

FULL PAPER

Open Access



Ground-based VLF wave intensity variations investigated by the principal component analysis

Barbora Bezděková^{1*} , František Němec¹ and Jyrki Manninen²

Abstract

Very low frequency wave intensity variations measured by the Kannuslehto station, Finland in the frequency range 0–12 kHz between 2016 and 2020 are analyzed by the principal component analysis (PCA). As the analyzed ground-based measurements are basically continuous, the length of individual basis vectors entering into PCA is fundamentally arbitrary. To better characterize both long- and short-period variations, two PCAs with different lengths of the basis vectors are eventually performed. Specifically, either daily frequency–time spectrograms or individual frequency spectra are chosen as the PCA basis vectors. Analysis of the first three principal components shows substantial variations of the wave intensity due to seasonal and local time effects. Intensity variations related to the geomagnetic activity characterized by Kp and AE indices and standard deviation of the magnetic field magnitude are less significant. Moreover, PCA allows one to distinguish between nighttime and daytime Kannuslehto variations and study them independently. Solar and geomagnetic activity effects on the daytime and nighttime measurements are discussed. Wave intensity variations related to substorm occurrence are also analyzed.

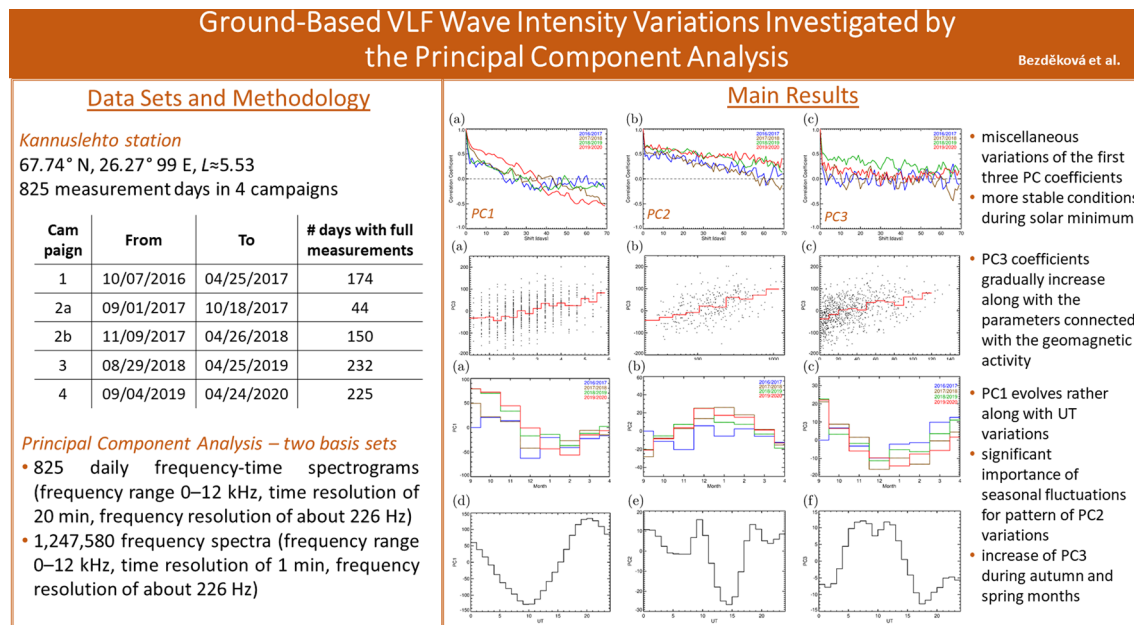
Keywords: ELF/VLF wave intensity, Principal component analysis, Geomagnetic activity, Solar activity, Seasonal variations

*Correspondence: barbora.bezdekova@mff.cuni.cz

¹ Faculty of Mathematics and Physics, Charles University, V Holešovičkách 747/2, 18000 Prague, Czech Republic

Full list of author information is available at the end of the article

Graphic Abstract



Introduction

Natural (i.e., non-anthropogenic) whistler mode waves observed on the ground are basically of two origins. Either they are generated directly near the Earth and are typically related with the lightning occurrence, or they originate in the magnetosphere and then propagate down to the Earth (Helliwell 1965). To exit through the bottom of the ionosphere, where the refractive index suddenly drops, the Snell's law tells us that the incident wave vectors need to be oriented nearly vertically down. For this reason, typically only whistler mode waves ducted for a considerable part of their propagation path are able to make it all the way to the ground (Helliwell 1965). Moreover, when penetrating through the ionosphere, waves suffer from a significant attenuation (especially on the dayside) (e.g., Graf et al. 2013). Due to the relatively strict conditions waves have to meet to make it through the ionosphere, their ground-based measurements are substantially limited (e.g., Sonwalkar 1995). However, once the waves exit the ionosphere, they can propagate considerable distances in the Earth–ionosphere waveguide before they are finally detected by a receiver.

Analysis of extremely and very low frequency (ELF, VLF) whistler mode waves is one of the crucial tools used to understand processes taking place in the Earth's inner magnetosphere. Due to its relatively simple availability, ground-based ELF/VLF wave measurements play an extraordinary role. Since effects occurring in the magnetosphere are typically closely linked, studying

the ELF/VLF waves also helps to reveal important facts about other phenomena. Although statistical studies are more suitable to describe global properties of processes affecting the whole system under various conditions, case studies are typically performed instead as data for such analyses are better available. However, long-term and climatic effects can be predominantly identified statistically. Many of the relevant statistical studies are based on measurements performed by a single ground-based station (e.g., Golden et al. 2009; Smith and Jenkins 1998; Smith et al. 2010). Multistation observations (e.g., Chrisan and Fraser-Smith 1996; Laaspere et al. 1964; Suzuki and Sato 1987) or a combination of ground-based and spacecraft observations (e.g., Hayakawa et al. 1977; Martinez-Calderon et al. 2016; Simms et al. 2019) are also sometimes used.

A comprehensive analysis of ELF/VLF signals observed predominantly by the Halley station, Antarctica covering 16 years (1992–2007) of measurements was performed by Smith et al. (2010). They analyzed electromagnetic wave data in the frequency range 0.3–10 kHz and identified that the lower part of this interval corresponds mainly to whistler mode chorus and hiss waves, while atmospherics from tropical lightnings dominate at higher frequencies. They further found out that the wave intensities vary substantially along with the local time and season of the year and that the whistler mode waves depend on the solar illumination of the ionosphere. They also compared Halley measurements with other Antarctica

stations and concluded that the chorus and lightning intensities decrease, while the auroral hiss intensity increases at higher latitudes. In addition, they reported a substantial enhancement of magnetospheric wave activity at the equinoxes and of chorus waves during geomagnetically disturbed periods.

An automatic procedure for detecting chorus and hiss waves observed between May 2000 and May 2010 at the Palmer station, Antarctica was introduced by Golden et al. (2011). Besides a presentation of the algorithm, the study focuses on the dependence of the wave occurrence on the geomagnetic activity described in terms of the Kp and AE indices. It is shown that the analyzed whistler mode waves occur preferentially during the times of an enhanced geomagnetic activity and their observations are also affected by local time and season of the year. While chorus was primarily observed in the dawn sector, hiss emissions were detected rather in the dusk sector. Both types of emissions were predominantly observed during winter months rather than during summer season.

Yonezu et al. (2017) reported a statistical analysis of measurements from three different ELF/VLF ground-based stations (Athabasca, Canada; Kannuslehto, Finland; Syowa, Antarctica). The results show that the simultaneous wave occurrence rate is higher in the dayside morning sector, suggesting possible MLT dependence of the wave longitudinal extent. Moreover, the simultaneous ELF/VLF emissions were observed primarily during periods with higher geomagnetic activity, described in terms of the AE and Dst indices. In addition, an enhancement of solar wind dynamic pressure was detected during simultaneous observations of hiss emissions. Similar results demonstrating that a magnetosphere compression can lead to an enhancement of whistler mode wave activity were also published by Manninen et al. (2016); Shiokawa et al. (2014).

One year of data from the Athabasca station, Canada were further used for a statistical study by Martinez-Calderon et al. (2015). The results show that ELF/VLF waves occur predominantly during the morning hours and they are correlated with a storm and substorm activity. The substorm activity acting from 2 days up to 1 h before the detection of ELF/VLF signals was found to be relevant for the wave occurrence. Solar wind parameters and geomagnetic activity were also confirmed to affect the ELF/VLF observations. Occurrence of chorus waves during geomagnetic storms was further studied by Smith et al. (2004a, 2004b) and Spasojevic (2014). These studies indicate that the amount of observed waves is enhanced during and after the geomagnetic storms.

The principal component analysis (PCA) is typically used to effectively reduce the dimensionality of an original data set while maintaining most of the original

information. It was successfully applied also to space physics problems. Variations of VLF wave intensities were recently studied using PCA by Bezdekova et al. (2021). The study is focused on the analysis of measurements performed by the French low-altitude spacecraft DEMETER, covering about 6.5 years of data. Variations of VLF intensities are described mainly in terms of the first principal component coefficient. Effects caused by overall geomagnetic activity, seasonal and longitudinal variations, and interplanetary shock arrivals were investigated.

In the present paper, we adapt the technique introduced by Bezdekova et al. (2021) to ground-based VLF measurements performed by the Kannuslehto station, Finland. Following the idea of the method, the first three principal component coefficients are used to describe seasonal and diurnal variations of the wave intensity, as well as the intensity variations related to the geomagnetic activity. A short review of the main idea of PCA, technical parameters of the Kannuslehto station and Sodankylä IMAGE magnetometer are described in “Data sets and methods”. Then, the obtained results are presented and discussed (in “Results” and “Discussion”). Main findings are summarized in “Conclusions”.

Data sets and methods

Measurements used in the present study were performed by the Finnish ground-based VLF receiver Kannuslehto located at 67.74° N, 26.27° E at *L*-shell about 5.53. The station is operated by Sodankylä Geophysical Observatory (SGO) in Sodankylä. Two orthogonal vertical magnetic loop antennas oriented in the north-south and east-west directions provide measurements in the frequency range 0.2–39 kHz. Their sizes are the same (10×10 m) and both have 10 turns, corresponding to effective areas of 1000 m². Sampling rate of data is 78,125 Hz. The measurements exhibit a wide dynamic range (up to 120 dB) and an extraordinary sensitivity (≈ 0.1 fT). The station typically operates during campaigns, which usually last several months (excluding summer, when the data would be polluted by significant lightning emissions). Data from altogether four campaigns were used in the present study to efficiently capture not only the seasonal variations, but also the variations given by the solar cycle and related phenomena. Campaigns between 2016 and 2020 were considered. Hence, considering prospective missing measurement time intervals, the used data set consists of 825 days fully covered by measurements. The individual campaigns used in the paper are summarized in Table 1.

Another data used in the present study are those of a ground-based magnetometer located in Sodankylä. It belongs to the IMAGE magnetometer chain, which

Table 1 Kannuslehto campaigns used in the present study

| Campaign | From | To | # days with full measurements |
|----------|------------------|-----------------|-------------------------------|
| 1 | 7 October 2016 | 25 April 2017 | 174 |
| 2a | 1 September 2017 | 18 October 2017 | 44 |
| 2b | 9 November 2017 | 26 April 2018 | 150 |
| 3 | 29 August 2018 | 25 April 2019 | 232 |
| 4 | 4 September 2019 | 24 April 2020 | 225 |

provides continuous measurements of the three components of the ambient magnetic field in Finland and in Northern Scandinavia. The time resolution of the data used in the study is 1 min, which is sufficient to provide information about low-frequency magnetic field variations related to the geomagnetic activity in the vicinity of the VLF receiver.

A suitable method for an analysis of a large data set when the data are assumed to be somehow correlated is the principal component analysis (PCA) (e.g., Jolliffe and Cadima 2016). This method was recently used by Bezděková et al. (2021) to handle data provided by a French low-altitude spacecraft DEMETER and a similar approach is applied in this study. A detailed description of the method is also included in the mentioned paper. For now, let us remind that PCA is able to reveal the most significant variations contained in the original data and it also provides weights of the individual variations. Its main purpose is to reduce a dimensionality of large data sets while preserving as much information as possible. A new basis of mutually orthogonal vectors, called principal components, is calculated within this routine. The principal components are linear functions of original vectors and they are sorted according to the variances of the original projections. Components with the highest variances are placed at first and then follow the others. Moreover, the variances correspond to the amount of the information from the original data set carried by the new variables. When frequency–time spectrograms are analyzed via PCA, each frequency–time spectrogram, i.e., a two-dimensional matrix, is considered as an original basis vector.

After computing a new PCA basis, a backward reconstruction of the original data set can be performed as a linear combination of the principal components. The more principal components are used, the more precise form of the original spectrogram is obtained. As discussed already by Bezděková et al. (2021), only a few first principal components are enough to capture the main characteristic of the original spectrogram. This is

due to the comparatively large amount of information from the original data set carried by the first several components. Hence, only a small number of the principal components (typically first two or three) describe the original data with a sufficient accuracy. In concrete calculations, one can effectively use a set of coefficients calculated from a decomposition of the original vectors into the principal components to characterize them. This is especially useful for more dimensional objects like frequency–time spectrograms. Let us further remind that when dealing with PCA in physical problems, a main problem is to understand the physical meaning of individual principal components. The present study aims not only to reveal possible effects affecting the ground-based VLF measurements (in terms of the principal components), but also to estimate how comparatively significant these effects are. Since PCA is essentially based on determining the deviations from an average behavior of the original data set, the mean value is subtracted from all original basis vectors before the start of the calculation.

When using PCA for ground-based (continuous) measurements, the question how to choose the basis vectors, i.e., how to split the more or less continuous measurements, arises. Two different approaches are applied in the present study, allowing us to study the variations of VLF intensities at different time scales. First, daily frequency–time spectrograms are used as basis vectors. The original data set hence consists of 825 frequency–time spectrograms in the frequency range 0–12 kHz with a time resolution of 20 min and a frequency resolution of about 226 Hz. The used frequency interval is lower than the actual frequency range of Kannuslehto measurements, to avoid problems due to signals coming from near Russian transmitters emitting at higher frequencies. The chosen basis is suitable mainly for a description of seasonal and long-period effects, such as variations due to solar cycle or changes related to periods with enhanced geomagnetic activity. However, when analyzing variations of the wave intensity on shorter time scales, the PCA performed using the daily frequency–time spectrograms as basis vectors is not useful. For this reason, it is necessary to choose another set of vectors and compute a new PCA basis with a finer time resolution. To perform such PCA, the Kannuslehto measurements are recalculated to obtain spectra with a time resolution of 1 min and a frequency resolution of about 226 Hz. These are used as a new original data set. Altogether, 1,247,580 spectra in the frequency range 0–12 kHz are analyzed. Note that the number of vectors used for this analysis is higher than it would correspond to the aforementioned 825 days of measurements. The reason is that in this analysis, it is

possible to include also measurements from incomplete measuring days which are omitted from the daily PCA.

Results

Entire day spectrograms as basis vectors

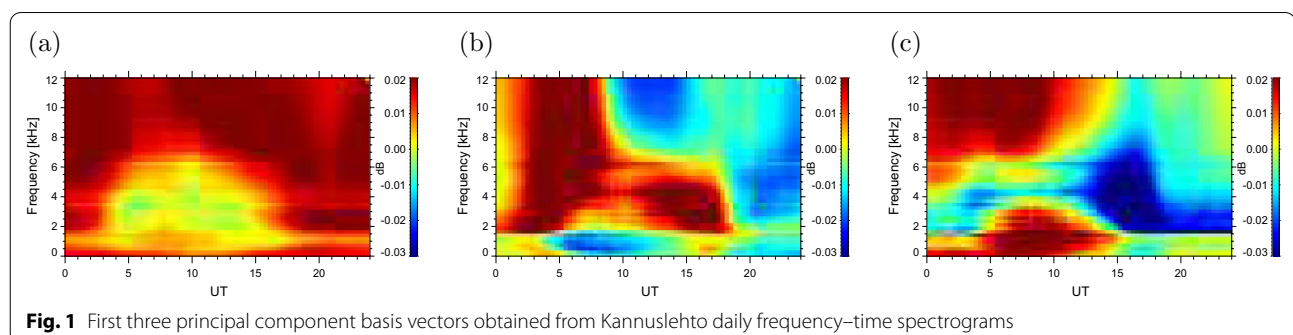
Following the approach introduced by Bezděková et al. (2021), the principal component basis vectors were plotted and manually checked to reveal their characteristic intensity patterns. First three principal basis vectors are shown in Fig. 1. They cover almost 60 % of the original information. Note that while the results are plotted as a function of universal time (UT), local time (LT) at Kannuslehto is larger by about 1.5 h. Moreover, magnetic local time (MLT) at Kannuslehto is larger by about 2.5 h with respect to UT. It is directly seen that the first principal component expresses mainly VLF wave intensity measured during the night, while the second principal component rather describes the dayside intensity, corresponding to about 10 % of the original information. The first principal component corresponds well to the average intensity profile obtained for this data set (not shown). Thus, as expected, the first principal component reveals main intensity profile features. These are mainly given by nightside measurements, where strong lightning generated whistlers occur. The first principal component corresponds to more than 40 % of the original information. The third principal component seems to express the intensity variations on the dawn side rather on the dusk side. Its physical interpretation needs to be explored more carefully (see below). It covers about 6 % of the original information. We note that the sudden intensity variations observed in the frequency spectra at about 1.5 kHz are likely related to the first cutoff frequency of the Earth–ionosphere waveguide (Budden 1961). While the wave above the cutoff frequency can propagate considerable distances in the waveguide, the waves at lower frequencies are essentially detectable only close to the ionospheric exit point.

To better interpret the first two principal components, which carry most of the original information, it is useful

to draw a scatter plot showing their mutual dependence and to investigate how a change of individual PC coefficients is related to frequency–time spectrogram features. This is done in Fig. 2.

Figure 2a shows a scatter of individual PC1 and PC2 values for each frequency–time spectrogram from the original data set. The points are color-coded according to individual measuring campaigns. The same color coding is used in all further plots. Blue points correspond to the campaign 2016/2017, brown to measurements obtained during campaign 2017/2018, green indicate measurements during 2018/2019 campaign, and red points were measured within 2019/2020 campaign. It is seen that the points corresponding to different campaigns are distributed over the whole range of obtained coefficient values equivalently. There is thus no “extraordinary” campaign with some preferred interval of PC1 or PC2 values. From this point of view, the individual campaigns are equal and they can be compared between each other, allowing us to assume that possible differences between them are of physical origin, not given by the data processing.

Evolution of the wave intensity given by the change of PC1 and PC2 is shown in Fig. 2b–e. The coordinates (in terms of PC1 and PC2) chosen for these figures are drawn in Fig. 2a by orange crosses along with the letter corresponding to an appropriate plot panel. It can be seen that positive values of PC1 coefficients correspond to a significant increase of the nighttime wave intensity (about 0–5 UT and 15–24 UT), while positive values of PC2 correspond to an increase of the daytime wave intensity (about 5–15 UT). The most intense spectrogram is hence obtained for large positive values of both PC1 and PC2 (Fig. 2c). This supports the idea suggested already by the principal component profiles shown in Fig. 1 that while the first principal component corresponds to the nighttime VLF measurements, the second principal component describes rather the daytime VLF measurements. Note again that the physical information related to the third principal component is more tricky and it will be discussed more in detail further.



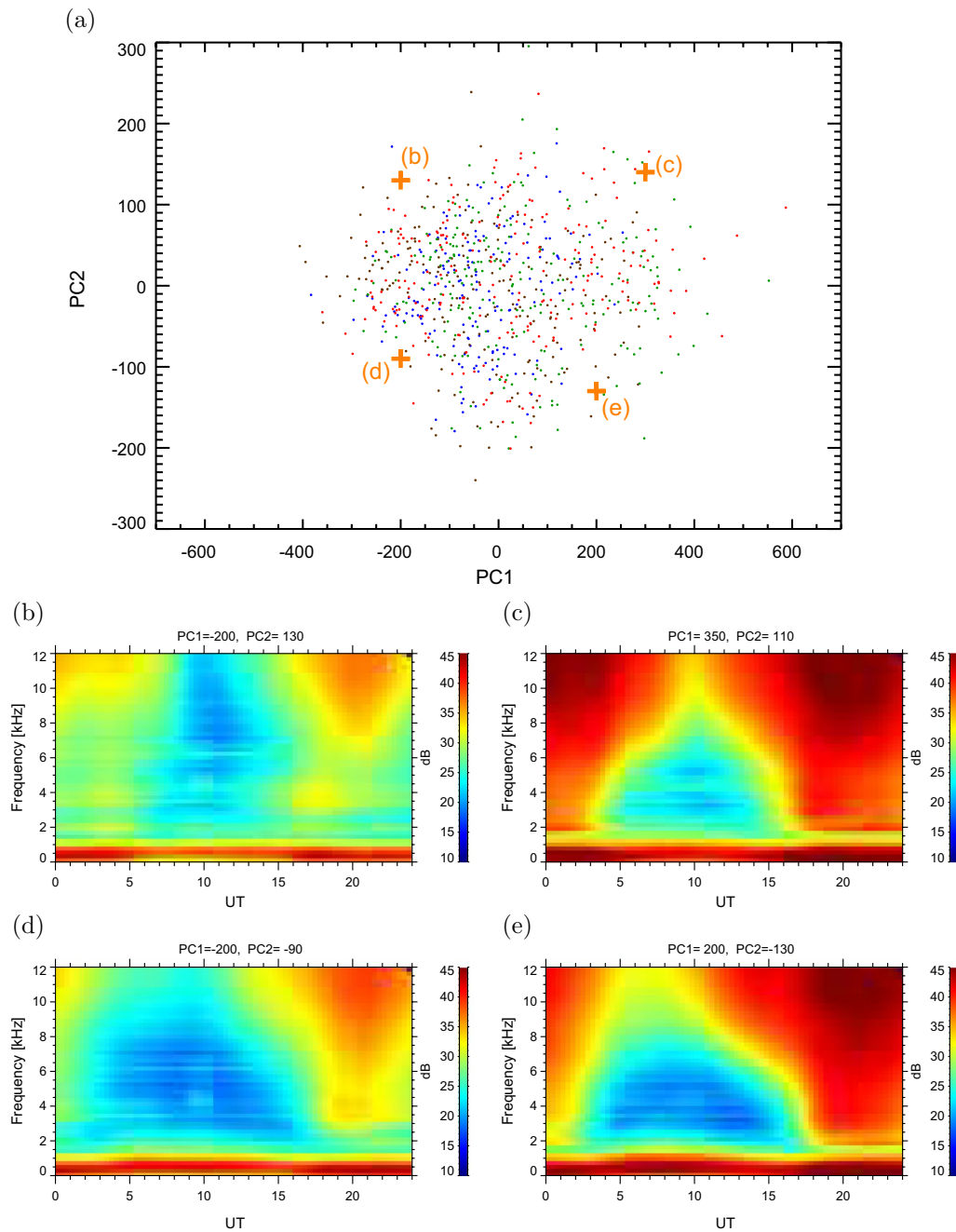


Fig. 2 **a** Scatter plot of individual principal component coefficients PC1 and PC2 obtained for daily frequency–time spectrograms. The colors of points distinguish individual Kannuslehto campaigns—blue points correspond to the 2016/2017 campaign, brown denote the 2017/2018 campaign, green indicate the 2018/2019 campaign, and red points were measured during the 2019/2020 campaign. **b–e** Frequency–time spectrograms obtained using the first two principal component coefficients. Their coordinates are plotted in **a** by orange crosses and correspond to **b** [−200, 130], **c** [350, 110], **d** [−200, −90], **e** [200, −130]

After getting an idea about the possible physical interpretation of at least the first two principal components, we aim to further investigate how individual PC coefficients vary during the season of the year. Since it is

obvious that in the ground-based measurements the seasonal dependence has a significant effect, it has to be somehow reflected by the principal components. Fig. 3a–c shows the mean values of PC1, PC2, and PC3

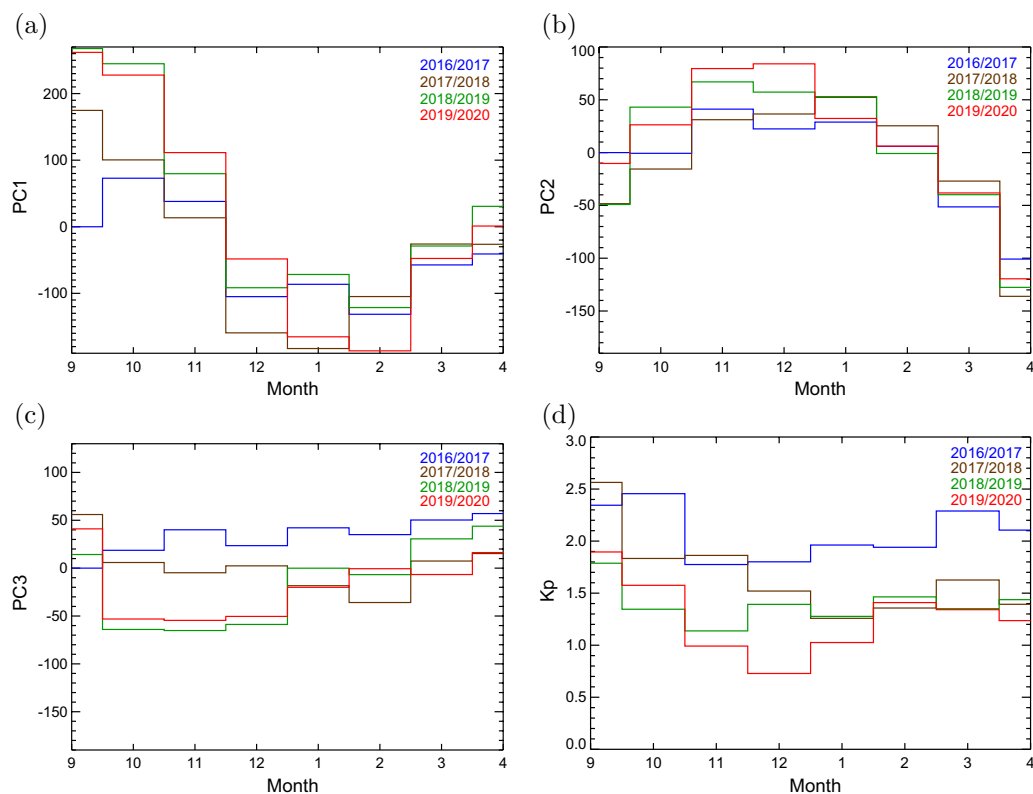


Fig. 3 Seasonal variations of the average (a) PC1, (b) PC2, (c) PC3 coefficients, and (d) Kp index

coefficients as a function of the months of campaigns. In addition, monthly average values of Kp index are shown in Fig. 3d, giving an idea about the variations of an overall geomagnetic activity during these months. The dependences are shown for each campaign separately and they are distinguished by different colors, following the color coding introduced along with Fig. 2a.

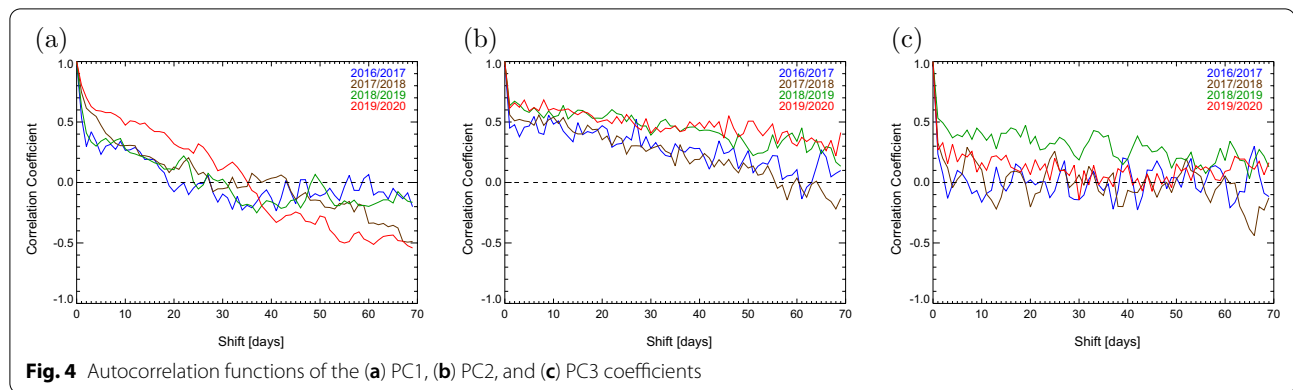
Figure 3a, b shows that PC1 and PC2 coefficients evolve in a completely different way. While the PC1 coefficients reach the highest values during autumn and spring months, the largest values of PC2 are reached in November or December, i.e., at months corresponding to or very close to the winter solstice. However, maximal values of both PC1 and PC2 in individual months are typically reached either for the 2018/2019 or 2019/2020 campaign. The trends obtained for both coefficients are in no way comparable with the Kp index variations shown in Fig. 3d.

Seasonal dependence of the PC3 coefficients shown in Fig. 3c is quite different in comparison to the previous PC coefficients. There is no pronounced maximum or minimum as in the previous cases and the maximal average PC3 coefficients in individual months are mainly reached for the 2016/2017 campaign. From this point of view, the

seasonal variations of the PC3 coefficients agree more with the Kp index dependence than for the other two PC coefficients. Although a direct correlation between the average PC3 coefficients and Kp indices is only approximate, the results shown in Fig. 3 indicate that if any principal component (out of the first three) could be related to the overall geomagnetic activity (in terms of Kp index), it is PC3.

The time scales at which individual PC coefficients evolve are analyzed in Fig. 4. It shows autocorrelation functions of the first three principal component coefficients for time lags from 1 to 70 days for individual Kanuslehto campaigns separately. Only the days of year when the data from all the four campaigns are available are used for this analysis.

The variations of autocorrelation functions obtained for PC1, PC2, and PC3 are significantly different. In the case of PC1 coefficients (Fig. 4a), the autocorrelations turn to be negative after around 30 days. In the case of the 2016/2017 campaign, it happens already after about 20 days. After becoming negative, the sign of correlation coefficients remains more or less the same for the rest of the investigated shift interval. The most significant change occurs for the 2019/2020 campaign, where the



difference between the positive (for short time lags) and negative (for long time lags) correlation coefficients is the largest. Autocorrelations obtained for the PC2 coefficients shown in Fig. 4b remain positive for basically the entire time lag interval, except for the 2017/2018 campaign values which turn to be negative after about 55 days, and they gradually decrease with increasing time lags. Autocorrelations of the 2016/2017 campaign are also negative at about 60 days time lag, but since they further reach positive values again this seems to be rather a random effect.

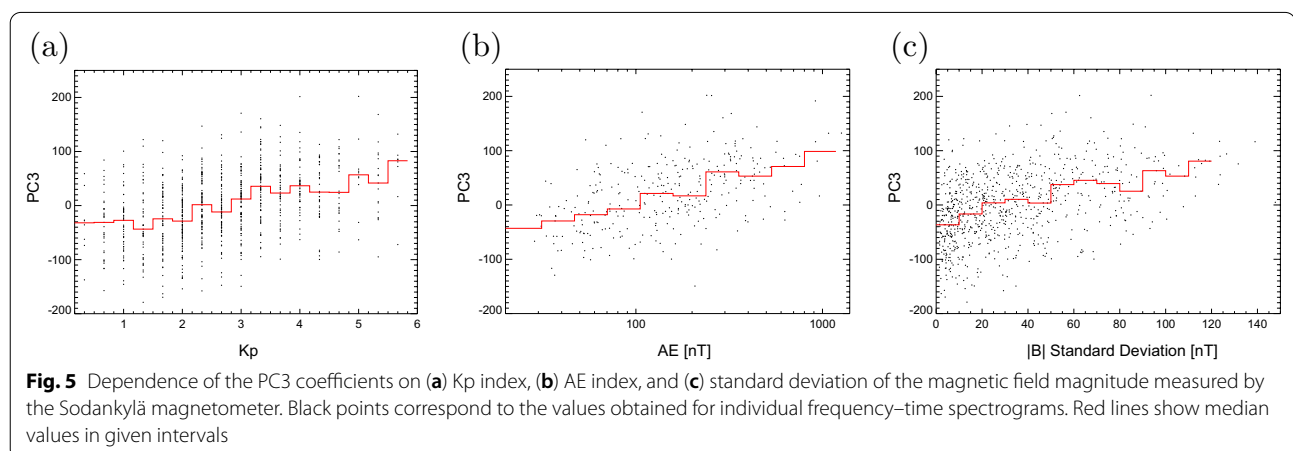
A completely different picture of autocorrelation functions is obtained for PC3 as shown in Fig. 4c. Autocorrelations obtained for the 2018/2019 campaign decrease only slowly towards zero and they remain positive for the whole analyzed interval of the time lags. Autocorrelation values obtained for other campaigns are lower and tend to fluctuate around zero. A similar behavior of the autocorrelation function is obtained for the Kp index (not shown).

Although the proper physical interpretation of PC3 has not been done yet, the previous results indicate that

it could carry, at least partially, information about wave intensity variations related to the geomagnetic activity. To confirm this hypothesis, it is necessary to find other relevant parameters which also provide information about or are affected by the geomagnetic activity. This is investigated further in Fig. 5, which shows the dependence of PC3 on the Kp index (Fig. 5a), AE index (Fig. 5b), and standard deviation of the magnetic field magnitude measured by the Sodankylä magnetometer (Fig. 5c).

It is clearly seen that all three dependences exhibit basically the same behavior—the PC3 coefficients gradually increase with given parameters. Given that all the three parameters are somehow connected with the geomagnetic activity, it is indeed reasonable to conclude that the PC3 coefficients increase along with geomagnetic activity.

When analyzing global effects which could influence the VLF wave intensity at Kannuslehto, we already mentioned the geomagnetic activity, predominantly described by Kp index. In this regard, it is important to note that the four analyzed campaigns took place during different phases of the solar cycle. The evolutions



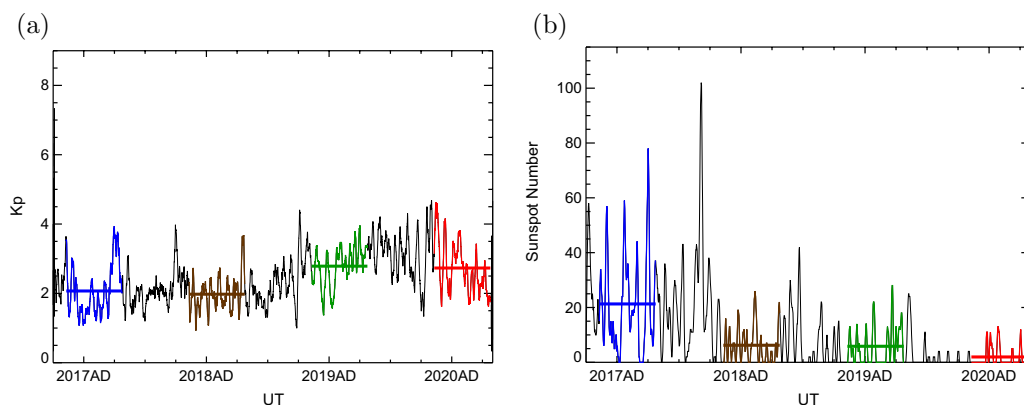


Fig. 6 Variations of (a) Kp index and (b) sunspot number during the years of the analyzed Kannuslehto campaigns. The horizontal lines show the mean values of given parameters within the analyzed campaign intervals

of both Kp index and sunspot number during the years of the investigated Kannuslehto campaigns are shown in Fig. 6. The intervals of campaigns which were used in the previous plots are drawn by the corresponding colors as introduced above. To better visualize the evolution of the parameters during the individual campaigns, mean values of the parameters over the campaign intervals are drawn by horizontal lines.

While the Kp indices during the first two campaigns (2016/2017, 2017/2018) were quite similar, their values increased for the latter two campaigns (2018/2019, 2019/2020). Similarly to the first two campaigns, in terms of the mean values the geomagnetic activity during these two campaigns was comparable. Fig. 6b shows that the highest solar activity occurred during the 2016/2017 campaign, then it significantly dropped, and it eventually reached the minimum during the 2019/2020 campaign (solar minimum was observed in December 2019).

Frequency spectra as basis vectors

As discussed above, to better characterize the wave intensity evolution on shorter time scales, PCA of individual frequency spectra with the time resolution of 1 min as basis vectors is used. First three principal components obtained are shown in Fig. 7. In this case, the physical interpretation of the principal components is more complicated. For now, let us only describe the profiles of the first three principal components depicted in Fig. 7. These three principal components carry almost 95 % of the original information. Most of the information is included in the first principal component (Fig. 7a), which carries about 81 % of the information. This component is almost constant at higher frequencies, but in the frequency range up to 2 kHz, where it is significantly lower, it decreases and drops close to zero at about 1.5 kHz. This is due to the fact that in the frequency range around 1.5 kHz the wave power is usually substantially higher than anywhere else, but for arbitrary Kannuslehto spectrograms it remains more or less same. Fig. 7b shows the second principal component. It can be seen that at higher

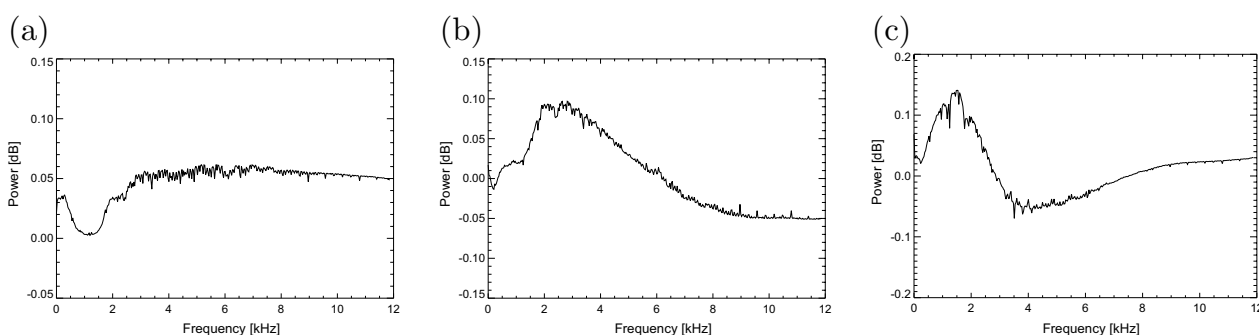


Fig. 7 First three principal component basis vectors obtained for individual frequency spectra

frequencies (above about 6 kHz) its sign turns to be negative. The second principal component carries about 9 % of the original information and it contributes significantly to the wave power in the frequency range between about 2 and 5 kHz where it is significantly increased. The third principal component shown in Fig. 7c reaches negative values in the frequency range between about 2 and 8 kHz and it also substantially increases at frequencies around 2 kHz. Out of the three components shown, it reaches the largest values of the wave power and it carries about 4% of the original information.

To better understand the physical meaning of the obtained principal components, it is again useful to draw a scatter plot and check how the frequency spectra vary with respect to the given PC coefficients. A scatter plot of PC1 and PC2 is depicted in Fig. 8 along with four reconstructed spectra corresponding to selected combinations of PC1 and PC2 coefficient values.

Due to the high number of the original frequency spectra (1,247,580), the scatter plot in Fig. 8a is depicted using a slightly different format to make the plot more comprehensible. It shows a number of individual original vectors associated to PC coefficients in given PC1–PC2 bins. The width of each bin is set to 10 in both dimensions. It is seen that the PC1, PC2 distribution is centered around zero, but it is not symmetric. Moreover, it seems that most of the frequency spectra are associated with negative or small positive PC2. This means that the increase of wave power observed for the second principal component (Fig. 7b) in the frequency range between about 2 and 5 kHz is not usual and the wave power in this range is typically rather decreased. A visual inspection of an arbitrary Kannuslehto frequency–time spectrogram confirms this interpretation. However, it remains unclear what this principal component in fact describes and if this can be indeed considered as a general feature of the original data set. The distribution of PC1 is roughly symmetric around zero, suggesting that the contribution of PC1 to the wave intensity can be both positive and negative. Considering that PC1 is almost constant, this is not a surprising result.

The effect of PC1 and PC2 coefficient values on the frequency spectra is seen in Fig. 8b–e. The values of PC1 and PC2 coefficients are chosen to correspond to extreme values. These are marked in the scatter plot (Fig. 8a) by the green crosses along with a letter of a corresponding panel in Fig. 8. It is worth mentioning that an arbitrary combination of PC1 and PC2 leads to a maximum value of the wave power in the frequency range up to about 1 kHz. Only the concrete patterns of these maxima vary. The increase in the frequency range between about 2 and 5 kHz observed in the spectrum of the second principal component is pronounced in the wave intensity

only if both PC1 and PC2 are positive. In other cases, PC1 makes this increase basically negligible. It is further seen that a positive PC1 coefficient makes the decrease of the wave power at about 1.5 kHz more obvious (Fig. 8c, e). Furthermore, the wave power at larger frequencies (above about 4 kHz) tends to be anticorrelated with the PC2 coefficients.

Figure 9 aims to identify possible controlling factors for the first three principal components. It shows dependences of PC1, PC2, and PC3 on month of the campaign (Fig. 9a–c) and on UT (Fig. 9d–f). While the dependences on month are drawn for each Kannuslehto campaign separately as the dependences for individual campaigns noticeably vary, the dependences on UT are drawn averaged over the campaigns, because they turned out to be almost identical for all campaigns.

As Fig. 9a–c show, PC1 and PC3 exhibit similar seasonal variations, while PC2 exhibits quite an opposite trend. PC1 and PC3 are increased during autumn and spring months, while their values are minimal during winter months. Minimal values of PC1 for individual campaigns are reached either in December or February and PC3 values are minimal either in December or January. On the contrary, the PC2 coefficients peak either in December or January and they are minimal for most of the campaigns in September. The PC coefficient dependences on UT obtained for the individual components are rather different. Notice that the local time at Kannuslehto is shifted with respect to the UT by about 1.5 h, i.e., the local noon corresponds to about 10:30 UT. The PC1 coefficient dependence exhibits two global extremes—minimum between about 9 and 11 UT and maximum between 20 and 21 UT. Considering the time shift, the global minimum obtained for the PC1 coefficients corresponds well to the Kannuslehto noon. Moreover, the obtained extremes are quite symmetric as their absolute values are almost identical. The dependences obtained for PC2 and PC3 are different. The PC2 coefficient values are typically rather positive or slightly negative during night and morning hours, and they become negative after 11 UT, reaching the minimum at about 15 UT. After 17 UT they turn to be positive again. Maximum values are reached between 9 and 10 UT and between 18 and 20 UT. Positive values of PC3 (Fig. 9c) are reached between 4 and 15 UT, peaking at about 8 and 11 UT, while minimal values occur between 17 and 18 UT.

Exploiting the fine time resolution of the original data set, it is possible to investigate how the PC coefficients are affected by a substorm occurrence. A substorm list used in the present study was provided by the SuperMAG network (Gjerloev 2012; Newell and Gjerloev 2011a, b). Results of this analysis are shown in Fig. 10. It shows the average time dependence of the first three

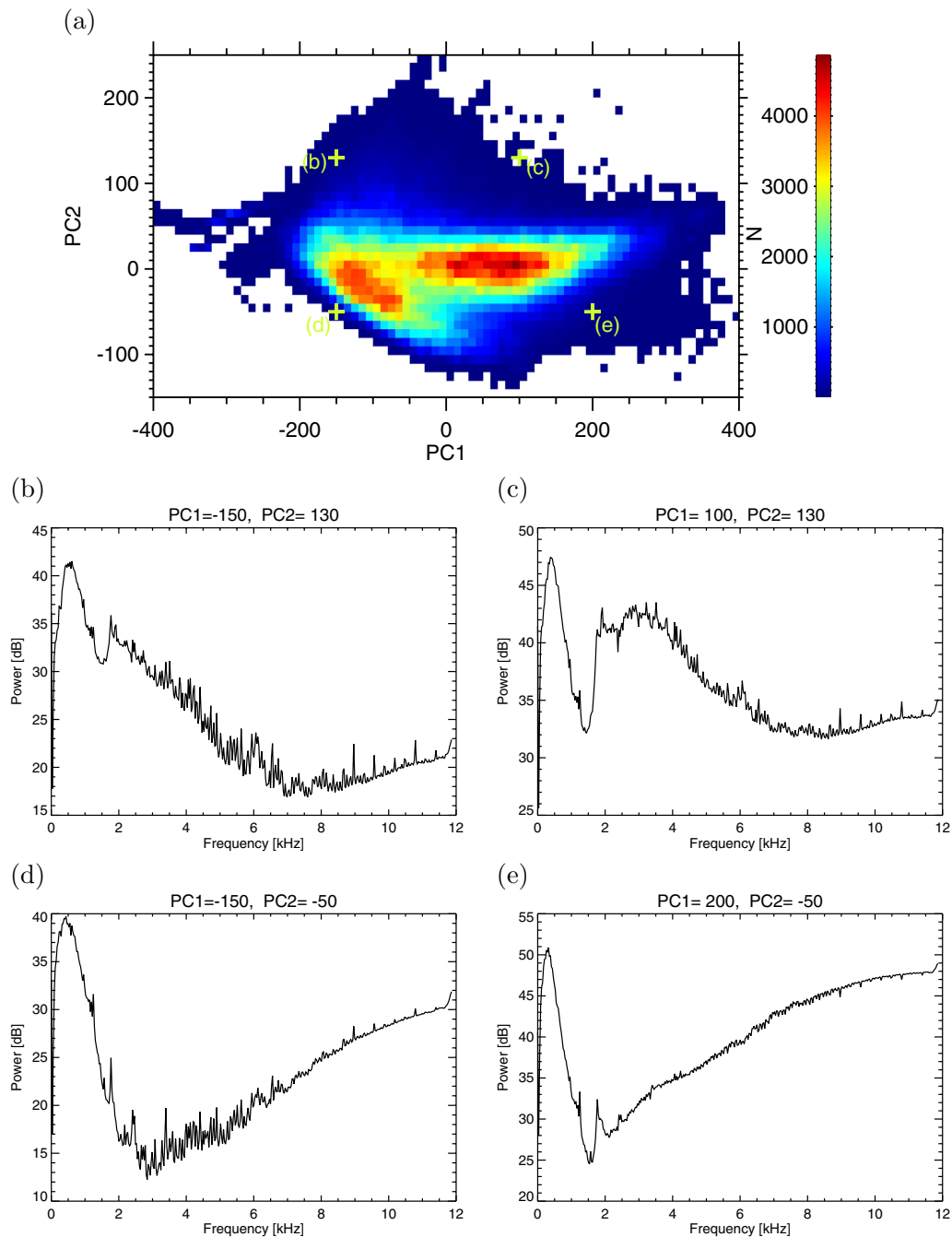


Fig. 8 **a** Scatter plot of individual principal component coefficients PC1 and PC2 obtained for frequency spectra. **b–e** Frequency spectra obtained using the first two principal component coefficients. Their coordinates are plotted in panel (a) by green crosses and correspond to **b** [– 150, 130], **c** [100, 130], **d** [– 150, – 50], **e** [200, – 50]

PC coefficients in the case when no substorms occurred between 6 h before and 6 h after the time of the measurement (black curves) and in the case when at least 16 substorms were detected in the given time interval (red curves). It is seen that the trend obtained for the PC1

coefficients (Fig. 10a) is very similar in both cases. Apart from the high increase of PC1 for the large number of substorms between 2 and 6 UT, the PC1 coefficients during large substorm numbers are rather lower than in no substorm situation. Remark that the profile obtained in

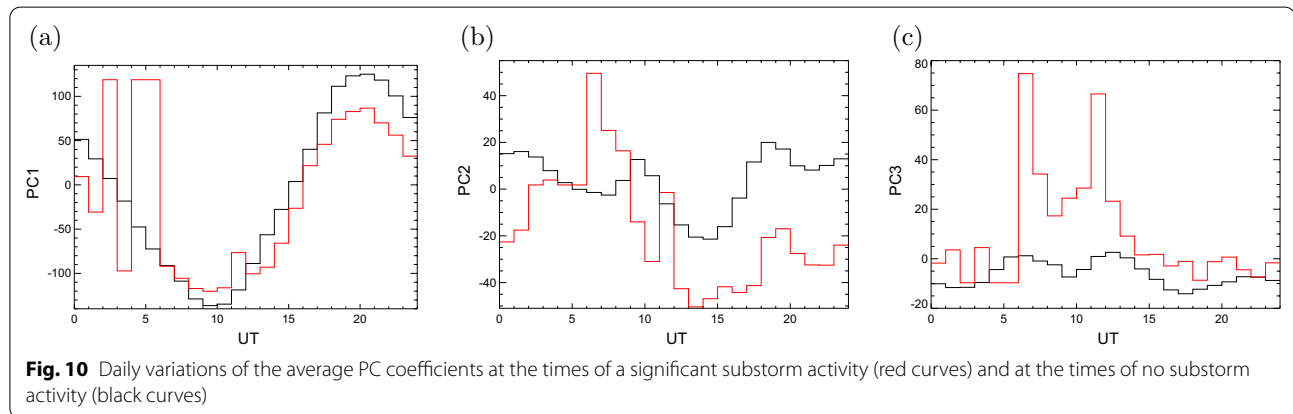
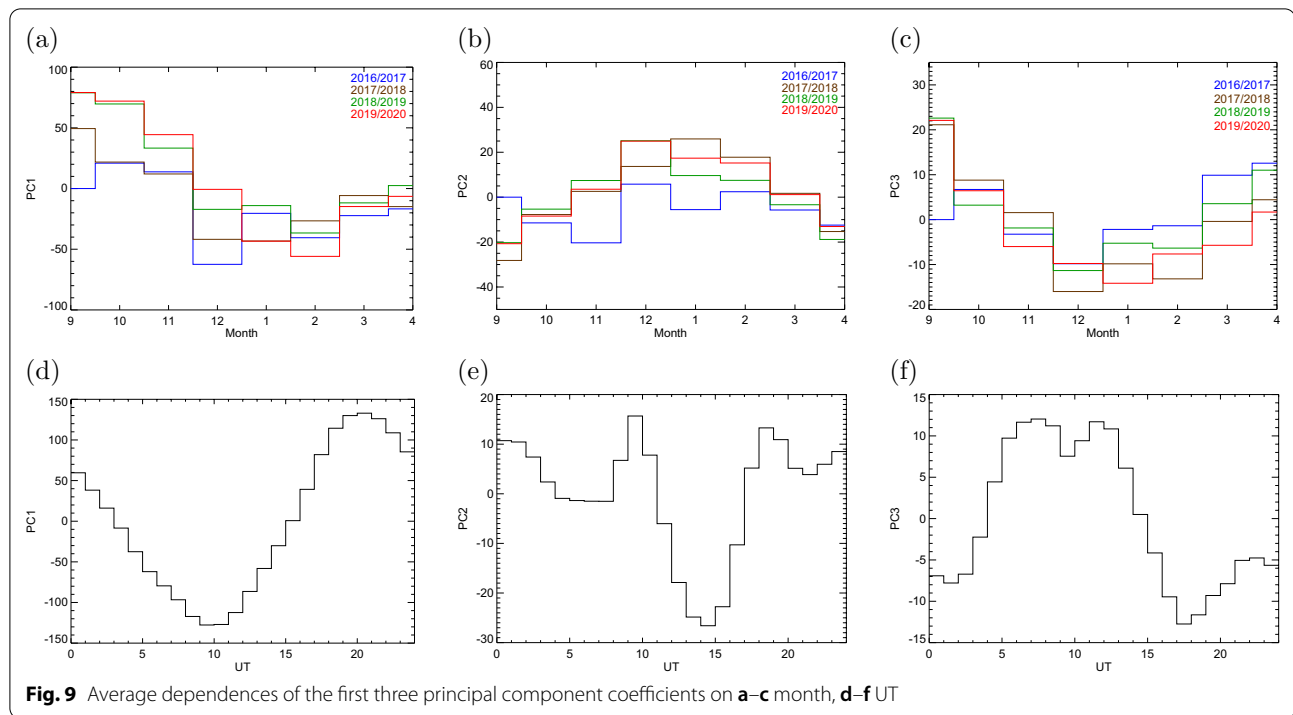


Fig. 10a is very similar to the overall UT dependence of PC1 shown in Fig. 9d. The substorm number thus does not significantly affect PC1.

This conclusion essentially holds also for the PC2 coefficients depicted in Fig. 10b. Again, the PC2 coefficients turn to be rather lower at the times of a significant substorms activity than in the case of no substorms. An exception is again the fine interval between about 5 and 9 UT where the PC2 values for large substorm numbers increase more than for no substorms. Similarly to PC1, especially variations of PC2 for no substorms correspond to the overall PC2 dependence on UT shown in Fig. 9e as a set of no substorms covers more than 36% of the original data set.

The situation for the PC3 coefficients is significantly different (Fig. 10c). The PC3 coefficients obtained at the times of a substantial number of substorms are mostly higher than the PC3 coefficients at the times of no substorms. Moreover, while in the case of no substorms the PC3 coefficients tend to be rather negative, the maximal average PC3 coefficients for a large amount of substorms is almost 80. These maximal average values of the PC3 coefficients are reached between 6 and 7 UT and from 11 to 12 UT.

Discussion

As discussed already by Bezděková et al. (2021), the crucial task when working with PCA is to give an appropriate physical interpretation of calculated principal

components. This, in turn, allows us to get an idea about the relative importance of individual factors affecting the VLF wave intensity measured by Kannuslehto. When using the basis of daily frequency–time spectrograms, the meaning of the first two principal components can be estimated quite easily. Already from the obtained principal components (i.e., the frequency–time spectrograms) forming the new vector basis, it is clear that the first principal component is related to the nighttime measurements, while the second principal component contains rather the information about daytime intensities. The VLF intensities measured by Kannuslehto during the night are typically significantly higher than the dayside intensities (intensity difference between daytime and nighttime measurements can at times exceed 20 dB), which is due to a lower ionospheric attenuation (e.g., Greifinger and Greifinger 1968). This is well consistent with the first principal component association with the nighttime intensities. In addition, there are other, more specific features, which are related to the higher principal components. These can at times significantly affect the wave intensity. When browsing through individual frequency–time spectrograms, one can notice that such features are rare, but eventually significant. From this point of view, it is somewhat surprising that the daytime intensity variations related to the second principal component correspond to only about 10% of the original information.

PCA further allows us to characterize the effect of the geomagnetic activity on the Kannuslehto measurements, albeit it is comparatively weaker than the diurnal variations. This is manifested by the fact that it takes the third PC to reveal the geomagnetic activity dependence, as the first two are essentially geomagnetic activity-independent. Moreover, if the geomagnetic activity effect already manifests, it causes the provable enhancement of the wave intensity. Considering the frequency–time spectrogram of the third principal component, it is further noteworthy that the effect is mostly related to the wave intensity measured on the dawn side. This supports the dawn–dusk asymmetry which also holds for observations of many whistler mode wave phenomena (e.g., Walsh et al. 2014, and references herein). Previous statistical studies of ground-based observations are consistent with this conclusion (e.g., Golden et al. 2011; Yonezu et al. 2017). The results further indicate that the variability of the dawn side intensities is larger than the variability of the dusk side intensities, as the higher principal components do not seem to be related with the dusk local time sector (not shown).

The analysis of autocorrelation functions obtained for individual PC coefficients and Kannuslehto campaigns provides an estimate of temporal scales on which the PC coefficients vary. At the same time, a comparison

of results obtained for different campaigns allows us to appreciate a role of the solar cycle. The miscellaneous behavior of each principal component coefficient affirms their different physical background.

Figure 4a shows that the autocorrelation functions of the PC1 coefficients significantly depend on the Kannuslehto campaign. The autocorrelation times are largest for the 2019/2020 campaign, corresponding to the solar cycle minimum and above-average Kp indices. Comparing the autocorrelations obtained for the PC1 coefficients during other campaigns with respective average sunspot numbers, it would indeed seem that larger solar activity corresponds to shorter autocorrelation times. This may be possibly interpreted in terms of specific solar events which disturb the—otherwise primarily season-controlled—wave evolution. A question why the PC1 autocorrelation functions during the 2019/2020 campaign become so negative, i.e., they basically exhibit a reverse behavior after about 30 days than before, remains unclear.

Behavior of the autocorrelation functions of the PC2 coefficients seems to be in this sense quite comparable to the PC1 coefficient results. It suggests that also the daytime VLF wave intensities have longer correlation times during the solar minimum. It means that during weaker solar activity the wave intensities remain similar for quite a long period, which makes them more stable. Quite surprisingly, these solar minimum periods seem to correspond also to larger average Kp index values.

Dependences of the autocorrelation functions obtained for the PC3 coefficients are quite different, with the corresponding correlation times being comparatively much lower. Moreover, during the 2018/2019 campaign the autocorrelation function behaves quite differently than in the other campaigns. We do not possess any direct explanation for that at the moment. However, the autocorrelations obtained for the other campaigns are generally well comparable with the autocorrelation functions of Kp indices derived during the times of these campaigns (not shown). This further supports the relation of the third principal component with the geomagnetic activity. However, one needs to keep in mind that a significant portion of the original information is also contained in higher principal components. The higher components are studied, the more complicated their interpretation is. As our results show that the geomagnetic activity is not the main parameter controlling the measured wave intensities, it cannot be assumed that its effects are described by a single component. Another point is that the overall geomagnetic activity itself is a complex property not describable by a single parameter.

Physical interpretation of the principal components obtained in the case of the frequency spectra as basis

vectors is more tricky. A possible approach to interpret the obtained principal components is to calculate correlations between the new and original basis vectors. The results are not shown in the paper since they highly correspond to the dependences shown in Fig. 9. It would thus seem that the first principal component describes the UT variations of the frequency spectra. This is supported also by the fact that the trend shown in Fig. 9d agrees well to the dependence of the correlation coefficients between the original basis vectors and PC1. Minimal average correlation coefficient calculated between the first principal component and the original frequency spectra was reached for the spectra in the UT interval between 9 and 10 UT and is about -0.54 , while maximal average value reached between 20 and 21 UT is about 0.07 . Note that when the average PC1 coefficients are negative, the average correlation coefficients are negative as well. The correlation coefficients calculated between the second principal component and the original frequency spectra (reaching average values from about -0.8 to -0.53) have almost identical trend in seasonal dependence as the average PC2 coefficients. Along with the dependence shown in Fig. 9b, this indicates that PC2 essentially corresponds to monthly variations of the wave power measured by Kannuslehto. This confirms that the seasonal and daily variations of the wave intensity are typically most significant factors controlling the wave intensity. Let us remark that many phenomena occurring in the system indeed exhibit seasonal variations, such as lightning activity or effects caused by different daylight hours (e.g., ionospheric absorption).

Regarding the physical interpretation of PC3, the situation is more complex. Several parameters, such as lightning occurrence rate, various geomagnetic indices, and interplanetary shock arrivals were verified as possible controlling factors for the PC3 coefficient values, but without success. Finally, a relation between the PC3 coefficient values and a substorm occurrence was identified. This would again suggest that PC3 could be somehow related with the geomagnetic activity and related processes. However, it cannot be straightforwardly regarded as a component related strictly to the overall geomagnetic activity. Nevertheless, a substorm occurrence can significantly increase the wave intensity in the frequency range around 1.5 kHz, which is mainly characterized by the third principal component. This supports the results presented in many previous papers that a substorm occurrence leads to the enhancement of whistler mode wave events (e.g., Meredith et al. 2002; Ripoll et al. 2021; Summers et al. 2004). Note that the substantial increase of the PC3 coefficients caused by the substorm occurrence was observed on the dayside (between 6 and 12 UT). Although substorms are nighttime phenomena,

observations of chorus waves on the dawn side caused by the substorm activity were previously presented by e.g., Spasojevic and Inan (2010); Tsurutani and Smith (1977).

Conclusions

VLF wave intensity variations measured by the Finnish ground-based Kannuslehto station in the frequency range 0 – 12 kHz were studied. Data from four different Kannuslehto campaigns obtained between 2016 and 2020 were analyzed by the principal component analysis. To better characterize both the long-period and short-period variations, two separate PCAs were performed. The first of them used daily frequency–time spectrograms with a time resolution of 20 min as basis vectors, while the second used individual frequency spectra with a time resolution of 1 min as basis vectors. Although a physical interpretation of the first two principal components of these two PCAs was quite different, effects of the overall geomagnetic activity turned to be related only to the third (and eventually higher) principal component.

In the case of PCA with daily spectrograms as basis vectors, the PC1 and PC2 can be related to nighttime and daytime wave intensity variations, respectively. The distinction of the nighttime and daytime measurements (in terms of PC1 and PC2) allows us to describe effects of the solar and geomagnetic activities on them separately. While the daytime VLF wave intensity variations seemed to be more stable during periods of the lower solar activity, the nighttime variations of the wave intensity remained during these periods stable over a shorter time.

The first and second principal components obtained for the PCA with frequency spectra as basis vectors were found to be related to the UT and monthly variations, respectively. We further showed that the substorm occurrence results in an increase of the wave activity at frequencies about 1.5 kHz, mainly between about 6 and 12 UT. It was further confirmed that the wave intensity is larger during periods with higher Kp and AE indices and at the times of larger fluctuations of local magnetic field magnitudes. Although the VLF wave intensity measured on the ground is somehow related to all these factors, our results demonstrate that the season of the year and local time are the most important.

Acknowledgements

BB and FN acknowledge the support of GACR grants 20-09671S and 21-01813S. BB would like to thank her brother Jaroslav Bezděk for many useful discussions and insights about PCA, especially about the interpretation of principal components, which allowed her to interpret the corresponding principal components properly.

Authors' contributions

Kannuslehto data used in the paper were provided and processed by JM; methodology, conceptualization and final form of the results were discussed and performed by FN and BB; paper was written by BB and revised by FN and JM. All authors read and approved the final manuscript.

Funding

This research was supported by GACR Grants 20-09671S and 21-01813S.

Availability of data and materials

The quick look plots of all VLF Kannuslehto campaigns are available at <https://www.sgo.fi/Data/VLF/VLFDData.php>. IMAGE magnetometer data can be downloaded from <https://space.fmi.fi/image/www/index.php?page=home>. Substorm list can be obtained at <https://supermag.jhuapl.edu/substorms/>.

Declarations

Competing interests

The authors declare that they are not aware of any competing financial or personal interests that could cause them embarrassment after the publication of the paper.

Author details

¹Faculty of Mathematics and Physics, Charles University, V Holešovičkách 747/2, 18000 Prague, Czech Republic. ²Sodankylä Geophysical Observatory, University of Oulu, Tähteläntie 62, 99600 Sodankylä, Finland.

Received: 28 November 2021 Accepted: 28 January 2022

Published online: 14 February 2022

References

- Bezděková B, Němec F, Parrot M, Kruparova O, Krupar V (2021) Using principal component analysis to characterize the variability of VLF wave intensities measured by a low-altitude spacecraft and caused by interplanetary shocks. *J Geophys Res Space Phys* 126(5):e2021JA029158
- Budden KG (1961) The Wave-Guide Mode Theory of Wave Propagation. Logos Press, Moscow
- Chrissan DA, Fraser-Smith AC (1996) Seasonal variations of globally measured ELF/VLF radio noise. *Radio Sci* 31(5):1141–1152
- Gjerloev JW (2012) The SuperMAG data processing technique. *J Geophys Res Space Phys* 117:9
- Golden DI, Spasojevic M, Inan US (2009) Diurnal dependence of ELF/VLF hiss and its relation to chorus at $L = 2.4$. *J Geophys Res Space Phys* 114:A5
- Golden DI, Spasojevic M, Inan US (2011) Determination of solar cycle variations of midlatitude ELF/VLF chorus and hiss via automated signal detection. *J Geophys Res Space Phys* 116(A3):A03225
- Graf KL, Lehtinen NG, Spasojevic M, Cohen MB, Marshall RA, Inan US (2013) Analysis of experimentally validated trans-ionospheric attenuation estimates of VLF signals. *J Geophys Res Space Phys* 118(5):2708–2720
- Greifinger C, Greifinger PS (1968) Theory of hydromagnetic propagation in the ionospheric waveguide. *J Geophys Res* 73(23):7473–7490
- Hayakawa M, Bullough K, Kaiser TR (1977) Properties of storm-time magnetospheric VLF emissions as deduced from the Ariel 3 satellite and ground-based observations. *Planet Space Sci* 25(4):353–368
- Helliwell RA (1965) Whistlers and related ionospheric phenomena. Stanford University Press, Stanford
- Jolliffe IT, Cadima J (2016) Principal component analysis: a review and recent developments. *Philos Trans R Soc A* 374(2065):20150202
- Laaspere T, Morgan MG, Johnson WC (1964) Chorus, hiss, and other audio-frequency emissions at stations of the whistlers-east network. *Proc IEEE* 52(11):1331–1349
- Manninen J, Kleimenova NG, Turunen T, Gromova LI (2016) Temporal behaviour of daytime VLF emissions caused by the solar wind and IMF disturbances: A case study. *Sunny Beach, Bulgaria, May 30–June 3, 2016*, page 38
- Martinez-Calderon C, Shiokawa K, Miyoshi Y, Keika K, Ozaki M, Schofield I, Connors M, Kletzing C, Hanzelka M, Santolik O, Kurth WS (2016) ELF/VLF wave propagation at subauroral latitudes: Conjugate observation between the ground and Van Allen Probes A. *J Geophys Res Space Phys* 121(6):5384–5393
- Martinez-Calderon C, Shiokawa K, Miyoshi Y, Ozaki M, Schofield I, Connors M (2015) Statistical study of ELF/VLF emissions at subauroral latitudes in Athabasca, Canada. *J Geophys Res Space Phys* 120(10):8455–8469
- Meredith NP, Horne RB, Iles RHA, Thorne RM, Heynderickx D, Anderson RR (2002) Outer zone relativistic electron acceleration associated with substorm-enhanced whistler mode chorus. *J Geophys Res Space Phys* 107(A7):29
- Newell PT, Gjerloev JW (2011) Evaluation of SuperMAG auroral electrojet indices as indicators of substorms and auroral power. *J Geophys Res* 116:A12
- Newell PT, Gjerloev JW (2011) Substorm and magnetosphere characteristic scales inferred from the SuperMAG auroral electrojet indices. *J Geophys Res Space Phys* 116:A12
- Ripoll J-F, Denton MH, Hartley DP, Reeves GD, Malaspina D, Cunningham GS, Santolik O, Thaller SA, Loridan V, Fennell JF, Turner DL, Kurth WS, Kletzing CA, Henderson MG, Ukhorskiy AY (2021) Scattering by whistler-mode waves during a quiet period perturbed by substorm activity. *J Atmos Sol Terr Phys* 215:105471
- Shiokawa K, Yokoyama Y, Ieda A, Miyoshi Y, Nomura R, Lee S, Sunagawa N, Miyashita Y, Ozaki M, Ishizaka K, Yagitani S, Kataoka R, Tsuchiya F, Schofield I, Connors M (2014) Ground-based ELF/VLF chorus observations at subauroral latitudes-VLF-CHAIN Campaign. *J Geophys Res Space Phys* 119(9):7363–7379
- Simms LE, Engebretson MJ, Clilverd MA, Rodger CJ (2019) Ground-based observations of VLF waves as a proxy for satellite observations: Development of models including the influence of solar illumination and geomagnetic disturbance levels. *J Geophys Res Space Phys* 124(4):2682–2696
- Smith AJ, Horne RB, Meredith NP (2004) Ground observations of chorus following geomagnetic storms. *J Geophys Res Space Phys* 109:A2
- Smith AJ, Horne RB, Meredith NP (2010) The statistics of natural ELF/VLF waves derived from a long continuous set of ground-based observations at high latitude. *J Atmos Terr Phys* 72(5):463–475
- Smith AJ, Jenkins PJ (1998) A survey of natural electromagnetic noise in the frequency range $f = 1–10$ kHz at Halley station, Radio atmospherics from lightning. *Antarctica: 1. J Atmos Terr Phys* 60(2):263–277
- Smith AJ, Meredith NP, O'Brien TP (2004) Differences in ground-observed chorus in geomagnetic storms with and without enhanced relativistic electron fluxes. *J Geophys Res Space Phys* 109:A11
- Sonwalkar VS (1995) Handbook of Atmospheric Electrodynamics, vol. II, chap. Magnetospheric LF-, VLF-, and ELF-Waves
- Spasojevic M (2014) Statistical analysis of ground-based chorus observations during geomagnetic storms. *J Geophys Res Space Phys* 119(10):8299–8317
- Spasojevic M, Inan US (2010) Drivers of chorus in the outer dayside magnetosphere. *J Geophys Res Space Phys* 115:A4
- Summers D, Ma C, Meredith NP, Horne RB, Thorne RM, Anderson RR (2004) Modeling outer-zone relativistic electron response to whistler-mode chorus activity during substorms. *J Atmos Sol Terr Phys* 66(2):133–146
- Suzuki H, Sato N (1987) Seasonal and diurnal variations of ELF emission occurrences at 750-Hz band observed at geomagnetically conjugate stations. *J Geophys Res Space Phys* 92(A6):6153–6158
- Tsurutani BT, Smith EJ (1977) Two types of magnetospheric ELF chorus and their substorm dependencies. *J Geophys Res* 82(32):5112–5128
- Walsh AP, Haaland S, Forsyth C, Keesee AM, Kissinger J, Li K, Runov A, Souček J, Walsh BM, Wing S, Taylor MGGT (2014) Dawn-dusk asymmetries in the coupled solar wind-magnetosphere-ionosphere system: a review. *Ann Geophys* 32(7):705–737
- Yonezu Y, Shiokawa K, Connors M, Ozaki M, Manninen J, Yamagishi H, Okada M (2017) Simultaneous observations of magnetospheric ELF/VLF emissions in Canada, Finland, and Antarctica. *J Geophys Res Space Phys* 122(6):6442–6454

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.