure 7b aims to investigate whether the effect on the PC1 coefficients depends on their initial values. It shows a relative change of PC1 coefficients in individual cases. Median values of PC1 relative changes are overplotted separately for low and high original PC1 coefficient values. Median values of relative changes for PC1 coefficients with original mean values below −20 are drawn by blue lines. Median values of relative changes for PC1 coefficients with mean values above 30 are drawn by red lines. It is apparent that the coefficients with different initial values behave differently. The relative change of PC1 coefficients is larger for lower initial coefficient cases than for higher initial coefficient cases. In fact, there seems to be no significant relative change for high initial coefficient values, as the respective median values (in red) after the shock arrival tend to fluctuate around zero.

The effect of additional FF shock characteristics on the PC1 evolution is analyzed in Figure 8. Intuitively, the change of PC1 coefficients should depend on the shock strength. This is defined as a ratio of IMF magnitudes measured before (B_1) and after (B_2) the shock arrival. Note that according to this definition, the FF shock B_2/B_1 ratio is always greater than one. To get sufficiently large data sets, the shocks were divided into two groups: the shocks with $B_2/B_1 < 1.5$ were considered as "weak", while the shocks with $B_2/B_1 > 2$ were considered as "strong". The shocks with $1.5 < B_2/B_1 < 2$ are omitted from the present analysis. Results obtained for the weak and strong shocks are shown in Figure 8a by blue and red colors, respectively. The plot format is the same as in Figure 6. It is seen that the PC1 coefficients corresponding to the strong shocks increase significantly more than the coefficients obtained for weak shocks.

Another worth mentioning factor affecting the shock geoefficiency and thus possibly leading to different PC1 behavior is the sign of IMF B_z component after the shock. Corresponding results are shown in Figure 8b. The PC1 coefficients obtained around FF shock arrivals are separated according to the B_z sign and plotted as blue (negative B_z) and red (positive B_z) points. Median values of the PC1 coefficients corresponding to negative and positive B_z are drawn by blue and red curves, respectively. The sign of the B_z component clearly affects the variations of PC1 coefficients. While their change around the shocks with positive B_z is rather small, the increase of PC1 coefficients around the shocks with negative B_z is quite substantial.
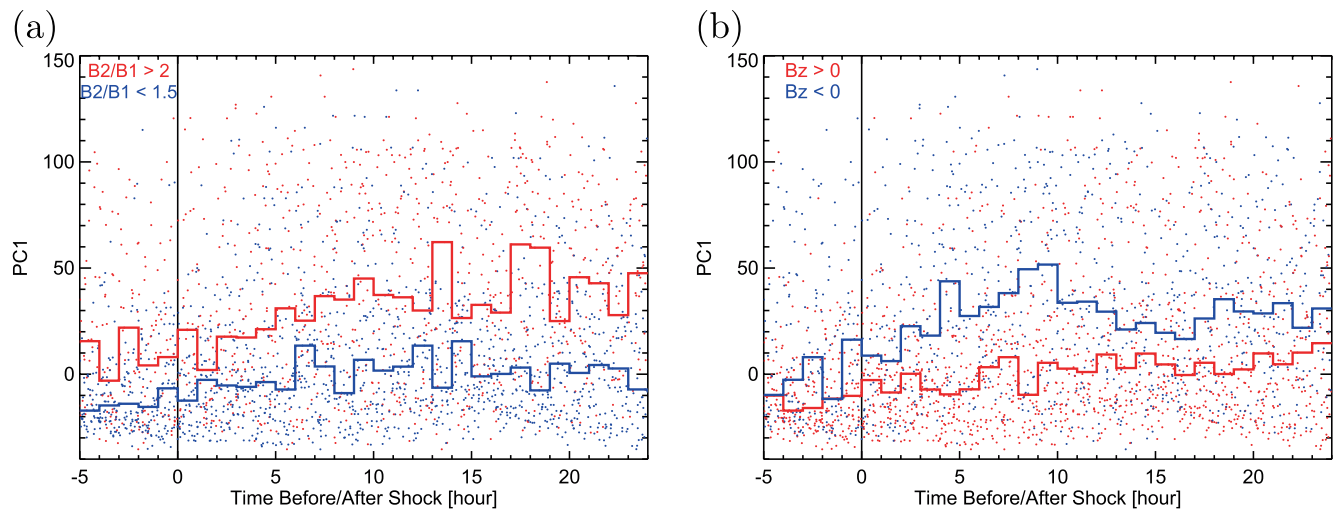
Figure 8. Variations of the daytime first principal component coefficients around fast forward (FF) shock arrivals separated according to the (a) shock strength and (b) B_z sign. The format of the plots is the same as in Figure 6.

5. Discussion

As mentioned above, a crucial task when dealing with the PCA is to find a physical interpretation of particular principal components. The first principal component includes most of the information from an original data set and it is thus regarded as the most important component. When aiming to describe basic characteristics of the original data set, this component should thus be sufficient. Considering the present case when the original data set consists of frequency-geomagnetic latitude spectrograms, being able to characterize each of them by a single number is truly an astonishing data set reduction. Although we are aware that our approach to the physical interpretation of the obtained principal components is rather simple, we showed that the first principal component at least roughly corresponds to the wave intensity measured during a given DEMETER frequency-latitude spectrogram. Additionally, unlike simple mean wave intensity, the first principal components effectively correspond to the weighted mean intensity, emphasizing the importance of frequency-latitude bins with the highest variabilities.

The suggested interpretation of the first principal component is further verified by investigating the variations of the PC1 coefficient values during different geomagnetic conditions and seasons of the year. The overall VLF wave intensity is expected to increase during geomagnetically disturbed periods, in particular during the daytime (Agapitov et al., 2014; Ma et al., 2017; Němec et al., 2010). During the nighttime, at least statistically and at the analyzed low altitudes, the wave intensity does not seem to be substantially affected by the geomagnetic activity nor shock arrivals, although some such cases were reported (Liu et al., 2017; Su et al., 2015). This corresponds well to the obtained dependences of PC1 coefficients on the *Kp* index. Moreover, the obtained longitudinal variations of dayside PC1 coefficients at different months are in agreement with the variations of the average power spectral density of electric field fluctuations obtained by Němec et al. (2020). All these results truly indicate that with a good accuracy the first principal component is affected by basic aspects of the overall wave intensity measured during given orbit and it can be applied to characterize intensity of a whole spectrogram.

The present analysis did not reveal any significant increase or decrease of PC1 coefficients after the slow shock arrivals, meaning that their effect on the VLF wave intensity measured by DEMETER is essentially negligible. This seems to be in agreement with several previous studies (e.g., Echer et al., 2003, 2004; Whang et al., 1996), which showed that the slow shocks are usually rare and comparatively weak near 1 AU.

The number of FR shocks found during the investigated time interval is the lowest from all analyzed shock types. Although at larger heliocentric distances IP shocks are commonly observed in forward-reverse pairs (Smith & Wolfe, 1976), at 1 AU this is usually not the case (Gosling et al., 1978). It was verified that the reverse and forward shocks found in the Wind spacecraft data are typically not directly associated and their

effect thus need not to be connected. The observed variations of PC1 coefficients around the time of FR shock arrivals suggest that they tend to be followed by a decrease of the wave intensity. However, the low amount of analyzed events (31) and the large scatter of obtained values make the statistical significance of this effect questionable.

The effect of FF shocks on the wave intensity is found to be the most significant. The wave intensity typically increases after the FF shock arrival. The magnitude of the increase depends on the wave intensity before the shock arrival. If the wave intensity was lower before the shock arrival, it increases more after the shock. On the other hand, in the case of high original wave intensity (larger PC1 coefficients), the shock arrival is not followed by such a substantial intensity increase. This supports an idea that there is a kind of 'saturation' for the wave intensity and it cannot be decreased arbitrarily.

The shock strength is another substantial factor affecting the wave intensity variations upon the shock arrival. The obtained results show that the stronger shock arrives (in terms of upstream/downstream magnetic field ratios), the higher increase of PC1 coefficients occurs. This is in a good agreement with intuitive expectations as well as with previously published papers suggesting a different geomagnetic response to shocks of various strengths (e.g., Ridley et al., 2006; Rudd et al., 2019; Wang et al., 2006). Moreover, it can be seen that the wave intensity increase is rather gradual, which suggests that the system response to the shock arrival happens rather due to the modification of solar wind parameters than due to the shock arrival itself.

It is thus understandable that not only the shock strength, but also the IMF orientation (i.e., the sign of B_z) plays a significant role for the PC1 coefficient response to the shock occurrence (Craven et al., 1985; Ogino et al., 1994). The wave intensity measured by DEMETER is generally more enhanced following the shocks with negative IMF B_z, corresponding to their overall larger geoeffectivity (Milan et al., 2015).

6. Conclusions

Variations of the overall VLF (up to 20 kHz) wave intensity detected by the French low-altitude DEMETER spacecraft were studied using the PCA. The obtained principal component coefficients were used to characterize every single frequency-geomagnetic latitude spectrogram measured by DEMETER. The behavior of the calculated PC1 and PC2 coefficients was at first verified by analyzing their effect on the frequency-latitude spectrogram appearance. The obtained results show that while the first principal component determines the overall wave intensity, the second principal component characterizes the basic intensity profile. Only PC1 coefficients were considered in the further analysis.

The analysis of dayside and nightside PC1 coefficients showed that while the wave intensity on the dayside depends significantly on the actual geomagnetic conditions, the nighttime wave intensities are basically independent of the geomagnetic activity. Moreover, the values of PC1 coefficients obtained during the local summer at geomagnetic longitudes corresponding to continents are increased, which is in agreement with the previous wave intensity analyses and can be attributed to lightning generated whistlers (Němec et al., 2020; Záhlava et al., 2018).

The effect of IP shock arrivals on the PC1 coefficient evolution was further studied. The results revealed that while there is basically no effect of slow shocks on the PC1 coefficients (and thus on the wave intensity), fast shocks can cause a significant change of the PC1 coefficient values. The most significant effect was found in the case of FF shocks. It was further shown that the change of PC1 coefficients depends on their initial values, being more significant for lower pre-shock PC1 coefficient values. If the PC1 coefficients are small before the shock arrival, they generally increase more. Additionally, the shock strength and IMF B_z also have significant impact on the wave intensity evolution. While the PC1 coefficients remain nearly unchanged after weak shocks ($B_2/B_1 < 1.5$) or in the case of positive IMF B_Z, substantial PC1 coefficient increase happens for strong shocks ($B_2/B_1 > 2$) followed by a southward oriented IMF.

Data Availability Statement

The whole DEMETER data set is available at https://sipad-cdpp.cnes.fr.

Journal of Geophysical Research: Space Physics

References

Agapitov, O. V., Artemyev, A. V., Mourenas, D., Kasahara, Y., & Krasnoselskikh, V. (2014). Inner belt and slot region electron lifetimes and energization rates based on AKEBONO statistics of whistler waves. *Journal of Geophysical Research: Space Physics*, *119*(4), 2876–2893. https://doi.org/10.1002/2014JA019886

Berthelier, J. J., Godefroy, M., Leblanc, F., Malingre, M., Menvielle, M., Lagoutte, D., et al. (2006). ICE, the electric field experiment on DEMETER. *Planetary and Space Science*, *54*, 456–471. https://doi.org/10.1016/j.pss.2005.10.016

Bezděková, B., Němec, F., Parrot, M., Santolík, O., & Kruparova, O. (2015). Magnetospheric line radiation: 6.5 years of observations by the DEMETER spacecraft. *Journal of Geophysical Research: Space Physics*, *120*, 9442–9456. https://doi.org/10.1002/2015JA021246

Blake, J. B., Kolasinski, W. A., Fillius, R. W., & Mullen, E. G. (1992). Injection of electrons and protons with energies of tens of MeV into L. *Geophysical Research Letters*, *19*(8), 821–824. https://doi.org/10.1029/92GL00624

Burlaga, L. F. (1971). Hydromagnetic waves and discontinuities in the solar wind. *Space Science Reviews*, *12*(5), 600–657. https://doi.org/10.1007/BF00173345

Colman, J. J., & Starks, M. J. (2013). VLF wave intensity in the plasmasphere due to tropospheric lightning. *Journal of Geophysical Research: Space Physics*, *118*(7), 4471–4482. https://doi.org/10.1002/jgra.50217

Craven, J. D., Frank, L. A., Russell, C. T., Smith, E. J., & Lepping, R. P. (1985). Global auroral responses to magnetospheric compressions by shocks in the solar wind - Two case studies. In Kamide, Y., & Slavin, J. A. (Eds.), *Solar wind magnetosphere coupling* (p. 367). https://doi.org/10.1007/978-90-277-2303-1_26

Echer, E., Alves, M. V., & Gonzalez, W. D. (2004). Geoeffectiveness of interplanetary shocks during solar minimum (1995–1996) and solar maximum (2000). *Solar Physics*, *221*(2), 361–380. https://doi.org/10.1023/B:SOLA.0000035045.65224.f3

Echer, E., Gonzalez, W. D., Vieira, L. E. A., Dal Lago, A., Guarnieri, F. L., Prestes, A., et al. (2003). Interplanetary shock parameters during solar activity maximum (2000) and minimum (1995–1996). *Brazilian Journal of Physics*, *33*(1), 115–122. https://doi.org/10.1590/S0103-97332003000100010

Fu, H. S., Cao, J. B., Mozer, F. S., Lu, H. Y., & Yang, B. (2012). Chorus intensification in response to interplanetary shock. *Journal of Geophysical Research*, *117*(A1), A01203. https://doi.org/10.1029/2011JA016913

Gonzalez, W. D., Joselyn, J. A., Kamide, Y., Kroehl, H. W., Rostoker, G., Tsurutani, B. T., & Vasyliunas, V. M. (1994). What is a geomagnetic storm? *Journal of Geophysical Research*, *99*(A4), 5771–5792. https://doi.org/10.1029/93JA02867

Gosling, J. T., Asbridge, J. R., Bame, S. J., & Feldman, W. C. (1978). Solar wind stream interfaces. *Journal of Geophysical Research*, *83*(A4), 1401–1412. https://doi.org/10.1029/JA083iA04p01401

Gotoh, K., Akinaga, Y., Hayakawa, M., & Hattori, K. (2002). Principal component analysis of ULF geomagnetic data for Izu islands earthquakes in July 2000. *Journal of Atmospheric Electricity*, *22*(1), 1–12. https://doi.org/10.1541/jae.22.1

Hattori, K., Serita, A., Yoshino, C., Hayakawa, M., & Isezaki, N. (2006). Singular spectral analysis and principal component analysis for signal discrimination of ULF geomagnetic data associated with 2000 Izu Island Earthquake Swarm. *Physics and Chemistry of the Earth, Parts A/B/C*, *31*(4–9), 281–291. https://doi.org/10.1016/j.pce.2006.02.034

Hayakawa, M., Hattori, K., & Ohta, K. (2007). Monitoring of ULF (ultra-low-frequency) geomagnetic variations associated with earthquakes. *Sensors*, *7*(7), 1108–1122. https://doi.org/10.3390/s7071108

Heppner, J. P. (1955). Note on the occurrence of world-wide S.S.C.'s during the onset of negative bays at College, Alaska. *Journal of Geophysical Research*, *60*(1), 29–32. https://doi.org/10.1029/JZ060i001p00029

Holappa, L., Mursula, K., & Asikainen, T. (2014). A new method to estimate annual solar wind parameters and contributions of different solar wind structures to geomagnetic activity. *Journal of Geophysical Research: Space Physics*, *119*(12), 9407–9418. https://doi.org/10.1002/2014JA020599

Hudson, M. K., Elkington, S. R., Lyon, J. G., Marchenko, V. A., Roth, I., Temerin, M., et al. (1997). Simulations of radiation belt formation during storm sudden commencements. *Journal of Geophysical Research*, *102*(A7), 14087–14102. https://doi.org/10.1029/97JA03995

Joliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical society of the Transactions of Royal Society A*, *374*(2065), 20150202. https://doi.org/10.1098/rsta.2015.0202

Kim, H.-J., Lyons, L. R., Ruohoniemi, J. M., Frissell, N. A., & Baker, J. B. (2012). Principal component analysis of polar cap convection. *Geophysical Research Letters*, *39*(11), L11105. https://doi.org/10.1029/2012GL052083

Kupařová, O., Maksimovic, M., Šafránková, J., Němeček, Z., Santolík, O., & Krupař, V. (2013). Automated interplanetary shock detection and its application to Wind observations. *Journal of Geophysical Research: Space Physics*, *118*(8), 4793–4803. https://doi.org/10.1002/jgra.50468

Li, J., Li, Q., Yang, D., Wang, X., Hong, D., & He, K. (2011). Principal component analysis of geomagnetic data for the Panzhihua earthquake (Ms 6.1) in August 2008. *Data Science Journal*, *10*, IAGA130–IAGA138. https://doi.org/10.2481/dsj.IAGA-20

Lin, J.-W. (2013). Is it possible to detect earlier ionospheric precursors before large earthquakes using principal component analysis (PCA)? *Arabian Journal of Geosciences*, *6*(4), 1091–1100. https://doi.org/10.1007/s12517-011-0419-z

Liu, N., Su, Z., Gao, Z., Reeves, G. D., Zheng, H., Wang, Y., & Wang, S. (2017). Shock-induced disappearance and subsequent recovery of plasmaspheric hiss: Coordinated observations of RBSP, THEMIS, and POES satellites. *Journal of Geophysical Research: Space Physics*, *122*(10), 421–510. https://doi.org/10.1002/2017JA024470

Liu, N., Su, Z., Zheng, H., Wang, Y., & Wang, S. (2018). Prompt disappearance and emergence of radiation belt magnetosonic waves induced by solar wind dynamic pressure variations. *Geophysical Research Letters*, *45*(2), 585–594. https://doi.org/10.1002/2017GL076382

Ma, Q., Mourenas, D., Li, W., Artemyev, A., & Thorne, R. M. (2017). VLF waves from ground-based transmitters observed by the Van Allen Probes: Statistical model and effects on plasmaspheric electrons. *Geophysical Research Letters*, *44*(13), 6483–6491. https://doi.org/10.1002/2017gl073885

Marshall, R. A., Inan, U. S., & Chevalier, T. W. (2008). Early VLF perturbations caused by lightning EMP-driven dissociative attachment. *Geophysical Research Letters*, *35*(21), L21807. https://doi.org/10.1029/2008GL035358

Meurant, M., Gérard, J.-C., Blockx, C., Hubert, B., & Coumans, V. (2004). Propagation of electron and proton shock-induced aurora and the role of the interplanetary magnetic field and solar wind. *Journal of Geophysical Research*, *109*(A10), A10210. https://doi.org/10.1029/2004JA010453

Milan, S. E., Carter, J. A., Korth, H., & Anderson, B. J. (2015). Principal component analysis of Birkeland currents determined by the active magnetosphere and planetary electrodynamics response experiment. *Journal of Geophysical Research: Space Physics*, *120*(12), 415–510. https://doi.org/10.1002/2015JA021680

Natali, M. P., & Meza, A. (2010). Annual and semiannual VTEC effects at low solar activity based on GPS observations at different geomagnetic latitudes. *Journal of Geophysical Research*, *115*(D18), D18106. https://doi.org/10.1029/2010JD014267

Němec, F., Santolík, O., Hospodarsky, G. B., Hajoš, M., Demekhov, A. G., Kurth, W. S., et al. (2020). Whistler mode quasiperiodic emissions: Contrasting Van Allen Probes and DEMETER occurrence rates. *Journal of Geophysical Research: Space Physics*, *125*(4), e2020JA027918. https://doi.org/10.1029/2020JA027918

Němec, F., Santolík, O., & Parrot, M. (2009). Decrease of intensity of ELF/VLF waves observed in the upper ionosphere close to earthquakes: A statistical study. *Journal of Geophysical Research*, *114*(A4), A04303. https://doi.org/10.1029/2008JA013972

Němec, F., Santolík, O., Parrot, M., & Rodger, C. J. (2010). Relationship between median intensities of electromagnetic emissions in the VLF range and lightning activity. *Journal of Geophysical Research*, *115*(A8), A08315. https://doi.org/10.1029/2010JA015296

Ogino, T., Walker, R. J., & Ashour-Abdalla, M. (1994). A global magnetohydrodynamic simulation of the response of the magnetosphere to a northward turning of the interplanetary magnetic field. *Journal of Geophysical Research*, *99*(A6), 11027–11042. https://doi.org/10.1029/93JA03313

Park, C. G., & Miller, T. R. (1979). Sunday decreases in magnetospheric VLF wave activity. *Journal of Geophysical Research*, *84*(A3), 943–950. https://doi.org/10.1029/JA084iA03p00943

Parrot, M. (1990). World map of ELF/VLF emissions as observed by a low-orbiting satellite. *Annales Geophysicae*, *8*, 135–145

Píša, D., Němec, F., Santolík, O., Parrot, M., & Rycroft, M. (2013). Additional attenuation of natural VLF electromagnetic waves observed by the DEMETER spacecraft resulting from preseismic activity. *Journal of Geophysical Research: Space Physics*, *118*(8), 5286–5295. https://doi.org/10.1002/jgra.50469

Ridley, A. J., De Zeeuw, D. L., Manchester, W. B., & Hansen, K. C. (2006). The magnetospheric and ionospheric response to a very strong interplanetary shock and coronal mass ejection. *Advances in Space Research*, *38*(2), 263–272. https://doi.org/10.1016/j.asr.2006.06.010

Rudd, J. T., Oliveira, D. M., Bhaskar, A., & Halford, A. J. (2019). How do interplanetary shock impact angles control the size of the geoeffective magnetosphere? *Advances in Space Research*, *63*(1), 317–326. https://doi.org/10.1016/j.asr.2018.09.013

Serita, A., Hattori, K., Yoshino, C., Hayakawa, M., & Isezaki, N. (2005). Principal component analysis and singular spectrum analysis of ULF geomagnetic data associated with earthquakes. *Natural Hazards and Earth System Sciences*, *5*(5), 685–689. https://doi.org/10.5194/nhess-5-685-2005

Smith, E. J., & Wolfe, J. H. (1976). Observations of interaction regions and corotating shocks between one and five AU: Pioneers 10 and 11. *Geophysical Research Letters*, *3*(3), 137–140. https://doi.org/10.1029/GL003i003p00137

Su, Z., Liu, N., Zheng, H., Wang, Y., & Wang, S. (2018). Multipoint observations of nightside plasmaspheric hiss generated by substorm-injected electrons. *Geophysical Research Letters*, *45*(20), 10–921. https://doi.org/10.1029/2018GL079927

Su, Z., Zhu, H., Xiao, F., Zheng, H., Wang, Y., Shen, C., et al. (2015). Disappearance of plasmaspheric hiss following interplanetary shock. *Geophysical Research Letters*, *42*(9), 3129–3140. https://doi.org/10.1002/2015GL063906

Su, Z., Zong, Q.-G., Yue, C., Wang, Y., Zhang, H., & Zheng, H. (2011). Proton auroral intensification induced by interplanetary shock on 7 November 2004. *Journal of Geophysical Research*, *116*(A8), A08223. https://doi.org/10.1029/2010JA016239

Šulić, D. M., Srećković, V. A., & Mihajlov, A. A. (2016). A study of VLF signals variations associated with the changes of ionization level in the D-region in consequence of solar conditions. *Advances in Space Research*, *57*(4), 1029–1043. https://doi.org/10.1016/j.asr.2015.12.025

Tsurutani, B. T., Park, S. A., Falkowski, B. J., Lakhina, G. S., Pickett, J. S., Bortnik, J., et al. (2018). Plasmaspheric hiss: Coherent and intense. *Journal of Geophysical Research: Space Physics*, *123*(12), 009–10. https://doi.org/10.1029/2018JA025975

Wang, C., Li, C. X., Huang, Z. H., & Richardson, J. D. (2006). Effect of interplanetary shock strengths and orientations on storm sudden commencement rise times. *Geophysical Research Letters*, *33*(14), L14104. https://doi.org/10.1029/2006GL025966

Whang, Y. C., Zhou, J., Lepping, R. P., & Ogilvie, K. W. (1996). Interplanetary slow shock observed from Wind. *Geophysical Research Letters*, *23*(10), 1239–1242. https://doi.org/10.1029/96GL01358

Yue, C., Chen, L., Bortnik, J., Ma, Q., Thorne, R. M., Angelopoulos, V., et al. (2017). The characteristic response of whistler mode waves to interplanetary shocks. *Journal of Geophysical Research: Space Physics*, *122*(10), 10047–10057. https://doi.org/10.1002/2017JA024574

Yue, C., Ma, Q., Jun, C. W., Bortnik, J., Zong, Q., Zhou, X., et al. (2020). The modulation of plasma and waves by background electron density irregularities in the inner magnetosphere. *Geophysical Research Letters*, *47*(15), e2020GL088855. https://doi.org/10.1029/2020GL088855

Záhlava, J., Němec, F., Santolík, O., Kolmašova, I., Hospodarsky, G. B., Parrot, M., et al. (2018). Longitudinal dependence of whistler mode electromagnetic waves in the Earth's inner magnetosphere. *Journal of Geophysical Research: Space Physics*, *123*(8), 6562–6575. https://doi.org/10.1029/2018JA025284

Zhima, Z., Cao, J., Liu, W., Fu, H., Wang, T., Zhang, X., & Shen, X. (2014). Storm time evolution of ELF/VLF waves observed by DEMETER satellite. *Journal of Geophysical Research: Space Physics*, *119*(4), 2612–2622. https://doi.org/10.1002/2013JA019237

Zhou, X., & Tsurutani, B. T. (1999). Rapid intensification and propagation of the dayside aurora: Large scale interplanetary pressure pulses (fast shocks). *Geophysical Research Letters*, *26*(8), 1097–1100. https://doi.org/10.1029/1999GL900173

Zong, Q.-G., Zhou, X.-Z., Wang, Y. F., Li, X., Song, P., Baker, D. N., et al. (2009). Energetic electron response to ULF waves induced by interplanetary shocks in the outer radiation belt. *Journal of Geophysical Research*, *114*(A10), A10204. https://doi.org/10.1029/2009JA014393