Abstract A clustering and classification method using a Gaussian mixture model (GMM) is used to summarize and simplify meteorological data from a numerical weather prediction (NWP) model. Each horizontal grid in the integration domain of the NWP model is characterized by a feature vector, which consists of a multivariable with multiple pressure levels. All horizontal grids at every forecast time are classified based on the GMM clustering. The classification results show that grids are clustered into air masses or disturbances with the same meteorological characteristics. This paper describes application of the proposed classification method as a framework to verify the forecast of surface solar irradiance from the NWP model. Satellite observation data are used as the reference so that verification can be performed over the integration domain of the NWP model for each air mass or disturbance that moves and changes shape over time. The mean square error (MSE) is decomposed into the square of the mean error and the MSE between variables centered on zero, the square root of which is called the centered root mean square error (CRMSE). The analyses are performed for forecast data over a 2 day forecast horizon. The change in mean error is not significant until the second day, whereas the CRMSE is maintained only during the first day. Each air mass has a different forecast error structure. The proposed framework clarifies the structure of the forecast error of the surface solar irradiance.

1. Introduction

Numerical weather prediction (NWP) models provide a large amount of meteorological forecast data. The datasets contain data about meteorological variables, which are often defined on grid points. Although much information can be obtained from these datasets, the data size is huge. However, there is redundancy in the datasets because there is a correlation between meteorological variables and atmospheric conditions on a grid are strongly connected to those on neighboring grids. When the meteorological fields are analyzed, changes in all meteorological variables in the three-dimensional grid space should be considered simultaneously. Thus, a method for simplifying and summarizing the data is needed in order to understand the characteristics of weather conditions in three dimensions.

Clustering is used to identify groups with similar characteristics with no prior information (Wilks, 2011). One aim of clustering is to provide summarized information in a simple form. Many meteorology and climatology studies have used clustering. Pernin et al. (2016) and Vrac et al. (2005) performed cluster or classification analyses of air columns to identify air masses in the global field. In their analysis vertical profiles of temperature and humidity feature air columns. Pernin et al. (2016) used copulas and Vrac et al. (2005) used a Gaussian mixture model (GMM) as core methods in their classification algorithms. In the present work, clustering with the GMM is used as a main part of the algorithm. The grids in the integration domain of the NWP model are classified into clusters with the same meteorological characteristics. As a result, air masses and disturbances can be marked up as clusters.

The forecast quality of the surface solar irradiance is verified by using the proposed method. Cloud is the most important factor determining surface solar irradiance. Cloud properties are related to the atmospheric conditions where the cloud is formed (Cotton et al., 2011), and cloud accompanies air masses and...
disturbances. Thus, the representation of cloud or surface solar irradiance should be evaluated following the air masses and disturbances. The results of the classification help to identify air masses and disturbances. Ohtake et al. (2015) and Perez et al. (2013) evaluated the forecast quality of the surface solar irradiance at fixed points, namely, at ground observation stations. However, the proposed approach gives the forecast quality for each air mass or disturbance, which enables us to understand the forecast quality information that accompanies air masses and disturbances. We show not only the evaluation of forecast quality for each air mass, but also differences in changes in forecast quality between different air masses as the forecast time progresses.

The rest of the paper is organized as follows. Section 2 presents information on the data used in this work. Clustering and classification with the GMM and the proposed algorithm are discussed in Section 3. The results of the proposed method and the framework for evaluating the forecast quality of surface solar irradiance are reported in Section 4. Section 5 gives the conclusions.

2. Data

2.1. Weather Forecast Data From the NWP Model

The grid point value (GPV) weather forecast data product provided by the Japan Meteorological Agency (JMA) from the mesoscale meteorological (MSM) model is used. The GPV forecast data of the MSM covers Japan and neighboring areas in East Asia (Figure 1). The JMA MSM model is an operational non-hydrostatic model based on a newly developed model called ASUCA (A System based on a Unified Concept for Atmosphere; Japan Meteorological Agency, 2014). Forecasting using ASUCA started in February 2017.

The meteorological variables on the three-dimensional grids in the data product are zonal, meridional, and vertical wind speed, geopotential, specific humidity, and temperature. The data for the variables at pressure levels of 300, 400, 500, 600, and 700 hPa are used. These levels are in the free atmosphere where friction from the ground can be ignored. These variables are used as the feature vector for the cluster analysis. The aim of the clustering is to survey atmospheric conditions, especially synoptic-scale weather phenomena. The variables for the lower troposphere and near surface are likely affected by smaller-scale phenomena characteristic of the local area and strongly related to the surface conditions and the topography. Thus, these variables are not used as features for cluster analysis.

Surface solar irradiance included in the data product is given as an averaged value over 1 h on horizontal grids. Forecast data for the surface solar irradiance started to be delivered in December 2017. The surface solar irradiance data is standardized using downward solar irradiance at the top of atmosphere. This standardized variable is called the clearness index (CI) (Woyte et al., 2007). Due to this modification, the diurnal and seasonal cycles related to the Earth’s rotation and revolution are reduced. The downward solar irradiance at the top of atmosphere is included in the AMATERASS data product mentioned below.

Figure 1. (a) Classification of data at FCST 3 with an initial time of 06:00 JST on May 3, 2018 and (b) filtered classification results. Colors represent clusters. Boundaries between clusters are shown as white lines.
The MSM forecast data product for the year 2018 is used. The forecast cycle of the MSM forecast for the year is eight forecasts per day. The initial times of the forecasts are every 3 h from 00:00 JST. The forecast horizon ranges from 1 to 39 h after the initial time. The forecast data intervals are 1 and 3 h for surface solar irradiance and three-dimensional variables, respectively. The forecast data at time $H$ JST is defined as the averaged value over 1 h from $H$ JST (i.e., from $H:00$ to $H:59$) in this work. For simplicity, the forecast time $f$ hours after the initial time is abbreviated as FCST $f$.

The horizontal resolution of the variables on the three-dimensional grid is about 10 km and that for the surface solar irradiance is about 5 km. The surface solar irradiance data is arranged on the 10 km grid using the smoothing and subsampling technique called image pyramids. In the procedure, the mean of the surface solar irradiance over $2 \times 2$ grids on a 5 km grid is computed and arranged on the 10 km grid, the center of which coincides with the center of the $2 \times 2$ grids. The latitude grid size is 253 and the longitude grid size is 241. The total number of horizontal grids is 60,973.

2.2. Solar Irradiance Data From Geostationary Satellite Observations

The AMATERASS dataset includes data about cloud properties, solar irradiance, meteorological variables, and photovoltaic electric output. The cloud properties and solar irradiance data are based on observations from the JMA geostationary satellite Himawari-8 (Bessho et al., 2016) and have a data interval of 2.5 min and spatial resolution of about 1 km over Japan. An advantage of the dataset is that full-resolution data for every observation can be provided. Fast computation of the surface solar irradiance is achieved by an algorithm for emulating the radiative transfer model using a neural network (Takenaka et al., 2011). The surface solar irradiance and the irradiance at the top of the atmosphere are used and rearranged on the 10 km horizontal grid to fit the three-dimensional MSM data. The averaged value over 1 h is used.

The present version of the algorithm does not include radiative effects of aerosols. The surface solar irradiance of the AMATERASS dataset has been evaluated by Damiani et al. (2018). They compared surface solar irradiance derived from AMATERRASS datasets and ground observations at four observation stations in Japan. They reported that the mean error (ME) ranged from 20 to 30 W/m², which means that Himawari-8 overestimated the ground-based observations; the root mean square error (RMSE) ranged from 80 to 100 W/m²; and the correlation coefficient was 0.95–0.96. Additionally, they evaluated the aerosol effect as 10–15 W/m² under clear-sky conditions measured with the ME. The surface solar irradiance data from the AMATERASS dataset is used as the reference for verifying the forecast quality of the NWP model. Thus, the verification results in this work may be different from those using ground observation data as a reference.

3. Methodology

3.1. Clustering With the GMM

The GMM is a linear combination of different Gaussian models. Clustering with the GMM is a model-based approach, in which the GMM is fitted to the test data, and the number of clusters is equal to the number of GMM components. The expectation–maximization (EM) algorithm is used to conduct maximum likelihood estimation for the GMM (Dempster et al., 1977). The EM algorithm consists of expectation and maximization steps, which are iterated until the likelihood reaches its maximum. In the EM algorithm, latent variables or missing variables are introduced to simplify the maximum likelihood solution. For mixture model problems, the latent variable is the indicator variable, which is a binary variable and represents from which mixture model components an observation comes. A detailed description of clustering with the GMM is given by Bishop (2006). The advantage of clustering with the GMM for meteorological data is that the correlation structure of the feature vector is considered, which decreases the redundancy of information in the feature vector.

When the GMM is used for classification, the results of the clustering are used as the classification rule. For a test data vector, the dissimilarity is measured with the probability that the test data come from a component in the GMM. The test point is assigned to the cluster with the highest probability.

To determine the number of components, the Bayesian information criterion (BIC) is used as the complexity measure. The BIC considers the fitness of the data for the estimated distribution and the complexity of the mode, and a smaller BIC indicates a more suitable model. However, the results of the analyses do not show the minimum for the testing range (Figure 2). The BIC decreases gradually as the number of components
increases. For clustering, the number of groups should not be so large as to become too complex. The procedure to determine the number of components is as follows.

1. The change in the BIC from \( n \) and \( n + 1 \) components is fitted with an inverse exponential curve as a function of the number of components, \( n \).
2. The ratios of the change in BIC to the change at \( n = 2 \) (i.e., the change from two to three components) are computed.
3. When the ratio at \( n \) becomes smaller than 0.1, \( n \) is determined as the number of components of the GMM, namely, the number of clusters.

Each feature has a different contribution to the classification. Removing features with lower contributions can increase the simplicity without degrading the classification. The relative importance of features for the classification can be evaluated using random forest classification (Breiman, 1996; Hastie et al., 2009; Louppe et al., 2013). The importance of a variable is measured with the mean decrease in impurity importance. The results of the classification of the GMM are used as the training data for the random forest, and the Gini index is used as the impurity measure.

The clustering with the GMM is performed twice in the proposed method. The classification rule and number of clusters are determined in the first clustering analysis with the GMM. The features selection using the random forest is made after the first clustering analysis. Then, the second clustering analysis is performed using a feature vector, which consists of only variables with higher importance. The classification rule and the number of clusters are updated. The results of the second cluster analysis are used in the following analyses.

The procedure for clustering and classification with the GMM is executed using Scikit-learn Python packages (Pedregosa et al., 2011).

### 3.2 Algorithm

The algorithm for the proposed method consists of three main parts (Figure 3).

![Figure 2](image-url)  
*(a) BIC as a function of the number of components of the GMM. (b) Change in BIC as the number of components increases. The number of components on the horizontal axis is that before the increase. The filled point is the point chosen according to the procedure.*

**Figure 3.** Overview of the forecast verification framework.
1. The GMM clustering is performed for data in each month. Data for four initial times, 00:00, 06:00, 12:00, and 18:00 JST, on every day of the month are used in the clustering. The training data are the forecast data 3 h after the initial time (FCST 3). Forecast data with large errors are poor training data because the classification rule is affected by the unrealistic weather conditions. It is assumed the forecast error does not inflate until FCST 3. One thousand horizontal grid points, which is about 1/60 of the total number of grids, are chosen randomly from the test data, and then feature vectors at randomly chosen grids from all test data are analyzed by the GMM clustering.

2. The results from the GMM clustering are applied as the classification rule. All grid points from all data in the month are classified according to the rule (Figure 1a).

3. The result of the GMM clustering is noisy. The two-dimensional median filter is applied five times to smooth out the data, and air masses appear (Figure 1b).

3.3. Evaluation of the Forecast Quality of Surface Solar Irradiance With the Moving Air Masses

The metrics used to measure the forecast quality over all grids in each air mass are the ME and the RMSE,

\[ \text{ME} = \frac{1}{n} \sum_{i=1}^{n} (f_i - s_i) = \bar{f} - \bar{s} \]

\[ \text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (f_i - s_i)^2} \]

where \( f_i \) and \( s_i \) are CIs from the NWP model and satellite observations at grid \( I = \{1, \ldots, n\} \) and \( n \) is the number of grids in a cluster domain.

MSE, which is the square of the RMSE, can be decomposed into four components (Murphy, 1988).

---

**Figure 4.** Summary of variables of the feature vector in each cluster. Panels (a) to (f) show features from C0 to C5. The variables are named with a letter from the abbreviation of the meteorological variable and the pressure level. The meteorological variables are abbreviated as follows: g, geopotential height; u, horizontal wind speed; t, air temperature. The pressure levels are abbreviated using the number of hundreds. For example, geopotential height at 300 hPa is abbreviated to g3.
Here, $sd_f$ and $sd_s$ are the standard deviation for the NWP forecast and satellite observation, respectively, and $r_{fs}$ is Pearson’s correlation coefficient.

Using the ratio of the standard deviation of CI from the NWP to that from the satellite observation, $\alpha = sd_f / sd_s$, MSE is written

$$\text{MSE} = (\text{ME})^2 + (1 + \alpha)sd_i^2 - 2\alpha sd_f sd_s r_{fs}$$

To remove the effect of ME from RMSE the centered root mean square error (CRMSE; RMSE for the centered or unbiased variable) is also used.

$$\text{CRMSE} = \sqrt{(\text{MSE} - \text{ME}^2)} = \sqrt{(1 + \alpha)sd_i^2 - 2\alpha sd_f^2 r_{fs}}$$

ME measures the difference in the mean surface solar irradiance and the cloud thickness over the domain. Ratio $\alpha$ measures the agreement of nonuniformity of the surface solar irradiance (cloud thickness) over the domain. The correlation coefficient summarizes the concordance of the surface solar irradiance between the NWP model and satellite observations at every grid in the domain. Consistency in cloud formation, cloud displacement, and differences in cloud thickness directly cause lower concordance.
4. Results and Applications

4.1. Detection and Tracing of Air Masses

The results of the classification in May 2018 are used in the analyses in this section. There is a seasonal cycle in the atmospheric conditions. To reduce the effect of the seasonal cycle on the cluster analysis, the cluster analysis is performed for each month. In May, the forecast quality of the surface solar irradiance from the MSM model is better than at other times during the year (Fonseca et al., 2020). However, in May, Japan often experiences aerosol events and there may be an uncertainty of about 10 W/m² in the reference data (Damiani et al., 2018).

The number of clusters is set to six according to the procedure for determining the components of GMM based on the BIC (Figure 2). Clusters 0, 1, ..., 5 are named C0, C1, ..., C5. The efficient features, which are chosen using feature selection using the random forest, are the following 12 meteorological variables: geopotential height at 300, 400, 500, 600, and 700 hPa; horizontal wind speed at 300 and 400 hPa; and temperature at 300, 400, 500, 600, and 700 hPa.

Distributions of efficient variables for the clustering are visualized using a boxplot in Figure 4. Each variable is centered on zero and standardized with its standard deviation. There are clear correlations between

Figure 6. Comparison of the mean of the CI between FCST 3 and the other forecast times for each atmospheric cluster. (a) FCST 6, (b) FCST 9, (c) FCST 27, (d) FCST 30, and (e) FCST 33. The green shaded boxes indicate that the difference is significant at the 95% confidence level with the two-tailed rank sum test.
variables at different pressure levels and the signs of the same variables at different levels are the same. This indicates that the clustering rule tends to emphasize the barotropic structure. Although the variables can be decreased further or the feature vector can be compressed into a low-dimensional vector, the decrease and compression are not performed in this work. This is because the GMM classifier considers the correlation structure between each pair of variables (Hastie 2009).

Figure 5 shows the classification results for forecast data in the daytime from FCST 3 to 33, for an initial time of 06:00 JST on May 3. The low-pressure system develops at FCST 3 (09:00 JST), moves eastward, and becomes mature after 24 h. Cluster C1 is west of the low-pressure system at the developing stage and lies in the low-pressure system at the mature stage. C1 is characterized by lower geopotential height and temperature and easterly wind speed. C2 appears from the north to east of the low-pressure system at the developed stage and surrounds the low-pressure system after the mature stage. C5, which has a higher temperature than C1, is in the southeast of the low-pressure system at the development stage, but is far from the low-pressure system at the mature stage. C4 appears in the western area of the low-pressure system.

Meteorological variables include the meridionally changing properties, such as the meridional

Figure 7. Comparison of ME between FCST 3 and the other forecast times for each atmospheric cluster. (a) FCST 6, (b) FCST 9, (c) FCST 27, (d) FCST 30, and (e) FCST 33. The green shaded boxes indicate the difference is significant at the 95% confidence level with the two-tailed rank sum test.
temperature or geopotential height gradients. Cluster masses tend to appear in a specific zone. C0 remains in the southern part of the domain. C3 tends to appear in the middle latitude of the domain, and the features in C3 are similar to the monthly means. The classification results for every 3 forecast hours from FCST 3 to FCST 33 are shown in Figure S1 of the Supporting Information.

### 4.2. Application for Evaluating the Forecast Quality of Surface Solar Irradiance

The forecast quality of the surface solar irradiance from the NWP is examined by applying the results of the classification. We focus on the changes in the forecast quality of each cluster of atmospheric conditions as the forecast time progresses. The verification at FCST3 is compared with that at other FSCTs. The results for data with an initial time of 06:00 JST are used, which cover daytime from 09:00 to 15:00 JST on two forecast days. The forecast day of the initial time is called Day d and the next day is called Day d + 1.

Results in this subsection are visualized using boxplots (Figures 4–9). Every boxplot is drawn as follows. The box is drawn from the first quartile to the third with a horizontal line at the median. Lower and upper whiskers are extended to −1.5 inner quartile range from the bottom of the box and +1.5 inner quartile range from

---

**Figure 8.** Comparison of CRMSE between FCST 3 and the other forecast times for each atmospheric cluster. (a) FCST 6, (b) FCST 9, (c) FCST 27, (d) FCST 30, and (e) FCST 33. The green shaded boxes indicate the difference is significant at the 95% confidence level with the two-tailed rank sum test.
The mean of the CI from the NWP model of each atmospheric cluster is shown in Figure 6. C4 has the strongest irradiance of all the clusters. There are grids with clear sky and thin clouds in this area (Figure 5). C3 has the smallest surface solar irradiance, although the length of the boxplot is longer, which indicates that there are variations in cloud type. Comparison of the surface solar irradiance of the same cluster between different forecast hours reveals that there is a diurnal cycle in the surface solar irradiance, in which the irradiance at 15:00 JST (i.e., FCST 9 and 33) becomes smaller than that at 09:00 JST and 12:00 JST. A similar diurnal cycle can be found in data from the satellite observations (figure is omitted).

Figure 9. Comparison of ratio $\alpha$ between FCST 3 and the other forecast times for each atmospheric cluster. (a) FCST 6, (b) FCST 9, (c) FCST 27, (d) FCST 30, and (e) FCST 33. The green shaded boxes indicate the difference is significant at the 95% confidence level with the two-tailed rank sum test.
The ME of every cluster is negative, which means that the CI from the NWP is underestimated compared with that from the satellite observations (Figure 7). The comparison of the ME between FCST 3 and other forecast times shows that the ME is maintained in almost all clusters over two forecast days. Only C3 shows a significant change at FCST 9 and 33, indicating that there is a diurnal cycle.

For the CRMSE, there is clear change in forecast quality with forecast horizon (Figure 8). The CRMSE is maintained during Day $d$. However, the forecast quality degrades for C0, C3, and C5 at all forecast times on the next day, as does C4 at 15:00 JST on the next day. The difference in the forecast quality for C1 and C2 is not substantial, although the CRMSE of both clusters on the next day is always larger than the reference forecast time. This is possibly because the number of samples is not large enough to draw statistically confident conclusions.

The change in the correlation coefficient contributes to the change in the CRMSE more than the change in ratio $\alpha$ (Figures 9 and 10). Ratio $\alpha$ does not change significantly.

| Cluster | Probability |
|---------|-------------|
| C0      | 100.0       |
| C1      | 100.0       |
| C2      | 100.0       |
| C3      | 100.0       |
| C4      | 100.0       |
| C5      | 100.0       |

Table 1: Probability of the joint event for ME

For the CRMSE, there is clear change in forecast quality with forecast horizon (Figure 8). The CRMSE is maintained during Day $d$. However, the forecast quality degrades for C0, C3, and C5 at all forecast times on the next day, as does C4 at 15:00 JST on the next day. The difference in the forecast quality for C1 and C2 is not substantial, although the CRMSE of both clusters on the next day is always larger than the reference forecast time. This is possibly because the number of samples is not large enough to draw statistically confident conclusions.

The change in the correlation coefficient contributes to the change in the CRMSE more than the change in ratio $\alpha$ (Figures 9 and 10). Ratio $\alpha$ does not change significantly.
The correlation coefficient decreases gradually with forecast time (Figure 10). Only the change in C2 is not significant at all forecast times on the next day.

The above analyses concern the distribution of each metric in different atmospheric clusters. The results can be interpreted as the persistency of the characteristics of forecast quality for each cluster. However, they do not address time sequential change in the forecast quality. The final analysis calculates the time sequential change of the probability that the difference in the metric between FCST 3 and other forecast time FCST \( f \) remains in the allowed range.

| Cluster | Probability |
|--------|-------------|
| FCST 3 | 100.0       |
| FCST 6 | 100.0       |
| FCST 9 | 100.0       |
| FCST 27| 100.0       |
| FCST 30| 100.0       |
| FCST 33| 100.0       |

The results suggest that the information about the forecast quality measured at a test time can help with decisions on whether the forecast data should be relied on after the test time. Additionally, the empirical uncertainty of the forecast for each air mass and disturbance can be learned from the results of the classification. Under the standard in this work, the information about forecast quality from both ME and CRMSE during the day is reliable, with probabilities higher than 80% for ME and 70% for CRMSE. However, the limitation of the analysis in this work is that the period of the data is for 1 year. To obtain more robust verification, data with longer periods should be analyzed. The verification results for the forecast in May are shown in Table 2.

5. Conclusions

We proposed a method to summarize and simplify the multidimensional and multivariable data from the NWP model. The method is based on clustering and classification with the GMM. The results indicate that the proposed method can be used to verify the forecast of surface solar irradiance from the NWP model.

The GMM classification can be used to detect air masses and disturbances that move and change shape over time. The forecast quality of the surface solar irradiance from the NWP model was investigated as an application of the proposed method. The forecast quality was evaluated for each cluster or moving air mass using satellite data as reference data.

The forecast quality was measured with the decomposed MSE, which allowed the factors in the forecast error to be evaluated separately. Information about the forecast quality measured with ME did not change greatly on two forecast days. However, CRMSE was only maintained during Day \( d \). The correlation term in CRMSE changed significantly with forecast time and different forecast quality characteristics were found for each atmospheric cluster.

The verification results for the forecast in May are shown in Table 2.
this paper; however, the method can be used for other months and seasons. The multi-year verification will be performed in future work.

Our method can also be used to analyze other meteorological and climatological data over other regions or globally. Modifications, such as the choice of variables and parameters in the algorithm, should be made according to the purpose of the analysis.

**Data availability statements**

Data for the global solar irradiance at the surface and at the top of atmosphere from the AMATERASS data product are used in this work. This data product is available for research, but user registration is required. To use the AMATERASS data product for research, please contact the Solar Radiation Consortium (http://amaterass.org, accessed 2020-5-25).

**References**

Bessho, K., Date, K., Hayashi, M., Ikeda, A., Imai, T., Inoue, H., et al. (2016). An introduction to Himawari-8/9 - Japan’s new-generation geostationary meteorological satellites. *Journal of Meteorological Society of Japan*, 94(2), 151–183. https://doi.org/10.2151/jmsj.2016-009

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. New York, NY: Springer.

Breiman, L. (1996). Random forests. *Machine Learning*, 45, 5–32.

Cotton, W. R., Bryan, G. R., & Van den Heever, S. C. (2011). *Storm and Cloud Dynamics 2nd Edition*. New York, NY: Academic Press.

Damiani, A., Irie, H., Horio, T., Takamura, T., Khatri, P., Takenaka, H., et al. (2018). Evaluation of Himawari-8 surface downwelling solar radiation by SKTNET observations. *Atmospheric Measurement Techniques Discussions*, 11(4), 2501–2521. https://doi.org/10.5194/amt-2017-440

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of Royal Statistical Society: Series B*, 39(1), 1–22. https://doi.org/10.2307/2296718

Fonseca, J. G. S. Jr., Uno, F., Ohtake, H., Oozeki, T., & Ogimoto, K. (2020). Enhancements in day-ahead forecasts of solar irradiation with machine learning: A novel analysis with the Japanese Mesoscale model. *Journal of Applied Meteorology and Climatology*, 59(5), 1011–1028. https://doi.org/10.1175/JAMC-D-19-0240.1

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning* (2nd ed.). New York, NY: Springer. https://doi.org/10.1007/978-0-387-84958-7

Japan Meteorological Agency (2014). The next generation nonhydrostatic model Asuca (in Japanese). NWP Section Tech. Supplement (Vol. 60), Japan Meteorological Agency, Tokyo, Japan. Retrieved from https://www.jma.go.jp/jma/kishou/books/nwpreport/60/No60_all.pdf

Louppe, G., Wehenkel, L., Sutera, A., & Geurts, P. (2013). Understanding variable importance in forests of randomized trees. *Advances in Neural Information Processing Systems*, 26, 431–439.

Murphy, A. H. (1988). Skill scores based on the mean square error and their relationships to the correlation coefficient. *Monthly Weather Review*, 116(12), 2417–2424. https://doi.org/10.1175/1520-0493(1988)116<2417:SSBOTES>2.0.CO;2

Ohtake, H., Fonseca, J. G. S., Takashima, T., Oozeki, T., Shimose, K., & Yamada, Y. (2015). Regional and seasonal characteristics of global horizontal irradiance forecasts obtained from the Japan meteorological agency mesoscale model. *Solar Energy*, 116, 83–99. https://doi.org/10.1016/j.solener.2015.03.020

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.

Perez, R., Lorenz, E., Pelland, S., Beauharnois, M., Van Knowe, G., Henker, K., et al. (2013). Comparison of numerical weather prediction solar irradiance forecasts in the US, Canada and Europe. *Solar Energy*, 94, 305–326. https://doi.org/10.1016/j.solener.2013.05.005

Perlman, J., Vrac, M., Crevoisier, C., & Chédin, A. (2016). Mixture model-based atmospheric air mass classification: A probabilistic view of thermodynamic profiles. *Advances in Statistical Climatology Meteorology and Oceanography*, 2(2), 115–136. https://doi.org/10.5194/acsmo-2-115-2016

Takenaka, H., Nakajima, T. Y., Higurashi, A., Higuchi, A., Takamura, T., Pinker, R. T., & Nakajima, T. (2011). Estimation of solar radiation using a neural network based on radiative transfer. *Journal of Geophysical Research*, 116, D08215. https://doi.org/10.1029/2010JD013337

Vrac, M., Chédin, A., & Diday, E. (2005). Clustering a global thermodynamic profile: A mixture decomposition of copulas. *Journal of Atmospheric and Oceanic Technology*, 22(10), 1445–1459. https://doi.org/10.1175/JTECH1179.1

Wilks, D. S. (2011). *Statistical Methods in the Atmospheric Sciences* (3rd ed.). New York, NY: Academic Press.

Woyte, A., Belmans, R., & Nijss, J. (2007). Fluctuation in instantaneous clearness index: Analysis and statistics. *Solar Energy*, 81(2), 195–206. https://doi.org/10.1016/j.solener.2006.03.001