Analysis of source regions and meteorological factors for the variability of spring PM$_{10}$ concentrations in Seoul, Korea

Jangho Lee, Kwang-Yul Kim[*]

School of Earth and Environmental Sciences, College of Natural Sciences, Seoul National University, Seoul 08826, Republic of Korea

ABSTRACT

CSEOF analysis is applied for the springtime (March, April, May) daily PM$_{10}$ concentrations measured at 23 Ministry of Environment stations in Seoul, Korea for the period of 2003–2012. Six meteorological variables at 12 pressure levels are also acquired from the ERA Interim reanalysis datasets. CSEOF analysis is conducted for each meteorological variable over East Asia. Regression analysis is conducted in CSEOF space between the PM$_{10}$ concentrations and individual meteorological variables to identify associated atmospheric conditions for each CSEOF mode. By adding the regressed loading vectors with the mean meteorological fields, the daily atmospheric conditions are obtained for the first five CSEOF modes. Then, HYSPLIT model is run with the atmospheric conditions for each CSEOF mode in order to back trace the air parcels and dust reaching Seoul. The K-means clustering algorithm is applied to identify major source regions for each CSEOF mode of the PM$_{10}$ concentrations in Seoul. Three main source regions identified based on the mean fields are: (1) northern Taklamakan Desert (NTD), (2) Gobi Desert and (GD), and (3) East China industrial area (ECI). The main source regions for the mean meteorological fields are consistent with those of previous study; 41% of the source locations are located in GD followed by ECI (37%) and NTD (21%). Back trajectory calculations based on CSEOF analysis of meteorological variables identify distinct source characteristics associated with each CSEOF mode and greatly facilitate the interpretation of the PM$_{10}$ variability in Seoul in terms of transportation route and meteorological conditions including the source area.

1. Introduction

Particulate matter under 10 μm, also known as PM$_{10}$ is well known for its negative effects on human health (Kampa and Castanas, 2008; Kassomenos et al., 2013). Issues of extreme PM$_{10}$ concentration events and its variability in Korea have been continuously raised for the last few years in connection with public health management (Oh et al., 2014; Kim et al., 2016). The PM$_{10}$ concentration in South Korea is largely affected by meteorological factors, long- and short-range transports from deserts and industrial areas in China, as well as local emission from South Korea itself (Kim et al., 2008; Lee et al., 2016).

Earlier research has identified major source regions and meteorological characteristics affecting the transport of PM$_{10}$ and its concentration over South Korea (Kim et al., 2008, 2010; Park et al., 2010). Well known source regions for PM$_{10}$ transported to South Korea are: the Gobi Desert, the Taklamakan Desert, the Loess Plateau, and the industrial areas in China. Numerous meteorological factors including wind velocity, sea level pressure, relative humidity, and temperature potentially affect the PM$_{10}$ concentration in Korea as well. Most recent

research dealing with the issues of PM$_{10}$ and yellow sand focused on specific extreme events (Lee et al., 2006; Kim et al., 2008).

The purpose of this study is to understand the variability of the springtime PM$_{10}$ concentrations in Seoul, Korea in the context of source regions and transport characteristics via atmospheric conditions. For this goal, variability of the springtime PM$_{10}$ concentrations in Seoul is decomposed into distinct modes together with their source regions and atmospheric conditions governing each mode based on detailed statistical and numerical analysis. Section 2 shows the data and method of analysis used in this study. The cyclostationary empirical orthogonal function (CSEOF) technique is employed to extract accurate evolutions of various meteorological variables in association with individual modes of PM$_{10}$ concentration in Seoul (Kim and Roh, 2010; Kim et al., 2012a, 2012b, 2013). Hybrid single particle Lagrangian trajectory (HYSPLIT) model is also used for back trajectory analysis (Draxler and Hess, 1997, 1998; Draxler et al., 1999; Stein et al., 2015). HYSPLIT model is capable of calculating back trajectory of air parcel and dust for given meteorological data at different pressure levels. By combining the HYSPLIT model and CSEOF analysis, meteorological patterns and main

source regions for individual modes of PM_{10} concentration will be probed. Section 3.1 presents the result of CSEOF analysis for the PM_{10} concentrations, while section 3.2 shows the results of regression analysis in CESOF space to identify major meteorological patterns responsible for each CSEOF mode of the PM_{10} concentration. Section 3.3 shows HYSPLIT back trajectory result, which in turn is used to identify the main source regions for each CSEOF mode. Statistical analysis including the K-means clustering is carried out to analyze the results.

2. Data and method of analysis

2.1. Data

Data used in this study are daily averaged PM_{10} concentration measured at 23 Korea Ministry of Environments (ME) stations during 2003–2012. The ME stations are spread fairly evenly across the city of Seoul in Korea. Meteorological variables are derived from the European Center for Medium-Range Weather Forecast (ECMWF) 6-hourly ERA-Interim product. Analysis is conducted over the East Asian region of 21°–60°N and 60°-150°E at the resolution of 1.5° during the spring season (March, April and May: hereafter, MAM).

2.2. CSEOF analysis

In the present study, cyclostationary empirical orthogonal function (CSEOF) analysis technique is used (Kim et al., 1996, 2015; Kim and North, 1997). CSEOF analysis writes data in space and time as

$$PM(r, t) = \sum_n LV_n(r, t)PC_n(t), \qquad t \in D, \tag{1}$$

where $LV_n(r, t)$ are called cyclostationary loading vectors (CSLVs), $T_n(t)$ are called principal component (PC) time series, and r, t, n, and D stand for space, time, mode number, and data interval, respectively. CSEOF analysis decomposes space-time data into orthogonal modes of evolution (CSLVs). The amplitudes of individual space-time evolutions are governed by corresponding PC time series that are mutually uncorrelated (and often nearly independent). Unlike conventional EOF analysis, of which the loading vectors are simply spatial patterns, CSEOF analysis renders each physical process as temporally evolving but periodically repeating spatial patterns:
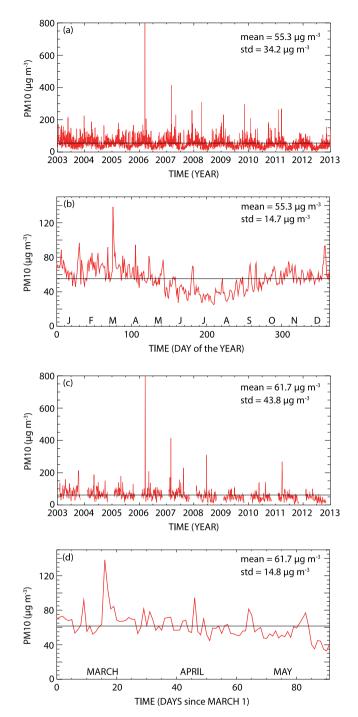
$$LV_n(r, t) = LV_n(r, t + d), \tag{2}$$

where d is called the nested period. In the present study, the nested period is set to 92 spring (MAM) days. Therefore, each CSLV consists of 92 spatial patterns.

2.3. Regression analysis in CSEOF space

Physical relationship between the PM_{10} concentrations in Seoul and the meteorological factors—air temperature, geopotential height, relative humidity, zonal wind, meridional wind, and vertical component of wind at 12 different pressure levels (1000hpa, 975hpa, 950hpa, 925hpa, 900hpa, 850hpa, 800hpa, 750hpa, 700hpa, 600hpa, 500hpa, 400hpa)—are investigated via regression analysis in CSEOF space. The PM_{10} concentration is selected as the target variable, while the 72 meteorological factors (6 variables × 12 pressure levels) are predictor variables. CSEOF analysis is conducted on each predictor variable. Then, the PC time series of each predictor variable are regressed onto the PC time series of the target variable as in (3):

$$PC_i(t) = \sum_n a_n^{(i)} PCP_n(t) + \varepsilon^{(i)}(t), \tag{3}$$

where $PC_i(t)$ denotes the target PC time series for mode i, and $PCP_n(t)$ denotes predictor PC time series for mode n, $\{a_n^{(i)}\}$ are regression coefficients, and $\varepsilon^{(i)}(t)$ is regression error time series, respectively. Regression patterns for each predictor variable are obtained as

Fig. 1. (a) The daily PM_{10} concentrations in 2003–2012 archived at the Ministry of Environment of Korea. (b) The annual cycle of daily PM_{10} concentrations. (c) The daily PM_{10} concentrations during spring (MAM). (d) The seasonal cycle of springtime PM_{10} concentrations. All the time series represent the average over the 23 stations in Seoul.

$LVPR_i(r, t)$ as in (4):

$$LVPR_i(r, t) = \sum_n a_n^{(i)} LVP_n(r, t), \tag{4}$$

where is CSLV for the predictor variable. Then, the target and predictor variables can be written as

$$\{T(r, t), P(r, t)\} = \sum_n \{LV_n(r, t), LVPR_n(r, t)\}PC_n(t), \tag{5}$$

namely, pairs of CSLVs are governed by identical PC time series. Then, the regressed patterns of the predictor variable, $LVPR_n(r, t)$, are
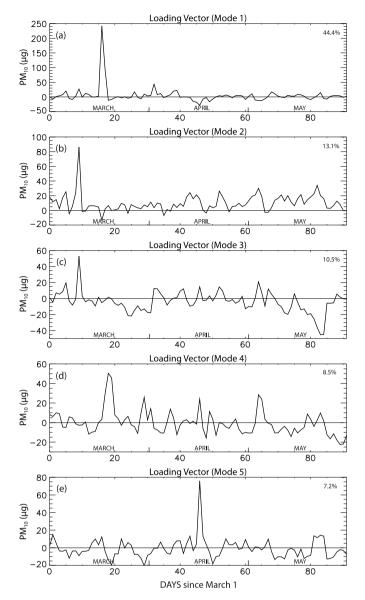
Fig. 2. The first five loading vectors of springtime daily PM_{10} concentrations averaged over the 23 ME stations: (a)–(e) CSEOF modes 1–5.

Fig. 3. The first five PC time series of springtime daily PM_{10} concentrations averaged over the 23 ME stations: (a)–(e) CSEOF modes 1–5.

regarded as physically consistent with the corresponding mode of the target variable, $LV_n(r, t)$. We can repeat this procedure for as many predictors as desired. As a result, we can make the loading vectors derived from different variables to be physically consistent with the loading vectors of the PM_{10} concentrations in Seoul (target variable).

2.4. HYSPLIT model

Back trajectories are computed using the Hybrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT, v4.7) model (Draxler and Hess, 1997, 1998; Draxler et al., 1999; Stein et al., 2015). HYSPLIT is a complete system for computing simple trajectories and complex dispersion and deposition simulations under a given meteorological setting. It allows us to specify geographic areas acting as the main source for PM_{10} and dust.

Model was run every 6 h during the spring period (MAM, 92 days); for each run, 3-day back trajectories are determined starting from Seoul (37.5°N, 127.0°N) at 500 m, 1 500 m, and 3 000 m above the ground level (AGL). These levels were selected as boundary layers (500–1 500 m: low levels, 3 000 m: mid levels), where the primary transportation of dust is observed in Seoul Korea (Choi et al., 2008).
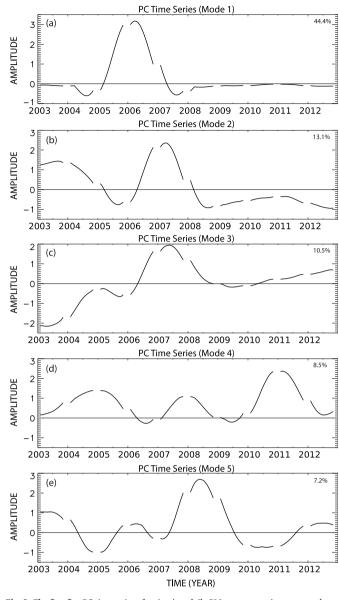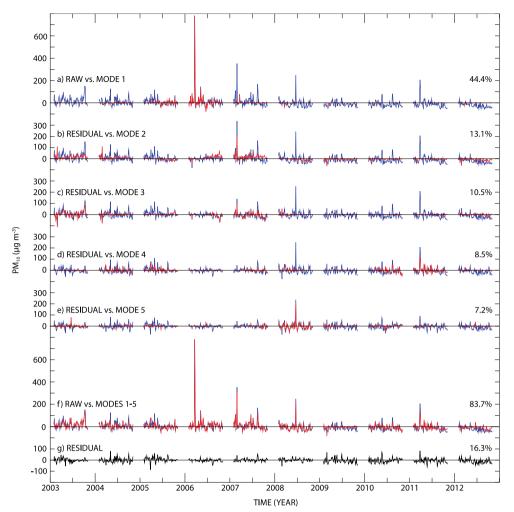
Since single back trajectory can be rather uncertain and insignificant, a large number of back trajectory episodes are generated for a more reliable identification of source areas.

2.5. K-means clustering

Cluster analysis or clustering is a statistical method of grouping a set of data in a way that data in same group is more similar than those of other groups. The K-means clustering is to divide n observations into k clusters, so that each of the n members belongs to a cluster that has the closest center from it. In other words, k clusters are determined such that the sum of distance of members of each cluster from its center is minimized. Euclidean distance is used based on the longitude and latitude of each trajectory. Mathematically, it can be written as

$$\arg\min_{S} \sum_{i=1}^{k} \sum_{x \in S_i} \left\| x - \mu_i \right\|^2,$$

(6)

where x is each member out of the n total observations, S_i is each set of clusters, and μ_i is the cluster center for S_i. Given a suitable number of k,

Fig. 4. Reconstruction of PM_{10} explained by the first five CSEOF modes. The first row is the raw time series versus reconstruction based on mode 1. The second row is the residual (raw – mode 1) versus mode 2, and so forth. The second to the last row is the raw time series versus reconstruction based on modes 1–5, and the last row is the residual after extracting the first five CSEOF modes. The numbers on the right are the percent variances explained. The ordinate scales are uniform for all the time series.

centers of k clusters are found iteratively based on the minimization problem (6). The parameter k is determined by using the elbow method. The elbow method calculates the RMSD (root mean square deviation) as a function of the number of clusters (k). When adding another cluster does not significantly improve the RMSD value of the population, the present k value is selected as the optimal number. More details can be found in Likas et al. (2003).

3. Results

3.1. PM_{10} concentration and CSEOF analysis

Fig. 1a shows the daily PM_{10} concentrations averaged over the 23 stations in Seoul for the data period (2003–2012). The average PM_{10} concentration in Seoul is 55.3 $\mu g\,m^{-3}$ during this period. Fig. 1b shows the composite (annual cycle) of daily PM_{10} concentrations averaged at the 23 stations in Seoul. As seen in the figure, the PM_{10} concentration in Seoul, on average, is highest in spring and is lowest in summer. Fig. 1c is the springtime (MAM) PM_{10} concentration. The corresponding seasonal (MAM) cycle exhibits that PM_{10} activity peaks in the middle of March and generally declines in April and May (Fig. 1d). Physical processes governing the variability of PM_{10} concentration may differ from one season to another. Thus, CSEOF analysis is carried out on the 92-day spring season (MAM), which exhibits the highest PM_{10} concentrations in Seoul. By using a 92-day nested period, physical evolutions during the spring season can be resolved.

Fig. 2 denotes the first five CSEOF loading vectors for the springtime PM_{10} concentrations in Seoul; the five modes together explain about

84% of the total variance. There is no significant variation among the stations. Therefore, station-averaged loading vectors are presented in Fig. 2. The first four modes show prominent peaks in March, whereas Mode 5 shows a prominent peak in April. Fig. 3 shows the PC time series for the five CSEOF modes, each of which describes the amplitude of corresponding loading vector. The first mode exhibits a prominent peak in 2006. Judging from the PC time series and the loading vector, Mode 1 seems to represent primarily a strong PM_{10} event in early 2006 (see also Fig. 4). Based on the PC time series in Fig. 3, PM_{10} concentration associated with Mode 2 had peaked in 2007 and decreased since then (see Fig. 4). Based on Mode 3, PM_{10} concentration in March had peaked in 2007 and is in an increasing trend now. On the other hand, PM_{10} concentration in May was significant in earlier period but has decreased significantly in 2007 and is in a decreasing mode now. In addition to the March peak, Mode 4 also exhibits sporadic peaks throughout the spring season, although the associated PM_{10} concentration generally decreases throughout the rest of the spring. According to the PC time series, PM_{10} events associated with Mode 4 were significant in spring of 2004–2005, 2007–2008, and 2010–2011 (see Fig. 4). Finally, the April peak associated with Mode 5 was strongest in 2008 (see Fig. 4). As shown in Fig. 4, the five CSEOF modes explain the majority of PM_{10} variability in Seoul. The remaining variability shows relatively weak variation of PM_{10} concentration during the record period (Fig. 4).
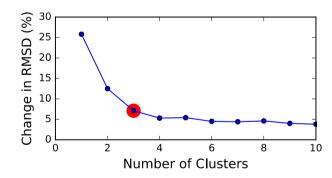
3.2. Regression analysis on CSEOF space

Regression analysis in CSEOF space has been conducted for various

Table 1
The R^2 values of regression and the maximum and minimum correlations between the averaged loading vector of PM_{10} concentrations in Seoul and the regressed loading vectors of meteorological variables.

Var	Level (hpa)	Corr Max	Corr Min	R^2	Var	Level (hpa)	Corr Max	Corr Min	R^2
Air Temp	400	0.274	−0.433	0.986	U Wind	400	0.493	−0.428	0.986
	500	0.260	−0.431	0.986		500	0.501	−0.430	0.986
	600	0.213	−0.420	0.986		600	0.507	−0.397	0.986
	700	0.263	−0.430	0.986		700	0.526	−0.387	0.986
	750	0.276	−0.461	0.986		750	0.426	−0.383	0.986
	800	0.307	−0.431	0.986		800	0.472	−0.398	0.986
	850	0.312	−0.451	0.985		850	0.396	−0.384	0.986
	900	0.256	−0.426	0.985		900	0.384	−0.412	0.986
	925	0.293	−0.439	0.985		925	0.387	−0.429	0.985
	950	0.348	−0.437	0.985		950	0.382	−0.430	0.985
	975	0.288	−0.438	0.985		975	0.387	−0.429	0.985
	1 000	0.254	−0.436	0.986		1 000	0.416	−0.432	0.985
Geo Potential Height	400	0.318	−0.465	0.986	V Wind	400	0.440	−0.461	0.985
	500	0.359	−0.475	0.986		500	0.442	−0.407	0.985
	600	0.385	−0.496	0.986		600	0.448	−0.378	0.985
	700	0.398	−0.510	0.986		700	0.444	−0.384	0.985
	750	0.472	−0.430	0.986		750	0.502	−0.344	0.985
	800	0.377	−0.418	0.986		800	0.400	−0.330	0.985
	850	0.452	−0.411	0.985		850	0.411	−0.335	0.985
	900	0.365	−0.390	0.985		900	0.412	−0.460	0.985
	925	0.415	−0.368	0.985		925	0.369	−0.369	0.985
	950	0.406	−0.347	0.985		950	0.383	−0.348	0.985
	975	0.397	−0.325	0.985		975	0.402	−0.332	0.985
	1 000	0.359	−0.376	0.985		1 000	0.394	−0.420	0.985
Relative Humidity	400	0.412	−0.366	0.986	Vertical Wind	400	0.394	−0.396	0.985
	500	0.412	−0.369	0.985		500	0.420	−0.418	0.985
	600	0.376	−0.404	0.985		600	0.441	−0.452	0.985
	700	0.338	−0.427	0.985		700	0.444	−0.459	0.985
	750	0.414	−0.417	0.985		750	0.426	−0.411	0.985
	800	0.407	−0.427	0.985		800	0.408	−0.478	0.985
	850	0.462	−0.413	0.985		850	0.435	−0.420	0.985
	900	0.478	−0.429	0.983		900	0.518	−0.415	0.985
	925	0.436	−0.396	0.985		925	0.463	−0.435	0.985
	950	0.438	−0.396	0.985		950	0.472	−0.435	0.985
	975	0.427	−0.391	0.985		975	0.508	−0.449	0.986
	1 000	0.420	−0.380	0.985		1 000	0.511	−0.436	0.986

Fig. 5. Percent change in RMSD with respect to the number of clusters. Red circle denotes the selected number of clusters ($k = 3$). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

meteorological variables with the PM_{10} concentration in Seoul as the target variable. Then, each loading vector of the PM_{10} concentration averaged over 23 stations in Fig. 2 is correlated with the matching regressed loading vectors of meteorological variables at each grid point. This results in a correlation coefficient at each grid point for each variable. The correlation map, in a sense, shows the geography of sensitivity to each meteorological variable for each CSEOF mode of PM_{10} concentrations. The R^2 value of regression for each variable is shown in Table 1 together with the maximum and minimum correlations. Average maximum correlation for all variables is 0.400 while average minimum correlation is −0.413, and average R^2 value is 0.985.

3.3. Cluster analysis on HYSPLIT model simulations

For each of the 72 variables (6 meteorological variables × 12 pressure levels), its 92-day mean spring patterns have been calculated by averaging the 10 years (2003–2012) of meteorological data. In the first step, 72-hr HYSPLIT back trajectory model is run with the mean field at every 6 h to generate 1 104 trajectories (6 hourly for 92 days on 3 arriving levels). This results in a dataset with a large number of source regions. In the second step, K-means cluster analysis is conducted with the source regions identified in the first step. To choose the best number of clusters (k) for grouping source regions, the elbow method based on RMSD percentage change has been utilized. For example, RMSD percentage change is very sluggish after $k = 3$ in Fig. 5; thus, $k = 3$ is chosen as an appropriate number of clusters to group source regions in this particular simulation. It has been noted that clustering of back trajectory locations could demonstrate different types of meteorological patterns (Dorling et al., 1992).

HYSPLIT back trajectory model is run with 6 different sets of meteorological fields to address the mean source regions, and the source regions for the first five CSEOF modes (see Table 2). Fig. 6 shows K-means cluster analysis result for the mean field, which is obtained by averaging the 10 years of reanalysis data on each of the 92 spring days (MAM). As seen in Fig. 6, three main sources governing the variability of PM_{10} concentrations in Seoul are: (1) northern Taklamakan Desert (NTD) (2) Gobi Desert (GD), and (3) East China industrial area (ECI). This result is consistent with previous research (Kim, 2008). Source regions for extreme ($\mu + \sigma$: exceeding one standard deviation above the

Table 2
Center location and percentage of members for three clusters under six different settings of meteorological fields (mean, mode 1, mode 2, mode 3, mode 4, and mode5). NTD, GD, ECI denotes the percentage of members in the source regions in the north Taklamakan Desert, the Gobi Desert, and the East China industrial region, respectively.
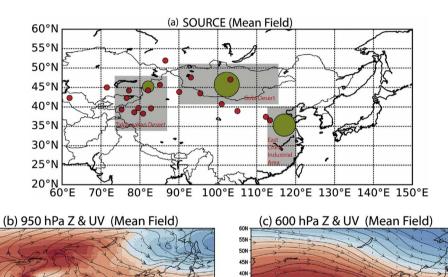
	Cluster 1 Center Location (lon, lat)		Cluster 2 Center Location (lon, lat)		Cluster 3 Center Location (lon, lat)		NTD	GD	ECI
Mean	82.1, 45.2	21%	117.1, 35.2	37%	102.6, 45.5	41%	21%	41%	37%
Mode 1	85.9, 45.9	27%	119.6, 33.7	40%	104.1, 45.6	32%	27%	32%	40%
Mode 2	94.3, 44.9	9%	110.7, 39.0	40%	116.9, 31.9	49%	0	9%	89%
Mode 3	85.0, 46.9	19%	117.2, 34.3	38%	102.4, 45.0	42%	19%	42%	38%
Mode 4	80.1, 42.2	37%	113.0, 35.2	24%	94.2, 47.8	38%	37%	38%	24%
Mode 5	84.5, 42.6	18%	113.5, 32.6	27%	100.3, 42.4	54%	19%	54%	27%

Fig. 6. (a) HYPLIT back trajectory result for the 10-year mean meteorological fields. The major source regions (green circled regions) are identified by K-means cluster analysis ($k = 3$) of back trajectories. Radius of each circle is proportional to the number of members in the cluster. Red dots denote source regions for extreme ($\mu + \sigma$) events of PM_{10} concentrations. The March-April-May climatological mean patterns of (b) 950-hPa geopotential height (shading) and wind (streamline), (c) 700-hPa geopotential height (shading), and wind (streamline), (d) air temperature, and (e) relative humidity (shading) and vertical velocity (contour: 0.02 interval from $\pm 0.02 \, Pa \, s^{-1}$). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

mean) PM_{10} concentration events are also identified as red dots. In the mean sense, NTD and GD are more important sources of extreme PM_{10} events in Seoul than ECI.
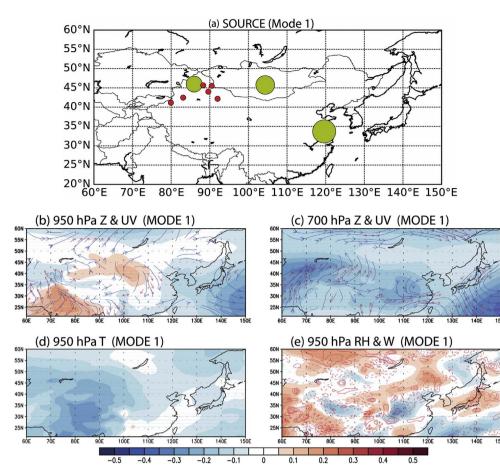
The mean flow patterns show that these source regions are on the path of the low- and mid-tropospheric westerly transport to Korea (Fig. 6b and c). Together with the strong temperature gradient (Fig. 6d) and relatively dry condition (Fig. 6e), these regions serve as an important source of PM_{10} observed in Seoul, Korea. It appears that the mean atmospheric patterns emphasize more of the source conditions and wind transport rather than the atmospheric conditions at the observing location (Seoul).

Figs. 7–11 show the correlation maps of the regressed loading vectors and the corresponding HYSPLIT back trajectory analysis results for the first five CSEOF modes. In order to generate the meteorological fields for the HYSPLIT model, the regressed loading vectors of meteorological variables for each CSEOF mode are added to the mean

fields on each day of the MAM period; this is equivalent to adding the anomalous fields to the mean fields for each CSEOF mode. Thus, panels (b)-(d) of Figs. 7–11 should be interpreted as the anomalous atmospheric conditions associated with the PM_{10} events for each CSEOF mode. These anomalous atmospheric conditions, and henceforth the actual source strength depicted in panel (a) of Figs. 7–11, fluctuate yearly according to Fig. 3.

As shown in Fig. 7a, HYSPLIT back trajectory model identifies source regions similar to that of the mean field in Fig. 6. A majority of PM_{10} events come from ECI, although extreme events, a few in number, originate from NTD. The low-tropospheric atmospheric conditions tend to show weak correlations (Fig. 7b, d, and 7e) with the loading vector of PM_{10} (Fig. 2a). The mid-tropospheric condition, on the other hand, shows a relatively significant correlation over the NTD region (Fig. 7c). The analysis result indicates that an increase in PM_{10} concentration in Seoul (as reflected in Fig. 2a) is associated with the decreased

Fig. 7. (a) HYSPLIT back trajectory result for the first CSEOF modes. Correlation map of (b) 950-hPa geopotential height (shading), and zonal and meridional wind (vector and streamline), (c) 700-hPa geopotential height (shading), and zonal and meridional wind (vector and streamline), (d) air temperature, and (e) relative humidity (shading) and vertical velocity (contour: 0.1 interval from $\pm$ 0.2) in the spring (MAM) season. Wind vectors with absolute correlations below 0.2 are not shown. The color bar at the bottom corresponds to panels (b)–(e). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

geopotential height and the anomalous southwesterly in the middle troposphere. This anomalous wind transports the PM_{10} produced over NTD northward, thereby increasing the transport of PM_{10} carried by the mean northwesterly (Fig. 6c). It should be noted that wind speed is increased over NTD, which might have contributed to an increased production of PM_{10}. Indeed, Figs. 3 and 4 indicate a strong PM_{10} event in 2006; this peak represents stronger mid-tropospheric wind and the severe dust storm observed in the Taklamakan Desert in March 2006 (Lee et al., 2010).

Fig. 8 shows the back trajectory results and correlation maps for the second CSEOF mode. The second CSEOF mode is related to short-range transports from the ECI region. Strong negative geopotential height anomaly to the northwest and strong positive geopotential height anomaly to the south of the Korean peninsula results in increased transport from ECI to Seoul (Fig. 8b and c). Because of the decreased relative humidity removal of dust from the atmospheric column becomes less effective (Fig. 8e), which provides a favorable condition for increased PM_{10} transport, ultimately increasing the PM_{10} concentration in Seoul. This is also reflected in the generally decreased vertical velocity over the ECI region. Temperature may exert little direct impact on the transportation and removal of dust, but it acts as an important factor for many physical processes (Yang, 2002). It seems that the decreased relative humidity is due to the increased air temperature (Fig. 8d).
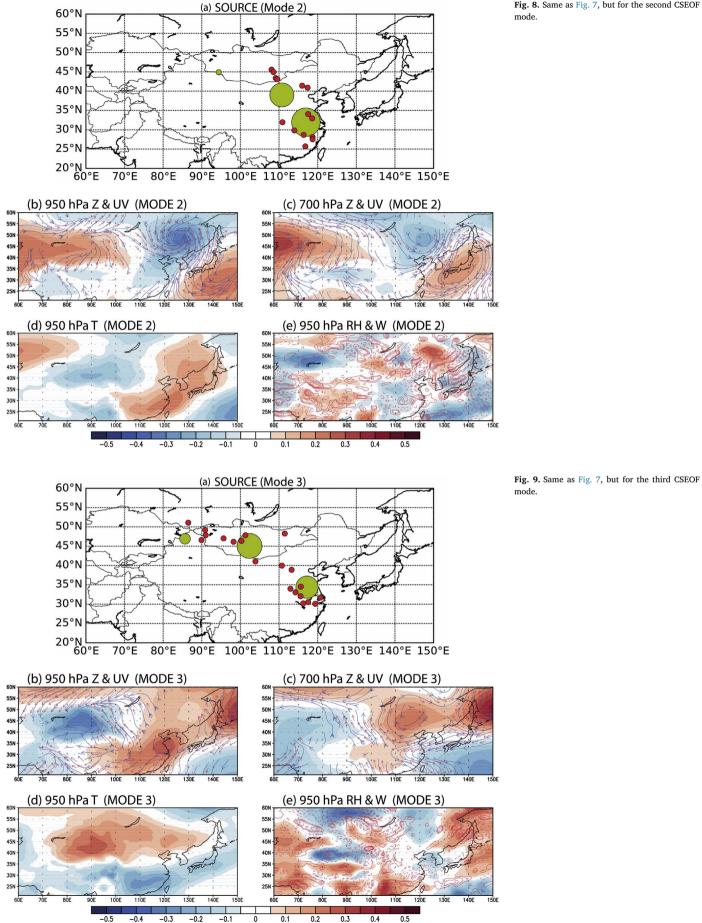
Fig. 9 shows the back trajectory result and correlation maps for the third CSEOF mode. As can be observed in Fig. 9a, the third CSEOF mode is related to increased extreme events from the ECI region. Significant positive correlation with the geopotential height field is shown in the ECI region (Fig. 9b). The positive geopotential height anomaly creates anomalous northeasterly from South Korea to ECI. This lower-tropospheric anomalous flow slows down the mean flow from ECI to Seoul (see Fig. 6b), which tends to increase the residence time of PM_{10}. The

source locations of extreme PM_{10} events near ECI also support this explanation; PM_{10} particles travelled only a short distance from ECI to Seoul for a three-day period. As the residence time increases, observed concentration of PM_{10} also increases, which leads to extreme PM_{10} events in Seoul.

The third CSEOF mode is also related to the increased transport from the GD region (compare Figs. 8a and 9a), which clearly distinguishes Mode 3 from Mode 2. In addition, only 4 of the 30 extreme events in Fig. 9a are on the same dates. It seems that the slower lower-tropospheric wind over GD increases the concentration of PM_{10} over the source region, which ultimately is transported to South Korea. Like the second mode, the pattern of air temperature in the third CSEOF mode shows consistency with that of relative humidity. Neither air temperature nor relative humidity exhibits any significant covariability with the third loading vector of the PM_{10} concentration in Seoul.

The fourth CSEOF mode is related with increased extreme events form NTD (Fig. 10a), and is significantly correlated with all the atmospheric variables (Fig. 10b–e). The anomalous high in the lower troposphere (Fig. 10b) and anomalous low in the middle troposphere (Fig. 10c) may represent a fairly cold atmosphere (Fig. 10d). Such a condition is often associated with the propagation of synoptic-scale cyclonic disturbances (see Fig. 12 in Kim et al., 2010). A comparison between Fig. 10b and c shows the characteristic baroclinic structure in the anomalous geopotential height field associated with a synoptic scale disturbance; the vortex center at the northeastern corner of the domain is further to the east at the lower tropospheric level. Specific humidity decreases to the west of the cyclonic vortex over East China and Korea (Fig. 10e). In other areas, relative humidity increases because the saturation specific humidity decreases owing to colder atmospheric condition. The tropospheric wind condition seems to establish a favorable condition for transporting PM_{10} from NTD to Seoul. As shown in Fig. 10c, middle tropospheric wind shows a dominant flow from NTD to
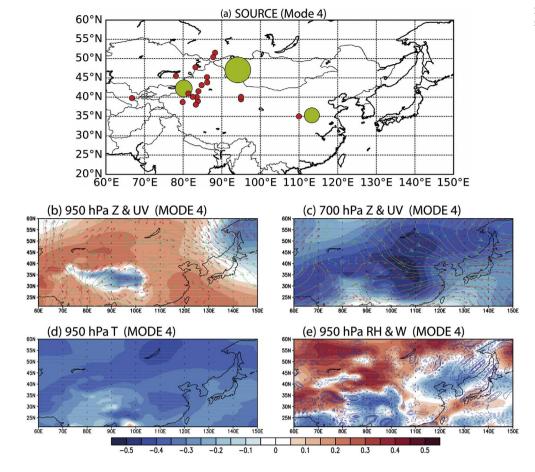
Fig. 8. Same as Fig. 7, but for the second CSEOF mode.

Fig. 9. Same as Fig. 7, but for the third CSEOF mode.

South Korea. In particular, downward motion (Fig. 10e) associated with the mid-tropospheric confluence (Fig. 10c) and the lower-tropospheric divergence (Fig. 10b) over the Korean peninsula may help increase the concentration of PM_{10} in Seoul.

A major source region for the fifth CSEOF mode is GD (Fig. 11a). Extreme PM_{10} events in Seoul associated with this mode also derive from GD. It appears that the relatively warm and dry condition over the source region (GD) is favorable for the production of PM_{10}, but correlation is fairly low. Furthermore the anomalous mid-tropospheric wind slows down the northwesterly mean wind, causing a slower transport of PM_{10} to Seoul. This slower transport with a higher production of PM_{10} seems to be responsible for a higher concentration of PM_{10} in Seoul. At the same time, transport of PM_{10} from NTD takes more time to reach Seoul and the chance of removal from the atmospheric column is increased. The lower tropospheric wind, on the other hand, causes a faster transit of PM_{10} produced over the ECI region. As a result, no extreme events from ECI are observed.

4. Summary and conclusions

CSEOF analysis is applied to the 10-year daily PM_{10} concentrations in Seoul, Korea to obtain uncorrelated evolutions of the PM_{10} concentrations and associated meteorological variables. The first five CSEOF modes explain ∼84% of the total variability. CSEOF analysis is conducted on six meteorological variables (geopotential, relative humidity, temperature, zonal wind, meridional wind, vertical wind) at 12 different pressure levels over East Asia. With the PM_{10} concentration as the target variable, regression analysis is conducted in CSEOF space with each meteorological variable as a predictor variable in order to understand the atmospheric conditions associated with each CSEOF mode of the PM_{10} concentrations in Seoul.
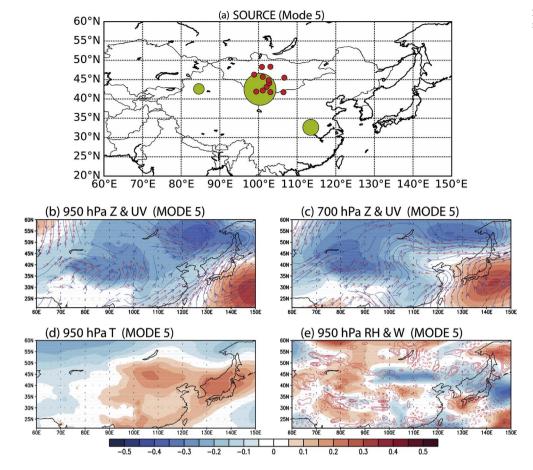
Correlation maps between the loading vector of the PM_{10} concentration and the regressed loading vectors of the meteorological variables show the general meteorological conditions and how strongly they are correlated with the variation of PM_{10} in Seoul for each CSEOF mode. These maps offer valuable insight into the meteorological conditions that are associated with the major modes of variability of the PM_{10} concentrations in Seoul. These maps also help understand prevailing source conditions and atmospheric conditions that lead to extreme PM_{10} events in Seoul, Korea.

HYSPLIT back trajectory model is used with the mean meteorological fields and those of each CSEOF mode to determine the source regions of the PM_{10} variability for each CSEOF mode. Starting from 3 different height levels (500, 1500, 3 000 m) in Seoul, 3-day back trajectories are computed every 6 h during the 92 days of winter for each CSEOF mode. For the endpoints (source locations) of the back trajectory model, K-means clustering algorithm is applied to identify major source regions.

As for the mean meteorological fields, source regions are consistent with those of previous research (Kim, 2008). Using the 10-year (2003–2012) averaged meteorological conditions, main source regions for the variability of the PM_{10} concentrations in Seoul are identified as: (1) northern Taklamakan Desert (NTD), (2) Gobi Desert (GD), and (3) East China industrial area (ECI). GD is the most frequent (41%) source region, whereas NTD is the least frequent (21%) one. On the other hand, most extreme events of PM_{10} in Seoul, in the climatological mean sense, originate from NTD, while ECI does not contribute significantly to the extreme PM_{10} events in Seoul. This difference between the frequencies of PM_{10} events and those of extreme events stem from source strengths and/or atmospheric conditions over the source region as well as the arrival location, which could not be elucidated in the present study. Under specific anomalous atmospheric conditions from the mean field, relative significance of PM_{10} contribution from each source region varies as described below.

Fig. 11. Same as Fig. 7, but for the fifth CSEOF mode.

The HYSPLIT back trajectory results for the first five CSEOF modes show different source/transport characteristics. Each CSEOF mode exhibits distinct atmospheric conditions, and henceforth, distinct transport characteristics. Due to the distinct atmospheric conditions, the source characteristics, including the source regions and frequencies for PM_{10} and extreme PM_{10} events, differ significantly from one mode to another. In a sense, specific atmospheric conditions are tied with specific source conditions from a climatological point of view.

As exemplified in the present study, source and transport characteristics of the PM_{10} variability in Seoul can be better understood in terms of CSEOF analysis followed by back trajectory analysis. Based on local measurements of PM_{10} concentrations, such as the PM_{10} concentration in Seoul, it is difficult to understand the state of the source region, atmospheric condition, or the characteristics of the transporting wind. Combining the CSEOF technique and the HYSPLIT back trajectory model greatly facilitates the analysis of local PM_{10} concentrations and their variability with respect to the source regions and the meteorological conditions affecting the transport of PM_{10}. It is also possible to delineate how each source region has contributed to the variation of the PM_{10} concentrations in Seoul including the extreme PM_{10} events. Uncertainty is inherent in the present analysis, since only the general and mean atmospheric conditions are investigated in interpreting the source and transport characteristics for each CSEOF mode. Inspection of the back trajectories of individual PM_{10} events with daily meteorological fields is needed to understand more specific reasons for the PM_{10} variability in Seoul. Nonetheless, this study provides a useful framework of analyzing variability of local PM_{10} concentrations.

Acknowledgments

References

Choi, H., Zhang, Y.H., Kim, K.H., 2008. Sudden high concentration of TSP affected by atmospheric boundary layer in Seoul metropolitan area during dust storm period. Environ. Int. 34 (5), 635–647.

Dorling, S.R., Davies, T.D., Pierce, C.E., 1992. Cluster analysis 1992: a technique for estimating the synoptic meteorological controls on air and precipitation chemistry—method and applications. Atmos. Environ. 26 (14), 2575–2581.

Draxler, R., Hess, G.D., 1997. Description of the HYSPLIT4 Modeling System.

Draxler, R., Hess, G.D., 1998. An overview of the HYSPLIT_4 modelling system for trajectories. Aust. Meteorol. Mag. 47 (4), 295–308.

Draxler, R., Stunder, B., Rolph, G., Stein, A., Taylor, A., 1999. HYSPLIT4 user's guide. NOAA Tech. Memo. ERL ARL 230, 35.

Kampa, M., Castanas, E., 2008. Human health effects of air pollution. Environ. Pollut. 151 (2), 362–367.

Kassomenos, P.A., Dimitriou, K., Paschalidou, A.K., 2013. Human health damage caused by particulate matter PM10 and ozone in urban environments: the case of Athens, Greece. Environ. Monit. Assess. 185 (8), 6933–6942.

Kim, J., 2008. Transport routes and source regions of Asian dust observed in Korea during the past 40 years (1965–2004). Atmos. Environ. 42 (19), 4778–4789.

Kim, K.-Y., North, G.R., 1997. EOFs of harmonizable cyclostationary processes. J. Atmos. Sci. 54, 2416–2427.

Kim, K.-Y., Roh, J.W., 2010. Physical mechanisms of the wintertime surface air temperature variability in South Korea and the near-7-Day oscillations. J. Clim. 23, 2197–2212.

Kim, K.-Y., North, G.R., Huang, J., 1996. EOFs of one-dimensional cyclostationary time series: computations, examples and stochastic modeling. J. Atmos. Sci. 53, 1007–1017.

Kim, H.-S., Chung, Y.-S., Lee, S.-G., 2008. Characteristics of aerosol types during large-scale transport of air pollution over the Yellow Sea region and at Cheongwon, Korea, in 2008. Environ. Monit. Assess. 184 (4), 1973–1984.

Kim, S.-W., Yoon, S.-C., Kim, J., Kang, J.-Y., 2010. Asian dust event observed in Seoul, Korea, during 29–31 May 2008: analysis of transport and vertical distribution of dust particles from lidar and surface measurements. Sci. Total Environ. 408 (7), 1707–1718.

Kim, Y., Kim, K.-Y., Kim, B.-M., 2012a. Physical mechanisms of European winter snow cover variability and its relationship to the NAO. Clim. Dynam. 40, 1657–1669.

Kim, Y., Kim, K.-Y., Jhun, J.-G., 2012b. Seasonal evolution mechanism of the East Asian winter monsoon and its interannual variability. Clim. Dynam. 41, 1213–1228.

Kim, Y., Kim, K.-Y., Park, S., 2013. Seasonal scale variability of the East Asian winter monsoon and the development of a two-dimensional monsoon index. Clim. Dynam. 42, 2159–2172.

Kim, K.-Y., Hamlington, B.D., Na, H., 2015. Theoretical foundation of cyclostationary EOF analysis for geophysical and climatic variables: concepts and examples. Earth Sci. Rev. 150, 201–218.

Kim, H.-S., Chung, Y.-S., Yoon, M.-B., 2016. An analysis on the impact of large-scale transports of dust pollution on air quality in East Asia as observed in central Korea in 2014. Air Qual. Atmos. Health 9 (1), 83–93.

Lee, B.-K., Lee, H.-K., Jun, N.-Y., 2006. Analysis of regional and temporal characteristics of PM 10 during an Asian dust episode in Korea. Chemosphere 63 (7), 1106–1115.

Lee, Y.-C., Yang, X., Wenig, M., 2010. Transport of dusts from East Asian and non-East Asian sources to Hong Kong during dust storm related events 1996–2007. Atmos. Environ. 44 (30), 3728–3738.

Lee, G.-S., Kim, P.-R., Han, Y.-J., Holsen, T.M., 2016. Atmospheric speciated mercury concentrations on an island between China and Korea: sources and transport pathways. Atmos. Chem. Phys. 16 (6), 4119–4133.

Likas, A., Vlassis, N., Verbeek, J.J., 2003. The global k-means clustering algorithm. Pattern Recogn. 36 (2), 451–461.

Oh, H.-R., Ho, C.-H., Kim, J., Chen, D., Lee, S., 2014. Influence of trans-boundary air pollution from China on multi-day high PM10 episodes in Seoul, Korea. AGU Fall Meet. Abstr. 1.

Park, S.-U., Choe, A., Park, M.-S., 2010. Estimates of Asian dust deposition over the Asian region by using ADAM2 in 2007. Sci. Total Environ. 408 (11), 2347–2356.

Stein, A.F., Draxler, R.R., Rolph, G.D., 2015. NOAA's HYSPLIT atmospheric transport and dispersion modeling system. Bull. Am. Meteorol. Soc. 96 (12), 2059–2077.

Yang, K.-L., 2002. Spatial and seasonal variation of PM10 mass concentrations in Taiwan. Atmos. Environ. 36, 3403–3411.