



Seasonal prediction of winter haze days in the north central North China Plain

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

¹Key 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), Nanjing University of Information Science & Technology, Nanjing, 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)

Received: 1 August 2016 – Published in Atmos. Chem. Phys. Discuss.: 16 August 2016

Revised: 26 October 2016 – Accepted: 31 October 2016 – Published: 30 November 2016

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). 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.

ter 2000, the combined effects of a rapid increase in total energy consumption and the influence of climate change intensified the haze pollution in central north China (Wang and Chen, 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 contributor of haze. Some new climatic studies 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, b; Q. Li et al., 2015). By weakening EAWM circulations, negative sea surface temperature (SST) anomalies over the subtropical western Pacific could significantly intensify WHD_{NCP} (Yin and Wang, 2016). Furthermore, the decline of preceding autumn (September–November) Arctic sea ice (ASI) has led to favorable environments for haze with stable atmosphere 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.

The climate variables in east Asia showed obvious characteristics of tropospheric biennial oscillation, based on which a new interannual increment approach was applied for short-term climate prediction (Wang et al., 2000, 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

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 and Liu, 2014), particularly after persistent heavy fog and haze in January 2013 (Zhang et al., 2014; Zhao et al., 2014). Af-

the observed information from the previous year and the features of biennial oscillation, the interannual variation and interdecadal trend could be captured well. In addition, the signals (i.e., variance) of the predictors and predictand were both amplified (Huang et al., 2014) and thus of benefit to improve the prediction skill. If the predictive objects (Y), e.g., haze days, were cross influenced by socioeconomic factors and climatic conditions, the predictand could be represented by $Y = YS + YC$, where YS and YC were the slowly varying socioeconomic and climatic components, respectively.

$$\begin{aligned} DY &= Y_t - Y_{t-1} = (YS_t + YC_t) - (YS_{t-1} + YC_{t-1}) \\ &= (YS_t - YS_{t-1}) + (YC_t - YC_{t-1}), \end{aligned}$$

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

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. After adding the predicted WHD_{NCP} DY to the observed WHD_{NCP} of the previous year, the interdecadal and socioeconomic components were contained in the final prediction. In prior studies, the DY approach has been used to explore the prediction of summer rainfall in China (Fan et al., 2008), heavy winter snow activity in northeast China (Fan and Tian, 2013), summer Asian–Pacific Oscillation (Huang et al., 2014) and winter North Atlantic Oscillation (Tian and Fan, 2015). Furthermore, some variables cross influenced by socioeconomic and climatic factors were predicted successfully using the DY approach, e.g., rice production in northeast China (Zhou and Wang, 2014) and the discoloration day for *Cotinus coggygia* 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} .

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

2 Data sets and methods

Monthly atmospheric data, such as geopotential height and surface air temperature (SAT), were derived from the National Centers for Environmental Prediction/National Center for Atmospheric Research (NCEP/NCAR) global reanalysis data set with a horizontal resolution of $2.5^\circ \times 2.5^\circ$ from 1979 to 2016 (Kalnay et al., 1996). The monthly mean extended

reconstructed SST data sets with a horizontal resolution of $2^\circ \times 2^\circ$ from 1979 to 2016 were obtained from the National Oceanic and Atmospheric Administration (NOAA) (Smith et al., 2008). ASI extent was calculated from the ASI concentration data, downloaded from the Hadley Centre with a horizontal resolution of $1^\circ \times 1^\circ$ 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, with a horizontal resolution of $0.5^\circ \times 0.5^\circ$ (Huug et al., 2003). The monthly Antarctic Oscillation (AAO) indices from 1979 to 2016 were also obtained from the Climate Prediction Center (Mo, 2000).

China ground observations from 39 NCP stations, collected by the National Meteorological Information Center of China four times per day from 1979 to 2016, were used to reconstruct the climatic WHD data (Yin and Wang, 2016). Here, haze 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 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, was ultimately expressed as a linear combination of K predictors (x_i) that 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 could be described as follows:

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

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

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

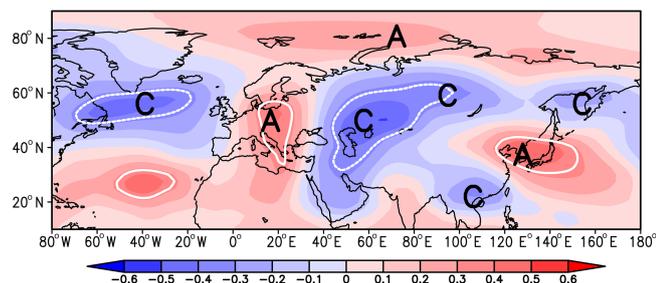


Figure 1. The CC between $\text{WHD}_{\text{NCP}} \text{DY}$ and geopotential height at 500 hPa (Z500) in winter from 1980 to 2013. The white curves indicate that the CC exceeded the 95 % confidence level. A and C represent anticyclone and cyclone, respectively.

The normalized data sets 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 were calculated for evaluation by simple fitting and leave-one-out cross validation.

3 The predictors and associated circulations

To choose the DY predictors, the correlated DY atmospheric circulations were identified, as shown in Fig. 1. The positive phase of the east Atlantic/west Russia (EA/WR) and west Pacific (WP; Barnston and Livezey, 1987) patterns and the negative phase of the Eurasia (EU; Wallace and Gutzler, 1981) pattern were obvious, and we took the anticyclone circulation over north China as an intermediary that 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 (EAJS), producing weaker cold air. Meanwhile, the water vapor transportation was also enhanced by an anomalous southeaster in the lower troposphere (figure omitted), creating favorable conditions for more WHD_{NCP} than in the previous year.

The pivotal local anticyclone over the NCP was the most important contributor; we therefore speculated that pre-autumn SAT DY around the NCP should be effective to impact $\text{WHD}_{\text{NCP}} \text{DY}$. There were significantly negative correlations between $\text{WHD}_{\text{NCP}} \text{DY}$ and pre-autumn SAT DY from the Sea of Japan to the Stanovoy Range (35–65° N, 130–140° E), the area mean of which was selected as predictor x_1 (Fig. 2). The correlation coefficient (CC) between $\text{WHD}_{\text{NCP}} \text{DY}$ and predictor x_1 was -0.47 , exceeding the 99 % confidence level. The features of negative EU and positive WP pattern could be identified clearly and the anomalous cyclone over south China and the South China Sea was significant in the circulations associated with predictor x_1 ($\times -1$) (Fig. 3).

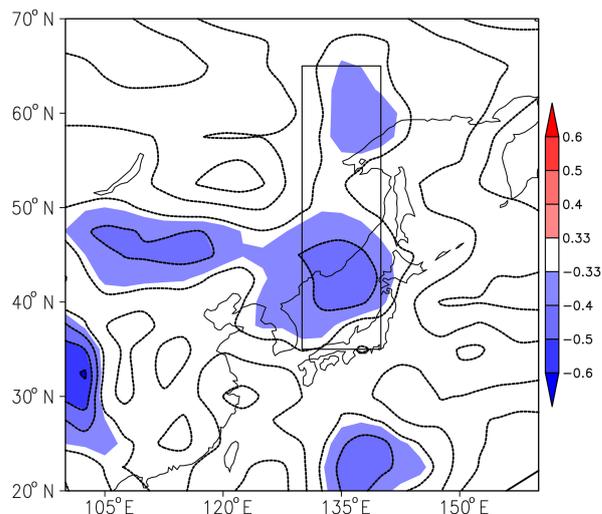


Figure 2. The CC between $\text{WHD}_{\text{NCP}} \text{DY}$ and SAT DY in autumn from 1980 to 2013. The shades indicate that the CC exceeded the 95 % confidence level, and the rectangle represents the selected region (35–65° N, 130–140° E) of predictor x_1 .

Although the associated land–air interaction, especially in the DY field, was complicated and still unclear, according to the analysis of Fig. 1, the horizontal and vertical diffusion of pollutant particles would be restricted efficiently.

The pre-autumn SST anomalies and their associated winter SST of the Pacific could influence WHD_{NCP} significantly via the air–sea interaction (Yin and Wang, 2016). Figure 4 shows the CC between $\text{WHD}_{\text{NCP}} \text{DY}$ and pre-autumn Pacific SST DY. The most significant CC distributed around the Gulf of Alaska (36–56° N, 130–170° W), and the area-averaged SST DY here was defined as predictor x_2 , whose CC with $\text{WHD}_{\text{NCP}} \text{DY}$ was 0.47 (above the 99 % confidence level). Chen and Wang (2015) found that the severe winter haze events in north China were closely related with the weaker and northward EAJS. The positive SST DY around the Gulf of Alaska could induce an obviously anomalous cyclone over eastern China and the adjacent ocean, and the stimulated easterly weakened the core of EAJS. Furthermore, there was a significantly anomalous southerly at the high latitude that restricted the cold activities from their source region and intensified the haze pollution over the NCP (Fig. 5).

Prior studies have documented that the triple SST pattern was a dominant mode of the northern Atlantic SST in autumn (Czaja and Frankignoul, 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, 2014). 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 $\text{WHD}_{\text{NCP}} \text{DY}$ and pre-autumn SST DY of the Atlantic was distributed

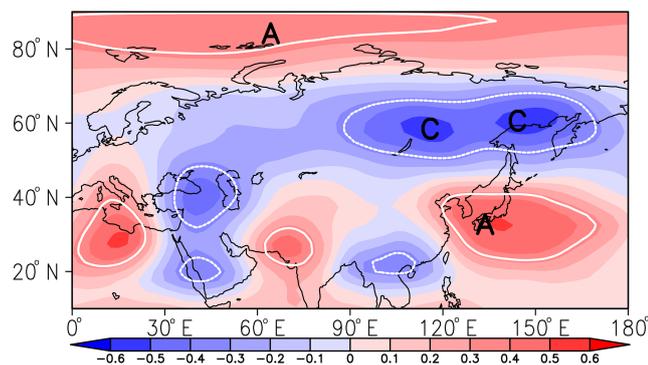


Figure 3. The CC between predictor x_1 ($\times -1$) and Z500 DY in winter from 1980 to 2013. The white curves indicate that the CC exceeded the 95 % confidence level. A and C represent anticyclone and cyclone, respectively.

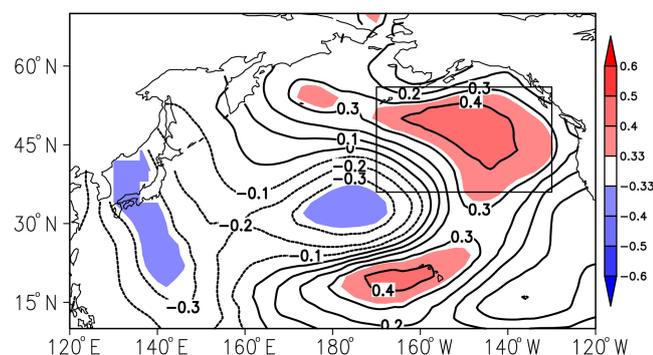


Figure 4. The CC between WHD_{NCP} DY and Pacific SST DY in autumn from 1980 to 2013. The shades indicate that the CC exceeded the 95 % confidence level, and the rectangle represents the selected region ($36\text{--}56^\circ\text{N}$, $130\text{--}170^\circ\text{W}$) of predictor x_2 .

in a “ $- + -$ ” pattern (Fig. 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 to predictor x_3 ($\times -1$) were the positive WP pattern, whose south center linked with a subtropical high (Fig. 7). The continental high and marine low was both weakened by the anomalous geopotential height from the lower to the 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} .

ASI decreased dramatically with significant variance and was a significant contributor influencing WHD in eastern China (Wang et al., 2015; Wang and Chen, 2016). The CC between pre-autumn ASI DY and WHD_{NCP} DY was calculated (Fig. 8) and was significantly positive around the Beaufort Sea ($73\text{--}78^\circ\text{N}$, $130\text{--}165^\circ\text{W}$). The area-averaged extent of ASI DY of the Beaufort Sea was selected as the fourth predictor (x_4), and its CC with WHD_{NCP} DY was

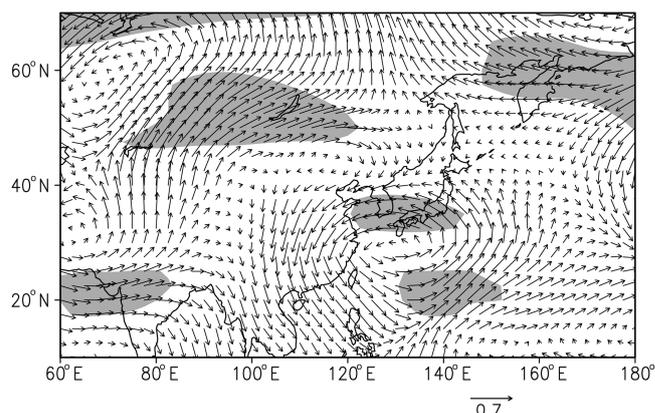


Figure 5. The CC between predictor x_2 and wind vector DY at 200 hPa in winter from 1980 to 2013. The shades indicate that the CC between the zonal wind DY and x_2 exceeded the 95 % confidence level.

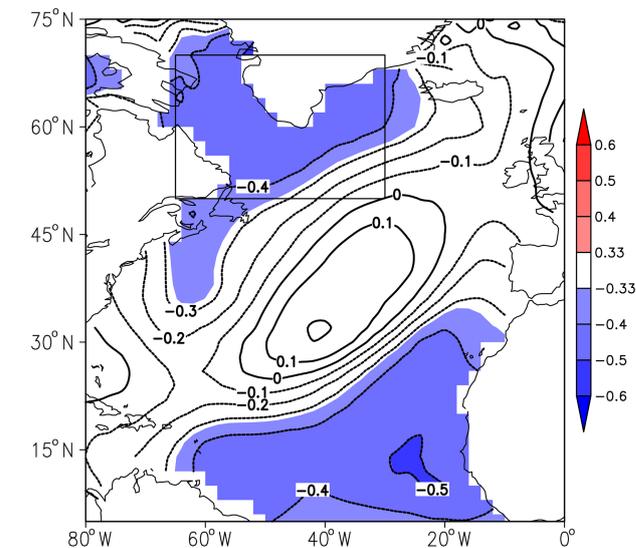


Figure 6. The CC between WHD_{NCP} DY and Atlantic SST DY in autumn from 1980 to 2013. The shades indicate that the CC exceeded the 95 % confidence level, and the rectangle represents the selected region ($50\text{--}70^\circ\text{N}$, $30\text{--}65^\circ\text{W}$) of predictor x_3 .

0.37 (above a 95 % confidence level). A positive center of geopotential height at 500 hPa was located over the Central Siberian and Mongolian plateaus, and negative centers were distributed zonally from southern China to the subtropical Pacific (Fig. 9). Thus, the EAJS was weakened by the induced easterly and shifted northward, which illustrated less cold activities over the NCP (Yang et al., 2002) and generated more haze days.

Following SST, the soil moisture was 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

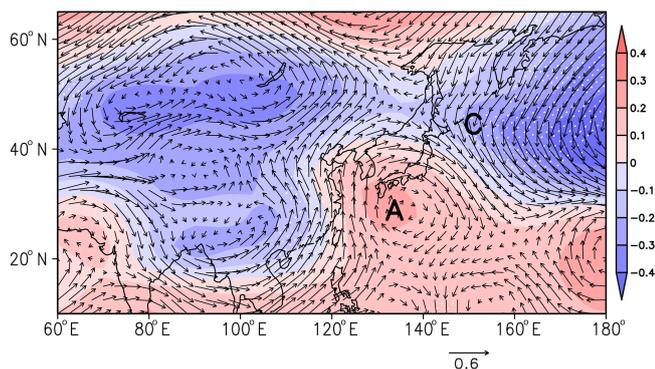


Figure 7. The CC between predictor x_3 ($\times -1$) and Z500 DY (shade)/850 hPa wind DY (arrows) in winter from 1980 to 2013. The dots indicate that the CC with meridional wind exceeded the 95 % confidence level. A and C represent anticyclone and cyclone, respectively.

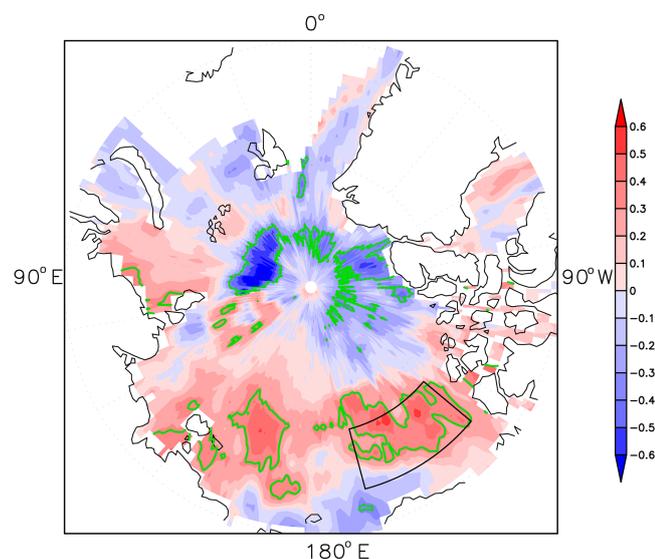


Figure 8. The CC between WHD_{NCP} DY and ASI DY in autumn from 1980 to 2013. The shades indicate that the CC exceeded the 95 % confidence level, and the box represents the selected region ($73\text{--}78^\circ\text{N}$, $130\text{--}165^\circ\text{W}$) of predictor x_4 .

(Yin et al., 2015a). Thus, 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 Economic Rim ($35\text{--}42^\circ\text{N}$, $117\text{--}127^\circ\text{E}$), 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 (Fig. 10). The CC between predictor x_5 and geopotential height at 500 hPa was distributed in a similar way as in Fig. 1. The positive EA/WR and WP phases and the negative EU phase was obvious and led to more WHD_{NCP} than in the previous year (Fig. 11). Being specific to local circulations, the cyclone over south China and the anticy-

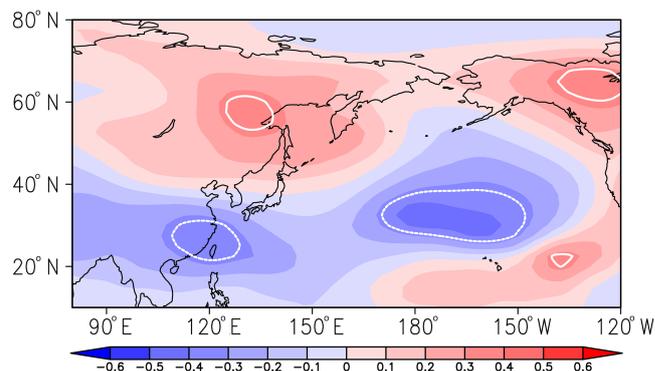


Figure 9. The CC between predictor x_4 and Z500 DY in winter from 1980 to 2013. The white curves indicate that the CC exceeded the 95 % confidence level.

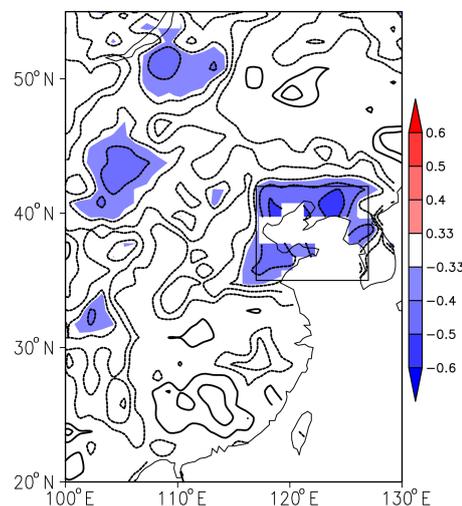


Figure 10. The CC between WHD_{NCP} DY and soil moisture DY in autumn from 1980 to 2013. The shades indicate that the CC exceeded the 95 % confidence level, and the rectangle represents the selected region ($35\text{--}42^\circ\text{N}$, $117\text{--}127^\circ\text{E}$) of predictor x_5 .

clone over the NCP and the west Pacific stimulated significant southeasters between them (figure omitted) that transported more moisture but decelerated the surface wind in the NCP. As shown in Fig. 12, the pre-summer soil moisture DY in the east of Mongolia ($48\text{--}52^\circ\text{N}$, $115\text{--}125^\circ\text{E}$) 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 (Fig. 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 the NCP could confine the vertical motion and the vertical diffusion

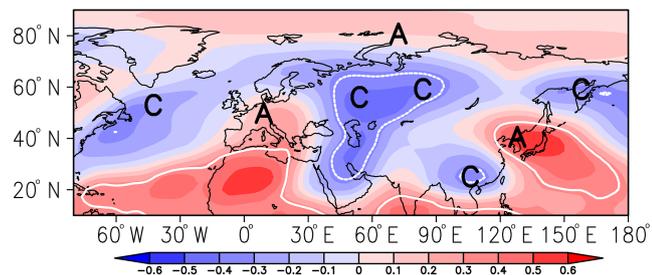


Figure 11. The CC between predictor x_5 ($\times -1$) and Z500 DY in winter from 1980 to 2013. The white curves indicate that the CC exceeded the 95 % confidence level. A and C represent anticyclone and cyclone, respectively.

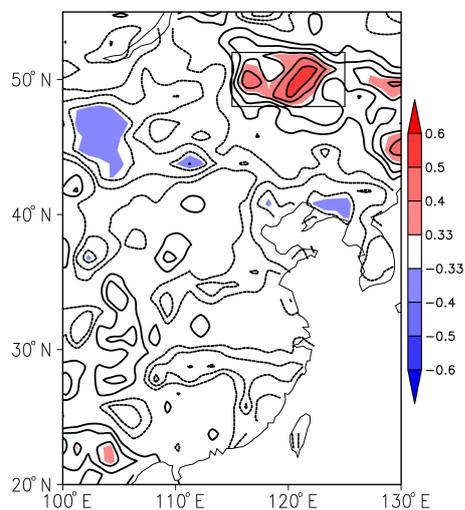


Figure 12. The CC between WHD_{NCP} DY and soil moisture DY in summer from 1980 to 2013. The shades indicate that the CC exceeded the 95 % confidence level, and the rectangle represents the selected region ($48\text{--}52^\circ \text{N}$, $115\text{--}125^\circ \text{E}$) of predictor x_6 .

of atmospheric particles and intensify the haze pollution over the NCP.

Recently, some studies documented that AAO could affect the east Asian climate through cross-equatorial flow, e.g., the Somali jet stream (Fan and Wang, 2004, 2006, 2007a, b). After the late 1990s, global sea level pressure and geopotential height at 300 hPa in boreal January were characterized by the concurrence of the Aleutian Low and the negative phase of the AAO (F. Li et al., 2015). We investigated the relationship between WHD_{NCP} DY and geopotential height at 850 hPa in the Southern Hemisphere and found that the distribution was remarkably similar to that of the negative phase of the AAO (Fig. 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 Fig. 15, the positive phases of the EA/WR and WP patterns were closely correlated with the negative phase of AAO and were responsible for more WHD_{NCP} than in the previous year. The anomalous

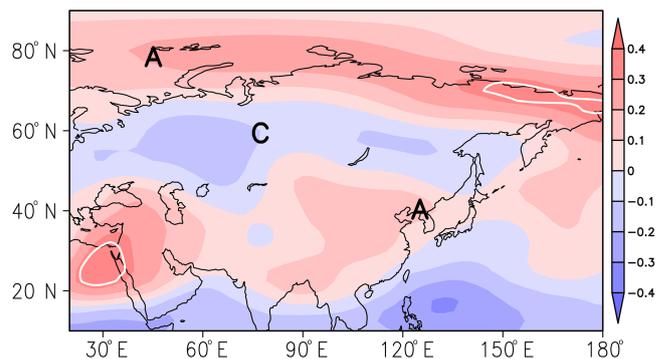


Figure 13. The CC between predictor x_6 and Z500 DY in winter from 1980 to 2013. The white curves indicate that the CC exceeded the 95 % confidence level. A and C represent anticyclone and cyclone, respectively.

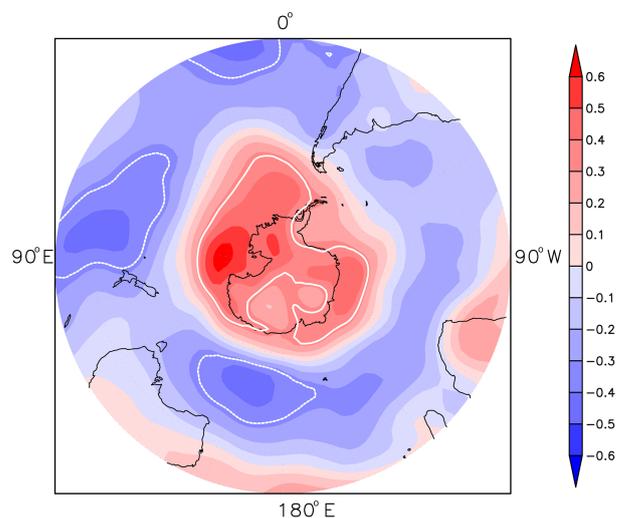


Figure 14. The CC between WHD_{NCP} DY and September–October Z850 DY from 1980 to 2013. The white curves indicate that the CC exceeded the 95 % confidence level.

anticyclone over the 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

In total, seven DY predictors ($x_1\text{--}x_7$) were chosen to build the seasonal prediction model (SPM) for WHD_{NCP} DY. Among the predictors were 21 types of pair combinations, of which only 5 pairs presented significant linear correlation. Thus, the multicollinearity would not be a problem when modeling with the MLR approach. Although the linear correlation between the predictand and each predictor

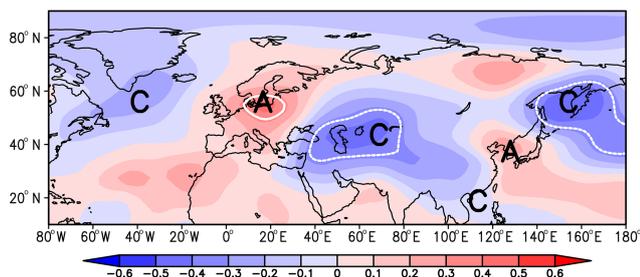


Figure 15. The CC between predictor x_7 ($\times -1$) and Z500 DY in winter from 1980 to 2013. The white curves indicate that the CC exceeded the 95 % confidence level.

was significant, the nonlinear 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.

The WHD_{NCP} DY showed obvious features of biennial oscillation (Fig. 16), 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 1). The percentage of the same sign (i.e., meaning 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 minimum WHD_{NCP} DY but did not adequately capture the continuous positive value after 2011 (Fig. 16a). The fitted WHD_{NCP} DY from 2011 to 2013 varied similarly with the one 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 nonlinear relationship with smooth functions. The SPM_{GAM} of WHD_{NCP} DY was as follows: $\text{DY} \times 10 = -2.164s(x_1) + 2.036s(x_2) - 1.721x_3 + 2.588s(x_4) - 2.157s(x_5) + 2.187x_6 - 2.506x_7$. 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 process the nonlinear 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, accounting for 54 % of the total variance (Table 1). The percentage of the same sign of SPM_{GAM} results was 73.5 %, which was close to the re-

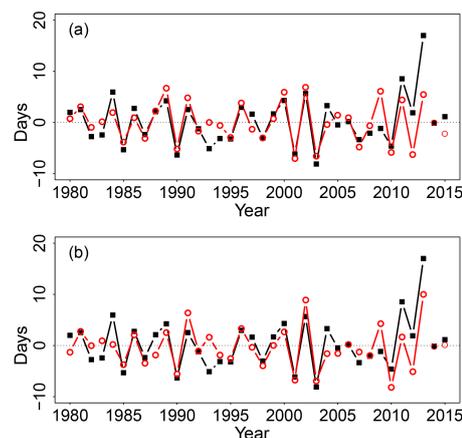


Figure 16. The temporal variation of measured (black) WHD_{NCP} DY, MLR (red, **a**) and GAM (red, **b**) cross-validation fitted WHD_{NCP} DY from 1980 to 2013. The results for 2014 and 2015 represent the measured (black square) and predicted (red hollow circle) WHD_{NCP} DY.

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.

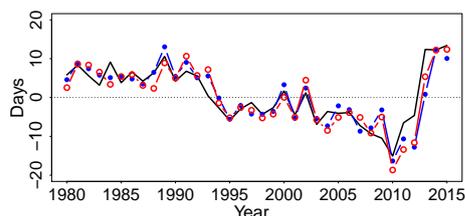
	MLR_s	MLR_{CV}	GAM_s	GAM_{CV}
RMSE	2.39	3.39	1.56	3.38
MAE	1.75	2.37	1.10	2.58
CC	0.87	0.72	0.95	0.74
EV	76 %	53 %	90 %	54 %
Bias ₁₄	0.09		−0.07	
Bias ₁₅	−3.33		−1.01	

sult 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 (Fig. 16b). The predicted bias in 2014 and 2015 was -0.07 and -1.01 , respectively, and the results were slightly better than those from SPM_{MLR} . The CC between the bias of SPM_{MLR} and SPM_{GAM} 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, and the result was the final predicted WHD_{NCP} in 2012. In Fig. 17, the simulated 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 simulated WHD_{NCP} was 0.89 (0.87) and 0.90 (0.88), respectively. Both of these

Table 2. The predictors and their meaning. CC indicates the correlation coefficient between predictor and WHD_{NCP} DY from 1980 to 2013.

Predictors	Meaning	CC
x_1	Pre-autumn SAT DY from the Sea of Japan to Stanovoy Range	-0.47
x_2	Pre-autumn SST DY around Gulf of Alaska	0.47
x_3	Pre-autumn SST DY to the south of Greenland	-0.50
x_4	Pre-autumn ASI extent DY of the Beaufort Sea	0.37
x_5	Pre-autumn soil moisture DY of the Bohai Economic 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

**Figure 17.** 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.

prediction models could capture the interannual and interdecadal trend and the extremums. The percentage of the same sign of the anomalies from the two models was 100 %, meaning these two models could predict the sign of the WHD_{NCP} anomaly successfully. The 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 WP patterns and the negative phase of the EU pattern intensified the haze pollution by inducing positive anomalies over the NCP and the Sea of Japan. Finally, seven leading predictors were selected and 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 the same sign 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 the same sign 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 consistency 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 pollution, 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 mechanisms via which the external forcings stimulated such anomalous circulations were worthy of further study.

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 included here. There was no doubt that the human activities, especially the energy consumption, were the fundamental drivers for the increase 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 socioeconomic component of WHD_{NCP} varied slowly. This assumption could support the seasonal prediction of haze days in most of the years but was still a compromise. In certain years, especially the recent years, this pollutant emission proportion varied rapidly, which needed to be taken into account. The preceding autumn energy consumption should be a good choice but is difficult to measure, and its DY could be introduced into the developed models directly to improve the predictive skill.

6 Data availability

Atmospheric data are available from the NCEP/NCAR data archive: <http://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis.html> (NCEP/NCAR, 2016). SST data are downloaded from <http://www.esrl.noaa.gov/psd/data/gridded/data.noaa.ersst.v4.html> (NOAA, 2016). ASI concentration data can be downloaded from the Hadley Centre: <http://www.metoffice.gov.uk/hadobs/hadisst/> (HadISST, 2016). The monthly gridded soil moisture data (CPC, 2016a) and the AAO indices (CPC, 2016b) can be downloaded from NOAA's Climate Prediction Center: <http://www.cpc.ncep.noaa.gov/products>. The ground observations are from the website <http://data.cma.cn/> (CMA, 2016).

Acknowledgements. This research was supported by the National Natural Science Foundation of China (grants 41421004 and 41210007) and CAS-PKU partnership program.

Edited by: A. Ding

Reviewed by: two anonymous referees

References

- Barnston, A. G. and Livezey, R. E.: Classification, seasonality and persistence of low frequency atmospheric circulation patterns, *Mon. Weather Rev.*, 115, 1083–1126, 1987.
- Chen, H. P. and Wang, H. J.: Haze days in North China and the associated atmospheric circulations based on daily visibility data from 1960 to 2012, *J. Geophys. Res.-Atmos.*, 120, 5895–5909, doi:10.1002/2015JD023225, 2015.
- CMA: China ground observation data sets, available at: <http://data.cma.cn/>, last access: 29 November 2016 (in Chinese).
- CPC: CPC Soil Moisture data sets, available at: <http://www.esrl.noaa.gov/psd/data/gridded/data.cpcsoil.html>, last access: 29 November 2016.
- CPC: CPC Daily Antarctic Oscillation Index, available at: http://www.cpc.ncep.noaa.gov/products/precip/CWlink/daily_ao_index/ao/ao_index.html, last access: 29 November 2016.
- Cressman, G.: An operational objective analysis system, *Mon. Weather Rev.*, 87, 367–374, 1959.
- Czaja, A. and Frankignoul, C.: Influence of the North Atlantic SST on the atmospheric circulation, *Geophys. Res. Lett.*, 26, 2969–2972, 1999.
- Ding, Y. H. and Liu, Y. J.: 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*, 57, 36–46, 2014 (in Chinese).
- Fan, K. and Tian, B. Q.: Prediction of wintertime heavy snow activity in Northeast China, *Chinese Sci. Bull.*, 58, 1420–1426, 2013.
- Fan, K. and Wang, H. J.: Antarctic oscillation and the dust weather frequency in North China, *Geophys. Res. Lett.*, 31, L10201, doi:10.1029/2004GL019465, 2004.
- Fan, K. and Wang, H. J.: Studies of the relationship between Southern Hemispheric atmospheric circulation and climate over East Asia, *Chinese J. Atmos. Sci.*, 20, 402–412, 2006.
- Fan, K. and Wang, H. J.: Simulation on the AAO anomaly and its influence on the Northern Hemispheric circulation in boreal winter and spring, *Chinese J. Geophys.*, 50, 397–403, 2007a.
- Fan, K. and Wang, H. J.: Dust storms in North China in 2002: A case study of the low frequency oscillation, *Adv. Atmos. Sci.*, 24, 15–23, 2007b.
- Fan, K., Wang, H. J., and Choi, Y. J.: A physically-based statistical forecast model for the middle-lower reaches of Yangtze River Valley summer rainfall, *Chinese Sci. Bull.*, 54, 602–609, 2008.
- Guo, W. D., Ma, Z. G., and Wang, H. J.: Soil moisture – an important factor of seasonal precipitation prediction and its application, *Climatic and Environmental Research*, 12, 20–28, 2007.
- HadISST: Hadley Centre Sea Ice and Sea Surface Temperature data sets, available at: <http://www.metoffice.gov.uk/hadobs/hadisst/>, last access: 29 November 2016.
- Hastie, T. J. and Tibshirani, R. J.: *Generalized Additive Models*, Chapman & Hall, London, UK, 1990.
- Huang, Y. Y., Wang, H. J., and Fan, K.: Improving the Prediction of the Summer Asian-Pacific Oscillation Using the Interannual Increment Approach., *J. Climate*, 27, 8126–8134, doi:10.1175/JCLI-D-14-00209.1, 2014.
- Huug, D., Huang, J., and Fan, Y.: Performance and Analysis of the constructed analogue method applied to US soil moisture applied over 1981–2001, *J. Geophys. Res.*, 108, 1–16, 2003.
- 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, K. C., Ropelewski, C., Wang, J., Jenne, R., and Joseph, D.: The NCEP/NCAR 40-year reanalysis project, *B. Am. Meteorol. Soc.*, 77, 437–471, doi:10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2, 1996.
- Li, F., Wang, H. J., and Gao, Y. Q.: Modulation of Aleutian low and Antarctic oscillation co-variability by ENSO, *Clim. Dynam.*, 44, 1245–1256, 2015.
- Li, Q., Zhang, R. H., and Wang, Y.: 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, 346–354, doi:10.1002/joc.4350, 2015.
- Mo, K. C.: Relationships between Low-Frequency Variability in the Southern Hemisphere and Sea Surface Temperature Anomalies, *J. Climate*, 13, 3599–3610, 2000.
- NCEP/NCAR: NCEP/NCAR Reanalysis data sets, available at: <http://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis.html>, last access: 29 November 2016.
- NOAA: NOAA Extended Reconstructed Sea Surface Temperature (SST) V4 data sets, available at: <http://www.esrl.noaa.gov/psd/data/gridded/data.noaa.ersst.v4.html>, last access: 29 November 2016.
- Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., Kent, E. C., and Kaplan, A.: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century, *J. Geophys. Res.*, 108, 4407 doi:10.1029/2002JD002670, 2003.
- Shi, X. L.: The impact of mid-high latitude oceans on climate change over north and northeast China, Ocean University of China, Qing Dao, China, 1 p., 2014.
- Smith, T., Reynolds, R., Peterson, T., and Lawrimore, J.: Improvements to NOAA's historical merged land–ocean surface temperature analysis (1880–2006), *J. Climate*, 21, 2283–2296, 2008.

- Tian, B. Q. and Fan, K.: A skillful prediction model for winter NAO based on Atlantic sea surface temperature and Eurasian snow cover, *Weather Forecast.*, 30, 197–204, 2015.
- Wallace, J. M. and Gutzler, D. S.: Teleconnection in the geopotential height field during the Northern Hemisphere winter, *Mon. Weather Rev.*, 109, 784–812, doi:10.1175/1520-0493(1981)109<0784:TITGHF>2.0.CO;2, 1981.
- Wang, H.-J. and Chen, H.-P.: Understanding the recent trend of haze pollution in eastern China: roles of climate change, *Atmos. Chem. Phys.*, 16, 4205–4211, doi:10.5194/acp-16-4205-2016, 2016.
- Wang, H. J., Zhou, G. Q., and Zhao, Y. An effective method for correcting the seasonal-interannual prediction of summer climate anomaly, *Adv. Atmos. Sci.*, 17, 234–240, 2000.
- Wang, H. J., Fan, K., Lang, X. M., Sun, J. Q., and Chen, L. J.: Advances in climate prediction theory and technique of China, Beijing, China Meteorological Press, 120–140, 2012 (in Chinese).
- Wang, H. J., Chen, H. P., and Liu, J. P.: Arctic sea ice decline intensified haze pollution in eastern China, *Atmos. Ocean. Sci. Lett.*, 8, 1–9, 2015.
- Wilks, D. S.: *Statistical methods in the atmospheric sciences*, Academic press, Oxford, UK, 1 p., 2011.
- Xiao, D., Li, Y., Fan, S. J., Zhang, R. H., Sun, J. R., and Wang, Y.: Plausible influence of Atlantic Ocean SST anomalies on winter haze in China, *Theor. Appl. Climatol.*, 122, 249–257, 2015.
- Yang, S., Lau, K. M., and Kim, K. M.: Variations of the East Asian Jet Stream and Asian–Pacific–American Winter Climate Anomalies, *J. Climate*, 15, 306–325, 2002.
- Yee, T. and Mitchell, N.: Generalized additive models in plant ecology, *J. Veg. Sci.*, 2, 587–602, 1991.
- Yin, Z. C. and Wang, H. J.: The relationship between the subtropical Western Pacific SST and haze over North-Central North China Plain, *Int. J. Climatol.*, 36, 3479–3491, doi:10.1002/joc.4570, 2016.
- Yin, Z. C., Yuan, D. M., Ding, D. P., and Xie, Z.: Statistical prediction based on meteorology of *Cotinus Cogggyria* leaves discoloration-day in the Fragrant hill, *Meteorological Monthly*, 40, 229–233, 2014 (in Chinese).
- Yin, Z. C., Wang, H. J., and Guo, W. L.: Climatic change features of fog and haze in winter over North China and Huang-Huai Area, *Science China Earth Sciences*, 58, 1370–1376, 2015a.
- Yin, Z. C., Wang, H. J., and Yuan, D. M.: 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, 1395–1400, 2015b (in Chinese).
- Zhang, R. H., Li, Q., and Zhang, R. N.: Meteorological conditions for the persistent severe fog and haze event over eastern China in January 2013, *Science China Earth Sciences*, 57, 26–35, doi:10.1007/s11430-013-4774-3, 2014.
- Zhao, N., Yin, Z. C., and Wu, F.: Characteristics of persistent fog and haze process and its forming reason in Beijing, *J. Meteorol. Environ.*, 30, 15–20, 2014 (in Chinese).
- Zhou, M. Z. and Wang, H. J.: Late winter sea ice in Bering Sea: predictor for maize and rice production in Northeast China, *J. Appl. Meteorol. Clim.*, 53, 1183–1192, 2014.