

## Algorithm construction test for interdisciplinary science

\*Norio Umemura<sup>1</sup>, Shuji Abe<sup>2</sup>, Yoshimasa Tanaka<sup>3,4</sup>, Shin ya Nakano<sup>5</sup>, Atsuki Shinbori<sup>1</sup>, Yasuo Saito<sup>3</sup>, Yasuyuki Minamiyama<sup>6</sup>

1. Institute for Space-Earth Environmental Research, Nagoya University, 2. International Center for Space Weather Science and Education, Kyushu University, 3. National Institute of Polar Research, 4. Joint Support-Center for Data Science Research, Research Organization of Information and Systems, 5. The Institute of Statistical Mathematics, 6. National Institute of Informatics

The preliminary project in ROIS (Research Organization of Information and Systems), which consists of several researchers in ISM (The institute of Statistical Mathematics), NIPR (National Institute of Polar Research), Nagoya University and Kyushu University, has developed the data-driven method to calculate the relationship among various events indicated by several scientific data, and has confirmed whether the output results are scientifically valid or not. First of all, we have applied NIPR data (time-series data: climatology data, environmental data, magnetic field data and so on, spatial distribution data: sea water temperature data and so on by multi-site collection, and frequency distribution data: mineral composition analysis and so on) as a sample case, and obtained the following results.

In the case of time-series data on 22 Dec. 2014, through correlation analysis, we confirmed that the correlation value between the variations of radiation obtained at Skarvsnes on Soya Coast, East Antarctica and the variation of UV (ultraviolet) obtained at Skallen on Soya Coast, East Antarctica was 0.4040, and the correlation value between the variations of solar radiation and geomagnetic field was -0.0024. This result suggests scientifically that the amount of heat flowing into the earth's surface from the sun has a high interaction with that of UV radiation, but has a low relationship with temperature and/or geomagnetic variations. In the case of frequency distribution data on 1986, with Earth Mover's Distance (EMD), we confirmed the similarity value  $EMD(P, Q)$  between the distributions of the meteorite component collected at Asuka Observational Base in Antarctica and Yamato Mountains in Antarctica was 2.23, but the  $EMD(P, Q)$  of each meteorite collected at the same Asuka Observational Base was 1028.05. In this meteorite case, we have not found the spatial deviation of meteorites with similar components. However, if there is a spatial deviation, it is also possible to extract whether a peculiar phenomenon, such as temperature fluctuation, has occurred or not at that particular point. That is, this result of data-comparison, data-relation, and data-segmentation can work as an interdisciplinary system for the investigation of scientific relationships even though data format and/or science field are different, as well as a conventional comparison system.

Actually, we are able to confirm the similarity with the matrix-like combinational calculations. These results has been already registered into database, and the preparation for storing hundreds of million results is in progress. Further, this calculation refers the actual data through metadata to work across several research institutions and universities. In 2020, we will build a comprehensive virtual network of data knowledge from the same and neighboring fields to completely different fields by increasing the number of data and algorithms and expanding to distribution data such as literature data and pharmaceutical analysis data, and public this system. Then, we will contribute to the realization of an interdisciplinary science society while fulfilling the mission of providing the research infrastructure to research institutions and universities.

Keywords: Data Science, Interdisciplinary Science, Data-Driven Science