ADVANCES IN ATMOSPHERIC SCIENCES, VOL. 30, NO. 5, 2013, 1343–1352

Statistical Guidance on Seasonal Forecast of Korean Dust Days over South Korea in the Springtime

Keon Tae SOHN*

Pusan National University, Busan, 609-735*, Korea*

(Received 26 May 2012; revised 12 November 2012; accepted 12 December 2012)

ABSTRACT

This study aimed to develop the seasonal forecast models of Korean dust days over South Korea in the springtime. Forecast mode was a ternary forecast (below normal, normal, above normal) which was classified based on the mean and the standard deviation of Korean dust days for a period of 30 years (1981–2010). In this study, we used three kinds of monthly data: the Korean dust days observed in South Korea, the National Center for Environmental Prediction in National Center for Atmospheric Research (NCEP/NCAR) reanalysis data for meteorological factors over source regions of Asian dust, and the large-scale climate indices offered from the Climate Diagnostic Center and Climate Prediction Center in NOAA. Forecast guidance consisted of two components; ordinal logistic regression model to generate trinomial distributions, and conversion algorithm to generate ternary forecast by two thresholds. Forecast guidance was proposed for each month separately and its predictability was evaluated based on skill scores.

Key words: Korean dust days, ternary forecast, logistic regression, NCEP/NCAR reanalysis data, large-scale climate indices

Citation: Sohn, K. T., 2013: Statistical guidance on seasonal forecast of Korean dust days over South Korea in the springtime. *Adv. Atmos. Sci.*, **30**(5), 1343–1352, doi: 10.1007/s00376-012-2112-x.

1. Introduction

Asian dust is a meteorological event in which dust particles are uplifted into the air by the strong surface wind in the source regions in the northern China and Mongolia, and are transported far away and deposited on the ground. Asian dust events that deposit dust in South Korea (called Korean dust events) may lead to severe economic and social damages. According to the National Institute of Meteorological Research (NIMR) in the Korea Meteorological Administration (KMA), the frequency and variation of Asian dust events are increasing gradually.

The source regions of Asian dust, that affect the Korean peninsula, cover most of northern China and Mongolia including the arid and semi-arid area of sand desert. Zhang et al. (2008) divided the source regions of Asian which were the main contributors to long-lived dusts in northern China and Mongolia, into five regions. Lim and Chun (2006) divided these source regions into three regions (Gobi region, Inner Mongolia, and the northeast China) and showed that the source regions of Asian dust that affect the Korean peninsula are gradually extending eastward. Chun et al. (2006) also showed that source regions of Asian dust have been expanding to Northeast China year by year because of the desertification in Northeast China due to drought, overgrazing and change of atmospheric circulation. Kurosaki and Mikami (2003) showed that the Asian dust events increased remarkably in the eastern part of the Asian Continent after 2000. Chun and Lim (2004) concluded that the intensity and occurrence of Asian dust have been increased gradually since 2000 and they have started to occur earlier in the year than before. Tian et al. (2007) examined the relationship between Asian dust frequency observed in Japan and that in northern China with long-term observations and the National Center for Environmental Prediction in National Center for Atmospheric Research (NCEP/NCAR) reanalysis data.

The KMA began to announce the seasonal forecast of Asian dust events since 2001. Since 2008, objective models have been utilized for the seasonal forecast of Korean dust days over South Korea. The forecast

*Corresponding author: Keon Tae SOHN, ktsohn@pusan.ac.kr

modeling is changed every year in order to improve forecast accuracy. Sohn et al. (2009a) applied a principal component regression model to the quantitative forecast and categorical forecast of Korean dust days (KDD). In this paper, KDD means the spatial average of monthly Korean dust days observed at 28 stations in South Korea. However, the skill scores were low and because the principle components were linear combinations of several kinds of data, they were difficult to explain. Sohn et al. (2010) and Park et al. (2011) considered the large-scale climate indices as additional predictors. Their accuracy of forecast is much improved; however, it was well-fit before the year 2000 but not after 2001.

This study aimed at improving the seasonal forecast model of KDD for each month in the springtime separately. In this study, we considered three kinds of data to examine the relationships among dust deposition, dust emission, and dust transport based on statistical analysis. The first one was meteorological data to estimate dust emission. Following the method of Lim and Chun (2006), we used meteorological factors which are considered to be important factors for the occurrence of Asian dust, in our study. As potential predictors, we used the NCEP/NCAR reanalysis data (hereafter, referred to as NNR) for ground temperature, rainfall amount, snowfall amount and ground wind speed. The second one was large-scale climate indices for the dust transport from source regions to the Korean peninsula. Schwing et al. (2002) and Kim et al. (2008) indicated that atmospheric and oceanic teleconnections govern many climate variations, the cause and effect relationships between these teleconnections are not well understood. In those cases, large-scale climate indices may be useful to explain the relationships. We examined whether large-scale climate indices can explain the dust transport from source regions to South Korea. The third one is the KDD.

In this study, we used a ternary forecast (below normal, normal and above normal) mode. It is well known that logistic regression model is useful to generate multinomial distribution when the target variable has a categorical value including ordinal scaling data (Myers et al., 2002). Sohn et al. (2009b) applied a logistic regression model to the categorical forecast of heavy snowfall. It is easy to explain the effect of predictors via odds ratio using this type of model. In this study, the logistic regression model generated probability forecast in the form of a trinomial distribution (p_0, p_1, p_2) where p_0 is P (below normal), p_1 is P (normal) and p_2 is P (above normal). The trinomial distribution was converted into a ternary forecast using thresholds.

Forecast guidance consists of the estimated ordinal logistic regression model and the conversion algorithm to generate categorical forecast. The predictability of forecast guidance is evaluated by assessment measures called skill scores. Which can be of various types (von Storch and Zwiers, 1999; Sohn and Park, 2008). In this study, skill scores for the binary case and ternary case were used to assess the forecast guidance and find the optimal guidance.

In section 2, three kinds of data are presented and analyzed. Temporal variations in KDD and NNR are examined. In section 3, the potential predictors are selected based on correlation analysis between KDD and predictors. Skill scores are introduced and forecast strategy and guidance are explained. The forecast guidance is proposed for each month separately and the results are summarized in section 4.

2. Data

2.1 *Korean dust days*

For this study, we used KDD over South Korea for a period of 30 years (1981–2010). Table 1 shows the 30-year mean of KDD for each month, from which it is clear that KDD mainly occurred (about 85%) in the springtime. Our study focused on KDD in the springtime. Figure 1 illustrates time series plots of KDD for three month (March–May). It is evident that the change patterns of KDD were different from each other. And the three KDD series were not significantly correlated. It led us to develop the forecast guidance for each month separately.

The interdecadal variations in KDD series given in Table 2, showed that the mean levels (0.44, 0.99, and 3.74) of KDD increased in March. As shown in Fig. 1, KDD increased remarkably after the middle of 1990s, and its significant slope was 0.145 days per year. The

Table 1. 30-year mean of KDD and occurrence share for each month (units: d).

Month	30-year mean	Occurrence share (%)
Jan	0.20	3.3
Feb	0.14	2.3
Mar	1.72	28.6
Apr	2.35	39.1
May	1.01	16.8
Jun	0.00	0.0
Jul	0.00	0.0
Aug	0.00	0.0
Sep	0.01	0.2
Oct	0.11	1.8
Nov	0.22	3.7
Dec	0.25	4.2
Total	6.01	100.0

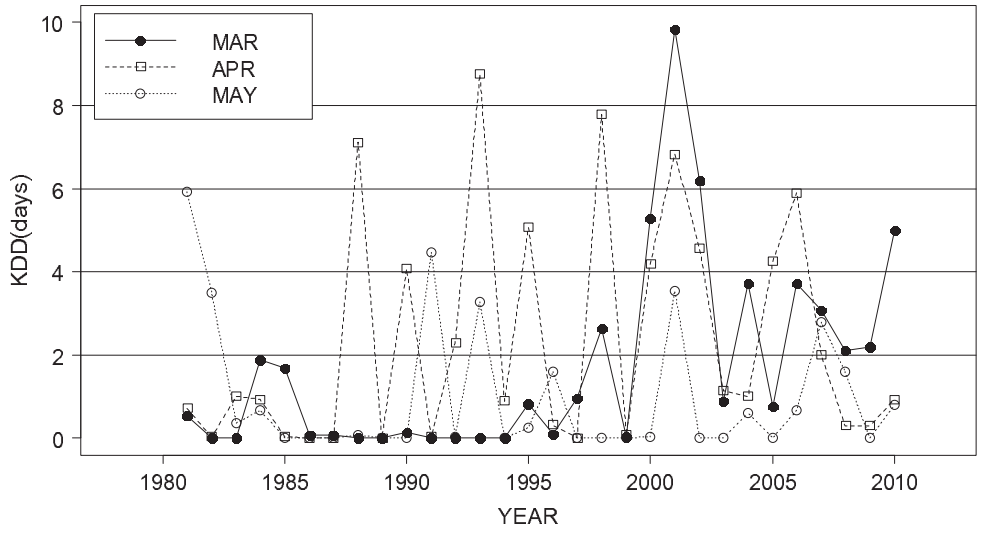

Fig. 1. Time series plots of Korean dust days over South Korea for March (dot), April (square) and May (circle).

Table 2. Inter-decadal variation of KDD: (M and STD are the 10-year mean and standard deviation, respectively; units: d).

	Period						Trend	
	1981–1990		1991–2000		2001–2010			
Month	M	STD	M	STD	M	STD	Slope	p-value
Mar	0.44	0.73	0.99	1.73	3.74	2.73	0.14	0.002
Apr	1.39	1.05	2.94	3.33	2.72	2.44	0.06	0.343
May	1.05	2.02	0.97	1.63	1.00	1.26	−0.03	0.314

Table 3. Classification rules and frequency table of KDD for each month (M and STD are the 30-year mean and standard deviation, respectively; units: d).

Month	M	STD	M − 0.5STD	M + 0.5STD	Below normal	Normal	Above normal
Mar	1.72	2.36	0.54	2.90	6	7	7
Apr	2.35	2.74	0.98	3.72	9	3	8
May	1.01	1.61	0.21	1.82	10	6	4

mean levels (1.39, 2.94, and 2.72) of KDD increased further in April in the 1990s and slightly decreased in the 2000s. However, almost no change occurred in the mean level (1.05, 0.97, and 1.00) of KDD in May. The changes in standard deviation were similar to the corresponding mean levels.

The target value in this study was a ternary forecast which was classified following the classification rules in Table 3, for each month in the springtime separately. KDD said to be below normal if KDD is less than (M −0.5×STD), above normal if greater than (M + 0.5×STD), and normal if otherwise, where M and STD are the 30-year mean and 30-year standard deviation respectively.

2.2 *Meteorological factors in source regions*

According to Lim and Chun (2006), source regions of Asian dust that affected the Korean peninsula in the springtime, are divided into three source regions (Fig. 2). The three regions are located in dry arid region (A, Gobi region: 35°–45°N, 100°–110°E), semi-arid region (B, Inner Mongolia region: 40°–45°N, 110°–120°E), and cultivated region (C, the northeast part of China: 40°–50°N, 120°–125°E).

The regional mean of NNR for monthly mean of ground temperature, monthly rainfall amount, monthly snowfall amount and monthly mean of ground wind speed over three source regions for a period of 30

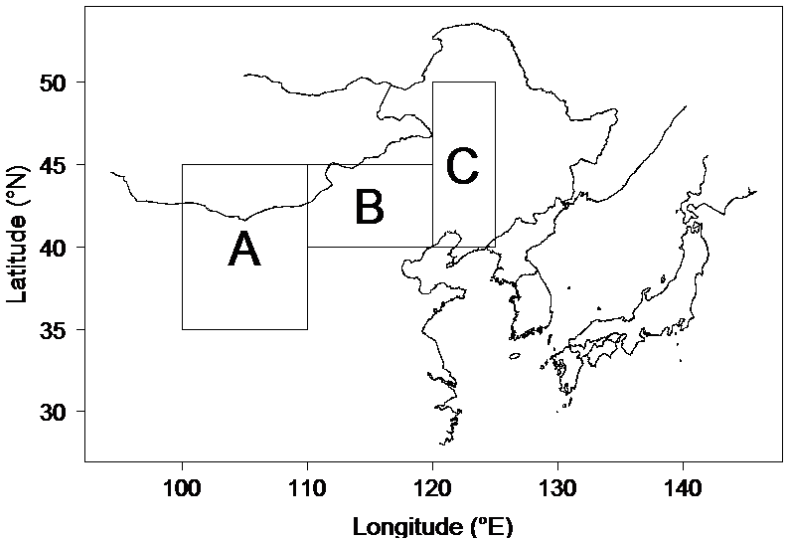

Fig. 2. Three source regions of Asian dust.

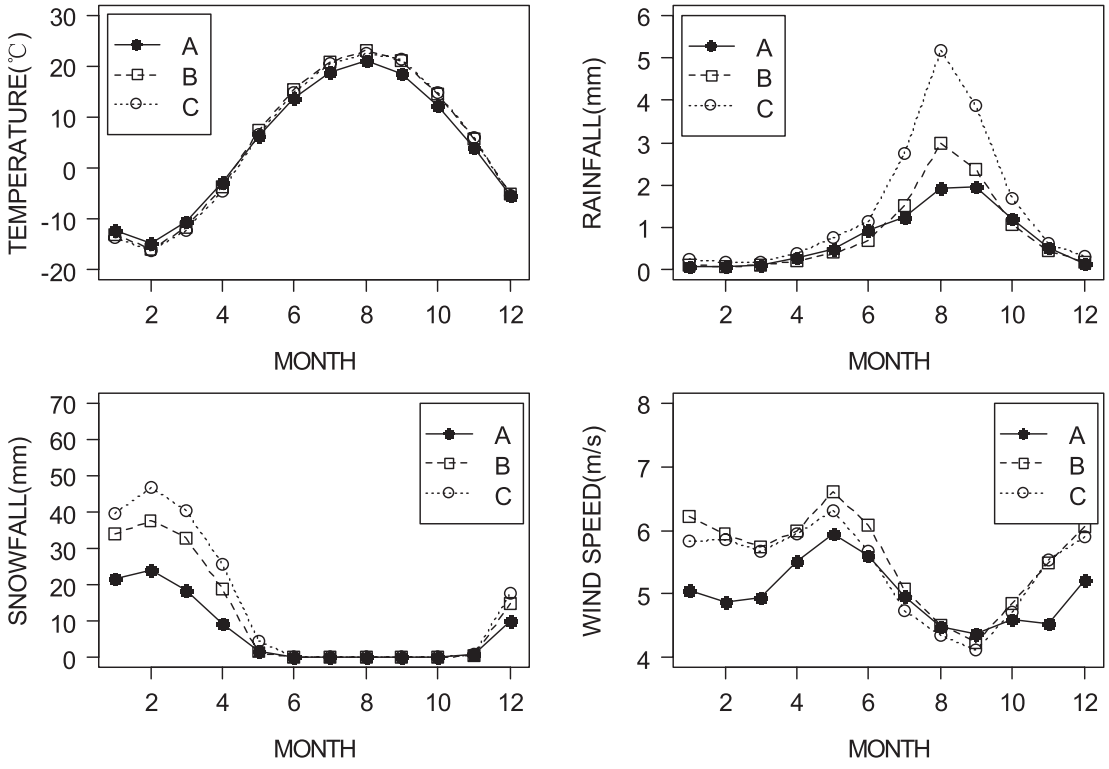

Fig. 3. 30-year means of NNR in three regions (A, B and C) for each factor and month.

years (1980–2009) were used as potential predictors. Figure 3 shows the 30-year means of NNR in three regions for each month. Ground temperatures in the three regions were similar to each other. Rainfall and snowfall amount in C were more than those in A and B, while the ground wind speed in B was higher than that in the others.

Table 4 shows the mean interdecadal variation in anomaly data of NNR. The ground temperature was found to have an increasing trend in all regions. Rainfall and snowfall amount showed a decreasing trend in all regions. Although the ground wind speed appeared to be decreasing in all regions, it did not show any significant trend. NNR indicated that all regions became drier and warmer, which might have increased floating dusts over the source regions; we considered these four

Table 4. Inter-decadal variation in anomaly data of NCEP/NCAR reanalysis data.

Period	Region	Factors			
		Temperature	Rainfall	Snowfall	Wind speed
1980–1989	A	−0.279	−0.006	2.019	0.558
	B	−0.376	0.052	3.550	0.478
	C	−0.409	0.133	4.553	0.036
1990–1999	A	0.359	0.011	1.279	−1.096
	B	0.195	0.096	2.784	−0.872
	C	0.073	0.140	4.132	−0.013
2000–2009	A	−0.080	−0.005	−3.298	0.539
	B	0.181	−0.148	−6.334	0.395
	C	0.336	−0.273	−8.684	−0.022

Table 5. Large-scale climate indices (Kim et al., 2008).

Abbr.	Full name with key reference
AO	Arctic Oscillation (Thompson and Wallace, 1998)
GML	Global Mean Land Ocean Temperature Index (Hansen et al., 1999)
MEI	Multivariate ENSO Index (Rasmusson and Carpenter, 1982)
NOI	Northern Oscillation Index (Schwing et al., 2002)
ONI	Oceanic Nino Index
PDO	Pacific Decadal Oscillation Index (Zhang et al. 1997)
PNA	Pacific/North American Pattern (Wallace and Gutzler, 1981)
SOI	Southern Oscillation Index (Rasmusson and Carpenter, 1982)
WPO	West Pacific Oscillation (Barnston and Livezey, 1987)

factors as predictors for KDD.

2.3 *Large-scale climate indices*

We also used nine climate indices (AO, GML, MEI, NOI, ONI, PDO, PNA, SOI, and WPO; see in Table 5) as additional predictors. They were obtained from the Climate Diagnostic Center (CDC) and Climate Prediction Center (CPC) in National Oceanic and Atmospheric Administration (NOAA) and updated on a monthly basis. The AO is an index of the pressure gradient between the polar and subpolar regions of the Northern Hemisphere. The GML is the global mean land/ocean temperature index. The MEI is calculated as the first principal component of six observed variables fields (SLP, zonal and meridional components of the surface wind, SST, SAT, total cloudiness fraction of the sky) over the tropical Pacific. The NOI is an index of climate variability based on the difference in SLP anomalies at the North Pacific High and near Darwin. The ONI is a three-month running mean of SST anomalies in the Niño3.4 region (5°N–5°S, 120°–170°W) based on the 1971–2000 base period. The PDO is the first principal component of monthly SST anomalies in the North Pacific Ocean. The PNA is an index derived from the formula of Wallace and Gutzler (1981) using 500-hPa geopotential height values. The SOI is an index of climate variability based on the difference in SLP anomalies at Tahiti and Darwin. The WPO identified through the application of a principal-component analysis involves changes to the wind pattern in the northwestern and north-central North Pacific.

3. Methodology

3.1 *Selection of predictors*

In NIMR (2011), forecast models were estimated using 30-year data (1981–2010), which were well-fit to the data before 2001, but did not show a good fit after 2001; this was because the patterns of KDD series changed after the middle of 1990s as shown in Fig. 1. To improve predictability, we decided to use 20-year data (1991–2010) for forecast model and 30-year data for classification rules.

As potential predictors, we use 12-month data (January-December in the previous year) of each factor because the KMA usually announces seasonal forecast for KDD in the springtime in February. The set of potential predictors contained as so many as 252 variables (3 regions×12 months×4 factors + 12 months×9 indices). To simplify the process, at first, we selected the secondary potential predictors (Table 6) among 252 variables, which were significantly correlated with KDD. Predictors were named by three components (region, factor, and month). For example, AT7 means

Table 6. Significant NNR factors and climate indices based on correlation analysis.

Month	Data period	NCEP/NCAR reanalysis data	Climate indices
Mar	Jan to Dec	AT7, BT7, CT7, AR3, BR1, BR7, CR1, CR9, AS10, AS12, BS12, CS12, BWS10	AO3, MEI3, NOI7, ONI4, PDO9, PNA7, SOI4, WPO5
Apr	Feb to Jan	AT5, AT10, CT7, BR5, AS3, CWS6	GML7, MEI7, NOI10, PNA9, SOI6, SOI7
May	Mar to Feb	CT4, AR12, AS3, AWS7, AWS11, BWS4, BWS7, CWS1, CWS7	AO5, NOI11, SOI5

Table 7. 3×3 cross table: values in parenthesis are weights and n_{ij} means the frequency that forecasted category is j when observed category is i.

	Forecast			
Observation	Below normal	Normal	Above normal	Total
Below normal	n_{00} (1.0)	n_{01} (0.5)	n_{02} (0.0)	$n_{0.}$
Normal	n_{10} (0.5)	n_{11} (1.0)	n_{12} (0.5)	$n_{1.}$
Above normal	n_{20} (0.0)	n_{21} (0.5)	n_{22} (1.0)	$n_{2.}$
Total	$n_{.0}$	$n_{.1}$	$n_{.2}$	N

the monthly mean of ground temperature over Region A in July in the previous year and AO3 means the Arctic oscillation index in March in the previous year.

3.2 *Assessment of forecast models*

The ternary forecast was assessed based on a 3×3 contingency table (see Table 7). We considered the thresholds that had good skill scores for both binary forecast and ternary forecast, since the above normal case was more serious than the normal or the below normal case. In this study, binary forecast was generated from ternary forecast in order to satisfy the concordance between two forecasts with merging the normal and the below normal cases into the normal cases.

In the binary forecast, we considered three skill scores based on a 2×2 contingency table. They were the hit rate (HR2), the probability of detection (POD2) and the false alarm rate (FAR2), and were computed by the following equations using the notation in Table 8.

$$\begin{aligned} \mathrm{HR2} &= (n_{00} + n_{01} + n_{10} + n_{11} + n_{22})/N \,, \\ \mathrm{POD2} &= n_{22}/(n_{20} + n_{21} + n_{22}) \,, \\ \mathrm{FAR2} &= n_{02}/(n_{02} + n_{12} + n_{22}) \,, \end{aligned}$$

where N is the number of total cases, HR2 is the exact forecast rate, POD2 is the conditional probability that above normal is forecasted given that above normal is observed, and FAR2 is the conditional probability that above normal is not observed given that above normal is forecasted. It is natural that the higher the values for HR2 and POD2 and the lower value for FAR2, the better the outcomes.

For the assessment of ternary forecast, we compute skill scores based on the 3×3 contingency table. Burrows (1991) assigned weights to the multicategorical forecast cases, so that a forecast that is different from the predictand in one category would be a better forecast than a forecast which is different in two categories, and so on. Following Burrows' procedure, we also used weights in Table 7, and computed the hit rate (HR3) and the weighted hit rate (wHR3) defined by the following equations using the notation in Table 7.

$$\begin{aligned} \mathrm{HR3} &= (n_{00} + n_{11} + n_{22})/N \,, \\ \mathrm{wHR3} &= (n_{00} + 0.5n_{01} + 0.5n_{10} + n_{11} + \\ &\quad 0.5n_{12} + 0.5n_{21} + n_{22})/N \,. \end{aligned}$$

It is natural that the higher the values for HR3 and wHR3, the better the outcomes. Note that HR3 is the hit rate in a 3×3 table whereas HR2 is the hit rate in

Table 8. 2×2 cross table with merging of below normal and normal to normal (n_{ij} means the frequency that forecasted category is j when observed category is i).

	Forecast		
Observation	Normal	Above normal	Total
Normal	$n_{00} + n_{01} + n_{10} + n_{11}$ (negative correction)	$n_{02} + n_{12}$ (false alarm)	$n_{0.} + n_{1.}$
Above normal	$n_{20} + n_{21}$ (miss)	n_{22} (hit)	$n_{2.}$
Total	$n_{.0} + n_{.1}$	$n_{.2}$	N

Table 9. 3×3 cross tables and skill scores of the maximum probability method and the threshold method for March.

	Forecast						
	Maximum prob. method			Threshold method			
Observation	Below normal	Normal	Above normal	Below Normal	Normal	Above normal	Total
Below normal	5	1	0	6	0	0	6
Normal	1	5	1	1	4	2	7
Above normal	0	2	5	0	0	7	7
Total	6	8	6	7	4	9	20

a 2×2 table.

To check the generalization ability of forecast guidance, we performed a leave-one-out cross-validation. The results of model validation were examined by the above skill scores on the basis of the 3×3 contingency table.

3.3 *Forecast guidance*

Ternary forecast was generated by a two-step process as follows. At first, an ordinal logistic regression model was utilized to generate trinomial distribution (p_0, p_1, p_2) using the following equations.

$$e_0 = \exp(b_0 + X); \quad e_1 = \exp(b_1 + X)\ ;$$
$$p_0 = e_0/(1 + e_0); \quad p_1 = e_1/(1 + e_1) - p_0\ ;$$
$$p_2 = 1 - p_0 - p_1\ ;$$

where b_0 and b_1 are constants, and X is a linear combination of final predictors.

Second, the generated trinomial distribution was converted into a ternary forecast. To generate a ternary forecast, we usually selected the category with the highest probability in the generated trinomial distribution. However, the threshold method might be better than the maximum probability method. For example, the skill scores of the maximum probability method were HR2 = 0.85, POD2 = 0.71, FAR2 = 0.17, HR3 = 0.75, wHR3 = 0.87, whereas those of the threshold method are HR2 = 0.90, POD2 = 1.00, FAR2 = 0.22, HR3 = 0.85 and wHR3 = 0.93 (Table 9). In this study, the threshold method was preferred to the maximum probability method. Two thresholds, T1 and T2, were used. A conversion of trinomial distribution (p_0, p_1, p_2) to a ternary forecast (0 = below normal, 1 = normal, 2 = above normal) were made using the followings:

IF $p_0 >$ T1 THEN
  forecast = 0 ;
ELSE IF $p_1/(p_1 + p_2) >$ T2 THEN
  forecast = 1 ;
ELSE forecast = 2 ;

Varying two thresholds from 0 to 1 independently, the above skill scores were computed for each pair of two thresholds and compared so as to determine the optimal thresholds. The proposed forecast guidance consisted of the estimated logistic regression model and the conversion algorithm using two optimal thresholds.

4. Results

We developed the statistical forecast guidance on ternary KDD in the springtime for each month separately. A three-grade ordinal logistic regression model was applied to generate a trinomial distribution using the final predictors given in Table 6. The optimal thresholds were determined based on skill scores. Some results were obtained in the following way.

First, the variable selection was performed by Fisher's scoring method offered by the statistical package called SAS. The final predictors, selected at the significance level 0.05, were the following: BS12, NOI7, and SOI4 for March; CT7, BR5, and NOI1 for April; and CT4, AR12, and AWS11 for May.

Second, the optimal thresholds were determined based on skill scores, which were 0.3 and 0.65 for March; 0.2 and 0.6 for April; and 05 and 0.5 for May.

Third, forecast guidance generated a binary forecast and a ternary forecast together. The proposed guidance for March was as follows:

$x = 0.1223\times$BS12 $- 1.8894\times$NOI7 $- 1.918\times$SOI4 ;
$e_0 = \exp(-8.5420 + x)$; $e_1 = \exp(-3.7485 + x)$;
$p_0 = e_0$ / $(1 + e_0)$; $p_1 = e_1$ / $(1 + e_1)$-p_0 ;
$p_2 = 1 - p_0 - p_1$.
IF $p_0 > 0.3$ THEN
  Ternary_forecast=0 ;
ELSE IF $p_1/(p_1 + p_2) > 0.65$ THEN
  Ternary_forecast =1 ;
ELSE Ternary_forecast =2 ;
IF Ternary_forecast=0 OR Ternary_forecast=1
  THEN Binary_forecast=1 ;
ELSE Binary_forecast=2 .

Table 10. Contingency tables for binary forecast using whole data (1: normal, 2: above normal).

	Forecast								
	March			April			May		
Observation	1	2	Total	1	2	Total	1	2	Total
1	11	2	13	10	2	12	16	0	16
2	0	7	7	0	8	8	0	4	4
Total	11	9	20	10	10	20	16	4	20

Table 11. Contingency tables for cross validation of binary forecast (1: normal, 2: above normal).

	Forecast								
	March			April			May		
Observation	1	2	Total	1	2	Total	1	2	Total
1	12	1	13	11	1	12	16	0	16
2	2	5	7	0	8	8	0	4	4
Total	14	6	20	11	9	20	16	4	20

Table 12. Contingency tables for ternary forecast using whole data (0: below normal, 1: normal, 2: above normal).

	Forecast											
	March				April				May			
Observation	0	1	2	Total	0	1	2	Total	0	1	2	Total
0	6	0	0	6	9	0	0	9	9	1	0	10
1	1	4	2	7	1	0	2	3	1	5	0	6
2	0	0	7	7	0	0	8	8	0	0	4	4
Total	7	4	9	20	10	0	10	20	10	6	4	20

Table 13. Contingency tables for cross validation of ternary forecast (0: below normal, 1: normal, 2: above normal).

	Forecast											
	March				April				May			
Observation	0	1	2	Total	0	1	2	Total	0	1	2	Total
0	5	1	0	6	9	0	0	9	9	1	0	10
1	1	5	1	7	1	1	1	3	1	5	0	6
2	0	2	5	7	0	2	6	8	0	0	4	4
Total	6	8	6	20	6	8	6	20	10	6	4	20

Table 14. Prediction results for the years 2011 and 2012.

	Year					
	2011			2012		
Month	Dust days	Actual	Forecast	Dust days	Actual	Forecast
March	2.5	above	normal	0.3	below	normal
April	0.3	below	below	0.0	below	below
May	5.7	above	above	0.0	below	normal

The proposed guidance for April was as follows:

$x = -3.5088\times\text{CT7} - 11.1385\times\text{BR5} + 2.2744\times\text{NOI10}$;
$e_0 = \exp(85.8864 + x)$; $e_1 = \exp(88.0663 + x)$;
$p_0 = e_0 \,/\, (1 + e_0)$; $p_1 = e_1 \,/\, (1 + e_1) - p_0$.
$p_2 = 1 - p_0 - p_1$;
IF $p_0 > 0.2$ THEN
Ternary_forecast=0 ;

ELSE IF $p_1/(p_1 + p_2) > 0.6$ THEN
Ternary_forecast =1 ;
ELSE Ternary_forecast =2 ;
IF Ternary_forecast=0 OR Ternary_forecast=1THEN
Binary_forecast=1 ;
ELSE Binary_forecast=2 .

The proposed guidance for May was as follows:

$x = 0.9213\times\text{CT4} + 41.0877\times\text{AR12} - 1.9697\times\text{AWS11}$;
$e_0 = \exp(x)$; $e_1 = \exp(4.176 + x)$;
$p_0 = e_0 \,/\, (1 + e_0)$; $p_1 = e_1 \,/\, (1 + e_1) - p_0$;
$p_2 = 1 - p_0 - p_1$;
IF $p_0 > 0.5$ THEN Ternary_forecast=0 ;
ELSE IF $p_1/(p_1 + p_2) > 0.\ 5$ THEN
Ternary_forecast =1 ;
ELSE Ternary_forecast =2 ;
IF Ternary_forecast=0ORTernary_forecast=1THEN
Binary_forecast=1 ;
ELSE Binary_forecast=2 ;

Forth, the results of binary forecast were summarized in 2×2 contingency tables (Table 10). According to contingency tables, the computed skill scores were as follows: HR2 = 0.9, POD2 = 1.0, FAR2 = 0.22 for March; HR2 = 0.9, POD2 = 1.0, FAR2 = 0.2 for April; and HR2 = 1.0, POD2 = 1.0, FAR2 = 0.22 for May. Table 11 shows the results of cross-validation for binary forecast as follows; HR2 = 0.85, POD2 = 0.71, FAR2 = 0.16 for March; HR2 = 0.85, POD2 = 0.75, FAR2 = 0.14 for April; and HR2 = 1.0, POD2 = 1.0, FAR2 = 0.22 for May. These scores indicate that each forecast guidance may be reliable and useful.

Fifth, the results of ternary forecast were summarized in the 3×3 contingency tables (Table 12). Based on cross tables, the computed skill scores were as follows: HR3 = 0.85, wHR3 = 0.93 for March; HR3 = 0.85, wHR3 = 0.93 for April; and HR3 = 0.9, wHR3 = 0.95 for May. Table 13 shows the results of cross-validation for ternary forecast and their skill scores were as follows: HR3 = 0.75, wHR3 = 0.88 for March; HR3 = 0.8, wHR3 = 0.9 for April; and HR3 = 0.9, wHR3 = 0.95 for May. These scores indicate that the proposed forecast guidance may be reliable.

Sixth, we applied the proposed models to forecast KDD in the years 2011 and 2012. The results are summarized in Table 14. In 2011, the actual category and forecasted category of KDD were (normal, above normal), (below normal, below normal) and (above normal, above normal) for March, April, and May, respectively. The rate of correct forecast was 67%. In 2012, they are (below normal, normal), (below normal, below normal) and (below normal, normal) for March, April and May respectively. Unfortunately, the rate of correct forecast was only 33% in this case. In April and May 2012, North Pacific high enlarged unusually to the Korean peninsula, which might have suppressed dust transport to South Korea.

5. Concluding remarks

To develop a separate ternary forecast of KDD for each month in the springtime, we considered the following strategy. First, we used 20-year data (1991–2010). Second, various skill scores (HR2, POD2, FAR2, HR3, and wHR3) were considered for the assessment of binary and ternary forecasts. Third, we used two thresholds to generate categorical forecast. Fourth, we propose the forecast guidance, which can be used to generate both the binary and the ternary forecasts together, in order to satisfy the concordance between two modes for each month.

The skill scores of proposed forecast guidance were found to be better than before. HR3 scores were 0.85 for March, 0.85 for April and 0.9 for May. We recommend that binary forecast might be more useful than ternary forecast because above normal cases are more important than normal and below normal cases.

In the future, we will investigate the relationships among large-scale climate indices in order to find how they affect Asian dust events.

Acknowledgements. This work is supported by the project "Development and Application of the Techniques on Asian Dust Monitoring and Prediction" of National Institute of Meteorological Research/Korea Meteorological Administration in 2011.

REFERENCES

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

Burrows, W. R., 1991: Objective guidance for 0-24 hour and 24-48 hour mesoscale forecast on lake-effect snow using CART. *Wea. Forecasting*, **6**, 357–378.

Chun, Y., and J. Y. Lim, 2004: The recent characteristics of Asian Dust and haze events in Seoul, Korea. *Meteor. Atmos. Phys.*, **87**, 143–152.

Chun, Y., S. W. Park, and Y. S. Jeong, 2006: *Development on the Asian Dust Monitoring and Prediction Techniques.* Report No. MR053A15, National Institute of Meteorological Research / Korea Meteorological Administration, 321pp.

Hansen, J., R. Ruedy, J. Glascoe and M. Sato, 1999: GISS analysis of surface temperature change. *J. Geophys. Res.*, **104**, 30997–31022.

Kim, Y., M. Kim, and W. Lee, 2008: An Investigation of large-scale climate indices with the influence on temperature and precipitation variation in Korea. *Atmosphere*, **18**, 83–95.

Kurosaki, Y., and M. Mikami, 2003: Recent frequent dust events and their relation to surface wind in East Asia. *Geophys. Res. Lett.*, **30**, doi:10.1029/2003GL017261.

Lim, J. Y., and Y. Chun, 2006: The characteristics of Asian dust events in Northeast Asia during the springtime from 1993 to 2004. *Global and Planetary Change*, **52**, 231–247.

Myers, R. H., D. C. Montgomery, and G. G. Vining, 2002: *Generalized Linear Models with Applications in Engineering and the Sciences.* John Wiley & Sons, New York. 342pp.

Park, G. J., K. T. Sohn, J. W. Kim, B. H. Park, and M. Y. Jeon, 2011: *Development and Application of the Techniques on Asian Dust Monitoring and Predictions.* Report No. 11-1360395-000243-01, National Institute of Meteorological Research/ Korea Meteorological Administration, 76pp.

Rasmusson, E. M., and T. H. Carpenter, 1982: Variations in tropical sea surface temperature and surface wind fields associated with southern oscillation/El Nino. *Mon. Wea. Rev.*, **110**, 354–384.

Schwing, F. B., T. Murphree, and P. M. Green, 2002: The northern oscillation index (NOI): a new climate index for the northeast Pacific. *Progress in Oceanography*, **53**, 115–139.

Sohn, K. T., and S. M. Park, 2008: Guidance on the choice of threshold for binary forecast modeling. *Adv. Atmos. Sci.*, **25**, 83–88, doi: 10.1007/s00376-008-0083-8.

Sohn, K. T., H. J. Lee, S. Lee, and B. Kim, 2009a: *Development and Application of the Techniques on Asian Dust Monitoring and Predictions.* Report No. 11-1360395-000172-01, National Institute of Meteorological Research/Korea Meteorological Administration, 110pp.

Sohn, K. T., J. H. Lee, and Y. S. Cho, 2009b: Ternary forecast of heavy snowfall in Honam area, Korea. *Adv. Atmos. Sci.*, **26**, 327-332, doi: 10.1007/s00376-009-0327-2.

Sohn, K. T., J. Lee, S. Lee, B. Kim, J. Yoon, E. Gam, and J. H. Park, 2010: *Development and Application of the Techniques on Asian Dust Monitoring and Predictions.* Report No. 11-1360395-000209-01, National Institute of Meteorological Research / Korea Meteorological Administration, 104pp.

Tian, S. F., M. Inoue, and M. Du, 2007: Influence of dust storm frequency in Northern China on fluctuations of Asian dust frequency observed in Japan. *SOLA*, **3**, 121–124.

Thompson, D. W. J., and J. M. Wallice, 1998: The Arctic Oscillation signature in the wintertime geopotential height and temperature fields. *Geophys. Res. Lett.*, **25**, 1297–1300.

von Storch, H., and F. W. Zwiers, 1999: *Statistical Analysis in Climate Research.* Cambridge University Press, Cambridge, 484pp.

Wallace, J. M., and D. S. Gutzler, 1981: Teleconnections in the 500 mb geopotential height field during the Northern Hemisphere winter. *Mon. Wea. Rev.*, **109**, 784–812.

Zhang, B., A. Tsunekawa, and M. Tsubo, 2008: Contributions of sandy lands and stony deserts to long-distance dust emission in China and Mongolia during 2000–2006. *Global and Planetary Change*, **60**, 487–504.

Zhang, Y., J. M. Wallace, and D. S. Battisti, 1997: ENSO-like interdecadal variability: 1900-93. *J. Climate*, **10**, 1004–1020.