Science of the Total Environment 848 (2022) 157699

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

A novel statistical-dynamical method for a seasonal forecast of particular matter in South Korea

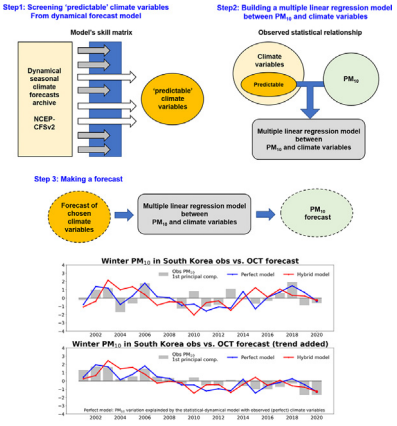
Jee-Hoon Jeong [a], Jahyun Choi [a], Ji-Yoon Jeong [a], Sung-Ho Woo [a], Sang-Woo Kim [b], Daegyun Lee [c], Jae-Bum Lee [c], Jin-Ho Yoon, Ph.D. [d,*]

[a] Faculty of Earth and Environmental Sciences, Chonnam National University, Gwangju 61186, Republic of Korea
[b] School of Earth and Environmental Sciences, Seoul National University, Seoul 08826, Republic of Korea
[c] Air Quality Forecasting Center, National Institute of Environmental Research, Incheon 22689, Republic of Korea
[d] School of Earth Sciences and Environmental Engineering, Gwangju Institute of Science and Technology, Gwangju 61005, Republic of Korea

HIGHLIGHTS

- Develop a novel method for a seasonal forecast of winter particular matter 10 in South Korea.
- The relationship between seasonal air quality and climate conditions was identified.
- Arctic and tropical climate factors were discovered as predictors of this model.
- This model is being used for the world's first semi-operational air-quality seasonal forecasting.

GRAPHICAL ABSTRACT



ARTICLE INFO

Editor: Anastasia Paschalidou

Keywords:
Air quality
Particular matter
Seasonal forecasting
Statistical model
Dynamical model

ABSTRACT

Societal concerns about air quality in East Asia are still growing despite country-level efforts to reduce air pollution emissions. In coping with this growing concern, the government and the public demand a longer-lead forecast of air quality to ensure sufficient response time until society prepares for countermeasures such as a temporary reduction of specific emission sources. Here we propose a novel method that produces skillful seasonal forecasting of wintertime (December to February) PM_{10} concentration over South Korea. The method is based on the idea that climate condition and air quality have co-variability in the seasonal time scales and that the state-of-art seasonal prediction model will benefit air quality forecasting. More specifically, a linear regression model is constructed to link observed winter PM_{10} concentration and climate variables where the predicted climate variables were furnished from NCEP CFSv2 forecast initialized during autumn. In this case, climate variables were selected as predictors of the model because they are not only physically related to air quality but also 'predictable' in CFS hindcast. Through analysis of retrospective forecasts of 20 winters for the period 2001–2020, we found this model shows statistically significant skill for the seasonal forecast of wintertime PM_{10} concentration.

* Corresponding author.
E-mail address: yjinho@gist.ac.kr (J.-H. Yoon).

http://dx.doi.org/10.1016/j.scitotenv.2022.157699
Received 19 May 2022; Received in revised form 14 July 2022; Accepted 25 July 2022
Available online 1 August 2022

1. Introduction

Air quality is a significant concern in East Asia, especially China and Korea. Fortunately, it has shown a gradual tendency to improve over the past two decades due to government regulations to reduce emissions. For example, the Korean government has regulated fugitive dust and exhaust gas emissions, encouraged the use of hybrid and electric vehicles, and stimulated a fuel source change toward natural gas (Ministry of Environment, 2016). Meanwhile, the Chinese government has legislated minimum emission standards to improve air quality (Clean Air Alliance of China, 2013). However, in recent years, the tendency has been slowed, and high-concentration particular matter (PM) episodes over South Korea have somewhat increased (Kim et al., 2017a; Park et al., 2021). Therefore, the air quality problem still poses tremendous social costs. In response to this, the government and the public demand a longer-lead forecast of bad air quality conditions to ensure long enough time for the society to mitigate its harmful impacts.

Air quality forecasting with lead times longer than even a day or two, however, is still a very challenging task. The most conventional method is using a chemical transport model, or air quality model with forecasted meteorological input (e.g., European Center for Medium Range Forecast (ECMWF)'s Copernicus Atmosphere Monitoring System; Morcrette et al., 2009). This forecast usually covers up to five days' lead time, which is bounded by a typical limit of weather forecast skill. Alternatively, the empirical relationship between meteorological and air quality variables has been used to develop statistical forecasting systems, but its practical skill is also very much limited to a few days (Koo et al., 2010). Even though overall skill for air quality is limited, there has been a growing need to extend the air quality forecast lead time to a few weeks and even to a season ahead for better management and preparation.

So far, such longer-lead seasonal forecasts of air quality have been done mostly through statistical methods. Since climate variables such as temperature, wind, and atmospheric stability affect the average behavior in air quality (Chattopadhyay et al., 2010; Lee et al., 2019; Yang et al., 2016; You et al., 2018), a statistical forecast model for air quality can be constructed by using observed lead-lag relationships between climate variables and air quality. For instance, Gao et al. (2010) and Sohn (2013) performed seasonal forecast of East Asian spring dust using its statistical association to climate variables. Yin and Wang (2017) attempted to predict winter haze in North China by constructing multiple linear regression (MLR) and the generalized additive model (GAM) using atmospheric circulation as a predictor. More recently, Shen and Mickley (2017) suggested a statistical model skillfully predicting US summertime ozone air quality based on large-scale climate patterns including the north Pacific SST in spring. Jeong et al. (2021) recently developed a statistical seasonal forecast model for $PM_{2.5}$ in East Asia using the statistical relationship and especially suggested that factors related to El Nino-Southern Oscillation (ENSO) and AO (Arctic Oscillation) provide predictability. Except for statistical models, successful seasonal forecasts based on numeric-based models have not been reported so far.

By virtue of tremendous advances in observation, data assimilation, and climate modeling techniques, the skill of seasonal climate forecasting by dynamical climate models has significantly improved over the past several centuries (Jeong et al., 2017). Although the skill of seasonal climate prediction is still modest in most extratropics, it has recently shown significant predictability in many regions directly or indirectly affected by ocean-atmosphere and land-atmosphere interactions (Baker et al., 2021; Jeong et al., 2017; Judt, 2020; Krishnamurthy, 2019). While predicting anthropogenic controls on air quality like human activities, emission, and policy is almost impossible, air quality linked to climate variabilities may be predictable to some extent. This study suggests a novel method for the seasonal forecast of regional PM concentration over South Korea based on this idea. Combining the empirical relationship between observed winter PM and climate variables and climate forecasts from a seasonal climate prediction system, we developed a statistical-dynamical forecast model for wintertime PM concentration of more than a month lead. This paper is organized as follows. Data and methods, including the model description, are described in Section 2. Results and discussion are provided in Sections 3 and 4.

2. Data and methods

2.1. Observation and forecast data

Monthly mean PM_{10} (particulate matter diameter 10 μm or less) concentrations at 153 stations over South Korea for the period 2001–2020 provided by the National Institute of Environmental Research of the Korean government are used. Finer particulates $PM_{2.5}$ have been receiving a lot of attention recently. However, in-situ observation of $PM_{2.5}$ available after 2015 in a large part of South Korea is too short in constructing the statistical-dynamical model targeted in this study. For the period 2015/16–2020/21, DJF PM_{10} and $PM_{2.5}$ show very similar inter-annual variability (the correlation coefficient between DJF PM_{10} and $PM_{2.5}$ is 0.97). Therefore, this study targets PM_{10}.

The target of this prediction model is the overall winter (DJF average) PM_{10} concentration throughout South Korea. PM_{10} concentrations in South Korea can sometimes surge due to yellow dust generated mainly in Mongolia or Inner Mongolia in China (Lee and Kim, 2018). In order to eliminate this effect, when calculating the DJF average, data on the occurrence of yellow dust (31 days out of 20 years of study) announced by the Korea Meteorological Administration were excluded. Since 153 monitoring stations include large cities, small towns, and rural areas, there is a significant difference in the average value for each station, especially in the early 2000s (Heo et al., 2017). There is also a large spatial inhomogeneity in the spatial distribution of the stations. Therefore, instead of using the average value of 153 stations, the principal component (PC) time-series of 1st Empirical Orthogonal Function (EOF) (Lorenz, 1956; Preisendorfer, 1988) mode is set as a value representative of PM_{10} concentration throughout South Korea.

For the seasonal forecast model of PM_{10}, multiple linear regression (MLR) model based on the historical relationship between climate variables and PM_{10} concentration for winter (December to February; DJF) is constructed. The observational climate variables such as near-surface air temperatures (SAT), mean sea level pressure (MSLP), and geopotential height at 300 hPa (GPH300) are from ERA5 reanalysis, which is the fifth generation ECMWF (European Centre for Medium Range Weather Forecasts) reanalysis of the global climate.

ERA5 is global reanalysis data calculated using the most advanced model and data assimilation system from a large amount of global observation data, and it is known that it has the best performance compared to other existing reanalysis products such as the JRA-55, CFSRv2, and MERRA-2 reanalysis (Hersbach et al., 2020; Jiang et al., 2021; Tarek et al., 2020). Especially, ERA5 benefits from Coupled Model Intercomparison Project5 (CMIP5) (Taylor et al., 2012) radiative forcing, a consistent historical set for sea-surface temperature (SST), sea-ice cover, and other boundary conditions, which provide improved variability and trend of surface air temperature, reduced precipitation bias, and improved hydrological balance (Hersbach et al., 2020). Due to these advantages, it is widely used in meteorological and climate research, and even in application fields (e.g., Aizpurua-Etxezarreta et al., 2022; Chen et al., 2021; Ding and Wu, 2020; Klingler et al., 2021; Yu et al., 2021). SST is taken from Extended Reconstructed SST, Version 5 (Huang et al., 2017).

The proposed statistical-dynamical seasonal forecast model utilizes the seasonal climate forecasts from the Coupled Forecast System model version 2 (CFSv2) (Saha et al., 2014), which consists of the atmosphere (horizontal resolution T126 of vertical 64 levels), ocean (horizontal resolution of 0.25 latitude by 0.5 longitude degree with vertical 40 levels), land surface, sea-ice model, and their interaction. As an operational seasonal forecast system of National Oceanic and Atmospheric Administration (NOAA), CFSv2 provides an extensive set of retrospective forecasts as well as real-time forecast products. The present study used both the retrospective 9-month forecasts (1982–2010) and real-time operational 9-month forecasts archived

(2011–2020). CFSv2 retrospective forecast runs four times a day (00, 06, 12, and 18 UTC) every five days from the 1st of January every year using the initial condition taken from the NCEP Climate Forecast System Reanalysis (Saha et al., 2010). The real-time operational forecast runs four times a day, every day from April 2011 with the operational Climate Data Assimilation System version 2 (Saha et al., 2014) initial conditions. We tested two sets of forecasts; one initialized in the second half of October (18th, 23th, and 28th) and the other initialized in the first half of November (2nd, 7th, and 12th), October and November forecast, respectively. Therefore, 12 ensembles were averaged for each prediction, a seasonal (DJF average) ensemble forecast. Forecasted anomalies were calculated with respect to 2001–2019 climatology.

2.2. Statistical-dynamical model

Fig. 1 provides the schematic of the statistical-dynamical model constructed in the present study. The final forecast product is provided by a multiple linear regression (MLR) model between observed winter PM_{10} concentration over South Korea (predictand) and climate variables (predictors). Climate variables are furnished by forecasted fields from CFSv2. While conventional statistical forecast models for air quality rely mostly on time-lagged relationships between predictand (PM_{10}) and predictors (climate variables) for preceding months or seasons (e.g., Jeong et al., 2021; Shen and Mickley, 2017), our model is based on simultaneous relationships. As will be described in more detail in the results, the variability of PM_{10} is related to various climate variables, and among them, the most efficient variable in a specific region that best explains PM_{10} variability is selected to compose the MLR. Additionally, our strategy is to select a climate variable for which CFSv2 shows sufficient (significant) skill and has a physical relevance to PM_{10} variability. The variables of sufficient skill are chosen by CFSv2's skill matrix estimated from its hindcast and forecast archive.

A hybrid (statistical-dynamical) forecast model is constructed based on predictors that have both a physical relation to PM_{10} and reliable CFSv2 predictability. A similar approach was adopted by Gillies et al. (2010) to establish a statistical relationship between local inversion and $PM_{2.5}$ over the Intermountain West and to produce a long-rage forecast using CFSv1 (Gillies et al., 2010). More practical details on building the MLR model are provided in the next section. By repeating retrospective forecast experiments with the developed model for the period 2001/02–2019/20, we examined the skill of the present model. In this study, forecast skill is defined as Pearson's correlation coefficient between forecast and observation (target variable) during the analysis period. More details on practical procedures are given in the following sections.

3. Results

3.1. Time-variation of PM_{10} over South Korea

Fig. 2a represents the spatial pattern of the leading EOF mode (EOF1 hereafter) of the winter PM_{10} concentration observed over 153 stations for the period 2001–2019 and associated PC time-series of EOF1 (PC1 hereafter). EOF1 shows positive values in most parts of South Korea. The largest values are detected, especially in the Seoul Metropolitan area, where a population of twenty-five million people resides. PC1 reveals a clear negative trend for the analysis period as well as considerable year-to-year variability. The time-variation of PC1 is almost consistent with the average PM_{10} concentration averaged over all 153 stations (red lines denoted as PM_{10}-avg in Fig. 2b). We set PC1 as the predictand of our forecast model.

Presumably, the negative trend in Fig. 2b resulted from government policies to suppress the source emission of PM (e.g., Heo et al., 2017). Because our primary goal is to forecast the year-to-year variation of PM_{10} concentration only associated with climate variability, we removed the linear trend for the studied period (2001–2020) from PM_{10} concentration as well as climate variables before building a prediction model.

3.2. Relationship between PM_{10} and climate variables

It is assumed here that the year-to-year variation of wintertime PM_{10} concentration is controlled largely by climate conditions. For instance, one of the well-known features is the reduced wind speed resulting in stagnant atmospheric conditions (Camalier et al., 2007; Jacob and Winner, 2009).

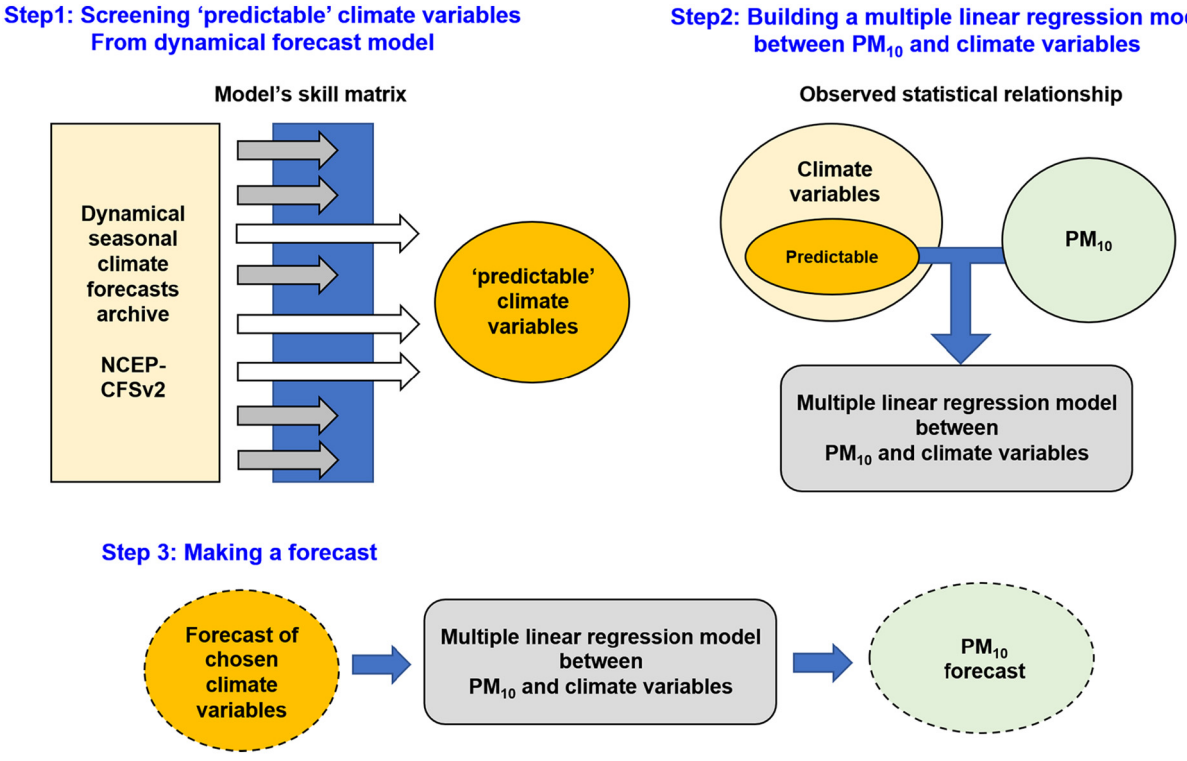


Fig. 1. Schematic diagram of the statistical-dynamical model for PM_{10} forecast.

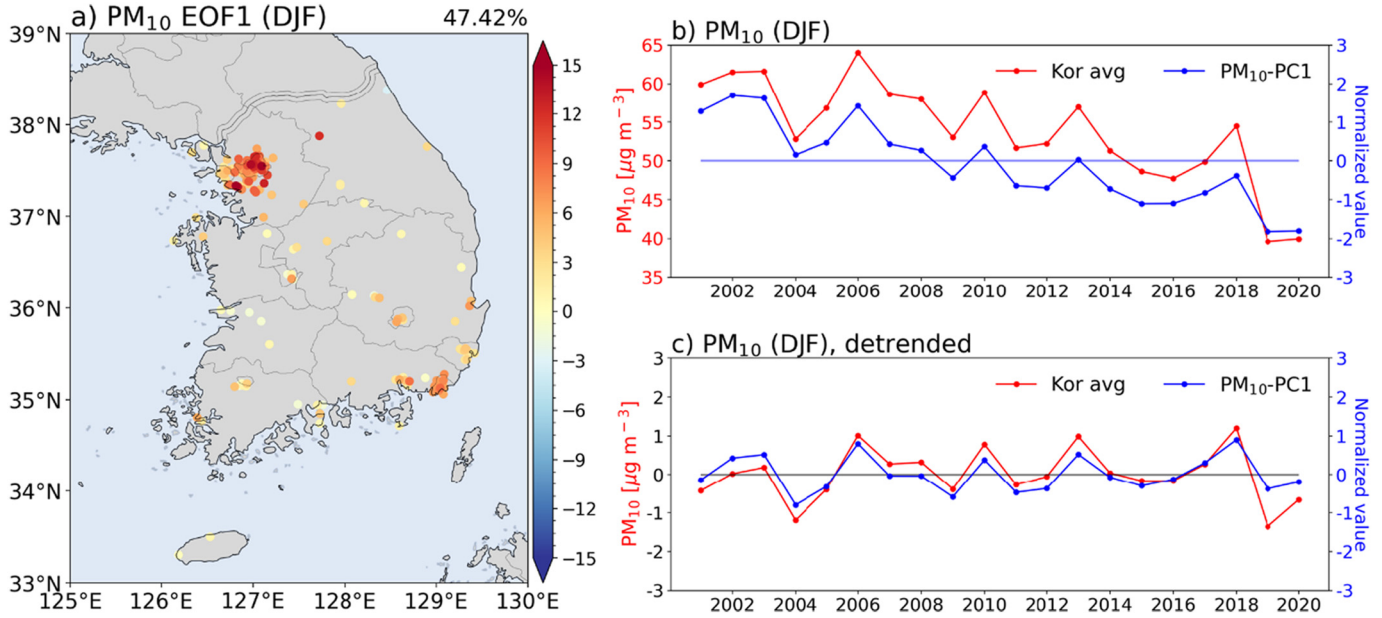


Fig. 2. a) The leading (1st) EOF of DJF average PM_{10} concentration for 153 stations in South Korea for the period 2001/02–2020/21, and b) associated leading principal component (PC1) time-series (blue line). PM_{10} concentration averaged for the 153 stations is also shown for comparison (red line). c) Same as b) but with removing linear trend.

Also, large-scale climate factors such as Siberian High intensity, Arctic Oscillation, ENSO, and Arctic-midlatitude teleconnection are known to affect the air quality in East Asia (Jeong and Park, 2017; Jeong et al., 2021; Ku et al., 2021; Wei et al., 2021; Yun and Yoo, 2019). To identify this, we examined the simultaneous anomaly correlation coefficient (ACC) between winter climate variables and PM_{10}_PC1 in Fig. 3. Key features associated with an anomalously high level of PM_{10} in South Korea are summarized as follows.

1) The positive and negative anomalies (correlation) of SAT over central Siberia to East Asia (Fig. 3a) and over the Barent-Kara Sea are the most distinct pattern. This is likely associated with the well-known wintertime seesaw pattern of SAT anomaly over the Arctic and Northern Eurasia. This is called the Warm-Arctic Cold-Eurasia pattern (Cohen et al., 2013) of SAT, which is suggested to be related to sea ice reduction over the Barents-Kara Sea (Honda et al., 2009; Kug et al., 2015). But, it is cold-Arctic and warm-East Asia in the correlation map of Fig. 3a.
2) MSLP anomalies in northern Siberia were apparent to be closely related to PM_{10} concentration in South Korea (Fig. 3b). The negative correlation in the Siberian region in Fig. 2b means the weakening of the Siberian high pressure in the winter when PM_{10} concentration is high. This relation implies that the East Asian winter monsoon is weakened, and the positive SAT anomaly over East Asia is favorable for the high PM_{10} concentration.
3) SAT and SST warming over the Chukchi Sea (Fig. 3a and d) may indicate a possible influence from atmospheric blockings, which often leads to

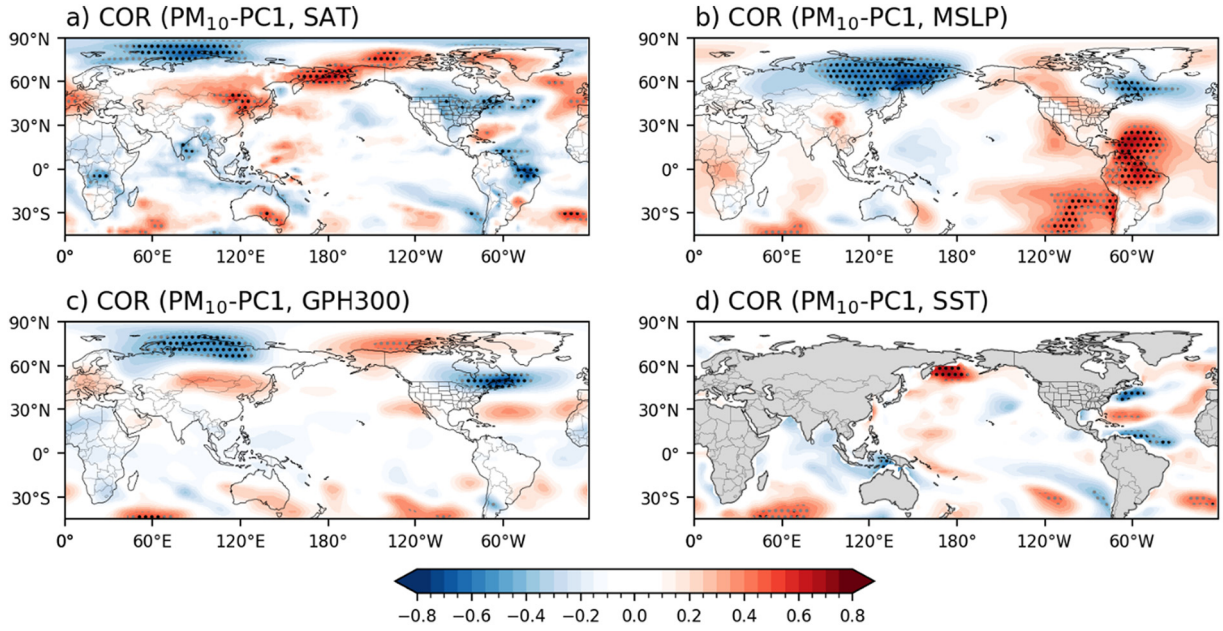


Fig. 3. The anomaly correlation coefficient between the PC1 of PM_{10} and a), b) MSLP, c) GPH at 300 hPa, and d) SST for winter for the period 2001/02–2019/20. Values statistically significant at 99 % (95 %) are indicated with black (gray) dots. All data were detrended before the analysis.

the stagnant condition causing pollutant episodes over East Asia (Kim et al., 2022; Takaya and Nakamura, 2005).

4) The remote influences from tropical ocean variabilities like El Niño-Southern Oscillation (ENSO), which presumably has a high impact on East Asian climate variability (Kim et al., 2017b; Woo et al., 2020), were not clearly found in this map. Instead, there are significant correlations over the Atlantic Ocean. Possibly, this is connected to the Rossby wave propagation from the tropical eastern Atlantic to the North Atlantic Ocean, further invoking the teleconnection between the Barents-Kara Sea and eastern Eurasia (Liu et al., 2022; Schubert et al., 2011).

Based on these contemporary correlation patterns, we develop a multiple linear regression model accounting for the year-to-year variation of PM_{10} with area-averaged indices of climate variables (predictors). This model is going to serve as an ideal model of maximum skill, which could be achieved by the climate-PM_{10} connection. Considering the forecast skill of CFSv2, the final form of MLR forecast model is decided. Its details are in Section 3.4.

3.3. Forecast skill of the dynamical climate model

In the above MLR model, if the chosen predictors are predictable by the dynamical climate model, then putting forecasted predictors into the regression model yields a PM_{10} forecast. This is a statistical-dynamical model. Therefore, the next step is to select those variables and regions in which the dynamical seasonal prediction model has good performance. To this end, we assessed the forecast skills of CFSv2 from the hindcast and forecast archives. Fig. 4 summarizes the CFSv2's skills for October forecast of key climate variables that potentially affect the PM concentrations. The key features are summarized as follows.

1) The temperature forecast skill is high in the tropical eastern Pacific, Indian Ocean, and Northwest Pacific (Fig. 4a). The skills of MSLP and GPH300 are significant in tropical regions, but become very low in most extratropics (Fig. 4b–c). Outside the tropic, there is a significant skill over the Barents-Kara Seas and North Pacific.
2) CFSv2 shows significant seasonal forecast skill of SST in almost all oceans (Fig. 4d).
3) There is almost no skill in South Korea for all variables.

According to previous studies (Jeong et al., 2021; Kumar et al., 2021), wintertime temperature over East Asia shows a significant negative correlation with seasonal PM concentration in association with the strength of East Asian winter monsoon circulation. Unfortunately, CFSv2 has very limited skills in South Korea and East Asia; temperature forecasts of CFSv2 show a marginal skill ($r = -0.04$) for the analysis period (Fig. 4e). Otherwise, it could be selected as a predictor for the hybrid model. Therefore, all selected variables are from outside East Asia.

3.4. Linear regression model and skill assessment

In the next step, the final form of the regression model was constructed by considering both the climate-PM_{10} relationships (Fig. 3) and the

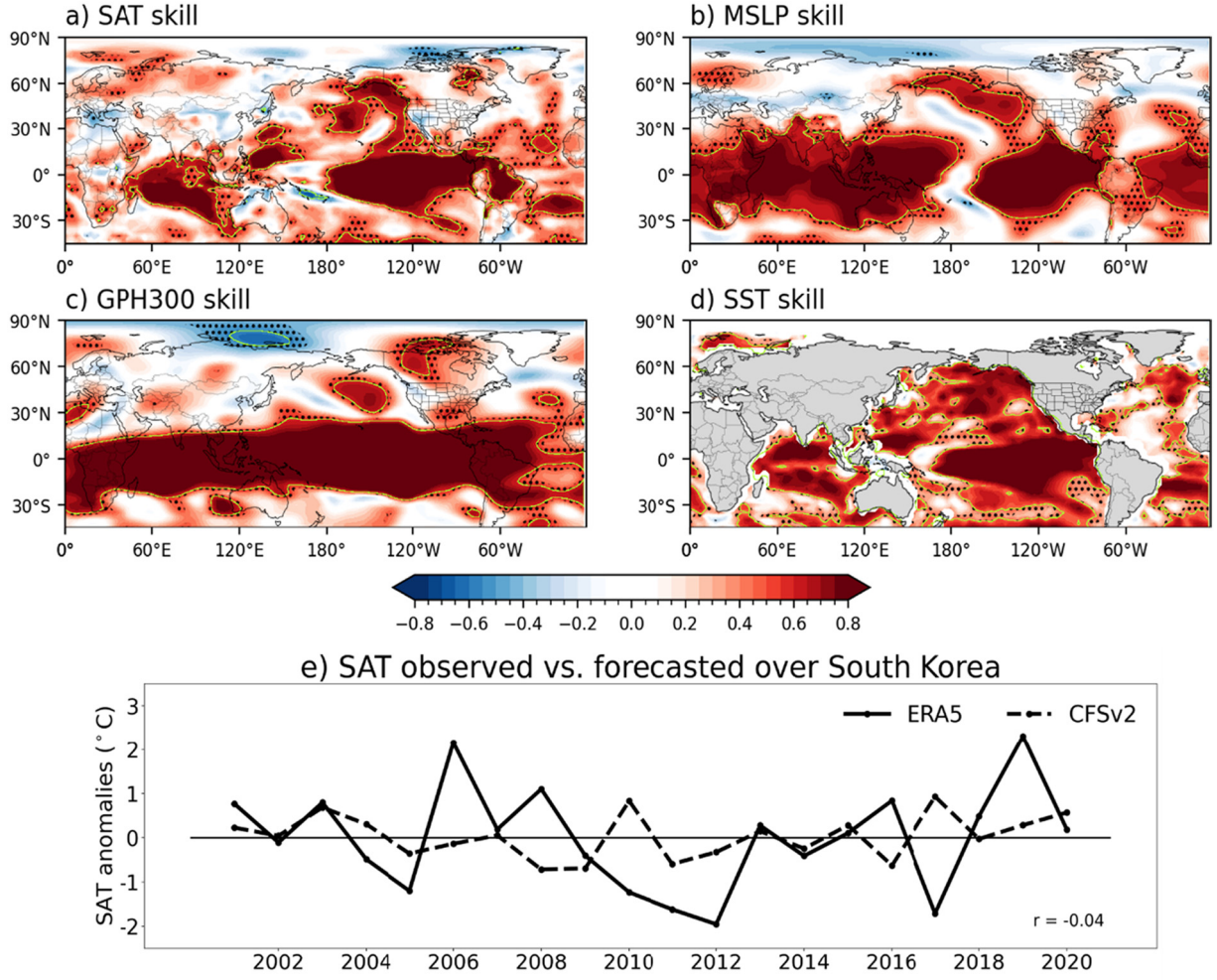


Fig. 4. CFSv2's skill (correlation coefficient between forecast and observation) for October forecast of winter a) SAT, b) MSLP, c) GPH at 300 hPa, and d) SST. Values statistically significant at 99 % (95 %) are outlined with a yellow-line (black-dotted pattern). e) Time-series of forecasted (CFSv2, dashed) and observed (ERA5, solid) SAT averaged over South Korea (34°–42°N, 124°–130°E).

corresponding forecast skill of CFSv2's (Fig. 4). Fig. 5 synthesizes the two factors altogether. Despite limited skills of CFSv2 over land and extratropics, we were able to find quite a few notable spots where both the ACCs and the skill are statistically significant for the objective selection of predictors. We applied three conditions for determining the predictors objectively: 1) both the ACC and CFSv2's skill need to be statistically significant, 2) the chosen predictors should be physically independent as much as possible, and 3) there should be some explainable physical relationship between the selected regional variables and the East Asian climate. We could identify three potential predictors: 1) MSLP over the South American continent near the South Atlantic (AT-MSLP1 in Fig. 5b), 2) GPH at 300 hPa over the Laptev Sea (LAP-GPH in Fig. 5c), and 3) SST over the Bering Sea (BE-SST1 in Fig. 5c). Several extratropical climate factors that were previously known to have strong influences on East Asian PM_{10}, like Siberian High and East Asian jet stream (Jeong and Park, 2017; Ku et al., 2021; Wei et al., 2021; Yun and Yoo, 2019), couldn't go through the screening of the CFSv2 skill and were not chosen for a predictor. Similarly, tropical and subtropical factors like ENSO and Western Pacific SST were not chosen despite high skills because there is no significant relationship with PM_{10} in South Korea. Note that CFSv2 has significant negative skill (i.e., forecasts opposite consistently) for the predictor LAP-GPH. Such significant negative skill is occasionally found in the dynamical seasonal climate forecast, which occurs when there is a potential skill, but the forecast model has misplaced the signal geographically. In most cases, this negative skill is ignored, but in case of multi-model ensemble seasonal forecasting, sometimes this negative skill is utilized by applying negative weight (DelSole et al., 2013; Wanders and Wada, 2015; Weigel et al., 2008) Therefore, with the idea of drawing out the maximum possible skill, we decided to use LAP-GPH for the forecast model.

Because the length of PM_{10} observations is not so long (20 winters), we choose two among the three above potential predictors for the regression model following the rules of thumb. LAP-GPH and BE-SST1 were finally chosen as predictors because it's the most statistically independent combination among the three potential predictors (see Table 1). They are almost independent in CFSv2 forecasts as well.

Table 1
Correlation coefficients between the observed PM_{10} concentration and climate variables for the potential predictor of MLR model of October forecast. Values in parentheses are for values from the CFSv2 hindcasts. Climate variables are denoted in Figs. 5 and 7. Values significant at 95 % confidence level are in bold.

	PM_{10}	LAP-GPH	AT-MSLP1	BE-SST1
PM_{10}	1	**−0.42**	**0.52**	**0.68**
LAP-GPH1		1	**−0.54** (−0.16)	−0.26 (−0.23)
AT-MSLP1			1	**0.56** (0.38)

Possible physical mechanisms liking the predictors and PM_{10} both for observation and forecast were examined (Fig. S1). LAP-GPH is associated with higher pressure over the Arctic, and warmer Arctic sea vs. colder eastern part of Eurasia pattern. This is similarly found in CFSv2 as well. This exhibits the anticorrelation between the Arctic Sea and eastern Eurasian continent, the Warm Arctic Cold Eurasia (Mori et al., 2014; Woo et al., 2020). BE-SST1 seemingly represents the tropical connection on East Asian winter monsoon — a positive relationship between eastern Pacific SST and East Asian temperature (Wang et al., 2000; Wang et al., 2010; Woo et al., 2020). It shows a positive SST pattern over the central Pacific, and its pattern resembles SST anomalies during the decaying phase of El-Nino. These physical linkages are found similarly in CFSv2 forecasts as well. BE-SST1 shows the correlation between the tropical Pacific and the Bering Sea created by atmospheric telecorrelation.

We first constructed a one-variable linear regression model with BE-SST1, then constructed an MLR using two variables, LAP-GPH and BE-SST1. The linear regression model with LAP-GPH and BE-SST1 as Eqs. 1–2 below.

$$y = a1 * BE_{SST1} + b1 \quad (1)$$

$$y = a2 * LAP_{GPH} + a3 * BE_{SST1} + b2 \quad (2)$$

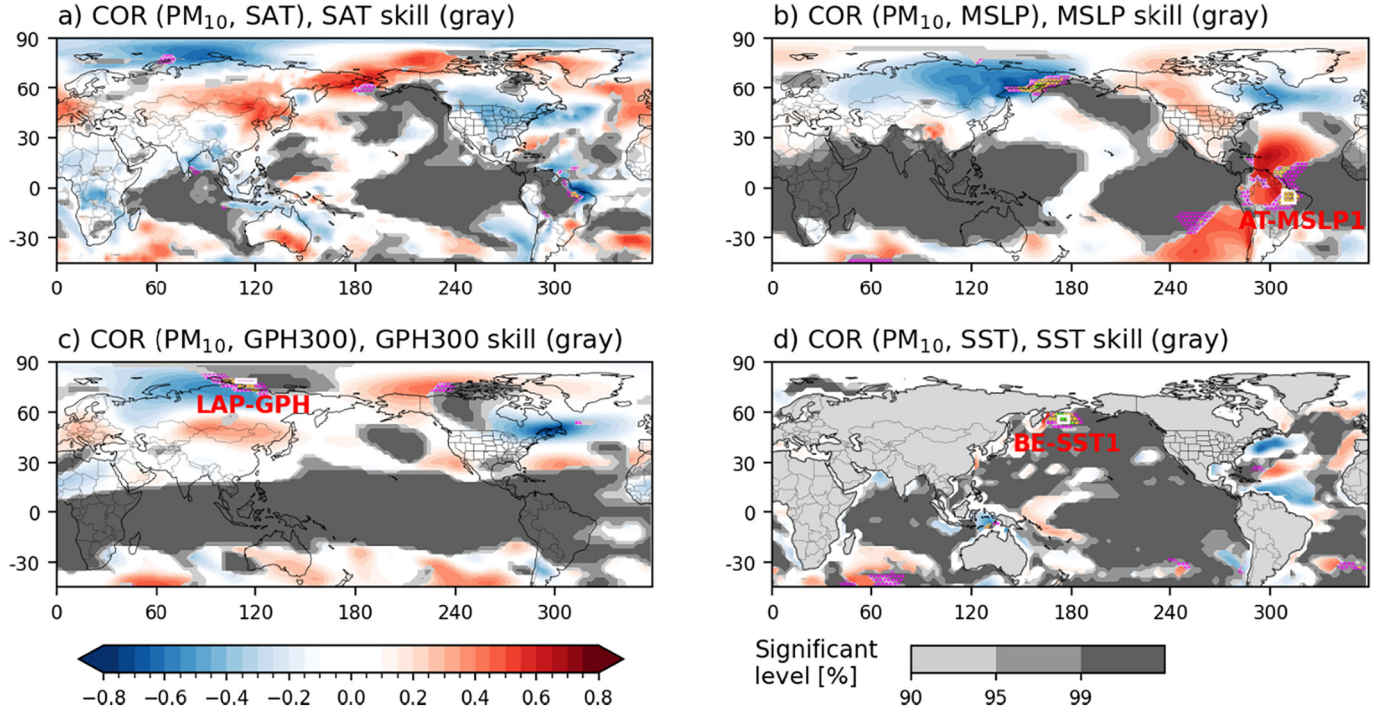


Fig. 5. Color shading represents the anomaly correlation coefficient (ACC) between the PC1 of winter PM_{10} and climate variables, same as Fig. 3. Gray shading highlights the region for which CFSv2's October forecast skill for corresponding climate variables are statistically significant (90, 95, and 99 % significance level). Dotted patterns (magenta, yellow, green) indicate the region where both the ACC and skill are statistically significant at 90, 95, and 99 %. The three potential climate predictors are indicated with rectangles with numbers.

When the data for the entire period were used, we got a1 = 0.05, b1 = 0.00, a2 = 0.06, a3 = 0.16, and b2 = 0.00. Note, in hindcast, the regression coefficients change slightly every year as the leaved-one-out cross-validation method (Hastie et al., 2009) was applied.

In the final step, applying every year's CFSv2 October forecast as predictors to the regression model, we produced PM_{10} forecasts for the period 2001/02–2020/21. Fig. 6 summarizes the results from 20 years of hindcasts. Here we compare observed PM_{10} (PC1, gray bars) and forecasted (red line). Here we compare the perfect model forecast for which the observed (i.e., perfect) LAP-GPH and BE-SST1 were used as the input of the regression model. This can be considered the best forecast achievable from the climate-PM_{10} relationship. Observed PM_{10} includes non-climatic factors, which are unpredictable in our model, so we used the perfect model forecast as a benchmark. The correlation coefficient between the perfect model forecast and observed PM_{10} is 0.62.

Despite the long lead-time, the statistical-dynamical model reasonably well captures the perfect model. For the regression model using one variable, BE-SST1, the correlation coefficient (r) between the perfect model and hybrid model is 0.49, which is statistically significant. If we include the linear trend for the period, which is not realistic for the real forecast though, then the correlation becomes 0.72. The MLR model using the two predictors, LAP-GPH and BE-SST1, has slightly lower skill (r = 0.40), but its significant. This suggests our method can provide a skillful seasonal PM_{10} forecast if the non-climatic factors can be realistically considered altogether.

In the same manner, we construct a statistical-dynamical model based on November (2nd, 7th, and 12th)'s forecast of CFSv2 and calculate hindcasts for the same period (Table 2). Fig. 7 summarizes the climate-PM_{10} relationship and CFSv2 prediction skills for November forecasts. Here we were able to identify three potential predictors: 1) temperature

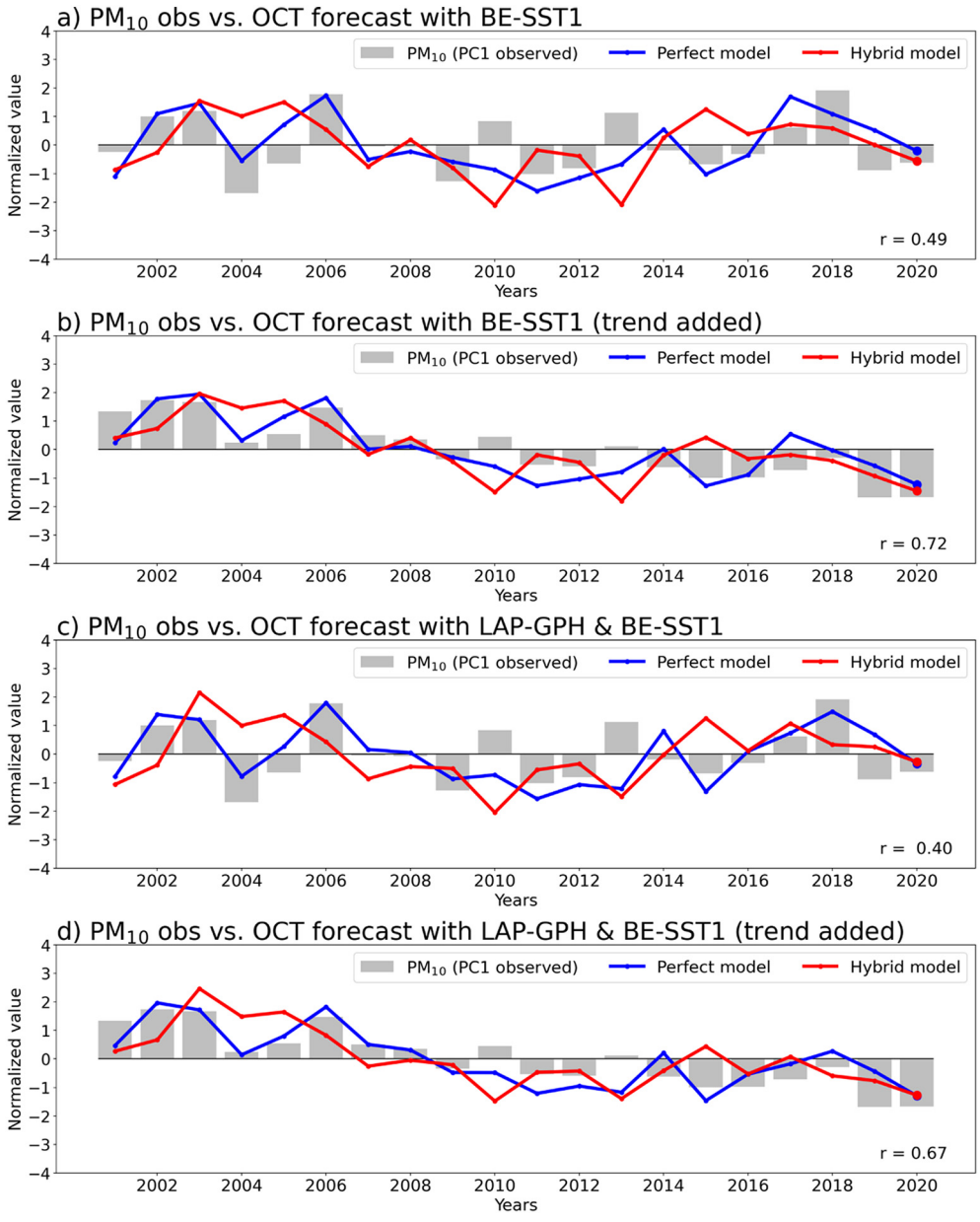


Fig. 6. Time-series of forecasted and observed winter PM_{10} concentrations in Korea for the October forecasts. Forecasts from the regression model with BE-SST1 (a–b) and the MLR model with LAP-GPH and BE-SST1 (c–d). Results for the detrended model (a, c) and with (b, d) a linear trend for the entire period added. The gray bars indicate the PC1 of observed PM_{10}. The blue line indicates the perfect model result, where the observed values were used for the input of the MLR model. The red line indicates the forecasted values from the statistical-dynamical model with the CFSv2's October forecast. Note all the forecasts were done with leave-one-out cross-validation.

Table 2
Same as Table 1 but for the November forecast.

	PM_{10}	KA-SAT	AT-MSLP2	BE-SST2
PM_{10}	1	**−0.51**	**0.61**	**0.69**
KA-SAT		1	−0.40 (−0.28)	−0.34 (−0.24)
AT-MSLP2			1	**0.55** (0.08)

over the Kara Sea (KA-SAT), 2) MSLP over the South American continent near the South Atlantic (AT-MSLP2, slightly different location compared to AT-MSLP1), and 3) SST over the Bering Sea (BE-SST2, slightly different location compared to BE-SST2) (Fig. 8). Through the same approach for the October forecast, BE-SST2 and KA-SAT, showing insignificant mutual correlation were chosen as the best predictors for the MLR model. The newly identified potential predictor, AT-MSLP2 may represent the relationship between the PM_{10} and tropical variability over the Western Pacific. It correlates positively with the pressure anomalies over southern China, the Western Pacific, the Indian Ocean, and the South American continent. However, it was excluded as predictors for the MLR model because correlates significantly with BE-SST2. As winter approaches, LAP-GPH in the Arctic was dropped from the potential predictor due to its low skill in CFSv2.

Similar to October forecast case, we first constructed a one-variable linear regression model with BE-SST2, which has highest correlation with observed PC1, then constructed an MLR using two variables, BE-SST2 and KA-SAT. The linear regression model with BE-SST2 and KA-SAT as Eqs. 3–4 below.

$$y = a1 * BE_{SST2} + b1 \quad (3)$$

$$y = a2 * BE_{SST2} + a3 * KA_{Temp} + b2 \quad (4)$$

When the data for the entire period were used, we got a1 = 0.32, b1 = 0.00, a2 = 0.11, a3 = −0.19, and b2 = 0.00. With a shorter (about 15 days) lead though, the model shows slightly lower skill compared to the October forecast, but still statistically significant. For the regression model using BE-SST2, the correlation coefficient between the perfect model and hybrid model is 0.44 and becomes 0.62 along with the linear trend. The MLR model using the two predictors, BE-SST2 and KA-SAT, has a comparable skill of 0.45, and 0.71 along with the linear trend.

In this study, considering the relatively short observation period (20 years) and limited skill of CFSv2, the model was trained using data from the entire period excluding the forecast year. Even though a substantial skill is achieved by applying such a leave-one-out cross-validation, it is highly possible that the performance of the model can change with time. To examine this aspect even a little, we established a regression model by setting the training period to 2001–2015, and conducted forecasts with the model for the test period 2016–2020. The results for the October forecast from this test are shown in Supplementary Fig. S2. The models still perform well in this case and exhibit very high skill in the testing period although the period is only five years.

4. Summary and discussion

Here we propose a statistical-dynamical model for seasonal forecasting of wintertime PM concentration over South Korea. By combining the observed statistical relationship between winter PM and climatic variables, and seasonal climate prediction products from NCEP CFSv2, this model produces a skillful seasonal forecast of wintertime PM_{10} concentration.

A chemistry-climate coupled model with realistic initialization and emission scenario will be the most desirable method providing a seasonal forecast of air quality as well as climate predictions. However, it's still far beyond our current modeling technology and observational capability. By virtue of tremendous advances in climate modeling techniques and observations, skillful seasonal forecasting from the dynamical climate model became feasible to some extent in recent decades (Alizadeh, 2022; Jeong et al., 2017). However, coupling complex chemistry to climate models and assimilations techniques is still in its infancy (Bocquet et al., 2015). The accurate estimation of emission for the chemistry-transport model is not sufficient yet. Until the performance of such a complete coupled model is sufficient, this intermediate-level model in the present study can be a good alternative. Although the hybrid model predicts PM variability that can be explained by only climate variability, it's relatively easy to implement only if a dynamical forecast model and its hindcast are available. As the dynamical model's skill improves, which is often the case with updates in model physics and dynamics and observations, the hybrid

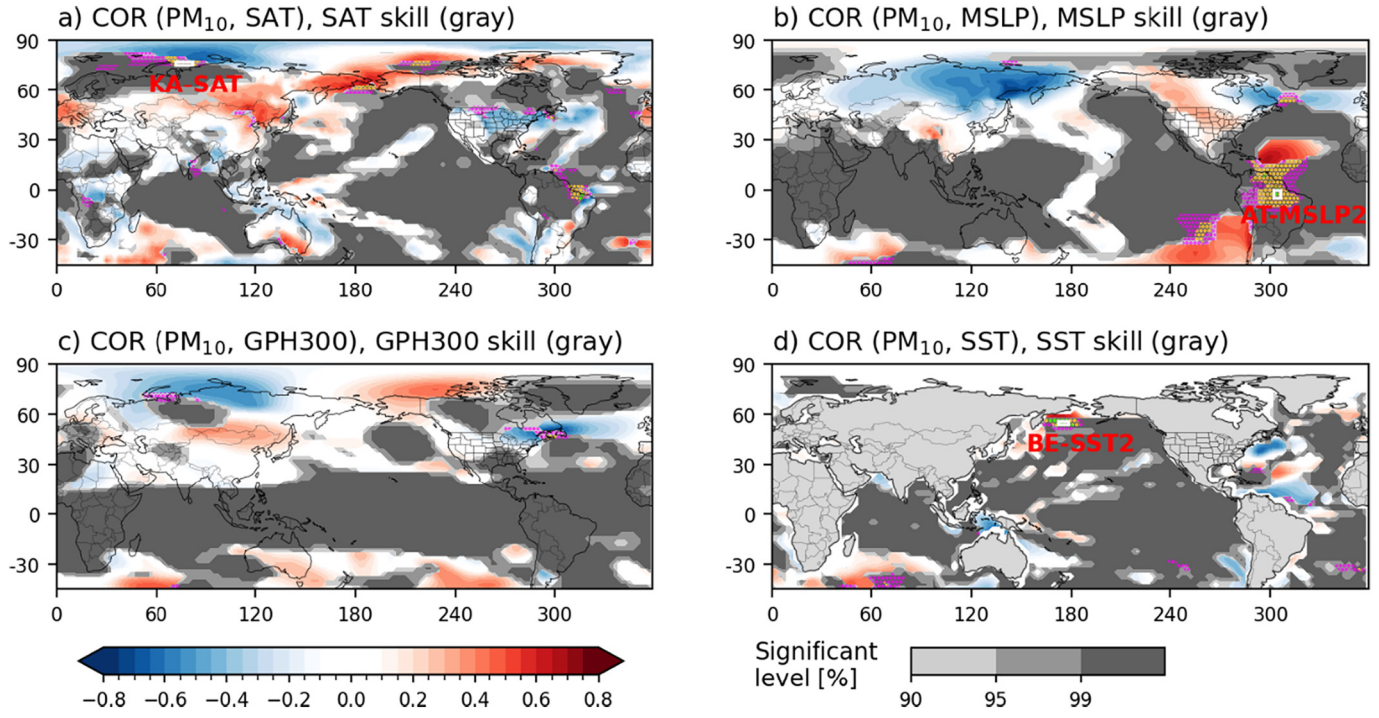


Fig. 7. Same as Fig. 5, but gray shadings are for the November forecasts/hindcast of CFSv2.

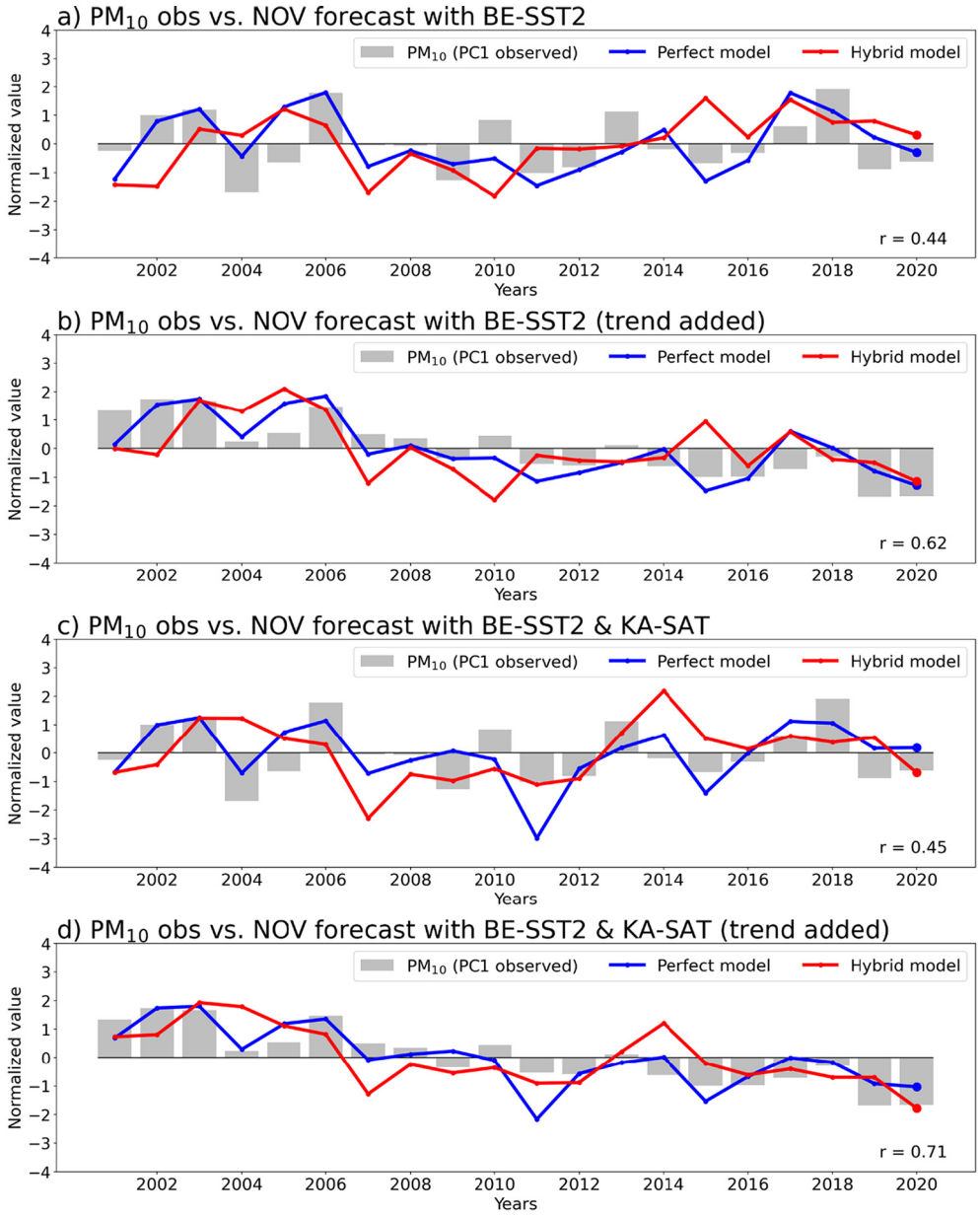


Fig. 8. Same as Fig. 6 but for the November forecast.

models' skill likely improves. Applying multi-model ensembles techniques may provide additional skill for this method.

CRediT authorship contribution statement

Jee-Hoon Jeong: Conceptualization, Investigation, Visualization, Writing – review & editing, Supervision. **Jahyun Choi:** Investigation, Methodology, Data curation, Visualization, Software, Writing – original draft. **Ji-Yoon Jeong:** Methodology, Data curation, Visualization, Software. **Sung-Ho Woo:** Methodology, Data curation, Visualization, Writing – review & editing. **Sang-Woo Kim:** Conceptualization, Writing – review & editing. **Daegyun Lee:** Data curation, Writing – review & editing. **Jae-Bum Lee:** Data curation, Writing – review & editing. **Jin-Ho Yoon:** Conceptualization, Investigation, Writing – review & editing, Supervision.

Data availability

Data will be made available on request.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by National Institute of Environment Research (NIER) funded by the Ministry of Environment (MOE) of the Republic of Korea (NIER-2019-01-02-048) and by Korea Environment Industry & Technology Institute (KEITI) through "Climate Change R&D Project for New Climate Regime.", funded by Korea Ministry of Environment (MOE) (RE202201655). Sung-Ho Woo was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2022R1I1A1A01063153). Jee-Hoon Jeong was supported by the NRF grant funded by the Korea government (MSIT) (NRF-2021R1A2C1011178).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2022.157699.

References

Aizpurua-Etxezarreta, M., Carreno-Madinabeitia, S., Ulazia, A., Sáenz, J., Saenz-Aguirre, A., 2022. Long-Term Freezing Temperatures Frequency Change Effect on Wind Energy Gain (Eurasia and North America, 1950– 2019). Sustainability 14, 5630. https://www.mdpi.com/2071-1050/14/9/5630.

Alizadeh, O., 2022. Advances and challenges in climate modeling. Clim. Chang. 170, 1–26. https://doi.org/10.1007/s10584-021-03298-4.

Baker, J.C., Castilho de Souza, D., Kubota, P.Y., Buermann, W., Coelho, C.A., Andrews, M.B., Gloor, M., Garcia-Carreras, L., Figueroa, S.N., Spracklen, D.V., 2021. An assessment of Land–atmosphere interactions over South America using satellites, reanalysis, and two global climate models. J. Hydrometeorol. 22, 905–922. https://doi.org/10.1175/JHM-D-20-0132.1.

Bocquet, M., Elbern, H., Eskes, H., Hirtl, M., Žabkar, R., Carmichael, G., Flemming, J., Inness, A., Pagowski, M., Pérez, Camaño J., 2015. Data assimilation in atmospheric chemistry models: current status and future prospects for coupled chemistry meteorology models. Atmos. Chem. Phys. 15, 5325–5358. https://doi.org/10.5194/acp-15-5325-2015.

Camalier, L., Cox, W., Dolwick, P., 2007. The effects of meteorology on ozone in urban areas and their use in assessing ozone trends. Atmos. Environ. 41, 7127–7137. https://doi.org/10.1016/j.atmosenv.2007.04.061.

Chattopadhyay, S., Gupta, S., Saha, R.N., 2010. Spatial and temporal variation of urban air quality: a GIS approach. J. Environ. Prot. 1, 264. https://doi.org/10.4236/jep.2010.13032.

Chen, L., Aalto, J., Luoto, M., 2021. Observed decrease in soil and atmosphere temperature coupling in recent decades over northern Eurasia. Geophys. Res. Lett. 48, e2021GL092500. https://doi.org/10.1029/2021GL092500.

Clean Air Alliance of China, 2013. State Council Air Pollution Prevention and Control Action Plan, Issue II. http://en.cleanairchina.org/product/6346.html.

Cohen, J., Jones, J., Furtado, J.C., Tziperman, E.L.I., 2013. Warm arctic, cold continents a common pattern related to arctic sea ice melt, snow advance, and extreme winter weather. Oceanography 26, 150–160. http://www.jstor.org/stable/24862104.

DelSole, T., Yang, X., Tippett, M.K., 2013. Is unequal weighting significantly better than equal weighting for multi-model forecasting? Q. J. R. Meteorol. Soc. 139, 176–183. https://doi.org/10.1002/qj.1961.

Ding, Z., Wu, R., 2020. Quantifying the internal variability in multi-decadal trends of spring surface air temperature over mid-to-high latitudes of Eurasia. Clim. Dyn. 55, 2013–2030. https://doi.org/10.1007/s00382-020-05365-5.

Gao, T., Xuebin, Z., Wulan, 2010. A seasonal forecast scheme for spring dust storm predictions in Northern China. Meteorological Applications 17, 433–441. https://doi.org/10.1002/met.175.

Gillies, R.R., Wang, S.-Y., Yoon, J.-H., Weaver, S., 2010. CFS Prediction of winter persistent inversions in the intermountain region. Weather Forecast. 25, 1211–1218. https://doi.org/10.1175/2010WAF2222419.1.

Hastie, T., Tibshirani, R., Friedman, J.H., 2009. The elements of statistical learning: data mining, inference, and prediction. Springer 2, 241–243. https://doi.org/10.1007/978-0-387-21606-5.

Heo, J., Park, J.-S., Kim, B.M., Kim, S.-W., Park, R.J., Jeon, H., Yoon, S.-C., 2017. Two notable features in PM10 data and analysis of their causes. Air Qual. Atmos. Health 10, 991–998. https://doi.org/10.1007/s11869-017-0488-6.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R.J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., Thépaut, J.-N., 2020. The ERA5 global reanalysis. Q. J. R. Meteorol. Soc. 146, 1999–2049. https://doi.org/10.1002/qj.3803.

Honda, M., Inoue, J., Yamane, S., 2009. Influence of low Arctic sea-ice minima on anomalously cold Eurasian winters. Geophys. Res. Lett. 36. https://doi.org/10.1029/2008GL037079.

Huang, B., Thorne, P.W., Banzon, V.F., Boyer, T., Chepurin, G., Lawrimore, J.H., Menne, M.J., Smith, T.M., Vose, R.S., Zhang, H.-M., 2017. Extended reconstructed sea surface temperature, version5 (ERSSTv5). J. Clim. 30, 8179–8205. https://doi.org/10.1175/JCLI-D-16-0836.1.

Jacob, D.J., Winner, D.A., 2009. Effect of climate change on air quality. Atmos. Environ. 43, 51–63. https://doi.org/10.1016/j.atmosenv.2008.09.051.

Jeong, J.-H., Lee, H., Yoo, J.H., Kwon, M., Yeh, S.-W., Kug, J.-S., Lee, J.-Y., Kim, B.-M., Son, S.-W., Min, S.-K., 2017. The status and prospect of seasonal climate prediction of climate over Korea and East Asia: a review. Asia-Pac. J. Atmos. Sci. 53, 149–173. https://doi.org/10.1007/s13143-017-0008-5.

Jeong, J.I., Park, R.J., 2017. Winter monsoon variability and its impact on aerosol concentrations in East Asia. Environ. Pollut. 221, 285–292.

Jeong, J.I., Park, R.J., Yeh, S.-W., Roh, J.-W., 2021. Statistical predictability of wintertime PM2.5 concentrations over East Asia using simple linear regression. Sci. Total Environ. 776, 146059. https://doi.org/10.1016/j.scitotenv.2021.146059.

Jiang, Q., Li, W., Fan, Z., He, X., Sun, W., Chen, S., Wen, J., Gao, J., Wang, J., 2021. Evaluation of the ERA5 reanalysis precipitation dataset over Chinese Mainland. J. Hydrol. 595, 125660. https://doi.org/10.1016/j.jhydrol.2020.125660.

Judt, F., 2020. Atmospheric predictability of the tropics, middle latitudes, and polar regions explored through global storm-resolving simulations. J. Atmos. Sci. 77, 257–276. https://doi.org/10.1175/JAS-D-19-0116.1.

Kim, H.C., Kim, S., Kim, B.-U., Jin, C.-S., Hong, S., Park, R., Son, S.-W., Bae, C., Bae, M., Song, C.-K., Stein, A., 2017a. Recent increase of surface particulate matter concentrations in the Seoul Metropolitan AreaKorea. Scientific Reports 7, 4710. https://doi.org/10.1038/s41598-017-05092-8.

Kim, S.-H., Sung, H.-J., Kim, S.-J., Baek, E.-H., Moon, J.-Y., Kim, B.-M., 2022. Contribution of Ural and Kamchatka blockings to the amplified warm arctic-cold Eurasia pattern under arctic sea ice loss and Eurasian cooling. J. Clim. 35, 4071–4083. https://doi.org/10.1175/JCLI-D-21-0635.1.

Kim, S., Son, H.-Y., Kug, J.-S., 2017b. How well do climate models simulate atmospheric teleconnections over the North Pacific and East Asia associated with ENSO? Clim. Dyn. 48, 971–985. https://doi.org/10.1007/s00382-016-3121-8.

Klingler, C., Schulz, K., Herrnegger, M., 2021. LamaH-CE: LArge-SaMple DAta for hydrology and environmental sciences for central Europe. Earth Syst. Sci. Data 13, 4529–4565. https://doi.org/10.5194/essd-13-4529-2021.

Koo, Y.-S., Yun, H.-Y., Kwon, H.-Y., Yu, S.-H., 2010. A Development of PM10 forecasting system. J. Korea. Soc. Atmos. Environ. 26, 666–682. https://doi.org/10.5572/KOSAE.2010.26.6.666.

Krishnamurthy, V., 2019. Predictability of weather and climate. Earth Space Sci. 6, 1043–1056.

Ku, H.-Y., Noh, N., Jeong, J.-H., Koo, J.-H., Choi, W., Kim, B.-M., Lee, D., 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. https://doi.org/10.1016/j.atmosenv.2021.118632.

Kug, J.-S., Jeong, J.-H., Jang, Y.-S., Kim, B.-M., Folland, C.K., Min, S.-K., Son, S.-W., 2015. Two distinct influences of Arctic warming on cold winters over North America and East Asia. Nat. Geosci. 8, 759–762. https://doi.org/10.1038/ngeo2517.

Kumar, N., Park, R.J., Jeong, J.I., Woo, J.-H., Kim, Y., Johnson, J., Yarwood, G., Kang, S., Chun, S., Knipping, E., 2021. Contributions of international sources to PM2.5 in South Korea. Atmos. Environ. 261, 118542. https://doi.org/10.1016/j.atmosenv.2021.118542.

Lee, H.-J., Kim, J.E., Cha, J.W., Song, S., Ryoo, S.-B., Kim, Y.P., 2019. Characteristics of long-lasting haze episodes observed in Seoul, South Korea, for 2009–2014. Theor. Appl. Climatol. 136, 55–64. https://doi.org/10.1007/s00704-018-2415-7.

Lee, J., Kim, K.-Y., 2018. Analysis of source regions and meteorological factors for the variability of spring PM10 concentrations in Seoul, Korea. Atmos. Environ. 175, 199–209. https://doi.org/10.1016/j.atmosenv.2017.12.013.

Liu, J., Huang, W., Zhang, Q., 2022. The quasi-biweekly oscillation of eastern China PM2. 5 in response to different Rossby wave trains over the Eurasian continent. Atmos. Res. 267, 105990. https://doi.org/10.1016/j.atmosres.2021.105990.

Lorenz, E.N., 1956. Empirical Orthogonal Functions and Statistical Weather Prediction. 1. Massachusetts Institute of Technology, Department of Meteorology Cambridge.

Ministry of Environment, 2016. Clean Air Conservation Act. https://elaw.klri.re.kr/eng_service/lawView.do?lang=ENG&hseq=41386.

Morcrette, J.J., Boucher, O., Jones, L., Salmond, D., Bechtold, P., Beljaars, A., Benedetti, A., Bonet, A., Kaiser, J., Razinger, M., 2009. Aerosol analysis and forecast in the European Centre for medium-range weather forecasts integrated forecast system: Forward modeling. J. Geophys. Res.-Atmos. 114. https://doi.org/10.1029/2008JD011235.

Mori, M., Watanabe, M., Shiogama, H., Inoue, J., Kimoto, M., 2014. Robust Arctic sea-ice influence on the frequent Eurasian cold winters in past decades. Nat. Geosci. 7, 869–873. https://doi.org/10.1038/ngeo2277.

Park, D.-H., Kim, S.-W., Kim, M.-H., Yeo, H., Park, S.S., Nishizawa, T., Shimizu, A., Kim, C.-H., 2021. Impacts of local versus long-range transported aerosols on PM10 concentrations in Seoul, Korea: An estimate based on 11-year PM10 and lidar observations. Sci. Total Environ. 750, 141739. https://doi.org/10.1016/j.scitotenv.2020.141739.

Preisendorfer, R., 1988. Principal Component Analysis in Meteorology and Oceanography. 17. Elsevier Sci. Publ, p. 425.

Saha, S., Moorthi, S., Pan, H.-L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D., 2010. The NCEP climate forecast system reanalysis. Bull. Am. Meteorol. Soc. 91, 1015–1058. https://doi.org/10.1175/2010BAMS3001.1.

Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., Behringer, D., Hou, Y.-T., Chuang, H.-Y., Iredell, M., 2014. The NCEP climate forecast system version 2. J. Clim. 27, 2185–2208. https://doi.org/10.1175/JCLI-D-12-00823.1.

Schubert, S., Wang, H., Suarez, M., 2011. Warm season subseasonal variability and climate extremes in the Northern Hemisphere: The role of stationary Rossby waves. J. Clim. 24, 4773–4792. https://doi.org/10.1175/JCLI-D-10-05035.1.

Shen, L., Mickley, L.J., 2017. Seasonal prediction of US summertime ozone using statistical analysis of large scale climate patterns. Proc. Natl. Acad. Sci. 114, 2491–2496. https://doi.org/10.1073/pnas.1610708114.

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

Takaya, K., Nakamura, H., 2005. Geographical dependence of upper-level blocking formation associated with intraseasonal amplification of the Siberian high. J. Atmos. Sci. 62, 4441–4449. https://doi.org/10.1175/JAS3628.1.

Tarek, M., Brissette, F.P., Arsenault, R., 2020. Evaluation of the ERA5 reanalysis as a potential reference dataset for hydrological modelling over North America. Hydrol. Earth Syst. Sci. 24, 2527–2544. https://doi.org/10.5194/hess-24-2527-2020.

Taylor, K.E., Stouffer, R.J., Meehl, G.A., 2012. An overview of CMIP5 and the experiment design. Bull. Am. Meteorol. Soc. 93, 485–498. https://doi.org/10.1175/BAMS-D-11-00094.1.

Wanders, N., Wada, Y., 2015. Decadal predictability of river discharge with climate oscillations over the 20th and early 21st century. Geophys. Res. Lett. 42, 10689–10695. https://doi.org/10.1002/2015GL066929.

Wang, B., Wu, R., Fu, X., 2000. Pacific-East Asian teleconnection: how does ENSO affect East Asian climate? J. Clim. 13, 1517–1536. https://doi.org/10.1175/1520-0442(2000)013%3C1517:PEATHD%3E2.0.CO;2.

Wang, B., Wu, Z., Chang, C.-P., Liu, J., Li, J., Zhou, T., 2010. Another look at interannual-to-interdecadal variations of the East Asian winter monsoon: the northern and southern temperature modes. J. Clim. 23, 1495–1512. https://doi.org/10.1175/2009JCLI3243.1.

Wei, W., Zhuang, B., Lin, H., Shu, Y., Wang, T., Chen, H., Gao, Y., 2021. Analysis of the Interactions between the 200 hPa Jet and Air Pollutants in the Near-Surface Layer over East Asia in Summer. Atmosphere 12, 886. https://www.mdpi.com/2073-4433/12/7/886.

Weigel, A.P., Liniger, M.A., Appenzeller, C., 2008. Can multi-model combination really enhance the prediction skill of probabilistic ensemble forecasts? Q. J. R. Meteorol. Soc. 134, 241–260. https://doi.org/10.1002/qj.210.

Woo, S.-H., Choi, J., Jeong, J.-H., 2020. Modulation of ENSO teleconnection on the relationship between arctic oscillation and wintertime temperature variation in South Korea. Atmosphere 11, 950. https://doi.org/10.3390/atmos11090950.

Yang, Y., Liao, H., Lou, S., 2016. Increase in winter haze over eastern China in recent decades: Roles of variations in meteorological parameters and anthropogenic emissions. J. Geophys. Res.-Atmos. 121, 13050–13065. https://doi.org/10.1002/2016JD025136.

Yin, Z., Wang, H., 2017. Statistical prediction of winter haze days in the North China plain using the generalized additive model. J. Appl. Meteorol. Climatol. 56, 2411–2419. https://doi.org/10.1175/JAMC-D-17-0013.1.

You, T., Wu, R., Huang, G., 2018. Differences in meteorological conditions between days with persistent and non-persistent pollution in Beijing, China. J. Meteorol. Res. 32, 81–98. https://doi.org/10.1007/s13351-018-7086-x.

Yu, S., Tett, S.F.B., Freychet, N., Yan, Z., 2021. Changes in regional wet heatwave in Eurasia during summer (1979–2017). Environ. Res. Lett. 16, 064094. https://doi.org/10.1088/1748-9326/ac0745.

Yun, S.-G., Yoo, C., 2019. The Effects of Spring and Winter Blocking on PM10 Concentration in Korea. Atmosphere 10, 410. https://www.mdpi.com/2073-4433/10/7/410.