

---

# Using neutron monitor network data to improve the detection of space weather events

T. Dudok de Wit<sup>1</sup> and A. A. Chilingarian<sup>2</sup>

<sup>1</sup> LPCE, CNRS and University of Orléans, 3A av. de la Recherche Scientifique, 45071 Orléans cedex 2, France, [ddwit@cnrs-orleans.fr](mailto:ddwit@cnrs-orleans.fr)

<sup>2</sup> Yerevan Physics Institute, Cosmic Ray Division, 2 Alikhanyan Brothers St., Yerevan 375036, Armenia, [chili@aragats.am](mailto:chili@aragats.am)

**Summary.** We present two multivariate statistical methods for improving the real-time detection of space weather events from neutron monitor network data. Both exploit the similarity in the evolution of the count rates, as recorded by different monitors. The same methods can also be used as a simple and yet robust tool for filling data gaps.

Neutron and muon monitor networks can be used as single multidirectional spectrograph, thereby improving the detection of Ground Level Enhancements (GLEs) and of signatures from approaching interplanetary disturbances (Usoskin et al., 1997; Mavromichalaki et al., 2004). The real-time analysis of such network data, however, is often hampered by a number of practical difficulties such as data gaps and noise. Here we show how multivariate statistical analysis methods known as blind source separation methods can help rapidly characterize network data while alleviating some of these problems.

## The blind source separation approach

Let us illustrate our approach using one year of hourly pressure-corrected neutron monitor data, downloaded from the website of the Solar-Terrestrial Division at Izmiran (<ftp://cr0.izmiran.rssi.ru/Cosray!/>). The neutron count rate measured at a location  $x$  and time  $t$  is  $c(x, t)$ . After eliminating all stations that exhibit large data gaps (more than 30% of the time) or experience stability problems, we are left with 43 stations. Small data gaps and spurious values are eliminated using the technique that will be described below.

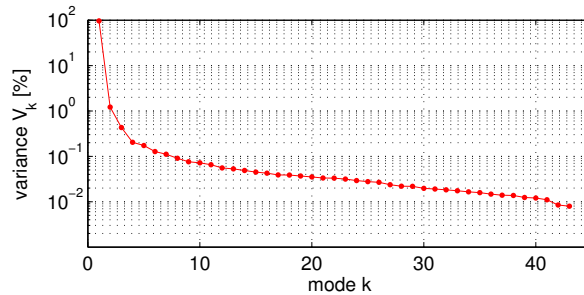
Our working hypothesis is that all count rates can be expressed as a linear combination of a few common regimes. This idea is supported by the remarkable similarity between count rate evolutions as observed at different stations. For that reason, we express the count rate as a linear combination of separable functions of space and time, hereafter called “modes”:  $c(x, t) = \sum_{i=1}^N A_i f_i(t) g_i(x)$  with the orthonormality constraint  $\langle f_i(t) f_j(t) \rangle = \langle g_i(x) g_j(x) \rangle = \delta_{ij}$ , where  $\delta_{ij}$  is the Kronecker symbol and  $\langle \cdot \rangle$  meaning ensemble averaging. This empirical decomposition is unique and is provided directly by the Singular Value Decomposition (SVD) (Chatfield & Collins, 1995). The weights are conventionally sorted in

decreasing order  $A_1 \geq A_2 \cdots \geq A_N \geq 0$ . The number of modes  $N$  equals the number of stations. Heavily weighted modes describe features that are observed coherently by many stations and so they are of prime interest. Detector and statistical noise on the contrary tends to be deferred to the last modes, because it is incoherent in time and in space. Since the method is data-adaptive, the same temporal modes are recovered regardless of the number of monitors, provided that these capture different aspects of dynamics.

A recent generalisation of the SVD, called Independent Component Analysis (ICA), regards the decomposition of multivariate data into a linear combination of modes that are not orthonormal, but independent (Hyvärinen & Oja, 2000). ICA is today often preferred to the SVD when it comes to handling blind source separation problems, in which a small number of “source” terms must be recovered from array data. Indeed, statistical independence generally is a more realistic assumption for disentangling different physical processes than decorrelation. The number of observables must of course exceed the number of modes.

### Application to neutron monitor data

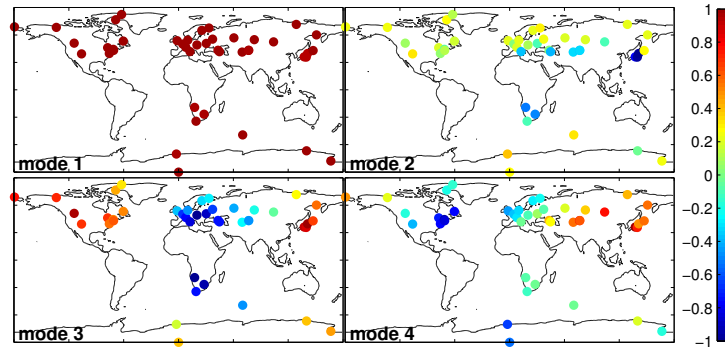
The fraction of the variance that is described by the  $i$ -th mode of the SVD is given by  $V_i = A_i^2 / \sum_{j=1}^N A_j^2$ . These values are plotted in Fig. 1, showing a characteristic distribution with a steep falloff followed by a flat tail. The largest modes capture salient features of the observed variability; three of them describe over 98.2% of the variance of the data. Modes from the tail in the contrary describe incoherent fluctuations that occur locally in time or at one particular station. The modes that are most likely to capture interesting physical features thus are the largest ones (Dudok de Wit, 1995). The number of such modes here is about 3 to 5.



**Fig. 1.** Fraction of the variance (in %) that is explained by each of the 43 modes obtained by SVD.

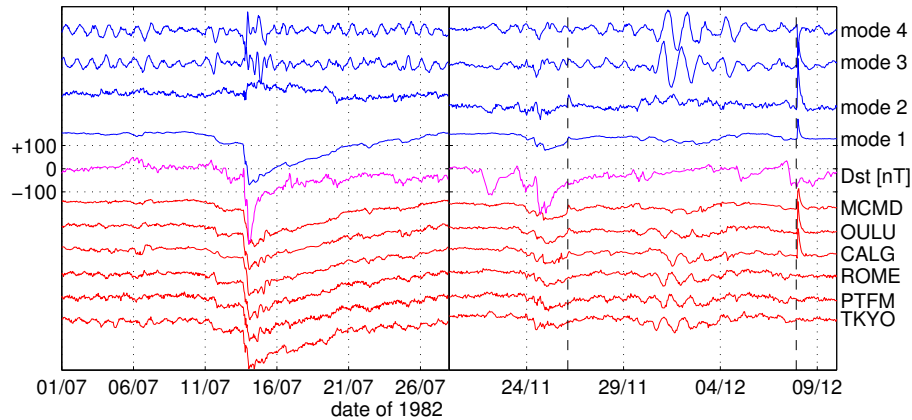
Preprocessing is an important issue. Since we are interested in fluctuations only, for each station we subtract the time average and normalise with respect to the standard deviation.

The spatial structure of the first 4 modes obtained by ICA is shown in Fig. 2. Mode 1 is a mere weighted average of stations (with weights ranging from 0.94 to 1) whereas mode 2 extracts the difference between low and high rigidity stations, with no longitudinal dependence. Mode 1 can therefore be interpreted as a large-scale flux (Belov, 2000) whereas mode 2 expresses a rigidity dependence. Modes 3 and 4 on the contrary exhibit a strong longitudinal dependence; they are also in quadrature, which means that they describe longitudinally



**Fig. 2.** Spatial structure of ICA modes 1 to 4. Each dot represents a monitor; the colour expresses the value of the  $k$ 'th spatial mode  $g_k(x)$  at that location, normalised to the maximum value of  $|g_k(x)|$ .

moving patterns (Aubry & Lima, 1995). Higher order modes in comparison have much less spatial coherence and are more difficult to interpret. The modes obtained by SVD are quite similar but their separating power is not as good.



**Fig. 3.** From top to bottom, temporal profiles of ICA modes 1 to 4, the  $D_{st}$  index and count rates of 6 stations with increasing rigidity. All but the  $D_{st}$  index are plotted with arbitrary units. Two GLEs are indicated with dashed lines.

The temporal structure of the first four ICA modes is shown in Fig. 3 for a large storm in July 1982 and two GLEs in November and December 1982. A first small Forbush decrease occurs on July 11th around 10:00 UT and a much larger one on July 13th at 16:00. The exact timing of their onset times is affected by the omnipresent solar diurnal variation. However, this modulation of the cosmic ray flux, which travels longitudinally as the Earth rotates, is almost entirely captured by modes 3 and 4. By reconstructing the data from the all but modes

3 and 4, the onset comes out more evidently, especially for high rigidity stations. Since in addition the solar diurnal variation is well captured by modes 3 and 4, one can determine its amplitude (defined as  $a(t) = \sqrt{f_3^2(t) + f_4^2(t)}$ ) and check that it drops before sudden storm commencements. Note that modes 3 and 4 are not equivalent to bandpass filtered count rates, since they exhibit occasional sharp transients. In particular, anisotropies associated with arriving shocks disrupt the regular oscillatory behaviour of these two modes and provide another criterion for detecting anomalous events. Such modes could therefore serve as an input for more elaborate cosmic-ray indices for space weather (Kudela & Storini, 2006).

GLEs manifest themselves as a short increase in modes 1 and 2. Two events (a weak one a strong one) are marked by dashed lines in Fig. 3. The increase observed in mode 2 implies that the contribution of GLEs versus that of the cosmic ray background is relatively more important at higher latitudes. A weak signature can also be found in modes 3 and 4 because the impact is not isotropic in longitude.

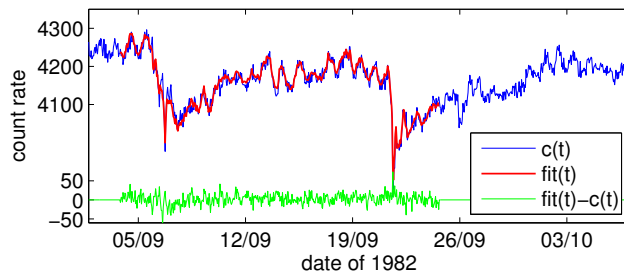
Interestingly, the power spectral density of temporal modes 1 and 2 reveals a power law scaling over more than two decades, which is almost totally devoid of the 24-hour modulation that dominates in modes 3 and 4. These two power laws have distinct spectral indices:  $\alpha_1 = -2.30 \pm 0.03$  for mode 1 and  $\alpha_2 = -1.30 \pm 0.03$  for mode 2. The first value is similar to the one found for the  $D_{st}$  index and attests a predominance of large-scale transients. Mode 2 in comparison has a much higher level of short-scale fluctuations; our interpretation is that this reflects the local conditions of the terrestrial environment. These two different spectral indices support the idea that modes 1 and 2 truly capture two different physical processes.

### Filling data gaps

Another potentially important application is filling the data gaps that plague neutron observations. The idea, which is further developed in (Kondrashov & Ghil, 2006), consists in exploiting the redundancy of the observations to replace missing or corrupted samples with values obtained from the most significant modes. First, all corrupted samples are flagged and replaced by some reasonable estimate. The SVD is then computed. The strongest modes are used to replace the flagged values with a new estimate, and the SVD is then computed again. Convergence usually occurs after 5 to 20 iterations. Two advantages of this method are its extreme simplicity (the number of significant modes is the only tunable parameter) and its excellent performance with measurements that exhibit coherent regimes. The result is illustrated in Fig. 4, in which one month of data from the Climax monitor were deliberately taken out and subsequently reconstructed by SVD. The good performance of the method is attested by the low variance of the residuals and their statistical independence.

### Conclusion and outlooks

Blind source separation methods offer a simple and yet powerful empirical approach for pre-processing in real time data from neutron monitor networks. Even though these methods are statistical, their modes convey a physical interpretation. Independent Component Analysis is well suited for characterizing common regimes, thereby enabling the robust identification of Forbush decreases and GLEs. The Singular Value Decomposition is better adapted for filling data gaps and resampling data.



**Fig. 4.** Original count rate of the Climax neutron monitor ( $c(t)$ ) and its value reconstructed by SVD ( $fit(t)$ ), using four dominant modes out of the 43, and the residual error ( $fit(t) - c(t)$ ).

These methods allow to investigate several other properties, and are presently being tested with data from the Aragats station (Chilingarian et al., 2005). First, due to the fact that both methods are linear, statistical hypothesis tests can be carried out on weak events. Another application is the correlated analysis of high-cadence data for the robust characterization of weak GLEs and anisotropic events. When using for example surface and underground detectors with different energy responses, differences in the amplitude of the spatial modes  $g_k(x)$  are automatically adjusted by the method. Finally, we are using these methods to detect anomalous measurements (drifting count rates, offset errors). A station or a channel that behaves differently from the others often can readily be identified as this difference will tend to be picked up by one single mode.

## References

- Aubry, N., Lima, R.: Spatio-temporal and statistical symmetries. *J. Stat. Phys.* **81**, 793–828 (1995).
- Belov, A.: Large-scale modulation: view from Earth. *Space Science Rev.* **93**, 79–105 (2000).
- Chatfield, C., Collins, A. J., Introduction to Multivariate Analysis. Chapman and Hall, London (1995).
- Chilingarian, A., Arakelyan, K., Avakyan, K. et al., Correlated measurements of secondary cosmic ray fluxes by the Aragats Space-Environmental Center monitors. *Nuclear Instr. Methods Phys. Res. A.* **543**, 483–496 (2005).
- Dudok de Wit, T.: Enhancement of multichannel measurements in plasma physics by biorthogonal decomposition. *Plasma Phys. Contr. Fusion*, **37**, 117–135 (1995).
- Hyvärinen, A., Oja, E.: Independent Component Analysis: algorithms and applications. *Neural Networks*, **13**, 411–430 (2000).
- Kondrashov, D., Ghil, M.: Spatio-temporal filling of missing points in geophysical data sets. *Nonl. Proc. Geoph.* **13**, 151–159 (2006).
- Kudela, K., Storini, M.: Possible tools for space weather issues from cosmic ray continuous records. *Adv. Space Res.* **37**, 1443–1449 (2006).
- Mavromichalaki, H., Yanke, V., Dorman, L., Iucci, N., Chilingaryan, A., Kryakunova, O.: Neutron monitor network in real time and space weather. In: Daglis, I. A. (ed), *Effects of Space Weather on Technology Infrastructure*. NATO Science Series II, **176**, 301–317 (2004).
- Usoskin, I. G., Kovaltsov, G. A., Kananen, H., Tanskanen, P.: The World Neutron Monitor Network as a tool for the study of solar neutrons. *Ann. Geoph.*, **15**, 375–386 (1997).