Classification and forecast of heavy rainfall in northern Kyushu during Baiu season using weather pattern recognition

Dzung Nguyen-Le,1* Tomohito J. Yamada1 and Duc Tran-Anh2

1Faculty of Engineering, Hokkaido University, Japan
2National Center for Hydro-Meteorological Forecasting, Hanoi, Vietnam

Abstract

In this study, the Self-Organizing Maps in combination with K-means clustering technique are used for classification of synoptic weather patterns inducing heavy rainfall exceeding 100 mm day$^{-1}$ during the Baiu season (June–July) of 1979–2010 over northern Kyushu, southwestern Japan. It suggests that these local extreme rainfall events are attributed to four clustered patterns, which are primarily related to the Baiu front and the extratropical/tropical cyclone/depression activities and represented by the intrusion of warm and moist air accompanied by the low-level jet or cyclonic circulation. The classification results are then implemented with the analogue method to predict the occurrence (yes/no) of local heavy rainfall days in June–July of 2011–2016 by using the prognostic synoptic fields from the operational Japan Meteorological Agency (JMA) Global Spectral Model (GSM). In general, the predictability of our approach evaluated by the Equitable Threat Score up to 7-day lead times is significantly improved than that from the conventional method using only the predicted rainfall intensity from GSM. Although the false alarm ratio is still high, it is expected that the new method will provide a useful guidance, particularly for ranges longer than 2 days, for decision-making and preparation by weather forecasters or end-users engaging in disaster-proofing and water management activities.

Keywords: self-organizing map; heavy rainfall; clustering; forecast; Baiu; Kyushu

Introduction

During the Baiu season (June–July; JJ) in northern Kyushu, southwestern Japan (Figure 1), heavy rain frequently occur causing flooding and serious damages to life and properties. Thus, it is critical to provide an early prediction and warning for such rainfall events in the region. Generally, issuing an alert for the occurrence of locally intensive rain is mainly based on the predicted rainfall amount, e.g. from numerical models. Nowadays, owing to the progress in computing performance and atmospheric modelling, numerical models can run operationally with horizontal resolution down to a few kilometres. This fine resolution allows to explicitly resolving small-scale processes such as deep convection and orography effect, improves dramatically the predictability of rainfall. Nevertheless, it has always been challenging to accurately forecast rainfall in practical, especially for ranges longer than 2 days, owing to the imperfection of models and constraint of computational expenses. Meanwhile, medium-range (3–7 days) forecasts are crucial for reducing the impact of extreme rainfall by providing more time for decision-making and preparation. However, the coarse spatial resolution of numerical models used in medium-range forecast, typically larger than 10 km, limits their ability to predict the rainfall amount exactly.

Rainfall has nonlinear relationships with various meteorological factors and many local heavy rainfall events correspond to large-scale atmospheric condition (e.g. Baiu front). Classification and identification of such synoptic weather patterns (WP) are thus not only fruitful for understanding the genesis of local heavy rainfall but also potentially improving the forecasting capability. In this framework, an artificial neural network learning mechanism such as self-organizing map (SOM; Kohonen, 1982) can be an effective tool for analysing the atmospheric data (e.g. Nishiyama et al., 2007; Ohba et al., 2015, 2016). In this study, the SOM approach is used to objectively classify the anomalous WPs inducing heavy rainfall in Fukuoka-Saga prefectures, northern Kyushu, during the Baiu season. Further, the results are used with the analogue method to predict the occurrence (yes/no) of heavy rainfall days over the region. The analogue method (Lorenz, 1969) is based on the assumption that if the current WP is similar to those of historical observation, the local rainfall can be similar to that of the past.

The reminder of this paper is organized as follows. Section 2 briefly describes the data and our method. The main results are given in Section 3. Finally, Section 4 presents conclusions.

© 2017 The Authors. Atmospheric Science Letters published by John Wiley & Sons Ltd on behalf of the Royal Meteorological Society.
Datasets and methodology

Datasets

The European Centre for Medium-Range Weather Forecasts ERA-Interim reanalysis (Dee et al., 2011) on a 0.75° grid during JJ 1979–2011 is utilized for atmospheric variables. We also use the gridded dataset of APHRO_JP V1207 (Kamiguchi et al., 2010) from 1979 to 2010 and the Japan Meteorological Agency (JMA) radar/rain gauge-analysed precipitation (R/A) (Nagata, 2011) from 2011 to 2016 as rainfall observation, in which rain gauge and radar estimation are compiled in a 0.05° and 1 km grid, respectively.

The scope of this study is to issue a warning of daily heavy rainfall using prognostic synoptic fields from JMA Global Spectral Model (GSM) and to evaluate the capability of our method up to 7-day lead times. We use atmospheric forecasts, specifically the operational TL959 (approximately 20-km resolution) and 60-vertical level version of GSM, which is executed for 96 h starting from 1200 UTC every day. Forecast is limited to JJ 2011–2016 and first 12 h is discarded. For comparison, predicted rainfall from GSM is also analysed.

Methodology

We applied SOM to daily-averaged atmospheric fields extracted from ERA-Interim, with a set of four variables was selected: 850-hPa zonal $U$ and meridional wind $V$, 850-hPa and 500-hPa equivalent-potential temperature $\theta_e$. Since $\theta_e$ is a thermodynamic parameter involving temperature and humidity, its distribution in the lower troposphere can be used to characterize the Baiu front activity (Tomita et al., 2011). Other studies (e.g. Ninomiya, 2000; Ninomiya and Shibagaki, 2007) also suggest that the differential advection of $\theta_e$ to the Japan Islands from the tropics that consists of a poleward moisture flux by the low-level jet (LLJ) is crucial to active the Baiu rainband. In addition, the importance of dry intrusion at the mid-level for heavy Baiu rainfall event is shown (e.g. Kato and Aranami, 2005; Kato, 2006). A quite similar selection of atmospheric variables was used by Ohba et al. (2015) to identify WPs that frequently provide heavy rainfall in Japan during the Baiu season. We also considered other candidates (e.g. 500-hPa geopotential high, mean sea-level pressure and vertically integrated moisture flux); however, selection of these variables was likely to degrade the performance of the classification and prediction.

Those four variables were obtained within a specific domain (24.75°–40.5°N, 121.5–141.75°E; Figure 1(a)) to form a total of 166 input vectors, corresponds to 166 heavy rainfall days were observed in JJ 1979–2010. Here, the heavy rainfall are defined when rainfall at least one grid box in Fukuoka-Saga prefectures exceeds 100 mm day$^{-1}$. By multiplying the dimension of the domain with the number of reanalysis fields, the size of each input vector is 4 (variables) $\times$ 28 (longitude points) $\times$ 22 (latitude points) = 2464. However, all elements are not at all independent due to spatial and inter-variable correlations among four fields, and the use of all elements in the training process is not an effective way in practice. To remove the redundant information contained in input vectors, principal component analysis (PCA) is applied for all four original reanalysis fields. Since both PCA and SOM are sensitive to the scales and thus dimensions of those fields, each variable was normalized prior to PCA. We only retain the empirical orthogonal functions that explain 90% of variance in the results. Then the corresponding principal components (PC) are digested into the SOM.
instead of the original data. The number of retained PCs is \( d = 31 \), which means the size of input vectors is reduced by a factor of 80. This improved both the independency of elements and the computational time.

Note that the SOM training process depends on the size of SOM lattice and various training parameters (e.g. learning rate, radius, and training length). Trial and error is thus used to obtain the most suitable SOM size and parameters. The quantization error (QE) and topographic error (TE) are also used as performance indices. As a result, the SOM configuration of \( 7 \times 7 \) hexagonal nodes (i.e. 49 WPs), a learning rate of 0.2 and a radius of 3 was selected. To keep the stability of SOM, the training length of two million steps was set. However, owing to concerns regarding the quality of SOM with a small number of samples, the bootstrap learning was incorporated into the SOM training process, in which the training sample was randomly drawn from the original training data. \( B = 1000 \) SOM maps generated by using the aforementioned configuration were compared. Accordingly, the map having the lowest QE and TE (Figure S1, Supporting information) and a relatively flat Sammon map (Figure S2), was defined as the master SOM.

Although SOM is powerful to project high-dimensionally nonlinear atmospheric features onto a visually understandable two-dimensional lattice, its drawbacks are unclear clustering boundaries between SOM nodes. To improve the clustering accuracy, the second stage is thus implemented, in which the SOM nodes are clustered again by the \( K \)-means method (Vesanto and Alhoniemi, 2000). However, the major disadvantage of \( K \)-means is the difficulty to decide an appropriate number of clusters \( K \). Here, this problem can be avoided by predetermining \( K \) from the master SOM by using the \( U \)-matrix (unified distance matrix) method (Ultsch and Siemon, 1990), which is similar to Nishiyama et al. (2007). The optimal clustering is determined based on the Davies-Bouldin index (DBI) proposed by Davies and Bouldin (1979). For more details on the SOM and \( U \)-matrix, one can refer to other studies (e.g. Nishiyama et al., 2007).

In the forecast phase, daily-mean of the four variables 850-hPa \( U \), \( V \), \( \theta_e \), and 500-hPa \( \theta_e \) on each day in JJ 2011–2016 are extracted from GSM over the same region in the SOM training. The GSM forecast is then normalized using the same values and is projected onto the same PCA space as used for ERA-Interim. Next, we calculate the Mahalanobis distances \( md_j \) (Mahalanobis, 1936), which is essentially the distance between the new input vector \( p = [p_1, p_2, \ldots, p_d] \) (\( d = 31 \)) and the centroid \( c_j = [c_{j1}, c_{j2}, \ldots, c_{jd}] \) of cluster \( j \) (1 ≤ \( j \) ≤ \( K \)), normalized by the standard deviation of the cluster in each dimension. Thus, it measures the number of standard deviations away the new vector is from the centroid of cluster \( j \). In PCA space, the \( md_j \) is \( \sqrt{\sum_{i=1}^{d} \left( \frac{P_i - c_{ij}}{\sigma_{ij}} \right)^2} \), where \( \sigma_{ij} \) is the standard deviation of cluster \( j \) in the \( i \)th dimension. Subsequently, if \( \min_{j} md_j \leq 6 \), the prognostic WP can be assigned to its best-match cluster \( c \), where \( c = \arg \min_{j} md_j \). Based on analogue method, the heavy rainfall will be predicted to occur in Fukuoka-Saga prefectures. To evaluate the predictability of our method, the probability of detection (POD), false alarm ratio (FAR) and equitable threat score (ETS) (see Appendix A2) were used.

Results

Classification

The clustering result (Figure S3) suggests four main different WPs inducing the heavy rainfall in Fukuoka-Saga...
prefectures during the Baiu season, which are shown in Figure 2. The existence of a Baiu front may be observed in both the Clusters 1 and 2 (C1 and C2), in which at the low-level, the southwestern Japan region including Kyushu is covered by an intrusion of remarkably warm and humid air with the LLJ. Meanwhile, at the upper troposphere, the tongue of high \( \theta_e \) air is separated west of northern Kyushu with the in-between lower \( \theta_e \) air is drier but warmer than that just to the east (see Figure S4). This low \( \theta_e \) air at 500hPa is located over the high \( \theta_e \) air at 850hPa, leading to the strong convective instability around northern Kyushu and heavy rainfall is brought in the region (Kato, 2006). Although having higher low-level \( \theta_e \), the rain-producing influence of the LLJ and warm moist tongue of C1 on the Japan Islands is less that of C2, possibly because their axis is more west–east. The area often experiences more than 50 mm day\(^{-1}\) of rainfall in C1 is relatively smaller than that in C2, including only mid- to northern Kyushu as well as the western edges of the Honshu and Shikoku Islands. Meanwhile, in C2, the high \( \theta_e \) accompanying the LLJ extends further from the southwest to the northeast elongated the Japan Islands; result in a larger rainband with moderate to heavy amount. Also note that in northern Kyushu, this cluster typically reproduces higher rainfall than C1.

Meanwhile, Cluster 3 (C3) is characterized by a relatively warm and moist air originating from the west to southwestern Japan and strong clockwise flow related to the Western North Pacific sub-tropical high. The lower \( \theta_e \) around northern Kyushu, which is produced from drier but warmer air (Figure S4), is also observed at the mid-level, causing the convective instability there; although it is not as strong as that of C1 and C2. Nishiyama et al. (2007) suggest that this WP leads to the genesis of local afternoon shower due to a local atmospheric circulation enhanced by the sunshine. The authors also show that it cannot cause intensive rainfall since water vapour is insufficient, corresponds to the impacts of C3 are only limited to northern Kyushu and westernmost Honshu with only few spots exceeding 100 mm day\(^{-1}\). On the other hand, C4 is linked to the low-level cyclonic circulation related to low pressure
system (i.e. extratropical/tropical cyclone/depression; ETCD) located around northern Kyushu. Although observed with the lowest $\theta_e$, the WPs of this cluster reproduce an abundant amount of rainfall over the broadest area of southwestern Japan, as compared to other clusters. The characteristic features of the ETCD, however, are still not clearly recognized due to the relatively coarse resolution of current reanalysis data.

Predictability

Figure 3 shows the maximum daily rainfall in Fukuoka-Saga prefectures in JJ 2011–2016 from observation and the GSM forecast (Figures 3(a)–(d)) and the heavy rainfall occurrence predicted based on prognostic WP from GSM (Figure 3(e)). The forecast skill of this new method and the traditional method based on the predicted rainfall intensity from GSM, are shown in Figure 4. Basically, the operational GSM failed to detect the outbreak of local heavy rainfall with very poor POD since its predicted rainfall amounts are usually much smaller than observations. It also provides high FAR, especially with the forecast range from 3 days. As a result, the overall ETS with a threshold of 100 mm day$^{-1}$ of GSM is 0.06 and 0.07 for 1 and 2 days forecast, respectively, and then decreases to around 0.1 for the longer forecast ranges. This shows that SOM not only significantly outperforms GSM for the same forecast ranges but also its medium-ranges (3–7 days) forecast are better than short-ranges (1–2 days) forecast from GSM.

Conclusions

The SOM in combination with K-means cluster are conducted for classification of WP causing the heavy rainfall exceeding 100 mm day$^{-1}$ during the Baiu season (JJ) of 1979–2010 over northern Kyushu, southwestern Japan. It results in four clustered patterns, which are primarily attributed to the Baiu front and the ETCD activities and characterized by high $\theta_e$ intrusion accompanying the LLJ or cyclonic circulation. These features are in good agreement with previous studies such as Nishiyama et al. (2007) and Ohba et al. (2015). The classification results are then implemented with the analogue method to forecast the occurrence (yes/no) of local heavy rainfall days in JJ of 2011–2016 using the prognostic synoptic fields from GSM. In general, the quantitative forecasting skill by the POD and ETS up to 7 days in advance under our approach is significantly improved than that from the conventional method based only on the predicted rainfall intensity from GSM, although the FAR is still high.

Note that our method is effective under the assumption that local rainfall is mostly controlled by the synoptic condition. It may be more difficult to apply to the region where rainfall process is mainly related to local factor such as the sea breeze circulation and local heating. Since the method can predict whether the heavy rainfall occurs but not being capable of providing the actual amount, questions may remain about its relative availability compared with other methods such as statistical downscaling by Ohba et al. (2016). Additionally, the forecast skill of SOM is still lower than that from the
JMA meso-scale model (MSM) that has a fine resolution of 5 km and the lead time up to 39 h (see Figure 4). Thus, there is still room for improvement. The spatial resolution of ERA-Interim is only 0.75°, which falls short for a precise description of heavy rainfall-inducing WPs, particularly for the ETCD activities. Given the weaknesses of the current dataset used as training sample for SOM, classification and predication results can be improved by using higher resolution and more accurate reanalysis and observations. Also further studies employed improved method should be considered. Nevertheless, in addition to previous studies, the present results encourage the idea of using weather pattern recognition for heavy rainfall prediction, by which it can provide a fruitful and first-order guidance, particularly for ranges longer than 2 days, for decision-making by weather forecasters or end-users engaging in disaster management activities.

Acknowledgements

The authors greatly appreciate three anonymous reviewers for their constructive comments. We also thank Dr. Le Duc of JAMSTEC for his critical suggestions. This research is supported by the Social Implementation of Climate change Adaptation Technology (SICAT) project of the Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT); and the Advancing Co-design of Integrated Strategies with Adaptation to Climate Change (ADAP-T) of the Japan International Collaboration Agency (JICA)/Japan Science and Technology Agency (JST) Science and Technology Research Partnership for Sustainable Development (SATREPS).

Supporting information

The following supporting information is available:

Appendix A1. Methodology

Figure S1. The Quantization Error (QE) and Topographic Error (TE) of 1000 SOM maps using bootstrap learning algorithm. The blue mark denotes the optimal master SOM.

Figure S2. Sammon map for the master SOM with the four corner nodes labelled.

Figure S3. Scaled U-matrix of the master SOM with the optimum clustering by K-means is shown.

Appendix A2. Forecast skill score

Appendix A3. Results of classification

Figure S4. Cluster-averaged (composited) of daily-mean (a), (c), (e) and (g) 500-hPa potential temperature $\theta$ (oK: shading) and (b), (d), (f) and (h) 500-hPa relative humidity ($\%$: shading) from Clusters 1–4 (C1–C4), respectively. The contours denote each cluster-averaged of daily-mean 500-hPa equivalent-potential temperature $\theta_e$.

References

Davies DL, Bouldin DW. 1979. A cluster separation measure. IEEE Transactions on Pattern Analysis and Machine Intelligence 1: 224–227.

Dee DP, Uppala SM, Simmons AJ et al. 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Quarterly Journal of the Royal Meteorological Society 137: 553–597. https://doi.org/10.1002/qj.828.

Kamiguchi K, Arakawa O, Kitoh A, Yatagai A, Hamada A, Yasutomi N. 2010. Development of APHRODITE, the first Japanese high-resolution daily precipitation product for more than 100 years. Hydrological Research Letters 4: 60–64. https://doi.org/10.3178/HRL-4.60.

Kato T. 2006. Structure of the band-shaped precipitation system inducing the heavy rainfall observed over northern Kyushu, Japan on 29 June 1999. Journal of the Meteorological Society of Japan 84: 129–153.

Kato T, Aranami K. 2005. Formation factors of 2004 Niigata-Fukushima and Fukui heavy rainfalls and problems in the predictions using a cloud-resolving model. Scientific Online Letters on the Atmosphere 1: 1–4.

Kohonen T. 1982. Self-organized formation of topologically correct feature maps. Biological Cybernetics 43: 59–69. DOI: https://doi.org/10.1007/BF00337288.

Lorenz EN. 1969. Atmospheric predictability as revealed by naturally occurring analogues. Journal of the Atmospheric Sciences 26: 639–646.

Mahalanobis PC. 1936. On the generalised distance in statistics. Proceedings of the National Institute of Sciences of India 2: 49–55.

Nagata K. 2011. Quantitative precipitation estimation and quantitative precipitation forecasting by the Japan Meteorological Agency. Technical review of RSMC Tokyo – Typhoon Center 13: 37–50 http://www.jma.go.jp/jma/jma-eng/jma-center/rsmc-hp-pub-eng/techrev/text13-2.pdf.

Ninomiya K. 2000. Large- and meso-$\alpha$-scale characteristics of Meiyu–Baiu front associated with intense rainfalls in 1–10 July 1991. Journal of the Meteorological Society of Japan 78: 141–157.

Ninomiya K, Shibagaki Y. 2007. Multi-scale features of the mei-yu–bairu front and associated precipitation systems. Journal of the Meteorological Society of Japan 85B: 103–122. https://doi.org/10.2151/jmsj.85B.103.

Nishiyama K, Endo S, Jinno K, Uvo CB, Olsson J, Berndtsson R. 2007. Identification of typical synoptic patterns causing heavy rainfall in the rainy season in Japan by a self-organizing map. Atmospheric Research 83: 185–200.

Ohba M, Nohara D, Yoshida Y, Kadokura S, Toyoda Y. 2015. Anomalous weather patterns in relation to heavy precipitation events in Japan during the Baiu season. Journal of Hydrometeorology 16: 688–701. https://doi.org/10.1175/JHM-D-14-0124.1.

Ohba M, Kadokura S, Nohara D, Toyoda Y. 2016. Rainfall downscaling of weekly ensemble forecasts using self-organising maps. Tellus A 68: 29293. https://doi.org/10.3402/tellusa.v68.29293.

Tomita T, Yamaura T, Hashimoto T. 2011. Interannual variability of the Baiu season near Japan evaluated from the equivalent potential temperature. Journal of the Meteorological Society of Japan 89: 517–537.

Ultsch A, Siemon HP. 1990. Kohonen’s self organizing feature maps for exploratory data analysis. In Proceedings of INNC’90, International Neural Network Conference. Kluwer Academic Publishers: Dordrecht; 305–308.

Vesanto J, Alhoniemi E. 2000. Clustering of the self-organizing map. IEEE Transactions on Neural Networks 11: 586–600. https://doi.org/10.1109/72.846731.