Solar Radiation and Weather Analysis of Meteorological Satellite Data by Tensor Decomposition

N. Watanabe [1], A. Ishida [2], J. Murakami [3], and N. Yamamoto [3,*]

[1] Advanced Course of Electronics and Information Systems Engineering, Kumamoto College, National Institute of Technology, Koshi, Japan; Email: ae22watanabe@g.kumamoto-nct.ac.jp (N.W.)
[2] Faculty of Liberal Arts, Kumamoto College, National Institute of Technology, Koshi, Japan; Email: ishida@kumamoto-nct.ac.jp (A.I.)
[3] Faculty of Electronics and Information Systems Engineering, Kumamoto College, National Institute of Technology, Koshi, Japan; Email: jun@kumamoto-nct.ac.jp (J.M.)
*Correspondence: naoki@kumamoto-nct.ac.jp (N.Y.)

***Abstract*—In this study, the data obtained from meteorological satellites were analyzed using tensor decomposition. The data used in this paper are meteorological image data observed by the Himawari-8 satellite and solar radiation data generated from Himawari Standard Data. First, we applied Higher-Order Singular Value Decomposition (HOSVD), a type of tensor decomposition, to the original image data and analyzed the features of the data, called the core tensor, obtained from the decomposition. As a result, it was found that the maximum value of the core tensor element is related to the cloud cover in the observed area. We then applied Multidimensional Principal Component Analysis (MPCA), an extension of principal component analysis computed using HOSVD, to the solar radiation data and analyzed the Principal Components (PC) obtained from MPCA. We also found that the PC with the highest contribution rate is related to the solar radiation in the entire observation area. The resulting PC score was compared to actual weather data. From the result, it was confirmed that the temporal transition of the amount of solar radiation in this area can be expressed almost correctly by using the PC score.**

***Keywords*—Tensor decomposition, Higher-Order Singular Value Decomposition (HOSVD), Multidimensional Principal Component Analysis (MPCA), meteorological image data, solar radiation data, Himawari-8**

I. INTRODUCTION

Earth observation satellites acquire data by remotely sensing the earth using sensors, which are onboard observation devices. Using this data, we can make regular and long-term observations over a wide area at once anywhere in the world. The various physical quantity data observed and acquired by this satellite are useful in people's lives. In recent years, the quality and quantity of satellite data have improved dramatically against the backdrop of technological innovation and an increase in the number of new companies entering the industry. Such development of the space utilization industry is expected to continue in the future. Our research aims to improve detailed feature extraction and processing speed by analyzing meteorological data from satellites using multidimensional data processing techniques.

In this research, we analyze this data using tensor decomposition. Tensor factorization is an extension of matrix factorization to tensors of 3rd-order and higher, where tensors are multidimensional arrays. There are two types of this decomposition: the Tucker decomposition, which decomposes the original higher-order tensor into the product of tensors and matrices, and the CANDECOMP/PARAFAC (CP) decomposition, which decomposes it into the sum of products of vectors. Their application areas are fuzzy modeling, signal processing, image processing, image classification, data analysis and so on [1]. Multidimensional Principal Component Analysis (MPCA) [2], which applies this decomposition, enables data analysis by obtaining principal components that represent the characteristics of the high-order tensor data to be analyzed [3, 4].

A multispectral image is a record of electromagnetic waves in multiple wavelength bands, and can represent the state of the earth's surface. In Ref. [5], tensor decomposition is applied to multispectral images acquired by Unmanned Aerial System (UAS) and used to correct cloud shadows. Furthermore, meteorological data may also be used as boundary conditions in building simulations. Rajput and Gahrooei *et al.* [6] proposed a method for creating a statistical weather model that can be used for simulations using Multiple Tensor on Tensor regression (MTOT) that applies tensor decomposition. By using this model, it is possible to predict changes in weather for each city and investigate the performance of buildings.

In this paper, we apply tensor decomposition and MPCA to image data and physical quantity data of solar

Manuscript received February 10, 2023; revised March 21, 2022, accepted May 10, 2023.

doi: 10.18178/joig.11.3.271-281

radiation acquired by meteorological satellites, and analyze the feature values and principal components obtained from these data. Section II explains the overview of the meteorological satellite data to be analyzed, and then Section III describes the tensor decomposition and Multidimensional Principal Component Analysis (MPCA) methods. Subsequently, in Section IV and Section V, we analyze features obtained from tensor decomposition and MPCA, respectively, and conclude in Section VI.

II. Overview of Acquired Metrological Data

In this study, meteorological data of weather images and solar radiation were obtained from two Research and Development Organization (R&D institutes) and used for analysis.

Firstly, the image data was obtained from the Himawari satellite data archive of the Information and Communications Technology (NICT) Science Cloud [7], which was provided by the National Institute of Information and Communications Technology (NICT). From this cloud, we can obtain image and video data for three different observation areas (Full-Disk, the Japan-area, and the Target area) taken by the Himawari-8 meteorological satellite every two minutes and 30 seconds. This time, we downloaded PNG-format color images of the Japan-area and handled them with the R package "imager" [8, 9].

Secondly, solar radiation data was provided by the interdisciplinary product provision service (P-Tree) of Japan Aerospace Exploration Agency (JAXA) [10]. With this service, geostationary meteorological satellite Himawari standard data and geophysical parameter data (Aerosol Property, Sea Surface Temperature, shortwave radiation, Chlorophyll-a, Cloud Property, Wild Fire, etc.) created from the standard data by JAXA can be obtained in near-real time. The data are updated every 10 minutes. In Section V, data related to solar radiation in NetCDF format was obtained from the geophysical data and used for the analysis [11]. The R package "ncdf4" [12] was used to handle the NetCDF data.

III. Tensor Decomposition Method and Multi-dimensional Principal Component Analysis

A. Tensors

The higher-order tensors mentioned in Section I represent multidimensional arrays, and the order of the tensor means the number of dimensions of the array. For example, a 3rd-order tensor is a 3-d array with component axes (modes) in the row, column, and depth directions. This time, we deal with a 3rd-order tensor whose row, column, and depth modes are 1-mode, 2-mode, and 3-mode, respectively.

Now, a 3rd-order tensor $\mathcal{A}$ of size $I_1 \times I_2 \times I_3$ as shown on the left side of Fig. 1 is defined by the following equation:

$$\mathcal{A} = \left(a_{i_1 i_2 i_3}\right), (i_n = 1,2,\ldots,I_n; n = 1,2,3), \qquad (1)$$

where $a_{i_1 i_2 i_3}$ is (i_1, i_2, i_3) component of $\mathcal{A}$.

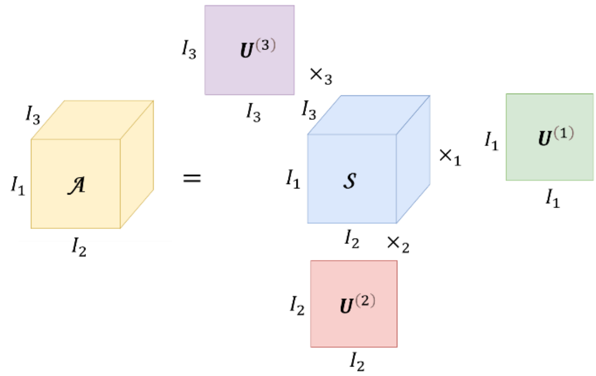


Figure 1. HOSVD of a 3rd-order tensor.

B. Higher-Order Singular Value Decomposition and N-Mode Product

More specifically, the Tucker decomposition described in Section I is a method of decomposing an original higher-order tensor into *n*-mode product of multiple matrices representing characteristics of each mode and a tensor called the core tensor (see below). Note that the *n*-mode product is an operation of multiplying a tensor and a matrix, which will be described later in this section.

Higher-Order Singular Value Decomposition (HOSVD) [13] is a technique often used to perform the Tucker decomposition. HOSVD is a generalization of Singular Value Decomposition (SVD) of matrices (2nd-order tensors) to decomposition of higher-order tensors of 3rd-order and higher. In this paper, we deal with 3rd-order tensor data, therefore HOSVD for that tensor is explained below.

The 3rd-order tensor $\mathcal{A}$ defined by Eq. (1) is decomposed by HOSVD as follows:

$$\mathcal{A} = \mathcal{S} \times_1 \boldsymbol{U}^{(1)} \times_2 \boldsymbol{U}^{(2)} \times_3 \boldsymbol{U}^{(3)}. \qquad (2)$$

In Eq. (2), $\mathcal{S}$ is a 3rd-order tensor with the same size as $\mathcal{A}$, called the core tensor, and corresponds to the singular value matrix of SVD. $\boldsymbol{U}^{(n)}$ represent orthonormal matrices of size $I_n \times I_n$, $(n = 1,2,3)$; $\times_n$, $(n = 1,2,3)$ denote operators of the *n*-mode products. Fig. 1 shows an image of this decomposition.

As for the *n*-mode product, the product in the case of the 3rd-order tensor is explained in [2, 13]. Now consider the *n*-mode product of a 3rd-order tensor $\mathcal{S}$ of size $I_1 \times I_2 \times I_3$ and each matrix $\boldsymbol{X}^{(n)}$, $(n = 1,2,3)$ of size $J_n \times I_n$. Firstly, a 1-mode product can be calculated from Eq. (3):

$$\left(\mathcal{S} \times_1 \boldsymbol{X}^{(1)}\right)_{j_1 i_2 i_3} = \sum_{i_1=1}^{I_1} s_{i_1 i_2 i_3} x_{j_1 i_1}, \qquad (3)$$
$$(i_n = 1, 2, \ldots, I_n; n = 1, 2, 3; j_1 = 1, 2, \ldots, J_1),$$

where the left side of the first line of Eq. (3) represents a (j_1, i_2, i_3) component of $\mathcal{S} \times_1 \boldsymbol{X}^{(1)}$. In the right side of Eq. (3), $s_{i_1 i_2 i_3}$ is a (j_1, i_2, i_3) component of $\mathcal{S}$, and $x_{j_1 i_1}$ is a (j_1, i_1) component of $\boldsymbol{X}^{(1)}$. From this calculation, a 3rd-order tensor $\mathcal{S} \times_1 \boldsymbol{X}^{(1)}$ of size $J_1 \times I_2 \times I_3$ can be obtained.

Similarly, 2-mode and 3-mode products are defined by the following Eqs. (4) and (5), respectively, and can be calculated as

$$\left(\mathcal{S} \times_2 \boldsymbol{X}^{(2)}\right)_{i_1 j_2 i_3} = \sum_{i_2=1}^{I_2} s_{i_1 i_2 i_3} x_{j_2 i_2}, \quad (i_n = 1, 2, \ldots, I_n; n = 1, 2, 3; j_2 = 1, 2, \ldots, J_2), \tag{4}$$

$$\left(\mathcal{S} \times_3 \boldsymbol{X}^{(3)}\right)_{i_1 i_2 j_3} = \sum_{i_3=1}^{I_3} s_{i_1 i_2 i_3} x_{j_3 i_3}, \quad (i_n = 1, 2, \ldots, I_n; n = 1, 2, 3; j_3 = 1, 2, \ldots, J_3). \tag{5}$$

For this calculation to hold, size of an n-mode of $\mathcal{S}$ and column size of $\boldsymbol{X}^{(n)}$ must be the same.

C. HOSVD Algorithm and N-Mode Matrix Unfolding

A calculation procedure of HOSVD in the case of a 3rd-order tensor is shown below [2, 13].

Algorithm 1: HOSVD of a 3rd-order tensor

[**Step 1**] Apply the n-mode matrix unfolding to a 3rd-order tensor $\mathcal{A}$ to obtain matrices $\boldsymbol{A}_{(n)}, (n = 1, 2, 3)$. This unfolding operation is described later in Algorithm 2.

[**Step 2**] By applying SVD, $\boldsymbol{A}_{(n)}, (n = 1, 2, 3)$ obtained in Step 1 are decomposed as follows:

$$\boldsymbol{A}_{(n)} = \boldsymbol{U}^{(n)} \boldsymbol{\Sigma}^{(n)} \boldsymbol{V}^{(n)\mathrm{T}}, (n = 1, 2, 3), \tag{6}$$

where $\boldsymbol{U}^{(n)}$ are the left singular matrices, $\boldsymbol{\Sigma}^{(n)}$ are the diagonal matrices with singular values on the diagonal, $\boldsymbol{V}^{(n)}$ are the right singular matrices, and T represents the transpose of a matrix.

[**Step 3**] From the 3rd-order tensor $\mathcal{A}$ and the matrices $\boldsymbol{U}^{(n)}$, $(n = 1,2,3)$ calculated in Step 2, calculate the core tensor $\mathcal{S}$ using the following equation:

$$\mathcal{S} = \mathcal{A} \times_1 \boldsymbol{U}^{(1)\mathrm{T}} \times_2 \boldsymbol{U}^{(2)\mathrm{T}} \times_3 \boldsymbol{U}^{(3)\mathrm{T}}. \tag{7}$$

[**Step 4**] Return the orthonormal matrices , $\boldsymbol{U}^{(n)}$, $(n = 1,2,3)$, the singular value matrices $\boldsymbol{\Sigma}^{(n)}$, $(n = 1,2,3)$, and the core tensor $\mathcal{S}$.

End

Next, we describe the n-mode matrix unfolding used in the first step of Algorithm 1. The matrix unfolding is an operation of transforming an original higher-order tensor into a matrix. In this paper, we use the matrix folding defined by Lathauwer *et al.* [13]. This algorithm for a 3rd-order tensor is shown below.

Algorithm 2: The n-mode matrix unfolding of a 3rd-order tensor

[**Step 1**] The matrix unfoldings of the following (i) to (iii) are performed.

(i) 1-mode matrix unfolding: As shown in Fig. 2(a), submatrices $\boldsymbol{A}_{i_2} = \left(a_{* i_2 *}\right), (i_2 = 1, 2, \ldots, I_2)$ are extracted from the 3rd-order tensor $\mathcal{A}$. Then, let $\boldsymbol{A}_{(1)}$ be a matrix in which the submatrices are arranged horizontally as

$$\boldsymbol{A}_{(1)} = \left(\boldsymbol{A}_1 \mid \boldsymbol{A}_2 \mid \cdots \mid \boldsymbol{A}_{I_2}\right). \tag{8}$$

(ii) 2-mode matrix unfolding: Extract submatrices $\boldsymbol{A}_{i_3} = \left(a_{** i_3}\right), (i_3 = 1, 2, \ldots, I_3)$ from $\mathcal{A}$ as shown in Fig. 2(b). Next, let $\boldsymbol{A}_{(2)}$ be a matrix obtained by transposing and horizontally combining the submatrices as

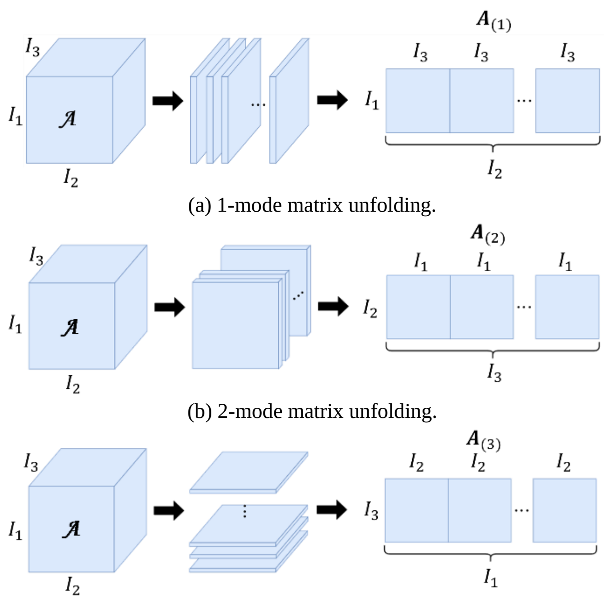
$$\boldsymbol{A}_{(2)} = \left(\boldsymbol{A}_1^{\mathrm{T}} \mid \boldsymbol{A}_2^{\mathrm{T}} \mid \cdots \mid \boldsymbol{A}_{I_3}^{\mathrm{T}}\right). \tag{9}$$

(iii) 3-mode matrix unfolding: Extract submatrices $\boldsymbol{A}_{i_1} = \left(a_{i_1 **}\right), (i_1 = 1, 2, \ldots, I_1)$ from $\mathcal{A}$ as shown in Fig. 2(c). Then, the same processing as in the 2-mode is performed to obtain a matrix $\boldsymbol{A}_{(3)}$ as

$$\boldsymbol{A}_{(3)} = \left(\boldsymbol{A}_1^{\mathrm{T}} \mid \boldsymbol{A}_2^{\mathrm{T}} \mid \cdots \mid \boldsymbol{A}_{I_1}^{\mathrm{T}}\right). \tag{10}$$

[**Step 2**] Return the matrices $\boldsymbol{A}_{(n)}, (n = 1, 2, 3)$.

End



(a) 1-mode matrix unfolding.

(b) 2-mode matrix unfolding.

(c) 3-mode matrix unfolding.

Figure 2. The n-mode matrix unfolding of a 3rd-order tensor.

In the above algorithm, the notation of submatrix $\left(a_{*i_2*}\right)$ in Step 1 means that a value of subscript i_2 is fixed and the other subscripts take all values. The other two notations have the same meaning.

D. Overview of the Multidimensional Principal Component Analysis

In matrix data with two modes of variables and individuals, principal component analysis (PCA) is a technique for obtaining principal components that represent characteristics of the data by synthesizing the variables. The individuals can be compared and classified using the obtained principal components. Eigenvalue Decomposition (EVD) and SVD are used to calculate PCA.

Multidimensional Principal Component Analysis (MPCA) is an extension of PCA to higher-order tensor data of three or more orders [2], and HOSVD is used for this calculation. In this paper, we use 3rd-order tensor data. Regarding the handling of modes of the data, two modes of the data are treated as variables, and the remaining one mode is treated as individuals. By applying MPCA, we can obtain principal components that show features of the data combining the variables of the two modes. Similar to PCA, it is possible to compare and classify the individuals using the principal components of interest. In Section V, MPCA is used to analyze time-series data of solar radiation.

E. Calculation of MPCA

Firstly, preprocessing of acquired 3rd-order tensor data is performed. At this time, if there are missing values in the obtained data, they are complemented by a method described later. If there are no missing values, the original data is used as is. Now let $\mathcal{C}$ be a 3rd-order tensor of size $I_1 \times I_2 \times I_3$.

Secondly, $\mathcal{C}$ is normalized by the following equation and transformed into a 3rd-order tensor $\mathcal{A}$ of size $I_1 \times I_2 \times I_3$.

$$a_{i_1 i_2 i_3} = \frac{c_{i_1 i_2 i_3} - \bar{c}_{i_1 i_2}}{s_{i_1 i_2}}, \quad (i_n = 1, 2, \ldots, I_n; n = 1, 2, 3). \tag{11}$$

In Eq. (11), $a_{i_1 i_2 i_3}$ shows an (i_1, i_2, i_3) component of $\mathcal{A}$, and $c_{i_1 i_2 i_3}$ denotes an (i_1, i_2, i_3) component of $\mathcal{C}$. $\bar{c}_{i_1 i_2}$ and $s_{i_1 i_2}$ are the mean and standard deviation, respectively, given by

$$\bar{c}_{i_1 i_2} = \frac{1}{I_3} \sum_{i_3=1}^{I_3} c_{i_1 i_2 i_3}, \quad s_{i_1 i_2} = \sqrt{\frac{1}{I_3} \sum_{i_3=1}^{I_3} (c_{i_1 i_2 i_3} - \bar{c}_{i_1 i_2})^2}, \quad (i_n = 1, 2, \ldots, I_n; n = 1, 2). \tag{12}$$

In the case of data where $s_{i_1 i_2} = 0$, only centering is performed using the following equation instead of Eq. (11).

$$a_{i_1 i_2 i_3} = c_{i_1 i_2 i_3} - \bar{c}_{i_1 i_2}, \quad (i_n = 1, 2, \ldots, I_n; n = 1, 2, 3). \tag{13}$$

The above preprocessing is performed with the 1-mode (row direction) and 2-mode (column direction) of $\mathcal{A}$ as variables, and the 3-mode (depth direction) as individuals. As can be seen from the calculations of Eqs. (11) to (13), the individual data of the 3-mode of $\mathcal{A}$ are standardized or centralized.

Thirdly, with $\mathcal{A}$ as input for Algorithm 1, HOSVD is used to obtain the orthonormal matrices $\boldsymbol{U}^{(n)}$, $(n=1,2)$ and the singular value matrices $\boldsymbol{\Sigma}^{(n)}$, $(n=1,2)$. $\boldsymbol{U}^{(1)}$ is a principal component matrix representing features of the 1-mode of $\mathcal{A}$, and $\boldsymbol{U}^{(2)}$ is a principal component matrix with characteristics of the 2-mode. By calculating the *n*-mode product of these two principal component matrices and $\mathcal{A}$, a tensor $\mathcal{B}$ with principal component scores of individuals of the 3-mode can be obtained as shown in the following equation.

$$\mathcal{B} = \mathcal{A} \times_1 \boldsymbol{U}^{(1)\mathrm{T}} \times_2 \boldsymbol{U}^{(2)\mathrm{T}}. \tag{14}$$

In Eq. (14), a subvector $\boldsymbol{b}_{i_1 i_2} = \left(b_{i_1 i_2 *}\right)$ of $\mathcal{B}$ stores each individual score of an (i_1, i_2) th principal component obtained using an i_1 th principal component of the 1-mode and an i_2 th principal component of the 2-mode. Note that $\left(b_{i_1 i_2 *}\right)$ indicates that values of indices i_1 and i_2 are fixed and an index of 3-mode take all values.

Lastly, contribution rates $r_k^{(n)}$, $(n=1,2)$ of *k*th principal components of the *n*-mode can be calculated by the following equation using diagonal components of the singular value matrices $\boldsymbol{\Sigma}^{(n)}$, $(n=1,2)$.

$$r_k^{(n)} = \frac{\sigma_k^{(n)2}}{\sum_{j=1}^{I_n} \sigma_j^{(n)2}}, (n=1,2), \tag{15}$$

where $\sigma_j^{(n)}$, $(j = 1, 2, \ldots, I_n; n = 1, 2)$ are *j*th diagonal elements of the matrices $\boldsymbol{\Sigma}^{(n)}$ shown in the following equation.

$$\boldsymbol{\Sigma}^{(n)} = \begin{bmatrix} \sigma_1^{(n)} & 0 & \cdots & 0 \\ 0 & \sigma_2^{(n)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{I_n}^{(n)} \end{bmatrix}, (n=1,2). \tag{16}$$

IV. Analysis of Metrological Image Data Using HOSVD

A. Meteorological Image Data Used

In this chapter, meteorological image data taken by the Himawari-8 described in Section II are acquired from the data archive of the cloud and analyzed using HOSVD.

In this analysis, image data from April 1 to 10, 2021 and May 1 to 10, 2021 at 12:00 am were used. Fig. 3 shows an example of acquired image data of size 2701 rows by 3301 columns. The area over Kyushu Island (hereafter simply write Kyushu) is the subject of the analysis, and data of the area was used by cutting out a submatrix of size 300 rows by 250 columns, starting from a component of 1451 row and 1051 column in the data of Fig. 3. An example of image data over Kyushu is shown in Fig. 4.

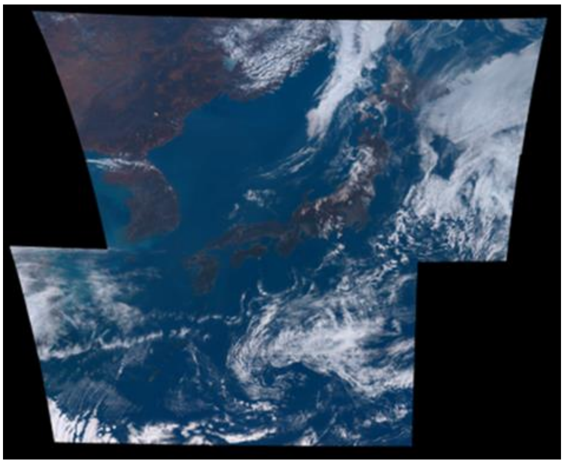


Figure 3. Metrological image data (Japan-area).

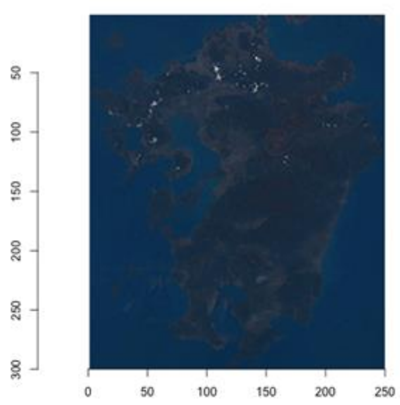


Figure 4. Metrological image data (Kyusyu).

The data used here is the RGB color image, which is a 3rd-order tensor data consisting of three gray-scale images with R, G, and B components. The data size is $300\times250\times3$. The function "HOSVD" provided by the rTensor package in R was used to calculate the HOSVD [14]. This function can compute an approximation called truncated HOSVD [15], which can be obtained as a core tensor of any size below the size of a 3rd-order tensor given to the function.

B. Analysis of Features of Meteorological Image Data Extracted from HOSVD

Firstly, we applied HOSVD to the data of Kyushu from April 1 to 10, 2021 at 12:00 am, and compared the distribution of the obtained core tensor element values.

The size of the core tensor acquired by HOSVD was set to $50\times50\times3$. The absolute values of the obtained tensor elements were taken, and they were sorted in descending order. The results of this calculation are shown in Fig. 5. In this figure, the vertical axis represents the absolute values of the core tensor elements, and the horizontal axis shows the rank of the absolute values.

Table I shows the values of rank 1 for each day in Fig. 5. In this table, the core tensor values are sorted in ascending order. The images over Kyushu at 12:00 am for 10 days in April 2021 are shown in Fig. 6. In this table, the core tensor value is the smallest on April 7, and the image in Fig. 6 shows that there is almost no cloud cover (clear weather) over the entire Kyushu on this day. In contrast, the largest core tensor value is found on April 4, which indicates that the entire area is covered by clouds. That is, as the core tensor values in Table I increase, the amount of cloud in the images in Fig. 6 also tends to increase.

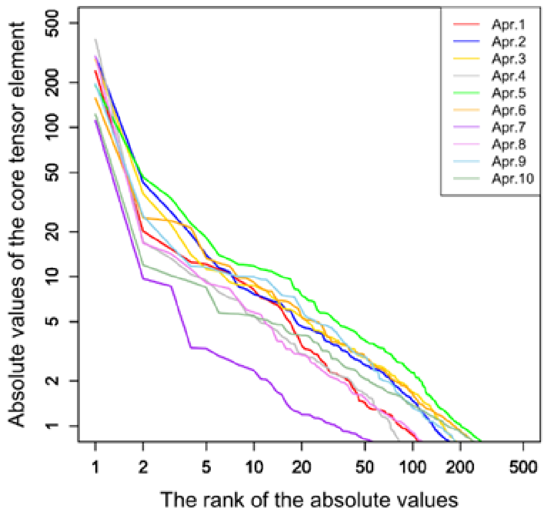


Figure 5. Distribution of absolute values of core tensor elements.

TABLE I. Values of Rank 1 in Fig. 5

Date	Maximum absolute value of core tensor elements
Apr. 7, 2021	111.11
Apr. 10, 2021	122.72
Apr. 6, 2021	156.71
Apr. 5, 2021	191.96
Apr. 9, 2021	195.61
Apr. 1, 2021	238.79
Apr. 3, 2021	286.38
Apr. 8, 2021	299.54
Apr. 2, 2021	299.87
Apr. 4, 2021	390.21

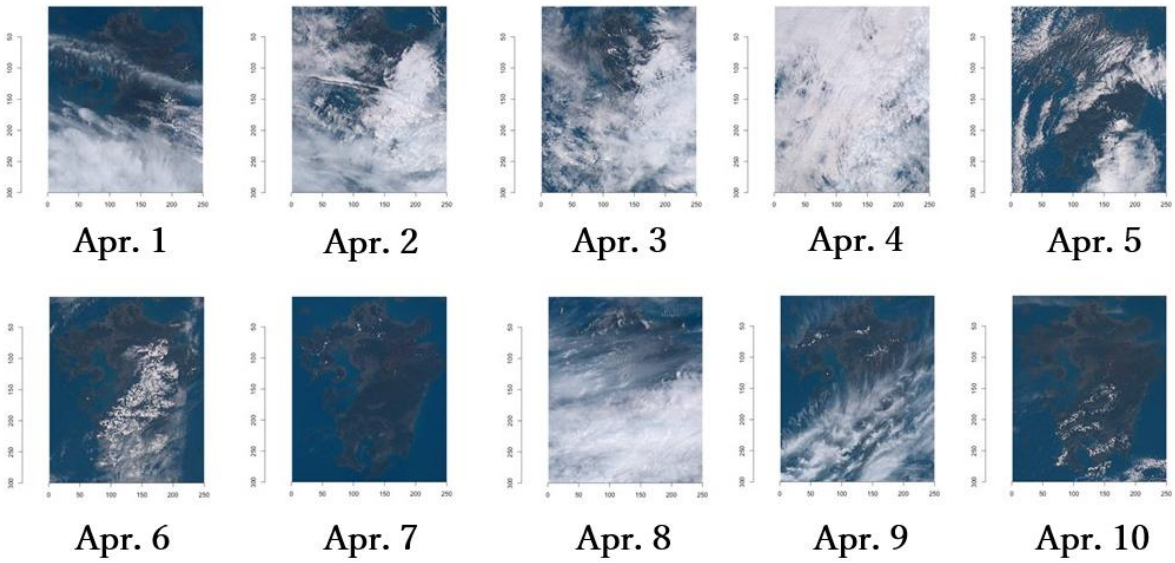


Figure 6. Image data over Kyushu (April 2021).

Furthermore, image data and core tensor values at 12:00 am for 10 days from May 1, 2021 are shown in Fig. 7 and Table II, respectively. From these results, it was found that the relationship between the core tensor value and the amount of cloud in the image has the same tendency as the results in April. Therefore, the maximum absolute value of the core tensor element obtained from HOSVD is related to the amount of cloud cover in the observation area.

Moreover, regarding the distribution after the second rank in Fig. 5, the values tend to be distributed lower on April 7 and 10, when there is little cloud cover. On the other hand, for other dates, the values are distributed highest on April 5, when the entire area is not covered with clouds, and it is considered that there is little relationship with cloud cover.

TABLE II. MAXIMUM VALUE OF CORE TENSOR ELEMENTS (MAY 2021)

Date	Maximum absolute value of core tensor elements
May 6, 2021	89.10
May 3, 2021	106.08
May 9, 2021	116.25
May 10, 2021	124.47
May 1, 2021	268.05
May 2, 2021	285.79
May 4, 2021	286.64
May 5, 2021	363.00
May 8, 2021	372.43
May 7, 2021	398.55

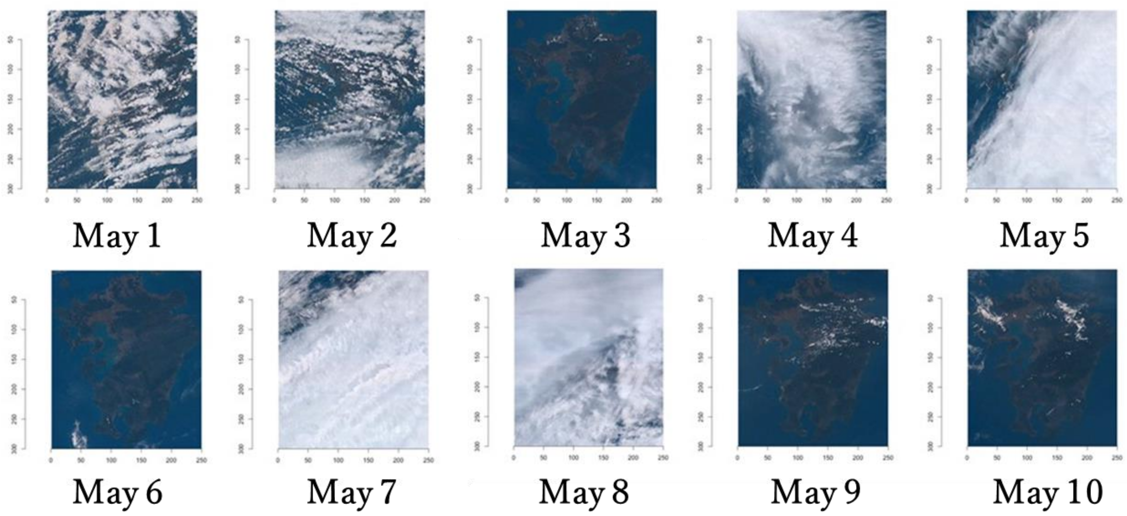


Figure 7. Image data over Kyushu (May 2021).

C. *Attempt to Determine Sunny/Cloudy Using Core Tensors*

In the previous section, it was found that the largest absolute value of a core tensor element is related to the cloud amount in the metrological image. Cloud amount is the percentage of clouds in the sky, and the Japan Meteorological Agency defines the cloud amount of 0 to 1 as clear, 2 to 8 as sunny, and 9 to 10 as cloudy. In this section, we attempted to determine the sunny/cloudy using the core tensor values described in the previous section. The devised algorithm for determination them is shown below.

Algorithm 3: Determination of Sunny/Cloudy

[**Step 1**] For each input data, the maximum absolute value of the core tensor element is obtained by HOSVD. Find the maximum value max and minimum value min from the obtained values and let the difference between them be $d = max - min$.

[**Step 2**] The two thresholds $Th1$, $Th2$ are obtained by the following equation:

$$\begin{aligned} Th1 &= min + 0.1d, \\ Th2 &= min + 0.9d. \end{aligned} \tag{17}$$

[**Step 3**] Using the threshold obtained in Step 2, data with core tensor values smaller than $Th1$ are determined to be clear weather, data with core tensor values larger than $Th2$ are determined to be cloudy, and otherwise are determined to be sunny.

End

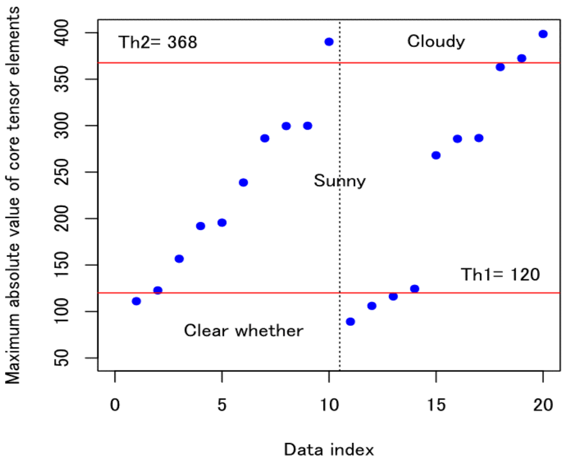


Figure 8. Determination result by Algorithm 3.

As input data for Algorithm 3, we used the image data of Figs. 6 and 7 shown in the previous section. The decision results of this algorithm are shown in Fig. 8. The horizontal axis represents indices attached to the data in Tables I and II. The indices from 1 to 10 represent the data in Table I and those from 11 to 20 correspond to the data in Table II. Based on thresholds $Th1 = 120$, $Th2 = 368$, the data on April 4 and May 7 and 8 were determined to be cloudy. Furthermore, the data on April 7, May 3, 6, and 9 were judged to be clear weather. All other data were determined to be sunny. With regard to these determination results, when the amount of cloud in the data in Figs. 6 and 7 was visually confirmed, it is considered that the determination was generally correct.

However, although this determination method is simple, it has the disadvantage that the threshold cannot be determined well when there are few patterns of cloud states in the input data. Therefore, when using this method to determine sunny/cloudy, it is necessary to increase the number of data as much as possible so that various cloud conditions are included in the input data. At least, by inserting data of the case where the observation area is mostly cloudless and the case where it is covered with clouds into the input data, as shown in Figs. 6 and 7, it is considered that the threshold value is set relatively well, and the judgment can be performed satisfactorily.

V. Analysis of Solar Radiation Data Using MPCA

A. *Solar Radiation Data Used*

Compared to data such as sea surface temperature, which cannot be obtained in areas with clouds, solar radiation can be obtained in the entire observation area, therefore there are few missing values, and it is considered easy to construct higher-order tensor data. In this chapter, the solar radiation data described in Section II is analyzed using MPCA.

This time, we obtained NetCDF data related to solar radiation from the JAXA Himawari Monitor. It contains data such as total atmosphere optical thickness, total atmosphere angstrom exponent, photosynthetically active radiation, shortwave radiation, ultraviolet-A radiation, and ultraviolet-B radiation [10, 11]. We extracted matrix data of the shortwave radiation (solar radiation) using the R package “ncdf4” and analyzed the obtained data.

B. *Configuration of Tensor Data for Solar Radiation*

Firstly, the solar radiation data described in above section is obtained for 24 hours. The data are observed every 10 minutes; therefore 144 data are acquired in 24 hours. Then, the matrix data of the solar radiation at each time is stacked in the 3-mode as shown in Fig. 9 to construct a 3rd-order tensor.

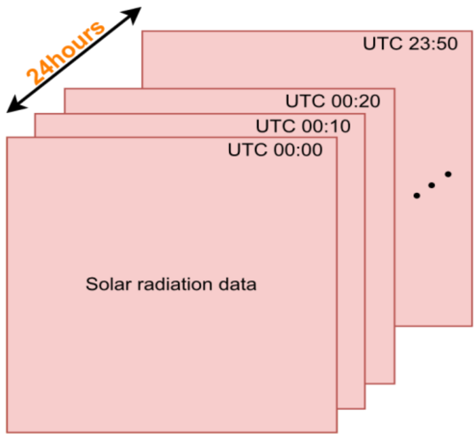


Figure 9. Structure of solar radiation tensor data.

Note that the size of the tensor in the 1-mode and 2-mode depends on the size of an observation area. In this paper, Kyushu and Kumamoto Pref. areas were targeted as the observation areas. The size of the original solar radiation data is 2401 rows and 2401 columns. For the data of the Kyushu area, the submatrix of size 71 rows and 61 columns was used with 521 rows and 981 columns of the original data as the origin. For the Kumamoto Pref. area, the submatrix with the size of 21 rows and 14 columns was obtained and used with the component of 538 rows and 1008 columns as the starting point. Fig. 10 shows an example of a color map of the solar radiation seen from above the Kyushu area created from an extracted data.

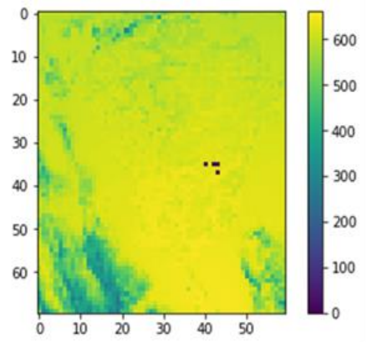


Figure 10. Color map of solar radiation seen from above Kyushu (12:00 on Jan. 1, 2022, Unit: W/m²).

In addition, the solar radiation data observed by the meteorological satellite may contain missing values due to unmeasured portions. Therefore, missing values are complemented for each data acquired at each time. This completion process repeats a process of searching for a missing value in matrix data at a certain time and replacing it with an average value of 8 neighboring elements of an element in which the missing value is found. This processing is performed for the matrix data at all times. After the processing, it is checked whether data for all times are available. Then, if there is no data at a certain time, an average matrix data is obtained from matrix data before and after that time, and a missing matrix data is replaced with the average data.

C. Analysis of Features of Solar Radiation Data Extracted from MPCA

We used solar radiation data for the Kyushu area for 24 hours from 9:00 am on January 1, 2022. The 3rd-order tensor data described in above section was constructed and MPCA was applied to it. The size of this tensor data is $71 \times 61 \times 144$.

The contribution rates of 1st principal components of the 1-mode (row direction) and 2-mode (column direction) of the tensor data obtained from the MPCA were calculated by (15) and was approximately 98.7% in both components. Therefore, we decided to compare the data at each time on the $(1,1)$ th principal component obtained by synthesizing these principal components.

This comparison can be performed by confirming a component value (principal component score) corresponding to each time stored in a partial vector $\boldsymbol{b}_{11} = (b_{11*})$ of the principal component score tensor $\mathcal{B}$ obtained from the calculation of (14). Hence, as shown in Fig. 11, it is visualized by a graph with the time on the horizontal axis and the component values of $\boldsymbol{b}_{11}$ on the vertical axis.

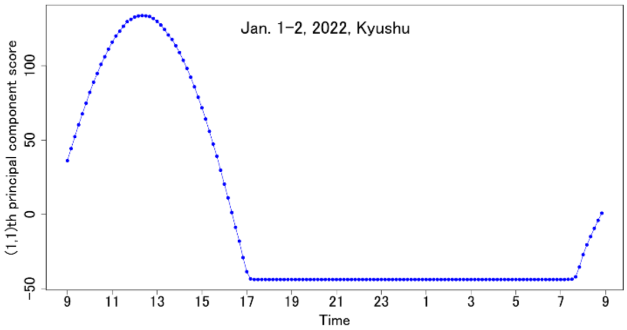


Figure 11. $(1,1)$ th principal component score at each time (Jan. 1-2, 2022, Kyushu).

From Fig. 11, it can be seen that the principal component score is large when the sun is rising and is small when the sun is setting, although exact times of sunrise and sunset are not known due to the wide area of Kyushu. Thus, the $(1,1)$ th principal component is considered to represent a feature reflecting the solar radiation in the entire observation area.

D. Comparison with Past Meteorological Data

To confirm whether the $(1,1)$ th principal component obtained in the previous section is a feature related to the solar radiation in the observed area, we compared it with past meteorological data. As calculation examples, we used the solar radiation data for the summer solstice (Jun. 21–22, 2021) and winter solstice (Dec. 22–23, 2021) in Kumamoto Pref. as the observation area. The size of the tensor data used for the MPCA calculation is $21 \times 14 \times 144$. Also, values of solar radiation observed on the earth were obtained from the website of "Past Meteorological Data Search" published by the Japan Meteorological Agency [16]. The values are shown in the columns of global solar radiation G in Tables III and IV.

Fig. 12(a) plots the $(1,1)$ th principal component scores at the summer and winter solstices. In this figure, the horizontal and vertical axes show time and the principal component scores, respectively. The blue and green plots correspond to the scores obtained from the data of summer and winter solstices, severally.

In this verification, we used hourly integrated global solar radiation observed by the Kumamoto Local Meteorological Observatory. Therefore, the principal component scores in Fig. 12(a) were integrated every hour. In Tables III and IV, the integrated values are shown as Isc. With reference to the calculation method of the daily radiant energy in [17], each value of Isc was calculated by the following equation when there are m pieces of score data sc_k, $(k = 1, 2, \ldots, m)$ within a certain hour.

TABLE III. COMPARISON OF INTEGRATED VALUE OF (1,1) TH PRINCIPAL COMPONENT SCORE AND GLOBAL SOLAR RADIATION (JUN. 21–22, 2021, KUMAMOTO PREF)

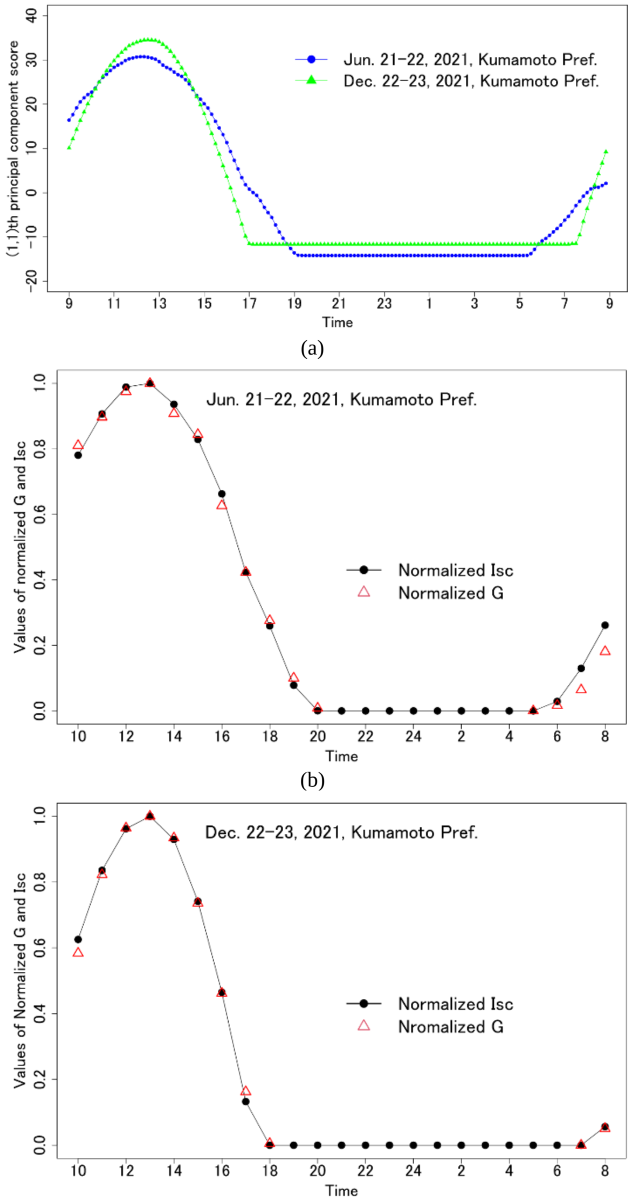
Day, Time	*Isc*	Normalized *Isc*	*G* [MJ/m^2]	Normalized *G*
Jun. 21				
10:00	7.42E+04	0.78	2.91	0.81
11:00	9.45E+04	0.91	3.22	0.90
12:00	1.08E+05	0.99	3.50	0.97
13:00	1.09E+05	1.00	3.59	1.00
14:00	9.92E+04	0.94	3.26	0.91
15:00	8.20E+04	0.83	3.03	0.84
16:00	5.53E+04	0.66	2.25	0.63
17:00	1.69E+04	0.42	1.52	0.42
18:00	−9.43E+03	0.26	0.99	0.28
19:00	−3.85E+04	0.08	0.36	0.10
20:00	−5.10E+04	0.00	0.03	0.01
21:00	−5.11E+04	0.00	None	
22:00	−5.11E+04	0.00		
23:00	−5.11E+04	0.00		
Jun. 22				
0:00	−5.11E+04	0.00		
1:00	−5.11E+04	0.00		
2:00	−5.11E+04	0.00		
3:00	−5.11E+04	0.00		
4:00	−5.11E+04	0.00		
5:00	−5.11E+04	0.00	0.00	0.00
6:00	−4.64E+04	0.03	0.06	0.02
7:00	−3.02E+04	0.13	0.23	0.06
8:00	−9.11E+03	0.26	0.65	0.18

TABLE IV. COMPARISON OF INTEGRATED VALUE OF (1,1) TH PRINCIPAL COMPONENT SCORE AND GLOBAL SOLAR RADIATION (DEC. 22–23, 2021, KUMAMOTO PREF)

Day, Time	*Isc*	Normalized *Isc*	*G* [MJ/m^2]	Normalized *G*
Dec. 22				
10:00	6.17E+04	0.63	1.15	0.58
11:00	9.65E+04	0.84	1.62	0.82
12:00	1.18E+05	0.96	1.90	0.96
13:00	1.24E+05	1.00	1.97	1.00
14:00	1.12E+05	0.93	1.84	0.93
15:00	8.10E+04	0.74	1.45	0.74
16:00	3.51E+04	0.47	0.91	0.46
17:00	−2.00E+04	0.13	0.32	0.16
18:00	−4.19E+04	0.00	0.01	0.01
19:00	−4.19E+04	0.00	None	
20:00	−4.19E+04	0.00		
21:00	−4.19E+04	0.00		
22:00	−4.19E+04	0.00		
23:00	−4.19E+04	0.00		
Dec. 23				
0:00	−4.19E+04	0.00		
1:00	−4.19E+04	0.00		
2:00	−4.19E+04	0.00		
3:00	−4.19E+04	0.00		
4:00	−4.19E+04	0.00		
5:00	−4.19E+04	0.00		
6:00	−4.19E+04	0.00		
7:00	−4.19E+04	0.00	0.00	0.00
8:00	−3.26E+04	0.06	0.10	0.05

$$Isc = \left(\frac{1}{m}\sum_{k=1}^{m} sc_k\right) \times 3600 \tag{18}$$

For example, consider the calculation of *Isc* for Jun. 21 at 10:00 am in Table III. First, an average score is obtained from $m = 6$ score data every 10mins from 9:10 to 10:00 in the blue plot in Fig. 12(a). Then, the *Isc* is obtained by multiplying the average score by 3600s. The *Isc* at other times can also be calculated in the same way.



(a) (b) (c)

Figure 12. Comparison between (1,1) th principal component score and past meteorological data (Kumamoto Pref.): (a) The (1,1) th principal component score at each time (Jun. 21–22, Dec. 22-23, 2021); (b) Comparison in summer solstice data (Jun. 21–22, 2021); (c) Comparison in winter solstice data (Dec. 22-23, 2021).

Then, to compare with actual meteorological data, the values of *Isc* and global solar radiation *G* were normalized so that the minimum value is 0 and the maximum value is 1. The normalized values are shown in Tables III and IV. In Fig. 12(b) and Fig. 12(c), the normalized values of *Isc* and *G* are plotted with black dots and red triangles, respectively, for comparison. From these results, it can be seen that a trend of change of the two normalized values is well matched. When the correlation coefficient was actually obtained from the data paired with these two values, a high correlation was confirmed with 0.997 in the case of Fig. 12(b) and 0.999

in the case of Fig. 12(c).

Conversely, the $(1,1)$ th principal component score itself was found to express characteristics of instantaneous solar radiation in the observation area, since the integrated value matches the global solar radiation. Originally, we dealt with short-wave radiation data with an energy intensity per unit time. Therefore, it was confirmed again that the feature value obtained by applying MPCA to the data is also instantaneous like this radiation.

VI. CONCLUSION

In this paper, we analyzed the weather image data and solar radiation data acquired by the weather satellite using tensor decomposition and MPCA. The results were as follows.

First, in the analysis of weather image data, multiple-colored images were obtained over Kyushu, and the 3rd-order tensor data was composed and analyzed for each image. The analysis used HOSVD, one of the tensor decompositions, and examined the characteristics of this data from the distribution of the obtained core tensor. As a result, it was found that the absolute value of the maximum value of the core tensor element is related to the amount of cloud in the observation area.

Therefore, we devised and implemented a simple algorithm to determine whether the observation area is sunny or cloudy using this value. The result of applying this algorithm to the data above Kyushu was almost correct. Note that this determination algorithm is an observation result from the sky because it uses data taken by the weather satellite, which may be different from the observation of cloud volume from the ground. Furthermore, since this algorithm uses only feature values related to cloud cover, it may be judged as sunny or cloudy even if it is raining or snowing in the observation area. If it is necessary to judge rain, snow, etc., other meteorological data must also be used.

Next, in the solar radiation analysis, we used the solar radiation data calculated by JAXA based on the data observed by meteorological satellites. The data used for the analysis was the 3rd-order tensor data acquired every 10 mins for 24 h in the observation area of Kyushu or Kumamoto prefecture. Then, MPCA using HOSVD was applied to this tensor data.

Analyzing the principal components that represent the characteristics of the calculation result data obtained from MPCA, it was found that the principal component with the highest contribution rate was related to the amount of solar radiation in the entire observation area. Looking at the results of comparison with actual meteorological data, it can be said that the time-series changes in the solar radiation in the observation area can be expressed almost correctly using this principal component.

As mentioned in Section I, tensor decomposition is used to process multispectral image data and statistical meteorological data in the fields of remote sensing and architecture and so on, and its application range is wide. We believe that our research can also play a role in these efforts.

Finally, the future work is to combine weather image data analysis with physical quantity data such as solar radiation to enable more accurate weather determination. In addition, we would like to work on similarity search of weather images and physical quantity data, and speeding up the search.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Nonoka Watanabe made a plan for the research, performed calculations and analyzed the results, and shared the writing of the manuscript of the paper; Akio Ishida shared the writing of the manuscript and made the final compilation; Jun Murakami supervised the research and shared the writing of the manuscript; Naoki Yamamoto supervised the research and shared the writing of the manuscript, and check the whole one; all authors had approved the final version.

FUNDING

This work was carried out in part of a grant from the Nozaki Wakaba Foundation under the theme of “Improvement of tensor decomposition calculation method and creation of multi-dimensional data processing educational content using it”.

ACKNOWLEDGMENT

Research product of the amount of solar radiation data (produced from Himawari-8) that was used in this paper was supplied by the P-Tree System, Japan Aerospace Exploration Agency (JAXA). The weather image data is collected with the NICT (National Institute of Information and Communications Technology) Science Cloud. We would like to thank those Agency and Institute.

REFERENCES

[1] L. D. Lathauwer, “A survey of tensor methods,” in *Proc. of IEEE International Symposium on Circuits and Systems*, 2009, pp. 2773–2776.

[2] J. Murakami, M. Hino, N. Yamamoto, and A. Ishida, *Introduction to Multidimensional Data Analysis Using the Statistical Software R: Hypothesis Testing, Analysis of Variance, and Principal Component Analysis*, Tokyo: Nisshin Shuppan, 2017, chap. 4, pp. 234–244. (in Japanese)

[3] N. Yamamoto, K. Matsuo, J. Murakami, A. Ishida, D. Furushima, S. Saito, and N. Kano, “Rehabilitation data analysis by Tucker-2 model and comparison with that by nonnegative Tucker-2 model,” in *Proc. of the International Conference on Mechanics and Mechatronics*, 2015, pp. 1140–1153.

[4] A. Ishida, K. Kawakami, D. Furushima, N. Yamamoto, and J. Murakami, “Analysis of relationships between amount of physical activity of patients in rehabilitation and their ADL scores using multidimensional PCA,” *Advances in Intelligent Systems and Computing*, vol. 690, pp. 147–158, 2017.

[5] S. Wang *et al.*, “Unmanned aerial system multispectral mapping for low and variable solar irradiance conditions: Potential of tensor decomposition,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 155, pp. 58–71, 2019.

[6] M. Rajput, M. R. Gahrooei, and G. Augenbroe, “A statistical model of the spatial variability of weather for use in building simulation practice,” *Building and Environment*, vol. 206, 108311, 2021.

[7] NICT science cloud, the Himawari-8 image data. [Online]. Available: https://sc-web.nict.go.jp/himawari/himawari8-image-archive.html
[8] R. C. Team. (2012). *R: A Language and Environment for Statistical Computing*, Vienna, Austria: R Foundation for Statistical Computing. [Online]. Available: https://www.R-project.org/
[9] S. Barthelme, D. Tschumperle, J. Wijffels, H. E. Assemlal, and S. Ochi. Package imager. [Online]. Available: https://cran.r-project.org/web/packages/imager/imager.pdf
[10] The p-tree system, the amount of solar radiation data produced from Himawari-8. [Online]. Available: https://www.eorcved.jaxa.jp/ptree/index_j.html
[11] R. Frouin and H. Murakami, "Estimating photosynthetically available radiation at the ocean surface from ADEOS-II global imager data," *Journal of Oceanography*, vol. 63, pp. 493–503, 2007.
[12] D. Pierce. Package ncdf4. [Online]. Available: https://cran.r-project.org/web/packages/ncdf4/ncdf4.pdf
[13] L. D. Lathauwer, B. D. Moor, and J. Vandewalle, "A multilinear singular value decomposition," *SIAM Journal on Matrix Analysis and Applications*, vol.21, no. 4, pp. 1253–1278, 2000.
[14] J. Li, J. Bien, and M. Wells. Package rTensor. [Online]. Available: https://cran.r-project.org/web/packages/rTensor/rTensor.pdf
[15] N. Vannieuwenhoven, R. Vandebril, and K. Meerbergen, "A new truncation strategy for the higher-order singular value decomposition," *SIAM Journal on Scientific Computing*, vol. 34, no. 2, pp. 1027–1052, 2012.
[16] The Japan meteorological agency, historical weather data search service. [Online]. Available: http://www.data.jma.go.jp/obd/stats/etrn/index.php
[17] Eko Instruments Co., Ltd. Pyranometer MS-80SH—Instruction Manual Ver. 1. [Online]. 27. Available: https://www.eko-instruments.com/media/xxsfunul/ms-80sh_manual-eng-08 2022.pdf