Using Weather Pattern Recognition to Classify and Predict Summertime Heavy Rainfall Occurrence over the Upper Nan River Basin, Northwestern Thailand

DZUNG NGUYEN-LE AND TOMOHITO J. YAMADA

Faculty of Engineering, Hokkaido University, Sapporo, Hokkaido, Japan

(Manuscript received 20 July 2018, in final form 8 February 2019)

ABSTRACT

In this study, self-organizing maps in combination with K -means clustering are used to objectively classify the anomalous weather patterns (WPs) associated with the summertime [May–June (MJ) and July–August–September (JAS)] heavy rainfall days during 1979–2007 over the Upper Nan River basin, northwestern Thailand. The results show that in MJ, intensive rains are mainly brought by the remarkable enhancement of the westerly summer monsoon. Meanwhile, westward-propagating tropical disturbances including tropical cyclones are the primary factors that reproduce heavy rainfall over the Upper Nan in JAS. These results also suggest that the occurrence time of local heavy rainfall is strongly related to the seasonal transition of the summer monsoon over the Indochina Peninsula. The classification results are then implemented with the perfect prognosis and analog method to predict the occurrence (yes/no) of heavy rainfall days over the studied basin in summer 2008–17 using prognostic WPs from the operational Japan Meteorological Agency Global Spectral Model (GSM). In general, the forecast skill of this approach up to 3-day lead times is significantly improved, in which the method not only outperforms GSM with the same forecast ranges, but also its 3-day forecast is better than the 1–2-day forecasts from GSM. However, the false alarms ratio is still high, particularly in JAS. Nevertheless, it is expected that the new approach will provide warning and useful guidance for decision-making by forecasters or end-users engaging in water management and disaster prevention activities.

1. Introduction

In Thailand and the Asian monsoon region in general, water-related disasters have taken many lives and devastated infrastructure such as bridges, roads, buildings, electrical grids, and telecommunications. To mitigate the damaging effects of such severe events, precise and timely rainfall prediction is crucial. One common way to improve the predictability of rainfall is the development and improvement of numerical weather prediction models. In recent years, thanks to the rapid progress in atmospheric numerical modeling and observational networks, the skill of rainfall prediction has been significantly improved. Nevertheless,

it has always been challenging to provide an accurate forecast of heavy rainfall in practice because of the imperfection of models and computational constraints. Thus, in addition to dynamical downscaling, various statistical techniques have been proposed (e.g., Friederichs and Hense 2007; Friederichs 2010; Maraun et al. 2010), because of their inexpensive cost, ease of use, and relatively comparable performance.

Typically, in the case of statistical rainfall prediction, a perfect prognosis (PP) method in which a statistical relationship is established between a local-scale observed “predictand” (i.e., rainfall) and simultaneously observed large-scale “predictors” (e.g., temperature, humidity, wind, or geopotential height) is taken and then applied to predicted predictors derived from a global model to yield forecasts of the predictand. A weather pattern (WP) recognition technique such as the self-organizing map (SOM), developed by Kohonen (1982, 2001), is one of the widely used methods to determine this relationship. Thus, previous studies (e.g., Gutiérrez et al. 2005; Nishiyama et al. 2007; Chattopadhyay et al. 2008; Borah et al. 2013;

Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/WAF-D-18-0122.s1>.

Corresponding author: Dzung Nguyen-Le, dzungnl@eng.hokudai.ac.jp

FIG. 1. (a) Terrain height (km) of the area used to define the weather patterns. The red box denotes the Upper Nan river basin. (b) Climatological annual cycle of area-maximum rainfall intensity (mm day^{-1}) during 1979–2007 over the Upper Nan river basin.

Ohba et al. 2016; Nguyen-Le et al. 2017) have proposed the use of a SOM-based analog technique for rainfall prediction. The traditional analog method (Lorenz 1969) is based on the assumption that if the current (targeted) WP is similar to those of historical observation, the local rainfall can be similar to that of the past. In this framework, Nguyen-Le et al. (2017) recently used the SOM in combination with *K*-means clustering to objectively classify the anomalous WPs causing heavy rainfall days over northern Kyushu, southwestern Japan, during the baiu season (June–July). The classification results are then used with the PP and analog method to predict the occurrence (yes/no) of heavy rainfall days in the region. The predictability of this relatively new method, which is a combination of an analog forecast and SOM (ASOM) up to 7-day lead times, is significantly improved from the traditional method using only the predicted rainfall intensity from an operational global model. However, it has been demonstrated that ASOM is effective in the mid-latitudes (e.g., Japan) where the rainfall process is primarily governed by synoptic-scale factors such as the baiu front. It may be more difficult to apply to the tropics (e.g., Indochina) where intense rainfall is often related to smaller-scale (i.e., mesoscale) factors such as tropical disturbances (TDs) (Takahashi and Yasunari 2006; Yokoi and Matsumoto 2008).

Here, to demonstrate the relevance of ASOM in the tropics, a work extending Nguyen-Le et al. (2017) to summertime (May–September) heavy rainfall over the Upper Nan River basin (17.375°N – 19.375°N , 99.875°E – 101.125°E) in northwestern Thailand (Fig. 1a) is conducted. The Nan River is the major tributary of the Chao Phraya River, the largest and the most important river

system in Thailand. The Sirikit Dam, the second biggest dam in Thailand, was constructed on the Upper Nan for the purposes of irrigation, flood control, and power generation. The Nan River basin and the Chao Phraya River basin in general were heavily affected by the massive 2011 flood and still face flood problems causing loss of lives and property damages almost every year. Furthermore, floods on the same scale as the 2011 one are expected to occur in the future because of global warming (Komori et al. 2012). This major flood damage is not only a domestic issue for Thailand, but also a global complication on account of its impact on industrial supply chains (Ye and Abe 2012). The objective of the present work is thus not only to potentially improve the forecasting capability, but is also fundamental in improving our understanding on the genesis of heavy rainfall (the root cause of the floods) over the studied basin.

The reminder of this paper is organized as follows. Sections 2 and 3 describe the datasets and the methodology, respectively. The classification and prediction results are then given in sections 4 and 5, respectively. Finally, section 6 presents the conclusions.

2. Datasets

The daily rainfall over the Upper Nan are derived from the Asian Precipitation–Highly Resolved Observational Data Integration Toward Evaluation of Water (APHRODITE) Monsoon Asia (APHRO_MA; Yatagai et al. 2012) and the Tropical Rainfall Measuring Mission (TRMM) Multisatellite Precipitation Analysis (TMPA), version 7, TMPA-3B42V7 (Huffman et al. 2007, 2010)

FIG. 2. Relative frequency histogram of the basin-maximum daily rainfall (mm day^{-1}) over the Upper Nan in (a) MJ and (b) JAS from 1979 to 2007. The bin width is set to 1 mm day^{-1} , and mean μ , standard deviation σ , and 95th percentile (95%tile) are provided in the legend.

datasets during 1979–2007 and 2008–17, respectively. APHRO_MA, which is provided by the Research Institute for Humanity and Nature (RIHN) and the Meteorological Research Institute of Japan Meteorological Agency (MRI/JMA), is created primarily from rain gauge data recorded from 1951 to 2007 and covers the entire monsoon Asian region (15°S – 55°N , 60° – 150°E) on a $0.25^{\circ} \times 0.25^{\circ}$ grid. TMPA-3B42V7 is a $0.25^{\circ} \times 0.25^{\circ}$ and 3-hourly merged multisatellite precipitation estimate available over a global belt ranging between 50°S and 50°N . This product is reproduced in four stages: 1) the microwave (MW) precipitation estimates are calibrated against TRMM Combined Instrument (TCI), which includes TRMM Precipitation Radar (PR); 2) the infrared (IR) precipitation estimates are created by using the calibrated MW precipitation; 3) the MW and IR estimates are combined; and 4) rescaling to monthly data is applied using the gauge dataset from the Global Precipitation Climatology Centre (GPCC).

It has been recognized that the seasonal progress of rainfall over the Upper Nan and Thailand in general is bimodal with a strong peak in early September and a weaker peak from late May to early June (Fig. 1b). Rainfall in the first half of the rainy season [May–June (MJ)] is brought by the westerly summer monsoon; while rainfall in the second half of the rainy season [July–August–September (JAS)] is primarily because of TDs including tropical cyclones (Takahashi and Yasunari 2006). In late June, there is a small dip in rainfall that corresponds to the monsoon break. Thus, the studied period is divided into two subseasons: MJ and JAS. Based on the relative frequency histogram of

the basin-maximum daily rainfall (Fig. 2), the days that heavy rainfall events occurred are derived using the following condition: at least one grid box over the Upper Nan recorded a daily accumulated rainfall exceeds the 95th percentile value that is 40 (45) mm in MJ (JAS). In 29 years (1979–2007) there were 92 and 131 heavy rainfall days recorded in MJ and JAS, respectively, which is equivalent to 3.2 and 4.5 days yr^{-1} each subseason.

The synoptic atmospheric circulation associated with each heavy rainfall day was obtained from the European Centre for Medium-Range Weather Forecasts interim reanalysis (ERA-Interim; Dee et al. 2011) on a specific domain (5.25° – 30°N , 84.75° – 120°E ; see Fig. 1a) with a horizontal resolution of 0.75° . Considering the WPs cause local heavy rains based on previous studies, mean sea level pressure (pmsl), 500-hPa geopotential height (z500), 850-hPa zonal wind (u850), and 850-hPa meridional wind (v850) are utilized. These four variables are also the essential elements in the surface and upper-air analysis, which are commonly used by Thai synopticians for weather forecasting. The pmsl pattern represents the central features of the weather system: highs, lows (including tropical cyclones), and cold fronts. The z500 map exhibits activity of the midlevel ridges and troughs, especially the western Pacific subtropical high (STH) that is an important component of the Asian monsoon system. Over the Indochina Peninsula, a weakening and eastward withdrawal of the STH is favorable for the enhanced rising motion and strong convection, whereas a strong and westward-extended STH leads to the suppressed convection (Zhang et al. 2002). Chan (2000) suggests that

anomalous 500-hPa flows that are responsible for steering tropical cyclones toward or away from a region are the main reason for the variability in the tropical cyclones activity over the western North Pacific basin. Finally, the monsoon flows and TD-like cyclonic vortices clearly play the most important role and the low-level wind fields (u850, v850) can be used to represent such factors. Additionally, these four variables are highly correlated with maximum rainfall over the studied basin (see Figs. S1 and S2 in the online supplemental material). We also considered other candidates (e.g., temperature and moisture). However, selection of these variables was likely to degrade the performance of the classification and prediction because of their limited forecast skill by the global model (Fig. S3), which is evaluated by the anomaly correlation coefficient (ACC) and root-mean-square error (RMSE) (see [appendix A](#)).

The scope of this study is to issue a warning of heavy rainfall days by using prognostic WPs from the JMA Global Spectral Model (GSM) and to evaluate the capability of ASOM up to 3-day lead times. We used atmospheric forecasts, specifically from the operational TL959 (approximately 20-km horizontal resolution) and 60 vertical-level version of GSM ([Japan Meteorological Agency 2016](#)), which are executed for 84 h starting from 0000 UTC (0700 LT) every day, in May–September 2008–17. For comparison, predicted rainfall intensity from GSM is also analyzed.

3. Methodology

a. Classification

The methodology applied in the present study was adopted from [Nguyen-Le et al. \(2017\)](#). We approach the classification problem of WPs, which induce heavy rainfall in the Upper Nan by using SOMs. The input vectors are obtained from the four daily averaged reanalysis fields of pmsl, z500, u850, and v850. All these fields at each heavy rainfall day compose an input vector that is fed into the SOM in the training phase. There are total of 93 and 131 input vectors, corresponding to the number of heavy rainfall days that were recorded in MJ and JAS during 1979–2007, respectively. By multiplying the dimension of the domain with the number of reanalysis fields, the size of each input vector is $4 \text{ (fields)} \times 41 \text{ (longitude points)} \times 28 \text{ (latitude points)} = 4592 \text{ (elements)}$. However, not all elements are independent because of spatial and intervariable correlations among the 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 the input vectors, principal component analysis (PCA) is

applied for all four original reanalysis fields. Since both PCA and SOMs 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 95% of variance in the results. Then the corresponding principal components (PCs) are digested into the SOM instead of the original data. The number of retained PCs is $d = 30$ and 34 for MJ and JAS, respectively, which means the size of an input vector is reduced by a factor of 153 and 135 in MJ and JAS, respectively. This improved both the independence of elements and the computational time.

The 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. For instance, we conducted several tests including different lattice sizes such as 4×4 , 5×5 , 6×6 , and 7×7 . The number of samples that fell into each node is considered, in which large values (e.g., higher than 10) in some lattice areas suggest that a larger size would be beneficial, whereas empty nodes indicate that the size is too big. The quantization error (QE) and topographic error (TE) are also used as performance indices. As a result, the SOM of 6×6 hexagonal nodes (i.e., 36 WPs), a learning rate of 0.2, and a radius of three was applied. To keep the stability of the SOM, a training length of two million steps was set. Because of concerns regarding the quality of the SOM with a small number of samples, bootstrap learning was incorporated into the SOM training process, in which the sample was randomly drawn from the training data. In total, 1000 SOM maps were generated with the aforementioned configuration in each subseason. Accordingly, the maps having the lowest QE and TE (Fig. S4) and relatively flat Sammon maps ([Sammon 1969](#)) (Fig. S5) are selected as the master SOMs.

Since the SOM was designed to preserve the topological structure of high-dimensional input space in its resulting two-dimensional lattice, a node will be more similar to its adjacent nodes rather than the nodes located farther away ([Vesanto and Alhoniemi 2000](#)). This sense of similarity can be quantified by using the unified distance matrix (*U*-matrix) measure. For a given node a *U*-matrix value is the average distance between the codebook vector of this node and those of the closest neighbors ([Ultsch and Siemon 1990](#)). Thus, to improve the clustering accuracy, the two-level approach can be implemented, in which the obtained lattice is further classified into clusters with respect to this measure. Here, the master SOMs are clustered again by using *K*-means clustering. The major drawback of *K*-means clustering is the difficulty in deciding

FIG. 3. Scaled U -matrix of the master SOM for (a) MJ and (b) JAS with the optimum clustering by K -means shown.

on an appropriate number of clusters K . To avoid this problem, K can be predetermined by analyzing the U -matrix. A potential cluster is identified with the nodes around a local minimum of the U -matrix while the boundaries between clusters correspond to the nodes whose U -matrix values are large relative to the neighboring local minima. Figures 3a and 3b show the U -matrix, whose values are scaled into the unit interval using their maximum and minimum values, of the master SOM in MJ and JAS, respectively. Accordingly, the number of clusters in MJ is four (Fig. 3a). By counting the local minima, there are three obvious clusters that correspond to three local minima at the top-left, top, and bottom-right areas. In addition, high values of the nodes around the bottom-left corner suggest that the WP from this area is relatively different to those from other areas. Apparently, the U -matrix in JAS can be divided into four distinct parts, which indicates $K = 4$ (Fig. 3b). Consequently, these classification results suggest there are four and four different WPs inducing heavy rainfall over the Upper Nan in MJ and JAS, respectively. For more details on the SOM and U -matrix, one can refer to Nishiyama et al. (2007) and Nguyen-Le et al. (2017).

b. Prediction

In the prediction phase, the classification results are implemented with the PP and analog method to produce forecasts of heavy rainfall over the Upper Nan by using the prognostic WPs from GSM. Daily-mean pmsl, z500, u850, and v850 on each day in MJ and JAS during 2008–17 are extracted from the GSM forecasts over the same region in the SOM training. These fields are then normalized by using the same values (means and standard deviations) and are projected onto the same PCA space as used for ERA-Interim to form a forecast vector $\mathbf{p} = [p_1, p_2, \dots, p_d]$

($d = 30$ and 34 for MJ and JAS, respectively). Next, we calculate the Mahalanobis distance md_j (Mahalanobis 1936), which is essentially the distance between $\mathbf{p}$ and the centroid $\mathbf{c}_j = [c_{1j}, c_{2j}, \dots, c_{dj}]$ of cluster j ($1 \leq j \leq 4$ and $5 \leq j \leq 8$ for MJ and JAS, respectively), normalized by the standard deviation of the cluster in each dimension. Thus, it measures the number of standard deviations away $\mathbf{p}$ is from the centroid of cluster j . In the PCA space, the md_j is $\sqrt{\sum_{i=1}^d [(p_i - c_{ij})/\sigma_{ij}]^2}$, where σ_{ij} is the standard deviation of cluster j in the i th dimension. Subsequently, if $\min_j \text{md}_j \leq 5.5$ the forecast vector $\mathbf{p}$ can be assigned to its best-match cluster $\mathbf{c}$, where $c = \arg\min_j \text{md}_j$. In other words, the prognostic WP is considered to be similar enough to the historical heavy rainfall-inducing WP of cluster $\mathbf{c}$. Based on the analog method, heavy rainfall is predicted to occur over the studied region. The threshold of $\min_j \text{md}_j$ is determined after conducting several sensitivity tests including its value varies from 3 to 8. For evaluation, the probability of detection (POD), false alarm ratio (FAR), and equitable threat score (ETS) (see appendix B) were used. The result of $\min_j \text{md}_j \leq 5.5$ has the best forecast skill scores, and it is the only result shown.

In addition to overall results, the differences in the performance of ASOM between years and clusters are evaluated. The ERA-Interim analysis on each heavy rainfall day is projected onto the master SOM to determine the cluster that the observed WP belongs to. The year-to-year variations are evaluated by calculating the skill scores for each year from 2008 to 2017. In this calculation, the prediction results in MJ and JAS are combined by reason of the limited number of samples in each subseason. We also assess the performance of ASOM for each individual cluster. Since the targeted basin is relatively small as compared to the big domain for the WPs, it is possible that the clusters

FIG. 4. (a),(d),(g),(j) Cluster-averaged daily-mean sea level pressure (hPa; shading) and 500-hPa geopotential height (m; contours); (b),(e),(h),(k) 850-hPa temperature (K; shading) and wind vectors (m s^{-1}); (c),(f),(i),(l) observed rainfall (mm day^{-1} ; shading) for C1–C4 in MJ. Percentages in (a), (d), (g), and (j) denote the occurrence frequency of each cluster. Numbers in (b), (e), (h), and (k) show the heavy rainfall days that occurred over the Upper Nan river basin in May and June, respectively. The red box in (c), (f), (i), and (l) denotes the Upper Nan river basin.

are correlated or represent the same heavy rainfall event during consecutive days. Thus, in this analysis, when heavy rainfall occurs, a forecast is still defined as a “hit” even if the predicted and observed WPs belong to different clusters.

4. Classification results

a. May–June

In MJ, the significant intensification of westerly flows observed in both clusters 2 and 3 (C2–C3) is the most recognizable WP causing heavy rainfall over the Upper Nan. At the low levels (850 hPa), summer monsoon westerlies prevail across the whole Bay of Bengal (BOB) and Indochina region (Figs. 4e and 4h). This is associated with the existence of a low pressure system over northern BOB and eastern India, which can be seen by a low and a deep north–south trough at sea level and midlevel, respectively (Figs. 4d and 4g). As a result, moderate to heavy rainfall occurs to the east side of the trough (i.e., Myanmar, northern Thailand, and Laos). In comparison with C2, the western ridge of the STH in C3, denoted by the 5865-gpm contour of

z500, is remarkably weaker and retreats eastward to the Philippines and the western North Pacific. Whereas, the ridge reaches the southeastern area of the Indochina Peninsula (i.e., central and southern Vietnam) in C2. The BOB low in C3 is also significantly more amplified than that in C2. Correspondingly, C3 reproduces more intensive rainfall in northern Myanmar and eastern Thailand, whereas the rainfall of C2 is higher along the southwestern coast of Myanmar.

The enhanced westerlies associated with the BOB low are also found in the least common C4 (Figs. 4j and 4k). Interestingly, both the low and westerlies are relatively stronger than those of C2–C3. However, heavy rainfall over the Upper Nan is attributed to the convergence of those westerlies and the trade easterlies associated with the STH. The STH expands westward to cover the northern South China Sea (SCS) and southern China, which results in warmer conditions over southwestern China and northernmost Myanmar. The midlevel trough over the northern BOB also extends eastward to reach Thailand. The studied region is thus sandwiched between these two systems. This WP provides heavy rainfall in northern Vietnam, Laos, and eastern Thailand.

FIG. 5. Relative frequency histogram of daily rainfall (mm day^{-1}) of all grid boxes inside the Upper Nan from (a) C1, (b) C2, (c) C3, and (d) C4 in MJ from 1979 to 2007. The bin width is set to 1 mm day^{-1} , and mean μ , standard deviation σ , and 95th percentile (95%tile) are provided in the legend.

On the other hand, C1 is characterized by the influence of a cold surge reflected by high pmsl (Fig. 4a) and low-level cold air (Fig. 4b) advection from southern China to the northern Indochina Peninsula. Heavy rainfall is thus brought by the convergence of the northeasterlies associated with this cold surge and the southwesterlies originating from the BOB. However, at the midlevels, a strong ridge of the STH still controls the entire SCS and most of the Indochina area. Thus, over Indochina, heavy rainfall is only found around northern Thailand and eastern Laos (Fig. 4c).

To further examine the resultant rainfall of C1–C4 in the Upper Nan, the relative frequency histogram of daily rainfall of all grid boxes inside the basin from each cluster is displayed in Fig. 5. This reveals that C2 reproduces the most extreme local rainfall with a 95th percentile of 60.9 mm day^{-1} (Fig. 5b). The mean and standard deviation of C2 are also considerably higher than other clusters, with the most intensive amount observed over southwestern basin (Fig. 4f). The 95th percentile of rainfall is about 45 mm day^{-1} from both C1 and C4, in spite of a higher mean for C2 (Figs. 5a and 5d). In C4, heavy rainfall is only recorded over southernmost Upper Nan (Fig. 4l), whereas it occurs over both the northern and southern area of the basin in C2 (Fig. 4c). Meanwhile, C3 has a 95th percentile of rainfall and is only marginally over 40 mm day^{-1} , although its mean is slightly higher than that of C4 (Fig. 5c). Relative similar to C4, the torrential rainfall of C3 is limited to the southern area of the studied basin (Fig. 4i).

b. July–August–September

In JAS, the most common WP has the characteristics of the westward-propagating TD (C5 and C8) that is in good agreement with previous studies (e.g., Takahashi and Yasunari 2006; Yokoi and Matsumoto 2008; Takahashi et al. 2015). Figures 6a,b and 6j,k show a low extended from the surface to the midlevels accompanied by an anomalous cyclonic circulation at 850 hPa near the studied region. The detailed characteristics of the TD, however, are relatively difficult to clearly recognize by considering only these fields. The

relative vorticity at 700 hPa (RV700) of C1 and C4 are thus provided in Figs. 6b and 6k, respectively. The 700-hPa level is used here to avoid the influence of mountainous terrain over Indochina. Figure 6k shows that RV700 of higher than $4 \times 10^{-5} \text{ s}^{-1}$, which is the threshold of a TD (Fudeyasu et al. 2006), is found over northern Laos and Vietnam. Furthermore, C5 is associated with an obvious low-level cyclonic circulation with RV700 of more than $5 \sim 6 \times 10^{-5} \text{ s}^{-1}$ (Fig. 6b), corresponding to an extreme low pressure center, both at sea level and midlevels, located around the studied basin (Fig. 6a). This WP indicates a landfalling tropical cyclone that is much stronger than the TD in C8. In comparison with other clusters, the two aforementioned clusters generally reproduce not only the most abundant amount of rainfall but also over the broadest area of Indochina, in which C5 provides much more rainfall over the Upper Nan and Laos than C8, whereas rainfall from C8 is notably higher over northern Vietnam (Figs. 6c and 6l).

In C6, local heavy rainfall is brought by an intensification of the westerly monsoon (Fig. 6e). This WP is relative similar to those of C2–C3, in which a low is also located over the northern BOB at the surface (Fig. 6d). However, at the midlevels over this low, a west–east trough extends from the BOB to cover western Thailand. To the east, a northeast–southwest-oriented trough exists between southeastern China and southern Indochina. The Upper Nan is located at the boundary between these two troughs. In addition to northeastern Thailand and northern Laos, this WP has a great influence on rainfall along the western coast of Myanmar (Fig. 6f).

Meanwhile, C7 points out the local heavy rainfall that is attributed to the convergence of the wet monsoon westerlies and the easterlies associated with the STH (Figs. 6g and 6h) similar to C4. The STH also extends significantly westward and withdraws southward to dominate not only southern China but also northeastern Indochina, which results in the high z500 and pmsl over these areas. As a result, the associated trade easterlies cover the northern Indochina Peninsula, while the mid- to southern Indochina

FIG. 6. (a),(d),(g),(j) Cluster-averaged daily-mean sea level pressure (hPa; shading) and 500-hPa geopotential height (m; contour); (b),(e),(h),(k) 700-hPa relative vorticity (10^{-5} s^{-1} ; shading) and 850-hPa wind vectors (m s^{-1}); (c),(f),(i),(l) observed rainfall (mm day^{-1} ; shading) for C5–C8 in JAS. Percentages in (a), (d), (g), and (j) denote the occurrence frequency of each cluster. Numbers in (b), (e), (h), and (k) show the heavy rainfall days that occurred over the Upper Nan river basin in July, August, and September, respectively. The red box in (c), (f), (i), and (l) denotes the Upper Nan river basin.

Peninsula is still prevailed upon by the monsoon westerlies. The wind convergence thus occurs over northern Thailand, Laos, and northern Vietnam; moderate to heavy rainfall is also limited to these regions (Fig. 6i).

The relative frequency histogram of daily rainfall of all grid boxes inside the Upper Nan from C5 to C8 is shown in Fig. 7. Unsurprisingly, C5 induces the heaviest rainfall with a mean and 95th percentile of more than 26 and 72 mm day^{-1} , respectively (Fig. 7a). Next, C6 and C8 have an equivalent mean and 95th percentile of 24 and 54 mm day^{-1} , respectively (Figs. 7b and 7d). Meanwhile, the 95th percentile of rainfall is only about

47 mm day^{-1} from C7, accompanied with the lowest mean of 19 mm day^{-1} (Fig. 7c). Note that the highest rainfall from all four clusters is recorded in the northeastern part of the Upper Nan and corresponds to the mountainous area (Fig. 1a), implying the impact of local topography on extreme rainfall.

c. Relationship between the occurrence time of local heavy rainfall days and the seasonal transition of the summer monsoon over the Indochina Peninsula

The Asian summer monsoon has often been referred to as the rainy phase of a seasonally changing pattern

FIG. 7. Same as Fig. 5 but for C5–C8 in JAS.

FIG. 8. Maximum rainfall intensity (mm day^{-1}) over the Upper Nan river basin during MJ 2008–17 derived from the observations (bars) and GSM forecast (lines) with (a) 1-, (b) 2-, (c) 3-day lead times. (d) The occurrence (yes/no) of the heavy rainfall day and its causing clustered WP (solid boxes denote yes) predicted by using atmospheric fields derived from the ERA-Interim reanalysis and GSM forecasts with 1–3-day lead times. Here, the forecast range of 0 day represents ERA-Interim.

and its onset (retreat) is a key indicator that characterizes the transition from dry (rainy) to rainy (dry) conditions. It is thus not really surprising to see the close relationship between the occurrence time of each cluster and the seasonal transition of the summer monsoon over Indochina. Zhang et al. (2002) shows that the summer monsoon outbreak over the Indochina Peninsula in around May is characterized by the arrival of the low-level tropical westerlies, the intensification of the BOB trough at the midlevels, and the eastward withdrawal of the STH. C2 seems to occur during this onset process when the STH is eastward retreating and only reaches to the easternmost area of the Indochina Peninsula. It also corresponds to an outbreak of rainfall over the Upper Nan. Next, C3 comes after the onset, in which the STH withdraws to the western North Pacific and monsoon westerlies cover Indochina entirely. Meanwhile, C1 mostly happens in May prior to the onset, when the STH still dominates Indochina. In contrast, the occurrence of C4 is often in June, which is latest in comparison with the other clusters.

In contrast, Nguyen Le et al. (2015) shows that the withdrawal of the summer monsoon over the northern Indochina Peninsula in September is accompanied by the westward extension of the STH and the associated trade easterlies. This is consistent with the occurrence time of C7 that is often seen late in the rainy season (September). However, C6 mostly occurs in July, after the break of monsoon in late June (Fig. 1b). In addition, C4 is mainly observed in July–August. These two clusters are synchronized with the peak of the Asian summer monsoon and the western North Pacific typhoon season (Wang and LinHo 2002). Later, C8, which is characterized by a weaker TD, appears in August–September.

5. Prediction results

a. Overall results

Figures 8 and 9 display the maximum daily rainfall over the Upper Nan during MJ and JAS of 2008–17 (10 years) from the observations and GSM forecasts (Figs. 8a–c and 9a–c) and local heavy rainfall occurrence

FIG. 9. Same as Fig. 8 but for JAS.

predicted by using ASOM with the prognostic WP from GSM (Figs. 8d and 9d), respectively. The forecast skill at the heavy rainfall thresholds of ASOM and the conventional method using the predicted rainfall intensity from GSM in MJ and JAS are then compared in Fig. 10. Basically, the operational GSM has difficulty in capturing the occurrence of local heavy rainfall with a relatively poor POD of less than 0.24 and 0.18 in MJ and JAS, respectively, since its predicted heavy rainfall peaks are often smaller than those of the observations. It also provides a quite high false alarm rate with the FAR higher than 0.54 and 0.63 for all forecast ranges in MJ and JAS, respectively. Consequently, the overall forecast skill of GSM estimated by the ETS is lower than 0.76 and 0.48 in MJ and JAS, respectively.

It is seen that the predictability of the occurrence of heavy rainfall over the Upper Nan can be significantly improved by using ASOM. It not only remarkably outperforms GSM for the same forecast ranges but also its longer-range (3 days) forecast is better than the short-range (1–2 days) forecast from GSM. In detail, the POD of ASOM is higher than 0.54 and 0.57 in MJ and JAS, respectively, for all forecast ranges. However, many overestimated forecasts still exist, with the FAR around 0.65 in MJ and varying from 0.7 to 0.75 in JAS. It may be

attributed to the fact that although the large-scale atmospheric conditions are highly favorable, heavy rain does not always fall inside the studied area. Also, although Indochina is known for a strong diurnal variation of rainfall and convection, heavy rainfall-inducing WP in this study is analyzed only on the basis of daily means. Another potential reason is the coarse resolution of ERA-Interim as the training sample for SOM puts a strong limit on the capability of obtaining an adequate classification result. Nevertheless, the ETS of ASOM in MJ exceeds 0.14 for a 1-day forecast and then decreases gradually to 0.13 for longer forecast ranges. However, in JAS, it is only 0.097 for the first day forecast and then decreases rapidly to 0.074 and 0.055 for the next two and three forecast days, respectively.

The diminished performance of ASOM in JAS is mainly due to the higher FAR, since its POD in JAS is only slightly lower than that in MJ. Note that GSM also performs worse in JAS, especially for short lead (1–2 days) forecasts. As shown in the previous section, heavy rainfall over the Upper Nan in MJ is mostly caused by an intensification of the large-scale monsoon westerlies, while in JAS it is associated with active TDs originating in the SCS and western North Pacific.

FIG. 10. Forecast skill score: (a),(b) POD, (c),(d) FAR, and (e),(f) ETS for the heavy rainfall occurrence over the Upper Nan river basin in (top) MJ and (bottom) JAS during 2008–17 from ASOM and the conventional method (GSM) with 1–3-day lead times.

Both ASOM and GSM thus generally fall short in capturing heavy rainfall days attributed to the mesoscale phenomena in JAS, as a result of their relatively coarse resolutions.

b. Difference between clusters

In this subsection, the performance of ASOM for each individual cluster from C1 to C8 during 2008–17 is evaluated (Fig. 11). In MJ, it is shown that C1, C2, and C3 have quite similar FARs. Thus, C2 has the highest ETS of over 0.275 for all forecast ranges, since its POD is remarkably better than those of other clusters. For a 1-day forecast, the ETS of C3 is 0.286, which is slightly higher than that of C2. However, its ETS decreases rapidly to 0.255 and 0.25 for the next two forecast ranges, consistent with the decrease (increase) in its POD (FAR). As a result of inferior POD (i.e., than those of the two aforementioned clusters), C1 has a lower ETS of 0.23 for lead days 1–2, and then decreases to 0.2 for 3-day lead time. Surprisingly, C4, which has an ETS of zero, indicates that ASOM has no forecast skill with the heavy rainfall causing by the WP of this cluster. This worst behavior may be due to an infrequent occurrence of C4 that happens only five times in 10 years from 2008 to 2017 (Table 1).

In JAS, ASOM has the highest skill for the WP of C8 with ETS values of 0.32, 0.33, and 0.26 for lead times

from 1 to 3 days, respectively, which are considerably higher than those from other clusters. Interestingly, this best performance is attributed to the lowest FAR, since the POD of C8 is lower to those of C5 and C6. Although having the highest POD, C6 possesses the lowest ETS values that are less than 0.17 for all forecast ranges, due to its worst FAR. The forecast skill of C5 is also largely suppressed in the same way. Whereas, in spite of having the lowest PDO, C7 still outperforms C6, especially in the forecast range of 3 days, thanks to a much better FAR.

In short, these results suggest that the performance of ASOM depends on the synoptic WPs that bring heavy rainfall events. More specifically, the skill score is higher if the WP is limited to one phenomenon such as intensified westerlies (C2 and C3) or TDs (C5 and C8). In contrast, ASOM has more difficulties in capturing the WP imposing the interaction between the monsoon westerlies and the trade easterlies (C4 and C7) or northeasterly winds (C1).

c. Year-to-year variability

We now assess the year-to-year variation in the performance of ASOM by calculating the POD, FAR, and ETS for each year from 2008 to 2017 (Fig. 12). It shows that the ETS generally varies from 0.05 to 0.1, except in 2009, 2010, 2012, and 2015. The most skillful prediction is

FIG. 11. Forecast skill score: (a),(b) POD, (c),(d) FAR, and (e),(f) ETS for the heavy rainfall occurrence over the Upper Nan caused by the WP of C1–C4 and C5–C8 in (top) MJ and (bottom) JAS during 2008–17 from ASOM with 1–3-day lead times.

found in the summer of 2009 when its ETS is remarkably higher than those in other years, resulting from the lowest FAR and the second highest POD. Next, the ETS in 2012 is 0.18 and 0.16 for 1–2-day forecasts, respectively, before dropping to 0.11 for the 3-day forecast. On the other hand, the lowest POD of approximately 0.3 leads to the second worst forecast skill in 2010. Finally, ASOM is the least skillful in the drought year of 2015 when there were only 9 heavy rainfall days (Table 1).

TABLE 1. Number of the occurrence (days) of the observed WP from each cluster C1–C8 during 2008–17.

Year	Cluster								Total
	C1	C2	C3	C4	C5	C6	C7	C8	
2008	3	3	6	1	6	5	3	2	29
2009	6	1	14	0	4	3	2	3	33
2010	5	0	1	0	3	9	7	8	33
2011	7	3	3	1	7	10	1	8	40
2012	0	7	9	0	2	3	3	9	33
2013	5	1	5	0	2	8	4	3	28
2014	8	1	1	1	2	6	4	4	27
2015	2	0	0	0	2	0	4	1	9
2016	9	2	0	2	3	6	1	6	29
2017	18	1	0	0	6	7	7	5	44
Total	63	19	39	5	37	57	36	49	305

To examine the source of the variability in efficiency of ASOM, the performance of GSM in predicting atmospheric circulation is evaluated. Figure 13 illustrates the annual mean of the ACC and RMSE for pmsl, z500, u850, and v850 during 2008–17, in which quite strong year-to-year fluctuations are seen in the first two variables. It is suggested that the forecast skill of ASOM (for the heavy rainfall occurrence) and GSM (for the pmsl and z500 fields) is roughly in phase. Accordingly, the good (bad) results of ASOM in 2009 and 2012 (2010) are consistent with the high (low) forecast skill of GSM. However, the reason(s) for the lowest skill score of ASOM in 2015 while GSM performs relatively well is still unclear. We speculate that this is because small-scale processes may play the dominant role in causing local heavy rainfall in this dry year. Other interesting features that may strongly affect the performance of ASOM are the sharp decrease (increase) in the forecast skill of GSM for pmsl (z500) in 2016–17. These issues should be investigated in further studies.

6. Conclusions

In this study, the ASOM method developed by Nguyen-Le et al. (2017) was applied to summertime

FIG. 12. Forecast skill score: (a) POD, (b) FAR, and (c) ETS for the heavy rainfall occurrence over the Upper Nan in each year from 2008 to 2017 from ASOM with 1–3-day lead times.

heavy rainfall over the Upper Nan River basin in northwestern Thailand. At first, the SOM was used for classification of WP inducing heavy rainfall in MJ and JAS of 1979–2007. The classification results were then implemented with the PP and analog method to predict the occurrence (yes/no) of local heavy rainfall days during 2008–17 by using the prognostic information from GSM. As a result, the forecast skill in ASOM up to 3-day lead times basically outperforms the traditional method using only the predicted rainfall intensity from GSM. This indicates that ASOM can be used in the tropics where the mesoscale factors strongly control the local rainfall, by which it can provide an early warning and guidance for decision-making by local forecasters or end-users engaging in disaster prevention activities.

However, there are several limitations in this work. For instance, heavy rainfall days are selected based on APHRO_MA and TMPA-3B42V7, which have their own limitations, especially for high-terrain areas

(Yatagai et al. 2012). Since ASOM can predict whether heavy rainfall events occur but is not capable of providing the actual amount, questions may remain about its relative availability compared with other methods such as statistical downscaling used by Ohba et al. (2016). There has also been a great concern about using GSM output as being representative of the conventional method for forecasting heavy rainfall, since its spatial resolution is relatively too coarse to reproduce the convective system, particularly in the tropics. In addition, the FAR of ASOM is still high. Therefore, there is still room for development. Given the weaknesses of the current ERA-Interim dataset used as the training sample for SOM, classification results can be improved by using higher resolution and more accurate analysis and observations. With respect to the PP approach, the skill of ASOM also benefits from the improvements in numerical weather prediction models by providing “better” prognostic WPs. Further studies employing advanced methods should be

FIG. 13. Annual mean of the ACC and RMSE between the GSM forecasts with 1–3-day lead times and ERA-Interim reanalysis over 5.25°–30°N, 84.75°–120°E for (a),(c) pmsl, (b),(d) z500, (e),(f) u850, and (g),(h) v850 from 2008 to 2017.

conducted. For instance, it can be expected that a “fuzzy” clustering (e.g., fuzzy K -means) will better describe the WPs imposing the interaction between different factors than the “hard” K -means clustering used in the present study. Finally, an update in the algorithm taking account of the subdaily variations should be considered.

Acknowledgments. The authors greatly appreciate two anonymous reviewers for their thoughtful and constructive comments. We also thank Dr. Le Duc of JAMSTEC and Mr. Duc Tran-Anh of VNHMS for their suggestions. This research is supported by 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); and the Social Implementation of Climate change Adaptation Technology (SICAT) project of the Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT).

APPENDIX A

Verification Indices for Atmospheric Circulation

For quantitative verification of GSM in predicting atmospheric variables, the anomaly correlation coefficient (ACC) and root-mean-square error (RMSE) were used.

a. Anomaly correlation coefficient

The ACC is the correlation between anomalies of forecasts and those of verifying values with the reference values, such as climatological values. Here, ACC is defined as follows:

$$\text{ACC} = \frac{\sum_{i=1}^N [(F_i - C_i) - (\overline{F_i - C_i})][(O_i - C_i) - (\overline{O_i - C_i})]}{\sqrt{\sum_{i=1}^N (F_i - C_i)^2 \sum_{i=1}^N (O_i - C_i)^2}},$$

where F_i , O_i , and C_i represent forecast, observation, and reference values, respectively. The climatological mean during 1981–2010 is used as a reference value. Here, N is the number of samples. If the variation pattern of the anomalies of forecast is perfectly coincident with that of the anomalies of verifying value, ACC will take the maximum value of 1. In turn, if the variation pattern is completely reversed, ACC takes the minimum value of -1 .

b. Root-mean-square error

RMSE is often used for representing the accuracy of forecasts, and is defined by

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (F_i - O_i)^2},$$

where F_i and O_i represent the forecast and observation, respectively; and N is the number of samples. If RMSE is closer to zero, it means that the forecasts are closer to the verifying values. For perfect forecast, RMSE is equal to zero.

APPENDIX B

Verification Indices for Heavy Rainfall Forecasts

For quantitative verification of the ASOM method and GSM in predicting the occurrence of heavy rainfall days over the Upper Nan, the probability of detection (POD), false alarm ratio (FAR), and equitable threat score (ETS) were computed, based on a 2×2 contingency table (hits: forecast yes, observation yes; false alarms: forecast yes, observation no; misses: forecast no, observation yes; and correct negatives: forecast no, observation no; Wilks 1995). All these indexes were calculated for the rainfall thresholds of 40 (45) mm day⁻¹ in May–June (July–August–September).

a. Probability of detection

The POD gives the fraction of rain occurrences that were correctly detected. It ranges from 0 to 1 with a perfect score of 1:

$$\text{POD} = \frac{\text{hits}}{\text{hits} + \text{misses}}.$$

b. False alarm ratio

The FAR measures the fraction of rain detections that were actually false alarms. It ranges from 0 to 1 with a perfect score of 0:

$$\text{FAR} = \frac{\text{false alarms}}{\text{hits} + \text{false alarms}}.$$

c. Equitable threat score

The ETS measures the fraction of observed and/or detected rain that was correctly detected, adjusted for the number of hits_{random} that could be expected purely due to random chance (e.g., it is easier to correctly forecast rain occurrence in a wet climate than in a dry

climate). The ETS ranges from $-1/3$ to 1 with 0 indicating no forecast skill and a perfect score of 1:

$$\text{ETS} = \frac{\text{hits} - \text{hits}_{\text{random}}}{\text{hits} + \text{misses} + \text{false alarms} - \text{hits}_{\text{random}}},$$

where

$$\text{hits}_{\text{random}} = \frac{(\text{hits} + \text{misses})(\text{hits} + \text{false alarms})}{\text{hits} + \text{misses} + \text{false alarms} + \text{correct negatives}}.$$

REFERENCES

- Borah, N., A. K. Sahai, R. Chattopadhyay, S. Joseph, S. Abhilash, and B. N. Goswami, 2013: A self-organizing map-based ensemble forecast system for extended range prediction of active/break cycles of Indian summer monsoon. *J. Geophys. Res. Atmos.*, **118**, 9022–9034, <https://doi.org/10.1002/jgrd.50688>.
- Chan, J. C., 2000: Tropical cyclone activity over the western North Pacific associated with El Niño and La Niña events. *J. Climate*, **13**, 2960–2972, [https://doi.org/10.1175/1520-0442\(2000\)013<2960:TCAOTW>2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013<2960:TCAOTW>2.0.CO;2).
- Chattopadhyay, R., A. K. Sahai, and B. N. Goswami, 2008: Objective identification of nonlinear convectively coupled phases of monsoon intraseasonal oscillation: Implications for prediction. *J. Atmos. Sci.*, **65**, 1549–1569, <https://doi.org/10.1175/2007JAS2474.1>.
- Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, <https://doi.org/10.1002/qj.828>.
- Friederichs, P., 2010: Statistical downscaling of extreme precipitation events using extreme value theory. *Extremes*, **13**, 109–132, <https://doi.org/10.1007/s10687-010-0107-5>.
- , and A. Hense, 2007: Statistical downscaling of extreme precipitation events using censored quantile regression. *Mon. Wea. Rev.*, **135**, 2365–2378, <https://doi.org/10.1175/MWR3403.1>.
- Fudeyasu, H., S. Iizuka, and T. Matsuura, 2006: Seasonality of westward-propagating disturbances over Southeast and South Asia originated from typhoons. *Geophys. Res. Lett.*, **33**, L10809, <https://doi.org/10.1029/2005GL025380>.
- Gutiérrez, J. M., R. Cano, A. S. Cofiño, and C. Sordo, 2005: Analysis and downscaling multi-model seasonal forecasts in Peru using self-organizing maps. *Tellus*, **57A**, 435–447, <https://doi.org/10.1111/j.1600-0870.2005.00128.x>.
- Huffman, G. J., and Coauthors, 2007: The TRMM Multisatellite Precipitation Analysis: Quasi-global, multiyear, combined-sensor precipitation estimates at fine scale. *J. Hydrometeorol.*, **8**, 38–55, <https://doi.org/10.1175/JHM560.1>.
- , R. F. Adler, D. T. Bolvin, and E. J. Nelkin, 2010: The TRMM Multi-satellite Precipitation Analysis (TMPA). *Satellite Applications for Surface Hydrology*, F. Hossain and M. Gebremichael, Eds., Springer, 3–22.
- Japan Meteorological Agency, 2016: Joint WMO Technical Progress Report on the Global Data Processing and Forecasting System and Numerical Weather Prediction Research Activities for 2016. Appendix to WMO Numerical Weather Prediction Progress Report, JMA, Tokyo, Japan, 60 pp., http://www.jma.go.jp/jma/jma-eng/jma-center/nwp/report/2016_Japan.pdf.
- Kohonen, T., 1982: Self-organized formation of topologically correct feature maps. *Biol. Cybern.*, **43**, 59–69, <https://doi.org/10.1007/BF00337288>.
- , 2001: *Self-Organizing Maps*. Springer-Verlag, <https://doi.org/10.1007/978-3-642-56927-2>, 502 pp.
- Komori, D., and Coauthors, 2012: Characteristics of the 2011 Chao Phraya River flood in Central Thailand. *Hydrol. Res. Lett.*, **6**, 41–46.
- Lorenz, E. N., 1969: Atmospheric predictability as revealed by naturally occurring analogues. *J. Atmos. Sci.*, **26**, 636–646, [https://doi.org/10.1175/1520-0469\(1969\)26<636:APARBN>2.0.CO;2](https://doi.org/10.1175/1520-0469(1969)26<636:APARBN>2.0.CO;2).
- Mahalanobis, P. C., 1936: On the generalised distance in statistics. *Proc. Natl. Inst. Sci. India*, **2**, 49–55.
- Maraun, D., and Coauthors, 2010: Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. *Rev. Geophys.*, **48**, RG3003, <https://doi.org/10.1029/2009RG000314>.
- Nguyen-Le, D., J. Matsumoto, and T. Ngo-Duc, 2015: Onset of the rainy seasons in the eastern Indochina Peninsula. *J. Climate*, **28**, 5645–5666, <https://doi.org/10.1175/JCLI-D-14-00373.1>.
- , T. J. Yamada, and D. Tran-Anh, 2017: Classification and forecast of heavy rainfall in northern Kyushu during Baiu season using weather pattern recognition. *Atmos. Sci. Lett.*, **18**, 324–329, <https://doi.org/10.1002/asl.759>.
- Nishiyama, K., S. Endo, K. Jinno, C. B. Uvo, J. Olsson, and R. Berndtsson, 2007: Identification of typical synoptic patterns causing heavy rainfall in the rainy season in Japan by a self-organizing map. *Atmos. Res.*, **83**, 185–200, <https://doi.org/10.1016/j.atmosres.2005.10.015>.
- Ohba, M., S. Kadokura, D. Nohara, and Y. Toyoda, 2016: Rainfall downscaling of weekly ensemble forecasts using self-organising maps. *Tellus*, **68A**, 29293, <https://doi.org/10.3402/tellusa.v68.29293>.
- Sammon, J. W., 1969: A nonlinear mapping for data structure analysis. *IEEE Trans. Comput.*, **C-18**, 401–409, <https://doi.org/10.1109/T-C.1969.222678>.
- Takahashi, H. G., and T. Yasunari, 2006: A climatological monsoon break in rainfall over Indochina—A singularity in the seasonal march of the Asian summer monsoon. *J. Climate*, **19**, 1545–1556, <https://doi.org/10.1175/JCLI3724.1>.
- , H. Fujinami, T. Yasunari, J. Matsumoto, and S. Baimoung, 2015: Role of tropical cyclones along the monsoon trough in the 2011 Thai flood and interannual variability. *J. Climate*, **28**, 1465–1476, <https://doi.org/10.1175/JCLI-D-14-00147.1>.
- Ultsch, A., and H. P. Siemon, 1990: Kohonen's self organizing feature maps for exploratory data analysis. *Proc. Int. Neural Network Conf. (INNC'90)*, Paris, France, Kluwer Academic Press, 305–308.
- Vesanto, J., and E. Alhoniemi, 2000: Clustering of the self-organizing map. *IEEE Trans. Neural Networks*, **11**, 586–600, <https://doi.org/10.1109/72.846731>.
- Wang, B., and LinHo, 2002: Rainy season of the Asian-Pacific summer monsoon. *J. Climate*, **15**, 386–398, [https://doi.org/10.1175/1520-0442\(2002\)015<0386:RSOTAP>2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<0386:RSOTAP>2.0.CO;2).
- Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences: An Introduction*. Academic Press, 467 pp.

- Yatagai, A., K. Kamiguchi, O. Arakawa, A. Hamada, N. Yasutomi, and A. Kitho, 2012: APHRODITE: Constructing a long-term daily gridded precipitation dataset for Asia based on a dense network of rain gauges. *Bull. Amer. Meteor. Soc.*, **93**, 1401–1415, <https://doi.org/10.1175/BAMS-D-11-00122.1>.
- Ye, L., and M. Abe, 2012: The impacts of natural disasters on global supply chains. ARTNeT Working Paper 115, Bangkok, ESCAP, 28 pp., <https://www.econstor.eu/bitstream/10419/64267/1/717874087.pdf>.
- Yokoi, S., and J. Matsumoto, 2008: Collaborative effects of cold surge and tropical depression-type disturbance on heavy rainfall in central Vietnam. *Mon. Wea. Rev.*, **136**, 3275–3287, <https://doi.org/10.1175/2008MWR2456.1>.
- Zhang, Y., T. Li, B. Wang, and G. Wu, 2002: Onset of the summer monsoon over the Indochina Peninsula: Climatology and interannual variations. *J. Climate*, **15**, 3206–3221, [https://doi.org/10.1175/1520-0442\(2002\)015<3206:OOTSMO>2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<3206:OOTSMO>2.0.CO;2).