

Seasonal Prediction of Winter Haze Days in the North-Central North China Plain

Zhicong Yin^{1,2}, Huijun Wang^{1,2,3}

¹[Key Laboratory of Meteorological Disaster, Ministry of Education / Joint International Research Laboratory of Climate and Environment Change \(ILCEC\) / Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters \(CIC-FEMD\)/Key Laboratory of Meteorological Disaster](#), Nanjing University of Information Science & Technology, Nanjing 210044, China

²Nansen-Zhu International Research Centre, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China

³Climate Change Research Center, Chinese Academy of Sciences, Beijing, China

Correspondence to: Zhicong Yin (yinzhc@163.com)

Abstract. Recently, the winter (December–February) haze pollution over the North-Central North China Plain (NCP) has become severe. By treating the year-to-year increment as the predictand, two new statistical schemes were established using the multiple linear regression (MLR) and the generalized additive model (GAM)-[approaches](#). By analyzing the associated increment of atmospheric circulation, seven leading predictors were selected to predict the upcoming winter haze days over the NCP (WHD_{NCP}). After cross validation, the root mean square error and explained variance of the MLR (GAM) prediction model was 3.39 (3.38) and 53% (54%), respectively. For the final predicted WHD_{NCP} , both of these models could capture the interannual and interdecadal trends and the extremums successfully. Independent prediction tests for 2014 and 2015 also confirmed the good predictive skill of the new schemes. The predicted bias of the MLR (GAM) prediction model in 2014 and 2015 was 0.09 (−0.07) and −3.33 (−1.01), respectively. Compared to the MLR model, the GAM model had a higher predictive skill in reproducing the rapid and continuous increase of WHD_{NCP} after 2010.

1. Introduction

In recent years, the North-Central North China Plain (NCP; 34–43°N, 114–120°E) has suffered from increasingly severe winter (December–February) haze pollution (Ding et al. 2014), particularly after persistent heavy fog and haze in January 2013 (Zhang et al. 2014; Zhao et al. 2014). After 2000, the combined effects of a rapid increase in total energy consumption

25 and the influence of climate change intensified the haze pollution in central North China (Wang et al. 2016). In conditions of heavy and slowly varying pollutant emissions, the fine particles in the atmosphere reach their saturation levels easily, and the climate conditions become another critical contributors of haze. Some new climatic studies ~~climatic findings have been~~ should be helpful for diagnosing seasonal predictors of winter haze days over the NCP (WHD_{NCP}). The East Asian winter monsoon (EAWM) has a significantly negative relationship with WHD_{NCP} (Yin et al. 2015a; Yin et al. 2015b; Li et al. 2015).
 30 By weakening EAWM circulations, negative Sea Surface Temperature (SST) anomalies over the subtropical western Pacific (~~SWP~~) could significantly intensify WHD_{NCP} (Yin et al. ~~2015~~2016). Furthermore, the decline of preceding autumn (~~September-November~~) Arctic Sea Ice (ASI) has led to favorable environments for haze, with high static stability and greatly intensified haze pollution in eastern China (Wang et al. 2015). Although recent studies on the changes in WHD_{NCP} and their associated mechanisms are new and still insufficient, they support the possibility of seasonal prediction.

35 The climate variables in East Asia showed obvious characteristics of tropospheric biennial oscillation (~~TBO~~), based on which, a new interannual increment approach was applied for short-term climate prediction (Wang et al. 2000; Wang et al. 2012). This new approach treated the year-to-year increment of a variable, i.e., the difference between the current and previous year (DY), as the predictand. Because the DY approach utilized the observed information from the previous year and the features of biennial oscillation~~TBO~~, the interannual variation and interdecadal trend could be captured well. In
 40 addition, the signals (i.e., variance) of the predictors and predictand were both amplified (Huang et al. ~~2015~~2014) and, thus, of benefit to improve the prediction skill. If the predictive objects (Y), e.g., haze days, were cross-influenced by socio-economic factors and climatic conditions, the predictand could be represented by $Y = YS + YC$, where YS and YC ~~are~~ were the slowly varying socio-economic and climatic components, respectively.

$$DY = Y_t - Y_{t-1} = (YS_t + YC_t) - (YS_{t-1} + YC_{t-1}) = (YS_t - YS_{t-1}) + (YC_t - YC_{t-1})$$

where the subscripts t and $t-1$ indicated d the current and previous years, respectively.

45 Commonly, the difference in pollutant emissions between current and previous year was very small, resulting in $(YS_t - YS_{t-1}) \approx 0$, so $DY \approx (YC_t - YC_{t-1})$. To some extent, the WHD_{NCP} DY reflected the fluctuations caused by climate variability~~change~~. After adding the predicted WHD_{NCP} DY to the observed WHD_{NCP} of the previous last-year, the interdecadal and socio-economic components were contained in the final prediction. In prior studies, the DY approach has

50 been used to explore the prediction of summer rainfall in China (Fan et al. 2008), heavy winter snow activity in Northeast China (Fan et al. 2013), summer Asian-Pacific Oscillation (Huang et al. 2014) and winter North Atlantic Oscillation (Tian et al. 2015). Furthermore, some variables cross-influenced by socio-economic and climatic factors were predicted successfully using the DY approach, e.g., rice production in Northeast China (Zhou et al. 2014) and the discoloration day for *Cotinus coggygria* leaves in Beijing (Yin et al. 2014). Considering the seriously negative impact of winter haze and the substantial need to predict WHD_{NCP}, we made it the goal of this study to apply the DY approach to the seasonal prediction of WHD_{NCP}.

55 The data and methods employed ~~are~~were introduced in section 2. Section 3 ~~described~~s the predictors and associated circulations. We ~~applied~~ied the DY approach to build the prediction models for WHD_{NCP} in section 4. In this section, the statistical models ~~are~~were built based on multiple linear regression (MLR) and generalized additive model (GAM). Then, leave-one-out cross-validation and independent tests ~~are~~were performed to assess the statistical schemes of WHD_{NCP} prediction.

60 2. Datasets and methods

Monthly atmospheric data, such as geopotential height (~~Z~~) and surface air temperature (~~TSAT~~), ~~are~~were derived from the National Centers for Environmental Prediction/National Center for Atmospheric Research (NCEP/NCAR) global reanalysis dataset with a horizontal resolution of 2.5°×2.5° from 1979 to 2016 (Kalnay et al. 1996). The monthly mean Extended Reconstructed SST datasets with a horizontal resolution of 2°×2° from 1979 to 2016 were obtained from the National
65 Oceanic and Atmospheric Administration (NOAA) (Smith et al. 2008). ASI extent was calculated from the ASI concentration data, downloaded from the Hadley Center with a horizontal resolution of 1°×1° from 1979 to 2016 (Rayner et al. 2003). The monthly gridded soil moisture data from 1979 to 2016 were downloaded from NOAA's Climate Prediction Center (~~CPC~~), with a horizontal resolution of 0.5°×0.5° (Huug et al. 2003). The monthly Antarctic Oscillation (AAO) indices from 1979 to 2016 were also obtained from the Climate Prediction Center~~CPC~~ (Mo et al. 2000).

70 China ground observations from 39 NCP stations, collected by the National Meteorological Information Center of China 4 times per day from 1979 to 2016, were used to reconstruct the climatic WHD data (Yin et al. ~~2015~~2016). Here,

haze ~~is-was~~ defined as visibility less than a certain threshold and relative humidity less than 90%. After excluding other weather phenomena affecting visibility, a day with haze at any time ~~is-was~~ defined as a haze day. Site WHD data were converted into grids after Cressman interpolation (Cressman, 1959), and then the WHD_{NCP} was computed as the mean value of the gridded data.

In this study, the statistical models were built based on MLR and GAM methods. The MLR approach, a model-driven method, ~~is-was~~ ultimately expressed as a linear combination of K predictors (x_i) that ~~can-could~~ generate the least error for prediction of \hat{y} (Wilks 2011). With coefficients β_i , intercept β_0 and residual ε , the MLR formula ~~can-could~~ be described as follows:

$$\hat{y} = \beta_0 + \sum_{i=1}^K \beta_i x_i + \varepsilon \quad (1)$$

The GAM approach ~~is-was~~ more advanced and was developed from MLR and the generalized linear model (GLM) (Hastie et al. 1990). This ~~data-driven~~ method ~~is-was~~ particularly effective at handling the complex non-linear and non-monotonous relationships between the predictand and the predictors, whose expressions ~~are-were~~ replaced by ~~unspecified~~ smooth functions (s). Similar to the generalized linear model GLM, the dependent variable in GAM ~~can-could~~ have different probability distributions, such as Gaussian, Poisson, and Binomial, any of which ~~could~~ be transferred by the link function (g). The GAM ~~is-was~~ data-driven rather than model-driven. The resulting fitted values ~~did~~ not come from an apriori model that was adopted by MLR and generalized linear model. The rationale behind fitting a nonparametric model ~~is-was~~ that the structure of data should be examined first to choose an appropriate smooth function for each predictor; ~~that is~~, the GAM ~~allowed~~ the data to determine the shape of the smooth function (Yee et al. 1991). The GAM ~~can-could~~ be written in the form:

$$g(\hat{y}) = \beta_0 + \sum_{i=1}^K \beta_i s(x_i) + \varepsilon \quad (2)$$

The normalized datasets from 1979 to 2013 were trained as the basic samples to fit the models, and those from 2014 to 2015 were treated as test data for independent prediction. Thereafter, the root mean standard error (RMSE), mean absolute error (MAE) and explained variance (~~EV~~) were calculated for evaluation by simple fitting and leave-one-out cross

validation.

3. The predictors and associated circulations

95 To choose the DY predictors, the correlated DY atmospheric circulations were identified, as shown in Figure 1. The positive phase of the East Atlantic/West Russia (EA/WR; [Barnston et al. 1987](#)) and [West Pacific-Japan \(WPJ; Barnston et al. 1987; Nitta 1987\)](#) patterns and the negative phase of the Eurasia (EU; Wallace et al. 1981) pattern were obvious, and we took the anti-cyclone circulation over North China as an intermediary [that leading led](#) to a more stable atmosphere [to analyze the associated physical process](#). The positive anomaly over the NCP could confine the particles within a thinner boundary layer by suppressing vertical movement and, [together with the cyclone; they could](#) induce an easterly to weaken the East Asian Jet Stream (EASJS), producing weaker cold air. Meanwhile, the water vapor transportation was also enhanced [by anomalous southeaster in the lower troposphere \(Figure omitted\)](#), creating favorable conditions for more WHD_{NCP} than in the previous year.

105 The pivotal local anti-cyclone over the NCP was the most important contributor; we therefore speculated that pre-autumn [FSSAT](#) DY around the NCP should be effective to impact WHD_{NCP} DY. There were significantly negative correlations between WHD_{NCP} DY and pre-autumn [FSSAT](#) DY from the Japan Sea to the Stanovoy Range (35–65°N, 130–140°E), the area mean of which was selected as predictor x_1 (Figure 2). The correlation coefficient (CC) between WHD_{NCP} DY and predictor x_1 was -0.47 , exceeding [thea](#) 99% confidence level. The [features of negative EU and positive WP pattern could be identified clearly and the anomalous cyclone over South China and South China Sea was significant in the](#) circulations associated with predictor x_1 ($\times -1$) ~~presented obvious features of the negative EU and positive PJ patterns (Figure 3), similar to those shown in Figure 1.~~ [Although the associated land-air interaction, especially in the DY field, was complicate and still unclear, according to the analysis of Figure 1, the horizontal and vertical diffusion of pollutant particles would be restricted efficiently.](#)

115 The pre-autumn SST anomalies [and their associated winter SST](#) of the Pacific could influence WHD_{NCP} significantly via the air-sea interaction (Yin et al. [2015e2016](#)). Figure 4 shows the CC between WHD_{NCP} DY and pre-autumn [Pacific](#) SST

DY. The most significant CC ~~was~~ distributed around the Alaska Gulf (36–56°N, 130–170°W), and the area-averaged SST DY here was defined as predictor x_2 , whose CC with WHD_{NCP} DY was 0.47, above the 99% confidence level. [Chen et al. \(Chen et al. 2015\)](#) found that the severe winter haze events in the North China were closely related with the weaker and northward EAJ. The positive SST DY around the Alaska Gulf [could induce obviously anomalous cyclone over eastern China and the adjacent ocean, and the stimulated easterly weakened the core of EAJ. Furthermore, there was significantly anomalous southerly at the high latitude that restricted the cold activities from their source region and intensified the haze pollution over NCP \(Figure 5\).](#) ~~closely correlated with the atmospheric teleconnection patterns, i.e., the positive phases of the EA/WR and PJ and the negative EU patterns intensified haze pollution over the NCP (Figure 5).~~

Prior studies have documented that the triple SST pattern was a dominant mode of the northern Atlantic SST in autumn [\(Czaja et al. 1999\)](#). When the pre-autumn SST anomalies were distributed in a “+–+” pattern from south to north, the subsequent EAWM was stronger, and the surface temperature of North China was lower (Shi 2009). [Xiao et al. \(Xiao et al. 2015\)](#) proved the SST anomalies over the North Atlantic from summer to the following winter exhibit a significant relationship with winter haze days on both decadal and interannual timescale. Similarly, the CC between WHD_{NCP} DY and pre-autumn SST DY of the Atlantic was distributed in a “–+–” pattern (Figure 6). The area-averaged SST DY of the northern center was defined as predictor x_3 , whose CC with WHD_{NCP} DY was –0.50, passing the 99% confidence test. The most obvious DY atmospheric circulations related with predictor x_3 ($\times -1$) were the ~~negative EU~~ [positive WP](#) pattern, whose south center linked with a subtropical high (Figure 7). The continental high and marine ~~low were~~ [low was](#) both ~~weaker~~ [weakened by the anomalous geopotential height form the lower to middle layer that,](#) leading to weaker EAWM and weaker cold air. The pressure gradient over the east coast of China also resulted in ~~asignificant~~ [southerly anomalies](#), indicating smaller surface wind and more moisture and resulting in more WHD_{NCP}.

ASI decreased dramatically with significant variance and was a significant contributor influencing WHD_{NCP} in eastern China (Wang et al. 2015; Wang et al. 2016). The CC between pre-autumn ASI DY and WHD_{NCP} DY was calculated (Figure 8) and was significantly positive around Beaufort Sea (73–78°N, 130–165°W). The area-averaged ASI extent DY of Beaufort Sea was selected as the fourth predictor (x_4), and its CC with WHD_{NCP} DY was 0.37, above a 95% confidence level. A positive center [of geopotential height at 500 hPa](#) was located over the Central Siberian and Mongolia Plateau, and negative

centers were distributed zonally from southern China to the subtropical Pacific (Figure 9). Thus, the EAJS was weakened by the induced easterly and shifted northward that illustrated less cold activities over NCP (Yang et al. 2002) and generated more haze days.

145 Following SST, the Ssoil moisture was ~~is~~ another important factor for seasonal prediction, ~~but only after SST~~ (Guo et al. 2007). The WHD_{NCP} was closely correlated with the moisture conditions due to the hygroscopicity of the atmospheric particles (Yin et al. 2015a). Thus, ~~the~~ the questions with respect to soil moisture were whether pre-summer or autumn soil moisture would be effective for seasonal prediction of WHD_{NCP} DY. The area-averaged pre-autumn soil moisture DY of the Bohai rim (35–42°N, 117–127°E), defined as predictor x_5 , showed a significantly negative correlation with WHD_{NCP} DY, i.e., the CC was -0.59, exceeding a 99% confidence test (Figure 10). The CC between predictor x_5 and Z500 (geopotential height at 500 hPa) was distributed in a similar way as in Figure 1. The positive EA/WR and ~~PJWP~~ phases and the negative EU phase was obvious and led to more WHD_{NCP} than in the previous year (Figure 11). Being specific to local circulations, the cyclone over South China and the anti-cyclone over NCP and West Pacific stimulated significant southeaster between them (Figure omitted) that transported more moisture but decelerated the surface wind in the NCP. As shown in Figure 12, the pre-summer soil moisture DY in the east of Mongolia (48–52°N, 115–125°E) also had a close relationship with WHD_{NCP} and with WHD_{NCP} DY. The area-averaged soil moisture DY in the east of Mongolia was defined as predictor x_6 , whose CC with WHD_{NCP} DY was 0.41, above a 95% confidence level. The negative EU pattern could be recognized from the associated atmospheric circulation with predictor x_6 (Figure 13). The anomalous geopotential height was distributed zonally at high latitude indicating that the meridional circulations that transported cold air were weak. The positive high over NCP could confine the vertical motion and the vertical diffusion of atmospheric particles, and which
160 intensified the haze pollution over the NCP.

Recently, some studies documented that Antarctic Oscillation (AAO) could affect the East Asian climate through cross-equatorial flow, e.g., the Somali jet (Fan et al. 2004; Fan et al. 2006; Fan et al. 2007a; Fan et al. 2007b). After the late-1990s, global sea level pressure and geopotential height at Z300 hPa in boreal January were characterized by the concurrence of the Aleutian low and the negative phase of the AAO (Li et al. 2014). We investigated the relationship
165 between WHD_{NCP} DY and geopotential height at Z850 hPa in the Southern Hemisphere and found that the distribution was

170 remarkably similar to that of the negative phase of AAO (Figure 14). Furthermore, the CC between the September–October AAO DY and WHD_{NCP} DY was -0.54 , exceeding a 99% confidence test. As shown in Figure 15, the positive phases of the EA/WR and PJWP patterns were closely correlated with the negative phase of Sep–Oct AAO and were responsible for more WHD_{NCP} than in the previous year. [The anomalous anti-cyclone over NCP and adjacent ocean not only led to stable atmosphere but also resulted in small wind and high humidity](#). Hence, the September–October mean AAO index was selected as the last predictor (x_7) to forecast the interannual increment of WHD_{NCP}.

4. The prediction models and validations

175 In total, seven DY predictors ($x_1, x_2, \dots, \text{and } x_7$) were chosen to build the seasonal prediction model (SPM) for WHD_{NCP} DY (Table 2). Among the predictors were 21 types of pair combinations, of which only 5 pairs presented significant linear correlation (Figure 16). Thus, the multicollinearity would not be a problem when modeling with the MLR approach. Although the linear correlation between the predictand and each predictor was significant, the non-linear interaction would also affect the WHD_{NCP} and should be taken into account. In this section, seasonal prediction models were established using MLR (SPM_{MLR}) and GAM (SPM_{GAM}) and validated in detail.

180 The WHD_{NCP} DY showed obvious features of biennial oscillation (Figure 1716), illustrating the DY approach was suitable for its prediction. The SPM_{MLR} of WHD_{NCP} DY was as follows: $\text{DY} \times 10 = -2.774x_1 + 2.582x_2 - 1.631x_3 + 2.528x_4 - 2.229x_5 + 2.555x_6 - 1.812x_7$. After leave-one-out cross validation, the RMSE_{CV} of SPM_{MLR} was 3.39 days, and the CC between fitted and observed WHD_{NCP} DY was 0.73, accounting for 53% of the total variance (Table 2). The percentage of same sign (PSS; i.e., same sign means the mathematical sign of the fitted and observed WHD_{NCP} DY was the same) was 79.4%. The SPM_{MLR} showed good ability to predict the negative and least-minimum WHD_{NCP} DY but did not adequately capture the continuous positive value after 2011 (Figure 176a). The fitted WHD_{NCP} DY from 2011 to 2013 varied similarly to that before 2010 and did not reflect the rapid rising trend after 2010. As an independent prediction test, the predicted bias, i.e., the predicted value minus the measurement, in 2014 was 0.09, illustrating good performance, but the bias in 2015 was larger, i.e., -3.33 .

We also applied the GAM approach to build a prediction model that would contain the non-linear relationship with

190 smooth functions. The SPM_{GAM} of WHD_{NCP} DY was as follows: $DY \times 10 = -2.164s(x_1) + 2.036s(x_2) - 1.721x_3 + 2.588s(x_4) - 2.157s(x_5) + 2.187x_6 - 2.506x_7$. During the simple fitting, the SPM_{GAM} performed very well. The RMSE was 1.56 days, and the CC between the fitted and observed WHD_{NCP} DY was 0.95. The SPM_{GAM} could fit the minimum (in 2003) and maximum (in 2013), and show the trend well, indicating an advantage to processing ~~ing~~ the non-linear relationship. After cross validation, the performance of SPM_{GAM} decreased dramatically, meaning that its stability was worse than that of SPM_{MLR} . The $RMSE_{CV}$ of SPM_{GAM} was 3.38 days and the CC between fitted and observed WHD_{NCP} DY was 0.74, 195 accounting for 54% of the total variance (Table 2). The [percentage of same signPSS](#) of SPM_{GAM} results was 73.5%, which ~~was~~ close to the result from SPM_{MLR} . The SPM_{GAM} also showed good ability to predict the negative and minimum WHD_{NCP} DY and better performance to fit the maximum in 2013 (Figure 176b). The predicted bias in 2014 and 2015 was -0.07 and -1.01 , [and the results that were slightly](#) better than those from SPM_{MLR} . The CC between the bias of SPM_{MLR} and SPM_{GAM} 200 from 1980 to 2013 was 0.83, above a 99.99% confidence level. If the SPM_{MLR} performed well in some years, the SPM_{GAM} also showed good ability in these years, and *vice versa*. We speculated that the reason was that some useful factors were not diagnosed and included here.

After adding the predicted WHD_{NCP} DY to the observed information in the previous year, the predicted WHD_{NCP} in the current year was obtained. For example, the predicted WHD_{NCP} DY in 2012 was added to the measured WHD_{NCP} in 2011, 205 and the result was the final predicted WHD_{NCP} in 2012. In Figure 178, the simulated ~~ive~~ WHD_{NCP} anomaly was fitted by cross-validation from 1980 to 2013 and predicted in 2014 and 2015. For SPM_{MLR} and SPM_{GAM} , the CC between the original (detrended) observed and simulative WHD_{NCP} was 0.89 (0.87) and 0.90 (0.88), respectively. Both of these prediction models could capture the interannual and interdecadal trend and the extremums. The [percentage of same signPSS](#) of the anomalies from the two models was 100%, meaning these two models could predict the sign of WHD_{NCP} anomaly successfully. The 210 SPM_{GAM} could simulate the abrupt rising trend in 2010 better than SPM_{MLR} , which was important for the prediction of recent years.

5. Conclusions and Discussions

In this paper, we treated the WHD_{NCP} DY as the predictand and built two prediction models using the MLR and GAM approach. In the DY atmospheric circulation, the positive phases of the EA/WR and [PJWP](#) patterns and the negative phase of the EU pattern intensified the haze pollution by inducing positive anomalies over the NCP [and Japan Sea](#). Finally, seven leading predictors were selected and ~~are~~ were listed in Table 2.

After cross validation, the $RMSE_{CV}$ and explained variance of SPM_{MLR} (SPM_{GAM}) was 3.39 (3.38) and 53% (54%). The [percentage of same signPSS](#) of these two prediction models was also similar, i.e., more than 73%. The WHD_{NCP} DY increased rapidly and persistently after 2010, and the SPM_{GAM} could capture this trend better. For the final predicted WHD_{NCP} , both of these [two](#) prediction models could capture the interannual and interdecadal trends and the extremums. The [percentage of same signPSS](#) of the anomalies from two models was 100%, and the SPM_{GAM} simulated the abrupt increase in 2010 better than SPM_{MLR} . The predicted bias of SPM_{MLR} (SPM_{GAM}) in 2014 and 2015 was 0.09 (−0.07) and −3.33 (−1.01), respectively. Both of these models performed well in [the](#) independent tests, but the biases of SPM_{GAM} were [slightly](#) smaller. [The consistence of these two models might indicate that, after including plentiful predictors, the linear relationship dominated the \$WHD_{NCP}\$ DY prediction. Actually, the studies about the associated physical mechanism, i.e., how the external forcings influenced haze pollutions, were new and still insufficient. In this paper, the underlying physical process was presented mostly from the way that the associated circulations impacted the \$WHD_{NCP}\$ DY. Thus, the physical mechanism that the external forcings stimulated such associated circulations needed further studies.](#)

Although these two statistical models performed well during most of the past 3 decades and could predict the WHD_{NCP} in 2014 and 2015 with small biases, they showed disadvantages when simulating the rapid rising trend after 2010. [The large abrupt change was a common challenge to the statistical models, including the DY approach, so the numerical model should be introduced into the prediction of haze pollution. At the same time, if the \$SPM_{MLR}\$ performed well in some years, the \$SPM_{GAM}\$ also showed good ability in these years, and *vice versa*. One possible reason could be that some useful factors, \[most notably the human activities\]\(#\), were not ~~diagnosed and~~ included here. \[There was no doubt that the human activities, especial the energy consumption, was the first driver for the increasing of haze pollution.\]\(#\) In this paper, we \[simply\]\(#\) assumed that the](#)

240 difference in pollutant emissions between current and previous years was very small and that the socio-economic component of WHD_{NCP} varied slowly. [This assumption could support the seasonal prediction of haze days in most of the years, but still was a compromise. Another possible reason might be that](#) in certain years, [especially the recent years](#), this pollutant emission proportion varied rapidly [that needed to be taken into account. The preceding autumn energy consumption should be a good choice, but difficult to be measured, and its DY could be introduced into the developed models directly to improve the predictive skill.](#)

Acknowledgement

245 This research was supported by the National Natural Science Foundation of China (Grants: 41421004 and 41210007) and CAS-PKU Partnership Program.

References

Barnston A G, Livezey R E. 1987. Classification, seasonality and persistence of low frequency atmospheric circulation patterns. *Mon. Wea. Rev.*, 115: 1083–1126.

Cressman G. 1959. An operational objective analysis system. *Mon. Wea. Rev.* 87(10): 367–374.

250 [Czaja A, Frankignoul C. 1999. Influence of the North Atlantic SST on the atmospheric circulation. *Geophys. Res. Lett.* 26: 2969–2972](#)

Ding Y H, Liu Y J. 2014. Analysis of long-term variations of fog and haze in China in recent 50 years and their relations with atmospheric humidity. *Sci. China Ser. D: Earth Sci.* 57: 36–46(in Chinese).

255 Fan K, Tian B Q. 2013. Prediction of wintertime heavy snow activity in Northeast China. *Chinese Science Bulletin.* 58(12): 1420–1426.

Fan K, Wang H J, Choi Y J. 2008. A physically-based statistical forecast model for the middle-lower reaches of Yangtze River Valley summer rainfall. *Chinese Science Bulletin.* 54(4): 602–609

Fan K, Wang H J. 2006. Studies of the relationship between Southern Hemispheric atmospheric circulation and climate over

- East Asia. Chinese journal of Atmosphere Sciences, 20: 402–412
- 260 Fan K, Wang H J. 2007a. Simulation on the AAO anomaly and its influence on the Northern Hemispheric circulation in boreal winter and spring. Chinese Journal of Geophysics, 50: 397–403
- Fan K, Wang H J. 2007b. Dust storms in North China in 2002: A case study of the low frequency oscillation. Adv. Atmos. Sci., 24: 15–23
- Fan K, Wang H J. 2004. Antarctic oscillation and the dust weather frequency in North China. Geophys. Res. Lett., 31,
265 L10201, doi:10.1029/2004GL019465
- Guo W D, Ma Z G, Wang H J. 2007. Soil moisture—an important factor of seasonal precipitation prediction and its application. Climatic and Environmental Research, 12 (1): 20–28
- Hastie T J, Tibshirani R J. 1990. Generalized Additive Models. Chapman & Hall, London.
- [Huang Y Y, Wang H J, Fan K. 2014. Improving the Prediction of the Summer Asian-Pacific Oscillation Using the Interannual Increment Approach. J. Climate, 27: 8126–8134, doi: http://dx.doi.org/10.1175/JCLI-D-14-00209.1](#)
- [Huang R H. 1992. The East Asia/Pacific pattern teleconnection of summer circulation and climate anomaly in East Asia. J. Meteor. Res., 6: 25–37](#)
- Huug D, Huang J, Fan Y. 2003. Performance and Analysis of the constructed analogue method applied to US soil moisture applied over 1981–2001. J. of Geophysical Research. 108: 1–16.
- 275 Kalnay E, Kanamitsu M, Kistler R, Collins W, Deaven D, Gandin L, Iredell M, Saha S, White G, Woollen J, Zhu Y, Leetmaa A, Reynolds R, Chelliah M, Ebisuzaki W, Higgins W, Janowiak J, Mo KC, Ropelewski C, Wang J, Jenne R, Joseph D. 1996. The NCEP/NCAR 40-year reanalysis project. Bull. Am. Meteorol. Soc., 77: 437–471, doi: 10.1175/1520-0477(1996)077<0437: TNYRP>2.0.CO; 2.
- Li F, Wang H J, Gao Y Q. 2015. Modulation of Aleutian low and Antarctic oscillation co-variability by ENSO. Climate
280 Dynamics, 44: 1245–1256
- Li Q, Zhang R H, Wang Y. 2015. Interannual variation of the winter-time fog–haze days across central and eastern China and its relation with East Asian winter monsoon. Int. J. Climatol. 36 (1): 346–354, doi: 10.1002/joc.4350.
- Mo K. C. 2000. Relationships between Low-Frequency Variability in the Southern Hemisphere and Sea Surface Temperature

Anomalies. *J. Climate*, 13, 3599–3610

285 [Nitta T. 1987. Convective activities in the tropical western Pacific and their impact on the Northern Hemisphere summer circulation. *J. Meteor. Soc. Japan*, 64: 373–390.](#)

Rayner N A, Parker D E, Horton E B, Folland C K, Alexander L V, Rowell D P, Kent E C, Kaplan A. 2003. Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.* 108 (14): 4407 doi: 10.1029/2002JD002670

290 Shi X L. 2014. The impact of mid-high latitude oceans on climate change over north and northeast China. Qing Dao: Ocean University of China, 1 pp.

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

295 Tian B Q, Fan K. 2015. A skillful prediction model for winter NAO based on Atlantic sea surface temperature and Eurasian snow cover, *Weather and Forecasting*. 30: 197–204

Wallace J M, Gutzler D S. 1981. Teleconnection in the geopotential height field during the Northern Hemisphere winter. *Mon. Wea. Rev.*, 109, 784–812, doi:10.1175/1520-0493(1981)109,0784:TITGHF.2.0.CO;2.

Wang H J, Chen H P, Liu J P. 2015. Arctic sea ice decline intensified haze pollution in eastern China. *Atmos. Oceanic Sci. Lett.*, 8 (1): 1–9

300 Wang H J, Chen H P. 2016. Understanding the recent trend of haze pollution in eastern China: role of climate change. *Atmos. Chem. Phys.*, 16: 4205–4211

Wang H J, Fan K, Lang X M, Sun J Q, Chen L J. 2012. Advances in climate prediction theory and technique of China, Beijing: China Meteorological Press. 120–140 (in Chinese).

305 Wang H J, Zhou G Q, Zhao Y. 2000. An effective method for correcting the seasonal-interannual prediction of summer climate anomaly. *Adv. Atmos. Sci.*, 17 (2), 234–240

Wilks D S. 2011. *Statistical methods in the atmospheric sciences*. Academic press, 1 pp.

[Yang S, Lau K M, Kim K M. 2002. Variations of the East Asian Jet Stream and Asian–Pacific–American Winter Climate Anomalies. *Journal of Climate*, 15\(3\): 306–325](#)

- 310 [Yee T, Mitchell N. 1991. Generalized additive models in plant ecology, *Journal of Vegetation Science*, 2\(5\): 587—602](#)
- [Xiao D, Li Y, Fan S J, Zhang R H, Sun J R, Wang Y. 2015. Plausible influence of Atlantic Ocean SST anomalies on winter haze in China. *Theor. Appl. Climatol.*, 122: 249—257](#)
- Yin Z C, Wang H J, Guo W L. 2015a. Climatic change features of fog and haze in winter over North China and Huang-Huai Area. *SCIENCE CHINA Earth Sciences*, 58(8): 1370–1376.
- 315 [Yin Z C, Wang H J, Yuan D M. 2015b. Interdecadal increase of haze in winter over North China and the Huang-huai area and the weakening of the East Asia winter monsoon. *Chin. Sci. Bull.* 60\(15\): 1395–1400 \(in Chinese\).](#)
- Yin Z C, Wang H J. ~~2015~~2016. The relationship between the subtropical Western Pacific SST and haze over North-Central North China Plain. *International Journal of Climatology*, [36\(10\): 3479—3491](#), DOI: 10.1002/joc.4570
- Yin Z C, Yuan D M, Ding D P, Xie Z. 2014. Statistical prediction based on meteorology of *Cotinus Coggygia* leaves discoloration-day in the Fragrant hill, *Meteorological Monthly*, 40 (2): 229–233 (in Chinese)
- 320 ~~Yin Z C, Wang H J, Yuan D M. 2015b. Interdecadal increase of haze in winter over North China and the Huang-huai area and the weakening of the East Asia winter monsoon. *Chin. Sci. Bull.* 60(15): 1395–1400 (in Chinese).~~
- Zhang R H, Li Q, Zhang R N. 2014. Meteorological conditions for the persistent severe fog and haze event over eastern China in January 2013. *Science China: Earth Sciences*, 57: 26–35. doi: 10.1360/972013-150
- Zhao N, Yin Z C, Wu F. 2014. Characteristics of persistent fog and haze process and its forming reason in Beijing. *J. Meteorol. Environ.* 30(5): 15–20 (in Chinese).
- 325 Zhou M Z, Wang H J. 2014. Late winter sea ice in Bering Sea: predictor for maize and rice production in Northeast China, *Journal of applied Meteorology and Climatology*, 53: 1183–1192

Table and Figure Captions:

330 **Table 1:** The RMSE, MAE, CC and [explained variance \(EV\)](#)~~EV~~ of MLR and GAM models, and predicted bias for 2014 and 2015. The subscripts “S” and “CV” indicated simple and cross-validation fitting.

Table 2. The predictors and their meaning. “CC” indicated the correlation coefficient between predictor and WHD_{NCP} DY

from 1980 to 2013.

Figure 1. The ~~correlation coefficient~~ (CC) between WHD_{NCP} DY and ~~geopotential height at 500 hPa (Z500)~~ Z500-DY in winter from 1980 to 2013. The white curves indicate that the CC exceeded the 95% confidence level. A and C represent anti-cyclone and cyclone, respectively.

Figure 2. The CC between WHD_{NCP} DY and ~~TSSAT~~ DY in autumn from 1980 to 2013. The shades indicate that the CC exceeded the 95% confidence level, and the rectangle represents the selected region (35–65°N, 130–140°E) of predictor x_1 .

Figure 3. The CC between predictor x_1 ($\times-1$) and Z500 DY in winter from 1980 to 2013. The white curves indicate that the CC exceeded the 95% confidence level. A and C represent anti-cyclone and cyclone, respectively.

Figure 4. The CC between WHD_{NCP} DY and Pacific SST DY in autumn from 1980 to 2013. The shades indicate that the CC exceeded the 95% confidence level, and the rectangle represents the selected region (36–56°N, 130–170°W) of predictor x_2 .

Figure 5. The CC between predictor x_2 and ~~Z500-wind vector DY at 200 hPa~~ DY in winter from 1980 to 2013. The ~~white curves shade~~ indicates that the CC ~~between the zonal wind DY and x_2~~ exceeded the 95% confidence level. ~~A and C represent anti-cyclone and cyclone, respectively.~~

Figure 6. The CC between WHD_{NCP} DY and Atlantic SST DY in autumn from 1980 to 2013. The shades indicate that the CC exceeded the 95% confidence level, and the rectangle represents the selected region (50–70°N, 30–65°W) of predictor x_3 .

Figure 7. The CC between predictor x_3 ($\times-1$) and Z500 DY (~~shade~~)/850 hPa wind DY (~~arrow~~) in winter from 1980 to 2013. The ~~dots white curves~~ indicate that the CC ~~with meridional wind~~ exceeded the 95% confidence level. A and C represent anti-cyclone and cyclone, respectively.

Figure 8. The CC between WHD_{NCP} DY and ASI DY in autumn from 1980 to 2013. The shades indicate that the CC exceeded the 95% confidence level, and the rectangle represents the selected region (73–78°N, 130–165°W) of predictor x_4 .

Figure 9. The CC between predictor x_4 and Z500 DY in winter from 1980 to 2013. The white curves indicate that the CC exceeded the 95% confidence level.

Figure 10. The CC between WHD_{NCP} DY and ~~Ssoil moisture~~ M DY in autumn from 1980 to 2013. The shades indicate that the CC exceeded the 95% confidence level, and the rectangle represents the selected region (35–42°N, 117–127°E) of predictor x_5 .

Figure 11. The CC between predictor x_5 ($\times-1$) and Z500 DY in winter from 1980 to 2013. The white curves indicate that the CC exceeded the 95% confidence level. A and C represent anti-cyclone and cyclone, respectively.

Figure 12. The CC between WHD_{NCP} DY and [Soil_Moisture](#) DY in summer from 1980 to 2013. The shades indicate that the CC exceeded the 95% confidence level, and the rectangle represents the selected region (48–52°N, 115–125°E) of predictor x_6 .

Figure 13. The CC between predictor x_6 and Z500 DY in winter from 1980 to 2013. The white curves indicate that the CC exceeded the 95% confidence level. A and C represent anti-cyclone and cyclone, respectively.

Figure 14. The CC between WHD_{NCP} DY and [September–October](#) Z850 DY from 1980 to 2013. The white curves indicate that the CC exceeded the 95% confidence level.

Figure 15. The CC between predictor x_7 ($\times -1$) and Z500 DY in winter from 1980 to 2013. The white curves indicate that the CC exceeded the 95% confidence level.

Figure 1716. The temporal variation of measured (black) WHD_{NCP} DY, MLR (red, a) and GAM (red, b) cross-validation fitted WHD_{NCP} DY from 1980 to 2013. The results for 2014 and 2015 represent the measured (black square) and predicted (red hollow circle) WHD_{NCP} DY.

Figure 1817. The temporal variation of measured (black) WHD_{NCP} anomaly from 1980 to 2015, MLR (blue) and GAM (red) simulative WHD_{NCP} anomaly, which was composed of cross fitted series from 1980 to 2013 and predicted values in 2014 and 2015.

Table 1: The RMSE, MAE, CC and [explained variance \(EV\)](#) of MLR and GAM models, and predicted bias for 2014 and 2015. The subscripts “S” and “CV” indicated simple and cross-validation fitting.

	MLR _s	MLR _{CV}	GAM _s	GAM _{CV}
RMSE	2.39	3.39	1.56	3.38
MAE	1.75	2.37	1.10	2.58
CC	0.87	0.72	0.95	0.74

EV	76%	53%	90%	54%
Bias₁₄		0.09		-0.07
Bias₁₅		-3.33		-1.01

385

390

Table 2. The predictors and their meaning. “CC” indicated the correlation coefficient between predictor and WHD_{NCP}

395 **DY from 1980 to 2013.**

Predictors	Meaning	CC
x_1	pre-autumn TSSAT DY from Japan Sea to Stanovoy Range	-0.47
x_2	pre-autumn SST DY around Alaska Gulf	0.47
x_3	pre-autumn SST DY to the south of Greenland	-0.50
x_4	pre-autumn ASI extent DY of Beaufort Sea	0.37
x_5	pre-autumn soil moisture DY of the Bohai rim	-0.59
x_6	pre-summer soil moisture DY in the east of Mongolia	0.41
x_7	September–October AAO index DY	-0.54

400

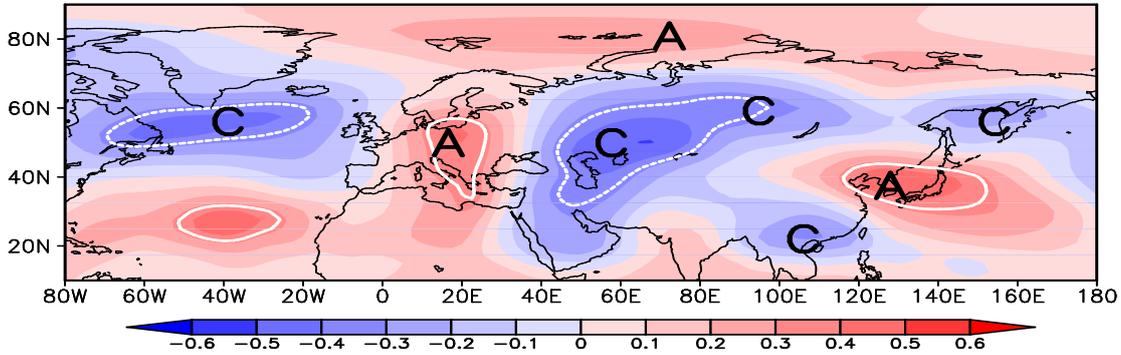


Figure 1. The correlation coefficient (CC) between $WHD_{NCP} DY$ and $geopotential\ height\ at\ 500\ hPa\ (Z500)Z500_{DY}$ in winter from 1980 to 2013. The white curves indicate that the CC exceeded the 95% confidence level. A and C represent anti-cyclone and cyclone, respectively.

405

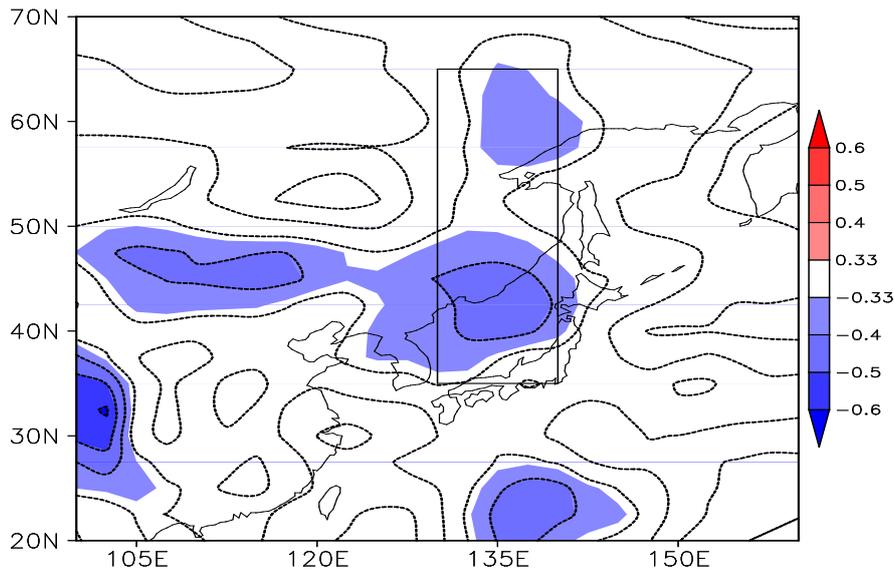
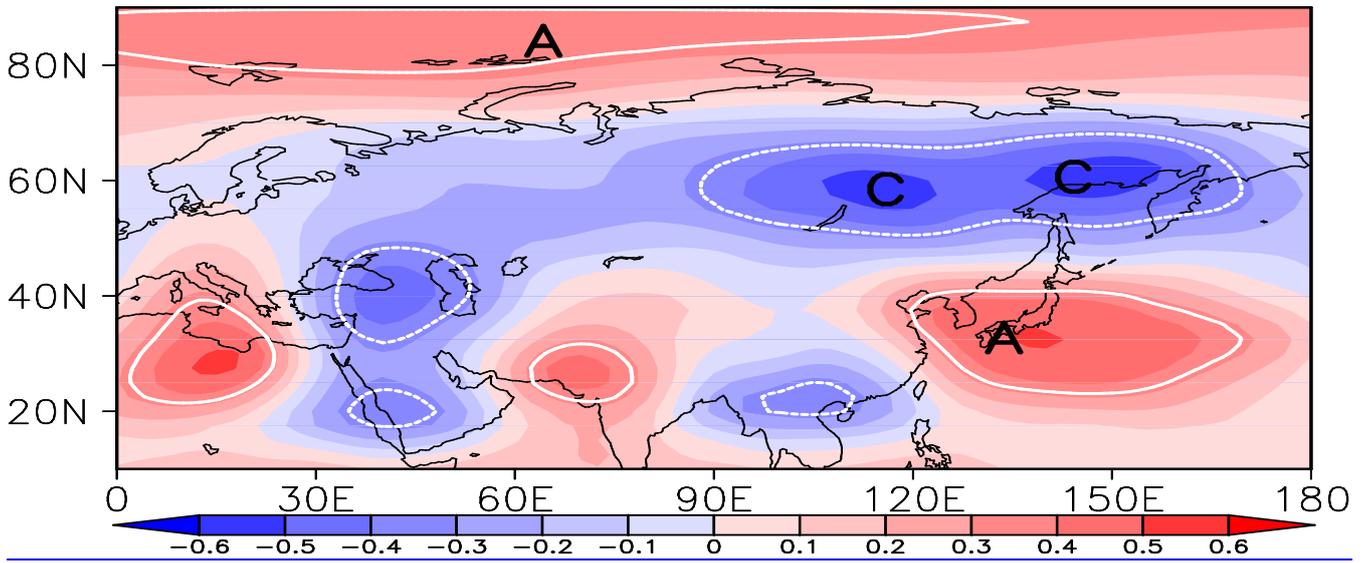
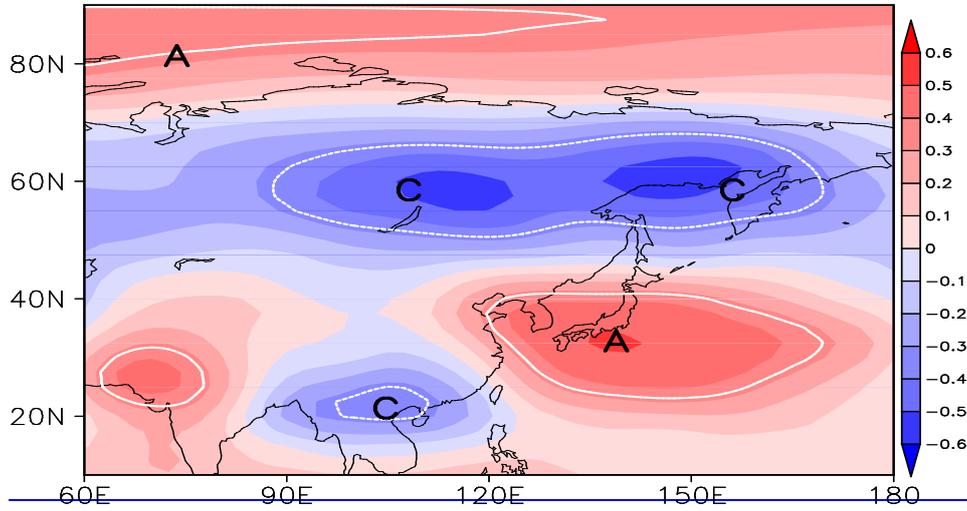


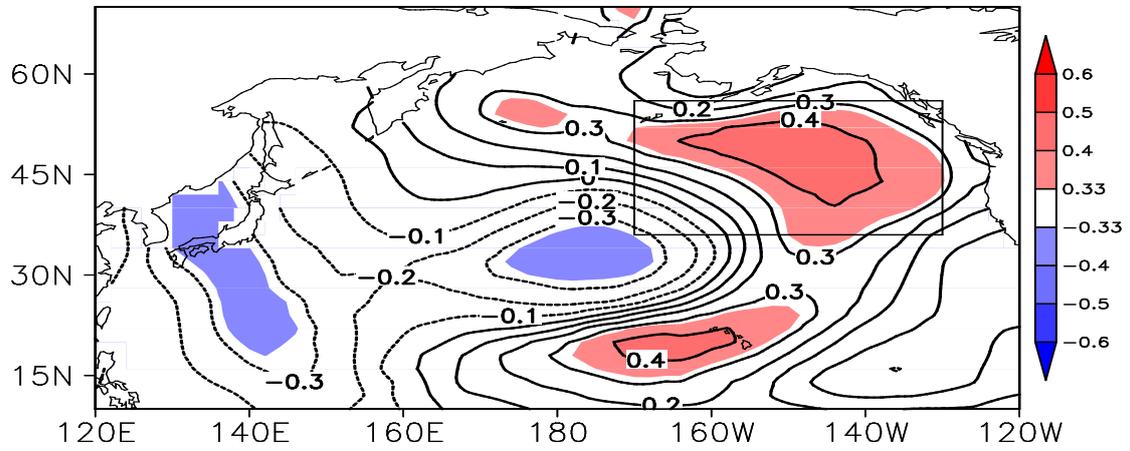
Figure 2. The CC between $WHD_{NCP} DY$ and $TSSAT_{DY}$ in autumn from 1980 to 2013. The shades indicate that the CC exceeded the 95% confidence level, and the rectangle represents the selected region ($35\text{--}65^\circ N$, $130\text{--}140^\circ E$) of predictor x_1 .

410

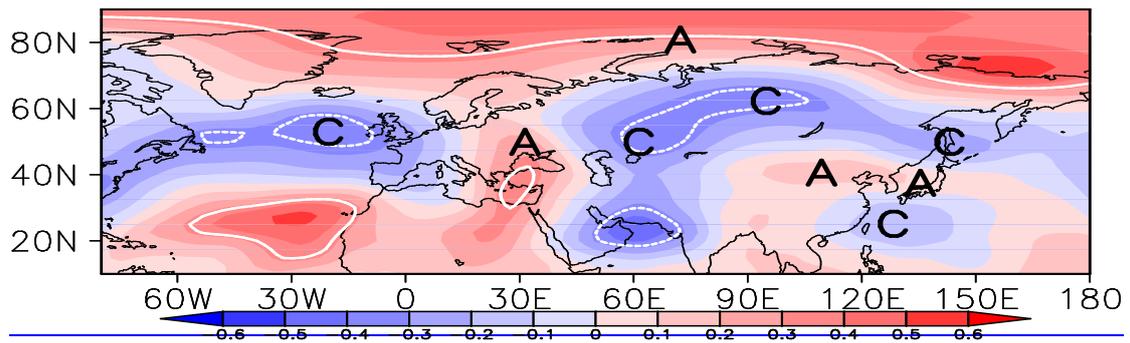


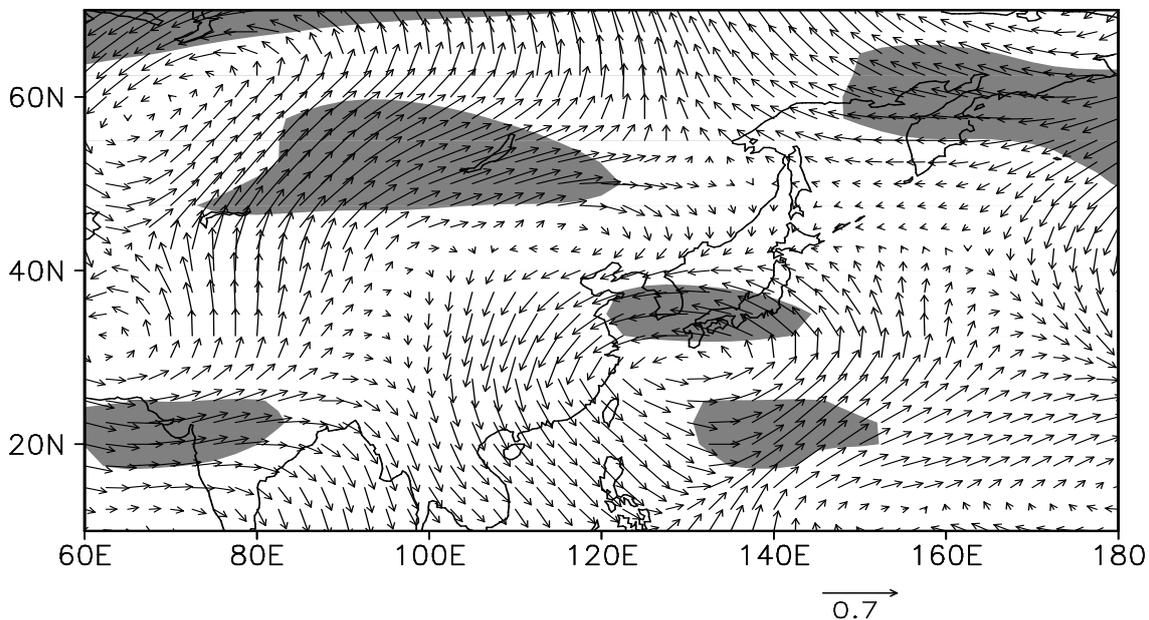
415

Figure 3. The CC between predictor x_1 ($\times -1$) and Z500 DY in winter from 1980 to 2013. The white curves indicate that the CC exceeded the 95% confidence level. A and C represent anti-cyclone and cyclone, respectively.

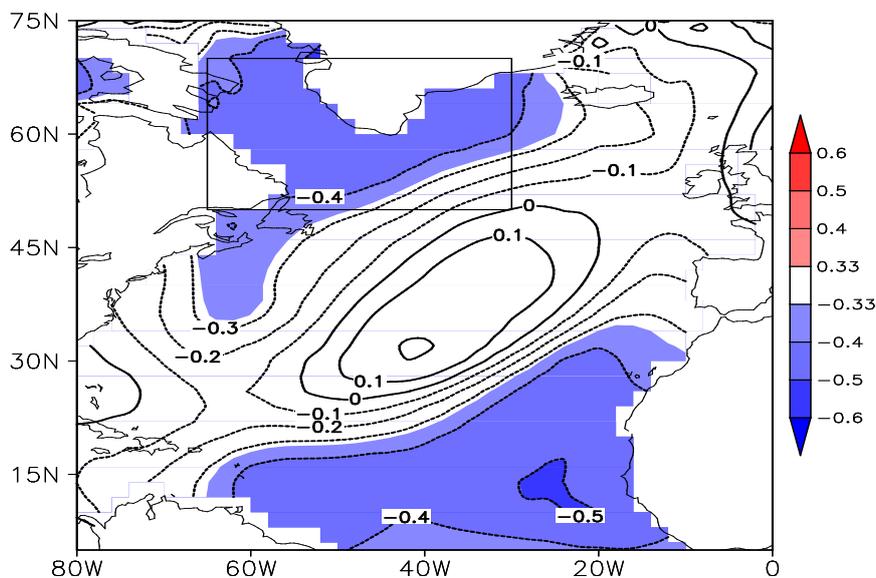


420 **Figure 4.** The CC between WHD_{NCP} DY and Pacific SST DY in autumn from 1980 to 2013. The shades indicate that the CC exceeded the 95% confidence level, and the rectangle represents the selected region ($36\text{--}56^{\circ}\text{N}$, $130\text{--}170^{\circ}\text{W}$) of predictor x_2 .





425 **Figure 5.** The CC between predictor x_2 and $Z500\text{-wind vector } DY \text{ at } 200 \text{ hPa } DY$ in winter from 1980 to 2013. The white curves shade indicates that the CC between the zonal wind DY and x_2 exceeded the 95% confidence level. **A** and **C** represent anti-cyclone and cyclone, respectively.



430 **Figure 6.** The CC between $WHD_{NCP} \text{ } DY$ and Atlantic SST DY in autumn from 1980 to 2013. The shades indicate that the CC exceeded the 95% confidence level, and the rectangle represents the selected region ($50\text{--}70^\circ N$, $30\text{--}65^\circ W$) of predictor x_3 .

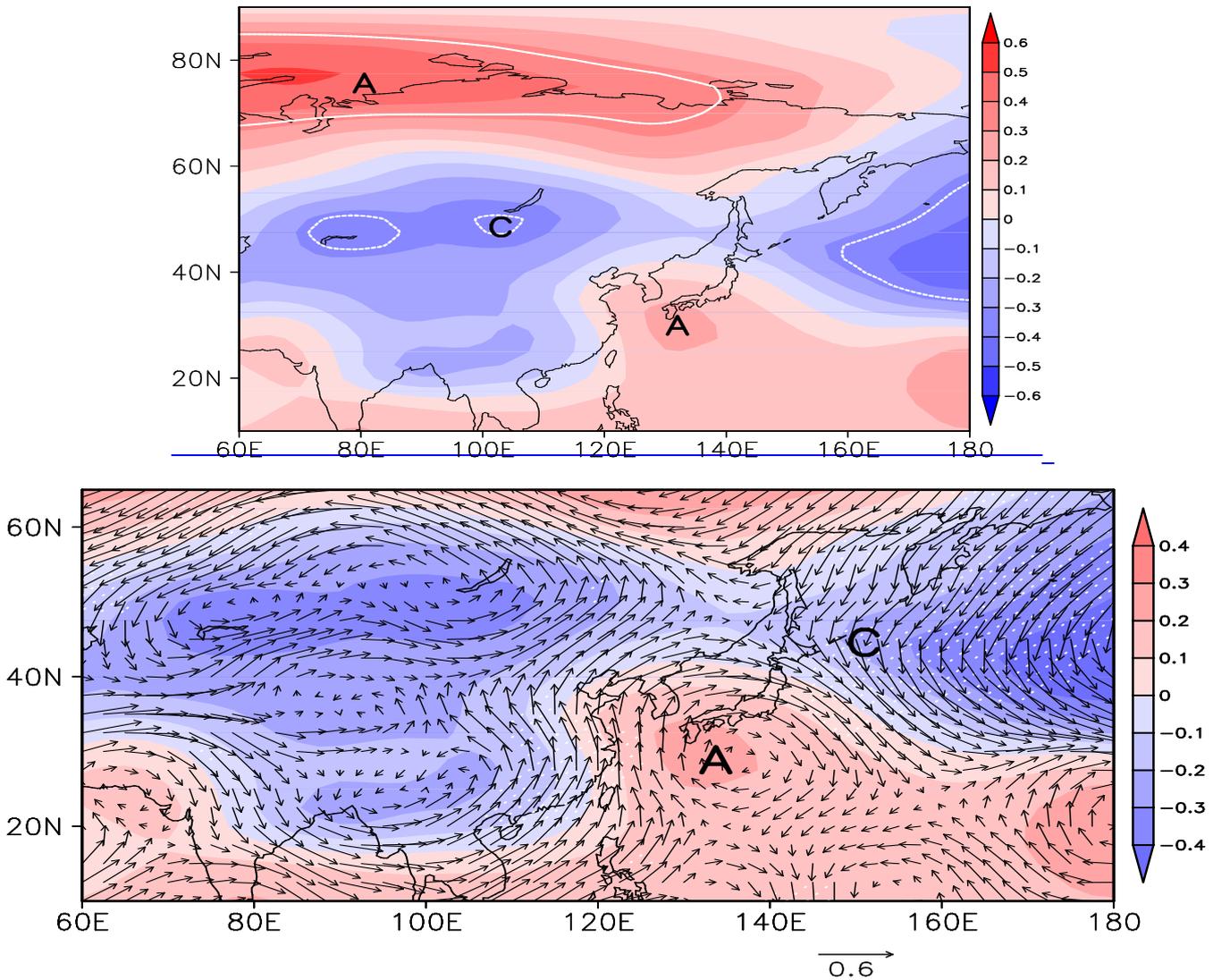
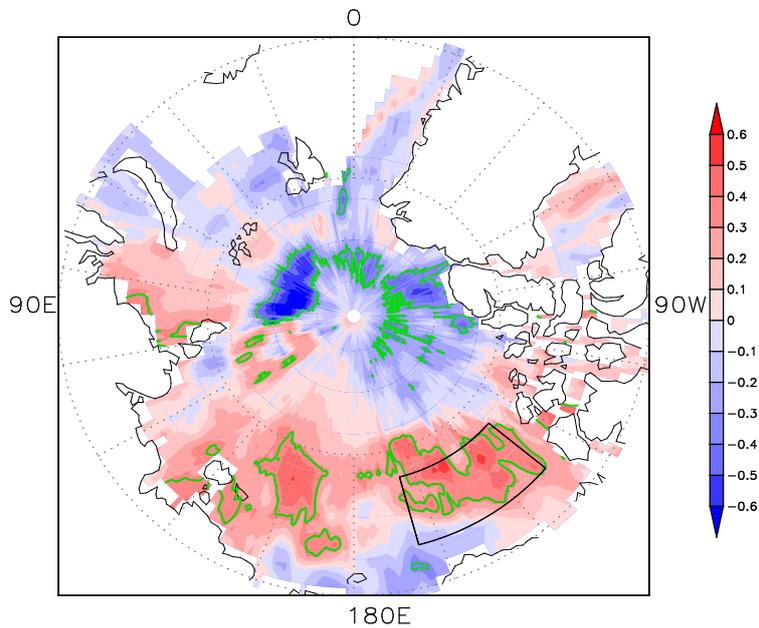
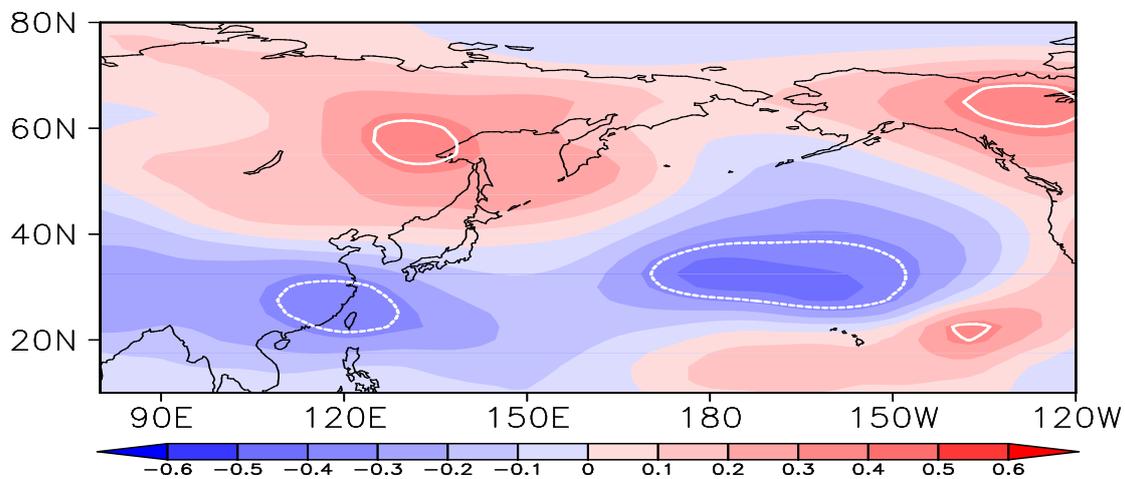


Figure 7. The CC between predictor x_3 ($\times -1$) and Z500 DY (shade)/850 hPa wind DY (arrow) in winter from 1980 to 2013. The dots white-curves indicate that the CC with meridional wind exceeded the 95% confidence level. A and C represent anti-cyclone and cyclone, respectively.

435



440 **Figure 8.** The CC between WHD_{NCP} DY and ASI DY in autumn from 1980 to 2013. The shades indicate that the CC exceeded the 95% confidence level, and the rectangle represents the selected region (73–78°N, 130–165°W) of predictor x_4 .



445 **Figure 9.** The CC between predictor x_4 and Z500 DY in winter from 1980 to 2013. The white curves indicate that the CC exceeded the 95% confidence level.

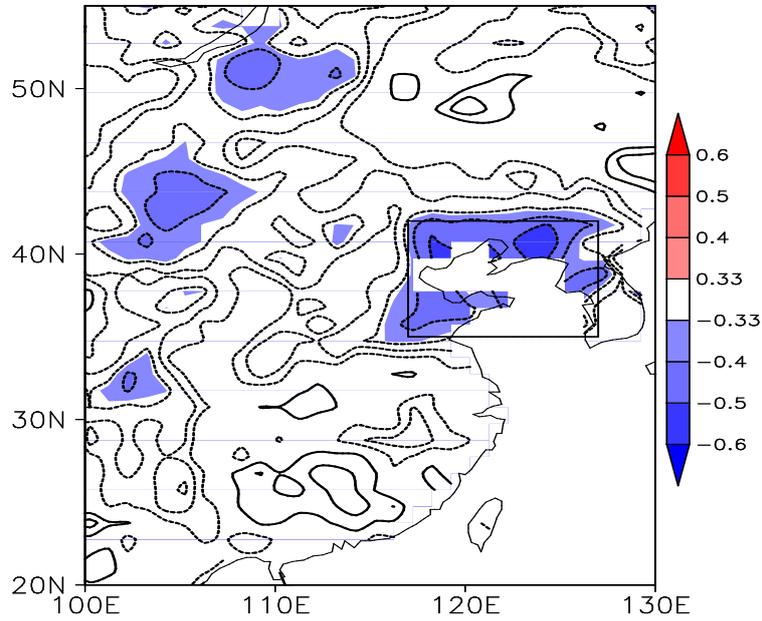


Figure 10. The CC between $WHD_{NCP} DY$ and $S_{soil\ moisture_M} DY$ in autumn from 1980 to 2013. The shades indicate that the CC exceeded the 95% confidence level, and the rectangle represents the selected region ($35\text{--}42^\circ N$, $117\text{--}127^\circ E$) of predictor x_5 .

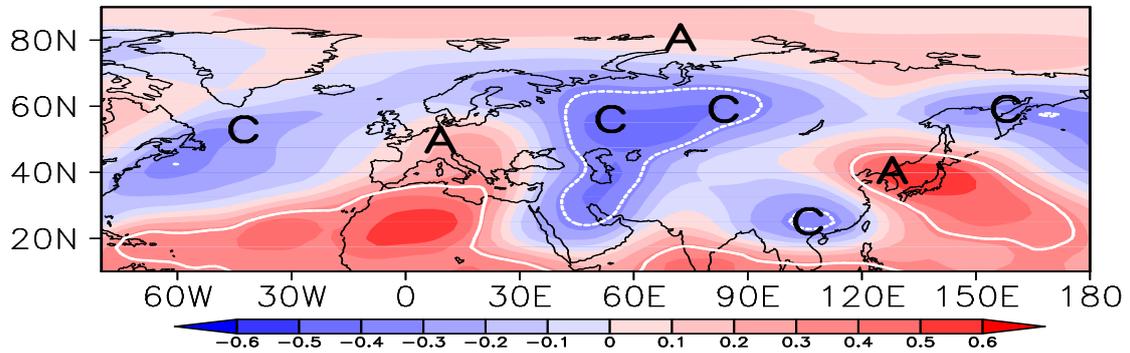
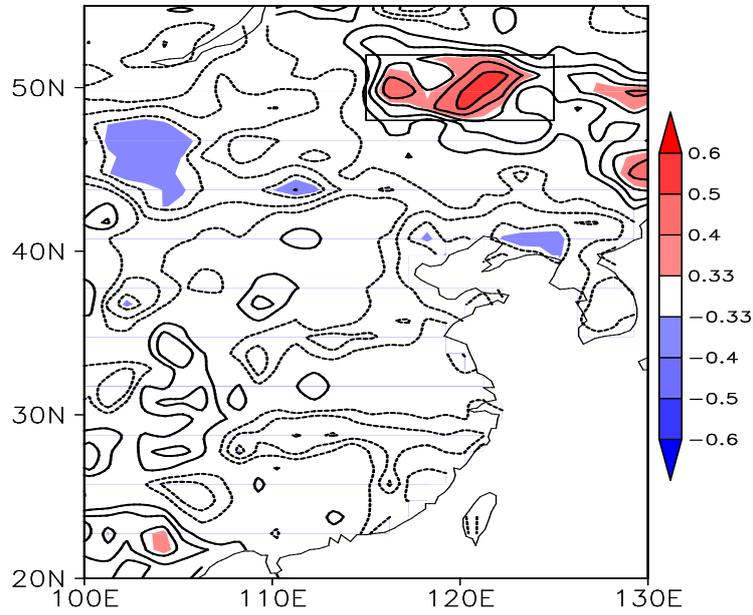


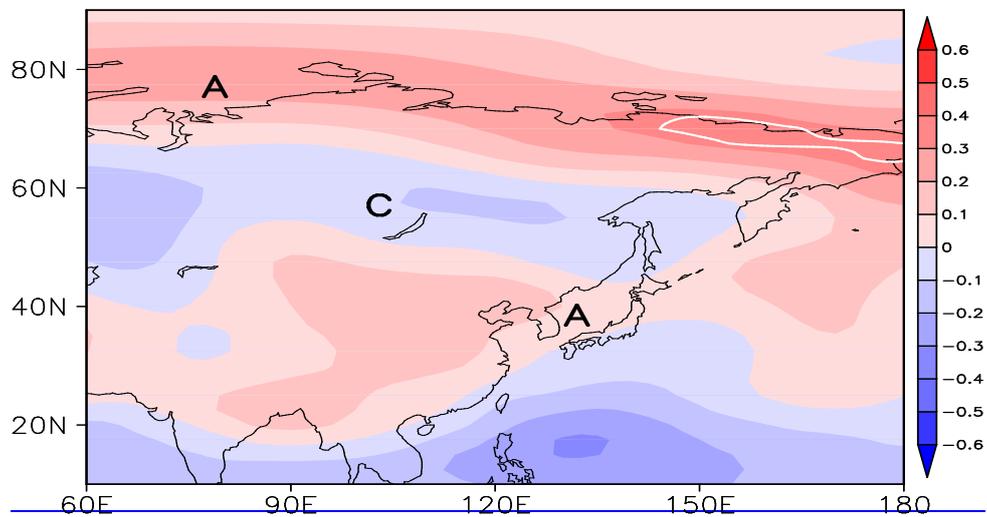
Figure 11. The CC between predictor $x_5 (\times -1)$ and $Z500 DY$ in winter from 1980 to 2013. The white curves indicate that the CC exceeded the 95% confidence level. A and C represent anti-cyclone and cyclone, respectively.

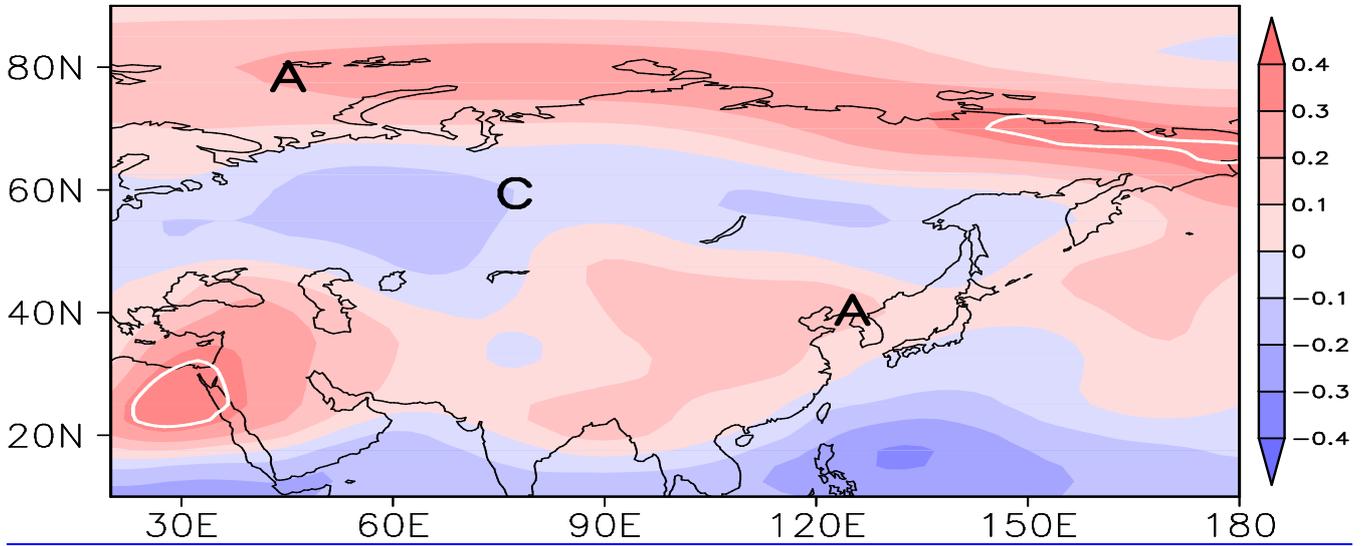
450



455

Figure 12. The CC between WHD_{NCP} DY and $S_{soil_Mmoisture}$ DY in summer from 1980 to 2013. The shades indicate that the CC exceeded the 95% confidence level, and the rectangle represents the selected region ($48\text{--}52^{\circ}\text{N}$, $115\text{--}125^{\circ}\text{E}$) of predictor x_6 .





460 Figure 13. The CC between predictor x_6 and Z500 DY in winter from 1980 to 2013. The white curves indicate that the CC exceeded the 95% confidence level. A and C represent anti-cyclone and cyclone, respectively.

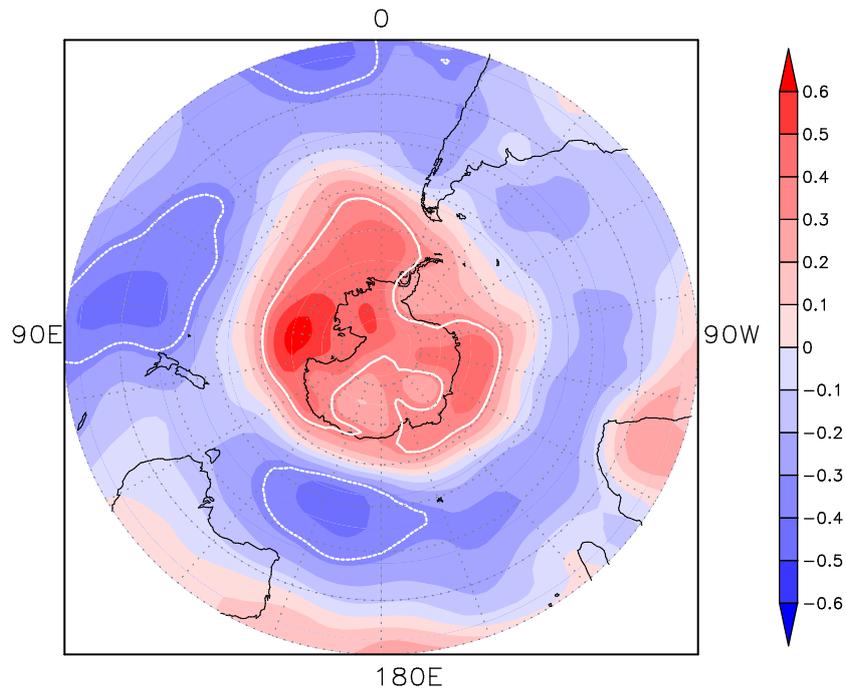
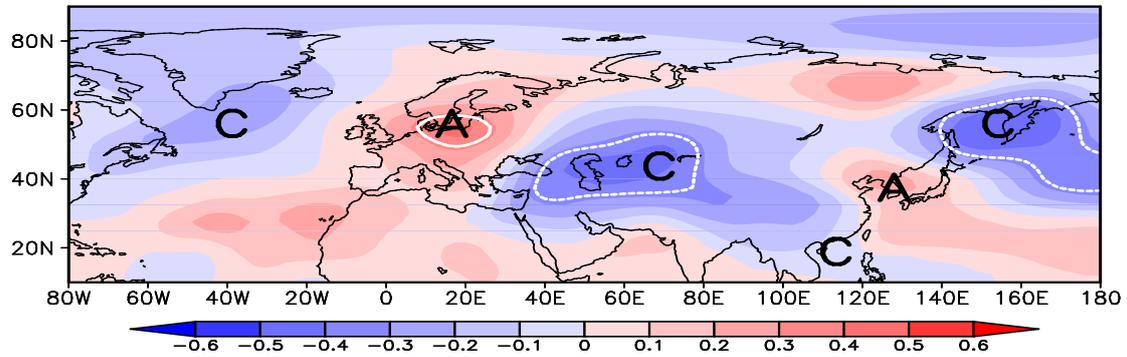
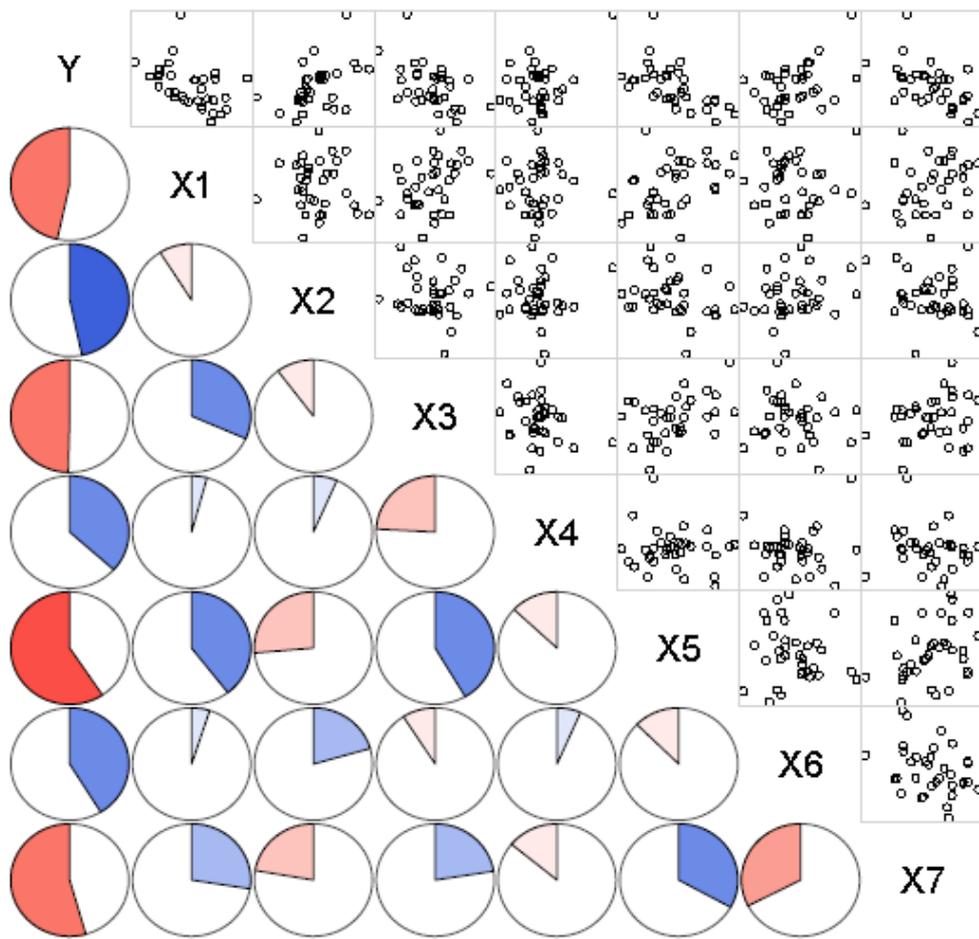


Figure 14. The CC between WHD_{NCP} DY and September-October Z850 DY from 1980 to 2013. The white curves indicate that the CC exceeded the 95% confidence level.



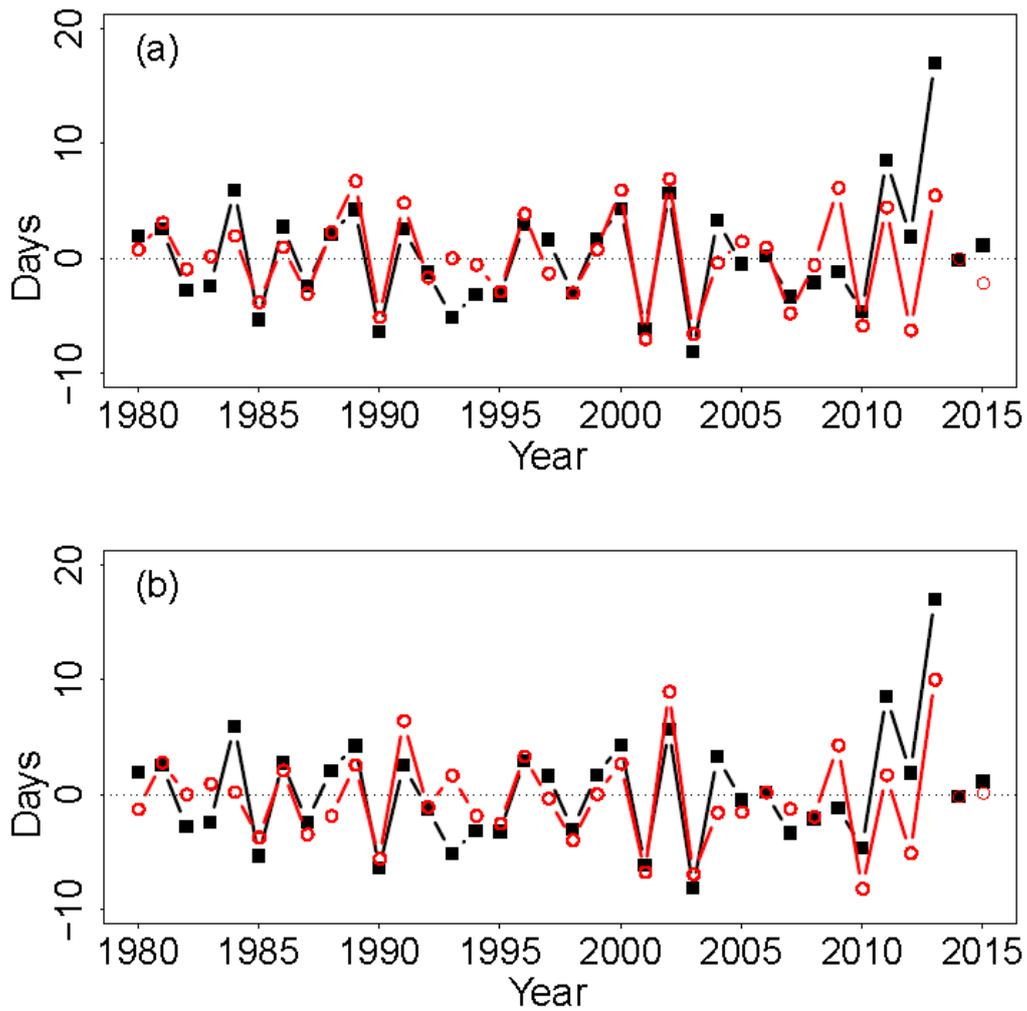
465

Figure 15. The CC between predictor x_7 ($\times-1$) and Z500 DY in winter from 1980 to 2013. The white curves indicate that the CC exceeded the 95% confidence level.



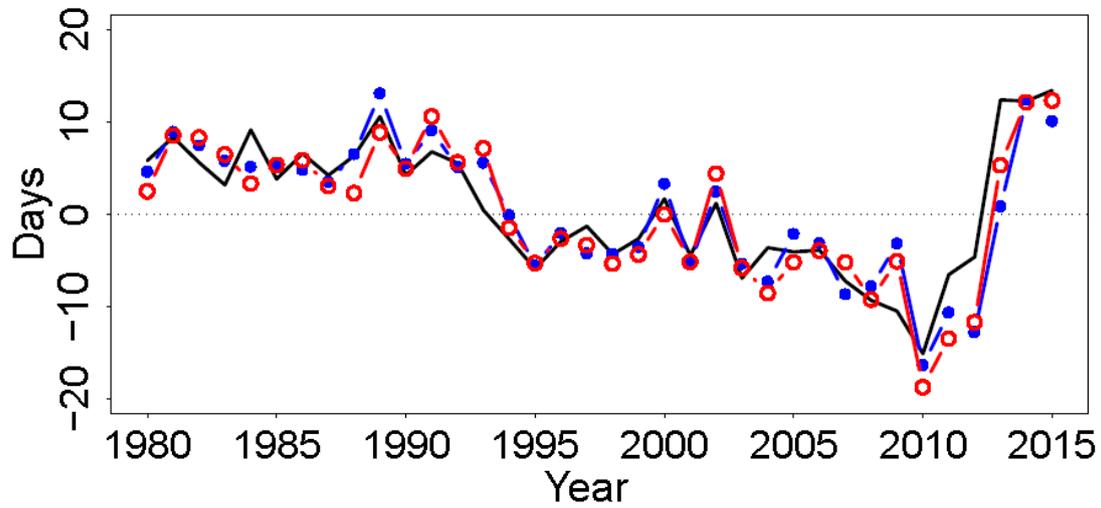
470

Figure 16. Correlogram of the dependent (Y) and independent ($x_1, x_2, \dots, \text{and } x_T$) variables, whose names were written on the diagonal. The lower panel shows the pie charts of correlation coefficients, representing the values by area and saturation, and showing positive/negative sign by blue/red, respectively. The upper panel shows the scatter plots.



475

Figure 17. The temporal variation of measured (black) WHD_{NCP_DY} , MLR (red, a) and GAM (red, b) cross-validation fitted WHD_{NCP_DY} from 1980 to 2013. The results for 2014 and 2015 represent the measured (black square) and predicted (red hollow circle) WHD_{NCP_DY} .



480 | **Figure 1817.** The temporal variation of measured (black) WHD_{NCP} anomaly from 1980 to 2015, MLR (blue) and
 | **GAM (red) simulative WHD_{NCP} anomaly, which was composed of cross fitted series from 1980 to 2013 and predicted**
 | **values in 2014 and 2015.**

485

490

495

Reply letter to the anonymous Referee #1

Received and published: 12 September 2016

General Comments:

500 The paper entitled “Seasonal prediction of winter haze days in the north-central North China Plain” selected seven predictors based on the analysis, and then used them to establish two statistical schemes for the prediction of the winter haze days over the north-central North China Plain. The two prediction models were demonstrated to have good ability to capture the interannual and interdecadal trends and the extremums of the haze days over the north-central North China. Thus, this study provides a good basis for the prediction of the haze days. I recommend the manuscript a minor revision before it be
505 published in the journal.

Specific Comments:

**(1) As is known, the human activity, especially the energy consumption, is the first driver for the increasing of the haze days in North China in recent years, while the climatic conditions may be
510 the second driver. So, how to take the human influence into account in the current prediction models? Some discussions about this issue are suggested to be included in the study.**

Reply:

515 There was no doubt that the human activities were the first driver and contributor for the increasing of haze days in China and should be taken into account, but it was quite difficult to gather the associated dataset. Our studies based on the assumption (or compromise) that the socio-economic component varied slowly between the current and previous year. Thus, the socio-economic terms could be neglected in the DY approach and were contained again by adding the previous measurement. Although the assumption was rough and simple, it indeed supports a way to the seasonal prediction of haze days.
520 This compromise might be unsuitable in certain years when this pollutant emission proportion varied

dramatically. Fortunately, the climate factors also contributed significantly and the developed models showed good performance. In the last section, following the kindly advice of the referee, we discussed the ideal scheme that used the preceding autumn energy consumption as a predictor.

Revision in the last paragraph:

525 At the same time, if the SPM_{MLR} performed well in some years, the SPM_{GAM} also showed good ability in these years, and *vice versa*. One possible reason could be that some useful factors, most notably the human activities, were not included here. There is no doubt that the human activities, especial the energy consumption, was the first driver for the increasing of haze pollution. In this paper, we simply assumed that the difference in pollutant emissions between current and previous years was
530 very small and that the socio-economic component of WHD_{NCP} varied slowly. This assumption could support the seasonal prediction of haze days in most of the years, but still was a compromise. In certain years, especially the recent years, this pollutant emission proportion varied rapidly that needed to be taken into account. The preceding autumn energy consumption should be a good choice, but difficult to be measured, and its DY could be introduced into the developed models directly to improve the
535 predictive skill.

**(2) The predictors are selected mainly based on the correlation analysis. The correlations may indicate some relations (phenomena) but do not really imply causality (reason). To confirm the reliability of the selected predictors for the prediction models, the physical mechanisms
540 underlying their relationships are suggested to be presented.**

Reply:

Actually, the studies about the associated physical mechanism, i.e., how the external forcings influenced haze pollutions, were new and still insufficient. In this paper, we selected 7 predictors and could not
545 present the physical mechanism that each external forcing stimulated such associated circulations.

Following the suggestions, we cited the latest reference about the impact of Pacific and Atlantic SST and the ASI on haze pollution, and also added some content and Figures about the way that the associated circulations impacted the WHD_{NCP} DY. Finally, we pointed out that the underlying physical mechanism about the external forcing needed further and deeper studies and some useful hints could be found in this paper.

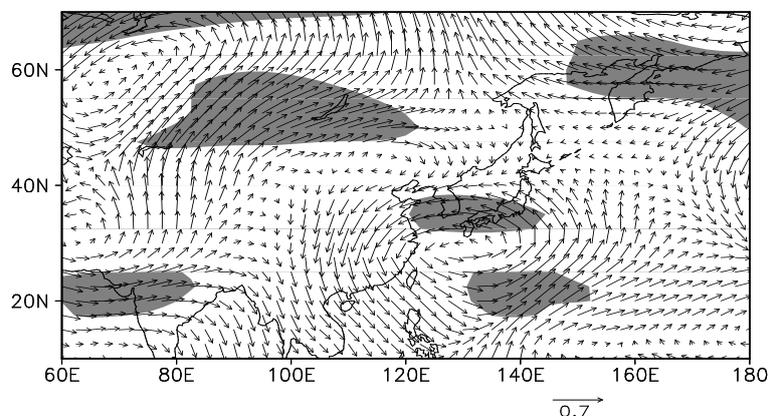
Revision for each predictor:

For Predictor x_1 , the following contents were revised:

The features of negative EU and positive WP pattern could be identified clearly and the anomalous cyclone over South China and South China Sea was significant in the circulations associated with predictor x_1 ($\times -1$) (Figure 3). Although the associated land-air interaction, especially in the DY field, was complicate and still unclear, according to the analysis of Figure 1, the horizontal and vertical diffusion of pollutant particles would be restricted efficiently.

For Predictor x_2 , the following contents were revised and Figure 5 was replaced:

Chen et al. (Chen et al. 2015) found that the severe winter haze events in the North China were closely related with the weaker and northward EAJS. The positive SST DY around the Alaska Gulf could induce obviously anomalous cyclone over eastern China and the adjacent ocean, and the stimulated easterly weakened the core of EAJS. Furthermore, there was significantly anomalous southerly at the high latitude that restricted the cold activities from their source region and intensified the haze pollution over NCP (Figure 5).



565

Figure 5. The CC between predictor x_2 and wind vector DY at 200 hPa in winter from 1980 to 2013. The shade indicates that the CC between the zonal wind DY and x_2 exceeded the 95% confidence level.

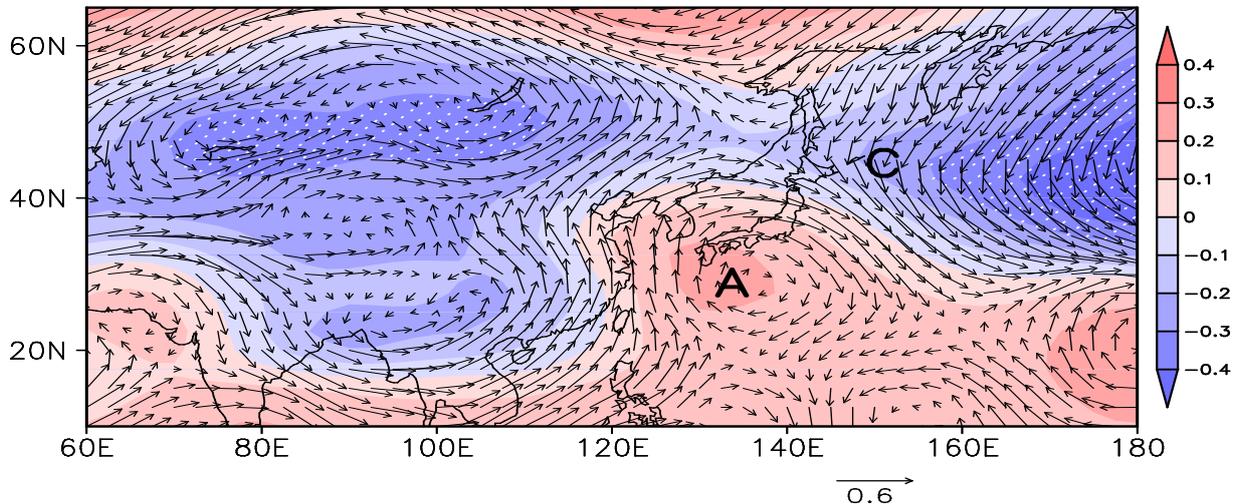
For Predictor x_3 , the following contents were revised, Figure 7 was replaced and a latest reference was cited:

570

Xiao et al (Xiao et al. 2015) proved the SST anomalies over the North Atlantic from summer to the following winter exhibit a significant relationship with winter haze days on both decadal and interannual timescale.

575

The most obvious DY atmospheric circulations related with predictor x_3 ($\times -1$) were the positive WP pattern, whose south center linked with a subtropical high (Figure 7). The continental high and marine low was both weakened by the anomalous geopotential height form the lower to middle layer that led to weaker EAWM and weaker cold air. The pressure gradient over the east coast of China also resulted in significant southerly anomalies, indicating smaller surface wind and more moisture and resulting in more WHD_{NCP}.



580

Figure 7. The CC between predictor x_3 ($\times -1$) and Z500 DY (shade)/850 hPa wind DY (arrow) in winter from 1980 to 2013. The dots indicate that the CC exceeded the 95% confidence level. A and C represent anti-cyclone and cyclone, respectively.

For Predictor x_4 , the following contents were revised:

585

Thus, the EAJS was weakened by the induced easterly and shifted northward that illustrated less cold activities over NCP (Yang et al. 2002) and generated more haze days.

For Predictor x_5 , the following contents were revised:

590

Following SST, the soil moisture is another important factor for seasonal prediction (Guo et al. 2007). The WHD_{NCP} was closely correlated with the moisture conditions due to the hygroscopicity of the atmospheric particles (Yin et al. 2015a).Being specific to local circulations, the cyclone over South China and the anti-cyclone over NCP and West Pacific stimulated significant southeaster between them (Figure omitted) that transported more moisture but decelerated the surface wind in the NCP.

For Predictor x_6 , the following contents and Figure 13 were revised:

595

The anomalous geopotential height was distributed zonally at high latitude indicating that the meridional circulations that transported cold air were weak. The positive high over NCP could confine

the vertical motion and the vertical diffusion of atmospheric particles and intensify the haze pollution over the NCP.

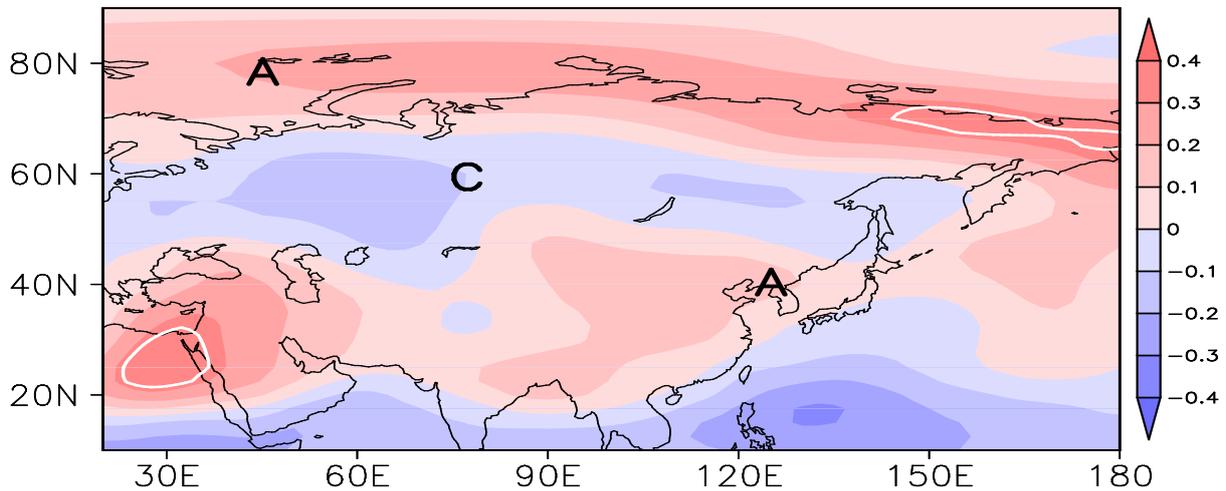


Figure 13. The CC between predictor x_6 and Z500 DY in winter from 1980 to 2013. The white curves indicate that the CC exceeded the 95% confidence level. A and C represent anti-cyclone and cyclone, respectively.

For Predictor x_7 , the following contents and were revised:

.....The anomalous anti-cyclone over NCP and adjacent ocean not only led to stable atmosphere but also resulted in small wind and high humidity.....

(3) There are far too many acronyms. This made me very confused when reading this paper. So, how can the authors expect the readers to remember all these acronyms in reading through the paper?

Reply

Yes, too many acronyms make the article difficult to read.

Revision in the last paragraph:

We have deleted the acronyms that used less than (including) thrice, such as Z, SWP, PSS, TBO, CPC,

GLM and EV.

615 **(4) The writing are needed to be further improved, and some sentences are needed to be rephrased to make them more clear. At several places, I cannot understand what is being conveyed. For example, “some new climatic finding have been helpful seasonal. . .”, “soil moisture is an important . . . but only after SST”, and so on. Please re-edit before submitting**

620 *Reply*

We have re-edited and improved the writing sentence by sentence, including the issues mentioned here and some others.

Revision related to the issues mentioned here, and the other revisions were addressed in the manuscript.

625 Some new climatic studies should be helpful for diagnosing seasonal predictors of winter haze days over the NCP (WHD_{NCP})

Following SST, the soil moisture is another important factor for seasonal prediction (Guo et al. 2007).

(5)Some references are cited in the text but not listed in the Reference section. Please check throughout the manuscript.

630

Reply

We checked throughout the manuscript and added the missed and some latest references.

Revisions:

The accessorial references were listed below.

635 Czaja A, Frankignoul C. 1999. Influence of the North Atlantic SST onthe atmospheric circulation. Geophys. Res. Lett. 26: 2969 –2972

Huang Y Y, Wang H J, Fan K. 2014. Improving the Prediction of the Summer Asian-Pacific Oscillation

Using the Interannual Increment Approach. *J. Climate*, 27: 8126—8134,
doi: <http://dx.doi.org/10.1175/JCLI-D-14-00209.1>

640 Yang S, Lau K M, Kim K M. 2002. Variations of the East Asian Jet Stream and Asian–Pacific–American
Winter Climate Anomalies. *Journal of Climate*, 15(3): 306—325

Yee T, Mitchell N. 1991. Generalized additive models in plant ecology, *Journal of Vegetation Science*,
2(5): 587—602

Xiao D, Li Y, Fan S J, Zhang R H, Sun J R, Wang Y. 2015. Plausible influence of Atlantic Ocean SST
645 anomalies on winter haze in China. *Theor. Appl. Climatol*, 122: 249—257

650

655

660

Reply letter to the anonymous Referee #2

Received and published: 21 September 2016

665 General comments:

By using the year-to-year increment as the predictands, the authors established a statistical model to predict the winter haze pollution over the North-Central North China Plain (NCP), in which seven predictors are selected and two schemes are employed. Cross validation shows that such model can successfully capture the interannual and interdecadal variabilities of winter haze days over the NCP and the extremums as well. The model based on the new approach of year-to-year increment is very skillful for the seasonal prediction of winter haze days in the NCP, and has greatly potential applications to the environmental pollutions. **However, more discussions are needed for the predictor selections and their possible physical processes in the successful seasonal prediction of winter haze days, so that the readers can better understand and apply this seasonal prediction model.** Minor revision is required before it is accepted.

670
675

Reply:

Actually, the studies about the associated physical mechanism, i.e., how the external forcings influenced haze pollutions, were new and still insufficient. In this paper, we selected 7 predictors and could not present the physical mechanism that each external forcing stimulated such associated circulations. Following the suggestions, we cited the latest reference about the impact of Pacific and Atlantic SST and the ASI on haze pollution, and also added some content and Figures about the way that the associated circulations impacted the WHD_{NCP} DY. Finally, we pointed out that the underlying physical mechanism about the external forcing needed further and deeper studies and some useful hints could be found in this paper.

680
685

General Revision and specific revision for x4, x5 and x7 (the specific revision for the other predictors would be presented under the specific issue):

In the last discussion section, we discussed the way that the external forcings impacted the WHD_{NCP} DY.

690Actually, the studies about the associated physical mechanism, i.e., how the external forcings influenced haze pollutions, were new and still insufficient. In this paper, the underlying physical process was presented mostly from the way that the associated circulations impacted the WHD_{NCP} DY. Thus, the physical mechanism that the external forcings stimulated such associated circulations needed further studies.....

695 **For Predictor x_4 , the following contents were revised:**

Thus, the EAJS was weakened by the induced easterly and shifted northward that illustrated less cold activities over NCP (Yang et al. 2002) and generated more haze days.

For Predictor x_5 , the following contents were revised:

Following SST, the soil moisture is another important factor for seasonal prediction (Guo et al. 2007).

700 The WHD_{NCP} was closely correlated with the moisture conditions due to the hygroscopicity of the atmospheric particles (Yin et al. 2015a).Being specific to local circulations, the cyclone over South China and the anti-cyclone over NCP and West Pacific stimulated significant southeaster between them (Figure omitted) that transported more moisture but decelerated the surface wind in the NCP.

For Predictor x_7 , the following contents and were revised:

705The anomalous anti-cyclone over NCP and adjacent ocean not only led to stable atmosphere but also resulted in small wind and high humidity.....

For publication, several major issues need to be addressed:

710 **1. Line 45: “climate change” usually indicates a long-term climate variation, but DY is more like interannual variability, so it’s not appropriate to use the phrase “climate change” here.**

Reply:

Yes, this detail indeed indicates different connotation.

Revision:

715 We changed the phrase to “climate variability”

2. Line 73-74 and Line 77: Why are the MLR and GLM called model-driven and datadriven methods, respectively? A brief description is highly encouraged.

720 *Reply:*

A brief description and related reference was supplemented.

Revision:

725 The GAM is data-driven rather than model-driven. The resulting fitted values do not come from an apriori model that was adopted by MLR and generalized linear model. The rationale behind fitting a nonparametric model is that the structure of data should be examined first to choose an appropriate smooth function for each predictor; that is, the GAM allow the data to determine the shape of the smooth function (Yee et al. 1991).

730 **3. Line 87, Line 97, Line 103 and many others: it should be noted that Pacific Japan (PJ) pattern is a summer teleconnection identified by Nitta (1987), not a winter one.**

Reply:

Thank you for the reminder. The WP pattern should be accurate following the description from the CPC's website.

735 The WP pattern is a primary mode of low-frequency variability over the North Pacific **in all months**. During winter and spring, the pattern consists of a north-south dipole of anomalies, with one center located over the Kamchatka Peninsula and another broad center of opposite sign covering portions of

southeastern Asia and the western subtropical North Pacific.

Revision:

740 The PJ pattern has been changed into WP pattern, and the anomalous cyclone over South China Sea was analyzed as a individual system. The particular revision could be seen in the revisions related the predictors.

4. Line 91-92: Why is the water vapor transportation enhanced?

745

Reply:

The reason was added in the paper.

Revision:

750Meanwhile, the water vapor transportation was also enhanced by anomalous southeaster in the lower troposphere (Figure omitted), creating favorable conditions for more WHD_{NCP} than in the previous year.

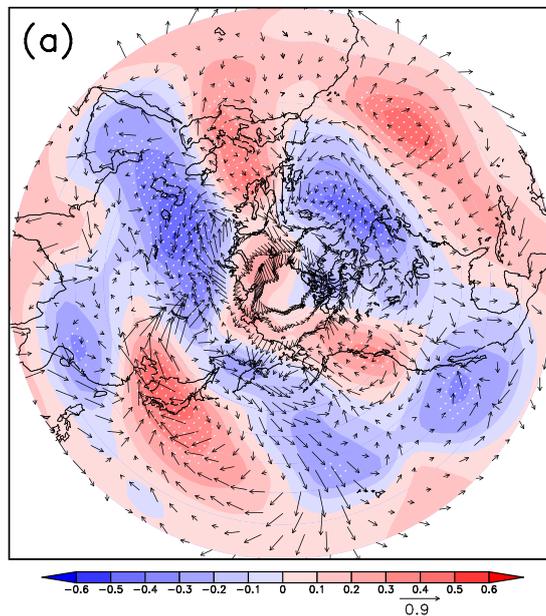


Figure S. The LCC between Z500 DY (shade)/surface wind DY (arrow) in winter and WDY from 1980

to 2013. The dots indicate LCC exceed the 90% confidence level.

755

5. Line 95: Region of the Japan Sea to the Stanovoy Range is chosen for the preautumn TS DY, however, from Figure 2, we can see the region from the Japan Sea to the south of Lake Baikal has larger correlation coefficient. Maybe it is better to use this region for predictor x1.

760 *Reply:*

Before submission, we tried the SAT DY from the Japan Sea to the south of Lake Baikal, but the performance was not as good as the selected one. We speculated that the reason might be that there were some internal interactions among the predictors. The balance among predictors was also a key point needed to be considered.

765 *Revision:*

Leave the predictor x1 as it was.

6. Line 97: As defined by Wallace and Gutzler (1981), EU pattern has three major nodes with their locations at (55N, 20E), (55N, 75E) and (40N, 145E). However, Figure 3 doesn't cover the whole area of EU pattern, so we cannot obviously see the negative EU features from Figure 3. Besides, Figure 3 has rather different features from Figure 1 except for the anticyclone anomalies over the South Japan.

775 *Reply:*

In this article, we followed the definition of EU from Wang et al (2016) and plotted a small Figure. Actually, if we expand the plotted area, the typical EU pattern as mentioned could be identified clearly. Thus, we re-plotted Figure 3 and revised the words to include more analysis about the physical process.

Wang H J, He S P. 2015b. The North China/Northeastern Asia Severe Summer Drought in 2014. *Journal of Climate*. 28(17), 6667–668

780 *Revision:*

For Predictor x_1 , the following contents and Figure 3 were revised:

The features of negative EU and positive WP pattern could be identified clearly and the anomalous cyclone over South China and South China Sea was significant in the circulations associated with predictor x_1 ($\times -1$) (Figure 3). Although the associated land-air interaction, especially in the DY field, was complicate and still unclear, according to the analysis of Figure 1, the horizontal and vertical diffusion of pollutant particles would be restricted efficiently.

785

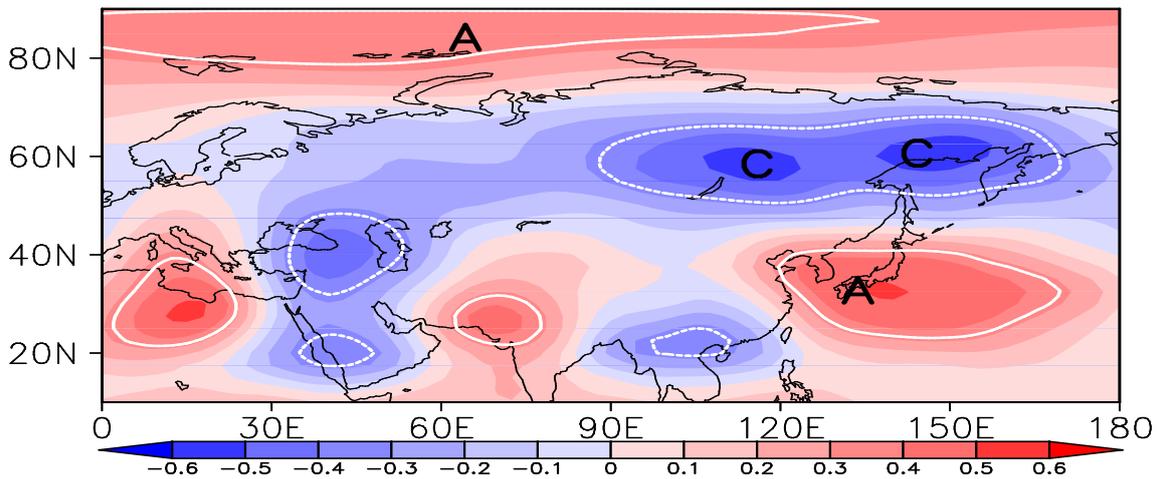


Figure 3. The CC between predictor x_1 ($\times -1$) and Z500 DY in winter from 1980 to 2013. The white curves indicate that the CC exceeded the 95% confidence level. A and C represent anti-cyclone and cyclone, respectively.

790

7. Line 103, Line 107: As for the PJ and negative EU pattern, they are not very clearly and significantly seen.

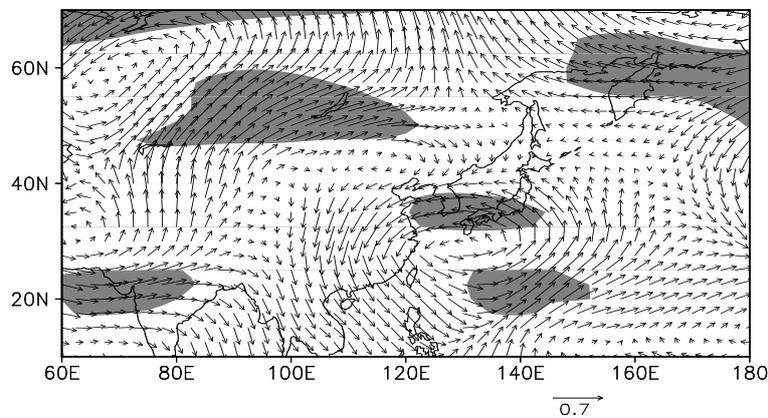
795 *Reply:*

Line 103 and 107 indicate the predictor x_2 (Figure 5) and x_3 (Figure 7), respectively. In the submitted manuscript, the EU and PJ patterns were not clear enough at the middle troposphere. Thus, we replaced the Figures and analyzed the associated wind at 200 hPa with x_2 (EAJS) and the associated WP pattern (and southeaster over East China) with x_3 . After revision, the associated circulations were significant and the physical process was clearer.

800 *Revision:*

For Predictor x_2 , the following contents were revised and Figure 5 was replaced:

Chen et al. (Chen et al. 2015) found that the severe winter haze events in the North China were closely related with the weaker and northward EAJS. The positive SST DY around the Alaska Gulf could induce obviously anomalous cyclone over eastern China and the adjacent ocean, and the stimulated easterly weakened the core of EAJS. Furthermore, there was significantly anomalous southerly at the high latitude that restricted the cold activities from their source region and intensified the haze pollution over NCP (Figure 5).

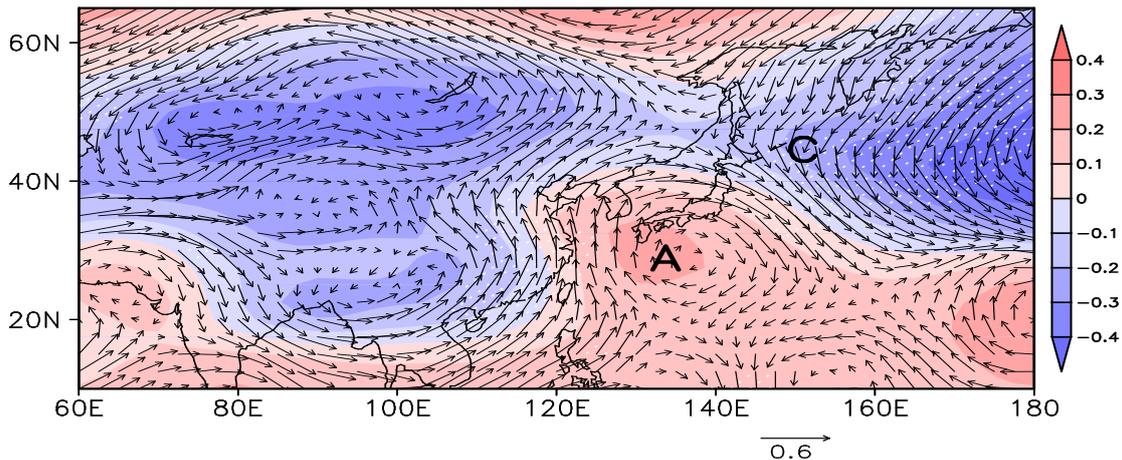


810 Figure 5. The CC between predictor x_2 and wind vector DY at 200 hPa in winter from 1980 to 2013. The shade indicates that the CC between the zonal wind DY and x_2 exceeded the 95% confidence level.

For Predictor x_3 , the following contents were revised, Figure 7 was replaced and a latest reference was cited:

815 Xiao et al (Xiao et al. 2015) proved the SST anomalies over the North Atlantic from summer to the following winter exhibit a significant relationship with winter haze days on both decadal and interannual timescale.

The most obvious DY atmospheric circulations related with predictor x_3 ($\times -1$) were the positive WP pattern, whose south center linked with a subtropical high (Figure 7). The continental high and marine low was both weakened by the anomalous geopotential height form the lower to middle layer that led to weaker EAWM and weaker cold air. The pressure gradient over the east coast of China also resulted in significant southerly anomalies, indicating smaller surface wind and more moisture and resulting in more WHD_{NCP} .



825 Figure 7. The CC between predictor x_3 ($\times -1$) and Z500 DY (shade)/850 hPa wind DY (arrow) in winter from 1980 to 2013. The dots indicate that the CC with meridional wind exceeded the 95% confidence level. A and C represent anti-cyclone and cyclone, respectively.

8. Line 110-113, Line 128-129: The correlation coefficients near the NCP are almost insignificant, how can the predictors x_3 and x_6 be so important?

830

Reply:

Line 110—113 and 128—129 related with the predictor x_3 (Figure 7) and x_6 (Figure 13), respectively. The issue about x_3 was addressed in Q7. The southerly anomalies at the lower troposphere were significant (dots in Figure 7) indicating smaller surface wind and more moisture. For predictor x_6 , we expand the plotted area; the typical EU pattern could be identified, but the correlation coefficients near the NCP were still insignificant. In consideration of distributed pattern, we still kept this predictor. In addition to the teleconnection, the local meteorological circulation and conditions were analyzed, including the weaker meridional circulations and vertical motion.

835

840

Revision:

For Predictor x_6 , the following contents and Figure 13 were revised:

The anomalous geopotential height was distributed zonally at high latitude indicating that the meridional circulations that transported cold air were weak. The positive high over NCP could confine the vertical motion and the vertical diffusion of atmospheric particles and intensify the haze pollution over the NCP.

845

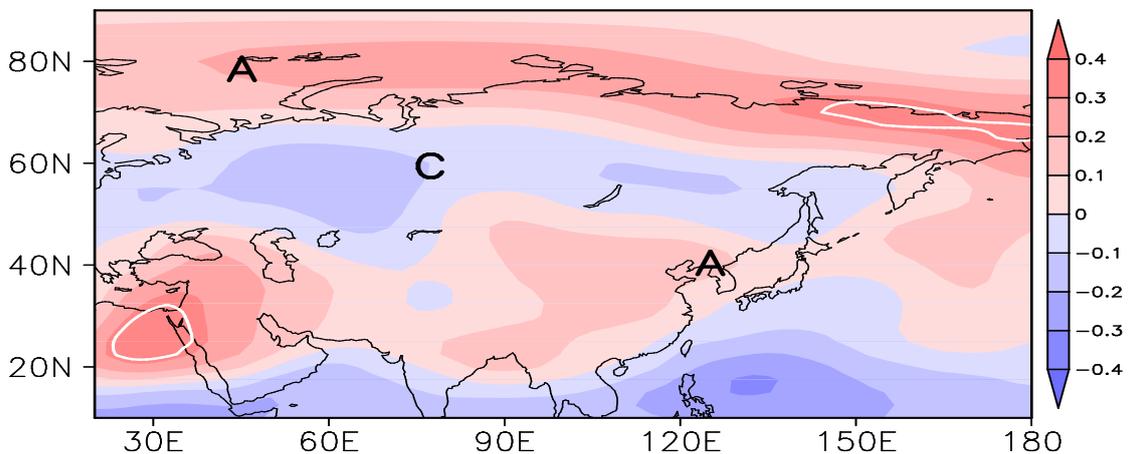


Figure 13. The CC between predictor x_6 and Z500 DY in winter from 1980 to 2013. The white curves indicate that the CC exceeded the 95% confidence level. A and C represent anti-cyclone and cyclone, respectively.

850

9. Line 143: Figure 16 is a little complicated, it's better to explain it briefly.

Reply:

Yes, Figure 16 is a little complicated and the information exceeded the necessary demand. In another word, the texts were enough. What we wanted to presented was that only 5 pairs of the predictors showed significant linear correlation and the multicollinearity would not be a problem when modeling with the MLR approach.

855

Revision:

Figure 16 was deleted.

860

10. Line 154: How to define the predicted bias?

Reply:

The definition was supplemented.

865

Revision:

.....As an independent prediction test, the predicted bias, i.e., the predicted value minus the measurement, in 2014 was 0.09.....

11. Line 180: The center of positive geopotential anomalies is actually located over the Japan Sea.

870

Reply:

The description was revised.

Revision:

.....positive anomalies over the NCP and Japan Sea.....

875

12. Line 182-184: The consistence of SPMMLR and SPMGAM indicates that the linear part dominates the WHDNCP predictions. At the same time, the failure of predicting the rapid rising trend after 2010 also implies that the DY method has some deficiencies in dealing with the large abrupt change. The authors should point them out.

880

Reply:

The reminder from the referee was important for the discussion. Thus, we enriched the discussion about the comparison between MLR and GAM, and the DY approach itself.

Revision:

885The consistence of these two models might indicate that, after including plentiful predictors, the linear relationship dominated the WHD_{NCP} DY prediction.....

.....The large abrupt change was a common challenge to the statistical models, including the DY approach, so the numerical model should be introduced into the prediction of haze pollution. At the same time, if the SPM_{MLR} performed well in some years, the SPM_{GAM} also showed good ability in these years, and *vice versa*. One possible reason could be that some useful factors, most notably the human activities, were not included here. There is no doubt that the human activities, especial the energy consumption, was the first driver for the increasing of haze pollution. In this paper, we simply assumed that the difference in pollutant emissions between current and previous years was very small and that the socio-economic component of WHD_{NCP} varied slowly. This assumption could support the seasonal prediction of haze days in most of the years, but still was a compromise. In certain years, especially the

890

recent years, this pollutant emission proportion varied rapidly that needed to be taken into account. The preceding autumn energy consumption should be a good choice, but difficult to be measured, and its DY could be introduced into the developed models directly to improve the predictive skill.....

900 **Technical corrections:**

13. Line 25-26: The sentence is a little awkward, and it needs modification.

Reply:

The sentence has been corrected.

905 *Revision:*

Some new climatic studies should be helpful for diagnosing seasonal predictors of winter haze days over the NCP (WHD_{NCP})

14. Line 38: The citation of Huang et al. 2015 is not present in the reference list.

910

Reply:

This citation has been merged into Huang et al. 2014

Revision:

.....the signals (i.e., variance) of the predictors and predictand were both amplified (Huang et al. 2014)
915 and.....

15. Line 48: The citation of Huang et al. 2014 is not present in the reference list.

Reply:

920 This citation of Huang et al. 2014 has been listed in the reference.

Revision:

Huang Y Y, Wang H J, Fan K. 2014. Improving the Prediction of the Summer Asian-Pacific Oscillation Using the Interannual Increment Approach. *J. Climate*, 27: 8126—8134, doi: <http://dx.doi.org/10.1175/JCLI-D-14-00209.1>

925

16. Line 91: “Asia” should be “Asian”.

Reply:

The error has been corrected.

930

Revision:

.....together with the cyclone; they could induce an easterly to weaken the East Asian Jet Stream (EAJS), producing weaker cold air.

17. Line 105: As for the “Prior studies”, some citations should be given.

935

Reply:

The citation below was given.

Czaja A, Frankignoul C. 1999. Influence of the North Atlantic SST on the atmospheric circulation. *Geophys. Res. Lett.* 26: 2969–2972

940

Revision:

Prior studies have documented that the triple SST pattern was a dominant mode of the northern Atlantic SST in autumn (Czaja et al. 1999).

18. Line 118-119: it needs to make clear what this sentence is talking about, geopotential height?

945

Reply:

Yes, it was talking about geopotential height and has been rewritten.

Revision:

.....The continental high and marine low was both weakened by the anomalous geopotential height
950 form the lower to middle layer that led to weaker EAWM and weaker cold air.....

19. Line 119: “EAJS” should be “EASJ”.

Reply:

955 We made a mistake when defining the acronym. It should be East Asian Jet Stream (EAJS), and we corrected them throughout the paper.

Revision:

.....together with the cyclone; they could induce an easterly to weaken the East Asian Jet Stream
(EAJS), producing weaker cold air.....
960

20. Line 160: “processing” should be “process”.

Reply:

The error has been corrected.

965 *Revision:*

.....and show the trend well, indicating an advantage to process the non-linear relationship.....

21. Line 165: “results that are” can be changed to “and the results are”.

970 *Reply:*

The error has been corrected.

Revision:

The predicted bias in 2014 and 2015 was -0.07 and -1.01 , and the results are slightly better than those from SPM_{MLR} .

975

22. Line 171: “simulative” to “simulated”.

Reply:

The error has been corrected.

980

Revision:

In Figure 17, the simulated WHD_{NCP} anomaly was fitted by cross-validation from 1980 to 2013 and predicted in 2014 and 2015.

23. Keep the tense consistent in the whole paper.

985

Reply:

The tense has been checked throughout and, then was kept consistent.