Supplementary Materials for

Modulations of synoptic weather patterns on warm-sector heavy rainfall in South China: Insights from high-density observations with principal component analysis

Wanju Li[1,2], Xueyan Bi[2*], Lifang Sheng[1], Yali Luo[3] , Jianhua Sun[4]

[1]College of Oceanic and Atmospheric Sciences, Ocean University of China, Qingdao 266100, China
[2]Institute of Tropical and Marine Meteorology, China Meteorological Administration, Guangzhou, 510080, China
[3]State Key Laboratory of Severe Weather, Chinese Academy of Meteorological Sciences, Beijing 100081 China
[4]Key Laboratory of Cloud-Precipitation Physics and Severe Storms, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China

*** Correspondence:**
Xueyan Bi
xybi@gd121.cn

Objective classification of synoptic patterns (PCT)

PCT is a dimensionality reduction method, mainly used in feature extraction, which is same as Empirical Orthogonal Function (EOF) in essence. But PCT seeks the largest variance of the principal components, and EOF seeks the smallest total error variance. PCT is mainly to find independent variables to replace the original related multivariate in order to compress data and simplify calculations. EOF is mainly for space-time decomposition, by decomposing into space function and time function to analyze the spatial structure of the actual field. The basic principle of PCT is to decompose the original high-dimensional data Z into two low-dimensional matrices F and A, $Z = FA^T$, where F is the main component and A is the load. All the main components are sorted according to the size of the characteristic value; that is, the contribution to the original data. Finally, the principal component F corresponding to the eigenvalues whose cumulative contribution to the original data Z exceeds a certain percentage is chosen to reduce the dimension (Li and Xiao, 2020).

To speed up the calculation of Principal Component Analysis (PCA), the daily mean geopotential height (GH) at the 850 hPa level are split into 10 subsets, and then the principal components (PCs) obtained from each subset are projected onto the remaining data. The PCT classification based on the Cost733class software package includes the following steps.

(1) The data of 850 hPa daily mean GH are standardized spatially. Each pattern's mean is subtracted from the data, and then the patterns are divided by their standard deviations.

(2) The standardized data are split into 10 subsets by selecting the data once every 10 days. For example, the first subset consists of the 1st, 11th, 21st, 31st, etc. days, and the second subset consists of the 2nd, 12th, 22nd, 32nd, etc. days.

(3) The principal components (PCs) are calculated using the singular value decomposition (SVD) for each subset. And the PCs of each subset are ordered according to the magnitude of their explained variances.

(4) An oblique rotation is applied to the PCs, employing an adaptation of the Gradient Projection Algorithm of Bernaards and Jennrich (2005). The main reason for using rotation is to facilitate the interpretation (Abdi and Williams, 2010). This transformation does not constrain the orthogonality, allowing for the PCs the freedom to better reflect the original data (Richman, 1981).

(5) The PC scores of each subset were projected onto the remaining data. Each circulation was assigned to a type with the highest corresponding loading.

(6) To evaluate the ten classifications based on the subsets, contingency tables are used. Subsequently CHI-square is calculated between each subset's types (every subset solution is compared with the other nine ones). The subset which gains the highest sum (of the nine comparisons) is selected, and only its types are returned.

Determining the number of synoptic patterns

The explained cluster variance (ECV) ranging from 0 to 1is selected to assess the performance of synoptic classification and to determine the number of classes (Hoffmann and Schlünzen, 2013;Philipp et al., 2014;Zong et al.,2021). ECV is defined as:

$$ECV = 1 - \frac{WS}{TS} \ , \qquad (1)$$

where *WS* is the sum of squares within synoptic patterns, and *TS* the total sum of squares:

$$WS = \sum_{j=1}^{k} \sum_{i \in C_j} D^2_{(Y_{i,} \overline{Y}_j)}, \qquad (2)$$

$$TS = \sum_{i=1}^{n} \sum_{l=1}^{m} (Y_{il} - \overline{Y}_l)^2, \qquad (3)$$

where k is synoptic patterns number, C_j is the pattern j, and the squared Euclidean distance $D^2_{(Y_{i,} \overline{Y}_j)}$ between an element and its centroid is defined as:

$$D^2_{(Y_{i,} \overline{Y}_j)} = \sum_{l=1}^{m} (Y_{il} - \overline{Y}_{jl})^2, \qquad (4)$$

where l is the time step ($l=1,2,\ldots,m$), Y_{il} is the respective data point, $\overline{Y}_{jl}$ is the estimate of the mean value for synoptic pattern j, $\overline{Y}_l$is the estimate of the total mean.

Then, the synoptic patterns number k can be determined by the increment of the ECV value (Ning et al., 2019):

$$\Delta ECV = ECV_k - ECV_{k-1}, \qquad (5)$$

The number of synoptic patterns k is finally determined when the ΔECV reaches the highest value, which suggests that the classification performance is improved substantially and tends to be stable (Ning et al., 2019).

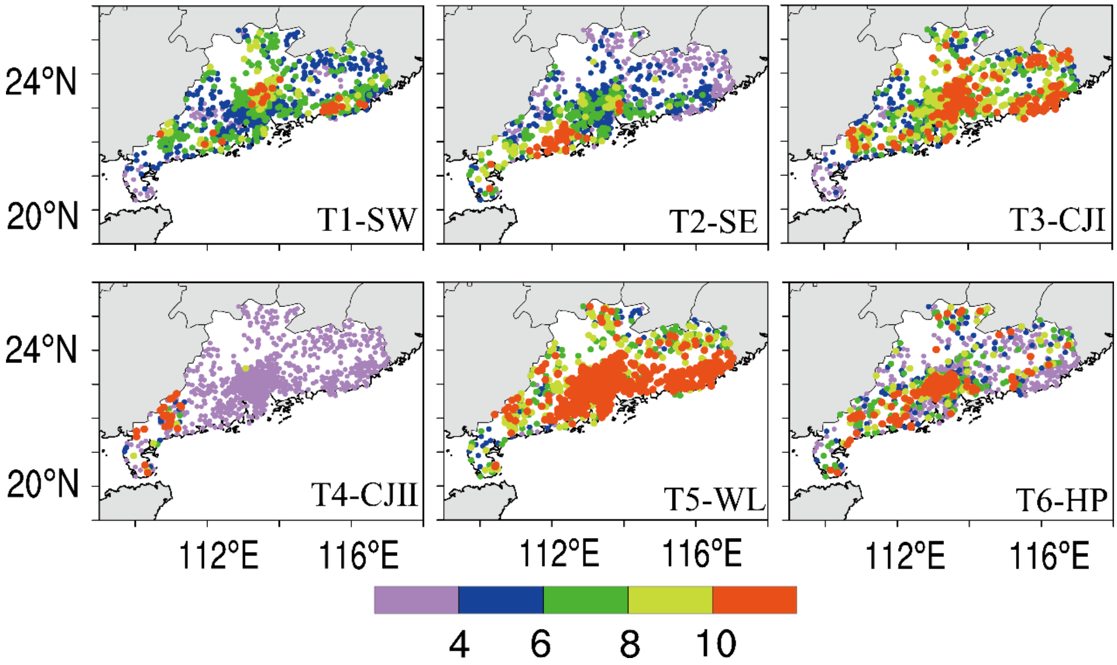

Fig.S1 Average daily precipitation of WSHR for six synoptic weather types (units: mm/d).

References

Abdi, H., and Williams, L. J.: Principal component analysis, Wiley Interdiscip. Rev. Comput. Stat., 2, 433-459, doi: 10.1002/wics.101, 2010.

Bernaards, C. A., and Jennrich, R. I.: Gradient projection algorithms and software for arbitrary rotation criteria in factor analysis, Educ. Psychol. Meas., 65, 676-696, doi: 10.1177/0013164404272507, 2005.

Hoffmann, P., and Schlünzen, K. H.: Weather pattern classification to represent the urban heat island in present and future climate, J. Appl. Meteorol. Climatol., 52, 2699-2714, doi: 10.1175/jamc-d-12-065.1, 2013.

Li, P., and Xiao, T. (2020). The diffusion and transport of PM2.5 under the polluted weather conditions during autumn and winter seasons in Chengdu. CHINA ENVIRONMENTAL SCIENCECE, 40(1), 63-75. doi: http://www.zghjkx.com.cn/CN/Y2020/V40/I1/63

Ning, G., Yim, S. H. L., Wang, S., Duan, B., Nie, C., Yang, X., Wang, J., and Shang, K.: Synergistic effects of synoptic weather patterns and topography on air quality: a case of the Sichuan Basin of China, Climate. Dyn., doi: 10.1007/s00382-019-04954-3, 2019.

Philipp, A., Beck, C., Esteban, P., Kreienkamp, F., Krennert, T., Lochbihler, K. U., Lykoudis, S. P., Pianko-Kluczynska, K., Post, P., and Alvarez10, D. R.: Cost733class-1.2 user guide, Augsburg Germany, 2014.

Zong Lian, Y. Yang, Hong Wang, Guicai Ning, Yubin Li, Zhiqiu Gao, Chao Liu, Linlin Wang, (2021) , Synoptic drivers of co-occurring surface ozone and PM2.5 pollution during summertime in eastern China, Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2020-596