Clim Dyn
DOI 10.1007/s00382-017-3757-z

Partial least regression approach to forecast the East Asian winter monsoon using Eurasian snow cover and sea surface temperature

Lulu Yu[1] · Zhiwei Wu[1] · Renhe Zhang[1] · Xin Yang[1]

Received: 27 December 2016 / Accepted: 6 June 2017

Abstract Seasonal prediction of the East Asian (EA) winter monsoon (EAWM) is of great significance yet a challenging issue. In this study, three statistical seasonal prediction models for the EAWM are established using three leading modes of the Eurasian snow cover (ESC), the first leading mode of sea surface temperature (SST) and the four leading modes of the combination of the ESC and SST in preceding autumn, respectively. These leading modes are identified by the partial-least square (PLS) regression. The first PLS (PLS1) mode for the ESC features significantly anomalous snow cover in Siberia and Tibetan Plateau regions. The ESC second PLS (PLS2) mode corresponds to large areas of snow cover anomalies in the central Siberia, whereas the third PLS (PLS3) mode a meridional seesaw pattern of ESC. The SST PLS1 mode basically exhibits an El Niño-Southern Oscillation developing phase in equatorial eastern Pacific and significant SST anomalies in North Atlantic. A strong EAWM tends to emerge in a La Niña year concurrent with cold SST anomalies in the North Atlantic, and vice versa. After a 35-year training period (1967–2001), three PLS seasonal prediction models are constructed and the 11-year hindcast is performed for the period of 2002–2012, respectively. The PLS model based on combination of the autumn ESC and SST exhibits the best hindcast skill among the three models, its correlation coefficient between the observation and the hindcast reaching 0.86. This indicates that this physical-based PLS model may provide another practical tool for the EAWM. In addition, the relative contribution of the ESC and SST is also examined by assessing the hindcast skills of the other two PLS models constructed solely by the ESC or SST. Possible physical mechanisms are also discussed.

Keywords Statistical model · East Asian winter monsoon · Seasonal prediction

1 Introduction

The East Asian (EA) winter monsoon (EAWM), a relative shallow system compared with its summer counterpart (Yang et al. 2002), regulates the major variation of wintertime EA climate (e.g., Tao 1957; Zhang et al. 1996; Wang et al. 2000; and many others). During the past few decades, extreme climate and weather events such as snowstorms and cold spells were far more prevalent than before, especially during the 2007–2015 winter (Wen et al. 2009; Seager et al. 2010; Wu et al. 2011; Li and Wu 2012), causing considerable economic losses as well as social impacts on people's daily life. Thus, how to predict the EAWM variations is of practical importance.

The EAWM system encompasses tropics, subtropics and mid-latitudes and exhibits complex temporal and spatial structures. Many atmospheric circulation factors have been recognized to contribute to the EAWM variations, such as Siberian-Mongolian High (SMH) (Takaya and Nakamura 2005a, b; Ma et al. 2012), Southern Hemisphere annular mode (SAM) (Wu et al. 2009), Arctic Oscillation (AO) and

This paper is a contribution to the special issue on East Asian Climate under Global Warming: Understanding and Projection, consisting of papers from the East Asian Climate (EAC) community and the 13th EAC International Workshop in Beijing, China on 24–25 March 2016, and coordinated by Jianping Li, Huang-Hsiung Hsu, Wei-Chyung Wang, Kyung-Ja Ha, Tim Li, and Akio Kitoh.

✉ Zhiwei Wu
zhiweiwu@fudan.edu.cn

[1] Institute of Atmospheric Sciences (IAS), Fudan University, Shanghai 200433, China

Published online: 13 June 2017

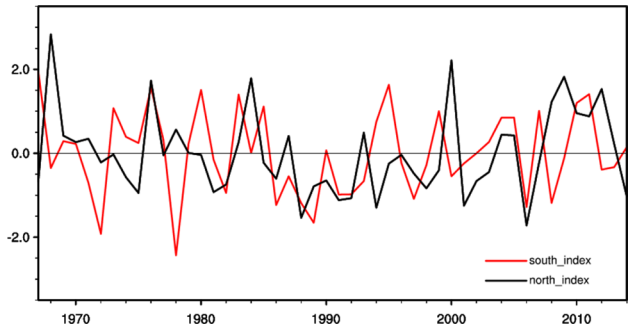

Fig. 1 Time series of the normalized East Asian winter monsoon index (EAWMI, Wang et al. 2010) for the northern (*black curve*) and southern modes (*red curve*), respectively (1967–2014). Winter refers to December–January–February (DJF) in this study

North Atlantic Oscillation (NAO) (Gong et al. 2001; Kim and Ahn 2012). Wang et al. (2010) identified two dominant modes of EA winter surface air temperature (SAT) (Fig. 1). A strong northern mode is characterized by an enhanced SMH (Fig. 2b) and a westward shift of the EA major trough (Fig. 2a), favoring a cold winter north of 40°N (Fig. 2c). A strong southern mode features a deepening EA trough (Fig. 2d–f) with cold air originated from western Mongolia, favoring a cold winter south of 40°N. Wu et al. (2011) also uncovered the third principal mode of EAWM, explaining only 8.7% of the total SAT variance but bringing about the catastrophic consequences such as the once-in-a-century snow storms in 2007/08 winter.

The predictability sources of the EAWM variations involve in air-sea-land interactions as well as the topographic forcing exerted by the Tibetan Plateau (TP) (Hsu

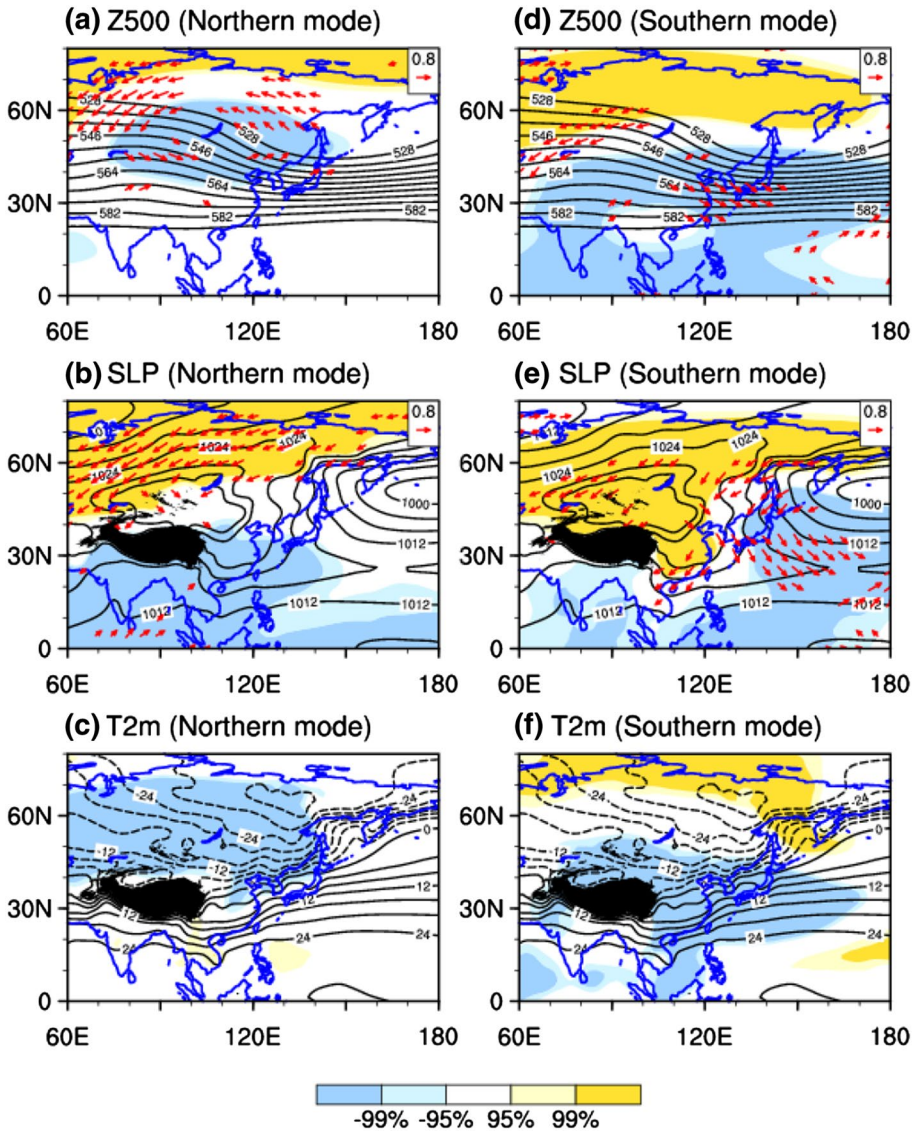

Fig. 2 The *left column* denotes the correlation pattern (*color shadings*) between the northern EAWMI and DJF **a** Z500, **b** SLP, **c** 2-m air temperature (T2m). *Contours* represent the DJF climatological mean **a** Z500 (10 gpm), **b** SLP (hPa), and **c** T2m (°C). The zonal and meridional components of the vectors represent, respectively, the correlation coefficients of the zonal and meridional winds with the northern EAWMI. The *color shadings* and vectors indicate the significant fields above the 95% confidence level. The *black shading* in Fig. 2b indicates the Tibetan Plateau region. The *right column* is the same as the left one except for the southern EAWMI

1987; Chang et al. 2006; Wang and Chen 2014). In terms of sea surface temperature (SST), El Niño-Southern Oscillation (ENSO), the major large-scale air-sea variability in the tropical Pacific, is viewed as an essential tropical forcing to the EAWM interannual variations (Zhang et al. 1996; Wang et al. 2000, 2010). Generally speaking, the EAWM tends to be weak in El Niño years, and vice versa (Li 1990). The key system linking the EAWM and ENSO is the anomalous Philippine Sea anticyclone (PSAC) (Zhang et al. 1996; Wang et al. 2000). Different from the tropical SST forcing, the extratropical SST is usually connected with the EAWM variation through its regulation on higher-latitude circulation systems such as SMH (Bates 2007), Aleutian Low (AL) (Bond et al. 2003), or the North Pacific Oscillation (NPO) (Wang et al. 2008; He and Wang 2013). Besides SST, snow cover in the northern hemisphere (NH) has been suggested to present the vital modulation on EA winter climate through the strong positive feedbacks related to albedo (Déry and Brown 2007; Cohen and Jones 2011) as well as the mean flow interaction processes across troposphere and stratosphere (Gong et al. 2003; Yu et al. 2016). These low-boundary forcings are considered as fundamental predictability sources for the EAWM variations (Charney and Shukla 1981; Trenberth and Hurrell 1994; Shukla 1998).

Owing to the complexity of the influencing factors and the intricate dynamical structures of the EAWM, seasonal prediction of the EAWM is deemed as a scientific challenge (Wang et al. 2009). Most dynamical seasonal predictions for the EAWM are unsuccessful (Shukla et al. 2000; Wang et al. 2009). Under this circumstance, statistical models are considered as an alternative way for the EAWM seasonal prediction owing to the linear connection between influential factors and the EAWM. Wu et al. (2011) established a physical-empirical (PE) model for the third leading mode of the EAWM which attains a promising skill in 2007/2008 winter. Lee et al. (2013) found that the dominant EAWM modes intimately related to ENSO is far more predictable by the statistical models than others. However, during the aforementioned linear statistical model establishment, the collinearity among the influential factors would make the regression coefficients unreliable or lead to the inaccurate prediction. Besides, most prior statistical strategies are based on SST. Lin and Wu (2012) suggested that it is possible that the snow cover effects are not properly presented in the dynamical models, which leads to the poor performance of dynamical prediction systems in winter, particularly the extreme winter conditions like 2009/10 winter. In light of these, we focus on the partial-least square (PLS) regression which can well eliminate the multicollinearity in an ordinary least square strategy. We also intend to identify the physical-based SST and Eurasian snow cover (ESC) modes to predict the strength of the EAWM and assess their relative contribution to the EAWM seasonal prediction.

In this paper, we attempt to answer two questions: (1) How are the EAWM variation connected with the previous ESC/SST anomaly? (2) What kind of contribution can the ESC/SST make to the EAWM variability? The whole paper is structured as following: Sect. 2 introduces the datasets and methodology used in this work. In Sect. 3, we analyze the prior autumn SST and snow cover anomalies during the abnormal EAWM years. The dominant ESC and SST modes preceding the EAWM are revealed by the PLS method in Sect. 4. Section 5 displays the PLS model results and highlights the contribution of ESC and SST in the EAWM prediction. Finally, conclusion and discussions are made in Sect. 6.

2 Data and method

The datasets utilized in the present study consists of: (1) the global land precipitation (PREC/L) data (Chen et al. 2002); (2) the monthly mean atmospheric fields, derived from the Japanese global atmospheric 55-year reanalysis data (JRA-55; Ebita et al. 2011); (3) monthly mean SST from Extended Reconstructed SST Version 3b (ERSST V3b; Smith et al. 2008); (4) Northern Hemisphere snow cover data at 7921 stations (derived online from http://climate.Rutgers.edu/snowcover/index, the Rutgers University Snow Lab). (5) Niño3.4 (5°N–5°S, 170°–120°W) index and unsmoothed Atlantic Multidecadal Oscillation (AMO) index (derived online from http://www.esrl.noaa.gov/psd/data/climateindices/list/). In this study, all data are validated for the period of 1967–2014. Winter refers to December–January–February (DJF), while autumn September–October–November (SON).

The EAWM index (EAWMI) used in this paper is according to the definition by Wang et al. (2010). This EAWMI is defined as 500 hPa geo-potential height averaged over region (30°–50°N, 110°–130°E). Wang et al. (2010) noted that the northern and southern modes of SAT can account for most of the total variance over the entire EAWM region but most EAWM circulation indices basically describe the southern mode (e.g., Guo 1983; Wu and Wang 2002; Yan et al. 2004). The EAWMI used in this work can well depict not only the southern SAT mode, but also the northern mode (Fig. 3). Its correlation coefficient with the southern SAT mode index (Fig. 1) reaches 0.70, while that with the northern mode index 0.64. This EAWMI also exhibits significant connection with other EAWM indices, i.e., defined by Guo (1983) (0.6), Wu and Wang (2002) (0.65) and Yan et al. (2004) (0.78). Besides, major circulation components of the EAWM are also well depicted by this EAWMI (Fig. 4).

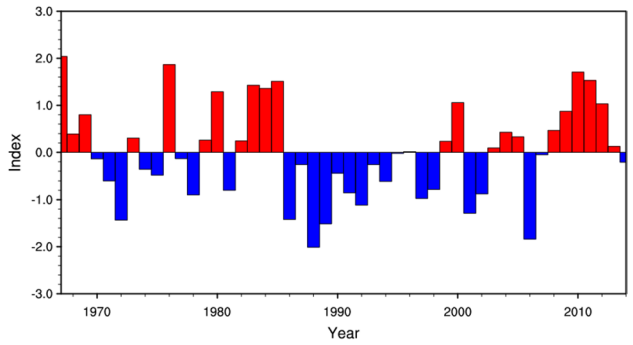

Fig. 3 Time series of the normalized combined EAWMI [Z500 averaged within the area (30°–50°N, 110°–130°E)] defined by Wang et al. (2010)

During a high EAWMI winter, the 300 hPa positive westerly anomalies over the EA jet stream core region concurrent with negative westerly anomalies over its pole-ward side would favor the cyclonic vorticity genesis (Fig. 4a) (Jhun and Lee 2004). This pattern may cause a deepening EA trough (Fig. 4b). The strong northwesterly surface wind anomaly intrudes along the eastern flank of enhanced SMH directly to the EA region (Fig. 4c), leading to extreme cold winter conditions (Fig. 4d) (Tao 1957). Meanwhile, poor precipitation mainly prevails over northern and East China.

The PLS method is firstly developed by Herman Wold in the 1960s (Wold 1966) for constructing predictive models when the variables are more than observation and highly collinear (Yeniay and Goktas 2002; Haenlein and Kaplan 2004). PLS has been widely used in computational biology (Tan et al. 2004), chemometrics (Haaland and Thomas 1988; Wold et al. 2001; Udelhoven et al. 2003) and neuroimaging (McIntosh and Lobaugh 2004). This technique was also applied in climate research for the purpose of statistical prediction (Smoliak et al. 2010; Zhang et al. 2011; Wu and Yu 2016; Song et al. 2016).

In this paper, we try to establish a EAWM prediction model based on the preceding autumn SST or ESC. The predictors X_{ij} is the autumn normalized SST or ESC variations preceding the EAWM (an i×j array, row i represents space times, also referred to as observations and column j a certain area of grid points or the number of predictors), while the predictand Y_i time series of the EAWMI. Note that the combined mode in Sect. 4 is actually a collection of SST and ESC data and the ESC data is interpolated to the Eurasian continent. Both the predictors and the predictand are standardized prior to performing the procedure. The analysis commences by calculating the correlation

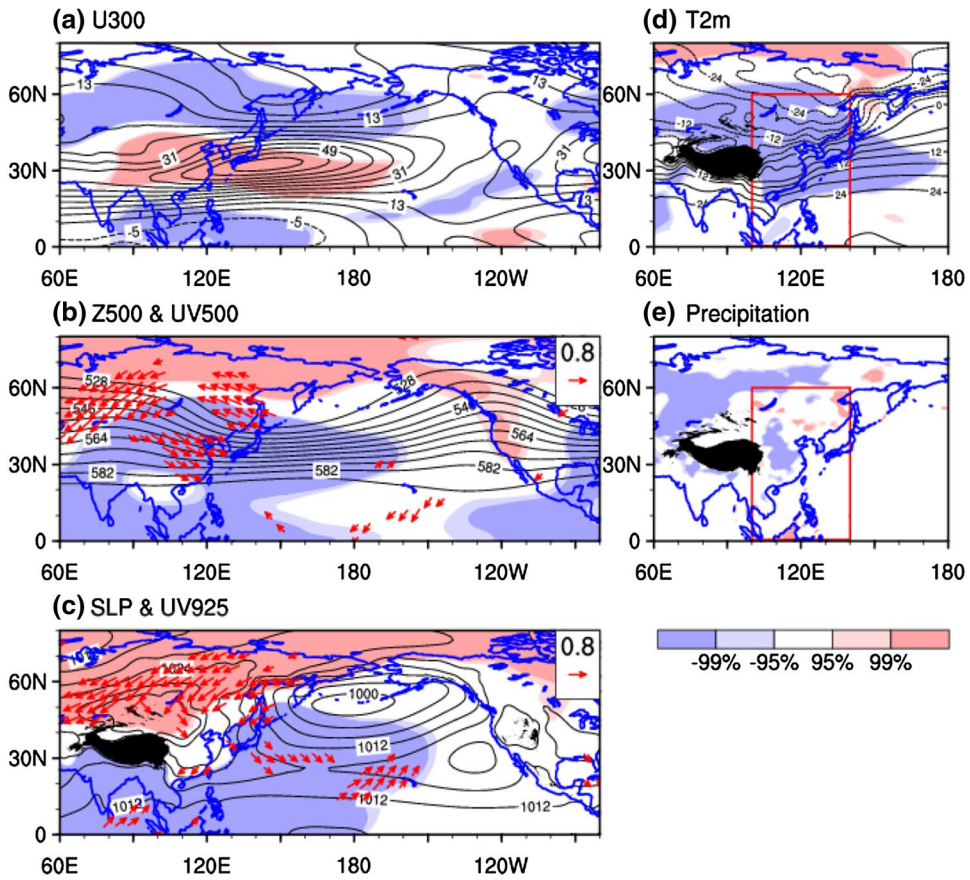

Fig. 4 The correlation pattern (*color shadings*) between the EAWMI in Fig. 3 and DJF **a** U300, **b** Z500, **c** SLP, **d** T2m, and **e** precipitation. *Contours* denote the DJF climatological mean **a** U300 (m s^{-1}), **b** Z500 (10 gpm), **c** SLP (hPa), and **d** T2m (°C). The zonal and meridional components of the vectors represent, respectively, the correlation coefficients of the zonal and meridional winds with the EAWMI. The *color shadings* and vectors indicate the significant fields above the 95% confidence level. The *red box* in **d** and **e** denotes the East Asia region (0–60°N, 90–140°E)

coefficients between Y_i and each observation of X_{ij} to get a correlation coefficients matrix b_j. Then each observation of X_{ij} is projected onto the b_j to obtain Z_1, the first predictor throughout the whole model establishment. Then build the regression equation between Z_1 and X_{ij}/Y_i respectively to reveal the residual fields for predictors X_{ij} and predictend Y_i. Next, the same procedure is performed on the residual fields of predictors X_{ij} and predictend Y_i till the majority of the covariance between X_{ij} and Y_i can be explained by $Z_1, \ldots, Z_k$ components, k representing the number of Z components during the regression. Note that $Z_1, \ldots, Z_k$ components are orthogonal to each other and arranged in order of decreasing variance of predictand Y_i. It should be mentioned that the optimal components for PLS models are determined by cross-validation using the training data. More detailed explanation and the mathematical framework of PLS is given by Song et al. (2016).

In the present example, the PLS method aims to find the leading modes of SST variations that best explain the covariance between SST and the EAWMI variations. This method is totally different from the principal component regression (PCR) or empirical orthogonal function (EOF) analysis. In other words, the SST mode uncovered by the PLS method can explain both variance of SST and the EAWMI, whereas that revealed from the EOF or PCR could only explain the variance of SST. Since both PCR and PLS are better to handle the collinearity among the regressors, regression model constructed by PLS has the higher predictive ability using the smaller number of factors than PCR (Yeniay and Goktas 2002).

3 Autumn SST/ESC anomalies preceding the abnormal EAWM

The general snow-monsoon and SST-monsoon relationships have been demonstrated in many works (Wang et al. 2000; Gong et al. 2003; Cohen and Jones 2011; He and Wang 2013). In this study, the anomalous EAWM years are defined when the normalized EAWMI is more than 0.8 or less than −0.8. This threshold value provides sufficient cases to contrast their snow cover features preceding the strong and weak monsoon winter. The strong EAWM years include: 1972, 1978, 1986, 1988, 1989, 1991, 1992, 1997, 2001, 2002, and 2006, whereas the weak years: 1967, 1976, 1980, 1983, 1984, 1985, 2000, 2009, 2010, 2011 and 2012.

To identify the snow cover and SST anomalies preceding the abnormal EAWM, Fig. 5 shows the autumn SST and snow cover composite differences between the strong and weak EAWM years (strong minus weak). Large areas of significant snow cover anomalies are mainly located in the central Eurasian continent with positive anomalies over the TP region and expand northeastward towards

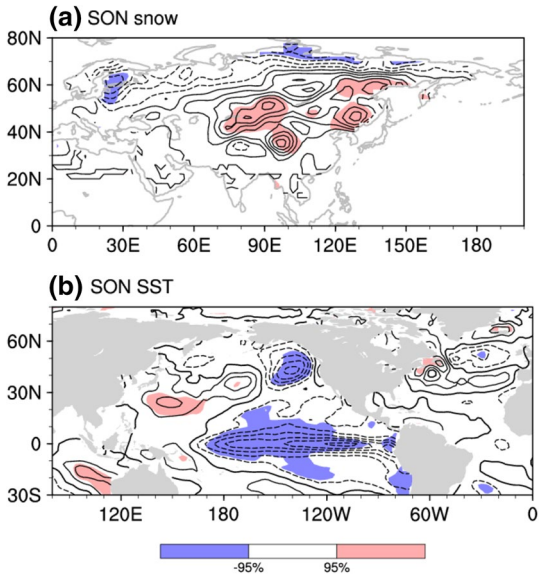

Fig. 5 Composite difference of September–October–November (SON) **a** snow cover and **b** SST between the strong and weak EAWM years (strong minus weak). A strong EAWM year refers to its corresponding EAWMI value greater than 0.8 times standard deviation, whereas a weak EAWM year less than −0.8 times standard deviation. The *shaded areas* exceed the 95% confidence level based on the Student's *t* test

northeastern Siberia (Fig. 5a). These crucial snow cover areas are basically consistent with those identified by previous results. For instance, Wang et al. (2010) found that snow cover in central Siberia and Russian far-east regions provides precursory conditions for the anomalous EAWM northern and southern mode, respectively. In addition, snow cover anomalies in due north of the Eurasian continent also signify precursors for the abnormal EAWM (Fig. 5a).

Figure 5b displays the autumn SST anomalies prior to the anomalous EAWM. The major SST difference between the strong and weak EAWM years is a pronounced ENSO-like pattern in North Pacific. This indicates that ENSO is the most significant predictability source in SST for the abnormal EAWM. Besides, North Atlantic and tropical Indian Ocean also show some significant SST signals in preceding autumn. The former basically exhibits an anomalous tri-pole SST pattern, while the latter is adjacent to the Maritime Continent.

To summarize, the SST and ESC anomalies preceding the abnormal EAWM imply that these autumn low-boundary forcings may play an essential role in the seasonal prediction of the EAWM variations. However, the specific details need further investigations, particularly the relative importance of SST and ESC in contribution to the EAWM seasonal prediction skill.

4 Dominant SST and ESC modes preceding the EAWM

Before establishing a statistical prediction model with the PLS regression (e.g., Zhang et al. 2011; Wu and Yu 2016), we need to identify the dominant SST and ESC modes preceded with the abnormal EAWM by using of the PLS regression with the EAWMI and SON SST (or ESC).

Figure 6 presents the first PLS (PLS1) mode, including prior autumn SST pattern and the corresponding time series for the 35-year period (1967–2001). The percentage numbers in Fig. 6a denote the significance of the PLS1 mode explaining the total SST (11%) and EAWMI (47.7%) variances, respectively. The PLS1 mode exhibits a tripole SST anomaly in the North Pacific (Fig. 6a). The correlation coefficient between the PLS1 time series and the SON Niño3.4 SST index attains 0.42, exceeding the 95% confidence level based on the Student's *t* test. In addition, the SST PLS1 mode also exhibits basin-wide negative anomalies in the North Atlantic, resembling an Atlantic Multi-decadal Oscillation (AMO) pattern (Enfield et al. 2001). The correlation coefficient between the PLS1 time series and the unsmoothed AMO index (the area-averaged SST anomaly over North Atlantic region) is 0.50. Note that we did a series of sensitive experiments by altering the SST domains and the number of leading PLS modes to reduce the dispensable information (noisy signals) to the minimum. Finally, it is found that the optimal SST region is within (120°E–360°, 5°N–80°N) and the PLS1 mode for SST is sufficient for the EAWM prediction skill based on results of cross-validation.

To further interpret the physical linkage between the SST PLS1 mode and the EAWM, Fig. 7 shows the seasonal evolution of SST patterns associated with the PLS1 mode from D(0)JF(1) to SON(1). The year in which the SST PLS1 mode occurs is denoted as year "0" (Fig. 6), whereas the next year as year "1". The significantly negative SST anomalies in the equatorial central-eastern Pacific sustain from D(0)JF(1) to MAM(1) (Fig. 7a, b) and weaken since JJA(1). This basically corresponds to an ENSO decaying phase. Meanwhile, the anomalously negative SST anomalies in the whole North Atlantic basin persist throughout the next year (Fig. 7).

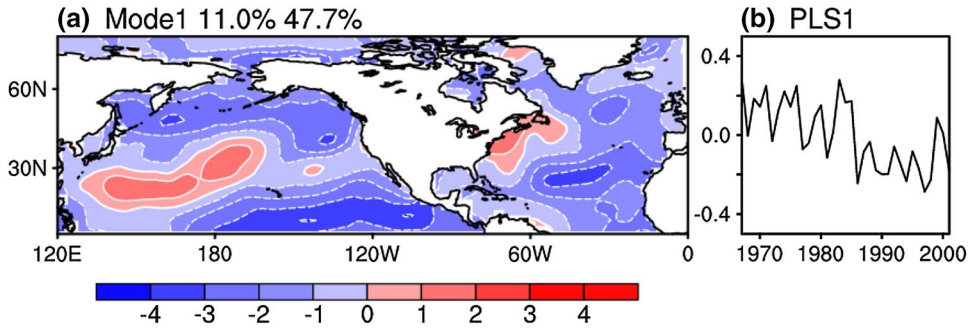

Fig. 6 **a** The first leading mode and **b** time series of SON sea surface temperature (SST) derived from the partial-least square (PLS) regression analysis (1967–2001). The percentage numbers represent the SST (*left*) and EAWMI (*right*) variance explained by the same SST mode, respectively

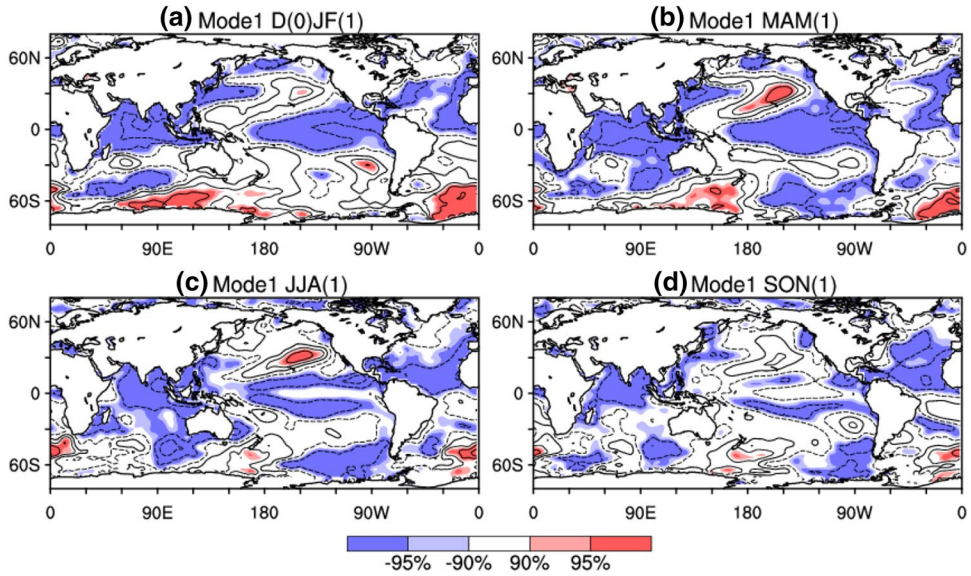

Fig. 7 **a** D(0)JF(1), **b** MAM(1), **c** JJA(1), **d** SON(1) SST anomalies (*contours* in units of K) regressed against the PLS1time series in Fig. 6b. The *light* and *dark shadings* denote 90 and 95% confidence levels. "0" denotes the simultaneous year, while "1" the next year. MAM refers to March–April–May and JJA June–July–August

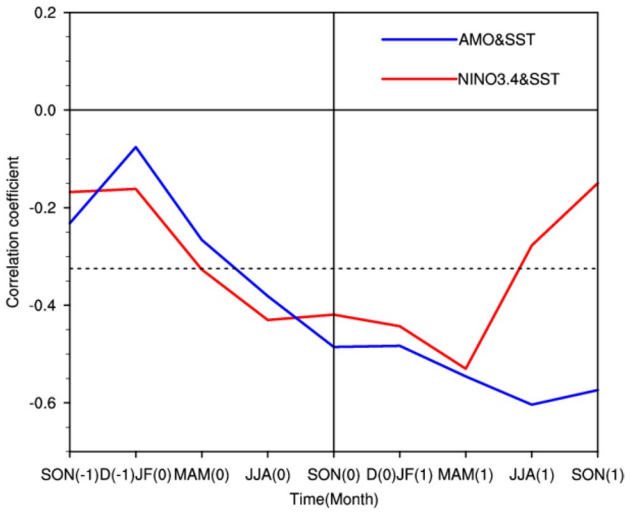

Fig. 8 The lead-lag correlation coefficients between SST PLS1 time series in Fig. 6b and the Niño3.4 index (*red curve*)/AMO index (*blue curve*) from SON(−1) to SON (1) respectively. The *horizontal line* represents the correlation coefficients significant at the 95% confidence level. The *vertical line* indicates SON(0), where the simultaneous correlations are shown

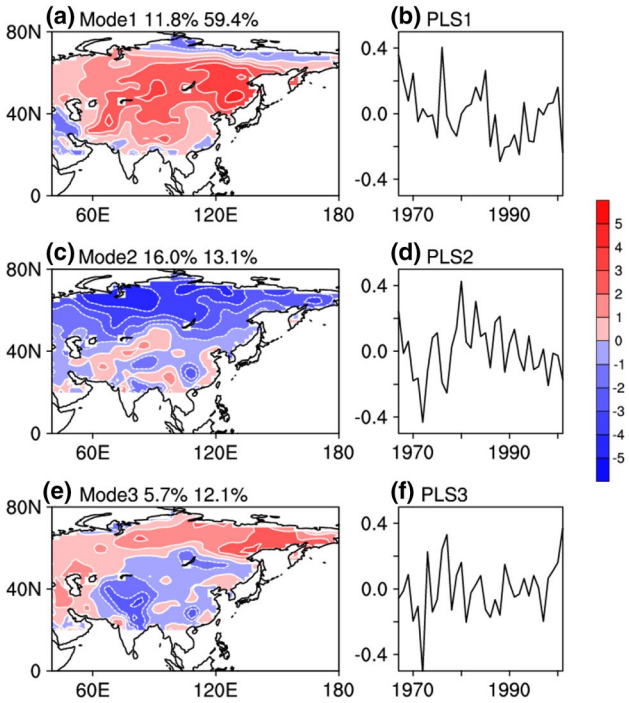

Fig. 9 The same as Fig. 6, but for the SON Eurasian snow cover (ESC)

The above results can also be discerned from the lead-lag correlations between PLS1 time series and the Niño3.4 index (red curve)/AMO index (blue curve) (Fig. 8). The PLS1 mode exhibits a stable and significant linkage with ENSO (or the Niño3.4 index) from SON(0) to JJA(1) (red curve). Such SST evolution tendency implies that the PLS1 SST pattern in SON(0) inclines to precede the ENSO mature phase and decay with the ENSO events. Wang et al. (2000) suggested that the key system connecting the ENSO and the EAWM is the PSAC. The northeasterly winter monsoon over southern EA tends to be weakened by the anomalous southwesterly winds to the west of the PSAC center. As for AMO, its correlation with the SST PLS1 mode is significant throughout the next year with Atlantic basin-wide SST anomalies on the EAWM interdecadal variations (Bates 2007).

Same as above, we did a series of sensitive experiments by using of the PLS regression with the EAWMI and SON ESC and found that the optimal ESC region is within (40°E–180°, 20°N–80°N) and the three leading PLS modes for ESC are sufficient for the EAWM prediction skill based on the cross-validation. Figure 9 presents the preceding autumn ESC patterns and the corresponding time series for 35-year period (1967–2001). The three leading ESC modes could explain 84.6%/33.5% of the EAWMI/ESC variances. The ESC PLS1 mode, explaining more than half of the EAWMI variance, is characterized by a continental-scale mono-sign ESC anomaly pattern (Fig. 9a). The extreme center is located in the central-eastern Siberia, expanding southwestward towards TP and the adjacent area. Preceding a strong EAWM, the excessive autumn ESC can reduce the amount of solar radiation absorbed by the ground due to the increasing albedo, and favor an intensified SMH over the mid-high latitudes (Fig. 10d) (Barnett et al. 1989). These results further verify those in Fig. 4c.

The ESC PLS2 and PLS3 modes, explaining 25.2% of the EAWMI variance, basically exhibit a north–south see-saw ESC anomaly pattern. Such opposite ESC anomaly distribution implies dissimilar snow cover impacts on the northern and southern Eurasia. Wang et al. (2010) proposed two distinct EAWM modes corresponding to two different snow cover forcings. A strong northern mode is usually preceded by excessive snow cover over central Siberia and a strong southern mode by reduced snow cover over the northeastern Siberia. The most evident ESC anomalies associated with the PLS2 mode are located in the region north of 40°N, resembling a northern mode. The ensuring winter circulation exhibits an obvious Z500 and sea level pressure (SLP) anomaly belt over the northwestern Pacific (Fig. 10b, e), favoring a zonal displacement of the EA grand trough. The PLS3 mode is characterized by significant snow cover anomalies in the northern Siberia, corresponding to a southern EAWM mode. The Z500 and SLP anomalies exhibit more meridional-oriented (Fig. 10c, f), favoring a meridional displacement of the EA grand trough and steering the cold air mass invasion to the southern EA.

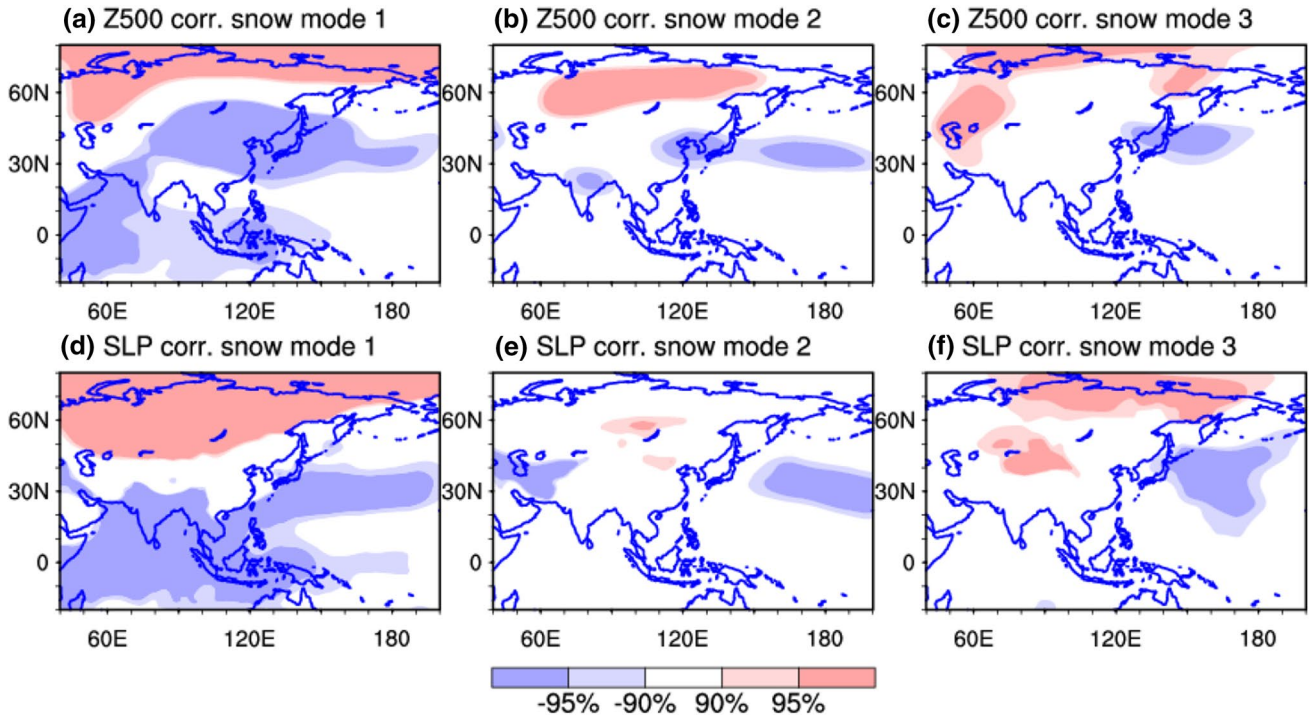

Fig. 10 The *left column* represents lag correlation between ESC PLS1 time series in Fig. 9b and DJF (**a**) Z500 (*color shadings*), **d** SLP (*color shadings*). *Color shadings* indicate the significant fields above the 90%/95% confidence level. The *middle and right column* are the same as the *left column*, but for time series of the second and third mode, respectively

The aforementioned result demonstrates that the anomalous ESC or SST may provide precursory conditions for the abnormal EAWM. To further analyze the co-mingled effects of the SST and ESC, Fig. 11 shows the four leading combined PLS modes and corresponding time series. Note that same as above, the sensitive experiments indicate that the four combined PLS modes are sufficient for the EAWM prediction skill based on the cross-validation. The combined modes account for 97.8% of the total EAWM variance. The first combined mode explains nearly 70% of the EAWM variance. Its combined pattern and evolution features (Figs. 11a, 12) basically resemble those associated with the sole SST or first ESC mode. The correlation coefficient for the first combined (Fig. 11b) and the SST PLS1 mode (Fig. 6b) reaches 0.92, whereas that for the first combined and the ESC mode (Fig. 9b) 0.81. The other three combined modes (Fig. 11c, e, g) accounting about 28% of the total EAWM variance also show clearly connection with corresponding sole ESC and SST modes and the correlation coefficients between combined PLS and SST(ESC) PLS all exceed the 95% confidence level based on the student's t test (Table 1).

To summarize, the preceding ESC and SST anomalies could provide potential predictability sources for the EAWM variations. To quantify the ESC or/and SST contribution to the EAWM seasonal prediction skill, we will construct the statistical prediction models based on the ESC, SST and combined PLS leading modes, respectively.

5 PLS prediction models

The PLS models are established using 1967–2001 as a training period, and then a 11-year (2002–2012) hindcast is conducted to examine the skill of these PLS models.

The optimal components for fitting PLS models are determined by cross-validation using the training data, seen in Sect. 4. We choose the first leading SST mode (Fig. 6), three leading ESC modes (Fig. 9) and four leading modes of the combination of ESC and SST (Fig. 11) to build the PLS models, respectively. The prediction equation is addressed as follows:

$$Y_{predict} = X \times Beta + Y_{residual}. \quad (1)$$

Two formulas are included in Eq. (1): $Y_{fit} = X \times Beta$, $Y_{predict} = Y_{fit} + Y_{residual}$. *Beta* is the PLS regression coefficients, a p + 1-by-1 matrix (p represents the number of predictors), containing intercept terms in the first row. X

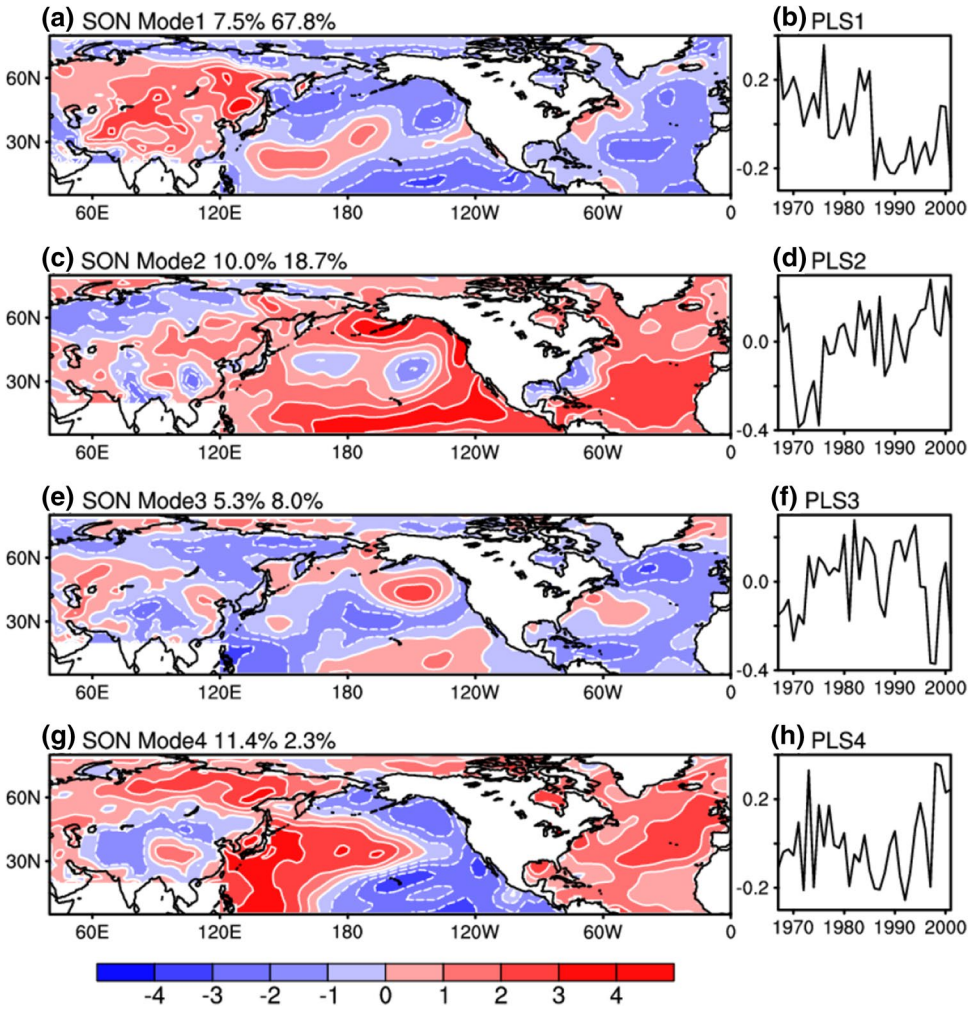

Fig. 11 The same as Fig. 6, but for the SON snow cover and SST, called combined mode

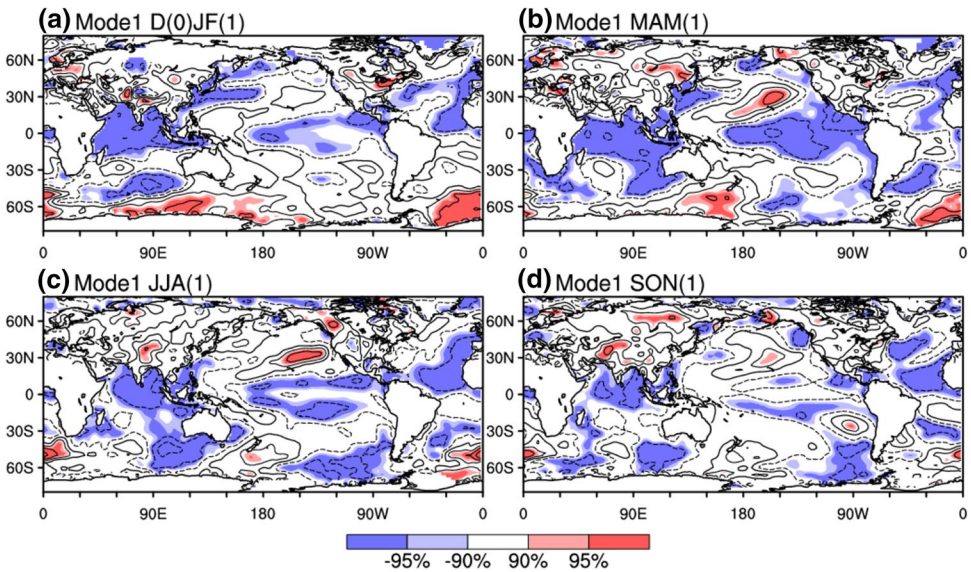

Fig. 12 Same as Fig. 7 but for time series of the SON combined PLS1 mode in Fig. 11

Table 1 Correlation between combined PLS and the corresponding sole SST/ESC PLS

Combined model	SST/ESC model	CORR
Combine_1	SST_1	0.92
	ESC_1	0.85
Combine_2	SST_2	0.96
	ESC_2	0.42
Combine_3	SST_3	0.73
	ESC_3	0.57
Combine_4	SST_4	0.9
	ESC_4	0.64

is the predictors, an n-by-p matrix (n represents the observations). Y_{fit} is the n-by-1 matrix, treated as the fitted value of the PLS model. $Y_{residual}$ represents the difference between the fitted value and the corresponding observation, $Y_{residual} = Y_{observe} - Y_{fit}$.

Take the EAWMI in 2002 as an example. The prediction of the EAWMI in 2002 uses the autumn SST (referred to as X, an n + 1-by-p matrix) from 1967 to 2002 as well as the EAWMI (referred to as Y, an n-by-1 matrix) for the period of 1967–2001, and so on. *Beta* comes from the 35-year (1967–2001) training process, during which the PLS method is performed with the observed autumn SST and EAWMI for the period of (1967–2001), $Y_{fit(1967-2001)} = X_{1967-2001} \times Beta$. After that, the SON SST data for the year 2002 is input into the equation, $Y_{fit2002} = X_{2002} \times Beta$. The $Y_{residual}$ for the year 2001 ($Y_{residual2001} = Y_{observe2001} - Y_{fit2001}$) obtained from the training period is added to "$Y_{fit2002}$" to get the predicted EAWMI in 2002 ($Y_{predict2002} = Y_{fit2002} + Y_{residual2001}$). The PLS prediction procedures using the ESC or combination of SST and ESC(SST + ESC) are the same as above. The "X" in Eq. (1) is replaced by ESC or SST + ESC.

Figure 13 presents the prediction skill of these PLS models. During the 35-year training period (1967–2001), all three models can well reconstruct the EAWMI time series (see Cor0 values at the left side of the vertical purple line). In general, the 11-year hindcasts (2002–2012) of these PLS models exhibit promising skills (see Cor1 values at the right side of the vertical purple line) which are superior to the ensemble mean of versions 3 and 4 of the Canadian Community Atmosphere Model (CanCM3/4) hindcasts from the newly developed North American Multi-model Ensemble (NMME) Prediction System. Table 2 shows the the correlation and root mean square error (RMSE) between model results and

Table 2 Correlation (CORR.) and root mean square error (RMSE) between observed and predicted values of CanCM3/4 and PLS models during the training and testing periods

Model	Training (1967–2001)		Testing (2002–2012)	
	RMSE	CORR	RMSE	CORR
SST model	0.79	0.69	0.71	0.74
Snow model	0.42	0.92	0.60	0.84
C_model	0.39	0.98	**0.59**	**0.86**
CanCM3/4	–	–	0.80	0.69

The bold values represent the best performer

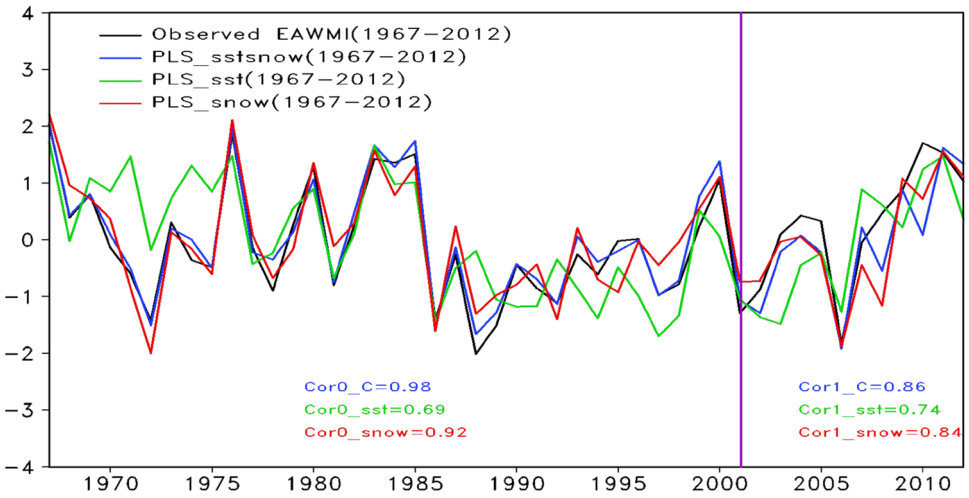

Fig. 13 Results of the three PLS regression model for predicting the EAWMI. The *black curve* represents the observed EAWMI (1967–2012). The results for the 1967–2001 period are the reconstructed EAWMI, while those for the 2002–2012 period are the 11-year hindcast. The *vertical purple curve* denotes 2002. The combined model output is the *blue curve*, SST-based model *orange curve*, and ESC-based model *green curve*. Corr0 represents the correlation coefficients between the observed and reconstructed EAWMI (1967–2001), while Corr1 the correlation coefficients for the observed and hindcast EAWMI (2002–2012)

observations during training and prediction period. The correlation coefficient for the observed and SST-based model predicted EAWMI is 0.74 (RMSE = 0.71), whereas that of the ESC-based model (Corr. = 0.84, RMSE = 0.60) is higher than the SST model. This highlights the importance of the ESC in the EAWM prediction. Moreover, if we use the combination of the ESC and SST to predict the EAWMI, the prediction skill reaches the highest (Corr = 0.86, RMSE = 0.59).

In general, all three PLS models show promising skills in the EAWM prediction, superior to the hindcast results from CanCM3/4. The ESC might contribute more to the prediction skill than SST as far as the latest 11-year hindcast is concerned. The PLS model established by the combination of the ESC and SST achieves the best performance.

6 Conclusion and discussion

Seasonal prediction of the EAWM has long been a challenging issue and receives increasing attention recently due to the higher frequency of the extreme weather events (Wen et al. 2009; Seager et al. 2010; Wang et al. 2010; Wu et al. 2011; Lin and Wu 2012). To identify potential predictability sources of the EAWM, we use the PLS regression to identify the leading SST and ESC modes preceding the abnormal EAWM. The SST PLS1 mode is usually associated with the ENSO cycle and SST anomalies in the North Atlantic. These autumn SST forcings may provide an essential predictability source for the EAWM (Zhang et al. 1998; Frankignoul and Sennechael 2007) or AO pattern (Kim and Ahn 2012). The ESC PLS1 mode exhibits continental-scale snow cover anomalies centered in the Siberian region and expanding towards Tibetan Plateau, favoring the enhancement (or weakening) of SMH. The ESC PLS2 and PLS3 modes correspond to the distinct northern and southern EAWM modes suggested by Wang et al. (2010), respectively. The four PLS modes of the combination of the ESC and SST is basically the admixture of the sole SST and ESC PLS modes.

To examine relative contributions of SST and the ESC, three PLS models are established based on a 35-year training period (1967–2001). The 11-year hindcast is performed for the period of 2002–2012 for each PLS model. Results manifest that the ESC-based model exhibits a higher skill than the SST-based, highlighting the importance of the autumn ESC in the EAWM prediction (Jhun and Lee 2004; Wang et al. 2010). Furthermore, the PLS model based on both the ESC and SST achieves the best performance among the three models. This combined PLS model may provide a useful reference for the practical seasonal prediction of the EAWM.

This work emphasizes effects of autumn snow cover and SST in seasonal prediction of the EAWM. It should be mentioned that besides the ESC and SST, other low boundary forcings, i.e., Arctic sea ice, are also considered as critical modulators on the NH winter climate (Wu et al. 2011; Li and Wu 2012; Liu et al. 2012). Then, how well would the PLS model predict the EAWM variations with the Arctic sea ice added? Besides, we use the autumn ESC and SST to predict the EAWM in this paper. If predictors in an earlier season, i.e., prior summer, are used in the PLS model, how well would the performance of the PLS model change? These are still outstanding issues and call for further studies.

Acknowledgements The authors are grateful to the Rutgers University Global Snow Lab for providing snow cover data. This work is jointly supported by the Ministry of Science and Technology of China (Grant Nos. 2016YFA0601801, 2015CB453201, 2015CB953904 and 2013CB430202) and the National Natural Science Foundation of China (NSFC) (Grant Nos. 41575075, 91437216 and 91637312).

References

Barnett TP, Dümenil K, Schlese U, Roeckner E, Latif M (1989) The effect of Eurasian snow cover on regional and global climate variations. J Atmos Sci 46:661–685

Bates GT (2007) Influence of the Atlantic multidecadal oscillation on the winter climate of East China. Adv Atmos Sci 24:126–135

Bond NA, Overland JE, Spillane M, Stabeno P (2003) Recent shifts in the state of the North Pacific. Geophys Res Lett 30:2183

Chang CP, Z Wang, H Hendon (2006) The Asian winter monsoon. In: Wang B (ed) Asian Monsoon, Praxis, pp 89–127

Charney JG, Shukla J (1981) Predictability of monsoons. In: Lighthill J, Pearce RP (eds) Monsoon dynamics. Cambridge University Press, New York, pp 99–109

Chen M, Xie P, Janowiak JE, Arkin P (2002) Global land precipitation: a 50-year monthly analysis based on gauge observations. J Hydrometeorol 3:249–266

Cohen J, Jones J (2011) A new index for more accurate winter predictions. Geophys Res Lett 38:L21701

Déry SJ, Brown RD (2007) Recent Northern Hemisphere snow cover extent trends and implications for the snow-albedo feedback. Geophys Res Lett 34:L22504

Ebita A et al (2011) The Japanese 55-year reanalysis "JRA-55": an interim report. SOLA 7:149–152

Enfield DB, Mestas-Nuñez AM, Trimble PJ (2001) The Atlantic multidecadal oscillation and its relation to rainfall and river flows in the continental US. Geophys Res Lett 28:2077–2080

Frankignoul C, Sennéchael N (2007) Observed influence of North Pacific SST anomalies on the atmospheric circulation. J Clim 20:592–606

Gong DY, Wang SW, Zhu JH (2001) East Asian winter monsoon and Arctic oscillation. Geophys Res Lett 28:2073–2076

Gong G, Entekhabi D, Cohen J (2003) Modeled Northern Hemisphere winter climate response to realistic Siberian snow anomalies. J Clim 16:3917–3931

Guo QY (1983) Relationship between the variations of East Asian winter monsoon and temperature anomalies in China. Chin J Appl Meteorol 5:218–225

Haaland DM, Thomas EV (1988) Partial least-squares methods for spectral analyses. 2. Application to simulated and glass spectral data. Anal Chem (US) 60(11):1202–1208

Haenlein M, Kaplan AM (2004) A beginner's guide to partial least squares analysis. Understanding Stat 3:283–297

He SP, Wang HJ (2013) Oscillating relationship between the East Asian winter monsoon and ENSO. J Clim 26:9819–9838

Hsu HH (1987) Propagation of low-level circulation features in the vicinity of mountain ranges. Mon Weather Rev 115:1864–1893

Jhun JG, Lee EJ (2004) A new East Asian winter monsoon index and associated characteristics of winter monsoon. J Clim 17:711–726

Kim HJ, Ahn JB (2012) Possible impact of the autumnal North Pacific SST and November AO on the East Asian winter temperature. J Geophys Res Atmos 117:D12104

Lee JY, Lee SS, Wang B, Ha KJ, Jhun JG (2013) Seasonal prediction and predictability of the Asian winter temperature variability. Clim Dyn 41:573–587

Li CY (1990) Interaction between anomalous winter monsoon in East Asia and El Nino events. Adv Atmos Sci 7:36–46

Li JP, Wu ZW (2012) Importance of autumn Arctic sea ice to northern winter snowfall. Proc Natl Acad Sci 109:E1898

Lin H, Wu ZW (2012) Contribution of Tibetan Plateau snow cover to the extreme winter conditions of 2009/10. Atmos Ocean 50:86–94

Liu J, Curry JA, Wang H, Song M, Horton RM (2012) Impact of declining Arctic sea ice on winter snowfall. Proc Natl Acad Sci USA 109:4074–4079

Ma T, Wu ZW, Jiang ZH (2012) How does coldwave frequency in china respond to a warming climate? Clim Dyn 39:2487–2496

McIntosh AR, Lobaugh NJ (2004) Partial least squares analysis of neuroimaging data: applications and advances. Neuroimage 23:S250–S263

Seager R, Kushnir Y, Nakamura J, Ting M, Naik N (2010) Northern Hemisphere winter snow anomalies: ENSO, NAO and the winter of 2009/10. Geophys Res Lett 37:L14703

Shukla J (1998) Predictability in the midst of chaos: a scientific basis for climate forecasting. Science 282:728–731

Shukla J, Marx L, Paolino D et al (2000) Dynamical seasonal prediction. Bull Am Meteorol Soc 81:2593–2606

Smith T, Reynolds R, Peterson T, Lawrimore J (2008) Improvements to NOAA's historical merged land-ocean surface temperature analysis (1880–2006). J Clim 21:2283–2296

Smoliak BV, Wallace JM, Stoelinga MT, Mitchell TP (2010) Application of partial least squares regression to the diagnosis of year-to-year variations in Pacific Northwest snowpack and Atlantic hurricanes. Geophys Res Lett 37:L03801

Song L, Duan W, Li Y, Mao J (2016) A timescale decomposed threshold regression downscaling approach to forecasting South China early summer rainfall. Adv Atmos Sci 33:1071–1084

Takaya K, Nakamura H (2005a) Mechanisms of intraseasonal amplification of the cold Siberian high. J Atmos Sci 62:4423–4440

Takaya K, Nakamura H (2005b) Geographical dependence of upper-level blocking formation associated with intraseasonal amplification of the Siberian high. J Atmos Sci 62:4441–4449

Tan Y, Shi L, Tong W et al (2004) Multi-class tumor classification by discriminant partial least squares using microarray gene expression data and assessment of classification models. Comput Biol Chem 28:235–243

Tao SY (1957) A study of activities of cold airs in East Asian winter. In: China Meteorological administration (ed) Handbook of short-term forecast. Meteorology Press, pp 60–92

Trenberth KE, Hurrell JW (1994) Decadal atmosphere-ocean variations in the Pacific. Clim Dyn 9:303–319

Udelhoven T, Emmerling C, Jarmer T (2003) Quantitative analysis of soil chemical properties with diffuse reflectance spectrometry and partial least-square regression: a feasibility study. Plant Soil 251:319–329

Wang L, Chen W (2014) An intensity index for the East Asian winter monsoon. J Clim 27:2361–2374

Wang B, Wu RG, Fu XH (2000) Pacific–East Asia teleconnection: how does ENSO affect East Asian climate? J Clim 13:1517–1536

Wang L, Chen W, Huang RH (2008) Interdecadal modulation of PDO on the impact of ENSO on the East Asian winter monsoon. Geophys Res Lett 35:L20702

Wang B, Lee JY, Kang IS et al (2009) Advance and prospectus of seasonal prediction: assessment of the APCC/CliPAS 14-model ensemble retrospective seasonal prediction (1980–2004). Clim Dyn 33:93–117

Wang B, Wu ZW, Chang CP, Liu J, Li JP, Zhou TJ (2010) Another look at interannual-to-interdecadal variations of the East Asian winter monsoon: the northern and southern temperature modes. J Clim 23:1495–1512

Wen M, Yang S, Kumar A, Zhang P (2009) An analysis of the large-scale climate anomalies associated with the snowstorms affecting China in January 2008. Mon Weather Rev 137:1111–1131

Wold H (1966) Estimation of principal components and related models by iterative least squares. Multivar Anal 1:391–420

Wold S, Sjöström M, Eriksson L (2001) PLSregression: a basic tool of chemometrics. Chemom Intell Lab Syst 58:109–130

Wu BY, Wang J (2002) Winter Arctic oscillation, Siberian high and East Asian winter monsoon. Geophys Res Lett 29:1897

Wu ZW, Yu L (2016) Seasonal prediction of the East Asian summer monsoon with a partial-least square model. Clim Dyn 46:3067–3078

Wu ZW, Li JP, Wang B, Liu X (2009) Can the Southern Hemisphere annular mode affect China winter monsoon? J Geophys Res 114:D11107

Wu ZW, Li JP, Jiang ZH, He JH (2011) Predictable climate dynamics of abnormal East Asian winter monsoon: once-in-a-century snowstorms in 2007/2008 winter. Clim Dyn 37:1661–1669

Yan HM, Duan W, Xiao ZN (2004) A study on relation between east Asian winter monsoon and climate change during raining season in China. J Trop Meteorol (in Chinese) 10:24

Yang S, Lau KM, Kim KM (2002) Variations of the East Asian jet stream and Asian–Pacific–American winter climate anomalies. J Clim 15:306–325

Yeniay O, Goktas A (2002) A comparison of partial least squares regression with other prediction methods. Hacettepe J Math Stat 31:99–101

Yu B, Lin H, Wu ZW, Merryfield WJ (2016) Relationship between North American winter temperature and large-scale atmospheric circulation anomalies and its decadal variation. Environ Res Lett 11:074001

Zhang RH, Sumi A, Kimoto M (1996) Impact of El Niño on the East Asian monsoon: a diagnostic study of the '86/87 and '91/92 events. J Meteorol Soc Jpn 74:49–62

Zhang Y, Norris JR, Wallace JM (1998) Seasonality of large-scale atmosphere-ocean interaction over the North Pacific. J Clim 11:2473–2481

Zhang HQ, Qin J, Li Y(2011) Climatic background of cold and wet winter in southern China: part 1. Observational analysis. Clim Dyn 37:2335–2354