AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Space Physics

RESEARCH ARTICLE

10.1002/2014JA020599

Key Points:

- First time applying the independent component analysis to geomagnetic activity
- Two first IC components give the annual contributions of CMEs and HSSs to GA
- ICA yields estimates of annual values of IMF strength and SW speed

Correspondence to:
L. Holappa,
lauri.holappa@oulu.fi

Citation:
Holappa, L., K. Mursula, and T. Asikainen (2014), A new method to estimate annual solar wind parameters and contributions of different solar wind structures to geomagnetic activity, *J. Geophys. Res. Space Physics*, *119*, 9407–9418, doi:10.1002/2014JA020599.

Received 8 SEP 2014
Accepted 17 NOV 2014
Accepted article online 25 NOV 2014
Published online 18 DEC 2014

A new method to estimate annual solar wind parameters and contributions of different solar wind structures to geomagnetic activity

L. Holappa[1], K. Mursula[1], and T. Asikainen[1]

[1] ReSoLVE Centre of Excellence, Department of Physics, University of Oulu, Oulu, Finland

Abstract In this paper, we study two sets of local geomagnetic indices from 26 stations using the principal component and the independent component (IC) analysis methods. We demonstrate that the annually averaged indices can be accurately represented as linear combinations of two first components with weights systematically depending on latitude. We show that the annual contributions of coronal mass ejections (CMEs) and high-speed streams (HSSs) to geomagnetic activity are highly correlated with the first and second IC. The first and second ICs are also found to be very highly correlated with the strength of the interplanetary magnetic field (IMF) and the solar wind speed, respectively, because solar wind speed is the most important parameter driving geomagnetic activity during HSSs while IMF strength dominates during CMEs. These results help in better understanding the long-term driving of geomagnetic activity and in gaining information about the long-term evolution of solar wind parameters and the different solar wind structures.

1. Introduction

Geomagnetic activity is produced in the interaction between the solar wind and the Earth's magnetic field. It has been studied systematically since the late nineteenth century using different geomagnetic indices. Most common geomagnetic indices are global indices such as *aa*, *Kp*/*Ap*, *Dst*, and *AE*, which are constructed from local indices, e.g., as weighted or normalized averages. For example, the *Kp* index is calculated from local *K* indices of 13 magnetic observatories located at mid-latitudes and subauroral latitudes. Local geomagnetic indices are mainly used to derive global indices, but the differences between local indices are rarely studied. This is surprising since there are over 200 magnetic observatories around the world continuously producing magnetic measurements, but the state of the Earth's magnetic field is often described by just one globally averaged number.

It has been known for a long time that global geomagnetic activity (measured, e.g., by the *aa* index) exhibits a dual peak structure during the solar cycle [*Chapman and Bartels*, 1940; *Newton*, 1948], the first peak during the solar maximum dominated by transient activity and the second peak during the declining phase related to recurrent activity. Later it became clear that the first peak is mainly produced by coronal mass ejections (CMEs) and the second peak mainly by high-speed streams (HSSs) [*Simon and Legrand*, 1986; *Gosling et al.*, 1991]. It is now known that there are significant differences between CME- and HSS-related geomagnetic activities. For example, CMEs are responsible for the largest geomagnetic storms [*Borovsky and Denton*, 2006], while HSSs dominate substorm activity [*Tanskanen et al.*, 2005]. Because of these differences, one can expect that average geomagnetic activity over suitably long time intervals can be decomposed into two components, one related to CME activity and the other related to HSS activity. *Richardson et al.* [2000, 2002] have identified times when CMEs and HSSs were present in the solar wind at 1 AU and studied the contributions of CMEs and HSSs to the *aa* index. They found that during solar maximum most *aa* activity is related to CMEs, while during declining phase and solar minimum most *aa* activity is related to HSSs. *Feynman* [1982] decomposed the annual *aa* index into two components, the "*R*" component being linearly related to the sunspot number and the residual "*I*" component defined as $I = aa - R$. While the *R* component is mainly produced by the CMEs, the *I* component is more closely related to HSSs. This decomposition is reasonable, but it assumes, e.g., that the CME contribution to geomagnetic activity strictly follows the sunspot number, which is poorly valid around solar maxima [*Richardson and Cane*, 2012].

Recently, we used the principal component analysis (PCA) method to extract information on the solar wind drivers of annually averaged geomagnetic activity using a set of local A_h indices [*Holappa et al.*, 2014].

Table 1. Stations and Their Geographic (GG) and Corrected Geomagnetic (CGM) Latitudes and Longitudes[a]

#	Station Name and Code	GG Latitude	GG Longitude	CGM Latitude	CGM Longitude
1	Alibag (ABG)	18.638	72.872	9.52	145.27
2	MBour (MBO)	14.384	−16.967	20.78	56.717
3	Kanoya (KNY)	31.420	130.882	24.17	202.020
4	Kakioka (KAK)	36.233	140.183	28.78	210.93
5	San Juan (SJG)	18.382	−66.118	29.27	5.02
6	Memambetsu (MMB)	43.907	144.193	36.56	214.56
7	Chambon-la-Foret (CLF)	48.017	2.267	43.67	79.94
8	Irkutsk (IRT)	52.167	104.450	46.78	176.67
9	Belsk (BEL)	51.837	20.792	47.41	96.38
10	Niemegk (NGK)	52.072	12.675	47.93	89.65
11	Hartland (HAD)	51.000	−4.483	47.99	75.55
12	Wingst (WNG)	53.743	9.073	50.05	87.31
13	Fredericksburg (FRD)	38.210	−77.367	50.07	356.16
14	Eskdalemuir (ESK)	55.317	−3.200	52.95	78.22
15	Victoria (VIC)	48.517	−123.417	54.04	294.56
16	Nurmijärvi (NUR)	60.508	24.655	56.69	102.78
17	Lerwick (LER)	60.133	−1.183	58.16	82.11
18	Sitka (SIT)	57.052	−135.335	59.82	278.10
19	Meanook (MEA)	54.615	−113.347	62.41	303.72
20	Sodankylä (SOD)	67.367	26.633	63.64	108.17
21	College (CMO)	64.867	−147.860	64.88	261.68
22	Abisko (ABK)	68.358	18.823	65.11	102.91
23	Leirvogur (LRV)	64.183	−21.7	65.46	68.57
24	Fort Churchill (FCC)	58.786	−94.088	69.61	330.03
25	Baker Lake (BLC)	64.333	−96.033	74.59	324.68
26	Thule (THL)	77.483	−69.167	86.00	36.77

[a]Stations are ordered according to their CGM latitudes.

We found that the first principal component (PC1) represents the global average of the A_h indices and correlates almost perfectly with the *Ap* index and that the second principal component (PC2) highly correlates with the annual fraction of high-speed streams in the solar wind. The PCA method, however, does not decompose geomagnetic activity into pure CME and HSS components. For example, the first PC representing global geomagnetic activity is a mixture of CME and HSS effects, which both contribute significantly to global geomagnetic activity [*Richardson and Cane*, 2012].

In this paper we develop the method further and show that the spatiotemporal information included in local indices of geomagnetic activity can be used to extract information about the independent contributions of HSSs and CMEs on geomagnetic activity without any external information about, e.g., solar activity or solar cycle phase. We also use this information to study the contributions of the two main solar wind parameters, the solar wind speed and the interplanetary magnetic field (IMF) intensity, to geomagnetic activity. This paper is organized as follows: Section 2 introduces the A_h and *IHV* (interhourly variability) indices used in this study. In section 3 the principal component analysis (PCA) method that we used earlier [*Holappa et al.*, 2014] is briefly reviewed and applied to A_h and *IHV* indices. The principal components are then processed using the independent component analysis (ICA) method in section 4. The relation of the two first independent components (ICs) to solar wind speed and IMF intensity, as well as to CME and HSS fractions, is discussed in section 5. Finally, conclusions are given in section 6.

2. Local Geomagnetic Indices and Other Data

We use two different measures of local geomagnetic activity: the A_h index [*Mursula and Martini*, 2007] and the *IHV* index [*Svalgaard and Cliver*, 2007]. The 3-hourly A_h index is analogous to *Ak*, the linearized *K* index [*Bartels et al.*, 1939], calculated from hourly data as the range of variation of the local horizontal magnetic field after removing the quiet day (*Sq*) variation. However, the quiet day variation cannot be fully removed from the data by any method and some amount of residual quiet day variation also remains in the A_h indices. In order to exclude the possibility that the residual *Sq* variation affects our results based on the A_h indices we also use *IHV* indices which are calculated using only local night sector data and are thus

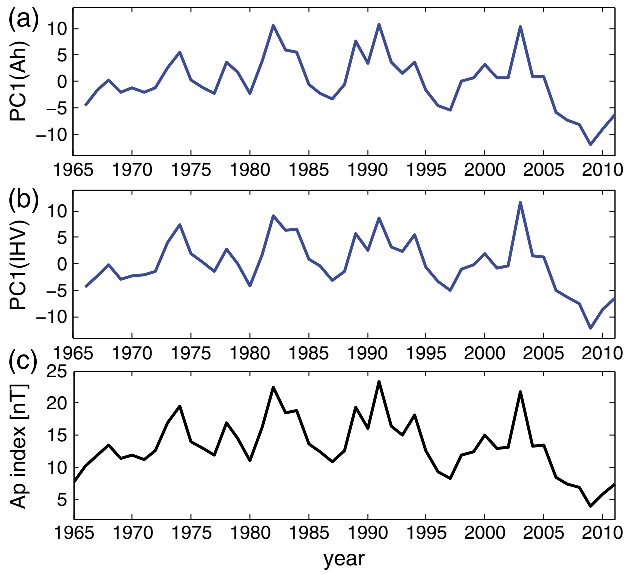

Figure 1. The first principal component of (a) A_{hs} indices and (b) IHV_s indices. (c) The annual averages of the *Ap* index.

practically unaffected by *Sq* variation. The daily *IHV* index is defined as the average of six absolute hourly differences of the local horizontal magnetic field around local midnight [*Svalgaard and Cliver*, 2007].

We use the A_h and the *IHV* indices of the 26 observatories listed in Table 1. The selection criteria for stations was high-quality and long-term continuity of their data sets and good global coverage. We only selected stations which have less than 20% of data missing for any year. We calculated the A_h and *IHV* indices for 1966–2011 (46 years) using hourly mean data obtained from World Data Center of Edinburgh [*World Data Center C1 for Geomagnetism*, 2011]. Before calculating the indices, we checked the baselines and excluded the outliers from the magnetic data by using a three-point median filter (for more details, see *Holappa et al.* [2014]). We also rescaled the A_h and *IHV* indices of the CLF station for years 1966–1971 because CLF recorded spot values instead of hourly means until the end of 1971, leading to excessively large A_h and *IHV* values in these years. For this, we calculated the averages of the ratios A_h(CLF)/A_h(NGK) and *IHV*(CLF)/*IHV*(NGK) in 1972–1981 and in 1962–1971 and multiplied A_h(CLF) and *IHV*(CLF) before 1971 by the corresponding ratios (0.8146 and 0.7736, respectively) so that the A_h(CLF)/A_h(NGK) and *IHV*(CLF)/*IHV*(NGK) ratios became continuous. (Note that NGK and CLF are geographically close to each other, which allows a meaningful comparison between the two stations.)

In addition to the magnetic data of ground stations, we use solar wind data from the OMNI database (http://omniweb.gsfc.nasa.gov/) and the classification of solar wind flow types by *Richardson and Cane* [2012]. There are three different solar wind types identified by *Richardson and Cane* [2012]: CMEs (including the cores of interplanetary CMEs and their related shocks and sheath regions), HSSs (corotating streams from coronal holes), and slow solar wind.

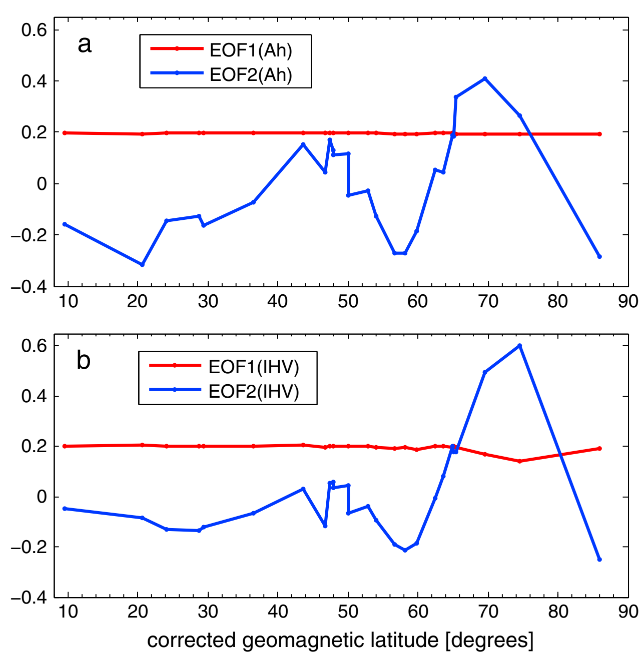

Figure 2. Two first EOFs of (a) the A_{hs} and (b) the IHV_s indices as a function of corrected geomagnetic latitude.

3. Principal Component Analysis Method

Principal component analysis [*Jolliffe*, 2005] is a statistical method, which can be used to represent a large number of correlated variables as linear combinations of a few uncorrelated variables called principal components. Here we apply PCA for annual means (46 years) of geomagnetic indices from 26 observatories. Before evaluating PCA we calculate the standardized annual means for each station separately

$$A_{hs} = \frac{A_h - \langle A_h \rangle}{\sigma}, \tag{1}$$

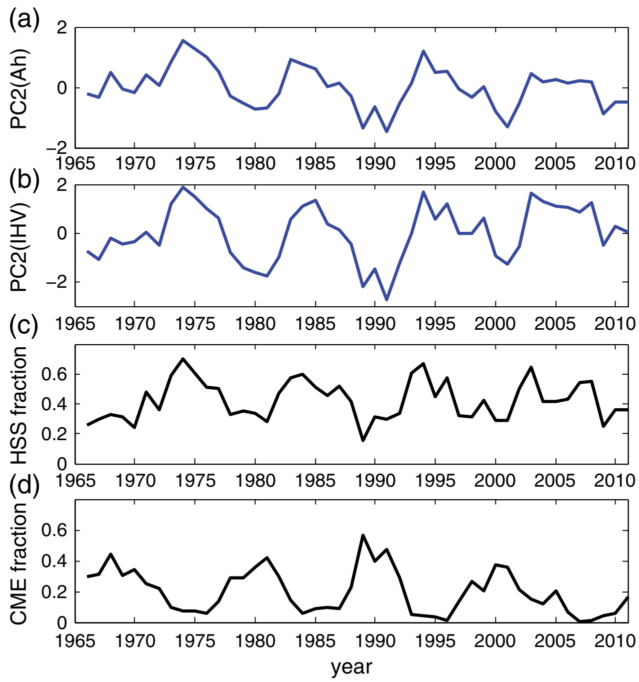

Figure 3. (a, b) The second PC of the A_{hs} and the IHV_s indices. (c, d) Yearly fraction of HSSs and CMEs.

where $\langle A_h \rangle$ is the mean and σ the standard deviation of the annually averaged A_h. We calculate the standardized annual mean IHV_s indices in the same way. Standardized annual means are then collected into the columns of the data matrix X (size 46×26). PCA can be evaluated using the singular value decomposition of the data matrix [see, e.g., *Hannachi et al.*, 2007]

$$X = UDV^T, \tag{2}$$

where U and V are orthogonal matrices ($UU^T = I$ and $VV^T = I$) and $D = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_{26})$ contains the so-called singular values of the matrix X. The column vectors of the 26×26 matrix V are called here the empirical orthogonal functions (EOFs). The principal components are obtained as the column vectors of the 46×26 matrix

$$P = UD. \tag{3}$$

The original variables can then be approximated as a linear combination of the K first principal components with weights given by EOFs as

$$X_{ij} = \sum_{k=1}^{K} P_{ik} V_{jk} \tag{4}$$

where X_{ij} is the value (standardized A_h index) of the jth variable (station) at the observation time (year) i. The variance of the kth PC is proportional to λ_k^2. Hence, the K first PCs include the following percentage

$$\frac{\sum_{k=1}^{K} \lambda_k^2}{\sum_{k=1}^{26} \lambda_k^2} \cdot 100\% \tag{5}$$

of the variance in the original variables.

3.1. The First PC

Figure 1 shows the first principal components of the A_{hs} and IHV_s indices (to be called PC1(Ah) and PC1(IHV)). One can see that there is an excellent agreement between the PC1s of the two indices. The respective EOF1(Ah) and EOF1(IHV) depicted in Figure 2 describe the latitudinal modes associated with the PC1s. As we found earlier [*Holappa et al.*, 2014], EOF1(Ah) is almost flat (independent of latitude), meaning that all stations contribute with roughly equal weights to PC1. Hence, the PC1(Ah) is very closely proportional to the average of the 26 A_{hs} indices. Also, the EOF1(IHV) is almost flat except for a small local minimum at the poleward boundary of the auroral oval (stations #24 and #25).

The PC1(Ah) and the PC1(IHV) correlate almost perfectly with the annual averages of the *Ap* index of the global geomagnetic activity (Pearson correlation coefficients and p values for zero correlation from Student's t test: cc(Ah) = 0.99, $p = 6.4 \cdot 10^{-34}$; cc(IHV) = 0.98, $p = 2.2 \cdot 10^{-31}$) which is also shown in Figure 1. Thus, the PC1(Ah) and PC1(IHV) also closely represent the mean global geomagnetic activity. The PC1(Ah) and PC1(IHV) already explain a large fraction of variance of the A_{hs} (95.6%) and the IHV_s indices (90.1%). Thus, at the annual timescale all stations at different latitudes observe roughly the same (mainly solar cycle related) long-term variation of geomagnetic activity.

10.1002/2014JA020599

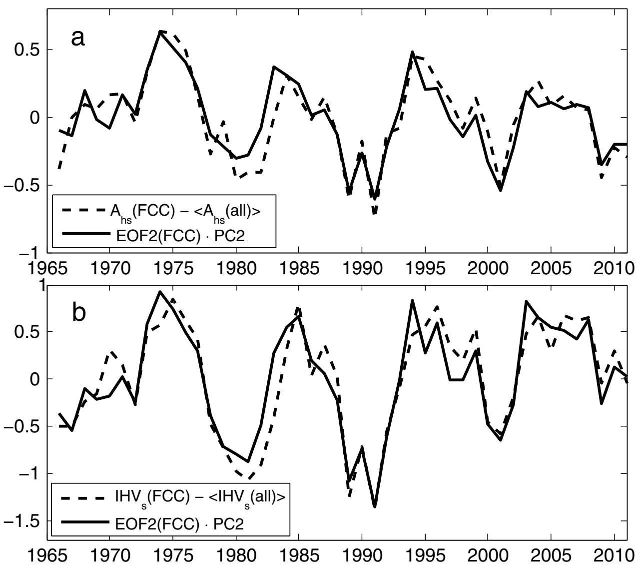

Figure 4. (a) The difference between the A_{hs} index of FCC station and the global average of the A_{hs} indices (solid line); and the second PC scaled by the EOF2 of FCC station (dashed line). (b) The same for IHV_s indices.

3.2. The Second PC

PC2(Ah) and PC2(IHV) are shown in Figure 3 and the associated EOF2(Ah) and EOF2(IHV) in Figure 2. As described above, the first PCs practically represent the annual global averages of the two indices. Therefore, the second PCs describe how these local indices at the individual stations deviate on an average from their global averages. For years of positive PC2(Ah) (PC2(IHV), respectively), the A_{hs} (IHV_s) indices of stations with positive (negative) EOF2 coefficients are higher (lower) than the globally averaged A_{hs} (IHV_s), and vice versa for years of negative PC2 values. This is demonstrated in Figure 4a which shows the difference between the A_{hs} index of FCC station and the average of all 26 A_{hs} indices. For any year, A_h(FCC) is expected to depart from the mean of all A_{hs} indices by PC2(Ah) times the EOF2 coefficient for FCC (EOF2(FCC) = 0.41). The 2nd PC scaled by 0.41 (also shown in Figure 4a) indeed explains the annual differences between the mean A_{hs} and Ah(FCC) very well. Figure 4b shows an analogous difference for *IHV*(FCC). One can see that PC2(IHV) scaled by 0.50 (EOF2(IHV) = 0.50 for FCC) explains the annual differences between *IHV*(FCC) and the global IHV_s very well.

Note that the PC2 only explains 1.8% (4.9%) of the total variance of the A_{hs} indices (IHV_s indices). Therefore, the annual deviations of individual station indices from the global average are not very large especially for stations whose EOF2 coefficients are close to zero (see Figure 2). However, the auroral stations at 65°–75° CGM latitudes (like FCC) with the greatest positive EOF2 coefficients can notably differ from the global average. For example, the absolute difference between IHV_s(FCC) and the global mean of IHV_s indices (Figure 4b) can be more than one (standard deviation), which is a large difference for annual means.

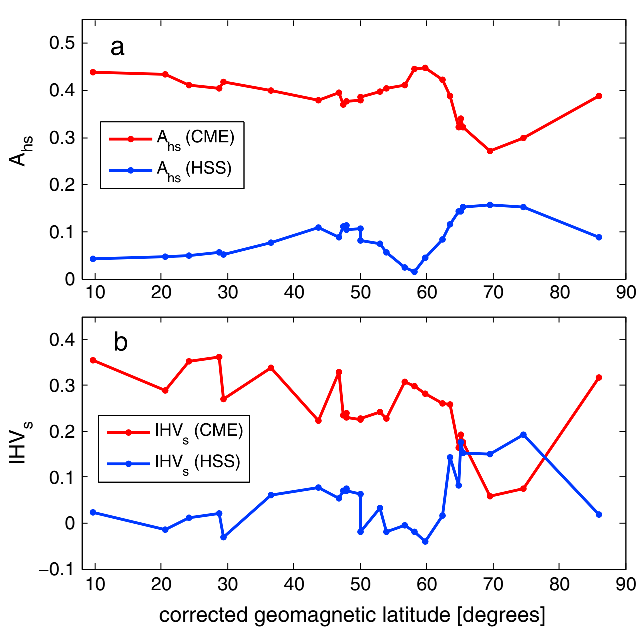

Figure 5. Averages of the (a) A_{hs} and (b) IHV_s indices during CMEs and HSSs.

As noted earlier [*Holappa et al.*, 2014] PC2(Ah) is very highly correlated ($cc = 0.82; p = 4.6 \cdot 10^{-12}$) with the annual time fraction of high-speed streams in solar wind. This can also be seen in Figure 3 which shows the annual fraction of HSSs in solar wind according to the classification of solar wind into three flow types [*Richardson and Cane*, 2012]. Figure 3 also shows the corresponding annual fractions of CMEs which are highly anticorrelated with the HSS fractions. Consequently, PC2(Ah) is anticorrelated with the CME fraction ($cc = -0.67; p = 3.5 \cdot 10^{-7}$). PC2(IHV) is

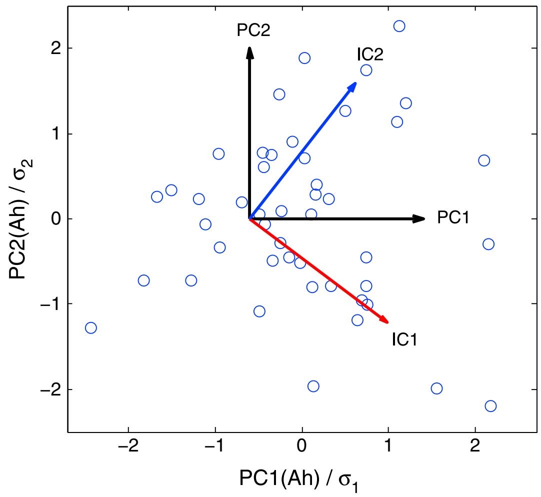

Figure 6. Scatterplot of the standardized first and second PCs of the A_{hs} indices (denoted in black). The red and blue arrows represent the row vectors of the rotation matrix A on which the PCs are projected.

also very highly correlated with the HSS fraction ($cc = 0.79$, $p = 8.2 \cdot 10^{-11}$) and anticorrelated with the CME fraction ($cc = -0.83; p = 9.7 \cdot 10^{-13}$).

Figure 5 shows the averages of the A_{hs} and IHV_s indices during CMEs and HSSs. Averages of the standardized 3-hourly values of A_h indices were calculated over those 3 h periods when only one solar wind type (CME or HSS) was present in the solar wind. Similarly, averages of the standardized daily values of the *IHV* indices were calculated over those local nights when only one solar wind type was present. As seen in Figure 5, there are clear latitudinal patterns in the A_{hs} and IHV_s indices during CMEs and HSSs. One can note the high similarity between the EOF2(Ah) (see Figure 2a) and the distribution of the A_{hs} indices during HSSs (Figure 5a). The distribution of the A_{hs} indices during CMEs is almost the mirror image of the HSS distribution. The IHV_s indices during CMEs and HSSs (Figure 5b) show roughly the same patterns as the corresponding A_{hs} indices. Also, the EOF2(IHV) (see Figure 2b) resembles the EOF2(Ah) (see Figure 2a) and matches with the distribution of the IHV_s indices during HSSs (see Figure 5b).

Because the second PCs of the A_{hs} and IHV_s indices correlate (anticorrelate) with the HSS (CME) fraction and the second EOFs match with the latitudinal distributions of the indices during HSSs (CMEs), one can conclude that PC2 is (mainly) caused by the latitudinally different response of local geomagnetic activity to CMEs and HSSs. Figure 5 shows that during HSSs the strongest values of A_{hs} and IHV_s indices are found at the auroral latitudes (65°–75°), while during CMEs the A_{hs} and IHV_s indices have a (local) maximum at subauroral latitudes (55°–63°). We showed earlier [*Holappa et al.*, 2014] that the relative contribution of HSS-driven substorms maximizes at the auroral latitudes while the relative effect of CME-driven substorms maximizes at subauroral latitudes (where substorms are observed especially during magnetic storms [*Tanskanen et al.*, 2002; *Hoffman et al.*, 2010]), which explains the subauroral minimum and the auroral maximum of EOF2. Since IHV_s indices only measure geomagnetic activity at the night sector, i.e., at the preferred local time sector of substorms, they are more sensitive to substorms (and therefore to HSSs) than the A_{hs} indices. This explains the slightly larger variation of IHV_s (HSSs) between the auroral maximum and the subauroral minimum (see Figure 5b). This also explains why EOF2(IHV) shows a higher auroral maximum than EOF2(Ah) (see Figure 2 and discussion later).

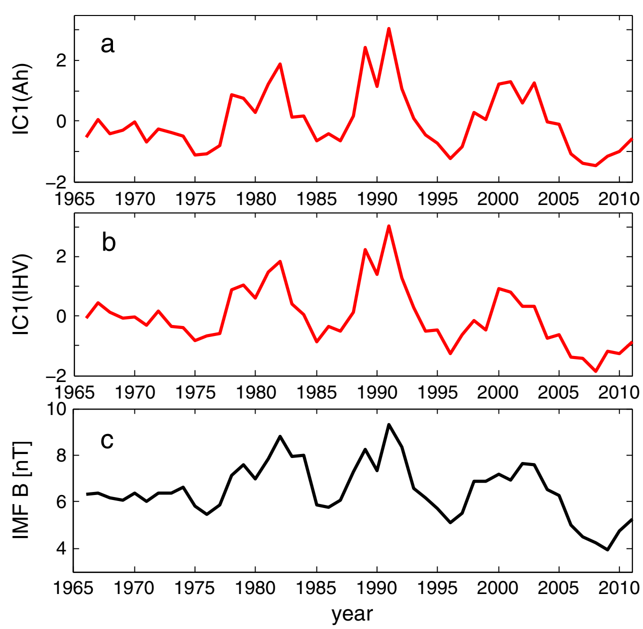

Figure 7. (a, b) The first ICs of the A_{hs} and the IHV_s indices. (c) The annual averages of the IMF strength B.

4. Independent Component Analysis Method

The basic idea of the independent component analysis (ICA) is analogous to that of the principal component analysis:

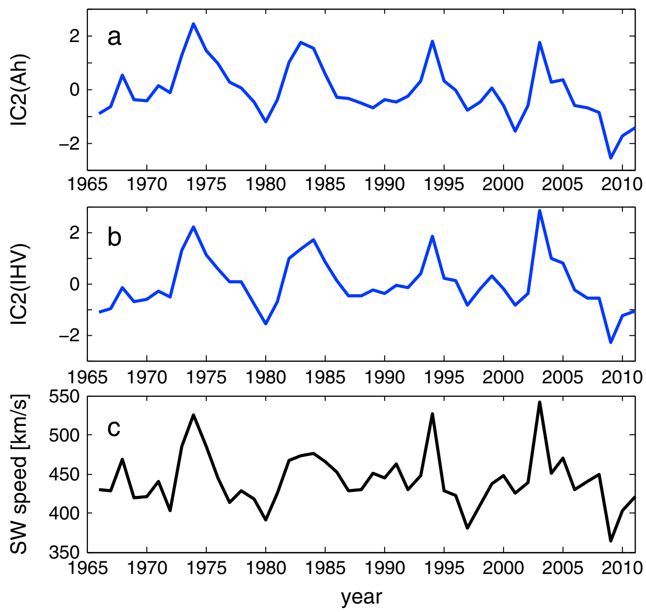

Figure 8. (a, b) The second ICs of the A_{hs} and the IHV_s indices. (c) The annual averages of the solar wind speed.

initially dependent variables are presented as a linear combination of statistically independent components. There are numerous ways to perform ICA [see, e.g., *Hyvärinen et al.*, 2001], but we use here the FastICA software package [*Hyvärinen*, 1999] (http://research.ics.aalto.fi/ica/fastica/).

While the principal components obtained by the PCA method are uncorrelated, they are not necessarily statistically independent. Actually, only if the principal components are Gaussian their uncorrelatedness also guarantees their statistical independence. To see if the two first principal components are independent or not, we first standardize them to unit variance by dividing them by their standard deviations σ_1 and σ_2. Using matrix notation, the standardized PCs are the columns of the matrix

$$P_s = P_2 Z \tag{6}$$

where the 46 × 2 matrix P_2 contains the two first columns of the matrix P of equation (3) and $Z = \mathrm{diag}\left(\sigma_1^{-1}, \sigma_2^{-1}\right)$. Figure 6 shows a scatterplot of the standardized PC1(Ah) and PC2(Ah). If the two PCs were statistically independent, the scatter pattern would be spherically symmetric. Clearly, this is not the case. One can see, e.g., that a positive value of PC1 implies either a large positive or a large negative value of PC2, and a negative value of PC1 implies a small value of PC2. The idea of the IC analysis is to find an orthogonal rotation of the principal components that makes the rotated components statistically as independent as possible. The rotation of the principal components can written as

$$S = AP_s^T, \tag{7}$$

where the orthogonal 2×2 matrix A is the so-called mixing matrix and the rows of 2×46 matrix S contain the independent components (with unit variances). The ICA algorithm finds the matrix A in an iterative process by minimizing the entropies of the independent components. The independent components are maximally non-Gaussian, because the Gaussian distribution has the greatest entropy among all distributions with the same variance.

The principal components are projected onto the basis defined by the row vectors of the matrix A which are shown in Figure 6 as IC1 and IC2. The matrix A calculated for A_{hs} indices performs a clockwise rotation by 37.4°, whence IC1(Ah) = 0.79· PC1$_s$(Ah) −0.61· PC2$_s$(Ah) and IC2(Ah) = 0.61· PC1$_s$(Ah) +0.79· PC2$_s$(Ah). For the IHV_s indices the rotation angle is 52.3°, whence IC1(IHV) = 0.61· PC1$_s$(IHV) −0.79· PC2$_s$(IHV) and IC2(IHV) = 0.79· PC1$_s$(IHV) +0.61· PC2$_s$(IHV).

Using equations (6) and (7) the original data matrix can be approximated as

$$X = P_2 V^T = P_s Z^{-1} V^T = S^T A Z^{-1} V^T, \tag{8}$$

where the row vectors in the matrix $AZ^{-1}V^T$ can be interpreted as the spatial modes (SM) corresponding to the two independent components (in analogous way with the matrix V^T in equation (2)). These spatial modes are obtained by rotation from the EOFs in V, but they are not orthogonal because the matrix AZ^{-1} is not orthogonal due to the different variances of the principal components. Equation (8) is analogous to equation (4) and simply states that the original data can be represented as the following linear combination

$$X_{ij} = \mathrm{IC1}(i) \cdot \mathrm{SM1}(j) + \mathrm{IC2}(i) \cdot \mathrm{SM2}(j), \tag{9}$$

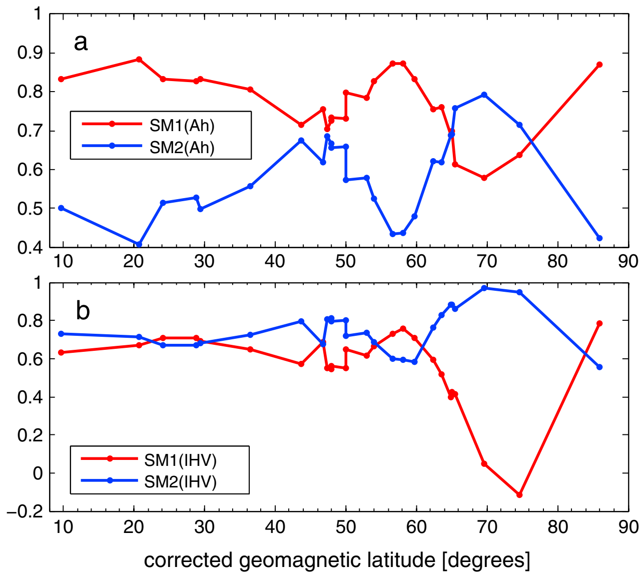

Figure 9. The spatial modes corresponding to the two ICs of (a) A_{hs} and (b) IHV_s indices as functions of corrected geomagnetic latitude.

where IC1(i) and IC2(i) are the two independent components for year i and SM1(j) and SM2(j) are the corresponding spatial mode coefficients for station j.

ICA could also be directly applied to the original data matrix, but the ensuing ICs are not ordered according to decreasing (or increasing) importance (fraction of total variance) and thereby do not reflect the physically most important processes. Rather, in this case, ICA tends to emphasize spikes in the data, which are highly non-Gaussian, but misses the physically relevant patterns. Instead, reducing first the dimension of the data by including only the two leading PCs in the ICA makes the two ICs also to include a large fraction (95%) of variance and the important physics.

4.1. The First and Second IC

Figures 7 and 8 show the first and second independent components for the two indices, respectively. One can see that the ICs of the two indices are very similar with each other, as expected from the similarity of the two first PCs of these indices. The correlations between the ICs of the two indices are very high: cc(IC1(Ah), IC1(IHV)) = 0.95, $p = 4.9 \cdot 10^{-23}$ and cc(IC2(Ah), IC2(IHV)) = 0.94, $p = 1.1 \cdot 10^{-21}$.

The spatial modes corresponding to the two ICs are depicted in Figure 9. One can see that the two spatial modes are almost mirror images of each other for both indices, especially for A_{hs}. However, the first spatial mode of *IHV* shows a very deep minimum at auroral latitudes, which is also related to the dip in EOF1(IHV) (Figure 2). Note also that the SM2(IHV) is generally larger than SM1(IHV). This means that the second IC has, on the average, a higher weight in the *IHV* indices than the first IC. This is opposite to A_h indices for which the first IC is dominating.

5. Relation to Solar Wind and IMF

Annual averages of the IMF intensity B and the solar wind speed v are plotted in Figures 7c and 8c, respectively. One can see that IC1(Ah) and IC1(IHV) are very highly correlated with the IMF intensity B with cc(IC1(Ah), B) = 0.90; $p = 4.2 \cdot 10^{-17}$ and cc(IC1(IHV), B) = 0.85; $p = 1.8 \cdot 10^{-12}$. The second ICs are, in turn, very highly correlated with the solar wind speed v: cc(IC2(Ah), v) = 0.82; $p = 4.7 \cdot 10^{-13}$ and cc(IC2(IHV), v) = 0.89; $p = 5.0 \cdot 10^{-16}$, or alternatively with v^2: cc(IC2(Ah), v^2) = 0.81; $p = 6.2 \cdot 10^{-12}$ and cc(IC2(IHV), v^2) = 0.89; $p = 2.8 \cdot 10^{-16}$. These correlations and the above ICA results expressed in equation (9) suggest that the annual averages of all local geomagnetic indices can be represented as a linear combination of the annual solar wind speed and the IMF strength with their own optimum relative weights for these two drivers.

Before presenting the results we note that, of course, it is not physically reasonable that momentary geomagnetic activity should depend on a linear combination of B and v (or v^2). Rather, the relation between geomagnetic activity and solar wind parameters is usually expressed in terms of different nonlinear coupling functions, e.g., Bv^2. There are also many coupling functions involving, e.g., solar wind density and IMF vector orientation, but at the annual timescale they do not correlate any better with global geomagnetic activity than the simple function Bv^2 [*Finch and Lockwood*, 2007]. The above ICA results and the earlier results regarding the nonlinear solar wind coupling functions can be understood as follows. During CMEs the coupling function Bv^2 is mainly enhanced above the mean value due to large values of B, with v remaining at the average level, while during HSSs the high values of Bv^2 are due to persistently high values of v, with B attaining average values [*Richardson et al.*, 2002; *Richardson and Cane*, 2012].

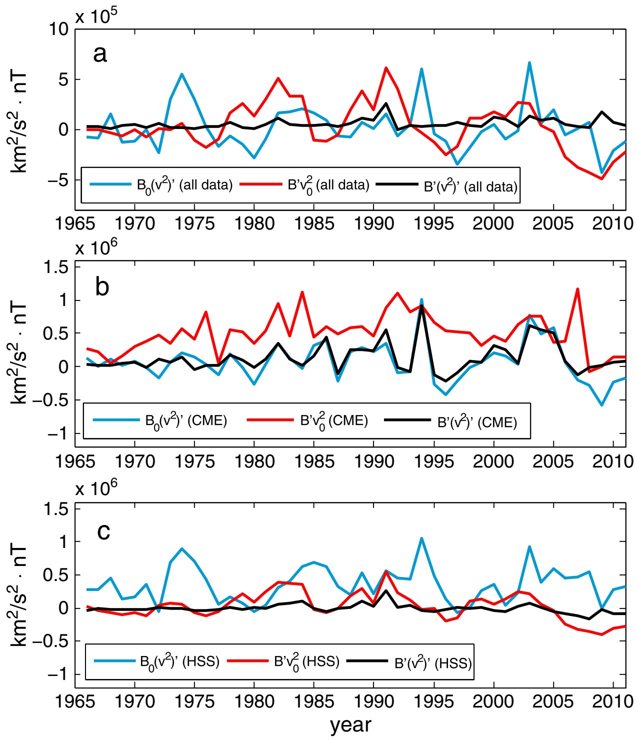

Figure 10. Annual averages of the three time-dependent terms in equation (10) contributing to the coupling function Bv^2 during (a) all times, (b) CME, and (c) HSS intervals.

To further test this hypothesis, we decompose hourly B and v^2 values into constant and fluctuating parts: $B = B_0 + B'$ and $v^2 = v_0^2 + (v^2)'$, where B_0 and v_0^2 denote the averages of B and v^2 in 1966–2011. Now we can write

$$Bv^2 = B_0 v_0^2 + B' v_0^2 + B_0 (v^2)' + B'(v^2)'. \quad (10)$$

The first term on the right-hand side determines the average value of the coupling function over the 46 year period ($B_0 v_0^2 = 1.3 \cdot 10^6$ nT·km^2/s^2, $B_0 = 6.4$ nT, $v_0 = 439$km/s), which, however, does not affect, e.g., the correlation between the coupling function and geomagnetic activity. Figure 10a shows the annual averages of the three last time-dependent terms on the right-hand side of equation (10) including all solar wind data. One can see that the third term $B'(v^2)'$ is overall rather small, suggesting that the fluctuations B' and $(v^2)'$ (and, in fact, also B and v^2) are rather uncorrelated. This also leads to the fact that, at the annual timescale, the functional form of the coupling function Bv^2 can indeed be effectively represented as a linear combination of B and v^2. Hence, at the annual timescale, geomagnetic activity has two components, one correlated with the IMF strength and the other with the solar wind speed. Both fluctuating terms in Figure 10a have approximately the same range of variation meaning that B and v^2 contribute to the variations of the coupling function Bv^2 roughly equally.

Figures 10b and 10c show the annual averages of the three time-dependent terms of equation (10) during CMEs and HSSs, respectively. One can see that the term $B' v_0^2$ clearly dominates over the two other terms during CMEs, while the term $B_0(v^2)'$ dominates during HSSs. Therefore, the IMF strength is indeed the dominant parameter driving global geomagnetic activity during CMEs, while the solar wind speed dominates during HSSs. Interestingly, in 1994 and 2003 all three terms are high during CMEs indicating that in these years CMEs carried strong magnetic fields and were very fast.

5.1. Relation to CMEs and HSSs

The ICA spatial modes in Figure 9 have quite similar latitudinal patterns as the average distributions of the A_{hs} and IHV_s indices during CMEs and HSSs depicted in Figure 5. This suggests that the IC1(Ah) and IC1(IHV) represent the CME contributions to these indices, while the second ICs represent the HSS contributions. The first ICs correlate well with the CME fraction (cc(IC1(Ah)) = 0.76, $p = 6.3 \cdot 10^{-10}$; cc(IC1(IHV)) = 0.81, $p = 5.6 \cdot 10^{-12}$) and the second ICs with the HSS fraction (cc(IC2(Ah)) = 0.73, $p = 7.8 \cdot 10^{-9}$; cc(IC2(IHV)) = 0.74, $p = 4.5 \cdot 10^{-9}$). However, it is not physical that the annual fractions of CMEs and HSSs in solar wind should determine the yearly levels of geomagnetic activity because the properties of CMEs and HSSs evolve from 1 year to another. For example, as shown clearly in Figure 10, the speeds and magnetic field strengths of CMEs and HSSs are different in different years. To take the varying properties of CMEs and HSSs into account, we estimate the CME and HSS contributions to global geomagnetic activity by calculating the quantities

$$C = \langle Ap \rangle_{\mathrm{CME}} \cdot f_{\mathrm{CME}} \quad (11)$$

$$H = \langle Ap \rangle_{\mathrm{HSS}} \cdot f_{\mathrm{HSS}}, \quad (12)$$

where $\langle Ap \rangle_{\mathrm{CME}}$ ($\langle Ap \rangle_{\mathrm{HSS}}$) is the annual average of the Ap index values observed during CMEs (HSSs) and f_{CME} (f_{HSS}) is the annual fraction of CMEs (HSSs) in the solar wind and plotting them in Figure 11. As expected, the

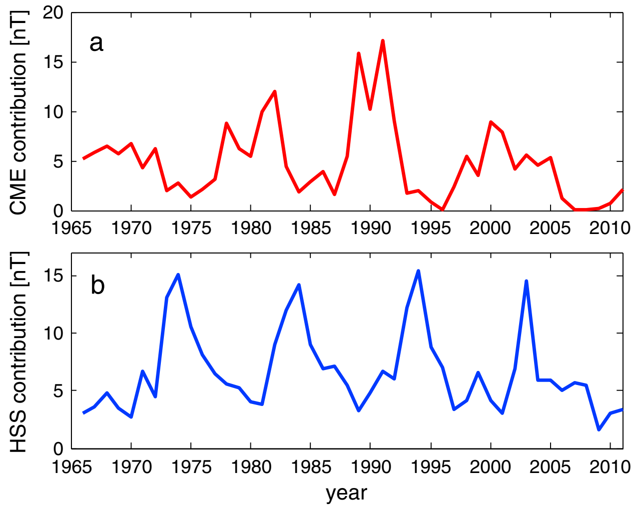

Figure 11. Annual (a) CME and (b) HSS contributions to the *Ap* index.

first ICs (see Figure 7) are very highly correlated with the CME contribution (cc(IC1(Ah)) = 0.92, $p = 5.7 \cdot 10^{-19}$; cc(IC1(IHV)) = 0.93, $p = 3.3 \cdot 10^{-20}$) and the second ICs (see Figure 8) with the HSS contribution (cc(IC2(Ah)) = 0.88, $p = 4.7 \cdot 10^{-16}$; cc(IC2(IHV)) = 0.90, $p = 4.9 \cdot 10^{-17}$). This gives strong evidence that the first and second ICs indeed represent the contribution of CMEs and HSSs, respectively, to geomagnetic activity. There are some small differences, e.g., between the second ICs and the HSS contribution, especially in 1989, when the HSS contribution shows a deep minimum but the second ICs only a shallow minimum. These differences are most likely related to the numerous gaps in the solar wind satellite measurements in 1980s and early 1990s, causing larger inaccuracy in solar wind classification and in the annual CME and HSS fractions at those times [*Richardson and Cane*, 2012].

Since the two ICs represent the CME and HSS contributions to geomagnetic activity, the corresponding IC spatial modes quantify the weights by which CMEs and HSSs contribute to the local geomagnetic activity at the different stations. Although the spatial modes of A_{hs} and IHV_s indices (see Figure 9) have a fairly similar latitudinal variation, the SM2(IHV) is at considerably higher level than SM2(Ah), indicating that the relative contribution of HSSs is, on an average, greater to the IHV_s indices than to the A_{hs} indices. Furthermore, the first spatial mode of IHV_s shows a very deep minimum at the poleward edge of the auroral oval, meaning that CMEs have a very small contribution to the IHV_s indices at these latitudes where geomagnetic activity is dominated by HSS-driven substorm activity in the night sector [*Tanskanen et al.*, 2005, 2011]. This is also consistent with the results by *Finch et al.* [2008] who showed that correlation between geomagnetic activity and solar wind speed maximizes in the night sector at auroral latitudes. On the other hand, the A_{hs} indices measure all local times and are thus not solely dominated by substorms even at auroral latitudes, which decreases the relative importance of HSSs in the A_{hs} indices. Because of the strong dominance of HSSs, the IHV_s indices at auroral latitudes have a slightly higher EOF2 and a smaller EOF1 (see Figure 2), as discussed in section 3.2.

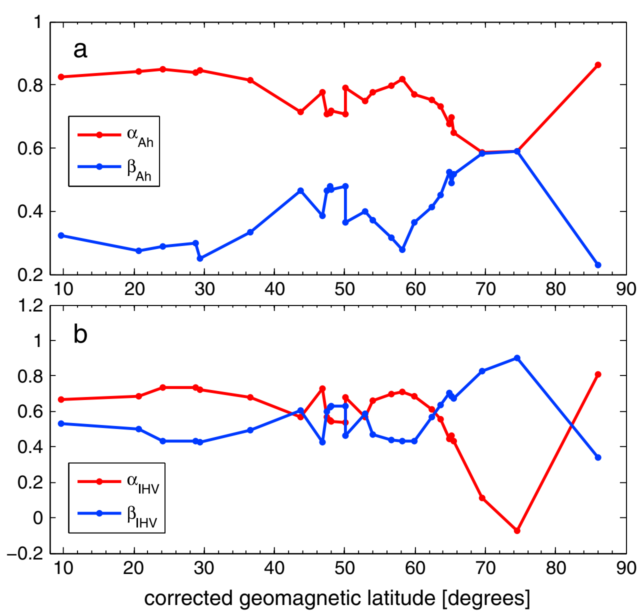

Figure 12. Least squares fit coefficients α and β (equations (13) and (14)) for (a) the A_{hs} indices and (b) the IHV_s indices.

In order to exclude the possibility that the spatial modes obtained by the independent component analysis are artifacts of the method, we have fitted coefficients α and β for A_{hs} and IHV_s indices of different stations so that

$$A_{hs} = \alpha_{Ah} B_s + \beta_{Ah} v_s^2 \quad (13)$$

$$IHV_s = \alpha_{IHV} B_s + \beta_{IHV} v_s^2, \quad (14)$$

where B_s and v_s^2 are standardized IMF strength and squared solar wind speed, respectively. The coefficients $\alpha_{Ah}(\alpha_{IHV})$ and $\beta_{Ah}(\beta_{IHV})$ are solved using the standard least squares fitting method. As seen in Figure 12, coefficients of equation (13) have the same latitudinal variation as the ICA spatial modes (equation (8)). Thus, the coefficients α_{Ah} (α_{IHV}) and β_{Ah} (β_{IHV}) obtained from the least squares fits are very similar with the first and second spatial mode coefficients of A_{hs} (IHV_s) indices, respectively. The only systematic difference is that the β_{Ah} coefficients are somewhat smaller than the coefficients of SM2(Ah). The fact that the least squares fit calculated using the measured solar wind data produces very similar results with the ICA (blind to solar wind data) gives great confidence on the results based on ICA method and their interpretation.

6. Conclusions

In this paper we have studied the spatiotemporal evolution of geomagnetic activity in 1966–2011 using local A_h and *IHV* indices of 26 stations covering a wide range of latitudes. We analyzed the indices using the principal component analysis method and confirmed that our recent results for the A_h indices [*Holappa et al.*, 2014] also hold for *IHV* indices, i.e., that the first PC describes global average geomagnetic activity and the second PC the deviations from the global average caused by high-speed streams.

We used the independent component analysis method to rotate the two first PCs into two independent components (ICs). The spatial modes of the two ICs clearly correspond to the distribution of the indices during CMEs (first mode) and HSSs (second mode). The two first ICs were found to match very well with the CME and HSS contributions to global geomagnetic activity. We also found that the first IC and the second IC correlate very highly with the IMF strength and the solar wind speed, respectively. This is due to the fact that high values of the IMF strength mainly dominate the (larger than average) driving of geomagnetic activity during CMEs, while high solar wind speed dominates the driving during HSSs.

We found essentially similar results both for A_h, which include all local times and for *IHV* indices, which only include the night sector. This shows that the residual *Sq* variation in the A_h indices has no major effect to the main results. It is also very reassuring that the same results can be found using indices which define geomagnetic activity quite differently: the A_h being a traditional range index, the *IHV* index using hourly absolute differences. Despite all these differences between the two indices, the PC and IC methods are able to find essentially the same information about the solar wind drivers. The combined PC/IC method presented here offers a new way to gain information about the relative occurrence of CMEs and HSSs and the long-term properties of solar wind, in particular the IMF strength and the solar wind speed. This improves our understanding of the long-term evolution of solar wind and the long-term driving of geomagnetic activity by the different solar wind structures.

Acknowledgments
We acknowledge the financial support by the Academy of Finland to the ReSoLVE Centre of Excellence (project 272157) and to project 257403. This work has benefited for collaborations and contacts within the COST ES1005 (TOSCA) Network Action (especially Working Group 2). The hourly magnetometer data were obtained from the World Data Center for Geomagnetism, Edinburgh (http://www.wdc.bgs.ac.uk/). The *Ap* index was obtained from World Data Center for Geomagnetism, Kyoto (http://http://wdc.kugi.kyoto-u.ac.jp/). The hourly values of the solar wind speed were obtained from the OMNI database (http://omniweb.gsfc.nasa.gov/). The list of the solar wind structures can be obtained by contacting Ian G. Richardson.

Yuming Wang thanks the reviewers for their assistance in evaluating this paper.

References

Bartels, J., N. H. Heck, and H. F. Johnston (1939), The three-hour-range index measuring geomagnetic activity, *Terr. Magn. Atmos. Electr.*, *44*, 411, doi:10.1029/TE044i004p00411.

Borovsky, J. E., and M. H. Denton (2006), Differences between CME-driven storms and CIR-driven storms, *J. Geophys. Res.*, *111*, A07S08, doi:10.1029/2005JA011447.

Chapman, S., and J. Bartels (1940), *Geomagnetism*, vol. 1, chap. 11, Oxford Univ. Press, New York.

Feynman, J. (1982), Geomagnetic and solar wind cycles: 1900–1975, *J. Geophys. Res.*, *87*(A8), 6153–6162.

Finch, I., and M. Lockwood (2007), Solar wind-magnetosphere coupling functions on timescales of 1 day to 1 year, *Ann. Geophys.*, *25*, 495–506, doi:10.5194/angeo-25-495-2007.

Finch, I. D., M. L. Lockwood, and A. P. Rouillard (2008), Effects of solar wind magnetosphere coupling recorded at different geomagnetic latitudes: Separation of directly-driven and storage/release systems, *Geophys. Res. Lett.*, *35*, L21105, doi:10.1029/2008GL035399.

Gosling, J. T., D. J. McComas, J. L. Phillips, and S. J. Bame (1991), Geomagnetic activity associated with Earth passage of interplanetary shock disturbances and coronal mass ejections, *J. Geophys. Res.*, *96*, 7831–7839, doi:10.1029/91JA00316.

Hannachi, A., I. T. Jolliffe, and D. B. Stephenson (2007), Empirical orthogonal functions and related techniques in atmospheric science: A review, *Int. J. Climatol.*, *27*, 1119–1152, doi:10.1002/joc.1499.

Hoffman, R. A., J. W. Gjerloev, L. A. Frank, and J. W. Sigwarth (2010), Are there optical differences between storm-time substorms and isolated substorms?, *Ann. Geophys.*, *28*(5), 1183–1198, doi:10.5194/angeo-28-1183-2010.

Holappa, L., K. Mursula, T. Asikainen, and I. G. Richardson (2014), Annual fractions of high-speed streams from principal component analysis of local geomagnetic activity, *J. Geophys. Res. Space Physics*, *119*, 4544–4555, doi:10.1002/2014JA019958.

Hyvärinen, A. (1999), Fast and robust fixed-point algorithms for independent component analysis, *IEEE Trans. Neural Networks*, *10*(3), 626–634.

Hyvärinen, A., J. Karhunen, and E. Oja (2001), *Independent Component Analysis*, 1st ed., Wiley-Interscience, New York.

Jolliffe, I. (2005), Principal component analysis, John Wiley, Hoboken, N. J., doi:10.1002/0470013192.bsa501.

Mursula, K., and D. Martini (2007), A new verifiable measure of centennial geomagnetic activity: Modifying the K index method for hourly data, *Geophys. Res. Lett.*, *342*, L22107, doi:10.1029/2007GL031123.

Newton, H. W. (1948), Sudden commencements in the Greenwich magnetic records (1879–1944) and related sunspot data, *Mon. Not. R. Astron. Soc.*, *5*, 159–185.

Richardson, I. G., and H. V. Cane (2012), Near-Earth solar wind flows and related geomagnetic activity during more than four solar cycles (1963–2011), *J. Space Weather Space Clim.*, *2*(26), A02, doi:10.1051/swsc/2012003.
Richardson, I. G., E. W. Cliver, and H. V. Cane (2000), Sources of geomagnetic activity over the solar cycle: Relative importance of coronal mass ejections, high-speed streams, and slow solar wind, *J. Geophys. Res.*, *105*, 18,203–18,214, doi:10.1029/1999JA000400.
Richardson, I. G., H. V. Cane, and E. W. Cliver (2002), Sources of geomagnetic activity during nearly three solar cycles (1972–2000), *J. Geophys. Res.*, *107*(A8), SSH 8-1–SSH 8-13, doi:10.1029/2001JA000504.
Simon, P. A., and J. P. Legrand (1986), Some solar cycle phenomena related to the geomagnetic activity from 1868 to 1980. II. High velocity wind streams and cyclical behaviour of poloidal field, *Astron. Astrophys.*, *155*, 227–236.
Svalgaard, L., and E. W. Cliver (2007), Interhourly variability index of geomagnetic activity and its use in deriving the long-term variation of solar wind speed, *J. Geophys. Res.*, *112*, A10111, doi:10.1029/2007JA012437.
Tanskanen, E., T. I. Pulkkinen, H. E. J. Koskinen, and J. A. Slavin (2002), Substorm energy budget during low and high solar activity: 1997 and 1999 compared, *J. Geophys. Res.*, *107*(A6), SMP 15-1–SMP 15-11, doi:10.1029/2001JA900153.
Tanskanen, E. I., J. A. Slavin, A. J. Tanskanen, A. Viljanen, T. I. Pulkkinen, H. E. J. HKoskinen, A. Pulkkinen, and J. Eastwood (2005), Magnetospheric substorms are strongly modulated by interplanetary high-speed streams, *Geophys. Res. Lett.*, *32*, L16104, doi:10.1029/2005GL023318.
Tanskanen, E. I., T. I. Pulkkinen, A. Viljanen, K. Mursula, N. Partamies, and J. A. Slavin (2011), From space weather toward space climate time scales: Substorm analysis from 1993 to 2008, *J. Geophys. Res.*, *116*, A00I34, doi:10.1029/2010JA015788.
World Data Center C1 for Geomagnetism (2011), *World Data Center for Geomagnetism*, Edinburgh, U. K. [Available at http://www.wdc.bgs.ac.uk/.]