Evaluation of the Twentieth Century Reanalysis Dataset in Describing East Asian Winter Monsoon Variability

ZHANG Ziyin[*1,2] (张自银), GUO Wenli[1] (郭文利), GONG Daoyi[2] (龚道溢), and Seong-Joong KIM[3]

[1]*Beijing Meteorological Bureau, Beijing* 100089

[2]*State Key Laboratory of Earth Surface Processes and Resource Ecology, Beijing Normal University, Beijing* 100875

[3]*Korea Polar Research Institute, Incheon* 406–840, *Korea*

ABSTRACT

The Twentieth Century Reanalysis (20thCR) dataset released in 2010 covers the period 1871–2010 and is one of the longest reanalysis datasets available worldwide. Using ERA-40, ERA-Interim and NCEP–NCAR reanalysis data, as well as HadSLP2 data and meteorological temperature records over eastern China, the performances of 20thCR in reproducing the spatial patterns and temporal variability of the East Asian winter monsoon (EAWM) are examined. Results indicate that 20thCR data: (1) can accurately reproduce the most typical configuration patterns of all sub-factors involved in the EAWM system, albeit with some differences in the main circulation fields over East Asia in comparison to ERA-40 reanalysis data; (2) is reliable and stable in describing the temporal variability of EAWM since the 1930s; and (3) can describe the high-frequency variability of EAWM better than the low-frequency fluctuations, especially in the early period. In conclusion, caution should be taken when using 20thCR data to study interdecadal variabilities or long-term trends of the EAWM, especially prior to the 1930s.

Key words: Twentieth Century Reanalysis dataset, East Asian winter monsoon, Siberian High, eastern China

1. Introduction

The East Asian winter monsoon (EAWM) is one of the most active systems in the Northern Hemisphere during boreal winter (December–January–February, DJF) (Chang and Lau, 1982; Ding and Krishnamurti, 1987). A strong (weak) EAWM usually leads to a severe (warm) winter in China and adjacent regions. There have been many efforts to understand and predict the influences of the EAWM on the winter or even summer climate over China and adjacent regions (Zhang et al., 1997; Compo et al., 1999; Chen et al., 2000; Jhun and Lee, 2004; Wang et al., 2009a; Wang and Chen, 2010; Zhou, 2011). The EAWM not only plays an essential role in controlling the winter climate over most of China, Mongolia, East Russia, the Korean Peninsula and most of Japan, but it also influences the current and subsequent ocean–atmosphere systems and modulates the global climate change to a certain extent (Lau and Li, 1984; Guo, 1994; Shi, 1996; Ji et al., 1997; Li and Mu, 2000; Huang et al., 2004; Huang et al., 2012).

Owing to a lack of widespread historical instrumental records, our understanding of the natural variability and evolution of the EAWM system and its components is limited. Over the past few decades, the European Centre for Medium Range Weather Forecasts (ERA-40) reanalysis data (Uppala et al., 2005) and the National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) reanalysis data (Kalnay et al., 1996) have played important roles in investigating the EAWM and its relationship to climate change in many studies. However, these two most popular reanalysis datasets

---

*Corresponding author: ZHANG Ziyin, zzy_ahgeo@163.com

only began during the mid-twentieth century, leaving many important climate events uncovered, and also limiting awareness of the variability of the EAWM and its circulation factors from a long-term perspective. The Twentieth Century Reanalysis (20thCR) Project is an effort led by the Physical Science Division of the Earth System Research Laboratory and the University of Colorado Cooperative Institute for Research in Environmental Sciences (CIRES) Climate Diagnostics Center to produce a reanalysis dataset extending back to the 1870s (Compo et al., 2011). However, it is unclear whether 20thCR data are reliable in describing the variability of EAWM since the 1870s. The purpose of this study, therefore, is to examine the ability of 20thCR reanalysis data in describing the spatial structure and temporal variability of the EAWM.

2.   Data and methods

The monthly sea level pressure (SLP), surface air temperature (T_s), zonal and meridional wind at 850 hPa (U850, V850), geopotential height field at 500 hPa (H500) and zonal wind at 200 hPa (U200) from four reanalysis datasets were used in the study, including 20thCR data covering the period 1871–2010 (Compo et al., 2011), NCEP–NCAR data since 1948 (Kalnay et al., 1996), ERA-40 data for the period 1957–2002 (Uppala et al., 2005), and ERA–Interim data since 1979 (Dee et al., 2011). To facilitate calculations and comparisons, the 20thCR and ERA-Interim reanalysis data were interpolated to share the same spatial resolution as the NCEP–NCAR and ERA-40 reanalysis datasets; namely, $2.5° \times 2.5°$. Moreover, a long-term Siberian High index derived from the Hadley Centre HadSLP2 dataset (Allan and Ansell, 2006) was also used. To examine the relationship and stability of the EAWM and the winter climate over eastern China, a seasonal temperature series from 71 meteorological stations located in eastern China (T71) covering the period 1880–2007 was also used (Wang et al., 2009b).

For the statistical and atmospheric circulation analyses carried out in the study, common methods such as correlation and regression analysis were applied. Moreover, a multivariate EOF analysis (MV–EOF) (Wang et al., 2008; Zhang et al., 2012) was used to examine the collaborative relations among all the sub-factors of the EAWM system. The time series corresponding to the first eigenvector from the MV–EOF analysis was then defined as a comprehensive index describing the variability of the entire EAWM system. To reduce the effects of low-frequency variation and examine whether or not the correspondence between the two time series is stable, the high-frequency correlat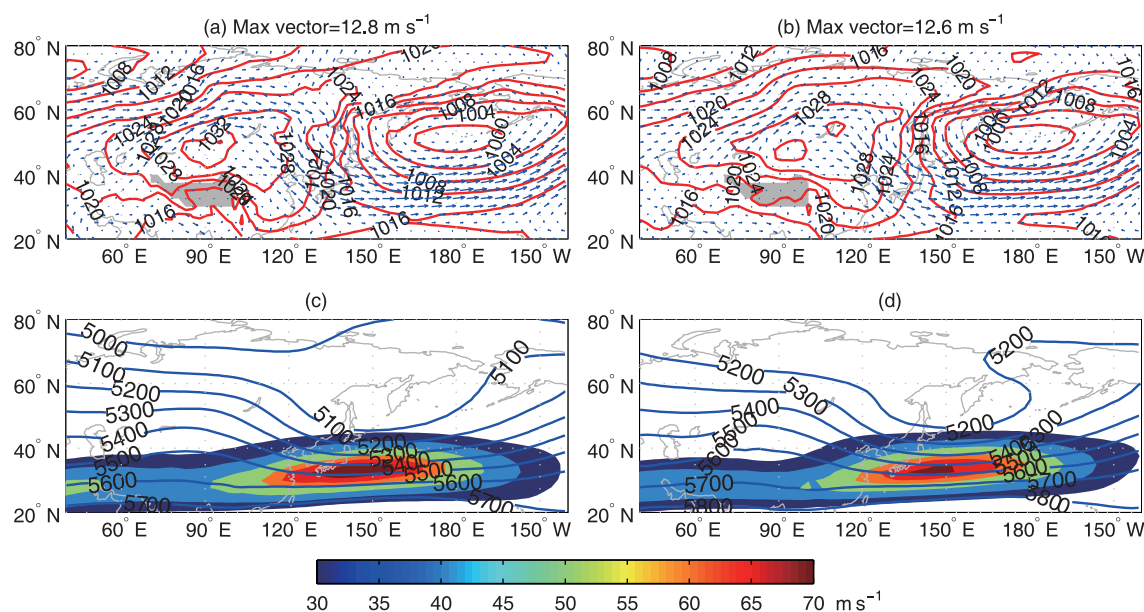ion of the high-pass filtered time series was also tested. Throughout this article, all time series were normalized to a base period of 1971–2000.

3.   Results

3.1   *Spatial features of the EAWM in* 20*thCR*

We examine whether 20thCR data are able to correctly capture the main circulation features of the EAWM, such as the Siberian High, low-level wind, the East Asian Trough and the Jet Stream. The Siberian High is an important part of the EAWM and exerts an obvious influence on the winter climate over most of East Asia; the climatological pattern of low-level wind corresponding to the Siberian High and the Aleutian Low is well known. Based on the ERA-40 reanalysis data, the winter climatological mean fields of SLP (contour lines) and low-level wind at 850 hPa (vectors) over East Asia are shown in Fig. 1a, which are consistent with previous understanding. Meanwhile, the climatological mean fields of SLP and low-level wind at 850 hPa derived from the 20thCR reanalysis data are shown in Fig. 1b. Intuitively, the spatial pattern of the SLP field and the low-level wind field described in 20thCR are consistent with ERA-40. The spatial correlation coefficient of the SLP field between ERA-40 and 20thCR is 0.977. All of the spatial correlation coefficients are significant at the 0.001 level throughout this article. The spatial correlation coefficient of the low-level wind field between ERA-40 and 20thCR is 0.939. Furthermore, the spatial correlations of U850 and V850 are 0.969 and 0.918, respectively, indicating that 20thCR can describe zonal wind better than meridional wind in the low-level wind field over East Asia and adjacent regions in winter.

Comparing H500 (contour lines in Figs. 1c and d) and U200 (colors in Figs. 1c and d) between ERA-40 and 20thCR over East Asia, the spatial patterns of 20thCR are highly consistent with ERA-40. The correlation coefficients of H500 and U200 between ERA-40 and 20thCR are 0.999 and 0.992, respectively. However, the H500 field of 20thCR is higher than that of ERA-40 (approximately 100 gpm), which indicates that the 20thCR dataset systematically overestimates the geopotential height in the middle troposphere in winter. Moreover, the East Asian Trough described in 20thCR is slightly deeper than that of ERA-40. In U200, the position of the westerly Jet Stream axis and the spatial shape are consistent, with an axis running from the Middle East to the Northwest Pacific with a high value center located from the East China Sea to the Northwest Pacific area. Similarly, there are some subtle differences between them; namely, the intensity of the central area of the East Asian Jet Stream in 20thCR is slightly weaker than that in ERA-40. Gen-

Fig. 1. Climatological mean fields of SLP (contour lines) and low–level wind at 850 hPa (vectors) based on (a) ERA-40 and (b) 20thCR data averaged from 1971 to 2000. Climatological mean fields of geopotential height at 500 hPa (contour lines) and zonal wind at 200 hPa (colors) based on (c) ERA-40 and (d) 20thCR data averaged over the same period. The zonal wind speed less than 30 m s^{-1} is omitted for clarity in (c) and (d).

erally, we conclude that the 20thCR data describes the spatial features of the EAWM system well, although there are some differences in the main circulations over East Asia when referenced to ERA-40 reanalysis data.
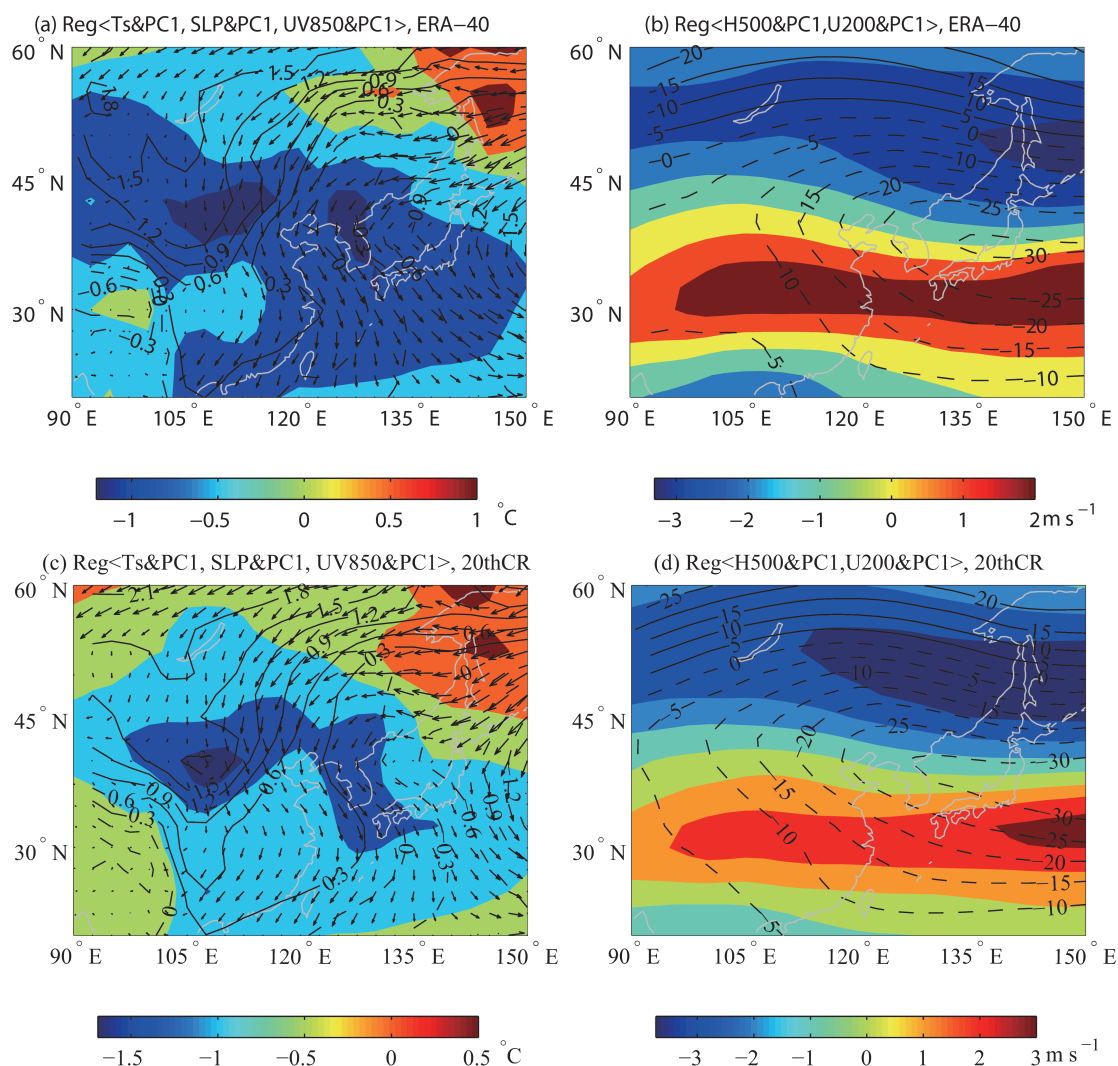
3.2 *Temporal features of the EAWM in* 20 *thCR*

It is important to know whether 20thCR data can capture the temporal variability of the EAWM because it is more complicated to reproduce the year-to-year changes in the strength of the EAWM than to describe a climatological spatial status. To reflect the integrated feature of the EAWM system with a unique index, a MV–EOF method was used to extract the dominant mode, synthesized from six variables, including SLP, T_s, U850, V850, H500 and U200 over East Asia (20°–60°N, 90°–150°E). The time series corresponding to the first eigenvector of the MV–EOF analysis (MV–PC1) was taken as the EAWM index, conferring an advantage over the winter climate in East China (Zhang et al., 2012). In addition to the 20thCR data, ERA-40, NCEP–NCAR and ERA-Interim reanalysis data were also used in the MV–EOF analysis for comparison. The explained variances of 20thCR are 19.7% for the analysis period of 1872–2010 and 25.6% for the period of 1949–2010. For ERA-40, the explained variance is 25.0% for the period 1958–2002, while for ERA-Interim it is 28.4% for the period 1980–2011. The high values of explained variance indicate that the first

eigenvector can capture the dominant spatiotemporal features of the EAWM system.

The distribution of DJF T_s (colors in Figs. 2a and c, °C), SLP (contour lines in Figs. 2a and c, hPa), winds at 850 hPa (vectors in Figs. 2a and c, m s^{-1}), H500 (contour lines in Figs. 2b and d, m) and U200 (colors in Figs. 2b and d, m s^{-1}) anomalies corresponding to a one-standard-deviation positive DJF EAWM index (MV–PC1) are plotted based on ERA-40 (upper) and 20thCR (lower) data from 1958 to 2002 in Fig. 2. The spatial patterns for all the variables in 20thCR (Figs. 2c and d) are highly consistent with ERA-40 (Figs. 2a and b). The spatial correlation coefficients of the regression fields of T_s, SLP, U850, V850, H500 and U200 between 20thCR and ERA-40 are 0.89, 0.99, 0.95, 0.84, 0.99 and 0.98, respectively. The high spatial correlations suggest that the 20thCR reanalysis data accurately captures the most typical features of the integrity of the EAWM system, which reflects the configuration patterns of all of the sub-factors involved in the EAWM system.

As previously mentioned, the time series corresponding to the first eigenvector of the MV–EOF analysis was taken as the EAWM indices based on 20thCR, NCEP–NCAR, ERA-40 and ERA-Interim, and they are shown in Fig. 3. The four EAWM indices are overall consistent with each other in both interannual variability and interdecadal fluctuations in the common period, with several strong anomalous events in
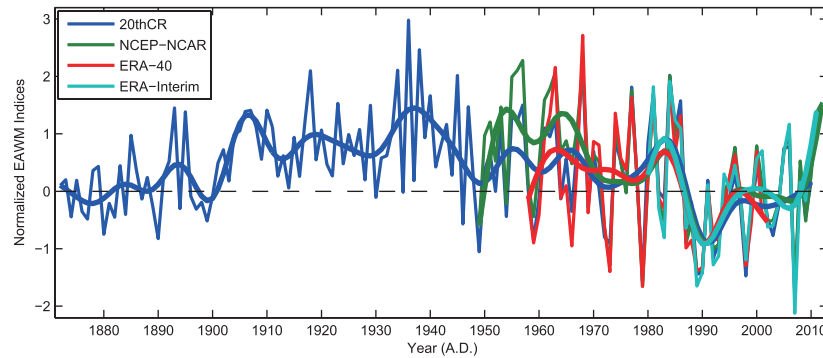
Fig. 2. Spatial distribution of the regression coefficients of (a) T_s, SLP and UV850 and (b) H500 and U200 in DJF upon the EAWM index (MV–PC1) based on ERA-40. The panels (c) and (d) are the same as (a) and (b), but for the 20thCR dataset. In (a) and (c), color shading denotes surface air temperatures ($^\circ$C), the isolines denote sea level pressure (hPa) and the vectors denote winds at 850 hPa (m s^{-1}). The maximum vectors in (a) and (c) are 1.40 and 1.44 m s^{-1}, respectively. In (b) and (d), color shading denotes zonal winds at 200 hPa (m s^{-1}) and the isolines indicate the geopotential height at 500 hPa (m). Data periods are 1958–2002. All of the values in this figure were scaled to correspond to one-standard-deviation of the EAWM index.

1967/68, 1976/77, 1980/81 and 1983/84, and several weak anomalous events in 1965/66, 1972/73, 1978/79, 1988/89 and 1997/98. On the interdecadal timescale (see the smoothed lines in Fig. 3), a highly positive phase dominated during the mid-1950s, mid-1960s and the early 1980s, and an obvious trough dominated in the 1970s and 1990s. The correlation coefficients between the four EAWM indices in each common period are shown in Table 1. The raw correlation coefficients between 20thCR and NCEP–NCAR, ERA-40 and ERA–Interim are 0.94, 0.94 and 0.95, respectively;

the high-frequency (<10 yr) correlations are 0.94, 0.94 and 0.96, respectively—all significant at the 0.01 level. The high correlations indicate that 20thCR is highly reliable and stable in capturing the temporal variability of the EAWM since the 1950s. However, how well do the 20thCR data describe the variability of the EAWM before the 1950s?

To examine the stability and reliability of 20thCR data in describing the variation in the EAWM before the 1950s, an additional analysis was performed based on the Siberian High index, which was derived from

Fig. 3. Four normalized EAWM indices derived from the 20thCR, NCEP–NCAR, ERA-40 and ERA-Interim reanalysis datasets based on the MV–EOF method. The smooth lines are the low–frequency (>10 yr) variations from a Butterworth filter for each curve.

Table 1. Correlation coefficients between each pair of EAWM indices.

		NCEP–NCAR	ERA-40	ERA–Interim
20thCR	r_1	0.94	0.94	0.95
	r_2	0.94	0.94	0.96
NCEP–NCAR	r_1	1	0.93	0.97
	r_2	1	0.94	0.95
ERA-40	r_1		1	0.92
	r_2		1	0.88

Note: All correlation coefficients are significant at the 0.01 level (two–tailed); r_1 and r_2 indicate raw correlation and high-frequency (<10 yr) correlation, respectively.
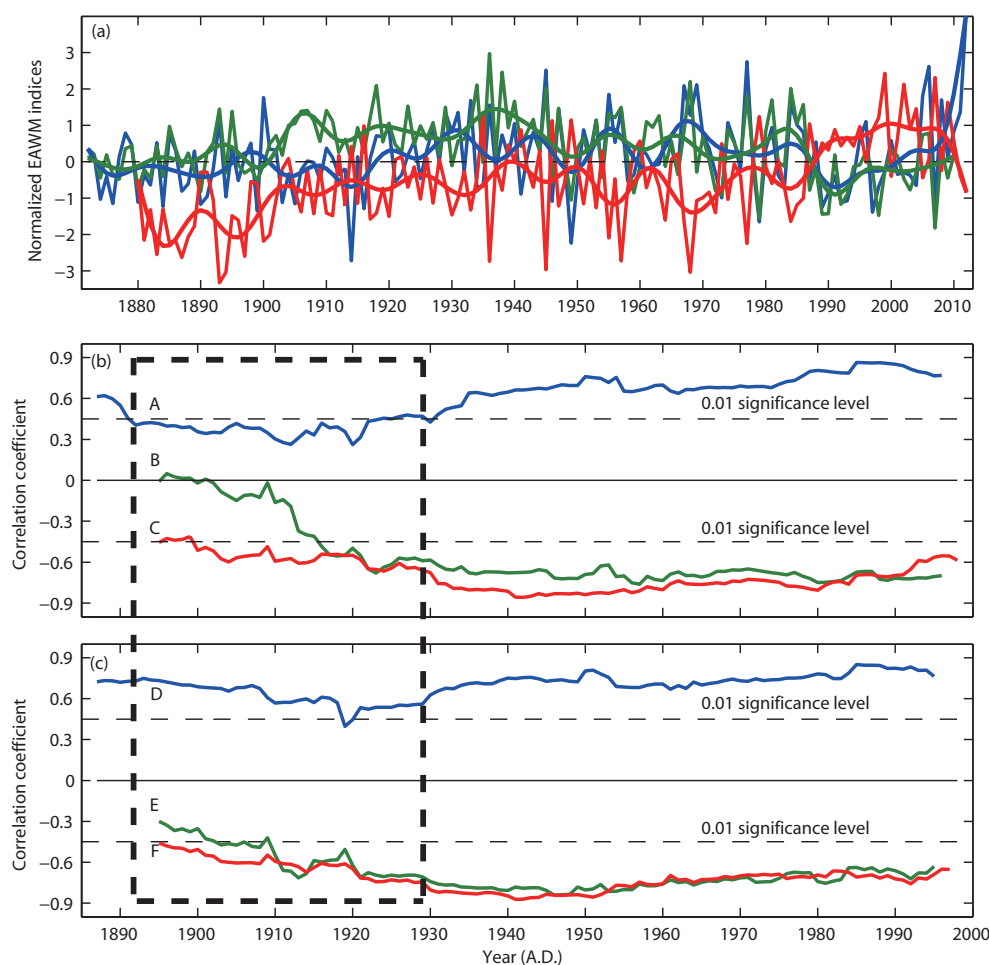
the HadSLP2 dataset and the temperature series from the 71-station records over eastern China from the 1870s and 1880s, respectively. The normalized time series of the EAWM index (green line), the Siberian High index (blue line) and the East China temperature (red line) are shown in Fig. 4a. The variability of EAWM generally coincides well with the Siberian High and is out-of-phase to East China temperature. The raw (r_1) and high-frequency (r_2) correlation coefficients between the EAWM and Siberian High are 0.60 and

0.72, respectively—significant at the 0.01 level (Table 2). This level of correlation indicates that they share stability and a high level of consistency from year to year. The Siberian High is one of the most important circulation components of the EAWM system, and it could be used as an indicator of the intensity of the EAWM system (Gong and Ho, 2002). A strong (weak) Siberian High often means strong (weak) EAWM activity and a cold (warm) winter in most parts of China and its surrounding areas. As a result, the variability

Table 2. Correlation coefficients among the three time series in different periods.

		SBH		T71	
		r_1	r_2	r_1	r_2
EAWM–20thCR	Entire period	0.60**	0.72**	−0.52**	−0.66**
	Post-1951	0.74**	0.73**	−0.72**	−0.70**
	Pre-1951	0.58**	0.70**	−0.23	−0.63**
SBH	Entire period	1	1	−0.45**	−0.69**
	Post-1951	1	1	−0.62**	−0.68**
	Pre-1951	1	1	−0.47**	−0.69**

Note: ** Significant at the 0.01 level; SBH denotes Siberian High index derived from HadSLP2 dataset; r_1 and r_2 indicate raw and high-frequency (<10 yr) correlations, respectively.

Fig. 4. (a) EAWM index (green) from 20thCR, Siberian High index (blue) from HadSLP2, and the temperature series (red) from East China. Panels (b) and (c) respectively show the raw and high-frequency (<10 yr) 30-yr running correlation coefficients between the EAWM and Siberian High (A, D), the EAWM and East China temperature (B, E), and the Siberian High and East China temperature (C, F).

of the Siberian High is consistent with the EAWM fluctuation. Consequently, the raw correlation coefficients between East China temperature and the EAWM and Siberian High are -0.52 and -0.45, respectively, and the high-frequency correlation coefficients are -0.66 and -0.69, respectively—all significant at the 0.01 level. The raw (high-frequency) correlation coefficients of the EAWM and the Siberian High during the periods of pre-1951 and post-1951 are 0.58 (0.70) and 0.74 (0.73), respectively. The raw (high-frequency) correlation coefficients of the EAWM and East China temperature during the periods of pre-1951 and post-1951 are -0.23 (-0.63) and 0.72 (0.70), respectively—all significant at the 0.01 level, except for -0.23. Generally, the correlation coefficients for the first period (pre-1951) are much lower than the correlation coefficients for the second period (post-1951), which indicates that the corresponding relationship between the

EAWM and the Siberian High and East China temperature became weaker earlier in history.

The 30-year running correlation coefficients between the EAWM and Siberian High (curve A), the EAWM and East China temperature (curve B), and the Siberian High and East China temperature (curve C) are shown in Fig. 4b. It can be seen that the correlation between the EAWM and Siberian High increased gradually with time, especially after the 1930s. Curves B and C are also characterized by a similar increase in correlation coefficients of East China temperature and the EAWM and Siberian High. It should be noted that, although both curves B and C decreased as they extended back into history prior to the 1930s, curve B dropped dramatically, especially prior to the mid-1920s, contrasting obviously with a gentle reduction of curve C. Furthermore, the high-frequency (<10 yr) correlation coefficients corresponding to curves A,

B and C are also plotted in Fig. 4c; namely, curves D, E and F. The most obvious difference between the raw and high-frequency correlation coefficient curves is the improvement in the correlation coefficients of the EAWM and Siberian High and East China temperature during the period before the 1930s, which suggests that the 20thCR dataset reproduces the interannual variability of EAWM better, especially during in the early period.

4.   Discussion and conclusion

Using ERA-40, NCEP–NCAR and ERA-Interim reanalysis data, as well as HadSLP2 data and instrumental temperature records over eastern China, we evaluated the spatial pattern and temporal variability of the EAWM described in the 20thCR dataset. The results showed that the 20thCR reanalysis data describe the spatial features of the EAWM system well, although there are some differences in the main circulation patterns over East Asia in comparison to the ERA-40 reanalysis data, such as a weaker Siberian High center, a slightly deeper East Asian Trough, and a slightly weaker East Asian Jet Stream center. The MV–EOF analysis also suggested that the 20thCR data reproduce the most typical configuration patterns of all sub-factors involved in the EAWM system reasonably well.

The EAWM index derived from the 20thCR is significantly correlated to that from the NCEP–NCAR, ERA-40, and ERA–Interim. This suggests that the 20thCR is highly reliable and stable in describing the temporal variability of EAWM since the 1950s. Compared with the longer time series of the Siberian High derived from HadSLP2 data and East China temperature from meteorological records, the 20thCR data reproduce the variability of EAWM in a credible way; however, a discernible discrepancy appears prior to the 1930s. According to the high-frequency correlation analysis, it can be concluded that the 20thCR dataset reproduces the high-frequency variability of EAWM better than the low-frequency variation, especially in the early period. In conclusion, caution should be taken when using 20thCR data to study the interdecadal variability or long-term trend of EAWM, especially prior to the 1930s.

REFERENCES

Allan, R., and T. Ansell, 2006: A new globally complete monthly historical mean sea level pressure data set (HadSLP2): 1850–2004. *J. Climate*, **19**, 5816–5842.

Chang, C. P., and K. M. Lau, 1982: Short-term planetary-scale interaction over the tropics and the mid-latitudes during northern winter. Part I: Contrast between active and inactive periods. *Mon. Wea. Rev.*, **110**, 933–946.

Chen, W., H. F. Graf, and R. H. Huang, 2000: The interannual variability of East Asian winter monsoon and its relation to the summer monsoon. *Adv. Atmos. Sci.*, **17**, 48–60.

Compo, G. P., G. N. Kiladis, and P. J. Webster, 1999: The horizontal and vertical structure of East Asian winter monsoon pressure surges. *Quart. J. Roy. Meteor. Soc.*, **125**, 29–54.

Compo, G. P., and Coauthors, 2011: The twentieth century reanalysis project. *Quart. J. Roy. Meteor. Soc.*, **137**, 1–28.

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.

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

Gong, D. Y., and C. H. Ho, 2002: The Siberian High and climate change over middle to high latitude Asia. *Theor. Appl. Climatol.*, **72**, 1–9.

Guo, Q. Y, 1994: Relationship between the variations of East Asian winter monsoon and temperature anomalies in China. *Quarterly Journal of Applied Meteorology*, **5**, 218–224. (in Chinese)

Huang, R. H., W. Chen, B. L. Yang, and R. H. Zhang, 2004: Recent advances in studies of the interaction between the East Asian winter and summer monsoon and ENSO cycle. *Adv. Atmos. Sci.*, **21**, 407–424.

Huang, R. H., J. L. Chen, L. Wang, and Z. D. Lin, 2012: Characteristics, processes and causes of the spatio-temporal variabilities of the East Asian monsoon system. *Adv. Atmos. Sci.*, **29**, 910–942, doi: 10.1007/s00376-012-2015-x.

Jhun, J., and E. Lee, 2004: A new East Asian winter monsoon index and associated characteristics of the winter monsoon. *J. Climate*, **17**, 711–726.

Ji, L. R., S. Q. Sun, K. Arpe, and L. Bengtsson, 1997: Model study on the interannual variability of Asian winter monsoon and its influence. *Adv. Atmos. Sci.*, **14**, 1–22.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-year reanalysis project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471.

Lau, K. M., and M. T. Li, 1984: The monsoon of

East Asia and its global associations-A survey. *Bull. Amer. Meteor. Soc.*, **65**, 114–125.

Li, C. Y., and M. Q. Mu, 2000: Relationship between East Asian winter monsoon, warm pool situation and ENSO cycle. *Chinese Science Bulletin*, **45**, 1448–1455.

Shi, N., 1996: Features of the East Asian winter monsoon intensity on multiple time scale in recent 40 years and their relation to climate. *Quarterly Journal of Applied Meteorology*, **7**, 175–182. (in Chinese)

Uppala, S. M., and Coauthors, 2005: The ERA-40 reanalysis. *Quart. J. Roy. Meteor. Soc.*, **131**, 2961–3012.

Wang, B., and Coauthors, 2008: How to measure the strength of the East Asian summer monsoon. *J. Climate*, **21**, 4449–4463.

Wang, L., and W. Chen, 2010: How well do existing indices measure the strength of the East Asian winter monsoon? *Adv. Atmos. Sci.*, **27**, 855–870, doi: 10.1007/s00376-009-9094-3.

Wang, L., R. H. Huang, L. Gu, W. Chen, and L. H.

Kang, 2009a: Interdecadal variations of the East Asian winter monsoon and their association with quasi-stationary planetary wave activity. *J. Climate.*, **22**, 4860–4872.

Wang, S. W., and Coauthors, 2009b: *Atlas of Seasonal Temperature and Precipitation Anomalies over China* (1880–2007). China Meteorological Press, 257pp.

Zhang, Y., K. Sperber, and J. Boyle, 1997: Climatology and interannual variation of the East Asian winter monsoon: Results from the 1979–95 NCEP/NCAR reanalysis. *Mon. Wea. Rev.*, **125**, 2605–2619.

Zhang, Z. Y., D. Y. Gong, M, Hu, and Y. N. Lei, 2012: Comparisons of the multiple East Asian winter monsoon indices and their relations to climate over eastern China. *Geographical Research*, **31,** 987–1003. (in Chinese)

Zhou, L. T., 2011: Impact of East Asian winter monsoon on rainfall over southeastern China and its dynamical process. *Int. J. Climatol.*, **31**, 677–686.