ENVIRONMENTAL RESEARCH LETTERS

LETTER • OPEN ACCESS

Dynamical-statistical method for seasonal forecasting of wintertime PM10 concentration in South Korea using multi-model ensemble climate forecasts

To cite this article: Jahyun Choi *et al* 2024 *Environ. Res. Lett.* **19** 064073

ENVIRONMENTAL RESEARCH
LETTERS

CrossMark

OPEN ACCESS

RECEIVED
19 October 2023

REVISED
2 May 2024

ACCEPTED FOR PUBLICATION
24 May 2024

PUBLISHED
7 June 2024

LETTER

Dynamical-statistical method for seasonal forecasting of wintertime PM10 concentration in South Korea using multi-model ensemble climate forecasts

Jahyun Choi[1], Sung-Ho Woo[1], Jin-Ho Yoon[2], Jin-Young Choi[3], Daegyun Lee[3] and Jee-Hoon Jeong[1,*]

[1] Faculty of Earth and Environmental Sciences, Chonnam National University, Gwangju, Republic of Korea
[2] School of Earth Sciences and Environmental Engineering, Gwangju Institute of Science and Technology, Gwangju, Republic of Korea
[3] Air Quality Forecasting Center, National Institute of Environmental Research, Incheon, Republic of Korea
[*] Author to whom any correspondence should be addressed.

E-mail: jjeehoon@jnu.ac.kr

Keywords: air quality, particulate matter, seasonal forecasting, multi-model ensemble, dynamical-statistical model

Supplementary material for this article is available online

Abstract

Climate conditions and emissions are among the primary influences on seasonal variations in air quality. Consequently, skillful climate forecasts can greatly enhance the predictability of air quality seasonal forecasts. In this study, we propose a dynamical-statistical method for seasonal forecasting of particulate matter (PM_{10}) concentrations in South Korea in winter using climate forecasts from the Asian Pacific Climate Center (APCC) multi-model ensemble (MME). We identified potential climate predictors that potentially affect the wintertime air quality variability in South Korea in the global domain. From these potential climate predictors, those that can be forecasted skillfully by APCC MME were utilized to establish a multiple-linear regression model to predict the winter PM_{10} concentration in South Korea. As a result of evaluating the forecast skill through retrospective forecasts for the past 25 winters (1995/96-2019/20), this model showed statistically significant forecast skill at a lead time of a month to a season. The skill of PM_{10} forecast from the MME was overall better than that from a single model. We also found that it is possible to improve forecast skills through optimal MME combinations.

1. Introduction

Air pollution caused by rapid industrialization, increased fossil fuel use, and pollutant transport is one of the most serious social problems in East Asian countries. Since fine particles, in particular, have a direct impact on human health, social costs for reducing them continue to increase. Since 2019, South Korea has been implementing a particulate matter (PM_{10}) seasonal management policy, which enforces emission reduction in industry, power generation, transportation, and daily life during periods of expected or occurring high-concentration events (Ministry of Environment 2020). Therefore, a seasonal forecast of PM_{10} several weeks to a month or more in advance is required to ensure sufficient time for early response to high-concentration events.

Air quality is affected by pollutant emissions and their interaction with weather and climate conditions. Predicting these emissions is complex, but numerical models can be used to predict how atmospheric pollutants will stagnate, deposit, diffuse, and disperse due to weather conditions. Chemistry transport models (CTMs), a numerical modeling technique, have been widely used for short-term air quality forecasts ranging from a few days to a week (Kukkonen *et al* 2012, Baklanov *et al* 2014, Sokhi *et al* 2022). For example, the Goddard Earth observing system composition forecast (GEOS-CF) based on the GEOS-Chem chemical model provides global 5 day forecasts of air quality (Keller *et al* 2021). Similarly, the Copernicus Atmospheric Monitoring Service (CAM; Peuch *et al* (2022) provides a 4 day forecast for PM_{10} for Europe.

Long-term forecasts such as several weeks or a month in advance are approached mostly through statistical methods based on statistical relationships between weather/climate conditions and air quality. Previous studies suggested that seasonal PM behavior in northeast Asia is sensitive to the air quality-climate relationship. The East Asian winter monsoon (Jeong and Park 2017), blocking and synoptic weather patterns (Lee *et al* 2020, Ku *et al* 2021), large-scale atmospheric circulation associated with El Niño (Jeong *et al* 2021), aerosol transport from the Tibetan Plateau (Li *et al* 2020a), and the remote influence of the Madden-Julian Oscillation (Jung *et al* 2022), are known to affect winter PM_{10} variability in East Asia and the Korean Peninsula. Inspired by these studies, Jeong *et al* (2021) developed a linear regression model to forecast the winter PM in South Korea based on the statistical relationship between East Asian PM and climate factors. Forecast skill showed a correlation of 0.8 with the target PM concentration. Jeong *et al* (2022) recently developed a dynamical-statistical model linking PM_{10} concentration and climate variables from a climate forecast model, showing a correlation of 0.4 with observed PM_{10}. A statistical model for seasonal forecasting of U.S. summer ozone concentrations utilized spring climate patterns (Shen and Mickley 2017) achieved a correlation of 0.67.

In recent decades, skillful climate forecasts on a monthly or seasonal time scale have become possible to some extent through advances in climate modeling techniques, increased observational data including satellites, and data assimilation technologies (Mariotti *et al* 2018, Smith *et al* 2019). In particular, the Asia-Pacific Economic Cooperation Climate Center (APCC) multi-model ensemble (MME) (Min *et al* 2017), the observing system research and predictability experiment (THORPEX) Interactive Grand Global Ensemble (Swinbank *et al* 2016), and the North American MME (Kirtman *et al* 2014) have recently improved climate forecast skill through the MME technique. Based on this notable improvement in seasonal climate forecast, this study attempts to apply the MME climate forecast to forecasting the winter PM_{10} in South Korea. We developed a multiple linear regression (MLR) model that uses climate variables from the APCC MME climate forecast as climate predictors. Section 2 describes the structure of the dynamical-statistical method. Section 3 presents the empirical relationship between observed winter PM_{10} concentration and climate variables, the forecast skill of the model, and the optimization of the model. Section 4 summarizes the main results and discusses the potential utilization of the model and further studies.

2. Methods

2.1. Dynamical-statistical model

In this study, we developed a dynamical-statistical hybrid model (hybrid model hereafter) to forecast the winter (December to February) PM_{10} concentration in South Korea. The winter average PM_{10} is predicted with data produced in October, which is a 2 month lead-time prediction from a climate forecasting perspective. The hybrid model is based basically on an MLR model between PM_{10} in South Korea and climate variables. More specifically, among the climate variables, only variables that can be predicted in APCC MME's seasonal climate forecast and are physically influential on PM_{10} are used as climate predictors (i.e. independent variables) of the MLR. Unlike conventional statistical models based on a lagged relationship between predictor and predictand, the hybrid model is constructed based on their simultaneous relationship.

There are three main steps to building this hybrid model. We first (1) examined the simultaneous correlation between observed climate variables and PM_{10} concentration over South Korea to identify potential climate predictors with statistically significant relevance. Then, (2) among the potential climate predictors, we extracted climate predictors for which the climate forecast (APCC MME) has sufficient skill. The skill matrix of the climate forecast is based on the correlation coefficient between the historical hindcast (retrospective forecasts) and observed climate variables. Applying these two criteria, the climate predictors that are 'predictable and related to', presumably influencing, South Korea's PM_{10} were selected. Finally, (3) a hybrid model via the MLR between these independent climate predictors and PM_{10}, can be established. Inputting the climate predictors forecasted from APCC MME into this hybrid model produces PM_{10} seasonal forecasts. Figure 1 shows a schematic diagram of the model and forecast method.

As a result of the first two (1)–(2) processes, we identified potential climate variables from various regions in the Northern Hemisphere. Out of these, six variables were chosen based on their presence within a minimum range of 4 degrees longitude and 2 degrees latitude. In the final step (3), these variables were further narrowed down to two for independent variables of the MLR models considering their availability over 25 year period (1995/1996–2019/20). This selection followed the rule of thumb (Peduzzi *et al* 1996) requiring at least 10 data samples per independent variable for regression models. In addition, we checked the cross-correlation of the two variables to select variables without multicollinearity problems.

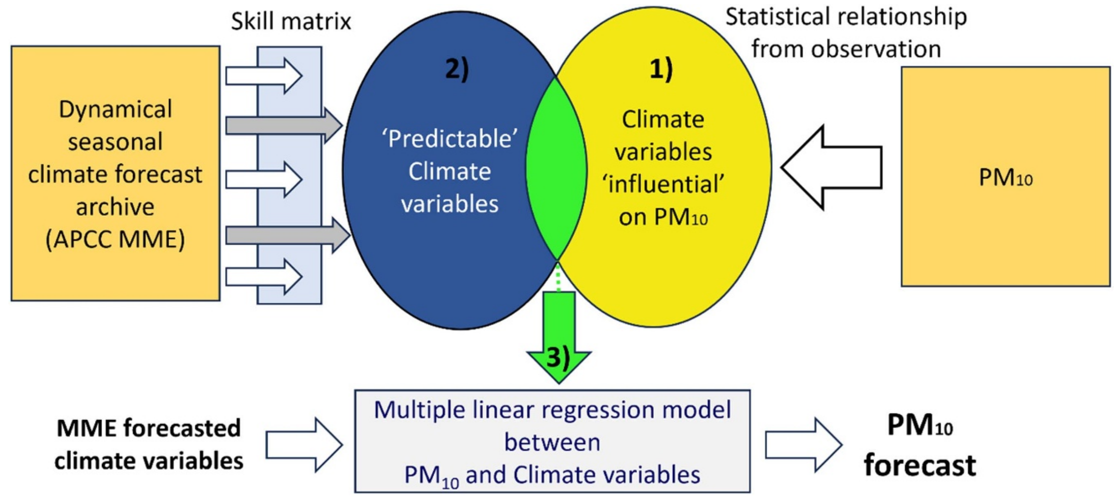

Figure 1. Schematic diagram of the dynamical-statistical model for PM_{10} forecast.

The chosen variables' spatial patterns and MLR model details are further described in sections 3.1 and 3.2.

For forecasting winter PM_{10}, the MLR uses the MME climate forecasts initialized in October. We employed the leave-one-out cross-validation technique over the 25 year study period to ensure robust model training and validation. For a given forecast year, the data of the remaining 24 years were used to train an MRL model, and the forecast was performed with the corresponding model. For convenience, the correlation between PM_{10} and climate variables and the skill matrix shown (figure 3) was calculated with data for the entire 25 years.

2.2. Observational data

The Korean Ministry of Environment has distributed PM_{10} concentration measured at 6-hour intervals since 2001. In this study, we used seasonal and monthly mean of PM_{10} from 153 stations for the period 2001/02–2019/20. In addition, PM_{10} concentrations at 12 stations in Seoul provided by the Seoul Metropolitan Government, are utilized for the period 1995/96–2019/20. The PM_{10} concentrations are strongly influenced by anthropogenic emissions as well as yellow dust transported from Mongolia or Inner Mongolia, northern China (Lee and Kim 2018). During large-scale yellow dust intrusion events, PM_{10} values of hundreds to even more than 1000 μm m^{-3} are observed for several days (Chung 1992). Therefore, to prevent the undue impact of extreme PM_{10} from excessively influencing the model configuration, we exclude yellow dust days declared by the Korea Meteorological Administration (a total of 37 days during the analysis period) when taking monthly averages. However, there is little difference in the results even if the yellow dust days are included.

The observation stations are in a variety of locations in South Korea, from large cities to rural areas, resulting in differences in the mean values by station. Due to the spatial heterogeneity of the stations, using an overall average value can be problematic for representativeness (Heo *et al* 2017). Therefore, the principal component time series (PC time series) of PC1, the first mode of the empirical orthogonal function (EOF) of 153 stations across South Korea, was used as an index indicating the PM_{10} variation. Figure 2(a) shows the leading EOF of South Korean PM_{10} (EOF1) for the period 2001/02–2019/20, which explains 48.87% of the total variability in winter PM_{10} across all the stations. EOF1 reveals a uniform anomaly pattern across South Korea, with high variability in major cities like Seoul and Busan. To extend the research period as much as possible, for the period 1995–2000, the PM_{10} averaged at 12 stations in Seoul were used as a proxy for the South Korean PM_{10}. This value closely matches South Korea's PC1 value with a correlation of 0.97 during the overlapping period 2001–2020 (figure 2(b)). The Seoul averages were standardized to align with the South Korean averages during the overlapping period.

During the study period, the PM_{10} in South Korea showed a clear decreasing trend (figure 2(b)), likely due to emission reduction policies in South Korea and neighboring countries like China and Japan since the late 1990s (Heo *et al* 2017, Dong *et al* 2020, Li *et al* 2020b, Ito *et al* 2021). Since the objective of the hybrid model is to predict the PM_{10} variation influenced by climate variability, we construct the PM_{10} forecast model with the linear trend removed for the model training period. The results of adding the trend to the model forecast are also shown for reference.

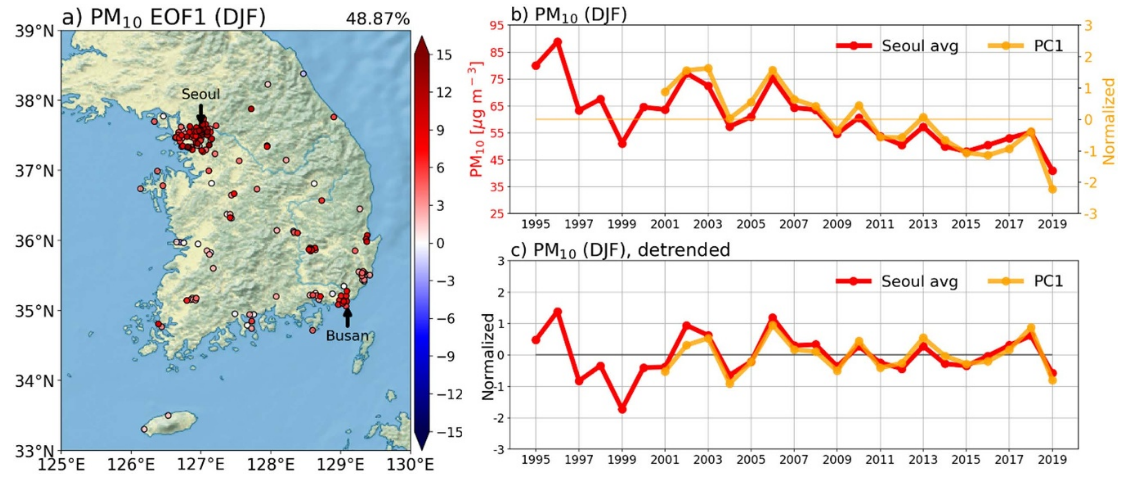

Figure 2. (a) The leading (1st) EOF of DJF average PM_{10} concentration for 153 stations in South Korea for the period 2001/02–2019/20, and (b) corresponding principal component (PC) time-series (yellow line). The red line shows the average PM_{10} concentration in winter at 12 stations in Seoul from 1995/96–2019/20. (c) Same as (b) but with removing linear trend.

Table 1. A summary of forecast and hindcast datasets from APCC MME.

Model	Period (HIND, FCST)	Ens. Size
APCC SCOPS	1995/96–2013/14, 2018/19–2019/20	10
BOM ACCESS-S2	1995/96–2019/20	27
CMCC SPS3.5	1995/96–2016/17	40
ECCC CANSIPSv2.1	1995/96–2019/20	20
KMA GLOSEA6GC3.2	1995/96–2016/17	12
NCEP CFSv2	1995/96–2010/11, 2015/16–2019/20	20
UKMO GLOSEA6	1995/96–2016/17	28
PNU-RDA CGCMv2.0	1995/96–2019/20	35

Observations (reanalysis) of atmospheric variables and sea surface temperature (SST) were used for the climate predictors in the model. The following variables were selected to capture representative features of the East Asian winter monsoon circulation, the high latitude-East Asia teleconnections, the tropical-East Asia teleconnections, and El Niño, which are known to be large-scale climatic conditions that influence winter PM_{10} behavior in East Asia. The following variables were used: geopotential height at 500 hPa (Z500), mean sea level pressure (MSLP), zonal and meridional wind at 200 hPa (U200 and V200, respectively), temperature at 850 hPa and 2 m (T850 and T2m, respectively) from ERA5 (Hersbach *et al* 2020), the 5th generation global climate reanalysis of the European Center for Medium-Range Weather Forecasts (ECMWF), and SST from Extended Reconstructed SST, Version 5 (ERSST, Huang *et al* (2017)). These variables are also available in the APCC MME, which will be discussed later, as they are eventually used in the forecast.

2.3. MME climate prediction

The climate variables used as climate predictors are obtained from the APCC MME seasonal forecast (Min *et al* 2017). APCC collects seasonal climate forecast products from 15 institutions in 11 countries every month and produces the MME climate forecast. Its hindcasts (retrospective forecasts from the most updated version of the model) or forecast archives (forecasts from the model version at the forecast made) for the past are provided as well. We used hindcasts/forecast archive (hindcast hereafter) data from eight models with as many hindcasts as possible in the target period of this study (from winter 1995/96–2019/20, table 1). Although the periods of hindcasts for participating models are different, all available hindcasts were used to maximize the sample size. Considering the operational use of the hybrid model, the winter climate predictions of APCC MME initialized in October were used in this study. We calculated the MME by averaging the ensembles of all models participating in this study.

3. Result

3.1. Potential climate predictors

Figure 3 illustrates the selection criteria for climate predictors based on their (1) strong relevance to PM_{10} in South Korea and (2) reliable prediction skill in the APCC MME. Color shading indicates the correlation between climate variables and PM_{10} concentrations

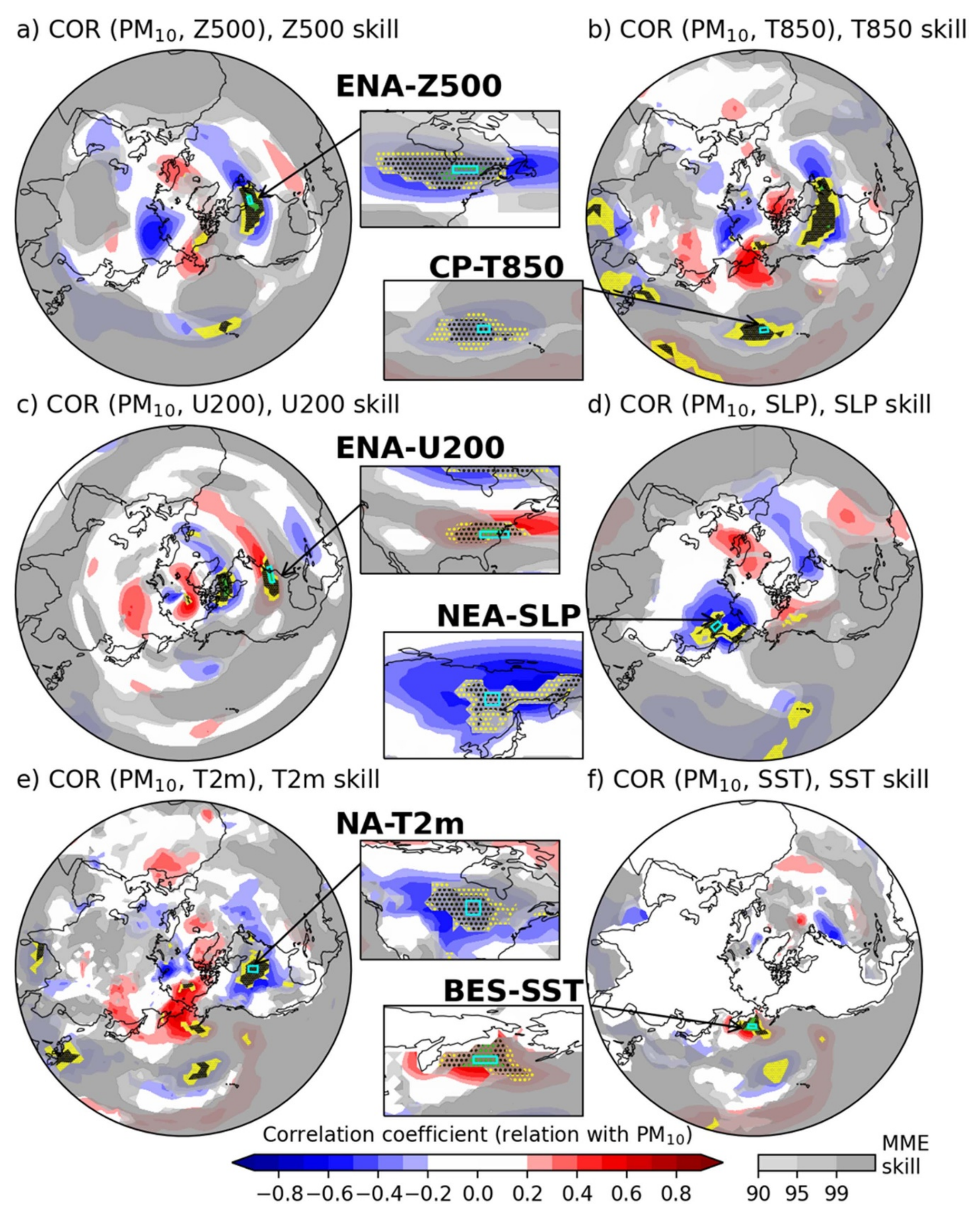

Figure 3. Color shading indicates the anomaly correlation coefficient (ACC) between the winter PM_{10} and climate variables. Gray shading highlights regions where APCC MME shows statistically significant (at the 90, 95, and 99% significance levels) forecast skills for that climate variable. The dotted pattern (yellow, black, green) indicates regions where both the ACC of PM_{10}-climate variables and forecast skills are statistically significant.

(criterion 1). The gray shading represents the forecast skill matrix (criterion 2) of the APCC MME, showing the correlation between the hindcasts and the observation. Potential climate predictors exist where statistically significant signals for both conditions (color and gray shading) overlap. For each variable, 2–3 potential climate variables are detected.

Several atmospheric teleconnection patterns are found for criterion 1 (shading). Focusing on signals connected to the East Asian region, 500 hPa geopotential height (Z500, figure 3(a)) shows a negative correlation over the Barents Sea region and a positive correlation over East Asia. There are positive temperature anomalies over northeast Mongolia-China (figures 3(b) and (e)). These are typical large-scale climate patterns found when East Asian winter monsoon circulation is weakened. This result is in agreement with the negative correlation between PM_{10} and the East Asian monsoon suggested by (Jeong and Park 2017). The weakening of the winter monsoon

Table 2. Domains and variables of the potential climate predictors.

Potential predictor	ENA-Z500	CP-T850	ENA-U200	NEA-SLP	NA-T2m	BES-SST
Longitude (°E)	282–288	183–187	277–287	128–134	267–272	174–182
Latitude (°N)	48–50	26–28	37–39	57–62	46–51	54–56

means the weakening of the cold northwesterly winds blowing from Siberia to East Asia, including South Korea. The weakening of northerly wind in East Asia are favorable conditions of atmospheric stagnation and the ventilation of air pollutants. The pattern is also similar to the teleconnection patterns related to warming in the Arctic and SST warming in the Atlantic (Jung *et al* 2017). This is not only due to the weakening of the seasonal wind, but also to the strengthening of the high-pressure circulation around South Korea, which prevents the spread of pollutants due to the stagnation of the atmosphere. Another major pattern found is the teleconnection between the tropical Pacific and the Atlantic. The warming in the western Pacific (figure 3(b)) and the positive SST anomaly pattern in the tropical Atlantic (figure 3(f)) appear to be related to the weakening of the East Asian winter monsoon induced by tropical-midlatitude teleconnection (Wang *et al* 2000, Ma *et al* 2018a, 2018b). These show that Criteria 1 is a good representation of the physical teleconnections affecting the East Asian climate and PM_{10} in South Korea.

For the MME prediction skill (figure 3, gray shading), statistically significant high skill is found mainly in the tropics, especially in the Pacific and Atlantic regions. This is a typical limitation of climate models but compared to a previous study that used single-model projections (Jeong *et al* 2022), MMEs significantly expand the region of significant skill from the tropics to the mid-latitudes.

Based on these results, we identified six potential predictors (ENA-Z500, CP-T850, ENA-U200, NEA-SLP, NA-T2m, and BES-SST). ENA-Z500 refers to Z500 in the eastern Americas, CP-T850 to T850 in the central Pacific, ENA-U200 to U200 in the eastern Americas, NEA-SLP to SLP in Northeast Asia, NA-T2m to T2m in North America, and BES-SST to SST in the Bering Sea. That satisfied both criteria: relevance to South Korea's PM_{10} variability and forecast skill of APCC MME for the climate variables (table 2). Note that local temperature may be the most effective predictor due to its direct impact on PM_{10} variability (Lee *et al* 2011, Kim 2019). However, it does not meet the MME predictability criteria (i.e. poor skill), so it was excluded as a potential predictor.

3.2. Dynamical-statistical model and its forecast skill

3.2.1. MME model

The six potential climate predictors can be combined to construct a total of 15 MLR models. As shown in figure 3 and supplementary figure S1, however, some of the six potential climate predictors appear to stem from the same teleconnection pattern and are therefore inter-correlated. To avoid the multicollinearity problem, we examined the cross-correlations between these climate predictors (figure S1) to select variables that were as statistically independent of each other as possible. Among these, three candidate MLR models were constructed using relatively independent variables (figure S2). We performed experimental forecasting for the past 25 years using the predictors obtained from APCC MME. For comparison, we also ran a hindcast experiment of perfect climate forecast model where we input each year's observed values to the climate predictors of MLR. The perfect climate forecast model's forecast represents the PM_{10} variation that can be explained by the climate variables, representing the maximum skill of the model.

The final MLR model selected (shows best skill and lowest correlation between two predictors) was the forecast model using ENA-U200 (Zonal wind at 200 hPa over eastern North America) and BES-SST (Bering Sea SST). The correlation between the two climate predictors (ENA-U200 and BES-SST) is 0.22. The MLR of this model is given in equation (1) below. Since the leave-one-out cross-validation was applied, the coefficients (i.e. MLR model) vary slightly from year to year, but for the sake of simplicity, we present the values obtained over the entire period here,

$$y = 0.49 * \mathrm{ENA} - \mathrm{U200} + 0.54 * \mathrm{BES} - \mathrm{SST} + 0.00. \quad (1)$$

Overall, the model demonstrates statistically significant skill over 25 year period, with correlation coefficients of $r = 0.504$ for the perfect model and $r = 0.439$ for observations. Including the linear trend raises these values to $r = 0.851$ and $r = 0.736$, respectively. The lower correlation for observed PM_{10} values is attributed to non-climatic factors like anthropogenic emissions.

To confirm that the performance of these models resulted from the physical relationship between climate conditions and PM_{10}, we examined the chosen two predictors' physical relevance to PM_{10}. Figure 4 shows the regression analyses of the two predictors on climate variables. First, it is found that ENA-U200 is associated with a wave teleconnection pattern propagating from the North Atlantic to the eastern Eurasian continent through the Barents-Kara Sea (figures 4(a) and (b)). The tripole pattern, triggered

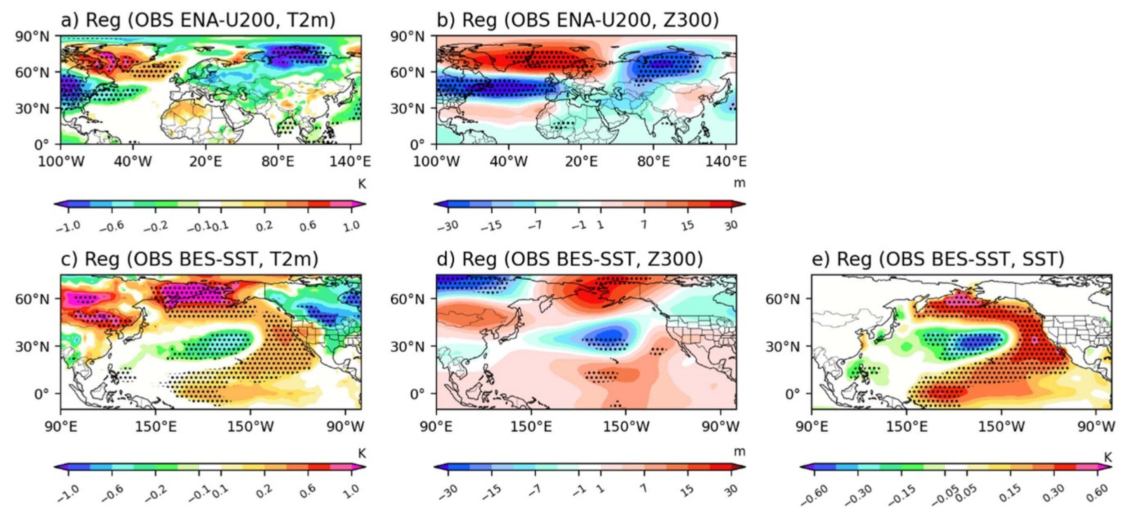

Figure 4. ENA-U200 regressed on (a) 2 m temperature (T2m) and (b) geopotential height at 300 hPa (Z300), and BES-SST regressed on (c) T2m, (d) Z300, and (e) sea surface temperature (SST). Values that are statistically significant at 90% are indicated with black dots.

by Rossby waves from North Atlantic SST anomalies, results in warmer temperatures and high pressure over East Asia, a weaker East Asian winter monsoon circulation (Ding and Yihui 1993, Chang and Lu 2012). The weakening of the prevailing northwesterly winds and high-pressure causes atmospheric stagnation, that is conducive to higher PM_{10} concentrations (Jeong and Park 2017). Second, the BES-SST-related anomalies indicate a remote influence from ENSO. Positive BES-SST is associated with high-pressure anomalies and higher temperatures in the Mongolia-East Asia region, further weakening the monsoon, as depicted in figures 4(c) and (d). The regression between BES-SST and SST (figure 4(e)) shows a pattern of teleconnection from the western Pacific to East Asia associated with tropical central Pacific warming during El Niño. In addition to the weakened East Asia winter monsoon, the tropical teleconnection delivers moist air masses from the western Pacific, affecting precipitation in East Asia (Ma *et al* 2018a). This directly affects air quality. These results indicate that the two predictors reflect high-latitude and tropical factors that influence East Asian climate, respectively. Their combined impact seems to influence PM_{10} variability in South Korea.

3.2.2. The effect of MME on forecast skill

The most important benefit of MME climate forecasting is to maximize climate predictability by offsetting errors in the modeling system through ensemble averaging across models. We aimed to assess whether the incorporation of MME climate forecasts into a hybrid model indeed enhances the predictability of PM_{10}, as compared to utilizing forecasts from individual models.

Figures 5 and 6 compare the PM_{10} forecast using the MME and the individual model's forecast (each ensemble) as climate predictors. Different MMEs and different individual models produce different predictions to some extent. However, overall, the MME models generally demonstrate the highest performance. It is interesting to note that one single model (ECCC) outperforms most MMEs. While it is difficult to conclude that this model is the best given the limited experimental periods, it does indicate the potential for additional skill when using the best model or combination of models. Various MME methods, like the super ensemble method (Krishnamurti *et al* 2009) that constructs optimal ensembles through MLR between model results and observations, and the weighted ensemble method (Kug *et al* 2008) that uses singular value decomposition, can further improve skill beyond the simple averaging approach used in this study.

To this extent, we attempted to predict the PM_{10} using the MME only consisting of models that exhibited the best skill for two climate predictors (ENA-U200 and BES-SST). Configuring the MME in this way could raise the issue of modifying the skill matrix of the MME. However, here, the MLR derived from the original MME was utilized without alteration, focusing only on enhancing performance through the optimal MME. First, Six climate models—APCC_SCOPS, BOM_ACCESS-S2, ECCC_CANSIPSv2.1, CMCC_SPS3.5, UKMO_GLOSEA6, KMA_GLOSEA6GC3.2 (referred to as TOP6)—were selected for their high predictability (table 3; r meeting Pearson's critical values of 0.05 or lower). The PM_{10} forecast was then produced by applying MME from these six models as input predictor values (figure 5, solid lines). Considering the differences in the predictabilities of ENA-U200 and BES-SST among the TOP6, we tested the PM_{10} forecasts using three different

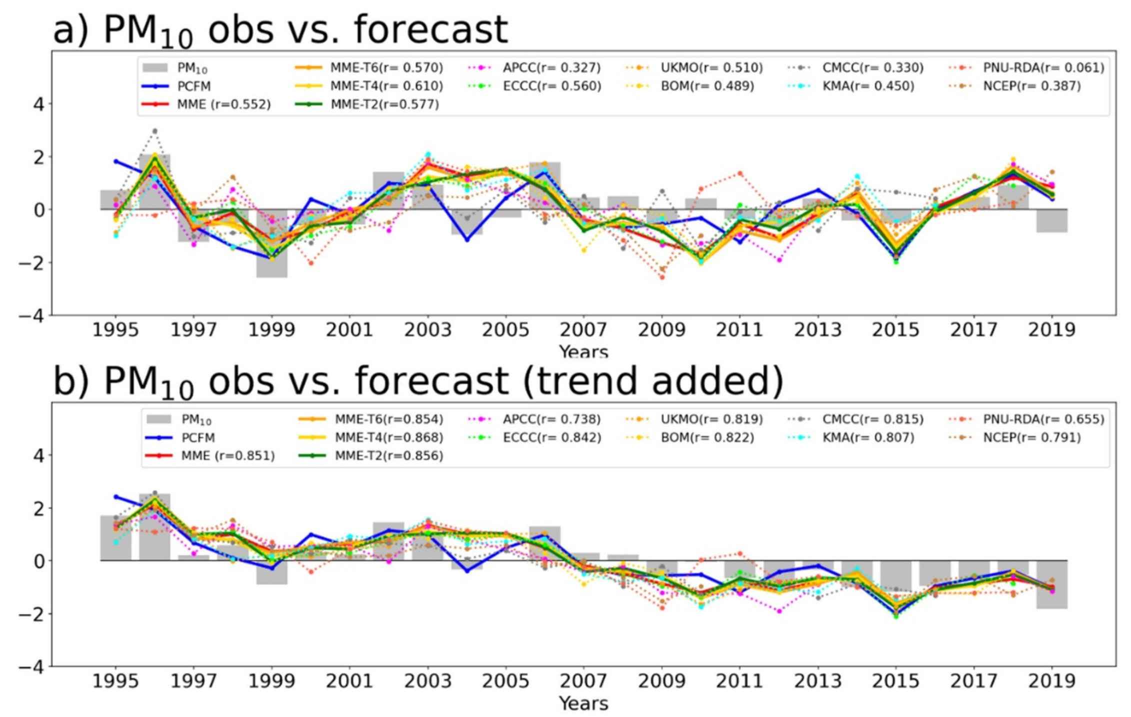

Figure 5. Forecasted and observed time series of winter PM_{10} concentrations in South Korea. Gray bars indicate PM_{10}. The blue line represents the perfect climate forecast model (PCFM)'s result. The red line represents the forecast from the dynamical-statistical model with MME. The dotted lines are the forecasts from individual participating single models, and the selected MMEs: MME-T2 (MME of ECCC_CANSIPSv2.1 and BOM_ACCESS-S2), MME-T4 (MME of CMCC_SPS3.5, UKMO_GLOSEA6 and MME -T2), and MME-T6 (TOP6 MME of APCC_SCOPs, KMA_GLOSEA6GC3.2 and MME-T4). All forecasts were conducted using leave-one-out cross-validation. (a) Result of the detrended model and (b) that with a linear trend added.

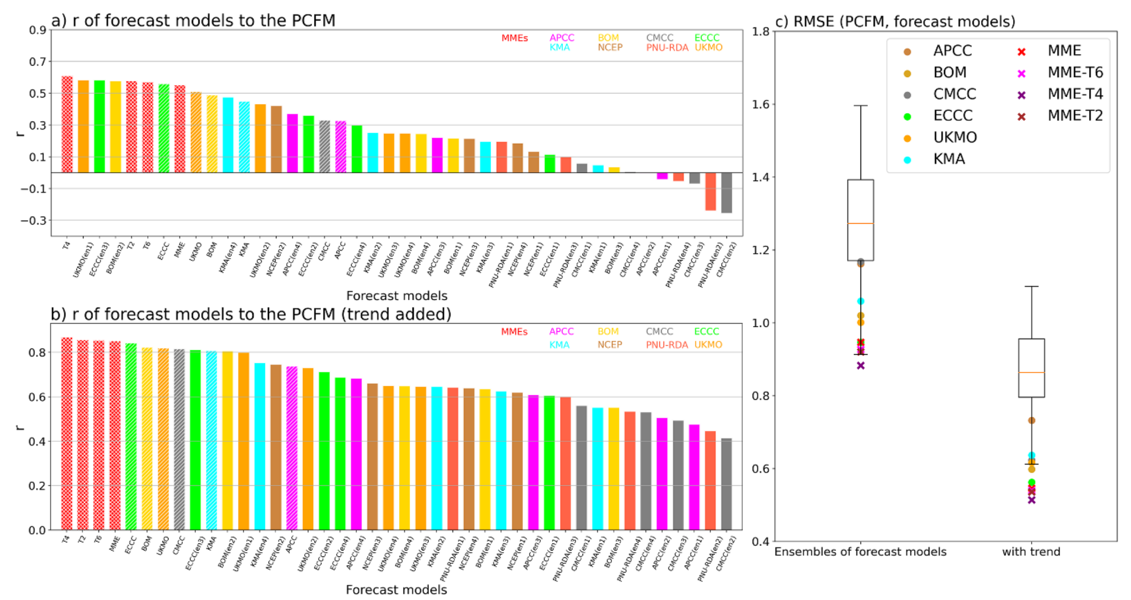

Figure 6. (a) Forecast skills of MMEs (MME, MME-T6, MME-T4, MME-T2) and ensemble averages from eight individual models. For equal comparisons across models, each model's ensemble average was calculated by randomly selecting an equal number (4) to produce 4 ensemble averages. (b) Result with liner trend added. (c) RMSE of MME models and TOP6. Boxes are RMSE of 32 ensembles. The orange line is the median RMSE of the ensembles.

combinations of the MME. These three combinations are as follows: MME-T6, comprising all TOP6 models; MME-T4, comprising BOM_ACCESS-S2, ECCC_CANSIPSv2.1, CMCC_SPS3.5, and UKMO_GLOSEA6 which exceeded 0.462 (r meeting Pearson's critical values of 0.02 or lower); MME-T2 comprising ECCC_CANSIPSv2.1, and BOM_ACCESS-S2, which exceeded 0.505 (r meeting

Table 3. The correlation coefficient between observation and predictors (ENA-U200 and BES-SST) predicted by the individual model and MME. An asterisk indicates that both predictors have a Pearson critical value of 0.05 or less, corresponding to a correlation coefficient of 0.396 or higher. Two asterisks signify that both predictors meet a Pearson critical value of 0.02 or less, with a correlation coefficient of 0.462 or higher. Three asterisks mean that both predictors satisfy a Pearson critical value of 0.01 or less, indicating a correlation coefficient of 0.505 or higher.

ENA-U200		BES-SST	
ECCC***	0.703	NCEP	0.647
UKMO**	0.676	BOM***	0.627
KMA*	0.653	ECCC***	0.572
BOM***	0.566	CMCC**	0.514
CMCC**	0.485	UKMO**	0.494
APCC*	0.417	PNU-RDA	0.432
PNU-RDA	0.346	APCC*	0.411
NCEP	0.323	KMA*	0.403
MME	0.694	MME	0.609
MME-T6	0.703	MME-T6	0.611
MME-T4	0.714	MME-T4	0.625
MME-T2	0.684	MME-T2	0.625

Pearson's critical values of 0.01 or lower). The forecast skills were ranked as MME-T4 (0.610), MME-T6 (0.570), and MME-T2 (0.577), and MME (0.552). RMSE analysis (figure 6(c)) confirmed that MME forecasts generally outperformed single model predictions, regardless of trend adjustments.

In addition, we evaluated the skill of the developed model using other conventional skill metrics such as mean absolute error (MAE, in figure S4) and mean absolute scaled error (MASE, in figure S5). The developed model shows useful skill both in terms of MAE and MASE, with the MME indicating better performance than the individual models. We also evaluated the skill of the developed model as a classification model to determine if it predicts high PM_{10} concentration years with ROC curve and its area under the curve (AUC) values (see details in figure S6 caption). MME shows fairly good performance with an AUC of 0.85, and MME-T2 and MME-T4, show a considerably high performance of 0.9. These results also demonstrate the usefulness of the developed model forecast.

4. Summary and discussion

Using the APCC MME climate forecasts and weather/climate-PM_{10} relationship, we developed a hybrid dynamical-statistical model that skillfully predicts winter PM_{10} in South Korea. By performing hindcast experiments for the past 25 years, we verified that this model has statistically significant forecast skill. Our findings suggest that the usage of MME benefits the forecast skill, and optimizing the MME can further boost the skill. This result demonstrates the possibility of skillful seasonal forecast of air quality.

Although the developed hybrid model uses dynamical models' forecast, it has distinct limitations of statistical models. It is difficult to reflect long-term fluctuations on interdecadal or longer time-scale due to the nature of statistical models. Additionally, the evolving relationship between climate and air quality complicates predictions amidst rapid environmental changes. However, clearly, finding predictable parts from MME prediction results and making local PM_{10} prediction possible is a significant advantage of the dynamical model. This strategy could extend to other regions and pollutants, opening new avenues for developing effective seasonal air quality forecasts.

While still in its infancy, it will eventually be possible to make simultaneous climate-air quality forecasts using models that combine atmospheric chemistry, climate, and human activities. This will allow for the consideration of emissions, atmospheric chemistry-climate interactions, and the effects of long-term climate variability that are currently excluded from this study. Further research is needed to integrate the continuously improving climate modeling and assimilation techniques, and rapidly developing artificial intelligence techniques.

Data and code availability statement

All data used in this study are freely accessible. The PM10 concentration data in Korea and Seoul can be accessed through the following link: www.airkorea.or.kr/web/last_amb_hour_data?pMENU_NO=123 and https://data.seoul.go.kr/dataList/OA-2218/S/1/datasetView.do#, respectively. The ERA5 atmospheric reanalysis data can be found here: https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels-monthly-means?tab=overview. The SST data from Extended Reconstructed SST, Version 5, can be assessed here: https://psl.noaa.gov/data/gridded/data.noaa.ersst.v5.html. The APCC MME seasonal forecasts/hindcasts can be downloaded from here: https://cliks.apcc21.org/dataset/mme/6-MON. The code is available by request via e-mail at jjeehoon@jnu.ac.kr.

Acknowledgments

This work was supported by the Korea Environment Industry & Technology Institute (KEITI) through the 'Climate Change R&D Project for New Climate Regime'., funded by the Ministry of Environment (MOE) of the Republic of Korea (RS-2022-KE002160); the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2021R1A2C1011178); and a grant from the National Institute of Environment Research (NIER), funded by the Ministry of Environment (MOE) of the Republic of Korea (NIER-2023-01-02-075). We thank Professor Tae-Won Park at Chonnam National University for his invaluable help in revising this manuscript. We also acknowledge that the APCC Multi-Model Ensemble (MME) Producing Centers for making their hindcast/forecast data available for analysis and the APEC Climate Center for collecting and archiving the data, as well as for producing APCC MME prediction.

ORCID iDs

Jahyun Choi https://orcid.org/0009-0003-0310-880X
Jin-Ho Yoon https://orcid.org/0000-0002-4939-8078
Jee-Hoon Jeong https://orcid.org/0000-0002-3358-3949

References

Baklanov A, Schlünzen K, Suppan P, Baldasano J, Brunner D, Aksoyoglu S, Carmichael G, Douros J, Flemming J and Forkel R 2014 Online coupled regional meteorology chemistry models in Europe: current status and prospects *Atmos. Chem. Phys.* **14** 317–98
Chang C-P and Lu M-M 2012 Intraseasonal predictability of Siberian high and east Asian winter monsoon and its interdecadal variability *J. Clim.* **25** 1773–8
Chung Y-S 1992 On the observations of yellow sand (dust storms) in Korea *Atmos. Environ.* A **26** 2743–9
Ding Y and Yihui D 1993 *Monsoons over China* vol 16 (Springer)
Dong Z, Wang S, Xing J, Chang X, Ding D and Zheng H 2020 Regional transport in Beijing-Tianjin-Hebei region and its changes during 2014–2017: the impacts of meteorology and emission reduction *Sci. Total Environ.* **737** 139792
Heo J, Park J-S, Kim B M, Kim S-W, Park R J, Jeon H and Yoon S-C 2017 Two notable features in PM 10 data and analysis of their causes *Air Qual. Atmos. Health* **10** 991–8
Hersbach H, Bell B, Berrisford P, Hirahara S, Horányi A, Muñoz-Sabater J, Nicolas J, Peubey C, Radu R and Schepers D 2020 The ERA5 global reanalysis *Q. J. R. Meteorol. Soc.* **146** 1999–2049
Huang B, Thorne P W, Banzon V F, Boyer T, Chepurin G, Lawrimore J H, Menne M J, Smith T M, Vose R S and Zhang H-M 2017 Extended reconstructed sea surface temperature, version 5 (ERSSTv5): upgrades, validations, and intercomparisons *J. Clim.* **30** 8179–205
Ito A, Wakamatsu S, Morikawa T and Kobayashi S 2021 30 years of air quality trends in Japan *Atmosphere* **12** 1072
Jeong J I and Park R J 2017 Winter monsoon variability and its impact on aerosol concentrations in East Asia *Environ. Pollut.* **221** 285–92
Jeong J I, Park R J, Yeh S-W and Roh J-W 2021 Statistical predictability of wintertime $PM_{2.5}$ concentrations over East Asia using simple linear regression *Sci. Total Environ.* **776** 146059
Jeong J-H, Choi J, Jeong J-Y, Woo S-H, Kim S-W, Lee D, Lee J-B and Yoon J-H 2022 A novel statistical-dynamical method for a seasonal forecast of particular matter in South Korea *Sci. Total Environ.* **848** 157699
Jung M-I, Son S-W, Kim H and Chen D 2022 Tropical modulation of East Asia air pollution *Nat. Commun.* **13** 5580
Jung O, Sung M-K, Sato K, Lim Y-K, Kim S-J, Baek E-H, Jeong J-H and Kim B-M 2017 How does the SST variability over the western North Atlantic Ocean control Arctic warming over the Barents–Kara Seas? *Environ. Res. Lett.* **12** 034021
Keller C A *et al* 2021 Description of the NASA GEOS composition forecast modeling system GEOS-CF v1.0 *J. Adv. Model Earth Syst.* **13** e2020MS002413
Kim M J 2019 Changes in the relationship between particulate matter and surface temperature in seoul from 2002–2017 *Atmosphere* **10** 238
Kirtman B P *et al* 2014 The North American multimodel ensemble: phase-1 seasonal-to-interannual prediction; Phase-2 toward developing intraseasonal prediction *Bull. Am. Meteorol. Soc.* **95** 585–601
Krishnamurti T N, Mishra A K, Chakraborty A and Rajeevan M 2009 Improving global model precipitation forecasts over India using downscaling and the FSU superensemble. Part I: 1–5-day forecasts *Mon. Weather Rev.* **137** 2713–35
Ku H-Y, Noh N, Jeong J-H, Koo J-H, Choi W, Kim B-M, Lee D and Ban S-J 2021 Classification of large-scale circulation patterns and their spatio-temporal variability during high-PM10 events over the Korean Peninsula *Atmos. Environ.* **262** 118632
Kug J-S, Lee J-Y, Kang I-S, Wang B and Park C-K 2008 Optimal multi-model ensemble method in seasonal climate prediction *Asia-Pac. J. Atmos. Sci.* **44** 259–67
Kukkonen J *et al* 2012 A review of operational, regional-scale, chemical weather forecasting models in Europe *Atmos. Chem. Phys.* **12** 1–87
Lee G, Ho C-H, Chang L-S, Kim J, Kim M-K and Kim S-J 2020 Dominance of large-scale atmospheric circulations in long-term variations of winter PM10 concentrations over East Asia *Atmos. Res.* **238** 104871
Lee J and Kim K 2018 Analysis of source regions and meteorological factors for the variability of spring PM10 concentrations in Seoul, Korea *Atmos. Environ.* **175** 199–209
Lee S, Ho C-H and Choi Y-S 2011 High-PM_{10} concentration episodes in Seoul, Korea: background sources and related meteorological conditions *Atmos. Environ.* **45** 7240–7
Li F, Wan X, Wang H, Orsolini Y J, Cong Z, Gao Y and Kang S 2020a Arctic sea-ice loss intensifies aerosol transport to the Tibetan Plateau *Nat. Clim. Change* **10** 1037–44
Li W, Shao L, Wang W, Li H, Wang X, Li Y, Li W, Jones T and Zhang D 2020b Air quality improvement in response to intensified control strategies in Beijing during 2013–2019 *Sci. Total Environ.* **744** 140776
Ma T, Chen W, Feng J and Wu R 2018a Modulation effects of the East Asian winter monsoon on El Niño-related rainfall anomalies in southeastern China *Sci. Rep.* **8** 14107
Ma T, Chen W, Nath D, Graf H-F, Wang L and Huangfu J 2018b East Asian winter monsoon impacts the ENSO-related Teleconnections and North American seasonal air temperature prediction *Sci. Rep.* **8** 6547
Mariotti A, Ruti P M and Rixen M 2018 Progress in subseasonal to seasonal prediction through a joint weather and climate community effort *npj Clim. Atmos. Sci.* **1** 4
Min Y-M, Kryjov V N, Oh S M and Lee H-J 2017 Skill of real-time operational forecasts with the APCC multi-model ensemble

prediction system during the period 2008–2015 *Clim. Dyn.* **49** 4141–56
Ministry of Environment 2020 Enforcement decree of the special act on the reduction and management of fine dust
Peduzzi P, Concato J, Kemper E, Holford T R and Feinstein A R 1996 A simulation study of the number of events per variable in logistic regression analysis *J. Clin. Epidemiol.* **49** 1373–9
Peuch V-H *et al* 2022 The copernicus atmosphere monitoring service: from research to operations *Bull. Am. Meteorol. Soc.* **103** E2650–68
Shen L and Mickley L J 2017 Seasonal prediction of US summertime ozone using statistical analysis of large scale climate patterns *Proc. Natl Acad. Sci.* **114** 2491–6
Smith D M *et al* 2019 Robust skill of decadal climate predictions *npj Clim. Atmos. Sci.* **2** 13
Sokhi R S *et al* 2022 Advances in air quality research – current and emerging challenges *Atmos. Chem. Phys.* **22** 4615–703
Swinbank R *et al* 2016 The TIGGE project and its achievements *Bull. Am. Meteorol. Soc.* **97** 49–67
Wang B, Wu R and Fu X 2000 Pacific–East Asian teleconnection: how does ENSO affect East Asian climate? *J. Clim.* **13** 1517–36