



# Principal Component Analysis of Ground Level Enhancement of Cosmic Ray Events

R. E. Ugwoke<sup>1,2</sup>, A. A. Ubachukwu<sup>2</sup>, J. O. Urama<sup>2</sup>, O. Okike<sup>3</sup>, J. A. Alhassan<sup>2</sup>, and A. E. Chukwude<sup>2</sup>

<sup>1</sup>Department of Physics, Federal University of Technology Owerri, Imo State, Nigeria; [romanusejike1971@gmail.com](mailto:romanusejike1971@gmail.com), [romanus.ugwoke@futo.edu.ng](mailto:romanus.ugwoke@futo.edu.ng)

<sup>2</sup>Department of Physics and Astronomy, University of Nigeria, Nsukka, Nigeria

<sup>3</sup>Department of Industrial Physics, Ebonyi State University, Abakaliki, Nigeria

Received 2022 August 14; revised 2023 January 13; accepted 2023 February 6; published 2023 March 10

## Abstract

We applied principal component analysis (PCA) to the study of five ground level enhancements (GLEs) of cosmic ray (CR) events. The nature of the multivariate data involved makes PCA a useful tool for this study. A subroutine program written and implemented in the R software environment generated interesting principal components. Analysis of the results shows that the method can distinguish between neutron monitors (NMs) that observed Forbush decreases from those that observed GLEs at the same time. The PCA equally assigned NMs with identical signal counts with the same correlation factor ( $r$ ) and those with close  $r$  values equally have a close resemblance in their CR counts. The results further indicate that while NMs that have the same time of peak may not have the same  $r$ , most NMs that had the same  $r$  also had the same time of peak. Analyzing the second principal components yielded information on the differences between NMs having opposite but the same or close values of  $r$ . NMs that had the same  $r$  equally had the tendency of being close in latitude.

*Key words:* Sun: particle emission – Sun: flares – (Sun:) solar-terrestrial relations – (ISM:) cosmic rays

## 1. Introduction

Cosmic ray (CR) studies have been carried out by a lot of researchers who sought an understanding of their origins, acceleration mechanisms, time profiles and their interactions in the heliosphere, magnetosphere as well as atmosphere. Their effects both on geomagnetic storms and lightning have provided a lot of insights that enable accurate predictions of weather and weather forecasts necessary for air transports, and for space explorations using satellites and cosmonauts. The time profile of CRs which either indicates a sudden (or sharp) drop in their intensity on arrival or a rapid increase in their intensity is always studied along with other geomagnetic indices. There have been cases of a concurrent sudden increase in intensity with time called ground level enhancement (GLE) in some neutron monitor (NM) stations along with a sharp decrease in intensity with time also known as Forbush decrease (FD) in other stations for a given event.

Researchers have employed different methods in their analysis of both FD and GLE but we do not know of any application of principal component analysis (PCA) in the study of GLE. PCA has however been used extensively in other fields of science and engineering, as a tool for investigation.

The technique of reducing the dimension of a multivariate data set such that the variance of the data set is maximized is called PCA (e.g., Faleriro et al. 2004; Okike & Collier 2011; Tharwat 2015; Jolliffe & Cadima 2016; Kassambara 2017). The hidden structures in the entire data are the principal

components (PCs) (Hayden 2018) and all the PCs are orthogonal (e.g., Jolliffe 2002; Okike & Collier 2011). The PCs retain almost all the information in the original data set (Dray 2008; Okike & Collier 2011; Jolliffe & Cadima 2016). The first PC (PC1) has the highest variance (e.g., Okike & Collier 2011; Forkman et al. 2019) and therefore contains most of the information in the original data set.

PCA is done either by the singular value decomposition method or by the method of utilizing the covariance matrix (Moore 1981; Muller et al. 2006; Dray 2008; Tharwat et al. 2015). The first two are sufficient for interpretation if their percentage variance is up to 80% (Okike & Collier 2011).

Identification of cosmic TeV gamma ray protons during extensive air showers was carried out by Faleriro et al. (2004) by the method of PCA. They did this by applying the PCA on the two-dimensional particle density fluctuations, which according to their report provided a decreasing sequence of covariance matrix eigenvalues with features that are not the same for different primary CRs. Muller et al. (2006) improved on the combination of PCA with visualization method. They were able to control all steps of the visualization pipeline using PCA results, thus removing the limitation of making the combination a preprocessing step of dimension reduction.

Okike & Collier (2011) carried out a multivariate study of FD simultaneity and were able to use PCA to discriminate between globally simultaneous and non-simultaneous FDs. Their work suggests that the structure of the intensity time

profile of FDs measured at different locations appears to be similar if there are strong positive correlations between raw data and PC1. Ear recognition using block-based PCA along with decision fusion by Tharwat (2015) achieved faster recognition performance compared to previous methods.

Zuska et al. (2019) also applied PCA to their assessment of the impacts of meteorological elements on the concentration of particulate matter (PM10). They were able to distinguish three PCs and how they affected PM10 in the four seasons of the year. Riadigos et al. (2020) applied PCA methods to the CR data of a newly installed muon detector to remove the atmospheric temperature effect on the data. The PC regression they applied captured at least 77% of the variability in the data due to the effect of the temperature of the air layers at the site of the detector.

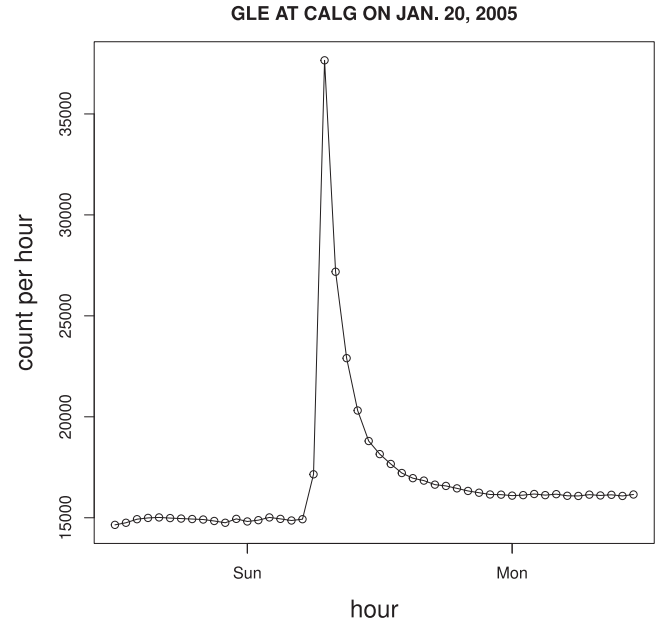
Okike & Alhassan (2022) investigated the relations between automatically selected FD, worldwide lightning frequency, sunspot number and other solar-terrestrial drivers using PCA. They were able to show the level of correlation of FD with these parameters. The authors also compared the result of their FD catalog with that of Pushkov Institute of Terrestrial Magnetism, Ionosphere, and Radio Wave Propagation, Russian Academy of Sciences (IZMIRAN) with PCA. They were able to identify similarities and differences between the FD correlations in the two catalogs with the said solar/geophysical parameters.

In the present submission, we present a PCA analysis of the GLEs from 1978 May 7; 1989 September 29; 1989 October 22; 2001 April 15 and 2005 January 20. These events are also known as GLEs 31, 42, 44, 60 and 69 respectively in McCracken et al. (2012). Given the numerous applications of PCA to scientific investigations, as have been cited above, we strongly believe that the application of the method to the analysis of GLE multivariate data will not only yield interesting results but open doors to further research on GLE. Figure 1(a) shows the time profile of a typical GLE event that occurred on 2005 January 20, which was observed at the Calgary (CALG) NM. The figure affirms that the Galactic cosmic ray (GCR) count before the GLE was stable and that the GLE commenced with a small increase in CR count and thereafter rose to peak count. The decay was equally quasi-exponential. Figure 1(b) is another example; this time FD occurred the same day as the GLE and was observed by the NM at Fort Smith (FSMT).

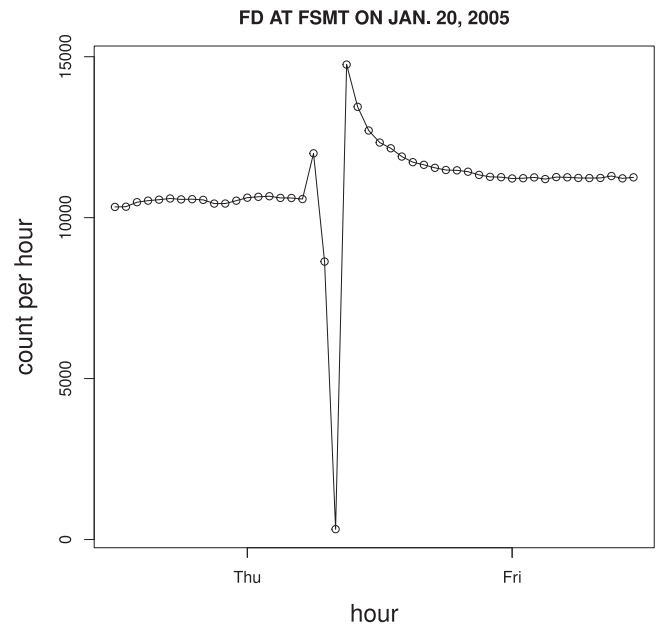
### 1.1. Theory of Principal Component Analysis

The under-listed equations by Jolliffe & Cadima (2016) succinctly describe the theory of PCA applied in this investigation. For an  $n \times p$  data matrix  $X$ , there are  $n$ -dimensional vectors  $x_1, x_2, \dots, x_p$ , with  $p$  numerical variables. The  $j$ th column is the vector  $x_j$  of observations on the  $j$ th variable.

Let  $X_a$  be a linear combination of the columns of the matrix having maximum variance such that



(a) GLE event of January 20, 2005 at CALG.



(b) FD observed by FSMT during the GLE of January 20, 2005.

**Figure 1.** (a) GLE event of 2005 January 20 at CALG. (b) FD was observed by FSMT during the GLE on 2005 January 20.

$$Xa = \sum_{j=1}^p a_j x_j, \quad (1)$$

$a$  is a vector of constants  $a_1, a_2, \dots, a_p$ . The variance is given by

$$\text{var}(Xa) = a'Sa, \quad (2)$$

where  $'$  is the transpose and  $S$  is the covariance matrix associated with  $Xa$ . Jolliffe (2002) and Jolliffe & Cadima (2016) recommended the normalization required for the solution to be

$$a'a = 1. \quad (3)$$

This ensures that

$$\text{var}(Xa) = a'Sa = \lambda a'a = \lambda, \quad (4)$$

where  $a$  is the eigenvector and  $\lambda$  is the eigenvalue. For linear combinations of uncorrelated  $Xa_k$ , having maximum variance,

$$Xa_k = \sum_{j=1}^p a_{jk}x_j. \quad (5)$$

This means that as long as  $k' \neq k$ ,

$$a_k'Sa_k = \lambda_k a_k'a_k = 0. \quad (6)$$

$Xa_k$  represents the PCs, while the elements of the eigenvector  $a_k$  are the PC loadings. In this work, the PC loadings are the correlation of the PC with the raw data. The elements of the linear combination  $Xa_k$  are the PC scores which in this work are the time series signal of the CR counts.

It has been a common practice to use mean-centered variables ( $x_j^*$ ) in PCA (Jolliffe 2002; Okike & Collier 2011; Jolliffe & Cadima 2016) where

$$x_j^* = x_{ij} - \bar{x}_j, \quad (7)$$

and  $\bar{x}_j$  is the mean value of the variable  $j$ . To avoid problems associated with the units of all the  $p$  variables which may not be the same, Jolliffe & Cadima (2016) recommend that in addition to mean centering of the variables,  $x_{ij}$  is divided by the standard deviation of the  $n$  observations of variable  $j$ . Thus

$$Z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}, \quad (8)$$

where  $s_j$  is the standard deviation. The matrix  $X$  is therefore replaced by  $Z_{ij}$  and  $Xa_k$  is replaced by  $Za_k$  so that

$$Za_k = \sum_{j=1}^p a_{jk}z_j. \quad (9)$$

This last equation results in PCs that do not change for any linear transformation of the units. In this case too, the normalization used is

$$\tilde{a}_k'\tilde{a}_k = \lambda_K, \quad K = 1, 2, \dots, p, \quad (10)$$

and not that in Equation (3). With this normalization, the coefficient of correlation ( $r$ ) between the  $k^{\text{th}}$  PC and the  $j^{\text{th}}$  variable is given as

$$r_{\text{var}}, \text{PC}_k = \sqrt{\lambda_k a_{jk}}. \quad (11)$$

Thus with the normalization  $\tilde{a}_k'\tilde{a}_k = \lambda_K$  instead of  $a'a = 1$ , the coefficient of the new loading  $\tilde{a}_k$  is the correlation between each original variable and the  $k^{\text{th}}$  PC.

## 2. Data and Method of Data Analysis

In the World-Wide Neutron Monitor Network <http://cro.izmiran.ru/common/links.htm> there are catalogs of raw CR data covering decades of CR observations. The website which is rich in pressure-corrected hourly and minute data is hosted by IZMIRAN. We selected composite hourly data from this website regarding NM stations spread across all latitudes.

Nineteen and twenty-three NMs that had composite data were studied in the 1978 May 7 and 1989 October 22 events respectively. In the 1989 September 29 event, we selected 27 NM stations. There were 28 NMs selected for the 2001 April 15 event while for the 2005 January 20 event, 22 NM stations were selected for study. In all the events under study, 48 hours of observation (12 hours before, 24 hours on the day of the event and 12 hours after the day of the event) were considered.

We employed AWK programming to convert the data frame from the above website to a data frame readable in an R software environment. R is software used in data science mainly for statistical analysis. Many packages are embedded in it and it can be obtained from <https://cran.R-project.org>. Venables et al. (2022) described R as an integrated suite of software facilities for data manipulation, calculation and graphical display.

Identifying a hidden pattern or information in a multivariate data set with correlated variables is the main reason for PCA of such data (e.g., Jolliffe 2002; Okike & Collier 2011; Kassambara 2017). We applied dimensionality reduction of the data set to variables of lower dimensions (PCs) through the following steps as described in Sheoran et al. (2020) as outlined below:

1. Centering the data set,
2. Computing the covariance and correlation matrix,
3. Computation of the eigenvectors and eigenvalues,
4. Computing the PC scores and loadings.

These PCs are orthogonal and carry nearly all the information from the original data set with PC1 having the maximum variance followed by PC2 up to the last one (e.g., Jolliffe 2002; Okike & Collier 2011; Kassambara 2017). The interpretation is based on the correlation of the original data set with the signal, see Okike & Collier (2011), since variables with common correlation values share some common information. A subroutine program we wrote and implemented in the R software environment generated the PC loadings and scores shown in Figures 1–10 and we used them for the interpretations.

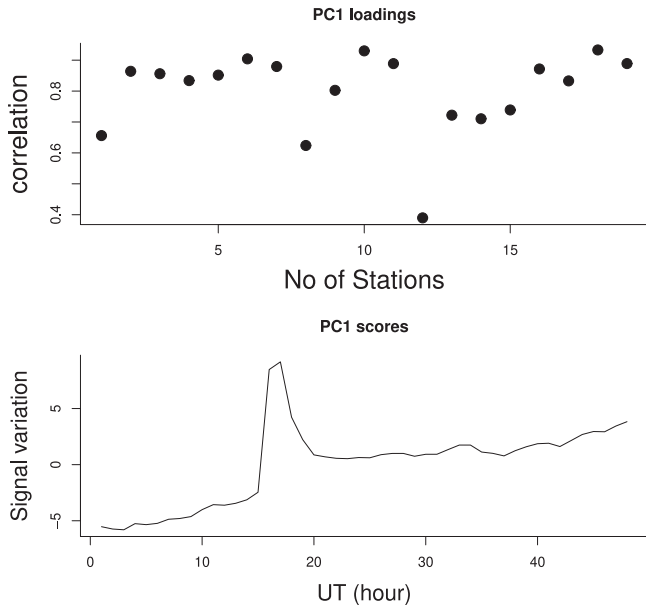


Figure 2. PC1 for GLE of 1978 May 7 event.

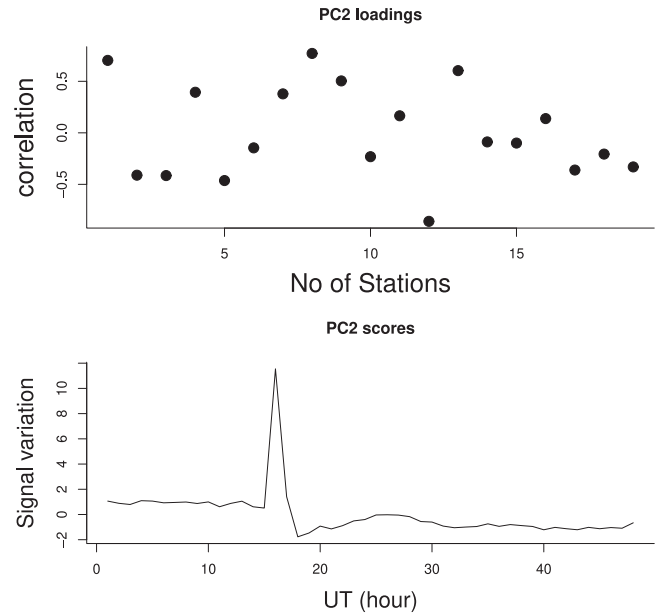


Figure 3. PC2 for GLE of 1978 May 7 event.

### 3. Results and Discussion

#### 3.1. Interpretation of the Correlation of the PCs with the Raw Data

The PCA results are presented in Figures 2–11. The upper and lower panels of each of these diagrams stand for the correlation coefficients of the PCs with the raw data and the variation patterns of the PCs respectively. The y-axes of the upper and lower panels are, thus, labeled “correlation” and “Signal variations” respectively. The physical interpretation/representation of the signal variations indicated in the lower panels depends on the values of the correlation coefficient ( $r$ ) in the upper panels of each of the diagrams. Strong signals will register high values of  $r$  (statistically significant) whereas weak signals will be associated with small or non-significant  $r$ . With regard to CR intensity variation which is the subject of the current work, PCs associated with large values of  $r$  might be global GLEs whereas those with smaller values of  $r$  are representative of local or station dependent GLEs. These global GLEs are seen at the same time by all the NMs irrespective of their locations. The signal variation plotted on the lower panel is the form or the average variation pattern of CR raw data observed by the NM stations at various locations on the Earth. Oh et al. (2008) and Okike & Collier (2011) would refer to such events as globally simultaneous CR variations. The PCA theory and some details of the mathematical relationship between the correlation of the PCs and the raw data are presented in Section 1.1.

As suggested by Okike & Collier (2011), another useful statistic, besides the correlation coefficient, used to decide the

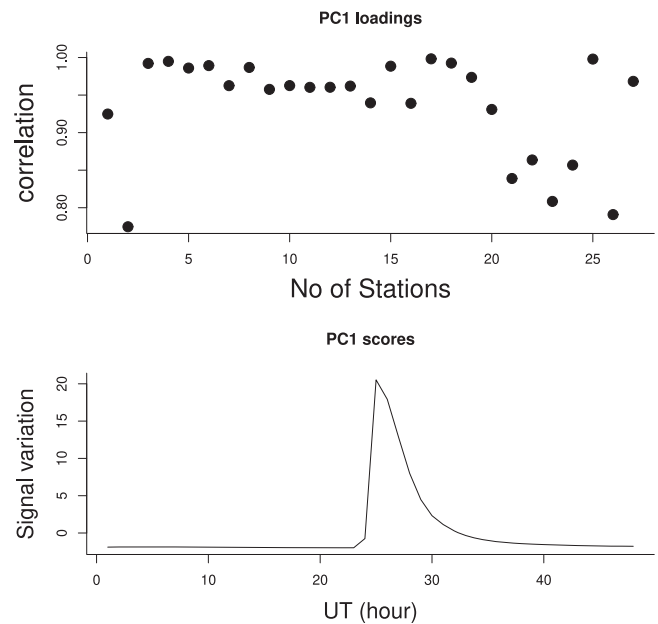


Figure 4. PC1 for GLE of 1989 September 29 event.

simultaneity of an event is the percentage variance. PC1 is usually the one with the highest variance and, as such, contains most of the information in the original data (Okike & Collier 2011; Kassambara 2017; Okike & Alhassan 2022). Any event whose PC1 is associated with about 80% variance would be taken as a strong and simultaneous GLE. For easy reference, the values of  $r$  associated with each PC1 of the five events analyzed here are presented in Tables 1, 3, 5, 7 and 9

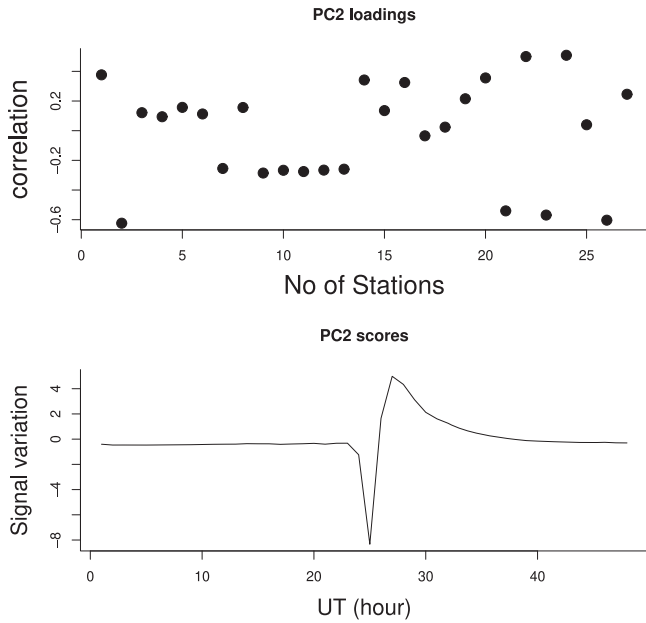


Figure 5. PC2 for GLE of 1989 September 29 event.

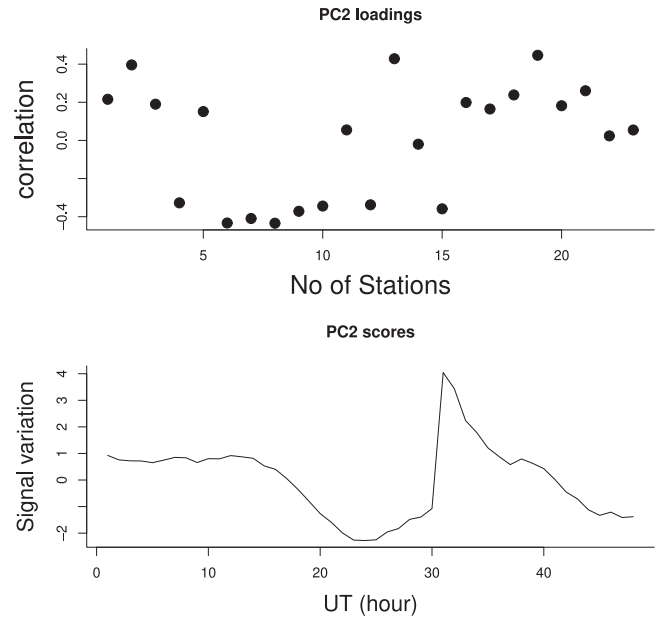


Figure 7. PC2 for GLE of 1989 October 22 event.

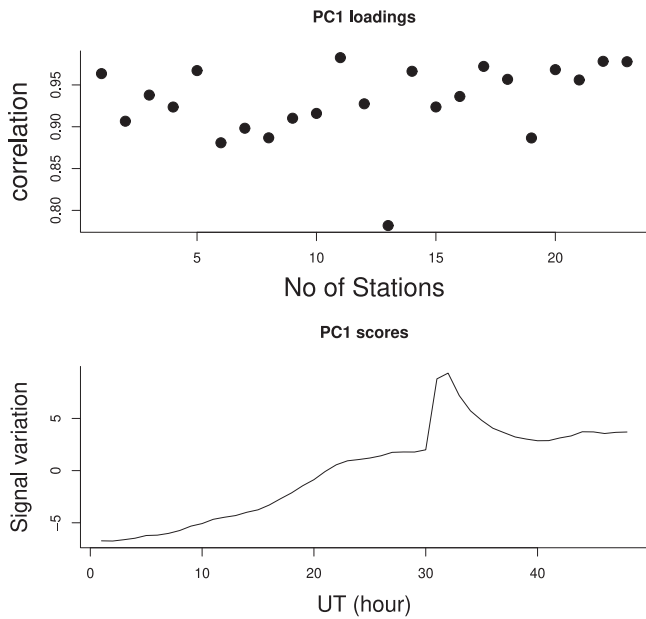


Figure 6. PC1 for GLE of 1989 October 22 event.

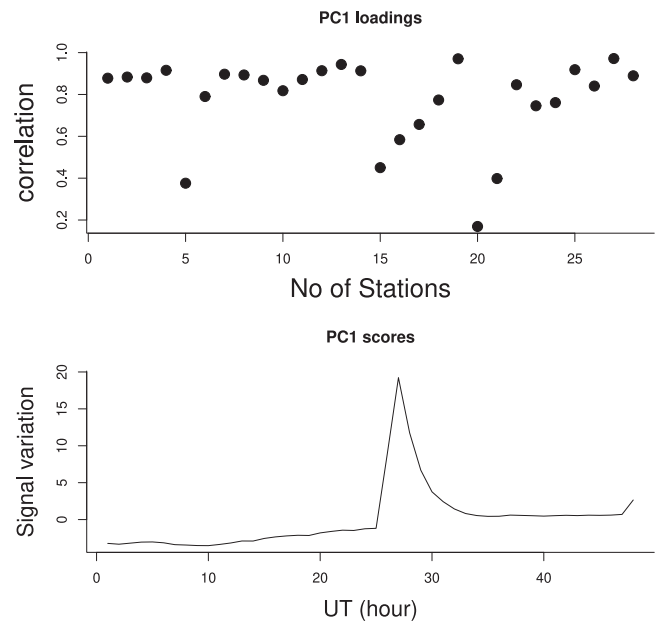


Figure 8. PC1 for GLE of 2001 April 15 event.

whereas the variances associated with all the PCs for each of the events are presented in Tables 2, 4, 6, 8 and 10.

### 3.2. Correlation of the First Principal Component (PC1) with Raw Data from 1978 May 7 GLE (GLE 31)

The PC1 for the event on 1978 May 7 is presented in Figure 2 and the quantitative results are reported in Tables 1 and 2. The total variance associated with PC1 is 65.52. This

implies that PC1 accounts for 65.52% intensity variation in the raw data at the time of this event. This value of variance associated with PC1 is less than the 80% benchmark of Okike & Collier (2011). The event may not, therefore, be regarded as globally simultaneous. The variance reported in Table 2 shows that about the first three PCs (65.52% + 19.58% + 0.07998% = 85.18%) are required to account for the CR intensity variation for this event

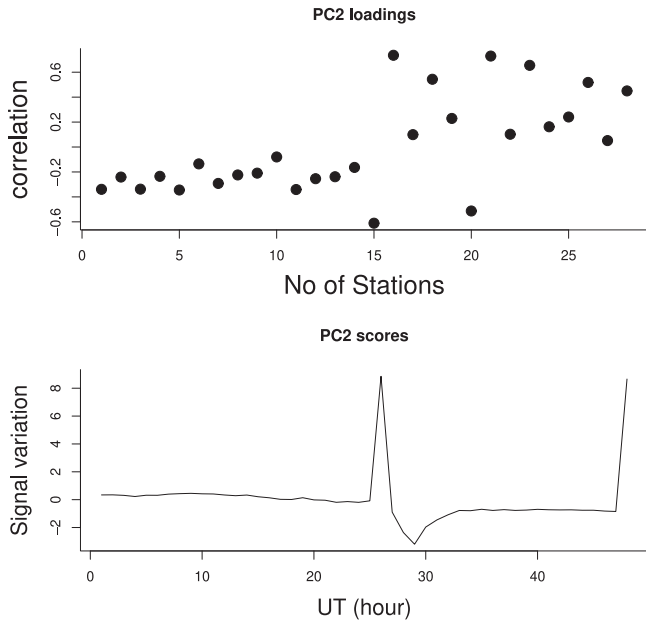


Figure 9. PC2 for GLE of 2001 April 15 event.

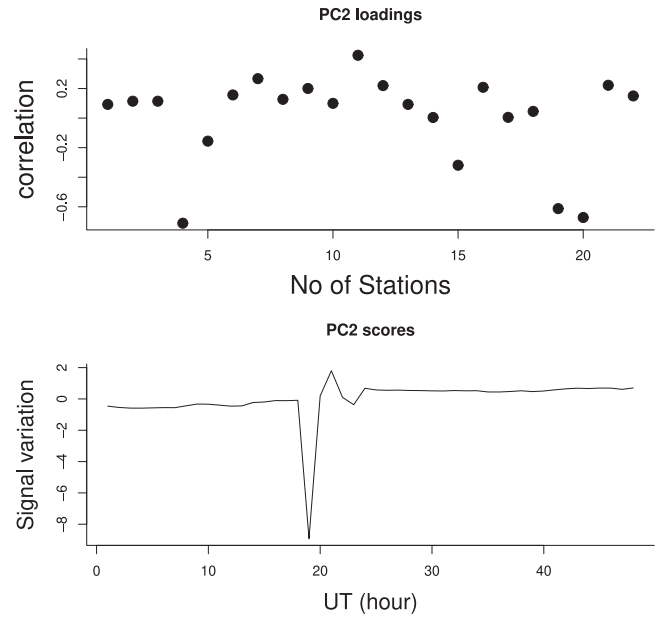


Figure 11. PC2 for GLE of 2005 January 20 event.

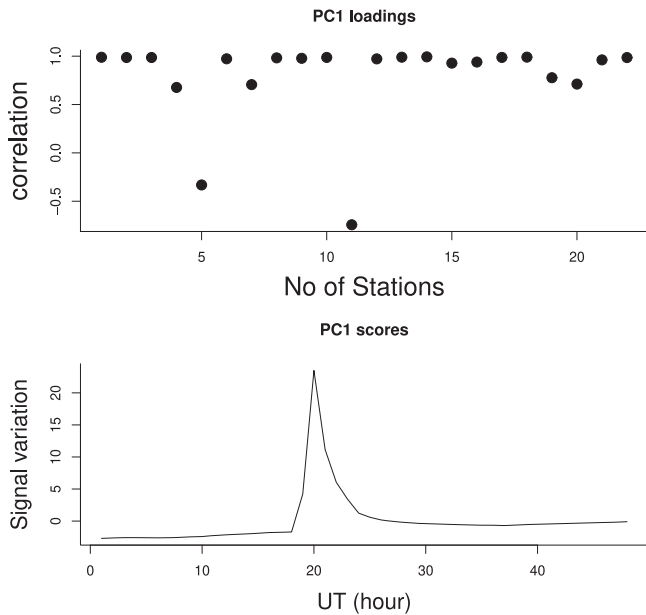


Figure 10. PC1 for GLE of 2005 January 20 event.

at all the stations. The contributions of each of the remaining 16 PCs in decreasing order (1.48%, 1.20%, 1.0%, 0.74%, 0.61%, 0.53%, 0.3%, 0.3%, 0.2%, 0.2%, 0%) are far less and may safely be regarded as noise. The general trend as seen from the signal in PC1 of this event is that the majority of the NMs had a gradual rise and fall in CR counts (in other words small FDs) before the main day of the event and never had an instant rise to peak level count. Many of them had pre-peak counts that are not evenly spaced.

Table 1

Full Names of the Places Where the NMs for the 1978 May 7 Event (GLE 31) are Located, their Short Names and the  $r$  Values of PC1

SN	NM Full Name	NM Short Name	$r$
1	APATITY	APTY	0.65
2	CALGARY	CALG	0.86
3	GOOSE BAY	GSBY	0.85
4	HERMANUS	HRMS	0.83
5	INUVIK	INVK	0.85
6	IRKUTSK	IRKT	0.90
7	JUNGFRAUJOCH	JUN1	0.88
8	KIEL	KIEL	0.60
9	LOMNICKY STIT	LMKS	0.80
10	MAGADAN	MGDN	0.92
11	MT WASHINGTON	MTWS	0.89
12	NOVOSIBIRSK	NVBK	0.40
13	OULU	OULU	0.70
14	POTCHEFSTROOM	PTFM	0.69
15	ROME	ROME	0.72
16	SANAE	SNAE	0.89
17	SOUTH POLE	SOPO	0.83
18	TIXIE BAY	TXBY	0.92
19	YAKUTSK	YKTK	0.90

Their decay also never reversed to the original background GCR counts.

Individual plots of CALG, HRMS, INVK and SOPO (plots not shown) which had  $r$  values of 0.86, 0.83, 0.85 and 0.83 respectively affirm that they all had enhancements in CR count before the peak. Their peaks occurred at 4:00 UT, 3:00 UT, 4:00 UT and 4:00 UT respectively. No instant jump to the peak



**Table 2**  
PCA Summary of the 1978 May 7 (GLE 31) Event

SN	Principal Component	Standard Deviation	Proportion of Variance	Cumulative Proportion
1	PC1	3.5282	0.6552	0.6552
2	PC2	1.9285	0.1958	0.8509
3	PC3	1.232 77	0.079 98	0.930 90
4	PC4	0.5304	0.0148	0.9457
5	PC5	0.477 65	0.012 01	0.957 71
6	PC6	0.436 08	0.010 01	0.967 72
7	PC7	0.373 67	0.007 35	0.975 07
8	PC8	0.341 56	0.006 14	0.981 21
9	PC9	0.315 74	0.005 25	0.986 46
10	PC10	0.236 67	0.002 95	0.989 40
11	PC11	0.2267	0.0027	0.9921
12	PC12	0.202 79	0.002 16	0.994 27
13	PC13	0.1950	0.0020	0.9963
14	PC14	0.145 01	0.001 11	0.997 38
15	PC15	0.132 99	0.000 93	0.998 31
16	PC16	0.119 15	0.000 75	0.999 06
17	PC17	0.104 77	0.000 58	0.999 64
18	PC18	0.072 19	0.000 27	0.999 91
19	PC19	0.040 99	0.000 09	1.000 00

was observed in the GLE event in these stations. The initial stage of the event however varied in them. All of them also confirm that the enhancement occurred after some series of FDs. Their decay was not all that quasi-exponential and they could not decay down to the initial GCR count before the GLE.

IRKT and YKTK had a common  $r$  value (0.90) and the same time of peak (4:00 UT). Both of them had quite odd CR counting characterized by rising and falling CR counts up to the peak and also during their decay. It is not only that they could not decay to the original GCR count but the rising and falling CR count continued after the decay. Their profile is much more similar than any other NM. However, TXBY and MGDN, with  $r=0.92$ , had the closest profile to the two of them.

SNAE and MTWS had the same value of  $r$  (0.89) and peak time (4:00 UT). Their profiles are also quite similar. In each of them, there was evidence of an FD before the GLE and they both had pre-peak counts. Whereas they had a quasi-exponential decay, they did not decay down to their previous GCR count before the event. GSBY, with  $r=0.85$ , had an initial gradual rise in CR count after which it jumped to the peak count at 4:00 UT. Its decay was gradual and not necessarily quasi-exponential. After the decay, it established a new level of GCR count that is much higher than what it was before the GLE.

In this event too, JUN1 had  $r=0.88$ . It had a more rapid rise to peak count than the NMs with  $r=0.89$  and equally a faster decay than MTWS. While SNAE and MTWS had a pre-peak count, JUN1 did not. JUN1 had peak at 03:00 UT.

PTFM and ROME, having  $r=0.69$  and  $0.72$  respectively, are almost the same in their GLE CR count. They also had multiple small FDs before the little enhancement. Their enhancement was quite minimal compared to the rest so far considered. In two of them, the count in CRs decayed instantly to a value close to the last part of the recovery phase of the precursory FD and started rising and falling again in the same pattern before the GLE. LMKS and OULU, with  $r=0.80$  and  $0.70$  respectively, also had instant rise to peak count. The two NMs recorded small FD before the GLE. The decay in LMKS was more rapid than others in this event. After about two decay counts, it came down to the GCR level. On the other hand, the decay in OULU was also two steps down to the GCR count but its first decay was too close to the peak count.

APTY and KIEL which had instant rise to peak count also had close  $r$  values equal to about  $0.60$  and  $0.65$  respectively. In them, the GCR count was relatively flat (without an FD) before the GLE event. The decay to GCR count which was almost instantaneous occurred the same way in them.

The count in NVBK was unexpected. Though it had  $r=0.40$  and should have been very much like APTY, it was a complete FD before a small enhancement. Being a deviation from the rest, it should have a negative correlation like those in GLE 69 that had FD when others were having GLE. Perhaps there may be an error in the data entry of the archive because the count before the peak in NVBK is a four-digit figure versus all others in the station have five-digit figures. It is also seen that except for JUN1, all the NMs with  $r$  less than  $0.85$  had a peak at 03:00 UT but the peak of NVBK occurred at 04:00 UT.

### 3.3. Correlation of the Second Principal Component (PC2) With Raw Data from 1978 May 7 GLE (GLE 31)

PC2 (Figure 3) of 1978 May 7 accounts for 19.58% of the information in the raw data (see Table 2). The relevant information needed in this PC is that their  $r$  values have counterparts such that their  $r$  signs are opposite and are equal or close to being equal.

APTY, KIEL and OULU had  $r=0.55$ ,  $0.55$  and  $0.51$  respectively. In PC1 they equally shared close values of  $r$ . All of them had an instant jump to a high peak count. Their decay equally took the same pattern. NVBK on the other hand had  $r=-0.58$  and an FD at the time others had enhancement in CR counts. Though it had positive  $r$  in PC1, it was the least  $r$  because of this departure from the trend seen in others.

CALG, GSBY and INVK in this PC2 had  $r=-0.45$  while SOPO and YKTK had  $r=-0.40$ . When these two groups are compared with those having  $r=0.40$  (JUN1) and  $0.50$  (LMKS), it is seen that each of these groups already had close values of  $r$  in PC1. The group with positive  $r$  had instant jump to peak count and almost dropped to the GCR count after two decay counts. On the other hand, those with negative  $r$  values did not instantaneously jump to peak count. In these latter

groups of NMs, the CR count also took several decay steps down to the GCR counts.

PTFM and ROME had  $r = -0.2$  and also had similar CR counts at their rising and decay phases. Both of them differed from SNAE and MTWS which had  $r = 0.20$  in having instant decay to almost the GCR count. SNAE and MTWS had gradual decay counts.

HRMS which had  $r = 0.40$  recorded an instant rise to peak count and almost decayed to the GCR count instantly, while MGDN with  $r = -0.30$  had several intermediate counts before the peak and also decayed gradually to GCR counts. In this PC2, having an opposite sign and being equal or almost equal suggests that a feature exists in them that is dissimilar.

### 3.4. Correlation of the First Principal Component (PC1) with Raw Data in the 1989 September 29 Event (GLE 42)

Figure 4 represents PC1 of the 1989 September 29 GLE event. The results of the GLE event are quantitatively shown in Tables 3 and 4. It is interesting to note that the correlation coefficient,  $r$ , of each station with PC1 is very high ( $>0.7$ , Table 3). Table 4 equally affirms that the variance associated with PC1 is very high (0.8825), higher than the 80% benchmark of Okike & Collier (2011). This is an indication that the event is very strong and globally observed by all the NMs no matter where they are located on Earth. The strong nature of this event as suggested by PC1 is in perfect agreement with the description of Moraal & Caballero-Lopez (2014).

The PC loading is shown as a signal representing how each of the signals from the individual NMs would appear. PC1 in this event assigned the same  $r$  to NMs having the same signal pattern. In general, the decay is quasi-exponential, though some tended to be more closely related than others.

Table 3 shows the correlation of PC1 with raw data. The values are approximated by values derived from the PC loading and score in Figure 3. In this table, all the NMs had high correlation coefficients, suggesting that their individual signals for the CR count are almost the same as the signal in Figure 3.

Individual plots of the NMs (plots not shown here) show that for those with  $r$  values from 0.97 to 1.0, their CR counts, and both the GCR and GLE counts up to the peak, are quite similar with little differences. THUL, CAPS and MGDN had  $r = 1.0$  and are all in the high-latitude Northern Hemisphere. They equally had the same profile and peaked at 12:00 UT. They differed from others that had  $r$  from 0.99 in having the first decay count that is too close to the peak. Similar to this group, with  $r = 1.0$ , are CALG, KIEL and NVBK (three of which had  $r = 0.99$ ). This group too had a pre-peak count that is too close to the peak value and in the upper mid-latitude in the Northern Hemisphere.

The  $r$  values of GSBY, INVK, TXBY and NWRK are 0.98, 0.98, 0.97 and 0.97 respectively. Only GSBY had a peak at 12:00 UT while the other three had theirs at 13:00 UT. All of

**Table 3**

Full Names of the Places where the NMs for 1989 September 29 Event (GLE42) are Located, their Short names and Their Correlation Coefficients ( $r$  values) of PC1

SN	NM Full Name	NM Short Name	$r$
1	APATITY	APTY	0.92
2	BEIJING	BJNG	0.73
3	CALGARY	CALG	0.99
4	CAPE SCHMIDT	CAPS	1.00
5	DEEP RIVER	DPRV	0.98
6	GOOSE BAY	GSBY	0.98
7	HERMANUS	HRMS	0.95
8	INUVIK	INVK	0.98
9	IRKUTSK2	IRK2	0.95
10	IRKUTSK3	IRK3	0.95
11	IRKUTSK	IRKT	0.95
12	JUNGFRAUJOCH	JUN1	0.95
13	JUNGFRAUJOCH2	JUNG	0.95
14	KERGUELEN	KERG	0.93
15	KIEL	KIEL	0.99
16	MCMURDO	MCMD	0.93
17	MAGADAN	MGDN	1.00
18	NOVOSIBIRSK	NVBK	0.99
19	NEWARK	NWRK	0.97
20	OULU	OULU	0.92
21	POTCHEFSTROOM	PTFM	0.84
22	SANAE	SNAE	0.86
23	TBILISI	TBLS	0.80
24	TERREADELIE	TERA	0.85
25	THULE	THUL	1.00
26	TSUMEB	TSMB	0.78
27	TIXIE BAY	TXBY	0.97

them had similar CR counting structure in their initial, main and decay phases. They are more like those that had  $r = 0.99$  except that their initial increase in CR count is greater compared with those that had  $r = 0.99$ . DPRV (with  $r = 0.8$ ) is just like this group except that its pre-peak count is closer to the peak just like those that had  $r = 0.99$ . Only INVK and TXBY are at high-latitude in the Northern Hemisphere while the rest (including DPRV) are in the mid-latitude Northern Hemisphere. Except for INVK, the other three are equal at close longitude.

IRKT, IRK2, IRK3, JUN1, HRMS, JUN1 and JUNG had  $r = 0.95$ . They had similar profiles; a relatively small enhancement in the first one hour of the GLE, followed by the peak (at 12:00 UT) and quasi-exponential decay. All the NMs in this category are in the mid-latitude Northern Hemisphere except HRMS which is in the Southern Hemisphere.

MCMD and KERG had  $r = 0.93$ . Their CR count varied slightly. Their initial count was very small, similar to those with  $r = 1.0$ . While MCMD had two other counts before the peak, KERG had one more count before the peak which is widely separated from the peak count. Their peaks did not occur at the same time and while KERG is at mid-latitude in



**Table 4**  
PCA Summary of the 1989 September 29 Event (GLE42)

SN	Principal Component	Standard Deviation	Proportion of Variance	Cumulative Proportion
1	PC1	4.8814	0.8825	0.8825
2	PC2	1.7358	0.1116	0.9941
3	PC3	0.327 75	0.003 98	0.998 11
4	PC4	0.153 56	0.000 87	0.998 98
5	PC5	0.120 43	0.000 54	0.999 52
6	PC6	0.067 44	0.000 17	0.999 69
7	PC7	0.064 94	0.000 16	0.999 84
8	PC8	0.038 61	0.000 06	0.999 90
9	PC9	0.027 89	0.000 03	0.999 93
10	PC10	0.022 25	0.000 02	0.999 95
11	PC11	0.019 99	0.000 01	0.999 96
12	PC12	0.017 13	0.000 01	0.999 97
13	PC13	0.012 22	0.000 01	0.999 98
14	PC14	0.011 81	0.000 01	0.999 98
15	PC15	0.009 556	0.000 000	0.999 990
16	PC16	0.009 313	0.000 000	0.999 990
17	PC17	0.008 321	0.000 000	0.999 990
18	PC18	0.006 835	0.000 000	0.999 990
19	PC19	0.006 02	0.000 000	1.000 000
20	PC20	0.005 639	0.000 000	1.000 000
21	PC21	0.005 134	0.000 000	1.000 000
22	PC22	0.004 496	0.000 000	1.000 000
23	PC23	0.004 206	0.000 000	1.000 000
24	PC24	0.003 346	0.000 000	1.000 000
25	PC25	0.003 096	0.000 000	1.000 000
26	PC26	0.002 781	0.000 000	1.000 000
27	PC27	0.002 326	0.000 000	1.000 000

the Southern Hemisphere, MCMD is at high-latitude in the Southern Hemisphere.

OULU and APTY are at high-latitude in the Northern Hemisphere. They are also at close longitude and had their peak at 13:00 UT. They have two peaks or two counts at their peaks in addition to a widely separated pre-peak count. Two of them had  $r=0.92$ . SNAE (with  $r=0.86$ ) and TERA (with  $r=0.85$ ) have identical profiles. Their GLE began with a very small increase in CR count and later two other steps in their counts before the peak that occurred at 14:00 UT. Both of them are at high-latitude in the Southern Hemisphere.

PTFM, TBLS, TSMB and BJNG had 0.84, 0.80, 0.78 and 0.73 as their respective  $r$  values. They had similar profiles characterized by a rapid rise to peak count and rapid decay. While TBLS and BJNG are at the lower part of mid-latitude in the Northern Hemisphere, PTFM and TSMB are at low-latitude in the Southern Hemisphere.

From the aforementioned it is seen that in PC1 there are differences in the rising pattern of some NMs with close  $r$  values. Looking at their decay patterns, these differences tend to disappear. The decay counts of all the NMs having  $r=0.97$  to 1.00 are quite similar and account for about 60% of them. All of the NMs having high  $r$  values arise from their decay

pattern being the same. Their different rises to peak count are also responsible for their various  $r$  values.

In GLE 42, 13 of the NMs that had their peak at 12:00 UT did not have  $r$  values that are the same. In addition, eight of the NMs that had peaks at 13:00 UT did not have similar  $r$  values. The same trend is observed in the NMs that had peaks at 14:00 UT. All these notwithstanding, NMs with the same  $r$  values tend to be simultaneous. This is in agreement with Okike & Collier (2011) who were able to distinguish between NMs with simultaneous FD and those that are not simultaneous.

### 3.5. Correlation of the Second Principal Component (PC2) with Raw Data in GLE 42

PC2 in this event (shown in Figure 5) accounts for 11.16% of the total intensity variation in the raw data as can be seen from its total variance (11.16%) in Table 4. In PC2 of this event, the features of the NMs' CR counts prior to the GLE and the main phase are reflected by their  $r$  values. The main phase here includes the moment of initial increase in CR counts up to the moment decay commenced.

IRKT, IRK2, IRK3, JUN1, JUNG, HRMS and TBLS with  $r=-0.3$  showed no sign of double peak or pre-peak count. In an hourly plot, double peaks are not obvious. KERG and MCMD, having  $r=0.3$ , showed features of seemingly double peaks with little time lag between them.

For the following NMs: DPRV, NVBK, TXBY and KIEL whose  $r$  are 0.11, 0.00, 0.20 and 0.10 respectively, their rising pattern of CR counts is similar with greater time lag between the pre-peak and peak count (or possibly between the two peaks) when compared with those having  $r=0.3$ . OULU with  $r=0.4$  had a larger time lag between the pre-peak and peak count than those with  $r=0.3$ .

TSMB, TBLS and BJNG all had  $r=-0.6$ . Each of them had an initial rise in CR count and rose from there to their peak count. In this case, they did not have pre-peak count. In comparison, those whose  $r$  is 0.4 (APTY, OULU, SNAE, TERA) had pre-peak count or possibly double peaks.

### 3.6. Correlation of the First Principal Component (PC1) with Raw Data in 1989 October 22 Event (GLE 44)

The PC1 of this event is as shown in Figure 6 while Tables 5 and 6 contain the quantitative results. The total variance associated with PC1 alone is 86.81. This also implies that 86.81% of the total intensity variation in the raw data in all the stations for this event is accounted for by PC1. Since this value exceeds the benchmark of Okike & Collier (2011), the event may be regarded as a globally simultaneous event. The contributions of the other PCs, especially PC3 to PC23 (1.74%, 1.35%, 0.62%, 0.19%, 0.13%, 0.09%, 0.06% and the rest are 0.00% ) suggest that they are noise.

The PC1 signal of the 1989 October 22 event (see Figure 6) represents the combined signals from all the NMs. Generally

**Table 5**

Full Names of the Places where the NMs for the 1989 October 22 Event (GLE 44) are Located, their Short Names and their Correlation Coefficients ( $r$  Values) of PC1

SN	NM Full Name	NM Short Name	$r$
1	APATITY	APTY	0.97
2	CALGARY	CALG	0.91
3	CLIMAX	CLMX	0.93
4	HERMANUS	HRMS	0.92
5	INUVIK	INVK	0.97
6	IRKUTSK2	IRK2	0.87
7	IRKUTSK3	IRK3	0.90
8	IRKUTSK	IRKT	0.89
9	JUNGFRAUJOCH	JUN1	0.91
10	JUNGFRAUJOCH2	JUNG	0.92
11	KERGUELEN	KERG	0.97
12	LOMNICKY STIT	LMKS	0.93
13	MCMURDO	MCMD	0.75
14	MAGADAN	MGDN	0.97
15	NOVOSIBIRSK	NVBK	0.93
16	NEWARK	NWRK	0.94
17	OULU	OULU	0.97
18	SANAE	SNAE	0.95
19	SOUTH POLE	SOPO	0.89
20	TERREADELIE	TERA	0.96
21	THULE	THUL	0.95
22	TIXIE BAY	TXBY	0.97
23	YAKUTSK	YKTK	0.97

the signal shows that there was a major FD on Saturday, 21 October and its recovery continued until the later part of Sunday when the GLE started. There were other small FDs as part of the recovery phase of the large FD. It is also seen from the signal that there was a pre-peak count. Usually, this signal is a representation of the manifestation of signals from the majority of the NMs, thus not all of them will have a pre-peak count.

APTY, KERG, MGDN, OULU, TXBY, INVK, YKTK and TXBY had  $r=0.97$  suggesting that their signal structures of the CR count are similar. All of them except MGDN had their peak at 19:00 UT while MGDN had its peak at 18:00 UT. Before the peak count, there was another count in all of them which is an enhancement of high value except in MGDN whose initial count was just a recovery from a minor FD prior to the enhancement. The CR counts before the peaks in INVK and TXBY are more widely separated than those in the rest of this group. Five of these NMs had similar decays which cannot be said to be quasi-exponential. MGDN is the only one that is at mid-latitude in the Southern Hemisphere while the rest are at high-latitude in the Northern Hemisphere. OULU whose  $r$  value is 0.98 has the same profile as APTY, INVK, KERG and TXBY NMs and is also at high-latitude in the Northern Hemisphere. TERA, which had  $r=0.96$ , is very much like those with  $r=0.97$  both in its rise to peak count and decay count. It however differed from them in the two minor FDs that

**Table 6**

PCA Summary of the 1989 October 22 Event

SN	Principal Component	Standard Deviation	Proportion of Variance	Cumulative Proportion
1	PC1	4.4683	0.8681	0.8681
2	PC2	1.423 55	0.088 11	0.956 19
3	PC3	0.631 73	0.017 35	0.973 54
4	PC4	0.556 92	0.013 49	0.987 02
5	PC5	0.378 25	0.006 22	0.993 24
6	PC6	0.2090	0.0019	0.9951
7	PC7	0.175 45	0.001 34	0.996 48
8	PC8	0.1439	0.0009	0.9974
9	PC9	0.115 68	0.000 58	0.997 96
10	PC10	0.100 44	0.000 44	0.998 40
11	PC11	0.086 89	0.000 33	0.998 73
12	PC12	0.082 79	0.000 30	0.999 03
13	PC13	0.068 35	0.000 20	0.999 23
14	PC14	0.064 00	0.000 18	0.999 41
15	PC15	0.060 78	0.000 16	0.999 57
16	PC16	0.056 62	0.000 14	0.999 71
17	PC17	0.043 54	0.000 08	0.999 79
18	PC18	0.034 80	0.000 05	0.999 84
19	PC19	0.033 51	0.000 05	0.999 89
20	PC20	0.028 85	0.000 04	0.999 93
21	PC21	0.026 38	0.000 03	0.999 96
22	PC22	0.023 99	0.000 03	0.999 98
23	PC23	0.019 01	0.000 02	1.000 00

are part of the recovery phase of the main FD. TERA is at high-latitude in the Southern Hemisphere.

NWRK, SNAE and THUL, all of which had  $r=0.95$ , have counts that are almost the same as the group above with  $r=0.97$ . The main difference between them is that in this group, the count before the peak (pre-peak count) and the peak count are much closer in time than in the former group. They too did not have at least a quasi-exponential decay. In this group, only NWRK is at the lower part of the mid-latitude in the Northern Hemisphere while the rest are at high-latitude in the Northern Hemisphere, with the exception of SNAE that is in the Southern Hemisphere.

LMKS, CLMX, HRMS and NVBK had counts that could not be said to be a GLE. Their  $r$  values are 0.93, 0.93, 0.92 and 0.93 respectively and their time of peak are 19:00 UT, 18:00 UT, 20:00 UT and 19:00 UT. Though the recovery phase of the prior large FD was completed in them, the additional rise in the CR count was a gradual increase up to the peak. The decay in each of them was also rising and falling CR counts. CLMX looked more like those that had  $r=0.95$ , however both its peak and the first decay count were too close. LMKs and CLMX are at the lower part of the mid-latitude in the Northern Hemisphere. While HRMS is at the mid-latitude in the Southern Hemisphere, NVBK is at the upper part of the mid-latitude in the Northern Hemisphere.

The signals of the CR count in IRKT, IRK2 and IRK3 are identical though their respective  $r$  values are 0.89, 0.87 and 0.90. The three of them apparently had what appears to be double peaks with their corresponding time of peaks as shown: IRKT (11:00 UT, 18:00 UT), IRK2 (10:00 UT, 18:00 UT) and IRK3 (10:00 UT, 18:00 UT). Their rising count could not show a full recovery from the large FD. These shared parameters could be because their latitudes and longitudes are almost the same. JUN1 and JUNG with  $r = 0.91$  and  $0.92$  respectively are among those that had what appears to be double peaks like those whose  $r$  value is between 0.89 and 0.90. Their slight difference is more in their decay.

CALG, MCMD and SOPO with  $r$  values equal to 0.91, 0.75 and 0.89 respectively all had instantaneous jump to high peak count. Though they all showed the large FD prior to the GLE, the minor FDs are barely noticeable. Their decays were equally quasi-exponential. The three of them had simultaneous peaks at 18:00 UT. In this event too, we see that NMs with the same  $r$  values tend to be simultaneous and 80% of NMs that had  $r$  greater than or equal to 0.93 had their peak at the same time (19:00 UT).

### 3.7. Correlation of the Second Principal Component (PC2) with Raw Data in 1989 October 22 Event (GLE 44)

We find that 8.81% of the total intensity variation in the raw data is accounted for by PC2 (Figure 7) in this event. In this PC2, CALG, MCMD and SOPO have  $r = 0.40$ ,  $0.43$  and  $0.43$  respectively. They all had instant rise to peak count. On the other hand, IRK2, IRK3 and IRKT also had  $r = -0.41$ ,  $-0.40$  and  $-0.41$  respectively. They are the exact opposite of the group above having  $r = 0.4$  in that they could not be said to have GLE as they had gradual rise from the previous FD and could not recover to the initial GCR count before the decay commenced.

Similar to this group having a negative correlation factor, NVBK, LMKS and HRMS, all of which had  $r = -0.35$ , had gradual rise in the CR counts up to the maximum count in each of them. They too did not decay to the initial GCR count. JUN1 and JUNG, whose  $r$  values are  $-0.39$  and  $-0.36$  respectively, share the same signal structure as other NMs with negative  $r$  above.

APTY, CLMX, NWRK, OULU and TERA had  $r = 0.2$  and all of them had clear GLE unlike those with negative  $r$  values. In addition, they had a relatively large initial increase in CR count before the peak. SNAE, which has  $r = 0.25$ , shared a similar profile as NMs with  $r = 0.2$ . In general, PC2 in this event shows that positive  $r$  was assigned to NMs that had clear GLE while negative  $r$  was for those that did not have clear GLE.

Table 7

Full Names of the Places Where the NMs for GLE 60 are Located, Their Short Names and Their Correlation Coefficient ( $r$  values) of PC1

SN	NM Full Name	NM Short Name	$r$
1	APATITY	APTY	0.88
2	CALGARY	CALG	0.89
3	CAPE SCHMIDT	CAPS	0.88
4	CLIMAX	CLMX	0.90
5	FORT SMITH	FSMT	0.40
6	HERMANUS	HRMS	0.80
7	INUVIK	INVK	0.90
8	IRKUTSK2	IRKT2	0.90
9	IRKUTSK	IRKT	0.89
10	JUNGFRAUJOCH	JUN1	0.80
11	KERGUELEN	KERG	0.88
12	KINGSTON	KGSN	0.90
13	KIEL	KIEL	0.94
14	LOMNICKY STIT	LMKS	0.90
15	MCMURDO	MCMD	0.48
16	MAGADAN	MGDN	0.60
17	MAWSON	MWSN	0.66
18	NAIN	NAIN	0.78
19	NORILSK	NRLK	1.00
20	NOVOSIBIRSK	NVBK	0.18
21	NEWARK	NWRK	0.40
22	OULU	OULU	0.89
23	SANAE	SNAE	0.76
24	SOUTH POLE	SOPO	0.78
25	TERREADELIE	TERA	0.90
26	THULE	THUL	0.89
27	TIXIE BAY	TXBY	1.00
28	YAKUTSK	YKTK	0.89

### 3.8. Correlation of the First Principal Component (PC1) with Raw Data from 2001 April 15 (GLE 60)

Figure 8 is the PC1 of GLE 60. The quantitative results of this event are in Tables 7 and 8. The respective variances associated with PC1, PC2 and PC3 are 64.67, 14.26 and 11.04. This shows that three of these PCs are required to account for 89.97% ( $64.67\% + 14.26\% + 11.04\%$ ) of the total intensity variation in the raw data. This is an indication that this is a weak GLE and is seen at different times at different NM stations.

The PC1 signal shows the combined effect of all the NMs' CR count. The correlation in PC1 signifies the relationship between the signal from each NM and the raw data; the higher the correlation value of each NM with the raw data signal, the closer the signal of the NM resembles the raw signal (PC1 signal).

In this event, NVBK had the least  $r$  (0.18). It observed several small FDs before the GLE which made it significantly different from any other NM in this group. It rose to instant peak count at 14:00 UT and had a quasi-exponential decay phase.

**Table 8**  
PCA Summary of the 2001 April 15 Event (GLE 60)

SN	Principal Component	Standard Deviation	Proportion of Variance	Cumulative Proportion
1	PC1	4.2553	0.6467	0.6467
2	PC2	1.9984	0.1426	0.7893
3	PC3	1.7582	0.1104	0.8997
4	PC4	1.202 87	0.051 67	0.951 40
5	PC5	0.964 76	0.033 24	0.984 64
6	PC6	0.457 27	0.007 47	0.992 11
7	PC7	0.290 61	0.003 02	0.995 13
8	PC8	0.229 15	0.001 88	0.997 00
9	PC9	0.1909	0.001 30	0.9983
10	PC10	0.115 87	0.000 48	0.998 78
11	PC11	0.089 87	0.000 29	0.999 07
12	PC12	0.078 65	0.000 22	0.999 29
13	PC13	0.066 83	0.000 16	0.999 45
14	PC14	0.060 54	0.000 13	0.999 58
15	PC15	0.049 28	0.000 09	0.999 67
16	PC16	0.045 96	0.000 08	0.999 74
17	PC17	0.042 12	0.000 06	0.999 81
18	PC18	0.039 13	0.000 05	0.999 86
19	PC19	0.031 28	0.000 03	0.999 90
20	PC20	0.027 08	0.000 03	0.999 92
21	PC21	0.025 25	0.000 02	0.999 95
22	PC22	0.020 95	0.000 02	0.999 96
23	PC23	0.018 34	0.000 01	0.999 97
24	PC24	0.017 54	0.000 01	0.999 99
25	PC25	0.012 89	0.000 01	0.999 99
26	PC26	0.0117	0.0000	1.0000
27	PC27	0.008 066	0.0000	1.000 000
28	PC28	0.006 679	0.0000	1.000 000

FSMT and NWRK whose  $r$  values are equal ( $r = 0.40$ ) have identical signal structure of the CR hourly count. In their individual plots that are not shown here, they reveal a small FD prior to the GLE. The plots are characterized by instant rise to peak count and quasi-exponential decay. FSMT displayed peak at 16:00 UT while NWRK had peak at 14:00 UT. MCMD featured the closest  $r$  value to FSMT and NWRK had a sharp rise to peak count which occurred at 17:00 UT. It is not known why it should have its peak later than all the other NMs. Its onset began with a rise to a relatively high count in the CR before it dropped to a value which could be called an FD. From this maximum depression of the FD, it rose instantly to the peak count.

MWSN, which also had instant rise to peak count at 14:00 UT and  $r$  value of 0.65, showed a small FD that is barely noticeable before the GLE. It decayed quasi-exponentially. The  $r$  value of MGDN was 0.60. It recorded similar CR count signal as MWSN but the small FD in MGDN is more outstanding than that in MWSN. MGDN also had peak at 14:00 UT.

Both SOPO and NAIN had  $r = 0.78$  and thus similar signal structure of the CR count. The plots for the two stations display a rise to peak count at 14:00 UT, followed by quasi-exponential

decay. The precursory FDs that were seen in other NMs did not show up in them.

HRMS and JUN1 also feature similar profiles characterized by rising and falling CR count during the recovery phase of a precursory large FD, a jump to peak count that occurred at 14:00 UT and decay phase that is not exponential. The profile from both NMs revealed another small FD besides the large FD before the GLE. The  $r$  for both stations is 0.78.

The  $r$  value of APTY, CAPS and KERG is 0.88. All the profiles show pre-peak counts and a peak count that occurred at 15:00 UT. The background GCR was relatively flat in all of them implying that they did not observe the precursory FD. The decay was quasi-exponential in the three NMs. With the  $r$  value of THUL, OULU, CALG and YKTK equal to 0.89, their profiles are almost the same as those characterized by  $r = 0.88$ . Apart from YKTK that had an initial small rise in CR count, others had instant rise to peak count that occurred at 14:00 UT. Their profiles did not show any FD just before the GLE and the decay was quasi-exponential in all the profiles.

There are two groups that had  $r = 0.90$  and both groups showed peaks at 14:00 UT. The first group which includes INVK, KGSN and TERA did not show any FD and rose to peak count instantaneously. Their decay was equally quasi-exponential. The second group includes LMKS, IRKT and CLMX. In them, there was a large FD that began a day before the GLE and a small FD just before the GLE. They rose to instant peak count at 14:00 UT from the recovery phase of the small FD. They too had a quasi-exponential decay phase.

NRLK and TXBY had the highest  $r$  value of 1.00. They both had two small FDs before the GLE and also pre-peak counts. Their peak occurred at 15:00 UT and they decayed quasi-exponentially. In nearly all the NMs in this event, equal  $r$  values resulted in simultaneous peak count and similar signal structure of the CR count. Apart from the rising phase of the GLE CR count, the decay phases of FSMT, MCMD, HRMS, JUN1, IRKT and LMKS are the reason their  $r$  values are low. This is because the decay shows a significant departure from the trend displayed by those with high  $r$  values. Almost all the NMs having  $r$  values from 0.88 to 0.90 are in the low-latitude region in the Northern Hemisphere. However, CAPS, INVK and THUL are at high-latitude in the Northern Hemisphere while KGSN and NWRK are at mid-latitude in the Southern and Northern Hemisphere respectively.

### 3.9. Correlation of the Second Principal Component (PC2) with Raw Data from GLE 60

PC2 (Figure 9) in this event accounted for 14.26% of the total intensity variation in the raw data. NMs in PC2 for this event having high  $r$  are also the ones that had low  $r$  values in PC1. In this PC2, NMs with positive high  $r$  values had an instant rise to peak count and their GCR counts before the GLE were relatively stable (flat). These NMs and their  $r$  are: NAIN

**Table 9**Full Names of the Places Where the NMs for GLE 69 are Located, Their Short Names and Their Correlation Coefficient ( $r$  Values) of PC1

SN	NM Full Name	NM Short Name	$r$
1	APATITY	APTY	1.00
2	CALGARY	CALG	1.00
3	CAPE SCHMIDT	CAPS	1.00
4	CLIMAX	CLMX	0.65
5	FORT SMITH	FSMT	-0.40
6	INUVIK	INVK	1.00
7	IRKUTSK2	IRK2	0.70
8	KERGUELEN	KERG	1.00
9	KINGSTON	KGSN	1.00
10	KIEL	KIEL	1.00
11	MCMURDO	MCMD	-0.70
12	MAGADAN	MGDN	1.00
13	MAWSON	MWSN	1.00
14	NAIN	NAIN	1.00
15	NORILSK	NRLK	0.95
16	NOVOSIBIRSK	NVBK	0.95
17	NEWARK	NWRK	1.00
18	OULU	OULU	1.00
19	SOUTH POLE	SOPO	0.80
20	TERREADELIE	TERA	0.70
21	THULE	THUL	1.00
22	YAKUTSK	YKTK	1.00

(0.46), THUL (0.50), SNAE (0.60), MGDN (0.75) and NWRK (0.75). Those having negative high  $r$  did not have the same stable GCR count before and after the GLE even though they had a sharp rise to peak count. These NMs and their  $r$  are: NVBK (-0.58) and MCMD (-0.65).

APTY, CAPS and KERG have  $r = -0.35$ . Three of them had pre-peak count which possibly could be a sign of a double peak structure that did not show up in hourly plots. YKTK having  $r = 0.40$  also had what looks like double peaks. The difference between it and these three with  $r = -0.35$  is that its second count at the peak is lesser than the first while in those three above, the first count was less than the second.

SOPO and MWSN had  $r = 0.1$ . They both have similar profiles characterized by stable (or flat) GCR count before and after the GLE. This means that there was no FD in them before the GLE. They also had an instant rise to peak count. JUN1 which had  $r = -0.1$  had an unstable GCR count before the GLE and after the decay. It showed a large FD and after the recovery phase of the FD, the GLE began with a gradual increment in the CR count for several hours before jumping to a peak count. The following NMs with  $r = -0.2$  have similar CR signal counting structures as JUN1: LMKS, IRKT and IRK3.

KGSN, INVK and CALG have  $r = -0.25$ . They had stable GCR counts before and after the GLE. Thus for several hours before the GLE they did not observe small or large FD. They too had an instant rise to peak count. KIEL with  $r = 0.3$  is the only NM having reversed  $r$  value compared to those NMs with

**Table 10**

PCA Summary of the 2005 January 20 Event

SN	Principal Component	Standard Deviation	Proportion of Variance	Cumulative Proportion
1	PC1	4.2424	0.818 10	0.8181
2	PC2	1.41786	0.091 38	0.909 47
3	PC3	0.97242	0.042 98	0.952 45
4	PC4	0.80133	0.029 19	0.981 64
5	PC5	0.52269	0.012 42	0.994 05
6	PC6	0.25495	0.002 95	0.997 01
7	PC7	0.15675	0.001 12	0.998 13
8	PC8	0.12213	0.000 68	0.998 80
9	PC9	0.10396	0.000 49	0.999 29
10	PC10	0.07308	0.000 24	0.999 54
11	PC11	0.06076	0.000 17	0.999 71
12	PC12	0.04694	0.000 10	0.999 81
13	PC13	0.03855	0.000 07	0.999 87
14	PC14	0.03397	0.000 05	0.999 93
15	PC15	0.02303	0.000 02	0.999 95
16	PC16	0.02006	0.000 02	0.999 97
17	PC17	0.01595	0.000 01	0.999 98
18	PC18	0.01368	0.000 01	0.999 99
19	PC19	0.01212	0.000 01	0.999 99
20	PC20	0.00917	0.000 00	1.000 00
21	PC21	0.006379	0.000 00	1.000 000
22	PC22	5.17e-18	0.000 00	1.000 000

$r = -0.25$ . KIEL did not show a stable GCR count both before and after the GLE. It had at least two small FDs just before the GLE.

### 3.10. Correlation of the First Principal Component (PC1) with Raw Data from 2005 January 20 (GLE 69)

Figure 10 is PC1 of the 2005 January 20 GLE. Tables 9 and 10 list the quantitative results of the GLE. The PCA summary for this event in Table 10 signifies that the proportion of variance for PC1 is 0.8181, implying that PC1 accounts for 81.8% of the total intensity variation in the raw data. In comparison with the event on 1989 September 29, the presented results of the PCA confirm that the GLE from 2005 January 20 is smaller (see Moraal & Caballero-Lopez 2014).

Fourteen out of twenty-two NMs considered for PCA have  $r = 1.00$  in PC1 and their peak occurred at 07:00 UT. This is about 63%. The NMs which included APTY, CALG, CAPS, INVK, KERG, KGSN, KIEL, MGDN, MWSN, NAIN, OULU, SNAE, THUL, YKTK and NWRK all had similar decay phases up to the day after the event. In other words their recovery phases were also similar. All the profiles exhibit quasi-exponential decay. Though there are noticeable differences in their rise to peak count, their decay counts are so similar that they are not distinguishable. Seven of these NMs are at high-latitude in the Northern Hemisphere, four at mid-latitude in the Northern Hemisphere, and two at mid- and high-latitude in the Southern Hemisphere.



Most of the NMs whose  $r$  values are equal to 1.00 had initial rises in CR count that are relatively small. The increase is much less in INVK, CAPS, NWRK, THUL and MGDN than in others. Their profile also shows that the GLE began with an FD as a precursor; the bigger the pre-increase in the CR count, the larger the FD. Those that revealed a small FD in addition to a large FD (for example MGDN, SOPO and THUL) did not show an initial rise in the CR count, instead they had an instant jump to the peak count.

NRLK and NVBK had  $r = 0.95$  and the profiles from both stations peak at 07:00 UT. Their decay phases are similar and slightly different from those with  $r = 1.00$  above. They also had initial increases in CR counts after the recovery phase of the FD before an instant jump to the peak. The steepness of these precursory counts was more in NVBK than in NRLK. NRLK equally had a pre-peak count which was not detected in NVBK.

IRK2 had  $r = 0.70$ . It never recovered like the previous ones considered before there was rise in CR count the same day. It had its peak at 06:00 UT. The decay in CLMX with  $r = 0.65$  was below the initial background GCR counts. This made it quite different from others. FSMT and MCMD, unlike any other NM, recorded negative correlation factors with  $r = -0.40$  and  $-0.70$  respectively. They recorded FD when others were observing enhancement of CRs. MCMD had greater value of negative correlation because after the maximum depression or drop in the CR counts, it tried to recover but the count dropped again to almost the maximum depression. From this later point, it registered an enhancement. MCMD therefore recorded two FDs and the first one is the larger FD.

Most of the stations having  $r \geq 0.95$  are at high-latitudes ( $60^\circ 10' - 76^\circ 50'$ ) in the Northern Hemisphere except NVBK (latitude  $54^\circ 80'$ ). All NMs that had two steps in their rise to peak also had  $r \geq 0.95$ . The two among them that have  $r = 0.95$  (NRLK and NVBK) also show identical CR count signal. MWSN at high-latitude ( $-67^\circ 60'$ ) in the Southern Hemisphere has  $r = 1.00$ . SOPO at latitude  $-90^\circ 0'$  and TERA at latitude  $-66^\circ 70'$  also have  $r = 0.80$  and  $0.70$ , respectively. They both have instant rise to peak count that occurred at 06:00 UT. SOPO did not exhibit an FD prior to the GLE. Stations that display instant rise to the CR peak count also have  $r \leq 0.80$  in GLE 69.

### 3.11. Correlation of the Second Principal Component (PC2) With Raw Data from GLE 69

PC2 (Figure 11) in this event accounts for 9.14% of the total intensity variation in the raw data. It has always been the case that close values of  $r$  reflect greater resemblance in the structure of the CR count. In this PC2, those that have  $r = 0.1$  such as APTY, CALG, CAPS, KERG, KIEL, MWSN and YKTK also had close CR count signals. While CAPS and INVK rose to peak counts instantly, the rest had initially small increases in

CR count before jumping to the peak count. FSMT is the only NM with close negative  $r$  ( $-0.2$ ) to these NMs. It had a complete large FD at the period others were having GLE. Its FD is a reverse of the GLE in terms of its CR counts. First, it dropped to a certain level before its maximum decent in the CR count just as the majority of these NMs first had an initial increase in CR count before the peak count. NAIN (who has  $r = 0.00$ ) and other NMs, which had  $r = 0.2$  (MGDN, NVBK and THUL), had similar CR count signal structures as those which had  $r = 0.1$ .

While MCMD has  $r = 0.40$ , NRLK has  $r = -0.40$ . NRLK is characterized by a large pre-peak count and later a peak count. At the same time, MCMD had a large FD. Even when it tried to recover, it dropped again to almost the level before the onset of recovery. After its recovery, the observed GCR count was higher than what it was before the GLE event. The decay of NRLK also established a new level of GCR count which is higher than it was before the GLE.

SOPO, TERA and CLMX have  $r = -0.65$ ,  $-0.68$  and  $-0.70$  respectively. There were no other NMs with high negative  $r$  value that could be compared with these values.

## 4. Summary and Conclusion

We have applied PCA to the GLE of CR events on 1978 May 7; 1989 September 29; 1989 October 22; 2001 April 15 and 2005 January 20. In this study, a combination of PC1 and PC2 sufficiently accounted for the information in the raw data in each of the events.

Our application of PCA to GLE studies yielded the following information:

1. PCA is well suited for the investigation of GLEs across multiple NM stations. It can, for example, clearly distinguish between strong/globally simultaneous GLEs and weak/non-simultaneous ones.
2. The time profile and structure of the CR count signal determine which of them will have the same or close values of  $r$ . In all the events, almost all the NMs with the same  $r$  values have the same time of peak. Having the same time of peak may not lead to equivalent  $r$  value but having the same  $r$  value in most cases resulted in the same time of peak.
3. NMs with identical signal structures of CR counts are assigned the same  $r$  value.
4. Most of the NMs that have the same  $r$  value tend to occupy close latitude range.
5. NMs that have FD at the time others observed GLE in an event can be distinguished from others as they usually have negative correlation factor as opposed to others that have positive correlation factor.
6. NMs with close  $r$  values equally have very similar CR profiles.

7. The features that are common to most or all the NMs in an event are what make  $r$  be higher in all the NMs. Such features could be that the majority of the NMs had instant rise to peak count or that they had the same decay pattern, or that they all had the same pre-peak count or a combination of two or more of these features.
8. Analyzing PC2 requires making a comparison between NMs having the same or close positive  $r$  with NMs having the same or close values of negative  $r$ . It usually yields information on the differences between NMs having opposite but the same or close values of  $r$ . The presented result of PCA on GLE is preliminary. It has the potential of stimulating a more detailed future work.
9. One of the major flaws of the PCA tool in the CR investigation is its inability to account for CR diurnal anisotropies in raw CR data. CR diurnal anisotropy has a significant influence on the time-intensity profiles/structure of CR signal. The results presented here should, therefore, be viewed with caution. In a future work, efforts will be made to remove the contributions of CR anisotropy before passing the data to the PCA algorithm. Our investigation has shown that PCA is an essential technique for analyzing multivariate GLE data. GLE research that involves several NM data should even begin with PCA because a lot of shared information by the NMs are promptly thrown out by the PCA.

### Acknowledgments

We are indebted to the Pushkov Institute of Terrestrial Magnetism, Ionosphere, and Radio Wave Propagation, Russian

Academy of Sciences (IZMIRAN) who hosts the website <http://cro.izmiran.ru/common/links.htm>. The raw data used in this investigation are from their data bank. This work was sponsored by the Nigeria Tertiary Education Trust Fund (TETFund) implemented by the Federal University of Technology Owerri, Imo State, Nigeria. We are indebted to the anonymous referee. His/her input greatly improved the manuscript.

### References

- Dray, S. 2008, *Computational Statistics & Data Analysis*, 52, 2228
- Falerio, M., Gomez, J., Munoz, L., et al. 2004, *ApJS*, 155, 167
- Forkman, J., Josse, J., & Piepho, H. 2019, *JABES*, 24, 289
- Hayden, L. 2018, <https://www.datacamp.com/community/tutorials/pca-analysis-r>
- Jolliffe, I. 2002a, *Principal Component Analysis* (2nd ed.; Berlin: Springer)
- Jolliffe, I., & Cadima, J. 2016, *Phil. Trans. R. Soc.*, A374, 20150202
- Kassambara, A. 2017, in *Practical Guide to Principal Component Methods in R* (STHDA)
- McCracken, K., Moraal, H., Shea, M., et al. 2012, *ApJ*, 761, 12
- Moore, B. 1981, in *IEEE Transactions on Automatic Control*, AC-26, No. 1
- Moraal, H., & Caballero-Lopez, R. 2014, *ApJL*, 790, 16
- Muller, W., Nocke, T., & Schumann, H. 2006, in *Proc. the 2006 Asia-Pacific Symp. on Information Visualization-Australian Computer Society Inco* 60, 121
- Oh, S., Yi, Y., & Kim, H. Y. 2008, *JGR*, 113, A01103
- Okike, O., & Alhassan, J. 2022, *EPJP*, 137, 317
- Okike, O., & Collier, A. B. 2011, *JASTP*, 73, 796
- Riadigos, I., Garcia-Castro, D., Gonzalez-Diaz, D., et al. 2020, *E&SS*, 7, 97
- Sheoran, O., Kumar, V., Poonia, H., & Malik, K. 2020, *IJEAT*, 9, 3
- Tharwat, A. A. I., Hassanien, A., & Schaefer, G. 2015, *IJAPR*, 9124, 246
- Venables, W., Smith, D., & Team, R. C. 2022, *An Introduction to R (Notes on R: A Programming Environment for Data Analysis and Graphics) Version 4.2.0 (2022-04-22)*, <https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf>
- Zuska, Z., Coplinska, J., Dacewicz, E., et al. 2019, *Sustainability*, 11, 6740