Define East Asian Monsoon Annual Cycle via a Self-Organizing Map-Based Approach

Lun Dai, Tat Fan Cheng, and Mengqian Lu

1Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, China
2Guangzhou HKUST Fok Ying Tung Research Institute, Guangzhou, China

Abstract Methods to define the East Asian monsoon usually involve ad hoc and diverse criteria. This study adopts a simple and coherent approach based on the self-organizing map to systematically derive the monsoon annual cycle based on 850-hPa wind fields. The derived onset dates in the warm season agree well with the literature despite lacking documentation for winter-season stages. Linear regression suggests an inverse relationship between the length of the Spring (Meiyu) stage and that of the Pre-Meiyu (Mid-summer) stage, suggesting predictability of the Meiyu and Fall onsets. The Aleutian Low, the western North Pacific subtropical high, and the upper-level westerlies play important roles in the Spring, Pre-Meiyu, and Meiyu onsets. Observations reveal strong rain belt events form right after the monsoonal onsets in the warm season. The proposed approach to deriving monsoon stages and onsets would be useful for monsoon climate studies and operational monsoon diagnosis and forecasting.

Plain Language Summary The East Asian monsoon profoundly affect the livelihoods of more than one-fifth of the global population. Given its impacts and complexity, many efforts have been taken to define the different monsoon stages’ arrival with diverse and inconsistent criteria. Instead of adding more complicated criteria, we adopt a machine-learning-based approach to derive the annual cycle of the East Asian monsoon stages in a simple and systematic way. The derived arrival timings of the warm-season stages are consistent with the literature despite lacking documentation for winter-season stages. Linear regression shows that the arrival of the preceding monsoon stages could inform that of the ensuing stage during the warm season. We find distinct weather systems that possibly explain the onset of monsoon stages. Further analysis indicates that heavy and persistent rain belt events in the warm season tend to form right after monsoonal wind stages’ onsets. The machine-learning-based approach to deriving the monsoon stages and onsets would be useful for monsoon climate research and practical use in operational monsoon diagnosis and forecasting, thereby short-term planning for agricultural activities and flood control.

1. Introduction

The East Asian monsoon has long been recognized as one of the most dynamic and complex monsoon systems on Earth (Wang, 2006). The East Asian monsoon stages, particularly in spring and summer, have been studied at length by the research community. As southwesterly winds first appear over Southeast China at the end of February, it is followed by the persistent spring rain in the mid-lower Yangtze River basin and South China in March and April. This stage is known as the first rainy stage in China (LinHo et al., 2008; Tian & Yasunari, 1998; Wang & LinHo, 2002). Starting from mid-May, a quasi-stationary rain belt over South China signifies the arrival of the pre-summer rainy (or Pre-Meiyu) season (Chiang et al., 2017; Ding & Chan, 2005; Tao & Chen, 1987). As the prevailing southwesterlies strengthen (G. T. J. Chen, 2004), the rain belt migrates northward to the mid-lower Yangtze River basin and South Japan in mid-June and stay there for a month, marking the arrival of the Meiyu/Baiu season in China and Japan. In mid-July, the monsoon front advances further north to affect the weather in the Huai and Yellow River basins and the Korean peninsula, implying the end of the Meiyu season and the start of mid-summer season (Chiang et al., 2017; Xu & Zipser, 2011). Meanwhile, another rain beltreviving over South China persists till the end of September due to typhoon activities and tropical disturbances (Chang & Chen, 1995; G. T. J. Chen, 2004; Cheng & Lu, 2020; Tian & Yasunari, 1998). During the winter monsoon season from November through March, northwesterly winds develop to the north of 20°N and become prevailing since mid-December (Chang et al., 2006; Jhun &
The northwesterlies are generally the strongest from mid-January to mid-February and decays afterward in the late winter (Kim et al., 2013).

Despite a holistic picture of the monsoon annual cycle in East Asia, the exact onset dates and durations of stages remain controversial due to diverse monsoon definitions, data source, and the intrinsic variability of the system (Luo et al., 2013; Martin et al., 2019; Tomita et al., 2011). Different criteria were proposed for monsoon onsets, which involve variables such as precipitation (Wang & LinHo, 2002; Xu et al., 2009), shear vorticity (Wang et al., 2008), equivalent potential temperature (Tomita et al., 2011), or a combination of them (LinHo et al., 2008; Luo et al., 2013). However, all these criteria-based methods were meticulously designed for one or a few monsoon stages only. For instance, four *ad hoc* criteria, with specific parameters for different variables, were proposed by LinHo et al. (2008) to define an onset date of the spring rain period in South China. Likewise, Xu et al. (2009) determined the Meiyu onset by applying complex rain band definitions. To the best of our knowledge, there is no method to date being proposed to define the monsoon annual cycle in a systematic and simple way. The simplicity and coherence of such a definition could be useful for both operational and research purposes. As such, we employ a neural-network-based algorithm, the self-organizing map (SOM), to classify the East Asian monsoon stages all year round. This method projects high-dimensional data onto a low-dimensional one and preserves the topological features (Kohonen, 2001), which has shown great potentials in classifying meteorological field patterns (Chiang et al., 2017; Dai et al., 2020; Johnson, 2013; Kong et al., 2017; Nishiyama et al., 2007; Pan & Lu, 2020). The results of the SOM-based approach are then assessed with the monsoon onset dates documented in the literature.

In line with the above motivations, this study consists of two parts. First, we introduce a SOM-based methodology to classify the monsoon annual cycle in East Asia and to derive the annual onset dates and explore their predictability. Second, we focus on exploring the attributable weather conditions before and after the onset of the warm-season monsoon stages, given their close relevance to devastating rainfall and floods. The relationship between the derived monsoonal wind stages and the occurrence of rain belt events (defined in Section 2.3) are also presented.

### 2. Data and Methods

#### 2.1. Data

Meteorological fields are retrieved from the fifth generation of the European Centre for Medium-Range Weather Forecast (ECMWF) atmospheric reanalysis data (ERA5) between 1979 and 2018 at 1° grid resolution (Copernicus Climate Change Service, 2017). Pressure-level fields at 25-hPa interval from 925 to 700 hPa are retrieved (except for the 725-hPa level). 850-hPa horizontal winds at 2.5° grid resolution from the National Centers for Environmental Prediction and the National Center for Atmospheric Research (NCEP/NCAR) reanalysis-1 (Kalnay et al., 1996) are also retrieved for results comparison. We estimate the equivalent potential temperature $\Theta_e$ (K) using the empirical equation by Bolton (1980), which can be found in the supporting information (SI) (Text S1).

The vertical mean of $\Theta_e$ (K) in the lower troposphere is computed from

$$\langle \Theta_e \rangle_{925-700} = \frac{\int_{925}^{700} \Theta_e \, dp}{\int_{925}^{700} dp}, \quad (1)$$

where $p_1$ and $p_2$ refer to 700 and 925 hPa, respectively. $\langle \cdot \rangle$ denotes the vertical mean. Likewise, the vertical mean of the horizontal advection of $\Theta_e$ in the lower troposphere is $\langle -V \cdot \nabla \Theta_e \rangle_{925-700}$, where $V$ (m s$^{-1}$) is the horizontal wind at the pressure level. The 200-hPa stationary wave is referred to as the 200-hPa geopotential height departure from the zonal mean.

#### 2.2. The SOM Training and Testing

We trained the SOM using the 5-day-moving-mean daily climatology of 850-hPa zonal and meridional winds over the East Asian monsoon region (110°E-140°E, 20°N-45°N, following Wang & LinHo, 2002). Not...
only that monsoon is traditionally defined by the seasonal reversal of near-surface winds (Krishnamurti et al., 2013), but also the 850-hPa wind fields can best describe the prevailing weather patterns during both summer and winter monsoon seasons. In contrast, the other key monsoonal variable, precipitation, is zero-bounded and thus deficient in depicting weather regimes in the dry season.

Since we adopt the Euclidean distance in the SOM algorithm, the wind fields, prior to the training, were weighted by the square root of the cosine of latitude to correctly account for the difference in grid area across latitudes (Johnson et al., 2008). To assess the performance, we conduct the false discovery rate (FDR) test (Benjamini & Hochberg, 1995; Wilks, 2006) on the number of indistinguishable pairs in the clustering results from 1 × 3 to 1 × 10 SOM configurations. In principle, the ratio of indistinguishable pairs is proportional to the number of clusters. As a result, we adopt the 1 × 8 configuration to capture all the important monsoon stages, as mentioned in Section 1, while ensuring a low ratio of indistinguishable pairs (Figure S1). A detailed illustration of the FDR test is given in the SI (Text S2) or can be found in our previous study (Dai et al., 2020).

2.3. Detection of Heavy and Persistent Rain Belt Events

Heavy and persistent rain belt events are identified to explore their relationship with monsoonal wind stages. The grid-based threshold for heavy rainfall is obtained by spatially smoothing the 75th quantile of the accumulated precipitation on wet days (greater than 1 mm/day) using Gaussian kernel smoothing (Lu et al., 2015). Heavy rainfall events with spatial disconnection (larger than 2° in longitudes or latitudes) and temporal discontinuity (more than 1 day) are excluded. Further, events having a zonal scale of at least 10° in longitude and temporal persistence of at least 3 days are classified as heavy and persistent rain belt events. Readers may find an example of a detected rain belt event and the local thresholds map in Figure S2.

3. The Annual Cycle of the East Asian Monsoon Stages and Their Onset Dates

3.1. Weather Patterns of the SOM-Derived Monsoon Stages

Eight distinct monsoonal wind stages are classified using the SOM approach (Section 2.2, Figure 1). Specifically, the clustered Patterns 1, 2, and 3 (Figures 1a–1c and 1i) exhibit anticyclonic circulations with strong northwesterlies to the north of 25°N-30°N and northeasterlies to the south. These wind regimes reveal the typical wintertime conditions documented in the literature (Jhun & Lee, 2004). In line with the circulations, the 330 K \( \theta_e \) contour line, which represents the nominal boundary between the tropical and polar air mass (Tomita et al., 2011), mainly resides at 20°N or further south, suggesting that the polar air masses encompass the entire East Asia in the three patterns (Figures 1a–1c). Notably, we identify Pattern 2 as the Mid-winter stage (Figure 1b) as it features the weakest precipitation intensity and the strongest northwest-erlies to the north of 25°N (Kim et al., 2013), alongside with the weakest precipitation intensity and the coldest-and-driest air masses in the region as the 320 K \( \theta_e \) contour line clears the southern boundary of East Asian land. The stage before the Mid-winter (i.e., Pattern 1) is thereby termed as the Early-winter (Figure 1a). Pattern 3 is identified as the Late-winter stage as southwesterlies emerge south of 30°N and mark the decay of the winter monsoon (Figure 1c).

Following the winter season, Pattern 4 unveils an overall weakened northwesterlies in the region with moderate southwesterlies and precipitation over South China (Figure 1d). Here comes the Spring stage. As the summer monsoon approaches, Pattern 5 orchestrates prevailing southwesterlies and a well-built rain belt in the southern part of the monsoon region (Figure 1e), which well aligns with the Pre-Meiyu weather patterns (e.g., Ding & Chan, 2005). Subsequently, Pattern 6 depicts the well-known Meiyu stage in which the quasi-stationary rain belt propagates northward to the mid-lower Yangtze River basin, South Japan and even the Korean Peninsula (Figure 1f). Meanwhile, southwesterlies become the strongest with tropical air masses (i.e., greater than 330 K in \( \theta_e \)) spreading all over the region. During the Mid-summer stage (Pattern 7), the southwesterlies and the Meiyu rain belt are much suppressed, while the rain band appears to emerge and persist in the Korean Peninsula, South China coast and the adjacent seas (Figure 1g). In the Fall stage (Pattern 8), southwesterlies almost disappear, followed by the developing northerlies and dry conditions (Figure 1h). This stage marks the transition of the weather regime from summer to winter monsoon.
Given the climatological wind patterns from the annual cycle (Figure 1), we derive the annual onset dates of all the monsoon stages via the following steps. We first determine the three nearest monsoon stages for each day, referring to the identified monsoon climatological calendar (Figure 1i). We then assign each day to the one among the three candidate stages based on the minimum Euclidean distance in the 5-day-moving-mean 850-hPa wind fields. After the assignment, each monsoon stage’s onset date is defined as the starting day when the stage first persists for at least 5 consecutive days. Compared to the results using thresholds of 3 or 7 consecutive days, the onset timings based on the 5-day threshold (Table S2) are closest to those documented in the literature (Table 1). We also repeated the above procedures using NCEP/NCAR reanalysis-1 for comparison. The results largely agree with that using the ERA5 reanalysis and are thus omitted here (Figure S3). A full table of annual onset dates is given in Table S1.

Notably, the standard deviation of the onset dates reaches 2–3 weeks from the Mid-winter to the Pre-Meiya stage, compared to fewer than 2 weeks from the Meiyu to the Fall stages (Table 1). This finding suggests a stronger interannual variability of the wintertime and springtime monsoon onsets. In particular, the standard deviation of the Meiyu stage’s onset date is only 9 days. Considering that the Meiyu stage represents the
Table 1
The Onset Dates of Monsoon Stages Derived in This Work and Previous Studies

| No. | References | Methods | Parameters | Data set | Time Period | Onset dates (mm/dd ± days) |
|-----|------------|---------|------------|----------|-------------|--------------------------|
| 1   | Kong et al. (2017) | SOM-based | 850-hPa winds | ERA5 | 1979–2018 (40-year) | 12/02 ± 20 |
| 2   | Lau and Yang (1997) | Criteria-based | P | APRHODITE | 1951–2007 (57-year) | 1/18 ± 22 |
| 3   | Tian and Yasunari (1998) | SOM-based | P | GPI | 1986–1994 (9-year) | 4/01 |
| 4   | Wang and LinHo (2002) | Criteria-based | P | 1,200 stations from GDS | 1980–1989 (10-year) | 2/25–3/01 |
| 5   | Ding and Chan (2005) | Criteria-based | P | CMAP | 1979–1998 (20-year) | 3/15 |
| 6   | LinHo et al. (2008) | Criteria-based | P, 500- and 200-hPa winds and OLR® | CMAP | 1979–2001 (23-year) | 3/12–16 ± 10 |
| 7   | Tomita et al. (2011) | Criteria-based | P, 850-hPa zonal winds and 850-hPa θe | GPCP & NCEP-NCAR | 1979–2002 (24-year) | 6/02 ± 8 |
| 8   | Luo et al. (2013) | Criteria-based | P | JRA | 1979–2008 (30-year) | 6/17 ± 11 |
| 9   | This study | SOM-based | P | NCEP-NCAR | 1979–2002 (30-year) | 6/17 ± 11 |

Note. Uncertainties are given based on one standard deviation if onset dates for individual years are available.

aPrecipitation (P). bOutgoing longwave radiation (OLR).
peak season of southwesterlies and monsoon fronts (Figure 1f), strong land-sea and meridional thermal contrasts are therefore required and are largely determined by periodic factors like the seasonally varying insolation, which in turn may explain the relatively weak variability of the Meiyu onset timing.

Onset dates of the warm-season monsoon stages are well-documented in the literature, and our results are in the closest agreement with the Spring, Pre-Meiyu, Meiyu and Mid-summer stages as reviewed in Ding and Chan (2005) (Table 1). Notably, the monsoon onset dates given in most of the listed studies are within one standard deviation of the onset dates derived in this work. As those studies worked on designing ad hoc criteria to ideally fit the weather regime during the concerned monsoon stage(s), the consistency in the onset timings suggests that the SOM-based approach can be a simple and robust alternative to determine the monsoon stages. We notice that the climatological onset dates derived by Kong et al. (2017) using the SOM approach on precipitation data are about 2 weeks behind our results, which may be attributable to only a half-year data (i.e., April to September) used in their study. As the Spring stage's onset can occur before April (Table 1), it is important to take the whole year into consideration when identifying monsoon stages. In addition, despite the lack of documented onset timings for winter-season stages, we believe the results carry some creditability given the cross-validated summer monsoon onsets and the largely consistent wintertime wind patterns discussed in Section 3.1.

3.3. Predictability of Monsoon Stages’ Onset Timing

The seasonal prediction of the monsoon onset, especially in the rainy season, is crucial but remains challenging. With the derived monsoon onset dates, we desire to know if the earlier monsoon stages can offer predictability on the subsequent stages. For this purpose, we explore the relationship between adjacent stages in terms of the onset time interval from one stage to another. The linear regression reveals that the onset time interval between the Spring and Pre-Meiyu \( T_{S,P} \) can explain 45% of the variance in that between the Pre-Meiyu and Meiyu \( T_{P,M} \) with a negative slope coefficient significant at the 0.01 level (Figure S4). That means, the longer the length of the Spring stage (i.e., \( T_{S,P} \)), the shorter the Pre-Meiyu stage (i.e., \( T_{P,M} \)) is expected. Likewise, an inverse relationship with \( R^2 \) of 0.51 is also found between the onset time interval from Meiyu to Midsummer \( T_{M,MS} \) and that from Midsummer to Fall \( T_{MS,F} \). The two regression models demonstrate the predictability of the Meiyu and Fall stages’ onset dates. For instance, once the onset dates of the Spring and Pre-Meiyu stages are observed, we can obtain the regressed onset time interval between the Pre-Meiyu to Meiyu stage using the regression equation (Figure S4a), and thereby the predicted onset date of the ensuing Meiyu stage. We note that other regression methods, such as polynomial and loess regression, provide no improved results despite their flexibility. And the inverse relationship between the onset timings of adjacent stages is not observed among the other pairs of stages.

4. Weather Regimes Associated With the Monsoon Stages’ Onset in the Warm Season

It is important to understand how the monsoon stages initiate in the first place and the weather systems at play. Given the critical hydrological impacts from warm-season stages, we limit our diagnosis to the weather regimes before and after the Spring, Pre-Meiyu, Meiyu, Mid-summer, and Fall onsets. We compute the daily composites and average them every 5 days (i.e., pentad) before and after the monsoon onset to highlight the prevailing weather with high-frequency noises removed. Here, pentad 0 denotes the period from the onset date (i.e., day 0) to 4 days after (i.e., day 4), pentad −1 refers to another five days before day 0, and a similar convention applies to others. The anomaly fields in the composites represent the deviations from the mean fields averaged from 15 days before and 9 days after the stage’s onset day. For a better presentation of the signals, the sets of lead-lag pentads in the composites are slightly different across different stages.

4.1. Spring (Mar/14 ± 16 Days) and Pre-Meiyu (May/04 ± 18 Days)

A strong Aleutian low is observed concurrent with a weak western North Pacific subtropical high (WNPSH) at least 2 weeks before the Spring onset (Figure 2a). As the onset approaches, the Aleutian Low notably
weakens while the WNPSH extends northward, manifesting an abrupt switch from negative to positive 850-hPa geopotential anomaly over North Pacific and Japan (Figure 2b). It is likely that the complete reversal of the pressure dipole anomaly from pentad −3 to pentad 0 (Figure 2b) leads to enhanced southerlies in East Asia seen in both raw and anomaly fields (Figures 2c and 2d) and thereby explain the onset of the Spring stage. The enhanced southerlies further advect warm-and-moist air masses (i.e., \((-V \cdot \nabla \theta) > 0\)) from the...
tropics into East Asia (Figures 2c and 2d). The instability generated from the warm-and-moist air at lower-levels (Figure 2c) and cold-and-dry air from the upper-level stationary trough (Figure 2e) contribute to spring rainfall over South China (Figures 2a and 2b).

Similar weather conditions progress into the Pre-Meiyu stage. The weakening of Aleutian Low continues (Figure S5a) while the WNPSH expands further north, which is again associated with a reversal pressure dipole anomaly from pentad −3 to pentad 0 (Figure S5b). As the western ridge of the WNPSH hovers over the South China Sea and the Philippine Sea, a continental low-pressure anomaly develops along with prevailing westerlies (Figures S5b and S5c), favoring rainfall formation in South China and the adjacent seas (Figures S5a and S5b). The enhanced rainfall signals are in line with more frequent monsoon rain belt events in the south (20°N-30°N) right after the Spring and Pre-Meiyu onsets (Figures S6a and S6b). This finding suggests a close relationship between rain belt events and the onset of the monsoonal wind stages.

Changes in the upper-level weather conditions around the onsets are also notable. Specifically, the upper-level westerly jet upstream (downstream) of the Tibetan Plateau significantly strengthens (weakens) in pentad 0 of the Spring and Pre-Meiyu stages (Figures 2f and S5f). The anomalous upper-level jet possibly results from the seasonal shift of the westerlies, from which the westerlies over the plateau intensifies and subsequently favors the downslope convergence and monsoon rainband during the early warm season (Chiang et al., 2019; Park et al., 2012).

4.2. Meiyu (Jun/12 ± 9 Days)

The preconditioned weather systems begin to differ as it approaches the Meiyu onset. As the influence of Aleutian low to East Asia vanishing, the WNPSH's western ridge rapidly strengthens and stretches to the northwest (Figure 3a), resulting in an anomalous high and drought conditions over the Philippine Sea in pentad 0 (Figure 3b). Since then, the westerly monsoon becomes the strongest and continuously fuels warm-and-moist air to the rain fields confined to the northwest of the WNPSH (Figures 3b, 3c and 3d). These observations reveal the typical weather regimes during the westward extension of the WNPSH (Cheng et al., 2019). The lower-level westerly monsoon over East Asia assist the northward vorticity advection and favor ascending motions via the Sverdrup balance (Rodwell & Hoskins, 2001), eventually spurring the Meiyu/Baiu rain belts in the Yangtze river basin and Japan (Figures 3a and S6c). We argue that the western extension of the WNPSH is accountable to the onset timing of the Meiyu stage, as the WNPSH are found influential to the Meiyu/Baiu front formation (Chang et al., 2000; Ninomiya, 1984; Wang & LinHo, 2002).

Aloft, the upper-level westerlies migrate northward to the northern edge of the Tibetan Plateau at 40°N (Figure 3e) and juxtapose with enhanced precipitation in the Yangtze river basin and South Japan during the Meiyu onset (Figures 3a and 3b). Given the rapid migration of the westerly jet right after the Meiyu onset (Figure 3f) and its dynamic controls on the East Asian summer monsoon widely reported by other studies (J. Chen & Bordoni, 2014; Park et al., 2012; Samppe & Xie, 2010), it is deemed another determinant of the onset timing of the derived Meiyu stage (Chiang et al., 2017; Kong et al., 2017).

4.3. Mid-Summer (Jul/19 ± 13 Days) and Fall (Sep/03 ± 12 Days)

Recall that the southwesterly monsoon starts to diminish since the Mid-summer stage (Figure 1g), we notice a sudden northward migration of the subtropical high in pentad 0 (Figure S7a), which manifests a pressure dipole consisting of a high (low) pressure anomaly in the north (south) (Figure S7b). In this case, the pressure dipole weakens the southwest monsoons over East Asia (Figures S7c and S7d). Moreover, the Meiyu rain belt vanishes while intense rainfall revives in the south since pentad 0 (Figures S7a and S7b). It is noteworthy that the Meiyu stage ends when the upper-level westerlies just clear the northern edge of the plateau (Figures S7e and S7f), suggesting the role of the westerlies on the demise of the Meiyu and the start of the Mid-summer (Chiang et al., 2017; Kong et al., 2017).

As the Fall stage begins, the western ridge of the WNPSH becomes less organized and extends further inland of East Asia (Figure S8a). The preconditioning of the Siberian high manifested from a high-pressure anomaly (Figure S8b) and anomalous northeasterlies (Figure S8d) becomes prominent over East Asia, causing
the advection of dry-and-cool air from the northeast (Figures S8c and S8d) and less rainfall (Figure S8b). Meanwhile, the westerlies return southward to the Tibetan Plateau (Figures S8e and S8f) probably due to the southward retreat of the maximum solar radiation and a stronger meridional temperature gradient by the cold air advection from the northeast (Jhun & Lee, 2004). These signals mark the end of the summer monsoon season and the transition to the wintertime weather regime.

Additionally, the latitudinal position of rain belts in the Mid-summer is mainly confined in the south (20°N–25°N) (Figure S6d), which is likely to result from the influence of tropical cyclones at this time (Cheng & Lu, 2020; Day et al., 2018; Pan & Lu, 2020). Rain belts then largely disappear in the Fall stage (Figure S6e) due to weaker moisture transports from the tropics. Further, the zonal mean precipitation intensity (Figures S6f–S6j) and the zonal extent (Figures S6k–S6o) of rain belts also undergo abrupt changes.
after monsoon onsets. These findings suggest the changes in the characteristics (e.g., location, extent, and intensity) of rain belt events are closely subsequent to the monsoonal wind stages’ onsets.

5. Discussion

In addition to the role of low-level circulations and westerly jets on monsoon onsets, the thermal forcing of the plateau was proposed to affect the EASM (Hsu & Liu, 2003; Yanai & Wu, 2006). However, our further analysis reveals no significant anomalies of the mid-tropospheric temperature over the plateau during the monsoon onset in the warm season (results not shown), which implies that the thermal forcing may be a minor factor for the interannual variability of summer monsoon onset.

As we learn a close relationship between monsoonal wind stages’ onsets and rain belt occurrence, the wind-rain belt relationship may be qualitatively understood through the Sverdrup balance (Rodwell & Hoskins, 2001). The planetary vorticity advection caused by warm-and-moist southerlies at lower-levels is approximately balanced by the stretching effect of air columns due to condensational heating, which feedbacks to stronger convection and thereby frontal rains. This explanation may be further supported by the frontal rain analysis by Day et al. (2018) and the finding from J. Chen & Bordoni (2014) that the meridional stationary eddy velocity is key for temperature advection and the Meiyu rainfall.

A recent study demonstrated higher predictive skills of the GloSea5 hindcast ensembles when employing the 850-hPa zonal wind index to determine the onset of the South China Sea Summer Monsoon (Martin et al., 2019). Thus, applying the SOM approach on the 850-hPa wind fields to derive the monsoon annual cycle could be useful for the operational monsoon forecasting. In addition, previous studies suggested that climate forcings, such as the El Niño–Southern Oscillation (Ding et al., 2020; Okada et al., 2017), sea surface temperature anomalies in the tropical Indian Ocean (He & Zhu, 2015) and greenhouse warming (Kitoh et al., 2013) could cause early or late onset of East Asian monsoon in past or future climates, which will be explored in our future work.

6. Conclusions

In the present study, we adopt a SOM-based approach on the 850-hPa wind fields to derive eight East Asian monsoon stages throughout a year in a systematic and coherent manner. The derived onset dates of the warm-season stages are cross-validated with the literature. Further, the preceding monsoon stages could inform the arrival of the ensuing stage during the warm season. Specifically, the onset time interval from the Spring to Pre-Meiyu (Meiyu to Mid-summer) can explain 45% (51%) of the variance in that from the Pre-Meiyu to Meiyu (Mid-summer to Fall) with a negative slope coefficient based on linear regression. In other words, the shorter the Spring (Meiyu) stage, the longer the Pre-Meiyu (Mid-summer) stage is expected. This finding reveals the predictability of the Meiyu and Fall onsets.

The weather diagnoses further reveal the interplay between the weakening of the Aleutian Low, the northward extension of the WNPSH and the upper-level westerlies' behavior during the Spring and Pre-Meiyu onset. Particularly, the westward extension of the WNPSH and the northward migration of the westerlies are likely to determine the Meiyu onset. In contrast, onsets of the Mid-summer and Fall stage coincide with a northwestward positioned WNPSH and thereby a high-pressure anomaly over the East Asian continent, which induces the demise of the summer monsoon. Finally, the observation of heavy and persistent rain belt events form right after the wind stages’ onsets suggest a close wind-rain belt relationship. The SOM-based approach to deriving the monsoon stages and onsets would be useful for monsoon climate research and practical use in operational monsoon diagnosis and forecasting, thereby short-term planning for agricultural activities and flood control.

Data Availability Statement

The meteorological data is retrieved from the ERA5 by the European Center for Medium-Range Weather Forecast (ECMWF) at https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5. The NCEP/NCAR Reanalysis 1 by NOAA Physical Sciences Laboratory is retrieved from https://psl.noaa.gov/data/grid-ded/data.ncep.reanalysis.html. The SOM training is executed by the R package “kohenon” (Wehrens, 2007).
The Gaussian kernel smoothing is implemented using an R package “smoothie” (Gilleland, 2013). The second criterion of the rain belt event detection involves using an R package “dbscan” (Hahsler et al., 2019). Some figures in this paper are generated using a Matlab package “M_Map” (Pawlowicz, 2020).

Acknowledgments
This research contributes to, and is financially supported by the National Natural Science Foundation of China funded project (No. 51709051) and the Hong Kong Research Grants Council funded project (Nos. 26200017 and 16201218). The authors thank three anonymous reviewers for their insightful comments to improve the manuscript. This work is part of Lun Dai’s and Tat Fan Cheng’s doctoral theses.

References
Benjamin, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bolton, D. (1980). The computation of equivalent potential temperature. *Monthly Weather Review*, 108(7), 1046–1053. https://doi.org/10.1175/1520-0493(1980)108<1046:TEPET>2.0.CO;2

Chang, C.-P., & Chen, G. T. J. (1995). Tropical circulations associated with Southwest Monsoon onset and westerly surges over the South China Sea. *Monthly Weather Review*, 123(11), 3254–3267. https://doi.org/10.1175/1520-0493(1995)123<3254:ctawsm>2.0.co;2

Chang, C.-P., Wang, Z., & Hendon, H. (2006). The Asian winter monsoon. In B. Wang (Ed.), *The Asian monsoon* (pp. 787. Berlin, Heidelberg: Springer-Verlag.

Chang, C.-P., Zhang, Y., & Li, T. (2000). Interannual and interdecadal variations of the East Asian summer monsoon and Tropical Pacific SSTs. Part I: Roles of the subtropical ridge. *Journal of Climate*, 13(24), 4310–4325. https://doi.org/10.1175/1520-0493(2000)13<4310:IIIVOT>2.0.CO;2

Chen, G. T. J. (2004). Research on the phenomena of Meiyu during the past quarter century: An overview. In C.-P. Chang (Ed.), *East Asian monsoon* (pp. 357–403). World Scientific. https://doi.org/10.1142/9789812701411_0010

Chen, J., & Bondoni, S. (2014). Orographic effects of the Tibetan Plateau on the East Asian summer monsoon: An energetic perspective. *Journal of Climate*, 27(8), 3052–3072. https://doi.org/10.1175/JCLI-D-13-00479.1

Cheng, T. F., & Lu, M. (2020). Moisture source-receptor network of the East Asian summer monsoon land regions and the associated atmospheric flow. *Journal of Climate*, 33(21), 9213–9231. https://doi.org/10.1175/jcli-d-19-08681.1

Cheng, T. F., Lu, M., & Dai, L. (2019). The zonal oscillation and the driving mechanisms of the extreme western North Pacific tropical high and its impacts on East Asian summer precipitation. *Journal of Climate*, 32(10), 3025–3050. https://doi.org/10.1175/JCLI-D-18-0076.1

Chiang, J. C. H., Fischer, J., Kong, W., & Herman, M. J. (2019). Intensification of the pre-Meiyu rainband in the late 21st century. *Geophysical Research Letters*, 46(13), 7356–7354. https://doi.org/10.1029/2019GL083383

Chiang, J. C. H., Swenson, L. M., & Kong, W. (2017). Role of seasonal transitions and the westerlies in the interannual variability of the East Asian summer monsoon precipitation. *Geophysical Research Letters*, 44(8), 3788–3795. https://doi.org/10.1002/2017GL072739

Copernicus Climate Change Service (2017). ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate. Copernicus Climate Change Service Climate Data Store (CDS). Retrieved from https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview

Dai, L., Cheng, T. F., & Lu, M. (2020). Summer monsoon rainfall patterns and predictability over Southeast China. *Water Resources Research*, 56(2), 1–21. https://doi.org/10.1029/2019WR025515

Day, J. A., Fung, I., & Liu, W. (2018). Changing character of rainfall in eastern China, 1951-2007. *Journal of Climate*, 31(20), 1023–1049. https://doi.org/10.1175/jcli-d-19-0438.1

Ding, Y., & Chan, J. C. L. (2005). The East Asian summer monsoon: An overview. *Meteorology and Atmospheric Physics*, 91(1–4), 117–142. https://doi.org/10.1007/s00703-005-0125-z

Ding, Y., Liang, P., Liu, Y., & Zhang, Y. (2020). Multiscale variability of Meiyu and its prediction: A new review. *Journal of Geophysical Research: Atmospheres*, 125(7), 1–28. https://doi.org/10.1029/2019JD031496

Gilleland, E. (2013). Package ‘smoothie’.

Hahsler, M., Piekenbrock, M., & Doran, D. (2019). Dbscan: Fast density-based clustering with R. *Journal of Statistical Software*, 91(1). https://doi.org/10.18637/jss.v091.i01

He, J., & Zhu, Z. (2015). The relation of South China Sea monsoon onset with the subsequent rainfall over the subtropical East Asia. *International Journal of Climatology*, 35(15), 4547–4566. https://doi.org/10.1002/joc.4305

Hsu, H. H., & Liu, X. (2003). Relationship between the Tibetan Plateau heating and East Asian summer monsoon rainfall. *Geophysical Research Letters*, 30(20), 30–33. https://doi.org/10.1029/2003GL017909

Jhu, J. G., & Lee, E. J. (2004). A new East Asian winter monsoon index and associated characteristics of the winter monsoon. *Journal of Climate*, 17(4), 711–726. https://doi.org/10.1175/1520-0442(2004)017<0711:ANAWMI>2.0.CO;2

Johnson, N. C. (2013). How many ENSO flavors can we distinguish?. *Journal of Climate*, 26(13), 4816–4827. https://doi.org/10.1175/JCLI-D-12-00649.1

Johnson, N. C., Feldstein, S. B., & Tremblay, B. (2008). The continuum of northern hemisphere teleconnection patterns and a description of the NAO shift with the use of self-organizing maps. *Journal of Climate*, 21(23), 6354–6371. https://doi.org/10.1175/2008JCLI2380.1

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., et al. (1996). The NCEP NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, 77(3), 377–407.

Kim, Y., Kim, K. Y., & Jhun, J. G. (2013). Seasonal evolution and the westerly jet in the Holocene East Asian summer monsoon. *Journal of Climate*, 30(7), 3343–3355. https://doi.org/10.1175/JCLI-D-16-0087.1

Krishnamurti, T. N., Stefanova, L., & Misra, V. (2013). Tropical meteorology: An introduction. New York, NY: Springer. https://doi.org/10.1007/978-1-4614-7409-8

Lau, K. M., & Yang, S. (1997). Climatological and interannual variability of the southeast asian summer monsoon. *Advances in Atmospheric Sciences*, 14(2), 141–162. https://doi.org/10.1007/s00376-997-0019-y

LinHo, L. H., Huang, X., & Lau, N. C. (2008). Winter-to-spring transition in East Asia: A planetary-scale perspective of the South China spring rain onset. *Journal of Climate*, 21(13), 3081–3096. https://doi.org/10.1175/2007JCLI1611.1

Kohonen, T. (2001). *Self-organizing maps*. Berlin, Heidelberg: Springer-Verlag.
Lu, M., Tippett, M., & Lall, U. (2015). Changes in the seasonality of tornado and favorable genesis conditions in the central United States. Geophysical Research Letters, 42(10), 4224–4231. https://doi.org/10.1002/2015GL063968

Luo, Y., Wang, H., Zhang, R., Qian, W., & Luo, Z. (2013). Comparison of rainfall characteristics and convective properties of monsoon precipitation systems over South China and the Yangtze and Huai River basin. Journal of Climate, 26(1), 110–132. https://doi.org/10.1175/JCLI-D-11-00100.1

Martin, G. M., Chevuturi, A., Comer, R. E., Dunstone, N. J., Scaife, A. A., & Zhang, D. (2019). Predictability of South China Sea Summer Monsoon onset. Advances in Atmospheric Sciences, 36(3), 253–260. https://doi.org/10.1007/s00376-018-8109-z

Ninomiya, K. (1984). Characteristics of Baiu front as a predominant subtropical front in the summer Northern Hemisphere. Journal of the Meteorological Society of Japan, 62(6), 880–894. https://doi.org/10.2151/jmsj1965.62.6_880

Nishiyama, K., Endo, S., Jinno, K., Bertacchi Uvo, C., 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(2–4), 185–200. https://doi.org/10.1016/j.atmosres.2005.10.015

Okada, Y., Takemi, T., Ishikawa, H., Kusunoki, S., & Mizuta, R. (2017). Future changes in atmospheric conditions for the seasonal evolution of the Baiu as revealed from projected AGCM experiments. Journal of the Meteorological Society of Japan, 95(4), 239–260. https://doi.org/10.2151/jmsj.2017-013

Pan, M., & Lu, M. (2020). East Asia atmospheric river catalog: Annual cycle, transition mechanism, and precipitation. Geophysical Research Letters, 47(15). https://doi.org/10.1029/2020gl089477

Rodwell, M. J., & Hoskins, B. J. (2001). Subtropical anticyclones and summer monsoons. Journal of Climate, 14(15), 3192–3211. https://doi.org/10.1175/1520-0442(2001)014<3192:SAASM>2.0.CO;2

Sampe, T., & Xie, S.-P. (2010). Large-scale dynamics of the Meiyu-Baiu rainband: Environmental forcing by the westerly jet. Journal of Climate, 23(1), 113–134. https://doi.org/10.1175/2009JCLI3128.1

Tao, X., & Chen, L. (1987). A review of recent research on the East Asian summer monsoon in China. In C.-P. Chang, & T. N. Krishnamurti (Eds.), Monsoon meteorology (pp. 60–92), New York, NY: Oxford University Press.

Tian, S.-F., & Yasunari, T. (1998). Climatological aspects and mechanism of spring persistent rains over Central China. Journal of the Meteorological Society of Japan, 76(1), 57–71. https://doi.org/10.2151/jmsj1965.76.1_57

Tian, S.-F., & Yasunari, T. (1998). Climatological aspects and mechanism of spring persistent rains over Central China. Journal of the Meteorological Society of Japan, 76(1), 57–71. https://doi.org/10.2151/jmsj1965.76.1_57

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(5), 517–537. https://doi.org/10.2151/jmsj.2011-507

Wang, B. (2006). The Asian monsoon. Berlin, Heidelberg: Springer-Verlag.

Wang, B., Wu, Z., Li, J., Liu, J., Chang, C.-P., Ding, Y., & Wu, G. (2008). How to measure the strength of the East Asian summer monsoon. Journal of Climate, 21(17), 4449–4463. https://doi.org/10.1175/2008JCLI2183.1

Wehrens, R. (2007). <Kohonen-Manual.Pdf>. Journal of Statistical Software, 21(5).

Xu, W., Zipser, E. J. (2011). Diurnal variations of precipitation, deep convection, and lightning over and east of the eastern Tibetan Plateau. Journal of Climate, 24(2), 448–465. https://doi.org/10.1175/2010JCLI3719.1

Yanai, M., & Wu, G.-X. (2006). Effects of the Tibetan Plateau. The Asian monsoon (pp. 513–549). Berlin, Heidelberg: Springer. https://doi.org/10.1007/3-540-37722-0_13