PROCEEDINGS OF SCIENCE

Data mining of cosmic-ray anisotropy observed with the Global Muon Detector Network

M. Kozai,[a,*] Y. Hayashi,[b] C. Kato,[b] K. Munakata,[b] Y. Masuda,[b] K. Iwai,[c] M. Rockenbach,[d] A. Dal Lago,[d] R. R. S. Mendonça,[d] E. Echer,[d] J. V. Bageston,[e] C. R. Braga,[f] H. K. Al Jassar,[g] M. M. Sharma,[g] M. L. Duldig,[h] J. E. Humble,[h] A. Kadokura,[a,i,j] R. Kataoka,[i,j] S. Miyake,[k] I. Sabbah,[l] P.-S. Mangeard,[m] T. Kuwabara[m] and P. Evenson[m]

[a] *Polar Environment Data Science Center, Joint Support-Center for Data Science Research, Research Organization of Information and Systems, Tachikawa, Tokyo 190-0014, Japan.*

[b] *Department of Physics, Shinshu University, Matsumoto, Nagano 390-8621, Japan.*

[c] *Institute for Space-Earth Environmental Research, Nagoya University, Nagoya, Aichi 464-8601, Japan.*

[d] *National Institute for Space Research (INPE), 12227-010 São José dos Campos, Brazil.*

[e] *Southern Space Coordination, National Institute for Space Research, P.O. Box 5021-97110-970–Santa Maria, RS, Brazil.*

[f] *George Mason University, 4400 University Drive, Fairfax, VA 22030, USA.*

[g] *Physics Department, Kuwait University, P.O. Box 5969 Safat, 13060, Kuwait.*

[h] *School of Natural Sciences, University of Tasmania, Hobart, Tasmania 7001, Australia.*

[i] *National Institute of Polar Research, Tachikawa, Tokyo 190-8518, Japan.*

[j] *Department of Polar Science, School of Multidisciplinary Sciences, The Graduate University for Advanced Studies, SOKENDAI, Tachikawa, Tokyo 190-8518, Japan.*

[k] *National Institute of Technology (KOSEN) Ibaraki College, Hitachinaka, Ibaraki 312-0011, Japan.*

[l] *Department of Laboratory Technology, College of Technological Studies, The Public Authority for Applied Education and Training, Shuwaikh, 72853, Kuwait.*

[m] *Bartol Research Institute and Department of Physics and Astronomy, University of Delaware,217 Shar Laboratory, Newark, DE 19716, USA.*

E-mail: kozai.masayoshi@nipr.ac.jp

Ground-based muon detectors are sensitive to anisotropies of cosmic rays at approximately 50 GeV and have been operated with great stability for over 10 years. The anisotropy is controlled by the space environment in various time scales and is an attractive target of data-driven approaches. We report on unsupervised machine-learning of anisotropy data accumulated by Global Muon Detector Network (GMDN). It will provide a comprehensive and statistical picture of cosmic-ray anisotropies excluding biases such as data selections relying on a visual inspection.

38th International Cosmic Ray Conference (ICRC2023)
26 July - 3 August, 2023
Nagoya, Japan

*Speaker

PoS(ICRC2023)1279

1. Introduction

Cosmic-ray anisotropy is an essential and attractive subject of cosmic-ray physics, interplanetary physics, and heliospheric physics as it is a manifestation of cosmic-ray propagation processes in the space environment. Diverse scientific phenomena in various time scales, such as interplanetary shocks (IP-shocks) and cyclic variations of the heliosphere, are involved in cosmic-ray anisotropy. A comprehensive study covering these variations is necessary to establish a statistical picture and approach a universal mechanism or physics behind the anisotropy. However, most of the previous works focus on a special subject, such as IP-shocks or an 11-year cycle variation of the heliosphere, separately. Such an approach can overlook a scientific finding because the space environment is a complex system where multiple phenomena evolve in parallel. Additionally, typical IP-shock studies are mainly case studies on limited events arbitrarily selected, while a small number of studies challenge establishing a statistical picture of IP-shock events[1–3].

Data mining is expected to be a powerful approach for an inclusive survey of diverse phenomena in anisotropy on a mathematical basis. It will lead to a statistically reliable picture and a discovery of a new profile of cosmic-ray anisotropy. This approach is also expected to be useful for reducing dimensions of anisotropy. Its multidimensional quantities interacting with each other via phase-space dynamics[4] have caused difficulty in establishing a physical interpretation of variation of anisotropy.

Global Muon Detector Network (GMDN) started its operation with two-hemisphere detectors at Nagoya (Japan) and Hobart (Australia) in 1992. SãoMartinho da Serra (Brazil) and Kuwait detectors were installed in 2006. GMDN achieved its initial completion in 2016 by expanding Kuwait detector to a comparable detection area (25 m^2) with Nagoya detector (36 m^2).

This study attempts an inclusive survey of anisotropy for both long- and short-term variations, based on unsupervised machine-learning. Harmonic components of anisotropy derived from GMDN data[5] are used as sample data. Preliminary results demonstrate the advantages of our approach.

2. Preparation of datasets

Muon counting-rate can be modeled in terms of spherical harmonics, as[6]

$$I_{i,j}(t) \approx \sum_{n=0}^{2}\sum_{m=0}^{n}\{\xi_n^{mc}(t)\left(c_{ni,j}^{m}\cos m\omega t_i - s_{ni,j}^{m}\sin m\omega t_i\right) + \xi_n^{ms}(t)\left(s_{ni,j}^{m}\cos m\omega t_i + c_{ni,j}^{m}\sin m\omega t_i\right) \tag{1}$$

where $I_{i,j}$ denotes our observable, deviation [%] of counting-rate recorded in the j-th directional channel of the i-th detector. Zero-level to take the deviation is set at an average counting-rate in each Bartels rotation (27 days). We use 13-17 channels compiled in each detector and total 60 channels in GMDN 4 stations. ξ_n^{mc} and ξ_n^{ms} are expansion coefficients of spherical harmonics in the order (n, m), representing cosmic-ray anisotropy. $c_{ni,j}^{m}$ and $s_{ni,j}^{m}$ are coupling coefficients expressing a response of $I_{i,j}$ to the anisotropy. They are derived by numerically calculating the geomagnetic and atmospheric propagation of cosmic rays with a specific rigidity, and then integrating them for the

rigidity. t is universal time and ω is the angular velocity of Earth's rotation. $\omega t_i = \omega t + \phi_i$ where ϕ_i is geographic longitude of the i-th detector. Anisotropy components ξ_n^{mc} and ξ_n^{ms} are derived by fitting the model function into observed deviation $I_{i,j}$. The best-fitting procedure, including corrections for atmospheric variations, is described in [6, 7] and practically used in previous works[3, 8–10]. One modification in this preliminary analysis from the precedents is the integration upper limit, set at 200 GV for $n > 0$ components, based on estimation by [11].

We performed the best-fitting for 60-channel data recorded every hour by GMDN from 2015 to 2022 and obtained 9 parameters in each time, as listed below.

- ξ_0^{0c}: Isotropic component of spherical harmonics representing the deviation of cosmic-ray density from average level.
- ξ_1^{0c}, ξ_1^{1c}, and ξ_1^{1s}: The 1st order harmonics, which represent uni-directional cosmic-ray flow and respectively correspond to z-, x- and y-components of the flow in this order.
- ξ_2^{0c}, ξ_2^{1c}, ξ_2^{1s}, ξ_2^{2c}, and ξ_2^{2s}: The 2nd order harmonics which represent bi-directional flow induced by spatial derivatives of the uni-directional flow and interplanetary magnetic field.

The best-fit parameters are obtained in the geocentric (GEO) coordinate system whose x-axis is fixed to 0 a.m. in LT, and then transformed to the geocentric solar ecliptic (GSE) coordinate system. The available hours with a small number of inactive channels are 45132, therefore our dataset size is 9 parameters × 45132 hours.

3. Methods and results

3.1 Cluster analysis

We categorize the anisotropy data using unsupervised machine-learning. The anisotropy at each time derived in the previous section is expressed as a single point in the parameter space with 9 dimensions, corresponding to 9 components of the anisotropy components. Each parameter in the dataset is standardized by the Z-score normalization for the whole period as a preprocessing. Then, 45132 points in the 9-dimensional space are grouped by the k-means clustering of Python scikit-learn library[12]. The number of clusters is set at 10, where the sum of squared errors of prediction is sufficiently decreased and the significant correlations with solar activity and IP-shocks discussed in the following sections appear. Each time in the dataset is provided with a cluster ID, 1, 2, 3,..., or 10, by this clustering. It is noted that the standardized parameters are restored to their original quantities after labeling the clusters, for scientific discussions in the following sections.

Figure 1 displays averages of the 0th and 1st anisotropy components in the individual clusters. Standard errors of averages are not plotted because they are so small that it's hard to see. Figure 1(a) shows averages of the 0th component, or density variation ξ_0^{0c}. Averages of the diurnal anisotropy, ξ_1^{1c} and ξ_1^{1s}, are displayed in Figure 1(b1) which is a so-called harmonic dial, where the upward direction corresponds to 0 a.m. (midnight) direction in LT while the right side represents 6 a.m. (morning). It has to be noted that the LT expression is exact in the GEO coordinate system, but now it deviates slightly because we transformed the anisotropy components into the GSE coordinate system. Figure 1(b2) shows averages of north-south anisotropy, ξ_1^{0c}, along with ξ_1^{1c} in a vector

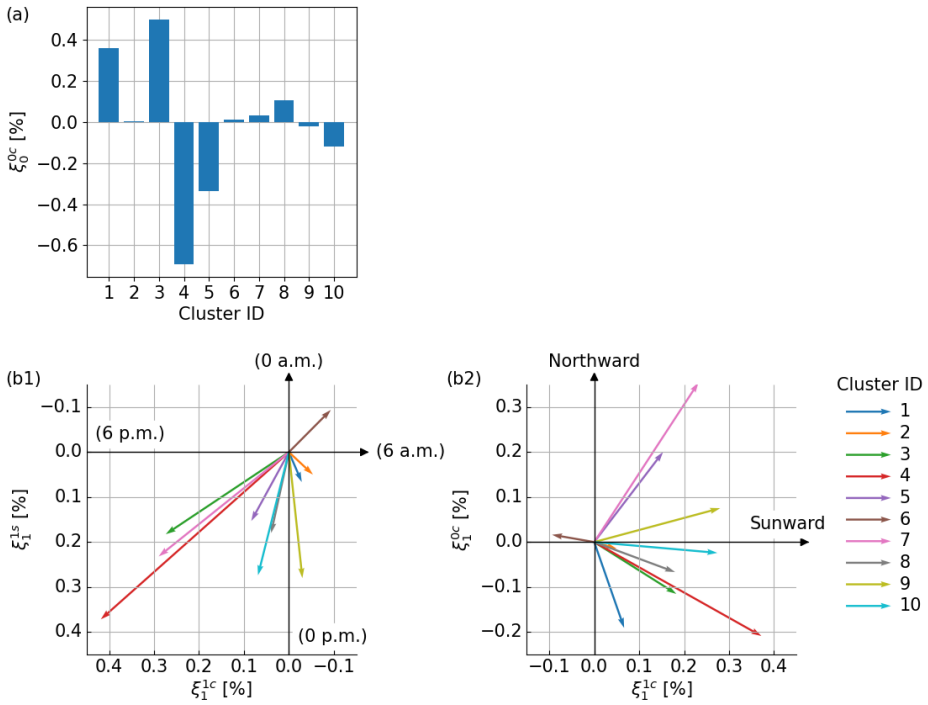

Figure 1: Averages of the 0th and 1st order anisotropy components in the clusters grouped by k-means clustering. (a) Average density deviation ξ_0^{0c} in each cluster from average level. (b1) Average vector of diurnal anisotropy (ξ_1^{1c} and ξ_1^{1s}) in each cluster, presented in a harmonic dial. (b2) Average of north-south anisotropy, ξ_1^{0c}, in each cluster presented in a vector diagram with ξ_1^{1c}

diagram. Averages of the 2nd order anisotropy (ξ_2^{0c}, ξ_2^{1c}, ξ_2^{1s}, ξ_2^{2c}, and ξ_2^{2s}) are displayed by intensity maps on spheres in Figure 2, because it cannot be expressed by a simple 2D or 3D vector.

The clusters 1, 3, 4, and 5 have enhanced deviations of density from the average level ($\xi_0^{0c} = 0$) in common. They also have relatively enhanced 1st order (diurnal or north-south) and 2nd order anisotropies. However, the polarities or directions of the enhancements are different from each other, possibly causing the grouping into different clusters. Clusters 7, 8, 9, and 10 have moderate or small magnitudes of the density deviation and 2nd order anisotropy, but have comparable amplitudes of the 1st order anisotropy with the clusters mentioned above (1, 3, 4, and 5). Remained clusters, 2 and 6, have negligible density deviations, but cluster 6 has a characteristic direction of the diurnal anisotropy while the 1st order anisotropy of cluster 2 has a negligible amplitude.

Most of the average diurnal anisotropies of the clusters orient in the range between 0 p.m. and 3 p.m. in Figure 1(b1), earlier LTs than ~6 p.m. expected by the diffusion-convection picture. This phase shift is known as a common feature in a period with the positive polarity of the solar dipole magnetic field[13] as it is in our analysis period (2015-2022).

3.2 Long-term variation of occurrence rates of clusters

The number of hours labeled with each cluster is counted in each year and then transformed into a yearly occurrence rate of the cluster by dividing it by the total active hours in the year. As shown in Figure 3 we can find clear solar-activity dependences of the occurrence rates. The occurrence

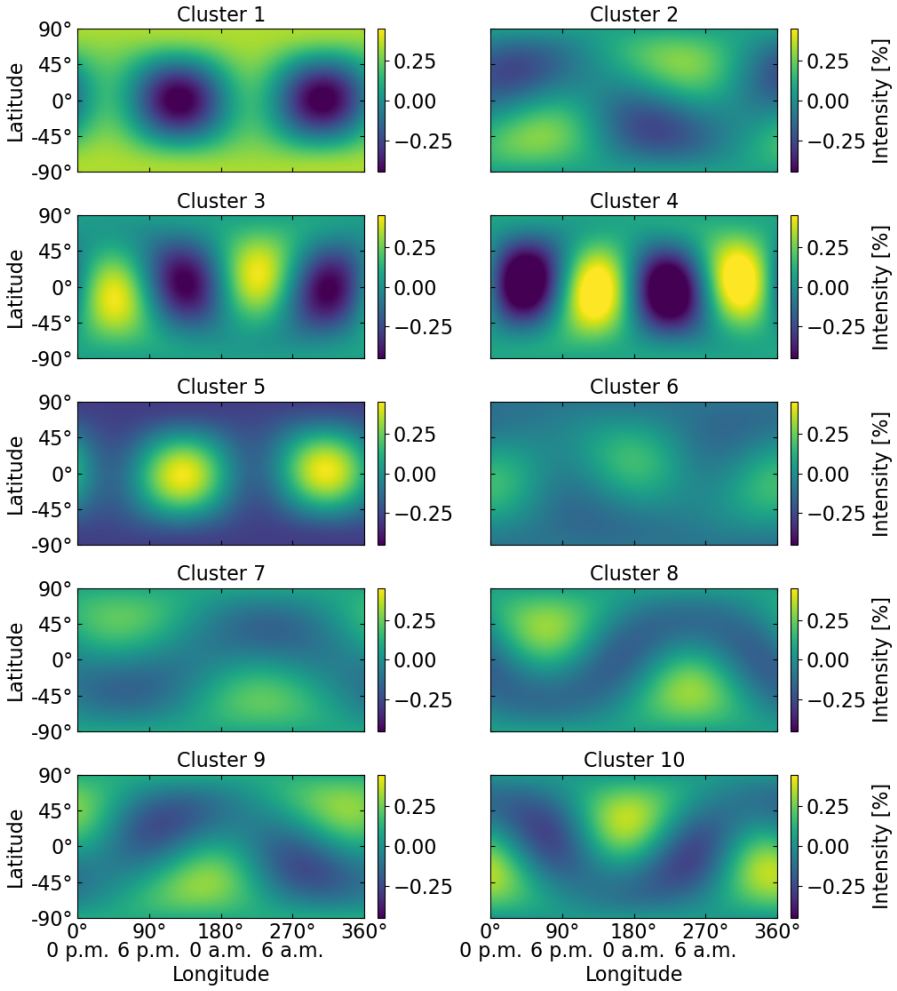

Figure 2: Average intensity maps of the 2nd anisotropy composed of ξ_2^{0c}, ξ_2^{1c}, ξ_2^{1s}, ξ_2^{2c}, and ξ_2^{2s} for individual clusters. Horizontal axes are GSE longitudes, and vertical axes are GSE latitudes.

rates of clusters 3, 4, and 7 are high near the solar activity maximum (just before 2015) and then decrease to their minimum around the solar activity minimum (2019-2020). Contrarily, clusters 2, 6, and 9 have low occurrence rates near the solar maximum and increase up to the solar minimum. Cluster 1 may also have a similar trend with relatively small significance.

The anisotropy vector simply averaged over time is proved to have its amplitude positively correlating with solar activity[13]. Our result is consistent with it overall, except for cluster 9 which has a relatively large amplitude of the 1st order anisotropy while negatively correlating occurrence rate with solar activity. Long-term variation of the heliosphere, or space climate, contains complex mechanisms such as coronal mass ejections, coronal holes, corotating interaction regions, heliospheric current sheet, and sector structure. Investigating correlations of individual clusters with such phenomena may lead to a new finding overlooked in previous works based on a simple mean of anisotropy.

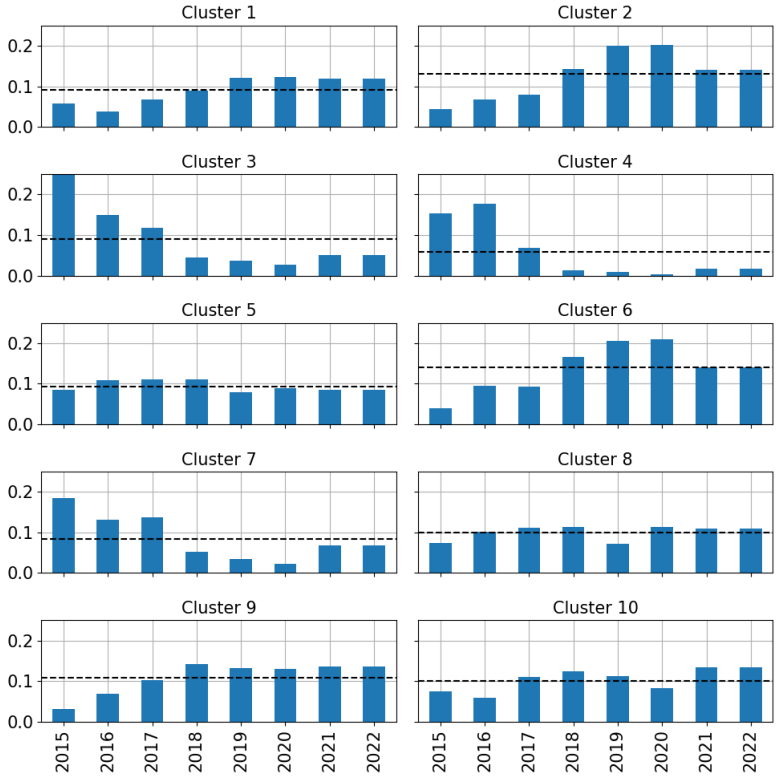

Figure 3: Yearly occurrence rate of each cluster. Horizontal dashed lines denote average occurrence rates in the whole analyzed period (2015-2022).

3.3 Occurrence rates of clusters around IP-shocks

Using an IP-shock list from CfA Interplanetary Shock Database[14], we investigate the temporal variation of the cluster occurrence rates before and after a shock arrival. Total of 81 shock events are identified in WIND satellite data in our analysis period (2015-2022). Yearly numbers of the events are 25, 6, 12, 10, 9, 4, and 15 respectively in 2015, 2016, 2017, 2018, 2019, 2020, and 2021. The event number is roughly high around the solar maximum but does not necessarily vary with the sunspot number.

We extract anisotropy data in each period of several days centered on a shock arrival at the Wind satellite. Timestamps of the data are re-assigned by hours relative to the shock arrivals with polarities negative before shock arrivals and positive after them. Then they are summed up for every 6 hours before and after shock arrivals. A cluster occurrence rate every 6 hours for all 81 events is calculated.

Figure 4 displays the occurrence rate for each cluster as a function of relative time to shock arrivals at the Wind satellite. Horizontal dashed lines are the average occurrence rates in the whole analyzed period (2015-2022) as well as Figure 3. It is notable that cluster 10 significantly enhances its occurrence rate, succeeding for ~1 day from just before shock arrivals. The occurrence rate of cluster 4 also increases from shock arrivals and keeps its enhancement over a few days. These 2 clusters, especially cluster 4, feature density depletions as shown in Figure 1(a). These enhancements are expected to represent Forbush decreases generally observed in IP-shock events.

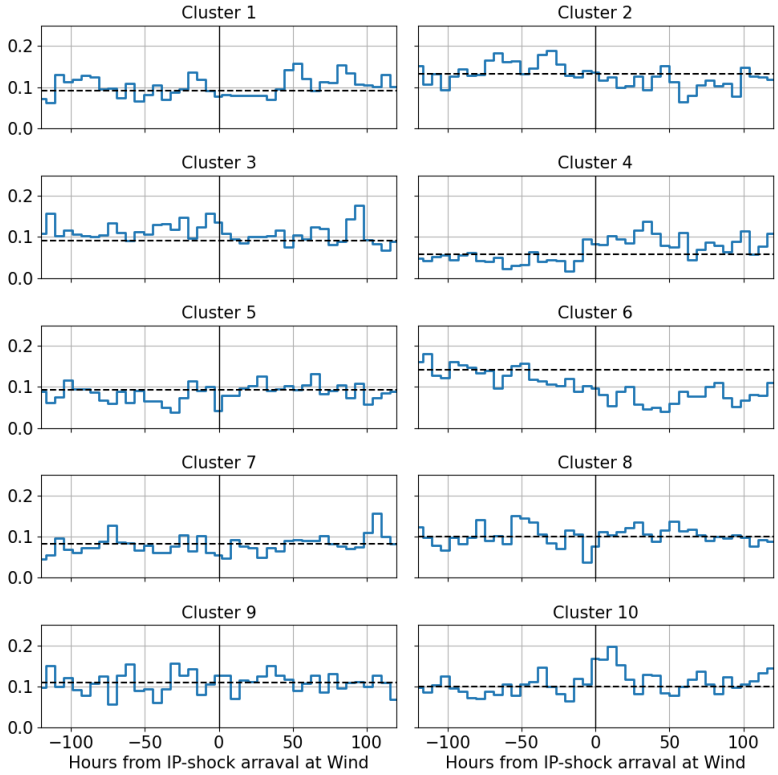

Figure 4: Occurrence rates of clusters every 6 hours before and after shock arrivals. Vertical lines at 0 hours correspond to 81 shock arrivals in the IP-shock list. Horizontal dashed lines denote average occurrence rates in the whole analyzed period (2015-2022).

Cluster 4 is also accompanied by an enhanced bi-directional flow along the Parker spiral (9 a.m. and 9 p.m.), as shown in Figure 2. Recently [15] reported an enhanced bi-directional flow inside a coronal mass ejection, and our result implies that such a bi-directional flow appears more frequently than expected. It is notable that cluster 4 also features the largest enhancement of the 1st order anisotropy. Clusters 2 and 6, both featuring negligible density deviations, show suppressions of their occurrence rates from before or just after shock arrivals, indicating density disturbances commonly appear in IP-shock events.

3.4 Conclusion

We demonstrated a data-mining approach for the study of cosmic-ray anisotropy by a preliminary analysis. The k-means clustering reduces 9 components of anisotropy into one of 10 cluster labels each time, allowing us to survey and discover an average feature of anisotropy. The temporal evolution of the cluster label shows close correlations with solar activities and IP-shock events. These results encourage future works to prove a new aspect and its physical background in anisotropy, which have been overlooked in previous works based on limited samples or simple averaging.

This work is supported by “Strategic Research Projects” grant from ROIS (Research Organization of Information and Systems). The observations are supported by Nagoya University with

the Nagoya muon detector, by INPE and UFSM with the São Martinho da Serra muon detector, by the Australian Antarctic Division with the Hobart muon detector, and by project SP01/09 of the Research Administration of Kuwait University with the Kuwait City muon detector.

References

[1] J. A. Lockwood, An investigation of the Forbush decreases in the cosmic radiation, Journal of Geophysical Research 65 (12) (1960) 3859–3880. doi:10.1029/JZ065i012p03859.

[2] M. Wada, T. Suda, Average features of cosmic ray variation associated with sudden commencement of magnetic strom, Scientific Papers of the Institute of Physical and Chemical Research 74 (1) (1980) 1–12.

[3] M. Kozai, K. Munakata, et al., AVERAGE SPATIAL DISTRIBUTION OF COSMIC RAYS BEHIND THE INTERPLANETARY SHOCK—GLOBAL MUON DETECTOR NETWORK OBSERVATIONS, The Astrophysical Journal 825 (2) (2016) 100. doi:10.3847/0004-637X/825/2/100.

[4] K. Munakata, K. Nagashima, A theory of cosmic ray anisotropies of solar origin, Planetary and Space Science 34 (1) (1986) 99–116. doi:10.1016/0032-0633(86)90107-8.

[5] GMDN collaboration, Global Muon Detector Network (GMDN) data.
URL http://hdl.handle.net/10091/0002001448

[6] M. Kozai, Space weather observations with the Global Muon Detector Network, Ph.D. thesis, Shinshu University (2015).
URL http://hdl.handle.net/10091/00018794

[7] Y. Okazaki, A. Fushishita, et al., Drift Effects and the Cosmic Ray Density Gradient in a Solar Rotation Period: First Observation with the Global Muon Detector Network (GMDN), The Astrophysical Journal 681 (1) (2008) 693–707. doi:10.1086/588277.

[8] A. Fushishita, T. Kuwabara, et al., PRECURSORS OF THE FORBUSH DECREASE ON 2006 DECEMBER 14 OBSERVED WITH THE GLOBAL MUON DETECTOR NETWORK (GMDN), The Astrophysical Journal 715 (2) (2010) 1239–1247. doi:10.1088/0004-637X/715/2/1239.

[9] A. Fushishita, T. Narumi, et al., DRIFT EFFECTS AND THE AVERAGE FEATURES OF COSMIC RAY DENSITY GRADIENT IN CIRS DURING SUCCESSIVE TWO SOLAR MINIMUM PERIODS, Advances in Geosciences 21 (2010) 199–210. doi:10.1142/9789812838209_0016.

[10] M. Kozai, K. Munakata, et al., The spatial density gradient of galactic cosmic rays and its solar cycle variation observed with the Global Muon Detector Network, Earth, Planets and Space 66 (1) (2014) 151. doi:10.1186/s40623-014-0151-5.

[11] K. Munakata, H. Miyasaka, et al., Long-term Variation of Cosmic-Ray Diurnal Anisotropy observed by a Network of Multi-directional Muon Telescopes in a wide range of Rigidity 2 (1997) 77.
URL https://ui.adsabs.harvard.edu/abs/1997ICRC....2...77M

[12] F. Pedregosa, G. Varoquaux, et al., Scikit-learn: Machine Learning in Python, MACHINE LEARNING IN PYTHON.

[13] K. Munakata, M. Kozai, et al., LONG-TERM VARIATION OF THE SOLAR DIURNAL ANISOTROPY OF GALACTIC COSMIC RAYS OBSERVED WITH THE NAGOYA MULTIDIRECTIONAL MUON DETECTOR, The Astrophysical Journal 791 (1) (2014) 22. doi:10.1088/0004-637X/791/1/22.

[14] Center for Astrophysics, Harvard and Smithsonian, CfA Interplanetary Shock Database.
URL https://lweb.cfa.harvard.edu/shocks/

[15] K. Munakata, M. Kozai, et al., Large-amplitude Bidirectional Anisotropy of Cosmic-Ray Intensity Observed with Worldwide Networks of Ground-based Neutron Monitors and Muon Detectors in 2021 November, The Astrophysical Journal 938 (1) (2022) 30. doi:10.3847/1538-4357/ac91c5.