ATMOSPHERIC AND OCEANIC SCIENCE LETTERS, 2011, VOL. 4, NO. 5, 281–286

The Definition and Classification of Extensive and Persistent Extreme Cold Events in China

PENG Jing-Bei[1] and BUEH Cholaw[1, 2]

[1] *The International Center for Climate and Environment Sciences, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China*

[2] *The State Key Laboratory of Numerical Modeling for Atmospheric Sciences and Geophysical Fluid Dynamics, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China*

Received 17 May 2011; revised 5 July 2011; accepted 6 July 2011; published 16 September 2011

Abstract Using the observed daily temperatures from 756 stations in China during the period from 1951 to 2009, extensive and persistent extreme cold events (EPECEs) were defined according to the following three steps: 1) a station was defined as an extreme cold station (ECS) if the observed temperature was lower than its 10th percentile threshold; 2) an extensive extreme cold event was determined to be present if the approximated area occupied by the ECSs was more than 10% of the total area of China (83rd percentile) on its starting day and the maximum area occupied by the ECSs was at least 20% of the total area of China (96th percentile); and 3) an EPECE was determined to be present if the extensive extreme cold event lasted for at least for eight days. 52 EPECEs were identified in this manner, and these identification results were also verified using other reliable data. On the basis of cluster analysis, five types of EPECEs were classified according to the spatial distribution of ECSs at their most extensive time over the course of the EPECE.

Keywords: EPECE, percentile, cluster analysis

Citation: Peng, J.-B., and C. Bueh, 2011: The definition and classification of extensive and persistent extreme cold events in China, *Atmos. Oceanic Sci. Lett.*, **4**, 281–286.

1 Introduction

Extensive and persistent low temperature events have occurred frequently over the last several years in China, resulting in devastating impacts on the social, economic, and environmental sectors of China. In January 2008, heavy freezing rains and snowstorms struck southern China, leading to a once-in-50-year (once-in-100-year in some areas) persistent extreme low temperature event (Tao and Wei, 2008). During that period, certain large-scale circulation patterns, such as blocking high over the Ural Mountains, persisted or frequently recurred (Tao and Wei, 2008). These persistent and recurrent circulation patterns allow, to some degree, for the potential to predict extensive and persistent extreme cold events (EPECEs) on an extended range time scale.

Since the 1990's, the frequencies and spatial distributions of extreme temperature and precipitation events in China and in the rest of the world have been a topic of significant interest, especially as scientists and decision makers pay increasing attention to the signs of both global warming and interdecadal climate variations (Horton et al., 2001; Klein Tank and Können, 2003; Zhai and Pan, 2003; Zhang et al., 2005; Qian and Zhang, 2007; Zhang and Wei, 2009). The disastrous event in January 2008, representative of a typical EPECE, aroused extensive scientific interest (Wen et al., 2009; Zhou et al., 2009; Bueh et al., 2011b). However, so far, EPECEs in China have not yet been systematically studied.

This study has two objectives: 1) to objectively identify EPECEs, and 2) to classify all EPECEs according to the spatial distributions of the extreme cold stations (ECS).

2 Data and methods

The dataset consists of two sets of observed temperatures: 1) the daily mean temperature (T_m) records at 756 National Standard Stations in China, as obtained from the Data and Information Center of the China Meteorological Administration, from 1951 to 2009; and 2) the homogenized daily mean temperature series at 549 National Standard Stations in China during the period from 1960 to 2008 (Li and Yan, 2009). Daily T_m anomalies were computed as the differences between the data and annual cycle defined as the 59-year mean. Cluster analysis was used to classify EPECEs.

3 Definition and validation

3.1 Definition

Two different terms have been used widely to designate extreme cold events: regional extreme cold events and site extreme cold events. The former is based on the daily surface air temperature (SAT) averaged over a region (Walsh et al., 2001), while the latter is based on SAT at each station (Yan et al., 2002; Zhang and Qian, 2011). In the literature, different lengths of time are also used to define the duration of extreme events. Zhang and Qian (2011), for example, adopted a criterion of five consecutive days as the length of time associated with their definition of a regional extreme low temperature event in China.

However, the definitions mentioned above are not appropriate for EPECEs in China, either in terms of spatial extent or duration length. As such, we need to introduce a

Corresponding author: PENG Jing-Bei, pengjingbei@mail.iap.ac.cn

new definition for EPECEs in China. For this purpose, EPECEs were defined in the following three steps. 1) A station was defined as an ECS if the observed T_m at that station was lower than its 10th percentile threshold. For a calendar day, the percentile was calculated from five-day windows centered on that calendar day for all years. Therefore, the percentile for each calendar day was determined from a total sample size of 290 days (58 years × 5 days). 2) Then, the area occupied by ECSs (S) for each day was identified. We divided the 756 stations into 1° × 1° grid boxes, as shown in Fig. 1. S equals the number of grid boxes that contain at least one ECS, as defined in Step 1. The advantage of this method is that it better represents the extent of ECS coverage because the 756 stations are nonuniformly distributed across China. This method has been used previously to determine the extent of snow cover for the Northern Hemisphere (Dewey and Heim, 1981) and the index of regional warm winters (Chen et al., 2009). An extensive extreme cold event was determined to be present if (a) S was found to constitute more than 10% of the total area of China (total number of grid boxes = 1012), the 83rd percentile of all S values; and (b) the maximum S was found to constitute at least 20% of the total area of China, the 96th percentile of all S values. 3) An EPECE was determined to be present if the extensive extreme cold event lasted for at least eight days. Each event was allowed up to two days that did not satisfy this criterion to still be classified as an event. From 1951/52 to 2008/09, eight days was about the 88th percentile of the duration of S that was larger than 10% of the total area of China.

According to this definition, a total of 52 EPECEs were identified during the period from 1951 to 2009, as displayed in Table 1.

3.2 Validation

In order to examine the robustness of the identification of EPECEs, we used the homogenized T_m at 549 stations (Li and Yan, 2009) to identify EPECEs again, following the same steps outlined in Section 3.1. A total of 43 winter EPECEs were identified during the period from 1960 to 2008 based on the homogenized data. As seen in Table 1, 43 winter EPECEs were also identified during the period from 1960 to 2008 based on the inhomogenized data. After a comparison between the two lists of EPECEs that were identified with the two datasets, 31 pairs of events (denoted by * in Table 1) were found to have identical start and end dates. As seen in Table 2, four pairs of events (denoted by ** in Table 1) have the same corresponding dates of maximum S, with the corresponding start and end dates only one day apart (see case numbers 1–4). Two pairs of events have the same corresponding start date, with the end dates being only one day apart (denoted by ** in Table 1, case numbers 5 and 7 in Table 2). One pair of events in March 1985 (denoted by ** in Table 1, case number 6 in Table 2) has the same dates of maximum S, with the start and end dates being one and six days apart, respectively. As such, 38 out of the 43 EPECEs (88.4%) can be consistently identified using these two different datasets.

However, five of the EPECEs (denoted by *** in Table 1) could be identified with the inhomogenized data but not with the homogenized data. On the other hand, an additional five EPECEs could only be identified with the homogenized data. After careful examination, it was found that the primary reason for this inconsistency was the result of the strict criteria of the maximum S and the eight-day persistence, which were difficult to satisfy simultaneously in the identification procedures used for the two different datasets. Thus, the identification of EPECEs detailed in Section 3.1 can be considered a robust and reliable method.

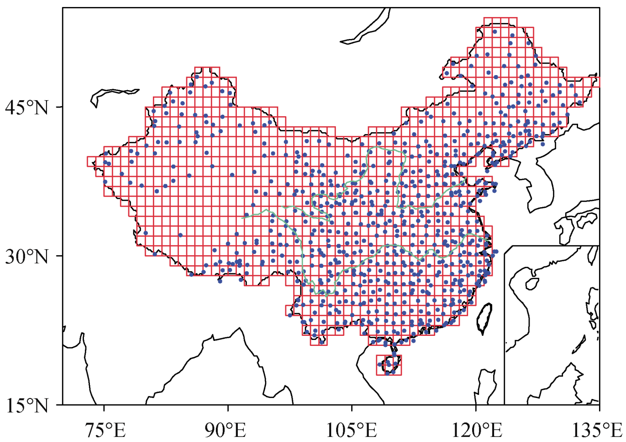

Figure 1 Distribution of 756 observation stations over China (blue dots), and 1° × 1° grid boxes (red lines).

4 Classification of EPECEs

EPECEs can be classified according to the spatial distributions of the ECSs at the time of maximum S, with each class corresponding to a special pattern of anomalous circulation. Cluster analysis is a flexible method that allows for the classification of observations objectively by comparing their similarities. All clustering algorithms require two basic criteria: to determine membership in a cluster and to determine separation between clusters. This ensures that each event belongs to one and only one cluster (Mo and Ghil, 1988). In other words, the inter-cluster distance should be maximal and intra-cluster distance should be minimal for optimal classification. Usually, when there are fewer clusters, the similarity of the observations within one cluster decreases, the maximum intra-cluster distance is larger and the minimum inter-cluster distance is smaller. Thus, when the maximum intra-cluster distance and the minimum inter-cluster distance are both stable, the partition is regarded to be optimal.

The Euclidean distance (dist) between the i-th and j-th EPECEs was defined as

$$\mathrm{dist}_{ij} = \sqrt{\sum_{m=1}^{756}(\delta_{mi} - \delta_{mj})^2}, i, j = 1, \cdots, 52, \quad (1)$$

where $\delta = \begin{cases} 1, \text{ECS} \\ 0, \text{the otherwise} \end{cases}$, and m is the serial number of the observation station.

Table 1 The dates of the beginning, maximum *S*, and end of the 52 EPECEs using temperature observations from 756 stations in China. The maximum *S*, duration, and cluster of each EPECE are also shown. Asterisks * (**) denote the 31 (7) events which were identified by both homogenized and inhomogenized data with identical (different) start and end dates. Asterisks *** denote the five events that could only be identified by the inhomogenized data and not by the homogenized data.

Number	The start and end dates of EPECE	Peak day	*S* in the peak day	Duration	Pattern
1	1–9 December 1952	3 December 1952	216	9	1
2	3–14 March 1954	5 March 1954	251	12	1
3	1–16 December 1954	9 December 1954	272	16	1
4	26 December 1954–17 January 1955	6 January 1955	274	23	1
5	7–25 December 1956	15 December 1956	233	19	1
6	5–19 Feburary 1957	11 Feburary 1957	368	15	1
7	5–16 March 1957	14 March 1957	304	12	4
8	4–12 January 1959	10 January 1959	309	9	2
9	17–26 December 1959	25 December 1959	248	10	2
10*	22 November–1 December 1960	26 November 1960	338	10	4
11*	10–17 January 1961	11 January 1961	294	8	5
12*	21–29 March 1962	22 March 1962	259	9	2
13*	20 November 1962–3 December 1962	28 November 1962	297	14	2
14*	8–27 Feburary 1964	21 Feburary 1964	345	20	1
15*	20 December 1966–17 January 1967	27 December 1966	343	29	1
16*	26 November 1967–15 December 1967	30 November 1967	390	20	1
17*	30 January 1968–22 Feburary 1968	7 Feburary 1968	354	24	1
18*	27 January 1969–7 Feburary1969	4 Feburary 1969	381	12	1
19*	13 Feburary 1969–4 March 1969	21 Feburary 1969	354	20	4
20**	25 Feburary 1970–25 March 1970	17 March 1970	353	29	1
21*	21–30 November 1970	29 November 1970	235	10	5
22***	27 January 1971–7 Feburary 1971	29 January 1971	209	12	2
23*	27 Feburary 1971–14 March 1971	7 March 1971	265	16	4
24*	3–11 Feburary1972	8 Feburary 1972	319	9	3
25*	20–27 March 1974	26 March 1974	244	8	2
26**	3–21 December 1974	14 December 1974	256	19	2
27*	7–23 December 1975	12 December 1975	399	17	1
28*	17–24 March 1976	19 March 1976	377	8	1
29**	10–27 November 1976	14 November 1976	412	18	1
30*	25 December 1976–15 January 1977	28 December 1976	415	22	1
31**	26 January 1977–10 Feburary 1977	30 January 1977	429	16	1
32*	9–18 Feburary 1978	15 Feburary 1978	337	10	1
33*	10–29 November 1979	18 November 1979	410	20	1
34**	29 January 1980–9 Feburary 1980	5 Feburary 1980	408	12	1
35*	1–10 November 1981	7 November 1981	463	10	1
36*	19 January 1984–10 Feburary 1984	6 Feburary 1984	263	23	5
37*	16–30 December 1984	24 December 1984	402	15	1
38*	16–24 Feburary 1985	18 Feburary 1985	219	9	4
39**	4–21 March 1985	9 March 1985	357	18	4
40*	6–17 December 1985	11 December 1985	321	12	4
41*	26 November 1987–7 December 1987	30 November 1987	424	12	1
42*	27 Feburary1988–8 March 1988	3 March 1988	344	10	1
43***	26 December 1991–2 January 1992	28 December 1991	351	8	3
44***	16–27 March 1992	17 March 1992	213	12	3
45*	14–24 January 1993	16 January 1993	343	11	3
46*	17–24 November 1993	21 November 1993	364	8	4
47*	14–22 December 1993	15 December 1993	218	9	3
48***	9–16 March 1994	14 March 1994	256	8	2
49*	17–24 Feburary 1996	20 Feburary 1996	298	8	3
50***	18–25 December 1999	21 December 1999	294	8	3
51**	24 January 2000–2 Feburary 2000	31 January 2000	276	10	3
52	14 January 2008–15 Feburary 2008	1 Feburary 2008	306	33	2

Table 2 The dates of the beginning, maximum *S*, and end of the seven EPECEs identified by both the homogenized and inhomogenized data, but with different start and end dates. The numbers in brackets in columns 2–4 are the non-zero differences between the dates of the beginning, maximum *S*, and end between the EPECEs, as determined by the two sets of data.

Number	Via inhomogenized data			Via homogenized data		
	Beginning	Maximum *S*	End	Beginning	Maximum *S*	End
1	25 Feburary 1970 (1)	17 March 1970	25 March 1970	26 Feburary 1970	17 March 1970	25 March 1970
2	3 December 1974	14 December 1974	21 December 1974 (1)	3 December 1974	14 December 1974	20 December 1974
3	10 November 1976	14 November 1976	27 November 1976 (1)	10 November 1976	14 November 1976	26 November 1976
4	26 January 1977 (1)	30 January 1977	10 Feburary 1977	27 January 1977	30 January 1977	10 Feburary 1977
5	29 January 1980	5 Feburary 1980 (5)	9 Feburary 1980 (1)	29 January 1980	10 Feburary 1980	10 Feburary 1980
6	4 March 1985 (1)	9 March 1985	21 March 1985 (6)	3 March 1985	9 March 1985	15 March 1985
7	24 January 2000	31 January 2000 (1)	2 Feburary 2000 (1)	24 January 2000	1 Feburary 2000	1 Feburary 2000

We used hierarchical cluster analysis (or minimum distance cluster analysis, Shi, 2002) to identify the seed points for the *K* cluster means, and used *K*-means (centroid) cluster analysis (Hartigan and Wong, 1979) to assign each EPECE. The maximum intra-cluster distances and minimum inter-cluster distances were calculated, and the cluster number varied from 25 to 2 (Fig. 2). The minimum inter-cluster distance remained unvaried as the cluster number varied from 7 to 3, and the maximum intra-cluster distance also remained unvaried as the cluster number varied between 5 and 4. Therefore, we considered it appropriate to classify 52 EPECEs into five clusters after additional visual inspection of all EPECEs (see Table 1).

The consistency in the distribution of the ECSs among each EPECE for a cluster was measured by f

$$f = \frac{\sum_{i=1}^{k} \delta_i}{k}, \quad (2)$$

where i denotes the serial number of a member in a cluster, and k is the total number of members in the cluster. Clearly, the distribution of f serves as the average distribution of the ECSs for a cluster (Fig. 3, right column), with $f = 1.0$ indicating that the station is the ECS for each member.

Cluster 1 is the largest of the five clusters, with 24 members, or approximately 46.2% of the total EPECEs. A representative example is shown in Fig. 3 (left top row). It is characterized by the nationwide distribution of ECSs at the time of maximum *S*. Extreme low temperatures dominated over most of China, except for northeastern China and the Tibetan Plateau. The degree of similarity in the extreme low temperature distribution among the 24 members of this cluster was high, which can be seen from the map of consistency in the distribution of ECSs (right top row).

Cluster 2 is the second largest cluster (nine members) and its ECSs were distributed mainly over northwestern China and to the south of the Yangtze River basin (Fig. 3, second row). The well-known EPECE that occurred in January 2008 belongs only to this cluster. In this cluster, the surface air temperature over northeastern China was apparently higher than normal.

Unlike Clusters 1 and 2, Cluster 3 is characterized by extreme low temperatures over the regions to the south of

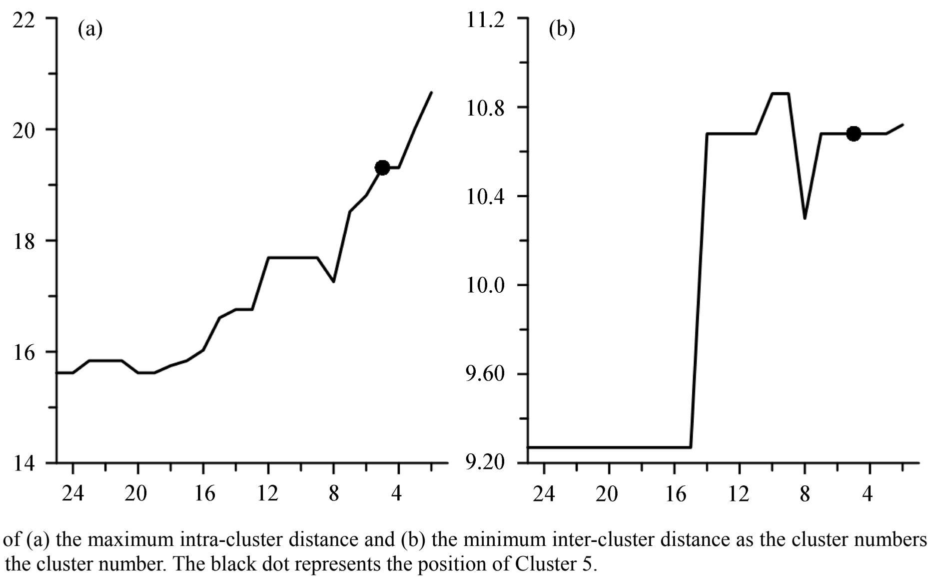

Figure 2 Variation of (a) the maximum intra-cluster distance and (b) the minimum inter-cluster distance as the cluster numbers varied from 25 to 2. The abscissa denotes the cluster number. The black dot represents the position of Cluster 5.

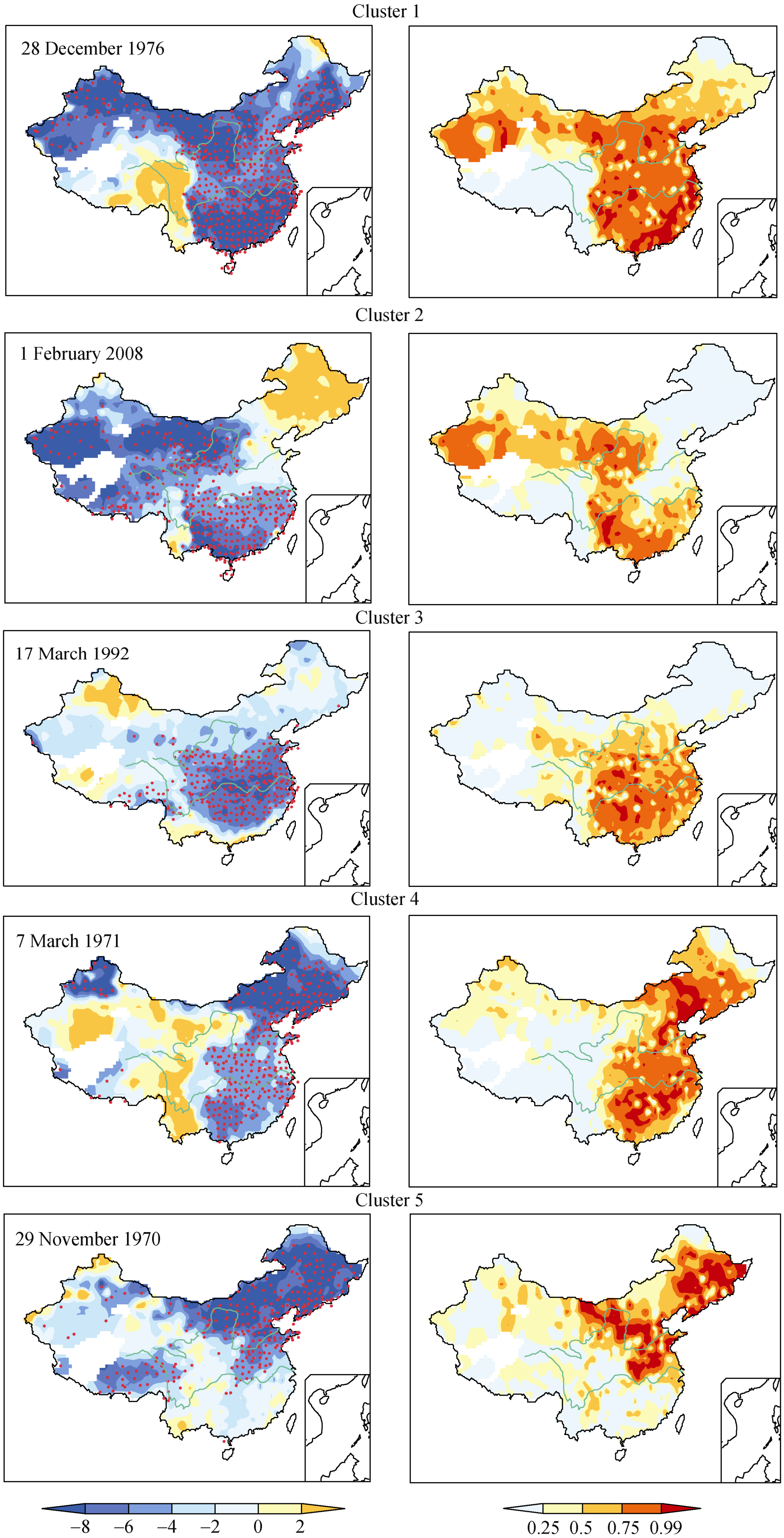

Figure 3 The distribution of T_m anomalies of the typical case (left column, Units: °C) and the consistency in distribution of ECSs f (right column) of the five clusters of EPECEs at the date of maximum S (shaded as in the legend, blank for missing data). The red dots in the left column represent the ECSs.

the Yellow River and to the east of the Tibetan Plateau (Fig. 3, third row). This cluster is the third largest and includes eight EPECEs.

Cluster 4 also has eight members, and its ECSs were distributed primarily over the eastern part of China, extending from northeastern China to southeastern China (Fig. 3, fourth row).

Cluster 5 is the smallest cluster (Fig. 3, bottom row), with only three members. The extreme low temperatures associated with this cluster occurred only over northeastern and northern China. Notably, the distributions of ECSs in Clusters 2 and 5 seem nearly opposite.

We speculate that the maximum *S*, which occupies at least 20% of the total area of China, was a tough criterion to meet for the EPECEs in Cluster 5, as the combined area of northeastern and northern China is finite; this is likely why few EPECEs were clustered into Cluster 5.

5 Discussion

Based on the daily mean temperatures at the 756 stations in China during the period from 1951 to 2009, wintertime EPECEs were defined and identified. In addition, all EPECEs were classified into five clusters according to the distribution of ECSs at the time of their maximum *S*.

Importantly, EPECEs are distinct from the cold waves that frequently affect East Asia (Ding and Krishnamurti, 1987; Takaya and Nakamura, 2005) in the following two aspects: 1) the extreme low temperature in the former is more persistent than the latter, and 2) the spatial extent of the ECSs in the definition of the former has a strict criterion, whereas it is not so strictly defined for the latter. Despite these distinctions, it was found that most of the EPECEs were recognized as cold waves at their start dates. Therefore, it is necessary to clarify the relationship between EPECEs and cold waves.

Though EPECEs have been identified in the present study, the typical circulation features associated with EPECEs and the key factors are still not clear. In fact, Ural blocking, the tilting of the East Asian trough axis, and the strength of stratospheric polar vortex have been recognized as causes of extreme cold events over China (Wang et al., 2009; Bueh et al., 2011a; Wei et al., 2011). Therefore, these factors should be studied in future investigations.

Acknowledgements. We are indebted to Profs. Ji Liren, Chen Lieting, and Dr. Wei Jie for giving us constructive advices on the definition and classification of EPECEs. This work was jointly supported by the National Key Technologies R&D Program of China (Grant No. 2009BAC51B02) and the Special Funds for Meteorology Scientific Research on Public Cause (Grant No. GYHY201106015).

References

Bueh, C., X. Y. Fu, and Z. W. Xie, 2011a: Large-scale circulation features typical of wintertime extensive and persistent low temperature events in China, *Atmos. Oceanic Sci. Lett.*, **4**, 235–241.

Bueh, C., N. Shi, and Z. W. Xie, 2011b: Large-scale circulation anomalies associated with persistent low temperature over southern China in January 2008, *Atmos. Sci. Lett.*, **12**, 273–280.

Chen, Y., G. Y. Ren, L. Wang, et al., 2009: Temporal change of warm winter events over the last 56 years in China, *J. Appl. Meteor. Sci.* (in Chinese), **20**, 539–545.

Dewey, K. F., and J. R. Heim, 1981: *Satellite Observations of Variations in Northern Hemisphere Seasonal Snow Cover*, NOAA Technical Report NESS 87, Washington D. C., 83pp.

Ding, Y., and T. N. Krishnamurti, 1987: Heat budget of the Siberian High and the winter monsoon, *Mon. Wea. Rev.*, **115**, 2428–2449.

Hartigan, J. A., and M. A. Wong, 1979: Algorithm AS 136: A *K*-means clustering algorithm, *Appl. Stat.*, **28**, 100–108.

Horton, E. B., C. K. Folland, and D. E. Parker, 2001: The changing incidence of extremes in worldwide and Central England temperatures to the end of the twentieth century, *Climatic Change*, **50**, 267–295.

Klein Tank, A. M. G., and G. P. Können, 2003: Trends in indices of daily temperature and precipitation extremes in Europe, 1946–99, *J. Climate*, **16**, 3665–3680.

Li, Z., and Z. Yan, 2009: Homogenized China daily mean/maximum/minimum temperature series 1960–2008, *Atmos. Oceanic Sci. Lett.*, **2**, 237–243.

Mo, K. C., and M. Ghil, 1988: Cluster analysis of multiple planetary flow regimes, *J. Geophys. Res.*, **930**, 10927–10952.

Qian, W. H., and W. W. Zhang, 2007: Changes in cold wave events and warm winter in China during the last 46 years, *Chinese J. Atmos. Sci.* (in Chinese), **31**, 1266–1278..

Shi, N., 2002: *The Multivariate Analysis Method for Meteorological Research and Forecast* (in Chinese), 2nd ed, China Meteorological Press, Beijing, 224–225.

Takaya, K., and H. Nakamura, 2005: Mechanisms of intraseasonal amplification of the cold Siberian High, *J. Atmos. Sci.*, **62**, 4423–4440.

Tao, S. Y., and J. Wei, 2008: Severe snow and freezing-rain in January 2008 in the southern China, *Climatic Environ. Res.* (in Chinese), **13**, 337–350.

Walsh, J. E., A. S. Phillips, D. H. Portis, et al., 2001: Extreme cold outbreaks in the United States and Europe, 1948–99, *J. Climate*, **14**, 2642–2658.

Wang, L., W. Chen, W. Zhou, et al., 2009: Interannual variations of East Asian trough axis at 500 hPa and its association with the East Asian winter monsoon pathway, *J. Climate*, **22**, 600–614.

Wei, K., W. Chen, and W. Zhou, 2011: Changes in the East Asian cold season since 2000, *Adv. Atmos. Sci.*, **28**, 69–79.

Wen, M., S. Yang, A. Kumar, et al., 2009: An analysis of the large-scale climate anomalies associated with the snowstorms affecting China in January 2008, *Mon. Wea. Rev.*, **137**, 1111–1131.

Yan, Z., P. D. Jones, T. D. Davies, et al., 2002: Trends of extreme temperatures in Europe and China based on daily observations, *Climatic Change*, **53**, 355–392.

Zhai, P. M., and X. H. Pan, 2003: Change in extreme temperature and precipitation over northern China during the second half of the 20th century, *Acta Geog. Sinica* (in Chinese), **58**, 1–10.

Zhang, T., and F. Y. Wei, 2009: Probability distribution of precipitation extremes during raining seasons in South China, *Acta Meteor. Sinica*, **67**, 442–451.

Zhang, X. B., E. Aguilar, S. Sensoy, et al., 2005: Trends in Middle East climate extreme indices from 1950 to 2003, *J. Geophys. Res.*, **110**, D22104, doi:10.1029/2005JD006181.

Zhang, Z. J., and W. H. Qian, 2011: Identifying regional prolonged low temperature events in China, *Adv. Atmos. Sci.*, **28**, 338–351.

Zhou, W., J. C. L. Chan, W. Chen, et al., 2009: Synoptic-scale controls of persistent low temperature and icy weather over southern China in January 2008, *Mon. Wea. Rev.*, **137**, 3978–3991.