

머신 러닝 기법을 활용한 극지/고위도 기후와  
우리나라 미세먼지의 연관성 분석 및 예측

Analysis and Prediction of Air Quality in South Korea  
related to Arctic/high-latitude Climate based on Machine  
Learning

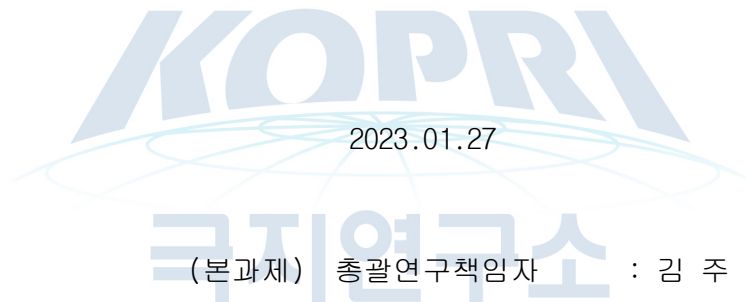


서울대학교

# 제 출 문

극지연구소장 귀하

본 보고서를 “북극 해양·해빙 변화에 기인한 북극과 한반도의 재해기상 현상 모델링 시스템(KPOPS-Earth)의 개발 및 활용” 과제의 위탁연구 “머신 러닝 기법을 활용한 극지/고위도 기후와 우리나라 미세먼지의 연관성 분석 및 예측” 과제의 최종보고서로 제출합니다.



2023.01.27

(본과제) 총괄연구책임자 : 김 주 흥

위탁연구기관명 : 서울대학교

위탁연구책임자 : 허 창 회

위탁참여연구원 : 허 진 우

“ : 현 동 규

“ : 박 인 규

“ : 김 가 영

## 보고서 초록

위탁연구과제명	머신 러닝 기법을 활용한 극지/고위도 기후와 우리나라 미세먼지의 연관성 분석 및 예측				
위탁연구책임자	허창희	해당단계 참여연구원수	4	해당단계 연구비	50,000천원
연구기관명 및 소속부서명	서울대학교 지구환경과학부		참여기업명		
국제공동연구	상대국명 :		상대국연구기관명 :		
요약(연구결과를 중심으로 개조식 500자 이내)				보고서 면수	50
<p>1. 미세먼지 농도 예측 모델 개발에 필요한 잠재적 입력자료를 데이터베이스로 구축</p> <ul style="list-style-type: none"> <li>- 한국과 중국 관측소에서 측정된 미세먼지 농도 자료를 자동으로 수집하여 데이터베이스로 구축</li> <li>- ERA5 재분석 자료와 KPOPS-Earth 기후모델로 생산한 과거 모의 자료를 데이터베이스로 구축</li> </ul> <p>2. 서울의 미세먼지 농도와 예측에 사용되는 잠재적 입력자료 간의 중장기 시간규모에서 상관관계 분석</p> <ul style="list-style-type: none"> <li>- 월 시간 규모에서 서울의 미세먼지 농도와 상관성이 높은 기상변수가 분포하는 지역이 한반도 주변 뿐만 아니라 북극과 고위도 지역까지 광범위하게 분포</li> <li>- 같은 변수라도 겨울철 월마다 서울의 미세먼지 농도와 높은 상관성이 나타나는 지역이 달라짐</li> <li>- 특정 지역의 기상변수 평균값을 입력자료로 사용하기보다 넓은 지역의 공간 분포를 학습할 수 있는 머신러닝 기법인 합성곱신경망(CNN)을 예측 모델에 사용하는 것이 합당</li> </ul> <p>3. CNN 알고리즘에 기반한 서울의 월평균 PM<sub>2.5</sub> 농도 예측 모델의 프로토타입을 개발</p> <ul style="list-style-type: none"> <li>- 2011년부터 2020년까지 겨울철(1, 2, 11, 12월)동안 CNN 알고리즘에 기반한 서울의 월평균 PM<sub>2.5</sub> 농도를 예측하는 모델을 개발</li> <li>- 학습에서는 ERA5 재분석 자료만 이용하며, 예측에서는 ERA5 자료를 이용하는 모델(CNN-ERA5)과 KPOPS-Earth 자료를 이용하는 모델(CNN-KPOPS)의 프로토타입을 개발</li> <li>- 입력자료에 대입되는 기상변수를 인위적으로 조정(각 변수 그리드 값에 PM<sub>2.5</sub> 농도와 상관계수값의 세제곱을 곱해서)해서 CNN-ERA5, CNN-KPOPS의 예측 성능을 향상시킬 수 있음</li> <li>- 관측값을 이용하는 CNN-ERA5 모델의 RMSE, 상관계수, 정확도, 탐지율, 오보율이 각각 3.56 <math>\mu\text{g m}^{-3}</math>, 0.73, 80%, 67%, 20%이고, 예측값을 이용하는 CNN-KPOPS 모델은 각각 4.44 <math>\mu\text{g m}^{-3}</math>, 0.46, 63%, 42%, 58%. CNN-KPOPS 모델의 예측 성능이 괜찮은 편</li> <li>- CNN-KPOPS 모델을 현업 예보에 활용할 수 있도록 설계하였으며, 추가 연구를 통해서 예측 성능을 높일 수 있을 것임</li> </ul>					
색 인 어 (각 5개 이상)	한 글	머신러닝, 합성곱신경망, PM <sub>2.5</sub> , 중장기 예측, 북극, 고위도			
	영 어	machine learning, convolutional neural network, PM <sub>2.5</sub> , mid-term prediction, long-term prediction, Arctic region, high-latitude region			

# 요 약 문

## I. 제 목

머신러닝 기법을 활용한 극지/고위도 기후와 우리나라 미세먼지의 연관성 분석 및 예측

## II. 연구개발의 목적 및 필요성

- 대기 중 미세먼지 농도가 높아지면 우리 몸에 해로울 뿐 아니라, 식생의 호흡 방해, 시정의 악화, 기계 결함의 원인이 되는 등 사회, 경제적으로 피해가 발생한다. 미세먼지 농도를 줄이기 위해 우리나라를 비롯한 여러 국가에서 대기오염물질 배출량 감축 정책을 펼치고 있다. 미세먼지 농도의 변동은 대기 상황의 변화, 중관 기상패턴과 밀접하게 관련되어 있다. 특히 북극, 고위도 지역의 기후 인자가 우리나라 미세먼지 농도에 영향을 끼칠 수 있다.
- 미세먼지 피해를 줄이기 위한 장기적 대응 및 저감정책 수립에 대한 국민의 요구가 커지고 있다. 이를 위해서 북극, 고위도 기후 인자의 영향까지 고려해서 한 달 이상의 장기간 시간규모에서 미세먼지 농도를 예측할 필요가 있다.
- 현재까지 여러 연구가 지상 관측 자료, 수치/기후모델과 대기화학 모델의 예측결과를 입력자료로 사용하여 우리나라의 권역별 미세먼지 농도를 예측하였다. 그러나 이들 연구는 대부분 수일 이내의 단기 예측을 목적으로 하였다. 이 연구에서는 극지, 고위도 지역의 대규모 대기순환장과 우리나라 미세먼지 농도 간의 연관성을 고려해서 한 달 혹은 계절 시간 규모에서 우리나라 미세먼지를 예측하고자 한다.
- 최근 다양한 종류의 머신러닝 기법을 이용해서 미세먼지 농도 예측 모델을 개발하려는 연구가 활발히 이루어지고 있다. 머신러닝 기법은 대기의 비선형적인 부분까지 학습할 수 있는 장점을 갖고 있다. 그런데 현재 개발된 머신러닝 기반 미세먼지 농도 예측 모델은 대부분 북극, 고위도 기후 인자의 영향을 고려하지 않으며, 수 일 이내의 단기 예측에 초점을 맞추고 있다.
- 미세먼지 피해를 최소화하기 위한 장기적 대응 및 저감 정책 수립을 위해 머신러닝 기법을 이용한 북극 및 고위도 지역의 기후 인자를 이용해서 중장기 미세먼지 농도 예측 모델을 개발할 필요가 있다.

## III. 연구개발의 내용 및 범위

- 관측/재분석 자료를 이용해서 중장기 시간 규모에서 우리나라 미세먼지 농도에 영향을 끼치는 북극, 고위도 지역의 기후 인자를 분석하고 이를 데이터베이스로 구축한다.
- 기후 모델링 시스템이 모의한 과거 자료에서 한반도 미세먼지 농도에 영향을 끼치는 북극, 고위도 기후 인자를 분석하고 이를 데이터베이스에 추가한다.

- 머신러닝 기법을 이용해 중장기(월-계절) 시간 규모의 한반도 미세먼지 농도 예측 모델의 프로토타입을 구축한다.

#### IV. 연구개발결과

- 관측/재분석 자료를 분석해서 한반도 미세먼지 농도에 영향을 끼치는 고위도 기후 인자를 분석
  - 환경부 산하 AirKorea에서 발간하는 확정 자료를 수집해서 우리나라 관측소의 미세먼지 농도를 데이터베이스로 구축한다.
  - 우리나라 미세먼지 농도는 겨울철에 높으며, 2001년부터 2010년까지 뚜렷하게 감소하고 있다. 이 연구에서는 2011년 이후 겨울철 미세먼지 농도를 예측하는 데에 초점을 맞춘다.
  - 중국 관측소의 미세먼지 농도는 aqicn.org 웹사이트에 공개되어 있으며, 이를 실시간으로 수집하는 시스템을 구축하여 데이터베이스로 구축한다.
  - 한반도 미세먼지 농도에 영향을 끼칠 수 있는 잠재적 기후 인자로 북극, 고위도 지역의 Arctic Oscillation, East Atlantic 패턴, Western Pacific 패턴, Pacific/North America 패턴, Polar/Eurasian 패턴의 지수를 수집하여 데이터베이스로 구축한다.
- 기후모델을 이용해서 한반도 미세먼지 농도에 영향을 끼치는 고위도 기후 인자를 평가
  - CESM 기후모델에 기반한 극지연구소의 기후모델(KPOPS-Earth)을 이용해 과거 모의 자료를 생산한다. 이 자료를 북극, 고위도 지역의 기후 인자로 데이터베이스를 구축한다.
  - KPOPS-Earth이 모의하는 결과를 관측자료로 간주하고 있는 ERA5 재분석 자료 비교했다. KPOPS-Earth 모델이 기후 인자를 현실적으로 모의하고 있음을 알았다.
  - KPOPS-Earth에서 모의한 기후 인자와 한반도 미세먼지 농도 상관성을 살펴본 결과, 한반도에서 멀리 떨어진 지역의 기후 인자가 미세먼지의 변화에 영향을 끼칠 수 있음을 알았다.
  - 기후 인자와 한반도 미세먼지 농도와의 상관성이 같은 겨울철이더라도 월에 따라서 달라지기 때문에 어느 지역의 평균값을 모든 월에 대해서 입력자료로 활용하면 안된다.
- 예측 인자의 잠재적 중요도 판단 및 미세먼지 예측에 머신러닝 기법 활용
  - 서울의 월평균 PM<sub>2.5</sub> 농도와 상관성이 높은 기후인자의 지역 평균값을 입력자료로 대입해서 랜덤 포레스트 기법에 기반해서 PM<sub>2.5</sub> 농도 예측 모델을 개발한다.
  - 랜덤 포레스트 모델의 결과를 보고 고위도 지역 예측 인자의 중요도를 판단하였으나, 이를 통해서 어느 입력변수가 더 유용한지 특정하기 어려웠다.
  - 합성곱신경망 알고리즘은 입력자료를 지역 평균값이 아닌 격자별로 받아들이기 때문에 기후인자의 공간적 특징을 학습하기 쉬운 장점을 갖고 있다.
  - 합성곱신경망 알고리즘을 이용해서 서울의 월평균 PM<sub>2.5</sub> 농도를 예측할 수 있는 모델의 프로토타입(CNN-ERA5, CNN-KPOPS)을 개발한다.
  - CNN-ERA5 모델의 미세먼지 모의 성능은 매우 훌륭하다. CNN-KPOPS 모델을 이용한 예측값도 이에 버금가는 성능을 보였다.
  - 합성곱신경망 모델의 입력 과정에서 각 변수별, 격자별로 PM<sub>2.5</sub> 농도와의 상관계수를 곱하는 작업을 적용해 예측 성능을 향상시켰다.

## V. 연구개발결과의 활용계획

- 본 과제에서 개발한 합성곱신경망 알고리즘 기반 서울 월평균  $PM_{2.5}$  농도 예측 모델은 아직은 현업에 사용할 정도의 수준이 아닌 프로토타입으로서 향후 예측 성능의 개선을 위해서 다양한 연구가 필요하다.
- 모델이 산출하는 중장기 미세먼지 예측결과를 활용해 국가 차원에서 미세먼지 피해 예방을 위한 장기적 대응 및 저감 정책을 효과적으로 계획, 시행하는데 기여할 수 있다.
- 본 과제에서 개발한 미세먼지 농도 예측 모델을 개선시켜서 현업 예보에 활용할 수 있다.



# S U M M A R Y

## I. Title

Analysis and prediction of air quality in the Republic of Korea using Arctic/high-latitude climate variables in machine learning algorithms

## II. Purpose and Necessity of R&D

- When particulate matters (PM) concentration is high, it have an adverse effect on human health and cause socio-economic damage such as reduction of visibility, mechanical failure in facilities and deterioration of plant growth. Many countries, including Korea, are implementing policies to reduce the emission of air pollutants. Many previous studies indicated that the variation of PM concentration is associated with climate change and synoptic weather pattern. In particular, Arctic and high-latitude climate variables can affect the PM concentration in Korea.
- The need for long-term response and reduction strategies for PM-related damage at the national is growing. To effectively plan and implement polices, it is crucial to predict long-term PM concentrations while taking into account the influence of Arctic and high-latitude climate variables
- In the previous studies, ground-observation data and reanalysis data have been used as input data for models to predict PM concentrations and grades. However, these models used data which are limited to East Asian countries so these models cannot consider the relationship between large-scale atmospheric circulation and PM concentration which previous papers suggest.

## III. Contents and Extent of R&D

- Ground-observation data and reanalysis, which have an influence on PM concentration in Korea on long-term time scale, have been saved in a database.
- The Arctic and high-latitude climate variables affecting PM concentration in Korea are analyzed in past data produced by climate modeling system and saved in a database.
- A prototype of a long-term (month-season) PM concentration prediction model in Korea using machine learning techniques was constructed.

## IV. R&D Results

- Analysis on the high-latitude climate factors that affect PM concentration in Korea using ground-observation and reanalysis data
  - PM concentration data at air quality monitoring station are collected as confirmed data published by AirKorea and saved in a database
  - PM concentration in Korea is high in winter, and there has been a strong

decreasing trend from 2001 to 2010, so it is reasonable to predict PM concentration in winter from 2011.

- PM concentration at Chinese air quality monitoring station is available on the aqicn.org website. A system to automatically collect PM concentration has been established to construct a database.

- The indices of Arctic Oscillation, East Atlantic pattern, Western Pacific pattern, Pacific/North America pattern, Polar/Eurasian pattern, which potentially affecting PM concentration in Korea, have been collected as high-latitude climate factors and saved in a database.

- The correlation between the indices of high-latitude climate factors and PM concentration in Korea is low, so other methods considering the impact of high-latitude climate factors are needed.

- Evaluation for Analysis on the high-latitude climate factors that affect PM concentration in Korea using a climate model

- We used a CESM-based model (KPOPS-Earth) to produce past simulated data and built a database of Arctic and high-latitude climate variables.

- By comparing the past data simulated by KPOPS-Earth with the ERA5 reanalysis data, which can be considered as observational data, we found that KPOPS-Earth can simulate climate factors realistically.

- When we examined the correlation between the climate factors simulated by KPOPS-Earth and PM concentration in Korea, we found that regions far from Korea can have an effect on PM concentration in Korea.

- The correlation between climate factors and PM concentration in Korea varies depending on the month, making it difficult to use simple average values of a certain region as input data for PM concentration prediction model.

- Determining importance of individual high-latitude prediction factor and utilizing machine learning techniques for PM concentration prediction

- We developed a random forest-based PM<sub>2.5</sub> concentration prediction model, using regional average values that show significant correlation with monthly average PM<sub>2.5</sub> concentration in Seoul as input data.

- We determined the importance of high-latitude prediction factor using the random forest model, but it was difficult to determine useful input variables.

- Convolutional neural networks (CNN) are convenient for learning spatial feature by taking in input data in the form of grid-level rather than regional average values.

- We developed a prototype of a CNN-based model (CNN-ERA5, CNN-KPOPS) for predicting monthly average PM<sub>2.5</sub> concentration in Seoul.

- CNN-ERA5 has good prediction performance, but CNN-KPOPS still requires improvement.



- In the input process of the CNN model, the performance of the prediction was improved by applying multiplication of the correlation coefficient between  $PM_{2.5}$  concentration and each variable, grid by grid.

#### V. Application Plans of R&D Results

- The CNN-based  $PM_{2.5}$  concentration prediction model developed in this project has the potential to improve prediction performance through various methods, and can serve as a basis for further research.
- The long-term prediction results of the model can be used to effectively plan and implement national-level policies for preventing damage from PM.
- The  $PM_{2.5}$  concentration prediction model developed in this project can be used in operational forecasts.



# 목 차

제 1 장 서론 .....	13
제 1 절 연구개발의 배경 .....	13
제 2 절 연구개발의 목적 .....	14
제 3 절 연구개발의 필요성 .....	15
1. 기술적 측면 .....	15
2. 경제, 산업적 측면 .....	15
3. 사회, 문화적 측면 .....	15
제 4 절 연구개발 내용 및 범위 .....	16
제 2 장 국내·외 기술개발 현황 .....	17
제 1 절 국내 연구동향 .....	17
제 2 절 해외 연구동향 .....	17
제 3 장 연구개발 수행 내용 및 결과 .....	18
제 1 절 미세먼지 농도 예측 모델에 이용 가능한 잠재적 예측 인자의 수집 및 분석 .....	18
1. 한국, 중국 관측소에서 측정한 미세먼지 농도 자료 .....	18
2. 동아시아 기후, 대기질에 영향을 끼치는 북극, 고위도 지역 기후 인자 수집 및 분석 .....	24
제 2 절 모델링 시스템 (KPOPS-Earth)의 기후 변수 모의 성능 평가 및 한반도 미세먼지와 중장기 시간 규모 상관성 분석 .....	27
1. KPOPS-Earth 과거 모의 자료 생산과 성능 평가 .....	27
2. KPOPS-Earth의 과거 자료 모의 성능 평가 .....	28
3. KPOPS-Earth 과거 모의 자료 기반 고위도 기후 인자와 한반도 미세먼지 간의 중장기 시 간 규모 상관성 분석 .....	32
제 3 절 머신러닝 기법을 활용한 미세먼지 농도 예측 모델 개발 .....	38
1. 랜덤 포레스트 모델 .....	38
2. 합성곱신경망 모델 .....	40
제 4 장 연구개발 목표 달성도 .....	45
제 1 절 연구목표 달성도 .....	45
1. 연구목표 및 연구 내용 .....	45
2. 연구개발 성과 .....	46
제 5 장 연구개발결과의 활용계획 .....	47
제 6 장 참고문헌 .....	48

# 표 목차

표 1. 미세먼지 농도 예측 모델 관련 국내 연구(DNN은 심층신경망, RNN은 순환신경망, CNN은 합성곱신경망, RF는 random forest, XGB는 extreme gradient boosting, LGB는 light gradient boosting을 의미) .....	17
표 2. 미세먼지 농도 예측 모델 관련 국외 연구(DNN은 심층신경망, RNN은 순환신경망, CNN은 합성곱신경망, RF는 random forest, XGB는 extreme gradient boosting, LGB는 light gradient boosting을 의미) .....	17
표 3. aqicn.org에서 수집한 자료를 측정한 중국 관측소 이름 .....	20
표 4. 각 등급별 일 평균 PM2.5 농도 기준, 이에 대한 퍼센타일과 월 평균 PM2.5 농도 기준 (*는 일의 자리에서 반올림 했음을 의미) .....	42
표 5. CNN 모델의 예측 성능(R0은 상관계수를 곱하지 않은 것, R1, R2, R3은 각각 상관계수의 1승, 2승, 3승을 의미) .....	43
표 6. 본 연구개발의 연차별 연구개발 목표 및 내용 .....	45
표 7. 본 연구개발의 정성적 성과 .....	46



# 그림 목차

그림 1. AirKorea에서 제공하는 관측소의 분포 및 관측소에서 측정된 PM2.5 농도	19
그림 2. aqicn.org에서 제공하는 대기오염물질 별 베이징 AQI	20
그림 3. 여러 권역을 하나로 통합시킨 미세먼지 예측 모델의 구조	21
그림 4. 한반도 권역 간 일평균 PM10 농도의 상관계수	21
그림 5. 클러스터링 기반 권역 분류(같은 광역 권역은 같은 색으로 표시)	21
그림 6. 서울의 월평균 PM10 농도의 시계열	22
그림 7. 서울의 겨울철 평균 PM10 농도	23
그림 8. 중국과 한국 관측소의 겨울 평균 미세먼지 농도를 이용하여 구한 첫 번째 empirical orthogonal function의 고유 벡터 (Lee et al. (2020)의 그림 2a)	23
그림 9. East Atlantic 패턴, West Pacific 패턴, Pacific/North American 패턴, Polar Eurasian pattern가 양의 상태일 때의 정규화된 500-hPa 지위고도장. (Climate Prediction Center는 이 패턴을 기준으로 각 기후 인자들의 지수를 계산함)	24
그림 10. Climate Prediction Center에서 제공하는 다양한 북반구 기후 인자의 지수 자료	25
그림 11. 한반도 고농도 발생 (a) 3일 전, (b) 2일 전, (c) 1일 전의 500-hPa 지위고도 아노말리 합성장	25
그림 12. KPOPS-Earth(KPOPS-CAM6) 모델의 도식도	27
그림 13. (a) KPOPS-Earth 과거 모의 자료, (b) ERA5 재분석 자료의 2011년부터 2020년까지 겨울철 평균 850-hPa 지위고도장과 (c) 그 차이	28
그림 14. (a) KPOPS-Earth 과거 모의 자료, (b) ERA5 재분석 자료의 2011년부터 2020년까지 겨울철 평균 850-hPa 동서바람장과 (c) 그 차이	29
그림 15. (a) KPOPS-Earth 과거 모의 자료, (b) ERA5 재분석 자료의 2011년부터 2020년까지 겨울철 평균 850-hPa 남북바람장과 (c) 그 차이	30
그림 16. (a) KPOPS-Earth 과거 모의 자료, (b) ERA5 재분석 자료의 2011년부터 2020년까지 겨울철 평균 850-hPa 온도장과 (c) 그 차이	31
그림 17. 서울의 월평균 PM2.5 농도와 (a) KPOPS-Earth, (b)ERA5 자료의 850-hPa 지위고도 간 상관계수의 공간 분포	33
그림 18. 서울의 월평균 PM2.5 농도와 (a) KPOPS-Earth, (b)ERA5 자료의 850-hPa 동서바람 간 상관계수의 공간 분포	34
그림 19. 서울의 월평균 PM2.5 농도와 (a) KPOPS-Earth, (b)ERA5 자료의 850-hPa 남북바람 간 상관계수의 공간 분포	35
그림 20. 서울의 월평균 PM2.5 농도와 (a) KPOPS-Earth, (b)ERA5 자료의 850-hPa 온도 간 상관계수의 공간 분포	36
그림 21. KPOPS-Earth 자료에서 (a) 1월, (b) 2월, (c) 11월, (d) 12월 평균 서울의 PM2.5 농도와 (a) KPOPS-Earth, (b)ERA5 자료의 850-hPa 지위고도 간 상관계수의 공간 분포	37

그림 22. 월평균 서울 PM10 농도와 ERA5 재분석 자료의 (a) 850-hPa 지위고도, (b) 850-hPa 온도, (c) 850-hPa 동서바람, (d) 850-hPa 남북바람, (e) 850-hPa 습도 간 상관계수의 공간장(점은 95% 신뢰수준에서 유의함). 빨간 사각형 안 유의한 지역의 평균값은 랜덤 포레스트 기반 PM10 농도 예측 모델의 입력 자료로 사용됨. .... 39

그림 23. 랜덤 포레스트 기반 PM10 농도 예측 모델(26개)의 결정계수 분포. .... 39

그림 24. 상위 25%( $R^2 > 0.3$ )의 성능을 보이는 랜덤 포레스트 기반 PM10 농도 예측 모델의 각 입력 변수별 중요도. .... 40

그림 25. K-fold와 bagging 앙상블 방법을 이용한 자료 분류. 여기서 K는 10, N은 9에 해당 ..... 41

그림 26. CNN 모델의 모식도. Raw\_data는 ERA5 재분석장 혹은 KPOPS 예측장, Correlation map은 ERA5 재분석장과 서울의 관측 PM2.5 농도와의 상관계수장으로 사용된 변수는 총 9개 ..... 42

그림 27. 관측 및 예측의 월 평균 PM2.5 농도. 회색은 관측 PM2.5 농도, 파란색 ×는 CNN-ERA5 모델의 예측, 빨간색 +은 CNN-KPOPS의 예측이다. .... 43

그림 28. CNN-ERA5, KPOPS 모델의 예측 농도와 관측 PM2.5 농도의 산포도. .... 44



# 제 1 장 서론

## 제 1 절 연구개발의 배경

대기 중 미세먼지 농도가 높아지면 인체에 해로울 뿐만 아니라 식생의 호흡 저해, 시정 저하, 기계 결함 발생 등 사회, 경제적으로 피해를 줄 수 있다. 이에 따라 전 세계적으로 미세먼지 농도를 낮추려고 노력하고 있다. 미세먼지 농도를 줄이기 위해 우리나라를 포함한 여러 국가에서 배출량 감축 정책을 펼치고 있는 가운데, 농도의 변동이 배출량뿐만 아니라 중관 기상 패턴 및 기후 변화와 관련 있다는 연구가 발표되고 있다. 특히 최근 연구들은 극지 및 고위도 지역 기후와 동아시아 지역 미세먼지 농도 사이의 연관성을 제시하고 있다. Wang et al. (2015)은 중국 동부에서 발생하는 겨울철 미세먼지 고농도 일의 빈도가 직전 가을의 해빙 면적과 음의 상관관계를 가지며, 가을 해빙 면적의 변동이 미세먼지 고농도 일 빈도의 경년 변동 및 10년 변동의 45-67%를 설명할 수 있음을 밝혔다. Zou et al. (2017)은 CESM 모의를 통해 2012년 가을에 발생한 해빙 면적의 감소와 초겨울 유라시아의 광범위한 적설 증가로 인해 바람장이 약해져 2013년 겨울에 중국 동부 지역에서 미세먼지가 축적되기 쉬웠음을 밝혔다. 이와 비슷하게 Kim et al. (2019)도 CESM 모의 결과를 분석해서 1월 한국의 PM<sub>10</sub> 농도와 바렌츠 카라 해의 해빙 면적 간에 상관성이 있음을 보여주었다. 해빙 이외에도 우랄 블로킹의 빈도수, 시베리아 고기압과 알루시안 저기압의 강도와 위치가 한국, 중국의 미세먼지 농도에 영향을 끼친다는 연구가 발표되었다 (Lee et al., 2020; Jeong et al., 2021; Oh et al., 2018).

최근 인공지능을 이용한 예측 모델 연구가 여러 분야에서 시도되고 있으며, 대기질 예측 모델 개발에도 적용되고 있다. 과거 연구에서는 미세먼지 농도와 입력 변수 간의 선형회귀 식을 이용하여 선형적으로 미세먼지 농도를 예측했지만, 머신러닝 러닝 기법을 이용하면 비선형적인 관계까지 학습시켜서 예측할 수 있다. 특히 대기장의 변화에는 여러 복잡한 형태의 비선형적인 특성이 있어서 머신러닝 기법이 미세먼지 농도와 대기장 간 비선형적인 관계를 효과적으로 학습할 수 있다. Hur et al. (2016)은 머신러닝 기법 중의 하나인 신경망 모델을 학습하여 미세먼지 등급을 70% 정도의 수준으로 모의하였다. 다만 이 연구는 입력자료의 시간 규모와 공간 규모가 짧아서 하루에 국한된 단기 예측만 수행할 수 있다. Ho et al. (2021a)은 신경망보다 시계열 자료를 학습하기에 적합한 순환신경망 기법을 사용해 하루, 이틀 뒤 서울 PM<sub>2.5</sub> 농도 예측 모델을 개발했다. 이 모델의 예측 성능은 AirKorea 예보 성능에 견줄 정도지만 북극, 고위도 지역의 기후 인자를 고려하지 못하고 단기 예측에 국한되는 한계를 갖고 있다.

## 제 2 절 연구개발의 목적

본 연구개발의 최종목표는 아래와 같다.

- 한반도 미세먼지 농도의 중장기 예측을 위한 기후 인자 선정, 확보 및 머신러닝 예측 모델의 프로토타입 구축

지금까지 많은 연구를 통해서 지상 관측자료, 역학 모델과 대기화학 모델의 예측결과 등을 입력자료로 사용하여 우리나라의 권역별 미세먼지 농도 및 등급을 예측하는 시스템을 갖추고 있다. 그런데 이들 모델이 사용하는 기상장은 동아시아 지역에 국한되어 있어서, 최근 연구를 통해서 알게 된 극지, 고위도 지역 대규모 대기순환장과 미세먼지 농도 간의 연관성을 고려하지 못한다. 게다가 현재 사용되고 있는 대부분의 미세먼지 예측 모델은 하루, 이틀 시간 규모의 단기 예측에 국한되었다. 월 단위 이상의 미세먼지 농도를 예측할 수 있는 모델은 아직 개발되지 않은 상황이다. 미세먼지 피해에 대한 국가 차원의 장기적 대응 및 저감 대책 수립에 대한 국민의 요구가 커지고 있으므로, 이를 효과적으로 계획하고 시행하기 위해서는 보다 긴 시간 규모의 미세먼지 예측이 필요하다. 이 과제에서는 중장기 시간 규모에서 변동하는 극지, 고위도 지역의 대규모 대기순환장이 동아시아 미세먼지 농도 변동에 끼치는 영향에 대한 메커니즘을 규명하고, 처음으로 월-계절 시간 규모의 미세먼지 농도 예측을 시도하고자 한다.



## 제 3 절 연구개발의 필요성

### 1. 기술적 측면

- 중장기 시간 규모에 적합한 머신러닝 기반 미세먼지 농도를 예측하는 모델 개발 필요
  - 일반적으로 머신러닝 기법은 몇 시간, 며칠 뒤 미세먼지 농도 예측에 쓰이므로 중장기 시간 규모의 미세먼지 농도 예측에 적합한 머신러닝 기법이 필요하다.
  - 국내 현업에서는 단기 예보만 하고 있으므로 중장기 예보 기술이 필요하다.

### 2. 경제, 산업적 측면

- 미세먼지로 인한 경제 피해 예방 및 대응책 수립 필요
  - 장기적으로 국가 차원에서 미세먼지 관련 정책을 만들기 위해 신뢰할만한 지표가 필요하다.
  - 보다 장기간의 시간 규모에서 미세먼지 농도를 전망할 필요가 있다.
- 미세먼지 배출 저감, 미세먼지 피해 예방 관련 산업의 가이드라인 필요
  - 중장기 시간 규모에서 미세먼지 배출을 예측해야 한다.
  - 미세먼지 배출 저감, 미세먼지 피해 예방 관련 산업의 수요와 공급을 예측할 필요가 있다.

### 3. 사회, 문화적 측면

- 미세먼지 문제의 해결을 원하는 국민의 니즈에 부합
  - 미세먼지 상황이 2000년 이후 크게 개선되었지만, 고농도 사례는 그대로여서 미세먼지 피해에 관한 국민의 관심이 증가하였다.
  - 미세먼지 문제 해결을 위해 중장기 시간 규모에서 미세먼지 농도가 증가하는 원인을 규명할 필요가 있다.



## 제 4 절 연구개발내용 및 범위

- 관측/재분석 자료를 분석해서 중장기 시간 규모에서 우리나라 미세먼지 농도에 영향을 끼치는 북극, 고위도 기후 인자를 분석하고 이를 데이터베이스로 구축하고자 함.
  - 한국, 중국 관측소에서 측정한 미세먼지 농도 자료를 수집하고 데이터베이스로 정리
  - 미세먼지 농도 예측에 필요한 북극, 고위도 지역의 기후 인자를 데이터베이스로 정리
  - 데이터베이스로 구축한 북극, 고위도 지역의 기후 인자와 미세먼지 농도 간 상관성 분석
  
- 기후 모델링 시스템이 모의한 과거 자료에서 한반도 미세먼지 농도에 영향을 끼치는 기후 인자를 분석한 뒤 이를 추출하여 데이터베이스로 구축하고자 함.
  - 기후모델링 시스템을 이용해서 과거 모의 자료를 생산하고 북극, 고위도 지역의 기후 인자를 추출하여 데이터베이스에 추가
  - 과거 모의 자료와 한반도 미세먼지 농도 간 중장기 시간 규모 상관성을 분석
  
- 머신러닝 기법을 이용해 중장기(월-계절) 시간 규모의 한반도 미세먼지 농도 예측 모델의 프로토타입을 구축하고자 함.
  - 랜덤 포레스트 기법을 이용하여 각 잠재적 예측 인자의 미세먼지 예측에의 중요도를 분석
  - 합성곱신경망 모델을 이용하여 북극 및 고위도 지역의 기후 기반 중장기 미세먼지 농도 예측 모델을 개발



## 제 2 장 국내외 기술개발 현황

### 제 1 절 국내 연구동향

머신러닝 기법을 활용하여 미세먼지 농도를 예측하는 모델을 개발하는 연구는 사용한 방법과 예보하고자 한 시간에 따라 분류할 수 있으며 이를 표 1에 정리했다. 2010년대의 연구(Hur et al., 2016; Park et al., 2018)에서는 머신러닝 기법 중에서도 가장 기초적인 DNN을 주로 사용했지만 2020년대에는 RNN, CNN, XGB, LGB 등 다양한 기법을 이용해 미세먼지 농도를 예측한다. 하지만 예측 모델들이 주로 예측하고자 하는 시간대는 하루에서 길게는 일주일로 중장기 시간 규모에서의 예측에 관한 연구가 부족한 실정이다. 또한, 연구들에서 사용하는 입력 자료는 극지, 고위도 등 한반도에서 멀리 떨어진 지역의 대기장 자료를 사용하고 있지 않다.

표 1. 미세먼지 농도 예측 모델 관련 국내 연구(DNN은 심층신경망, RNN은 순환신경망, CNN은 합성곱신경망, RF는 random forest, XGB는 extreme gradient boosting, LGB는 light gradient boosting을 의미)

논문	사용한 기법	예보 시간	예측한 변수	분석한 지역
Hur et al. (2016)	DNN	Daily	PM <sub>10</sub>	서울
Park et al. (2018)	DNN	Daily	PM <sub>10</sub>	서울
Kim et a. (2019)	RNN	Daily	PM <sub>10</sub> , PM <sub>2.5</sub>	서울과 광역시
Yang et al. (2020)	CNN+RNN	Daily to a week	PM <sub>2.5</sub>	서울
Ho et al. (2021)	RF, XGB, LGB	Daily	PM <sub>2.5</sub>	서울
Sayed et al. (2021)	CNN	Daily to a week	PM <sub>10</sub> , PM <sub>2.5</sub>	전국
Kim et al. (2022)	RNN	Daily	PM <sub>10</sub> , PM <sub>2.5</sub>	서울

### 제 2 절 국외 연구동향

국외에서 수행된 미세먼지 농도 예측 모델 관련 연구를 표 2에 정리했다. 국외 연구도 국내 연구와 마찬가지로 중장기 시간 규모의 미세먼지 농도를 예측하기보다 몇 시간, 몇 일 뒤 농도를 예측하는 연구가 많다. 입력 자료도 주변 지역의 관측소에서 측정된 미세먼지 농도, 대기 변수, aerosol optical depth 자료 등을 이용하기 때문에 멀리 떨어진 지역이 끼치는 영향을 고려할 수 있는 모델 구조가 아니다.

표 2. 미세먼지 농도 예측 모델 관련 국외 연구(DNN은 심층신경망, RNN은 순환신경망, CNN은 합성곱신경망, RF는 random forest, XGB는 extreme gradient boosting, LGB는 light gradient boosting을 의미)

논문	사용한 기법	예보 시간	예측한 변수	분석한 국가
Li et al. (2017)	RNN	Hourly	PM <sub>2.5</sub>	중국
Huang and Kuo (2018)	CNN+RNN	Daily	PM <sub>2.5</sub>	중국
Zhao et al. (2019)	DNN, RNN	Daily	PM <sub>2.5</sub>	중국
Doreswamy et al. (2020)	RF, XGB	Daily	PM <sub>10</sub> , PM <sub>2.5</sub>	대만
Masood et al. (2020)	DNN	Daily	PM <sub>2.5</sub>	인도
Faraji et al. (2022)	CNN+RNN	Daily	PM <sub>2.5</sub>	이란

## 제 3 장 연구개발수행 내용 및 결과

### 제 1 절 미세먼지 농도 예측 모델에 이용 가능한 잠재적 예측 인자의 수집 및 분석

머신러닝 기법은 수많은 데이터 간 상관관계를 선형적, 비선형적으로 파악하는 방법이어서 이를 활용하면 미래에 특정 변수가 어떻게 변할지를 예측할 수 있다. 그렇기에 머신러닝 기반 예측모델 개발의 가장 첫 단계는 예측하고자 하는 변수(예측 변수)를 정하는 것이고 다음 단계는 이 변수와 연관이 있는 다른 변수(입력 변수)를 선정한 뒤 예측 변수와 입력 변수의 데이터를 수집하는 것이다. 본 과제의 1차년도 주목표는 예측 변수, 입력 변수 선정과 그 데이터를 수집하여 데이터베이스를 구축하고 변수들의 특성을 분석하는 것이다. 당시에는 모델의 구체적인 구조가 정해지지 않아 최대한 다양한 변수를 실험실 서버에 데이터베이스로 구축했다.

#### 1. 한국, 중국 관측소에서 측정된 미세먼지 농도 자료

##### 가. 미세먼지 농도 자료의 수집과 데이터베이스 구축

본 과제에서 개발할 머신러닝 기반 모델의 목표는 서울 미세먼지 농도를 예측하는 것이다. 예측 변수로 쓰일 서울 관측소의 미세먼지 농도 자료를 수집하고, 서울 이외의 미세먼지 농도 자료가 입력 변수로 쓰일 수 있어서 한국 내 모든 관측소의 미세먼지 농도 자료를 수집해야 한다. AirKorea는 2023년 기준 약 400여 개의 관측소에서 측정된 대기오염물질 농도를 실시간으로 웹사이트에 그림 1와 같이 공개하고 있다. 실시간 대기오염물질 농도 자료에는 이상치가 있을 수 있어서 AirKorea는 실시간 농도 자료에 quality control를 적용하여 주기적으로 대기오염물질 농도의 최종확정 측정자료를 엑셀, csv 형태로 배포하고 있다. 최종확정 측정자료는 2001년부터 2022년까지 시간 단위로 제공되고 있으며 각각 관측소마다 관측 기간이 다르다. 2001년부터 2022년까지의 최종확정 측정자료를 실험실 서버에 수집한 후 모델의 입력 변수에 적합한 형태로 후처리 작업을 하여 한국 관측소 미세먼지 농도 자료를 데이터베이스로 구축했다. 하지만 2001년부터 2015년까지의 PM<sub>2.5</sub> 농도 자료는 AirKorea에서 제공되지 않아 보건환경 연구소에 자료를 요청해 서울의 25개 구에 위치한 도시대기 측정 관측소 자료를 수집하여 데이터베이스로 구축했다.

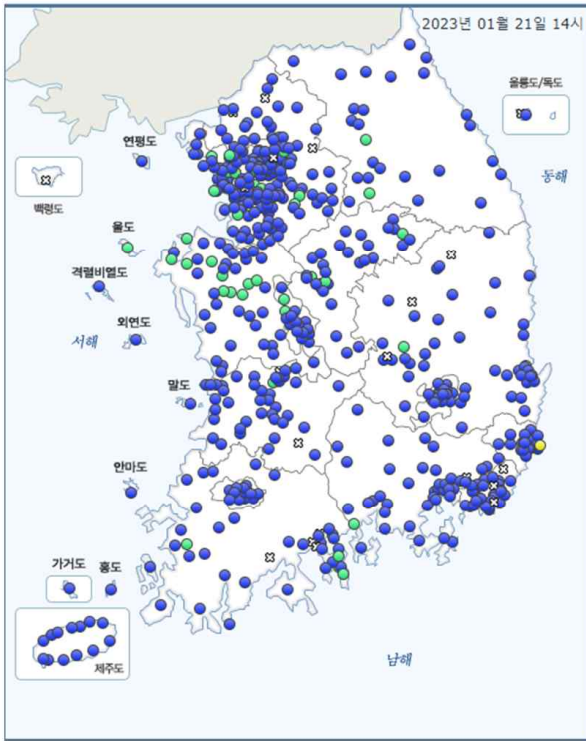


그림 1. AirKorea에서 제공하는 관측소의 분포 및 관측소에서 측정된 PM<sub>2.5</sub> 농도

중국 대기오염물질 농도 자료는 그림 2처럼 aqicn.org 웹사이트에 실시간으로 공개되고 있지만 AirKorea처럼 최종확정 측정자료를 파일 형태로 제공되진 않는다. 중국 대기오염물질 자료를 수집하려면 웹사이트에 올려진 실시간 자료를 바로 다운로드하는 방법밖에 없다. 실험실 서버 내에 중국 대기오염물질 자료를 실시간으로 수집하는 웹크롤링 코드를 개발하고 이를 특정 시간에 실행되도록 설계했다. 이렇게 수집된 중국 대기오염물질 자료의 단위는 대기질 지수 (air quality index, AQI), AirKorea에서 제공하는 PM<sub>2.5</sub> 농도 자료의 단위는  $\mu\text{g m}^{-3}$ 이로 서로 단위가 다르다. AQI는 대기질이 좋은지 나쁜지를 단순히 농도로 제공하기보다 이를 기반으로 특정 단계를 정해 일반인이 현재 대기질 상태가 어떠한지 쉽게 이해하도록 도와주며, 어떤 행동을 해야 하는지 알려주는 지표이다. AQI의 정의는 각 국가마다 다르므로 중국에서 사용하는 AQI 계산식을 이용하여 AQI 지수를  $\mu\text{g m}^{-3}$ 로 변환했다. 자동으로 중국 대기오염물질 농도 자료를 수집한 후 AQI 지수를 한국에서 쓰이는 농도 단위로 변환하는 것으로 실험실 내 중국 대기오염물질 데이터베이스를 구축했다. 데이터베이스에는 2014년부터 2022년까지 총 31개 중국 관측소의 PM<sub>2.5</sub>, PM<sub>10</sub> 농도 자료가 수집되었으며 31개의 관측소 이름은 표 3와 같다.

# 베이징의 대기 오염 : 실시간 대기질 지수 (AQI)

베이징 北京 | 상하이 上海 | 광둥 广州 | 청두 成都 | 홍콩 香港 | 난징 (南京) 南京 | 가장 가까운 도시의 위치를 찾 | 당신의 도시를 검색하세요



그림 2. aqicn.org에서 제공하는 대기오염물질 별 베이징 AQI

표 3. aqicn.org에서 수집한 자료를 측정한 중국 관측소 이름

중국 관측소 이름					
Beijing	Datong	Huhehaote	Qingdao	Shijiazhuang	Zhenzhou
Changchun	Fushun	Jinan	Qinhuangdao	Tianjin	
Chengdu	Fuzhou	Lanzhou	Qiqihaer	Weifang	
Chifeng	Guangzhou	Lianyungang	Rizhao	Wenzhou	
Chongqing	Haerbin	Mudanjiang	Shanghai	Wuhan	
Dalian	Hefei	Nanjing	Shenyang	Yantai	

## 나. PM 농도 자료의 시공간적 특성 및 상관성 분석

머신러닝 기법을 사용하는 예측 모델의 성능은 자료의 양과 질에 크게 좌우된다. 대기오염물질의 측정 기간은 길어야 20년이므로 자료의 양이 적어서 머신러닝 학습 과정에 문제가 생길 수 있다. 자료의 양이 적을 때 자료의 수를 늘리기 위한 다양한 방법이 있으며 그중 하나로 상관성이 높은 지역의 자료를 모두 학습에 이용하는 방법이 있다. 예를 들어 서울의 미세먼지 농도를 예측하기 위해 서울의 미세먼지 농도만 예측 변수로 사용하는 모델과 그림 3와 같이 서울 미세먼지 농도와 높은 상관성을 지닌 인천, 경기 지역의 미세먼지 농도까지 평균한 수도권 권역 미세먼지 예측 모델을 생각해 볼 수 있다. 후자의 모델은 전자보다 학습 과정에서 서울 말고도 인천, 경기 자료까지 이용할 수 있으므로 자료의 수가 3배 증가한다. 어떤 지역끼리 묶어서 자료의 수를 늘릴 수 있는지 살펴보기 위해 17개 권역 간 일평균 미세먼지 농도의 상관계수를 확인했고, 그 결과는 그림 4와 같다. 두 지역의 상관계수가 최소 0.8 이상일 때 하나의 지역으로 분류하면 17개의 권역이 수도권청권 (서울, 인천, 세종, 대전, 경기, 충북, 충남), 강원경상권 (대구, 울산, 부산, 강원, 경북, 경남), 전라제주권 (광주, 전북, 전남, 제주)의 3개의 광역 권역으로 분류되었다. 추가로 군집 분류에 주로 이용되는 클러스터링 방법을 이용하여 같은 분

석을 수행하면 그림 5과 같이 17개의 권역이 수도권+영서 (서울, 인천, 경기북부, 경기남부, 영서), 경상권 (대구, 울산, 부산, 경북, 경남), 전라권 (광주, 전북, 전남), 영동권 (영동), 제주권 (제주)의 6개의 광역 권역으로 분류되었다.

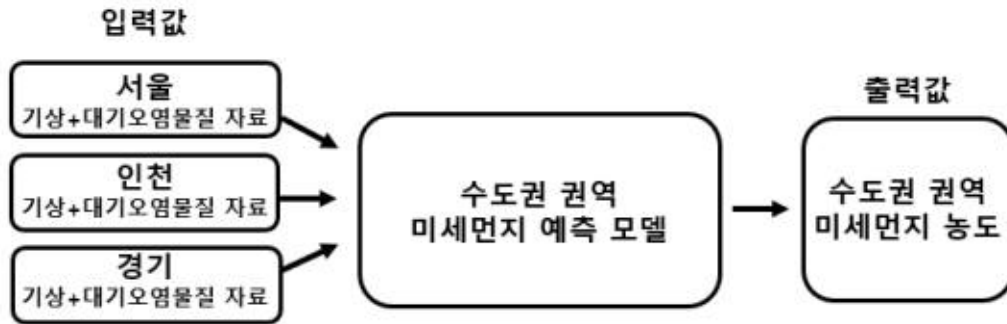


그림 3. 여러 권역을 하나로 통합시킨 미세먼지 예측 모델의 구조

권역	서울	경기	인천	충남	충북	대전	세종	광주	경북	대구	경남	울산	부산	전북	광주	전남	제주
서울	1.0	1.0	1.0	0.9	0.9	0.9	0.8	0.8	0.7	0.7	0.7	0.6	0.6	0.8	0.7	0.7	0.5
경기	1.0	1.0	1.0	0.9	0.9	0.9	0.8	0.8	0.7	0.7	0.7	0.6	0.6	0.8	0.7	0.7	0.5
인천	1.0	1.0	1.0	0.9	0.8	0.8	0.8	0.8	0.7	0.7	0.7	0.6	0.6	0.8	0.7	0.7	0.5
충남	0.9	0.9	0.9	1.0	0.9	0.9	0.9	0.8	0.8	0.7	0.7	0.7	0.7	0.9	0.8	0.8	0.6
충북	0.9	0.9	0.8	0.9	1.0	0.9	0.8	0.9	0.9	0.8	0.8	0.7	0.7	0.9	0.8	0.8	0.6
대전	0.9	0.9	0.8	0.9	0.9	1.0	0.9	0.8	0.8	0.8	0.8	0.7	0.7	0.9	0.9	0.8	0.6
세종	0.8	0.8	0.8	0.9	0.8	0.9	1.0	0.7	0.7	0.7	0.7	0.6	0.6	0.8	0.8	0.7	0.5
광주	0.8	0.8	0.8	0.8	0.9	0.8	0.7	1.0	0.9	0.8	0.8	0.8	0.8	0.8	0.7	0.8	0.6
경북	0.7	0.7	0.7	0.8	0.9	0.8	0.7	0.9	1.0	0.9	0.9	0.9	0.9	0.8	0.8	0.9	0.7
대구	0.7	0.7	0.7	0.7	0.8	0.8	0.7	0.8	0.9	1.0	0.9	0.9	0.9	0.8	0.8	0.9	0.7
경남	0.7	0.7	0.7	0.7	0.8	0.8	0.7	0.8	0.9	0.8	1.0	1.0	1.0	0.8	0.8	0.9	0.8
울산	0.6	0.6	0.6	0.7	0.7	0.7	0.6	0.8	0.9	0.9	1.0	1.0	1.0	0.7	0.8	0.9	0.7
부산	0.6	0.6	0.6	0.7	0.7	0.7	0.6	0.8	0.9	0.9	1.0	1.0	1.0	0.7	0.8	0.9	0.8
전북	0.8	0.8	0.8	0.9	0.9	0.9	0.8	0.8	0.8	0.8	0.8	0.7	0.7	1.0	0.9	0.9	0.7
광주	0.7	0.7	0.7	0.8	0.8	0.9	0.8	0.7	0.8	0.8	0.8	0.8	0.8	0.9	1.0	0.9	0.8
전남	0.7	0.7	0.7	0.8	0.8	0.8	0.7	0.8	0.9	0.9	0.9	0.9	0.9	0.9	0.9	1.0	0.8
제주	0.5	0.5	0.5	0.6	0.6	0.6	0.5	0.6	0.7	0.7	0.8	0.7	0.8	0.7	0.8	0.8	1.0

그림 4. 한반도 권역 간 일평균 PM<sub>10</sub> 농도의 상관계수

특자연구소



그림 5. 클러스터링 기반 권역 분류(같은 광역 권역은 같은 색으로 표시)

한반도 내 미세먼지 농도는 공간적으로 다른 특성을 보일 뿐만 아니라 시간적 특성이 다를 수 있다. 계절마다 미세먼지 농도와 다른 변수 간의 상관성이 다를 수 있는데 이 경우 여름철 자료를 학습하면 겨울철 농도 예측에 부정적인 영향을 줄 수 있다. 또한, 머신러닝 기반 미세먼지 농도 예측 모델의 예측 기간이 어느 계절이냐에 따라 예측 성능이 다를 수 있다. 그래서 미세먼지 농도의 시간적 특성을 분석하고 이후 예측 모델의 구조를 결정해야 한다. 가장 먼저 서울 PM<sub>10</sub> 농도의 특성이 월별로 얼마나 다른지 확인했다. 그림 7은 서울의 월평균 PM<sub>10</sub> 농도의 시계열이다. PM<sub>10</sub> 농도는 대체로 봄철, 겨울철에 높고, 여름철, 가을철에 낮다. 봄철, 겨울철에는 북서풍 계열의 바람으로 중국으로부터 수송되는 대기오염물질의 영향을 받아 PM<sub>10</sub> 농도가 높은 것으로, 여름철, 가을철에는 중국으로부터의 수송이 줄어들며 강수가 많아 세정 작용으로 PM<sub>10</sub> 농도가 낮아진 것으로 보인다. 봄철 PM<sub>10</sub> 농도가 가장 높지만 황사의 영향이 크기 때문에 그다음으로 PM<sub>10</sub> 농도가 높은 겨울철(1, 2, 11, 12월)에 대해 미세먼지 농도 예측 모델을 만드는 것이 합리적이다.

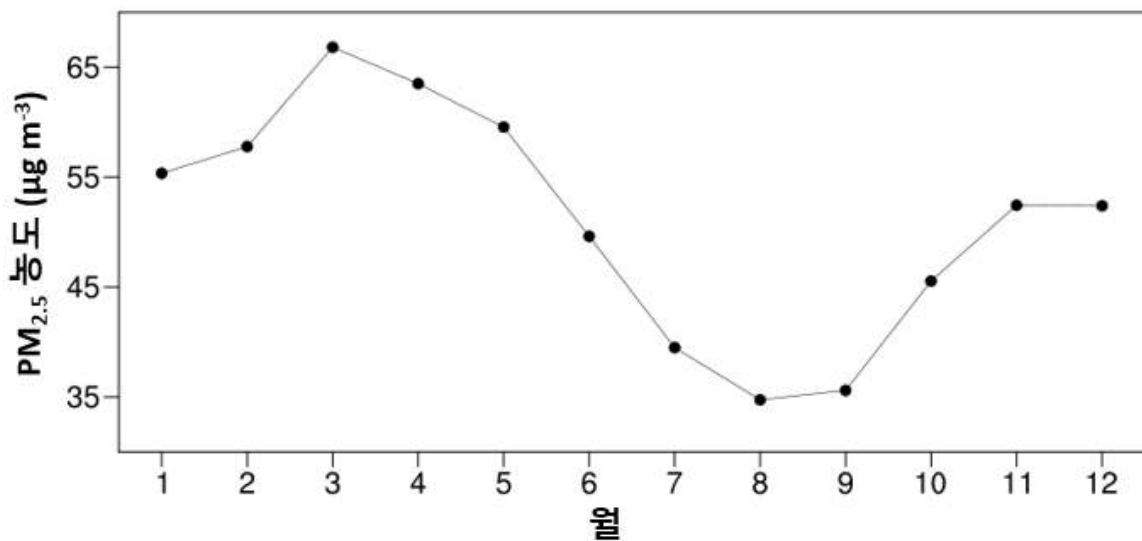


그림 6. 서울의 월평균 PM<sub>10</sub> 농도의 시계열

서울 PM<sub>10</sub> 농도의 시간적 특성으로 겨울철 평균(1, 2, 11, 12월) PM<sub>10</sub> 농도의 연도별 변화를 살펴보았다(그림 8). 2000년대에는 농도가 60 µg m<sup>-3</sup>였지만 2010년대에는 농도가 50 µg m<sup>-3</sup>로 서울의 대기질이 크게 개선되었다. 주목할만한 점은 2000년대부터 2010년대까지 PM<sub>10</sub> 농도가 60 µg m<sup>-3</sup>에서 50 µg m<sup>-3</sup>로 천천히 감소한 것이 아니라 2010년과 2011년 사이에 계단식으로 감소했다는 것이다. 또한, 2019년과 2020년 PM<sub>10</sub> 농도는 40 µg m<sup>-3</sup>로 2018년 이후 다시 계단식으로 감소했다. 2000년대와 2010년대의 PM<sub>10</sub> 농도 감소의 원인은 국가에서 시행한 대기오염 물질 배출량 감소 정책 때문으로 보이고(Ho et al., 2021b), 2019년과 2020년의 PM<sub>10</sub> 농도 감소의 원인은 COVID-19의 영향으로 인한 배출량 감소로 보인다. 2019년과 2020년의 PM<sub>10</sub> 농도는 COVID-19라는 특수한 상황으로 농도가 낮아졌다는 점을 고려하면 2000년대와 2010년대의 서로 다른 특성을 지닌 시기라 볼 수 있다. 본 과제에서 개발할 미세먼지 농도 예측 모델의 궁극적인 목표는 미래의 미세먼지 농도를 예측하는 것이기에 현재와 가까운 2010년대 자료인 2011년부터 2020년까지 자료를 사용하여 미세먼지 농도 예측 모델을 개발하고자 했다.

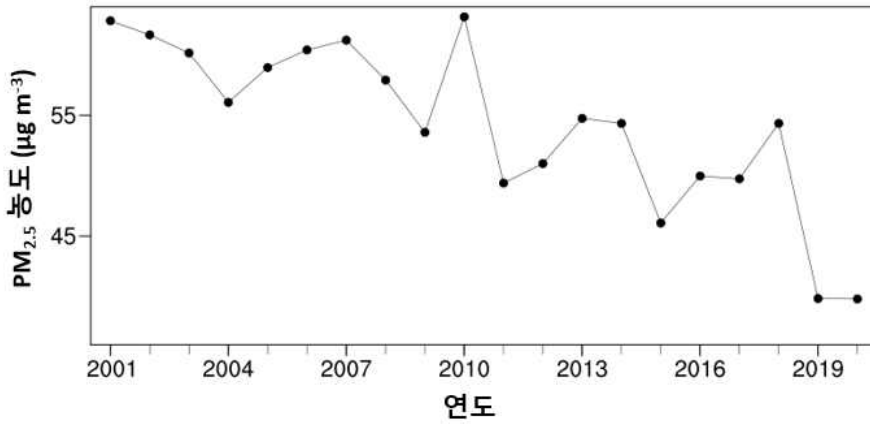


그림 7. 서울의 겨울철 평균 PM<sub>10</sub> 농도

선행연구에 따르면 중국의 미세먼지 농도는 한국의 미세먼지 농도 간 연관성이 있다. 일 시간 규모에서는 한국과 중국의 미세먼지 농도 간 양의 상관관계가 있으며, 보통 한국의 미세먼지 고농도가 발생하기 하루, 이틀 전 중국의 미세먼지 농도가 먼저 증가한다(Lee et al., 2011, 2013; Oh et al., 2015). 반면, 월 시간 규모에서 empirical orthogonal function 방법으로 중국과 한국의 미세먼지 농도의 연관성을 살펴보면 음의 상관관계가 보인다(그림 8). 시간 규모에 따라 한반도 미세먼지 농도와 중국 관측소에서 측정된 미세먼지 농도 간 상관성이 다르다.

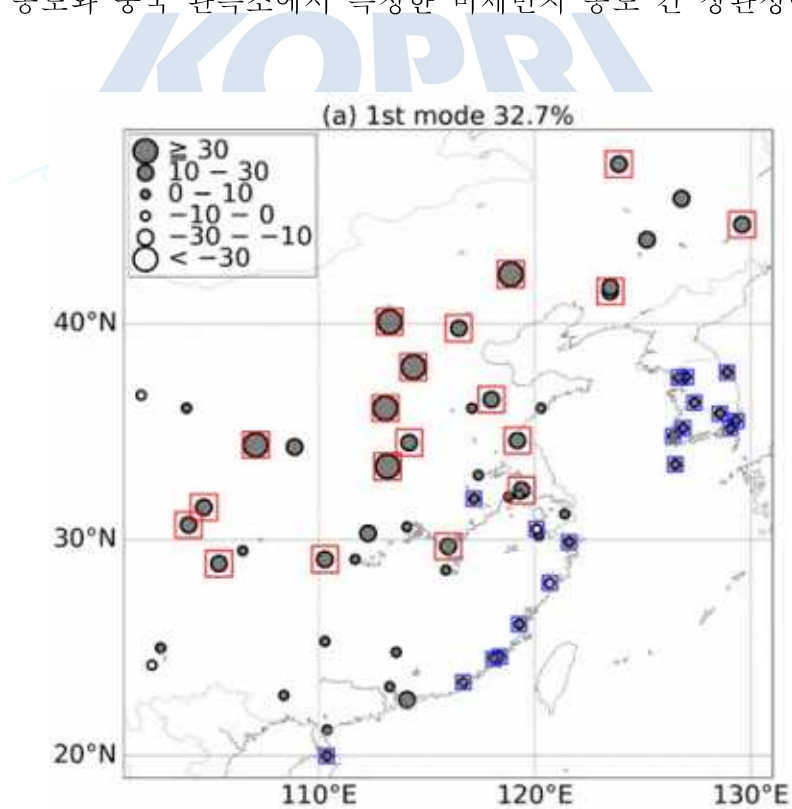


그림 8. 중국과 한국 관측소의 겨울 평균 미세먼지 농도를 이용하여 구한 첫 번째 empirical orthogonal function의 고유 벡터 (Lee et al. (2020)의 그림 2a)



## 2. 동아시아 기후, 대기질에 영향을 끼치는 북극, 고위도 지역 기후 인자 수집 및 분석

### 가. 선행연구 조사 및 정리

다양한 기후 인자가 증장기 시간 규모에서 동아시아 기후에 영향을 끼쳐 한반도의 대기 조건을 바꿀 수 있다. 이렇게 바뀐 대기 조건은 한반도 대기오염물질의 배출, 축적, 환기에 영향을 끼쳐 최종적으로 기후 인자는 한국의 대기질에 영향을 끼친다. 한반도 대기질에 영향을 끼칠 수 있는 기후 인자를 살펴본 후 잠재적 예측 인자가 될 수 있는 기후 인자로 시베리아 고기압, 알루시안 저기압, Arctic Oscillation (AO) 패턴, East Atlantic (EA) 패턴, Western Pacific (WP) 패턴, Pacific/North America (PNA) 패턴, Polar/Eurasian (PE) 패턴을 선정했다 (Jia et al., 2015.; Jeong and Park, 2017; Zou et al., 2017; Oh et al., 2018; Yun and Yoo 2019). 시베리아 고기압, 알루시안 저기압은 동아시아 겨울 몬순 시스템의 한 축을 담당하는 요소로 이 두 가지 기후 인자는 동아시아 대기질에 직접적인 영향을 줄 것으로 보인다. AO 인덱스는 극소용돌이의 강화, 약화를 나타내는 지수로 태평양의 제트기류에 영향을 끼쳐 한반도 대기질이 변할 수 있다. EA, WP, PNA, PE 패턴(그림 9)은 북반구 대기순환장과 연관된 기후 인자로 한반도 주변 지역의 대기순환장에는 큰 영향을 끼치지 않기 때문에 한반도 대기질에 직접적 영향을 끼치지 않을 것으로 보인다.

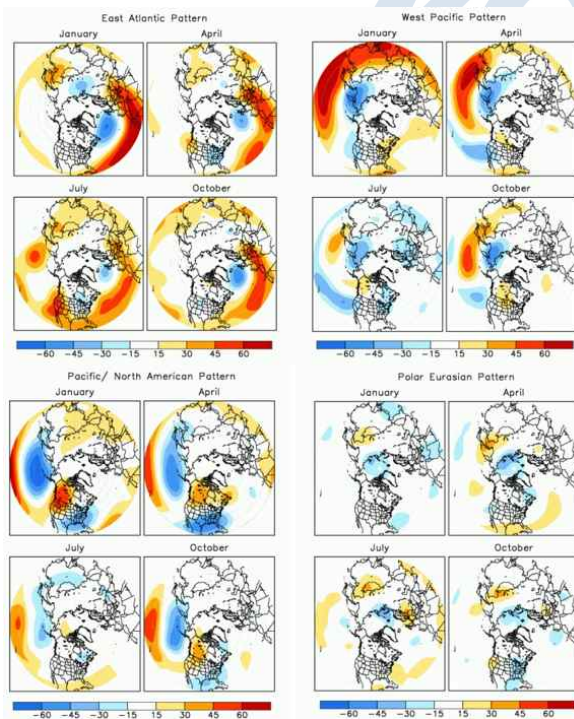


그림 9. East Atlantic 패턴, West Pacific 패턴, Pacific/North American 패턴, Polar Eurasian pattern가 양의 상태일 때의 정규화된 500-hPa 지위고도장. (Climate Prediction Center는 이 패턴을 기준으로 각 기후 인자들의 지수를 계산함)

### 나. 북극, 고위도 기후 인자의 수집과 데이터베이스 구축

시베리아 고기압, 알루시안 저기압의 세기는 각각 시베리아 지역, 알루시안 저기압의 평균 해면기압값으로 정의할 수 있다. 이 두 변수는 재분석 자료를 이용하여 직접 계산한 값을 데이터베이스로 구축했다. AO와 EA, WP, PNA, PE 패턴은 미국의 Climate Prediction Center에서 제공하는 다양한 기후 인자의 지수 자료를 수집하여 데이터베이스로 구축했다 (그림 10).

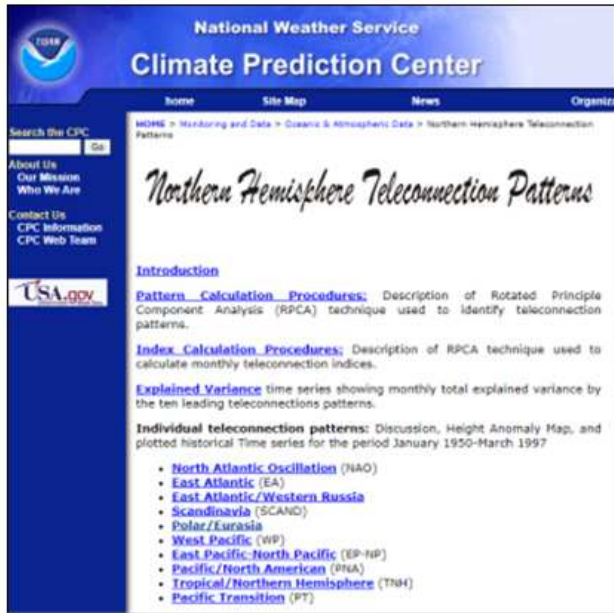


그림 10. Climate Prediction Center에서 제공하는 다양한 북반구 기후 인자의 지수 자료

다. 기후 인자와 한반도 미세먼지 간의 중장기 시간 규모에서의 상관성 분석

시베리아 고기압의 강도는 동아시아 지역의 풍속, 풍향을 바꾸어 한반도 미세먼지 농도에 영향을 끼칠 수 있다. Oh et al. (2018)의 결과에 따르면, 일 시간 규모에서 알루시안 저기압이 강할수록 환기효과를 방해해 한반도의 고농도 미세먼지 사례가 발생하는데(그림 11) 중장기 시간 규모에서도 알루시안 저기압의 강도와 한반도 미세먼지 간 상관성이 있었다. AO, EU 패턴, WP 패턴, PNA 패턴, EP 패턴 지수와 서울 PM<sub>10</sub> 농도 간 중장기 상관성 분석한 결과, 유의미한 상관성은 나타나지 않았다. 기존에 알려진 기후 패턴을 대표하는 지수를 사용하는 것보다 알루시안 저기압의 강도처럼 특정 지역에서 대기순환장의 형태나 세기를 고려하는 식으로 북극, 고위도 영향을 고려하는 것이 미세먼지 예측에 더 적합할 것이다. 머신러닝 기법 중 심층 신경망이나 순환신경망을 사용한다면 특정 지역의 평균값이나 Hur et al. (2016)처럼 코사인 유사도를 사용하는 것이 바람직할 것으로 보이며, 합성곱신경망을 이용한다면 북극, 고위도 지역에 해당하는 모든 지역의 격자별 변수값을 사용하는 것이 합당하다.

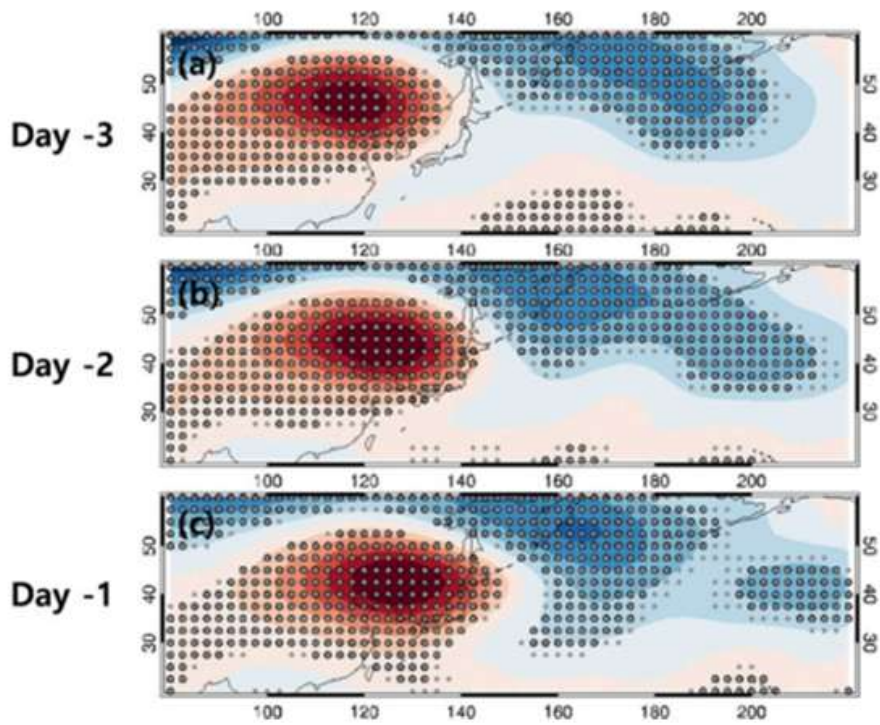


그림 11. 한반도 고농도 발생 (a) 3일 전, (b) 2일 전, (c) 1일 전의 500-hPa 지위고도 아노말리 합성장



## 제 2 절 모델링 시스템 (KPOPS-Earth)의 기후 변수 모의 성능 평가 및 한반도 미세먼지와의 중장기 시간 규모 상관성 분석

### 1. KPOPS-Earth 과거 모의 자료 생산과 성능 평가

#### 가. KPOPS-Earth 과거 모의 자료의 데이터베이스 구축

본 과제의 궁극적인 목표는 현업에 활용 가능한 미세먼지 농도 예측 모델을 만드는 것이다. 현업에서 미래의 미세먼지 농도를 예측하기 위해서는 실시간으로 모델이 예보한 대기 변수를 입력 자료로 사용해야 한다. 하지만 예측에 앞서 머신러닝 기법으로 한반도 미세먼지 농도와 모델 자료 간 상관성을 학습할 때는 이미 관측된 미세먼지 농도 자료와 과거 모의 자료를 이용해야 한다. 머신러닝 기반 미세먼지 농도 예측 모델 개발을 위해 KPOPS-Earth 모델 시스템을 사용했으며 그 구조와 자료의 생산 과정은 그림 12과 같다. KPOPS-Earth 모델은 Community Earth System Model v2 모델 기반으로 대기모델은 Earth를, 지면 모델로는 CLM5로 구성된다. 해수면 온도/해빙 경계조건은 2010년 이전에 CFSv1 자료를, 2011년부터는 CFSv2를, Earth의 대기 초기조건으로는 ERA5 재분석 자료를, 지면 초기조건으로는 CLM5의 20년 시물레이션 자료를 사용했다. 극지연구소 서버에 KPOPS-Earth 모델 시스템을 구축한 후 2000년부터 2022년까지 1, 2, 10, 11, 12월 1일을 기준으로 120일 동안 모의하여 11개의 앙상블로 과거 모의 자료를 생산했다. 연직 자료로 1000, 925, 850, 500, 300, 100, 50 hPa층의 지위고도, 동서바람, 남북바람, 연직바람, 온도가, 지상 자료로 대기경계층 고도, 해면기압, 기압, 2m 온도, 해수면 온도, 해빙 면적이 있으며 이들 자료를 연구실 서버에 데이터베이스로 구축했다.

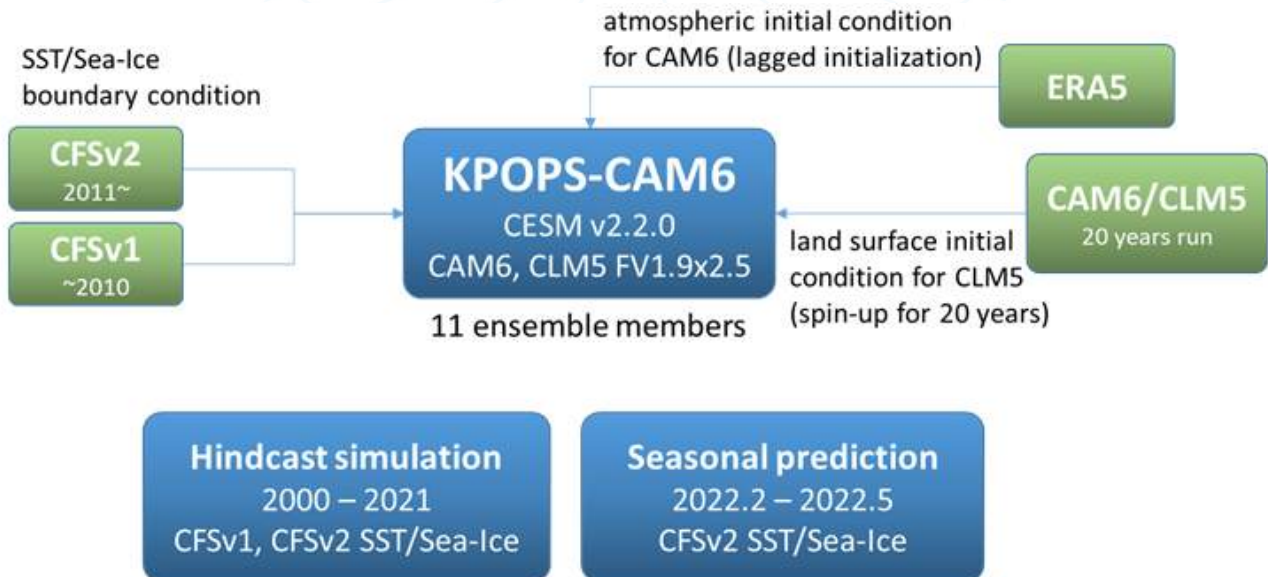


그림 12. KPOPS-Earth(KPOPS-CAM6) 모델의 도식도

## 2. KPOPS-Earth의 과거 자료 모의 성능 평가

KPOPS-Earth 자료를 예측 모델의 입력 변수로 사용하기 전에 KPOPS-Earth 모델의 성능을 확인했다. KPOPS-Earth의 과거 모의 자료와 관측 자료라고 볼 수 있는 ERA5 재분석 자료로 2011년부터 2020년까지 겨울철(1, 2, 11, 12월) 평균장을 비교하는 방법으로 성능을 평가했다. 성능 평가에서는 4개의 변수(850-hPa 지위고도, 850-hPa 동서바람, 850-hPa 남북바람, 850-hPa 온도)를 사용했다.

KPOPS-Earth 과거 모의 자료와 ERA5 재분석 자료의 850-hPa 지위고도의 공간분포를 보면(그림 15), 두 자료 모두에서 중위도 지역의 값이 높지만 고위도로 갈수록 그 값이 감소한다. 이러한 감소 경향은 태평양, 대서양에서 더욱 뚜렷하게 나타났다. 두 자료에서 공간장 차이를 살펴보면, KPOPS-Earth는 유럽 남부 지역에서 아프리카 북부 지역까지 850-hPa 지위고도를 과대모의하고 유럽 북서부 지역에서 과소모의 하는 경향이 있다. 북태평양 지역에서도 과대모의가 나타나며, 특히 히말라야 산맥이 위치한 지역에서는 심하게 과대모의 되고 있다. 동아시아 지역에서는 KPOPS-Earth, ERA5 간 850-hPa 지위고도 차이가 적다.

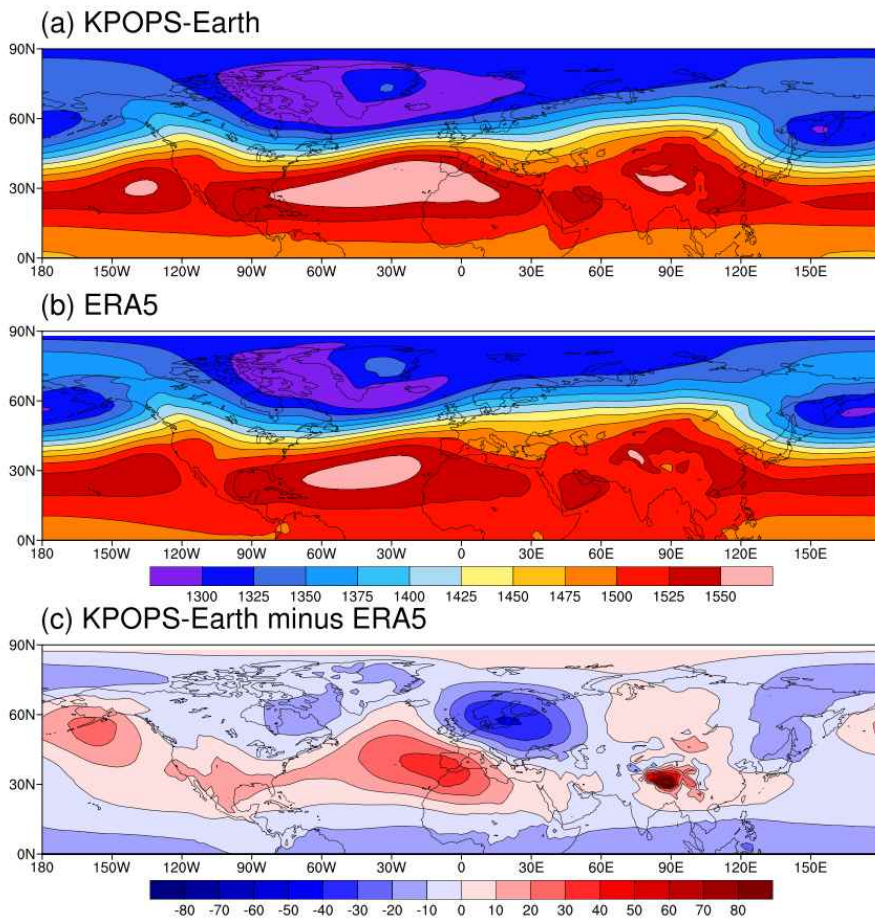


그림 13. (a) KPOPS-Earth 과거 모의 자료, (b) ERA5 재분석 자료의 2011년부터 2020년까지 겨울철 평균 850-hPa 지위고도장과 (c) 그 차이

KPOPS-Earth와 ERA5 자료의 850-hPa 동서바람 공간장(그림 16)에서 나타나는 공통점은 저위도에서는 동풍 계열의 바람이, 중위도에서는 서풍 계열의 바람이 분다는 것이다. 이러한 동풍과 서풍은 북대서양, 북태평양 지역에서 그 강도가 강한데 이는 상대적으로 육지보다 해상에서 지표면의 마찰이 적기 때문이다. 동서바람장의 차이를 보면, 적도 지역은 동풍 아노말리, 중위도 지역은 서풍 아노말리가 있다. 즉 KPOPS-Earth는 저위도 무역풍과 중위도 편서풍을 더욱 강하게 모의하고 있다. 이러한 과대모의는 동아시아 지역에서 적게 나타났다.

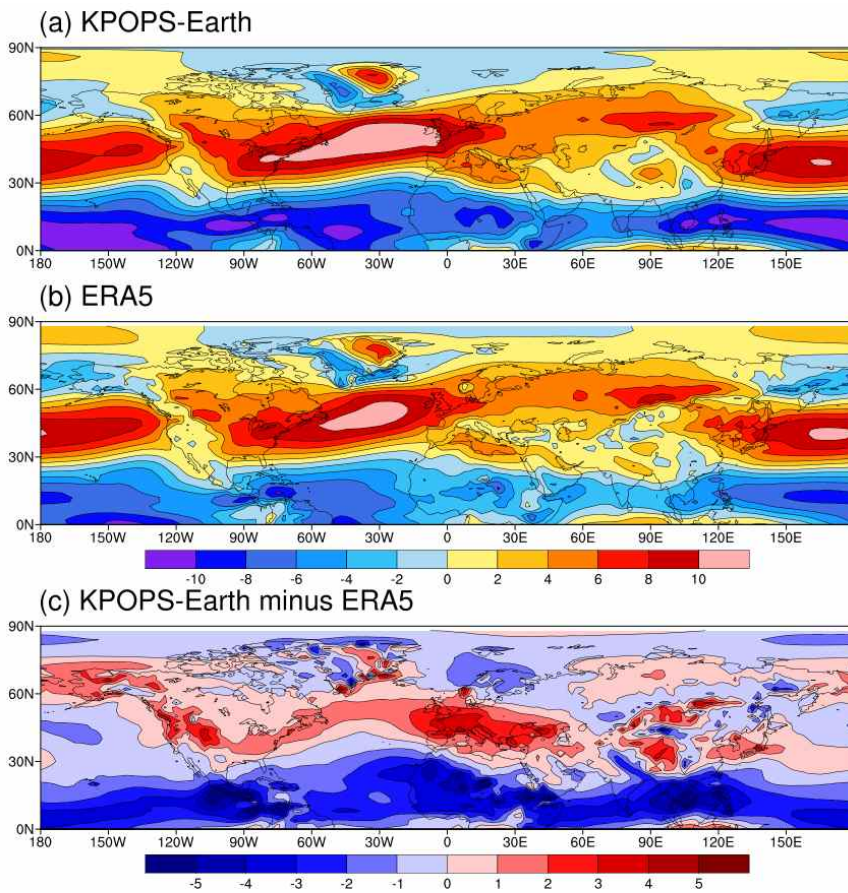


그림 14. (a) KPOPS-Earth 과거 모의 자료, (b) ERA5 재분석 자료의 2011년부터 2020년까지 겨울철 평균 850-hPa 동서바람장과 (c) 그 차이

KPOPS-Earth와 ERA5 자료의 850-hPa 남북바람의 공간장(그림 15)은 850-hPa 지위고도, 동서바람의 공간장보다 훨씬 복잡하다. 대체로 대륙의 서쪽에는 남풍이, 동쪽에는 북풍이 불고 있다. 이는 겨울철 북반구의 육지에서는 고기압이 발달하기 때문이다. 동아시아 지역에 초점을 맞추면 겨울철 시베리아 고기압이 발달하여 그 동편에 위치한 동아시아 지역에는 북풍이 분다. 이와 같은 특징이 KPOPS-Earth의 850-hPa 남북바람 공간장에서 잘 모의되고 있다. 하지만 두 자료의 차이를 보면, KPOPS-Earth는 동아시아 지역의 북풍을 과대모의하고 있다. 북풍이 강하면 차갑고 맑은 공기가 한반도에 유입될 수 있으므로 미세먼지 농도 예측 모델의 성능에 영향을 끼칠 수도 있다.

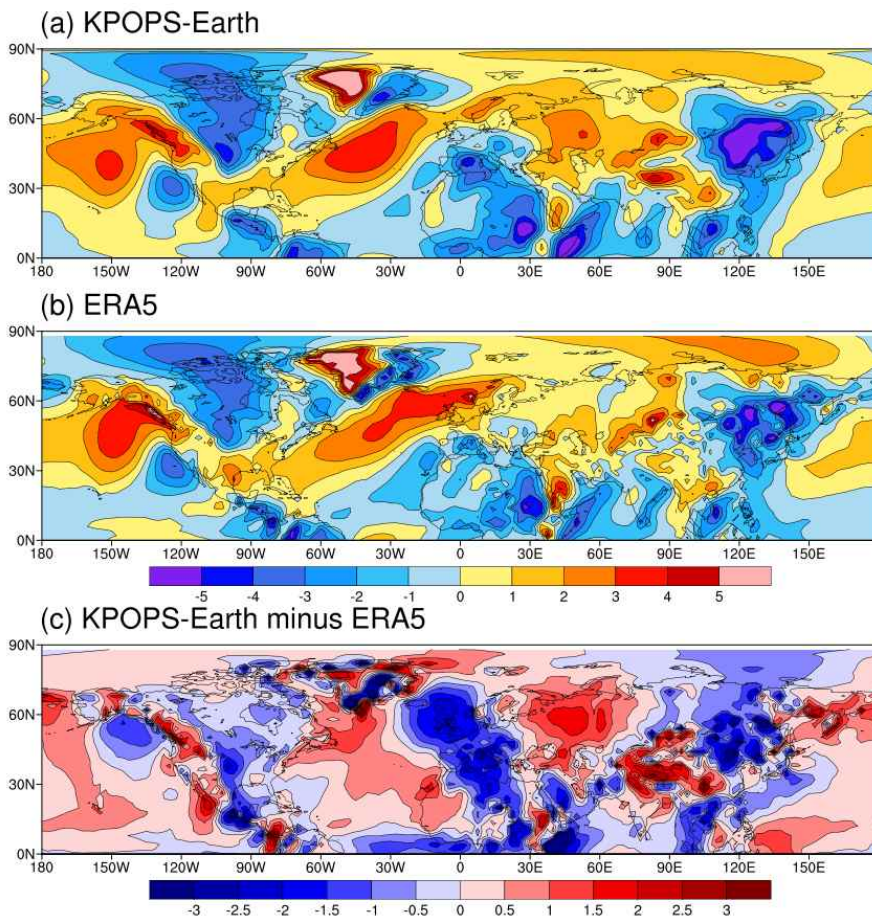
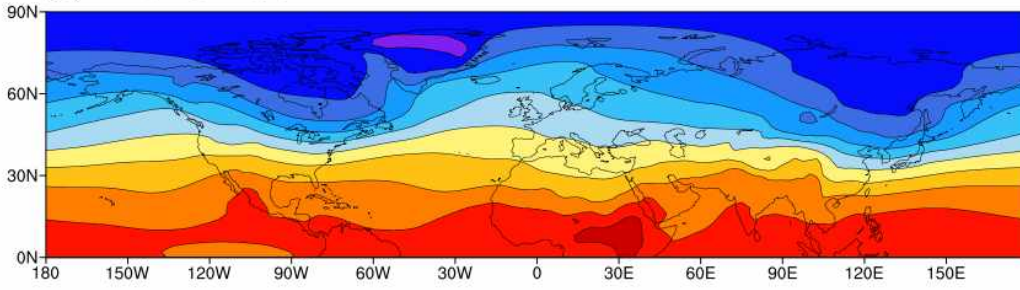


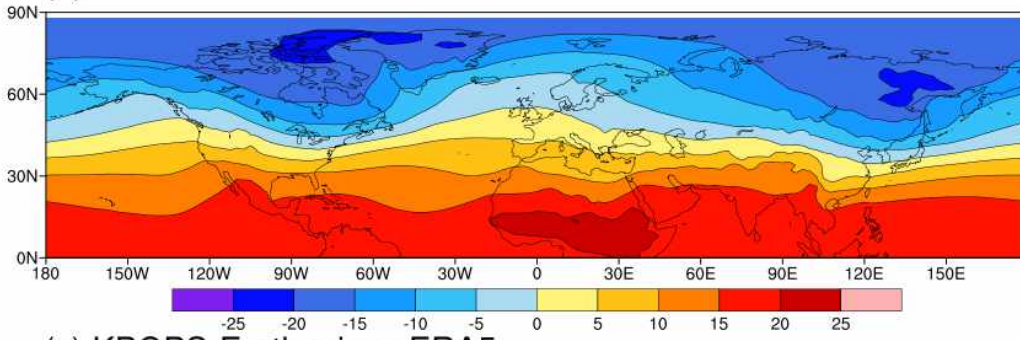
그림 15. (a) KPOPS-Earth 과거 모의 자료, (b) ERA5 재분석 자료의 2011년부터 2020년까지 겨울철 평균 850-hPa 남북바람장과 (c) 그 차이

그림 16은 KPOPS-Earth와 ERA5 자료의 850-hPa 공간장으로 저위도에서는 온도가 높고, 고위도에서는 온도가 높다. 같은 위도를 기준으로 비열이 낮은 대륙의 온도가 더 낮다. 두 자료의 공간장은 전반적으로 유사하지만 실제 차이를 보면 KPOPS-Earth가 온도를 과소모의 한다. 히말라야 산맥 지역, 그린란드 지역과 같이 해발고도가 높은 지역에서 과소모의 하는 경향이 심하다.

(a) KPOPS-Earth



(b) ERA5



(c) KPOPS-Earth minus ERA5

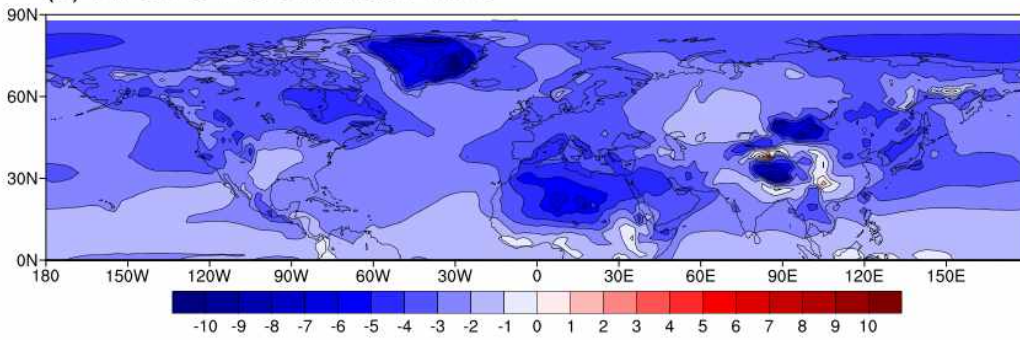


그림 16. (a) KPOPS-Earth 과거 모의 자료, (b) ERA5 재분석 자료의 2011년부터 2020년까지 겨울철 평균 850-hPa 온도장과 (c) 그 차이

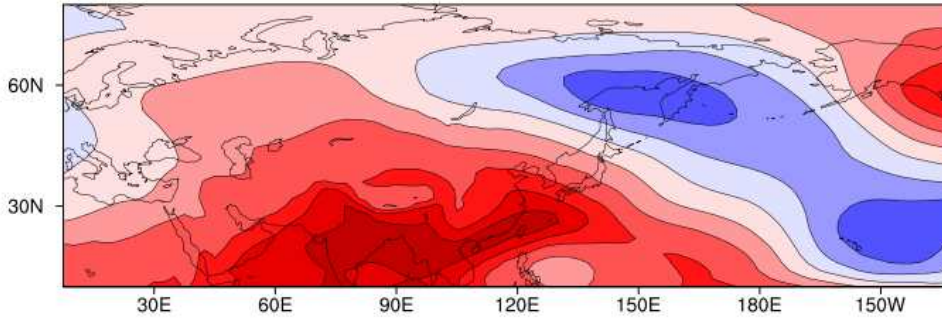


### 3. KPOPS-Earth 과거 모의 자료 기반 고위도 기후 인자와 한반도 미세먼지 간의 중장기 시간 규모 상관성 분석

머신러닝 기반 예측 모델 개발의 첫 단계로 어떤 변수를 입력 변수로 사용할지 선별해야 한다. Ho et al. (2021a)에서는 머신러닝 기반 미세먼지 농도 예측 모델을 개발할 때 선형 상관성이 높은 변수를 우선 모델의 입력 변수로 선별했다. 본 과제에서도 입력 변수 선별을 위해 월 평균 서울 미세먼지 농도와 다양한 대기 변수 간의 선형 상관계수의 공간장을 확인했다. 먼저 서울의 월평균 PM<sub>2.5</sub> 농도와 KPOPS-Earth 과거 모의 자료 간 상관계수의 공간장을 확인해보고 같은 분석을 ERA5 재분석 자료로 수행하여 두 결과를 비교했다. 서울 PM<sub>2.5</sub> 농도와 대기 변수 간 상관성이 어떻게 모의 되고 있는지 확인하고 KPOPS-Earth가 상관성을 관측과 얼마나 유사하게 모의하는지 확인하려고 한다.

KPOPS-Earth 자료에서 서울의 월평균 PM<sub>2.5</sub> 농도와 850-hPa 지위고도 간 상관계수의 공간장을 보면 인도부터 중국 남부 지역까지 양의 상관계수가 분포하고, 오호츠크 해와 캄차카 반도 지역에서는 음의 상관계수가 나타나며, 이 음의 아노말리는 태평양 지역까지 펼쳐져 있다(그림 17a). Hur et al. (2016)에서 따르면, 일 시간 규모에서 서울 PM<sub>2.5</sub> 농도가 높을 때는 양의 850-hPa 지위고도 아노말리가 한반도와 동해까지 영향을 끼친다. 월 시간 규모에서 이러한 특징은 나타나지 않으며, 한반도에서 멀리 떨어진 지역과의 상관성이 크게 나왔다. 즉 월 시간 규모에서는 한반도에서 멀리 떨어진 지역의 대기순환이 한반도에 영향을 끼쳐 한반도 미세먼지 농도가 변할 수 있다. ERA5 자료로 같은 분석을 한 결과(그림 17b), 인도부터 중국 남부까지의 양의 상관계수가 분포하는 것은 KPOPS-Earth 자료의 결과와 같았다. 하지만 음의 상관계수의 중심이 오호츠크 해와 캄차카반도 지역보다 좀 더 북쪽에 나타나고 중심의 강도가 더 강했다. KPOPS-Earth는 한반도 미세먼지 농도와 850-hPa 지위고도의 상관성을 관측과 유사하게 모의했다.

(a) KPOPS-Earth



(b) ERA5

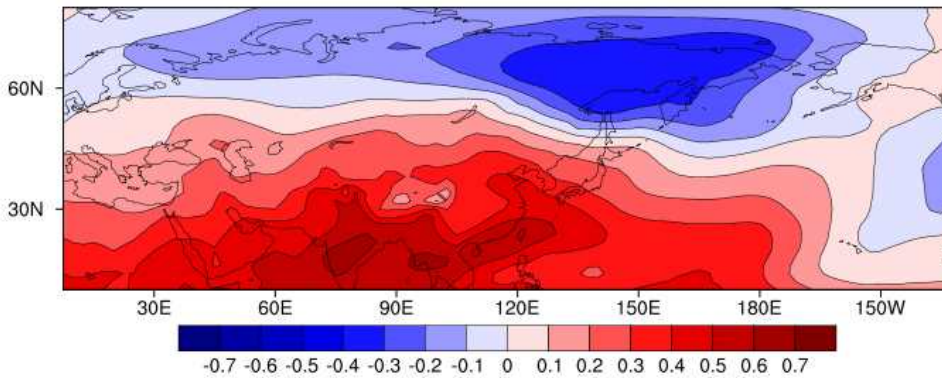
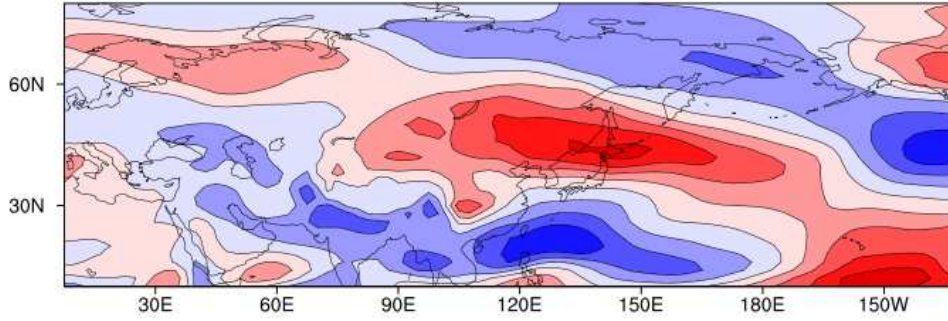


그림 17. 서울의 월평균  $PM_{2.5}$  농도와 (a) KPOPS-Earth, (b)ERA5 자료의 850-hPa 지위고도 간 상관계수의 공간 분포

일 시간 규모에서는 하층 풍속의 감소로 환기효과가 약해지거나, 중국 북부의 대기오염물질이 하층의 서풍 기류를 타고 한반도로 수송되면 고농도 미세먼지 사례가 발생한다고 알려져 있다(Oh et al. 2015). 하지만 KPOPS-Earth 자료에서 서울의 월평균  $PM_{2.5}$  농도와 850-hPa 동서바람 간 상관계수의 공간 분포를 보면(그림 18a) 한반도 남쪽에는 음의 상관계수, 한반도 북쪽에는 양의 상관계수, 더 먼 북쪽에는 다시 약한 음의 상관계수가 나타났다. 이를 해석하자면 월평균  $PM_{2.5}$  농도가 높을 때 한반도에서 상관계수가 거의 변하지 않으므로 동서바람의 약화로 미세먼지 농도가 증가하는 것을 설명하기 어렵다. 반대로 중국 북부와 한반도 지역 주변에 양의 상관계수가 나타나지 않아 하층의 편서풍이 더 강해진 것도 아니기 때문에 중국 북부 지역의 대기오염물질이 한반도로 더 쉽게 수송되었다고 보기도 어렵다. 시간 규모에서 따라 고농도 미세먼지 사례 발생 기작이 다를 수 있다. ERA5 자료의 결과에서도(그림 18b) 한반도 북쪽에는 양의 상관계수가, 남쪽에는 음의 상관계수가 나타나 KPOPS-Earth는 서울  $PM_{2.5}$  농도와 850-hPa 동서바람의 상관성을 잘 모의했다.

(a) KPOPS-Earth



(b) ERA5

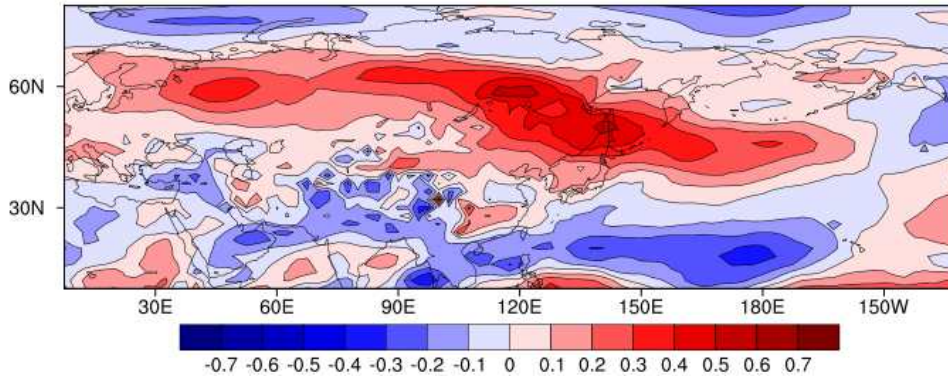
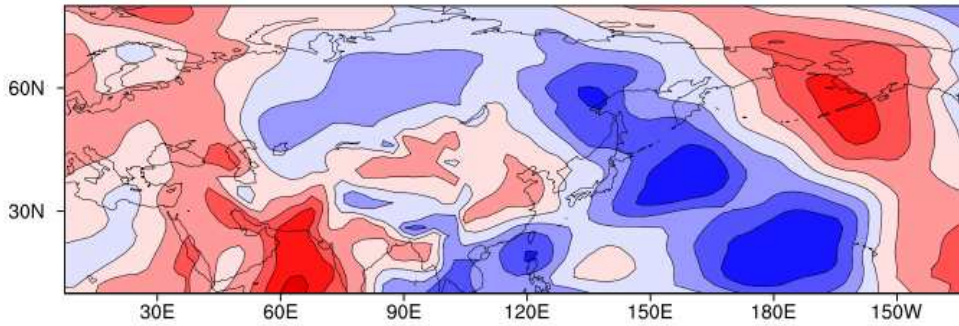


그림 18. 서울의 월평균  $PM_{2.5}$  농도와 (a) KPOPS-Earth, (b)ERA5 자료의 850-hPa 동서바람 간 상관계수의 공간 분포

KPOPS-Earth 자료에서 월평균 서울  $PM_{2.5}$  농도와 850-hPa 남북바람 간 상관계수의 공간 분포를 보면(그림 19a) 중국 북부 지역부터 서해상까지는 약한 양의 상관계수가 나타나고, 음의 상관계수 지역이 오호츠크 해부터 남동쪽으로 태평양까지 분포한다. 이보다 더 동쪽에는 양의 상관계수 지역이 알루시안 저기압 지역에서 남동쪽으로 펼쳐져 있다. ERA5 자료의 결과에서는(그림 19b) 중국 북부 지역에서 서해상으로 약한 양의 상관계수가 나타나는 것이 유사하다. 하지만 한반도 북동쪽에 음의 상관계수 지역이 있지만 남동쪽 방향으로 펼쳐져 있지 않으며, 알루시안 저기압 지역의 양의 상관계수 지역은 KPOPS-Earth 자료와 달리 남서쪽으로 펼쳐져 있다. KPOPS-Earth 모델은 월평균 한반도 미세먼지 농도와 850-hPa 남북바람 간 상관성을 관측과 다소 다르게 모의하는 것으로 보인다.

(a) KPOPS-Earth



(b) ERA5

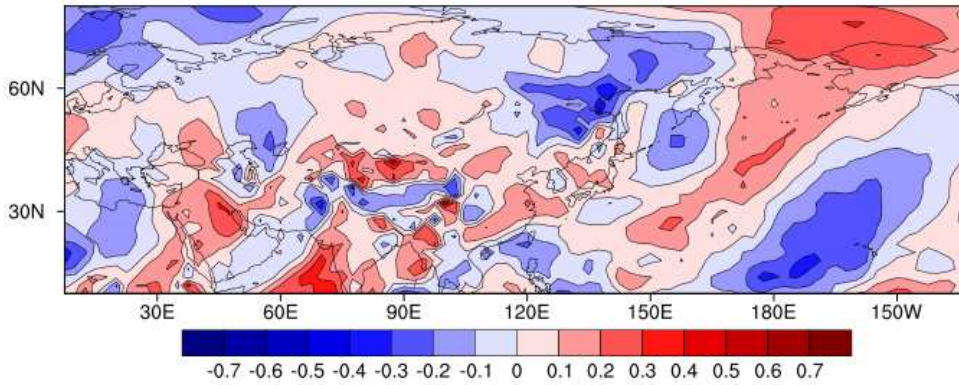
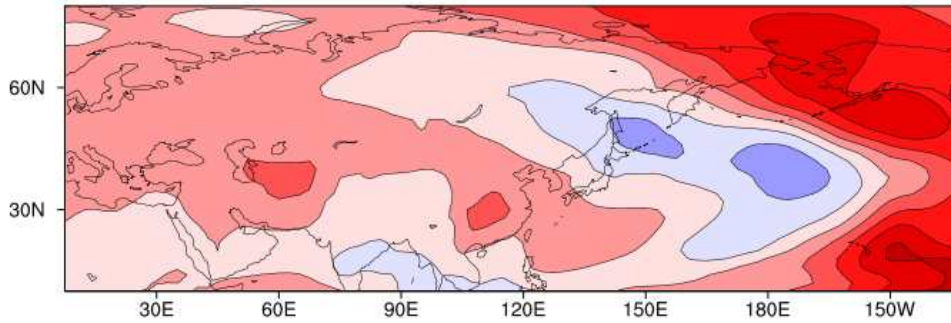


그림 19. 서울의 월평균  $PM_{2.5}$  농도와 (a) KPOPS-Earth, (b)ERA5 자료의 850-hPa 남북바람 간 상관계수의 공간 분포

일 시간 규모에서 한반도 미세먼지 농도가 높을 때 한반도를 중심으로 양의 850-hPa 아노말리가 나타난다. 하층의 따뜻한 공기는 한반도의 연직 안정도를 증가시키므로 수직 방향으로 환기효과가 약해져 미세먼지 농도가 축적되기 쉬우며, 지상의 온도가 증가해 2차 생성이 활발해져 미세먼지 생성이 많아진다고 알려져 있다. KPOPS-Earth 자료에서 서울의 월평균  $PM_{2.5}$  농도와 850-hPa 온도 간 상관계수의 공간 분포를 보면(그림 20a) 한반도 주위에 양의 상관계수가 나타나  $PM_{2.5}$  농도가 높을 때 온도가 높은 경향이 있다. 하지만 상관계수의 크기가 0.2보다 작아 강한 연관성은 없는 것으로 보인다. 양의 상관계수의 중심이 중국 남부 지역에 위치하는 것 또한 일 시간 규모의 결과와 다르다. 흥미롭게도 한반도에서 멀리 떨어진 척치해에 강한 양의 상관계수가 나타나 북극, 고위도의 온도가 한반도 미세먼지 농도와 연관이 있을 수 있다. ERA5 자료의 결과에서는(그림 20b) 중국 남부와 한반도 주변 지역에 양의 상관계수가 나타나며, 척치해 주변에서 양의 상관계수가 나타난 것이 KPOPS-Earth의 결과와 유사하며, 다른 점으로는 유럽의 온도와 서울  $PM_{2.5}$  농도 간 양의 연관성을 좀 더 강하게 모의 되었다. 대체적으로 KPOPS-Earth 모델은 미세먼지 농도와 850-hPa 온도의 관계를 잘 모의하는 것으로 보인다.

(a) KPOPS-Earth



(b) ERA5

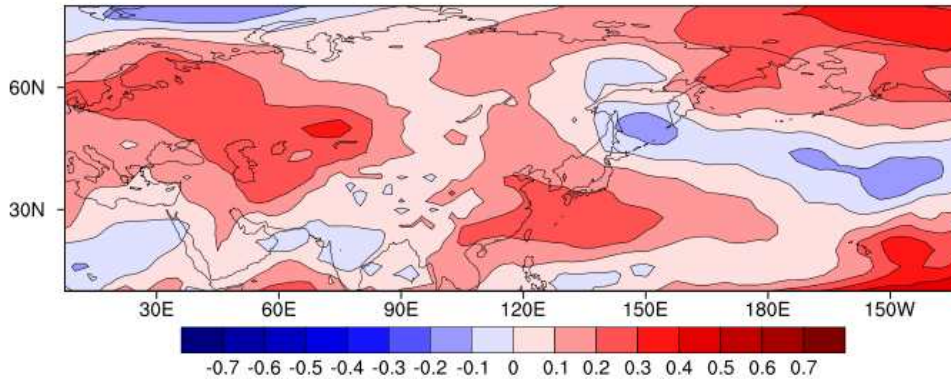


그림 20. 서울의 월평균 PM<sub>2.5</sub> 농도와 (a) KPOPS-Earth, (b)ERA5 자료의 850-hPa 온도 간 상관계수의 공간 분포

Kim et al. (2019)에서 1월 평균 PM<sub>10</sub> 농도가 이전 기간의 북극 해빙 면적 간 상관성이 있다는 것을 보여주지만, 1월이 아닌 달의 평균 PM<sub>10</sub> 농도와 해빙 면적 간 상관성은 적었다. 즉 어떤 달이냐에 따라 한반도 미세먼지 농도와 연관성이 큰 지역이 다를 수 있고, 연관성은 크더라도 상관계수의 부호가 다를 수 있다. 지금까지 상관성 분석을 수행한 기간은 1, 2, 11, 12월로 월별로 평균 미세먼지 농도와 대기 변수 상관관계를 확인할 필요가 있다. KPOPS-Earth 자료에서 1월, 2월, 11월 12월에 대해 월평균 서울 PM<sub>2.5</sub> 농도와 850-hPa 지위고도 간 상관계수의 공간 분포를 확인해보았다(그림 21). 1월에 유라시아 대륙에서는 약 위도 45도를 경계로 남쪽에는 양의 상관계수, 북쪽에는 음의 상관계수가 나타났다. 2월의 경우 유라시아 대륙 남쪽에서 양의 상관계수가 나타나지만 유라시아 대륙 북서쪽에서는 음의 상관계수가, 북동쪽에는 양의 상관계수가 나타났다. 또한, 알루시안 저기압 지역부터 남동쪽으로 음의 상관계수 지역이 펼쳐져 있어 1월의 결과와는 달랐다. 11월을 보면 음의 상관계수 지역이 시베리아 지역부터 동쪽으로 오호츠크 해, 태평양까지 펼쳐져 있으며 고위도 지역에서는 양의 상관계수 지역만이 나타났다. 12월에는 11월의 음의 상관계수 지역이 좀 더 북쪽에 위치하며 인도부터 중국 남부 지역까지 강한 양의 상관계수 지역이 보인다. 어느 기간이든 한반도 남서쪽의 양의 상관계수 지역이 나타나는 것만 같을뿐 어떤 달이냐에 따라 상관계수의 공간 분포가 크게 달랐다. 상관성이 높은 지역을 단순히 평균해서 예측 모델의 입력 자료로 사용하기에는 한계가 있다.

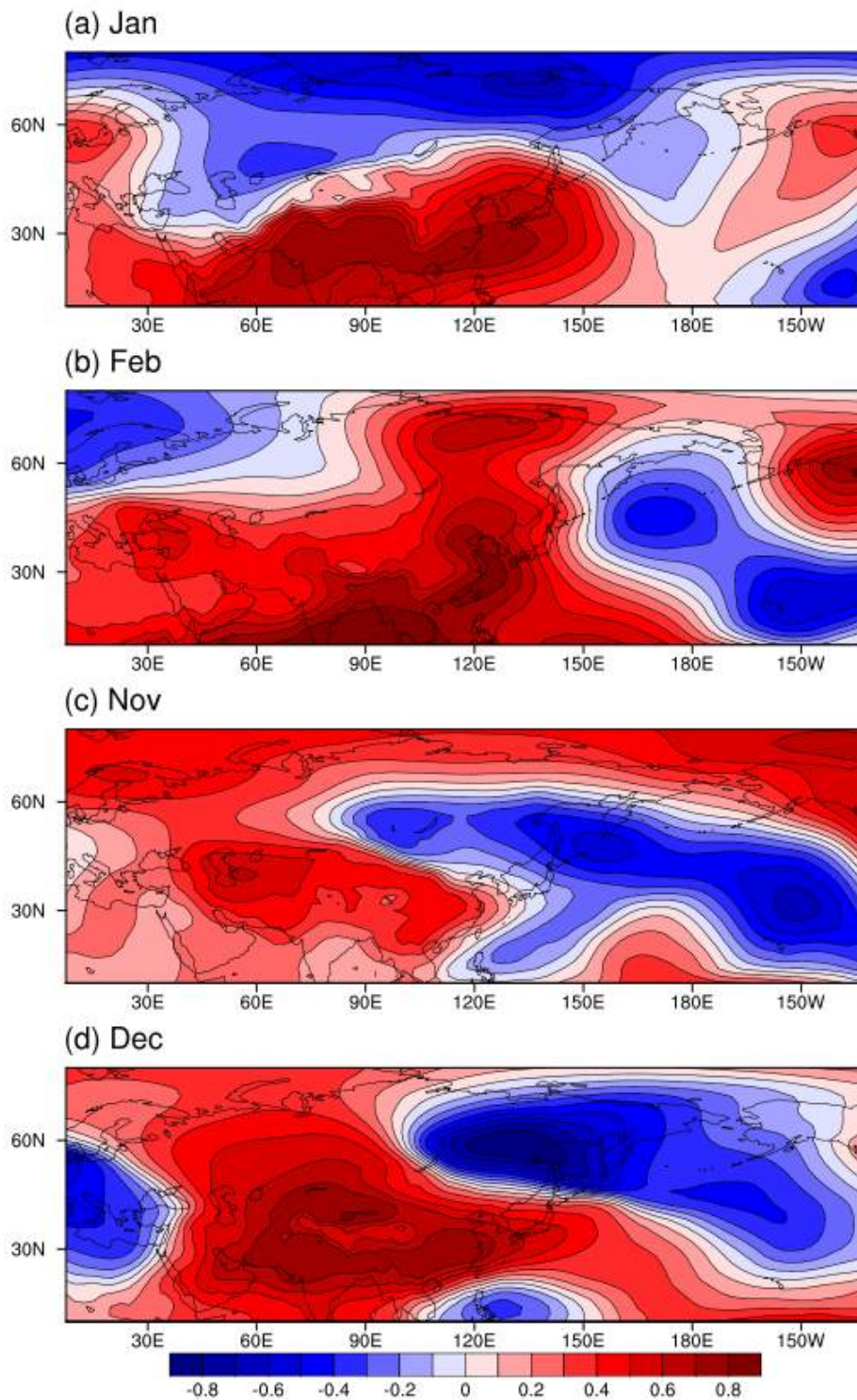


그림 21. KPOPS-Earth 자료에서 (a) 1월, (b) 2월, (c) 11월, (d) 12월 평균 서울의  $PM_{2.5}$  농도와 (a) KPOPS-Earth, (b) ERA5 자료의 850-hPa 지위고도 간 상관계수의 공간 분포

### 제 3 절 머신러닝 기법을 활용한 미세먼지 농도 예측 모델 개발

#### 1. 랜덤 포레스트 모델

랜덤 포레스트 기법은 머신러닝 기법 중에서도 기초적인 방법으로 대기과학 분야에서는 대기오염물질의 농도나 대기 변수를 예측하는데 많이 쓰인다. 랜덤 포레스트 모델은 여러 개의 디시전 트리 모델로 구성되어 있다. 디시전 트리 기법은 입력 변수를 받은 후 단계적으로 다양한 조건문을 설정하고 그 조건에 따라 예측 변수의 결과값을 산출하는 방법이다. 단순히 선형적인 방법으로는 어느 입력 변수가 중요하다고 판단하기 어려워서 랜덤 포레스트 기반 PM<sub>10</sub> 농도 예측 모델을 개발하여 입력 변수의 중요도를 파악하고자 했다. 이 분석의 목적은 중요한 입력 변수를 파악한 뒤 이 변수들을 좀 더 진보한 머신러닝 기법인 순환신경망 혹은 합성곱신경망 기반 한반도 미세먼지 농도 예측 모델의 입력 자료로 활용하는 것이다.

랜덤 포레스트 모델의 입력 자료 선정을 위해 2011년부터 2020년까지 1, 2, 11, 12월에 대해 월평균 서울 PM<sub>10</sub> 농도와 ERA5 자료의 850-hPa 지위고도, 온도, 동서바람, 남북바람, 습도 간 상관계수의 공간 분포를 확인하였다(그림 22). 각각 자료들의 격자값을 입력 자료로 사용할 수 있지만, 이 경우 계산량이 너무 많기 때문에 상관성을 보이는 지역의 변수값을 입력 자료로 사용한다. 850-hPa 지위고도, 온도, 동서바람, 남북바람, 습도 순으로 유의한 상관성을 보이는 지역은 3, 2, 4, 3, 5개의 총 17개 지역으로 각 지역의 평균값을 랜덤 포레스트 모델의 입력 변수로 선정했다. 랜덤 포레스트 모델은 이 입력 변수를 이용해 월평균 서울 PM<sub>10</sub> 농도를 예측하도록 학습했다. 또한, 랜덤 포레스트 모델의 학습 초기조건에 변화를 주어 총 40개의 앙상블 멤버를 만들었다. 랜덤 포레스트 예측 성능 확인을 위해 랜덤 포레스트에서 예측한 PM<sub>10</sub> 농도와 관측 PM<sub>10</sub> 농도의 결정계수를 그림 23에 나타냈다. 이때 40개의 앙상블 멤버 중 14개의 모델은 모든 월평균 PM<sub>10</sub> 농도를 평균값으로 예측하는 것보다 성능이 좋지 않아 분석에서 제외했다. 나머지 26개의 모델에서 결정계수의 중위값은 0.25 정도로 가장 성능이 좋은 모델의 결정계수는 약 0.5이다.

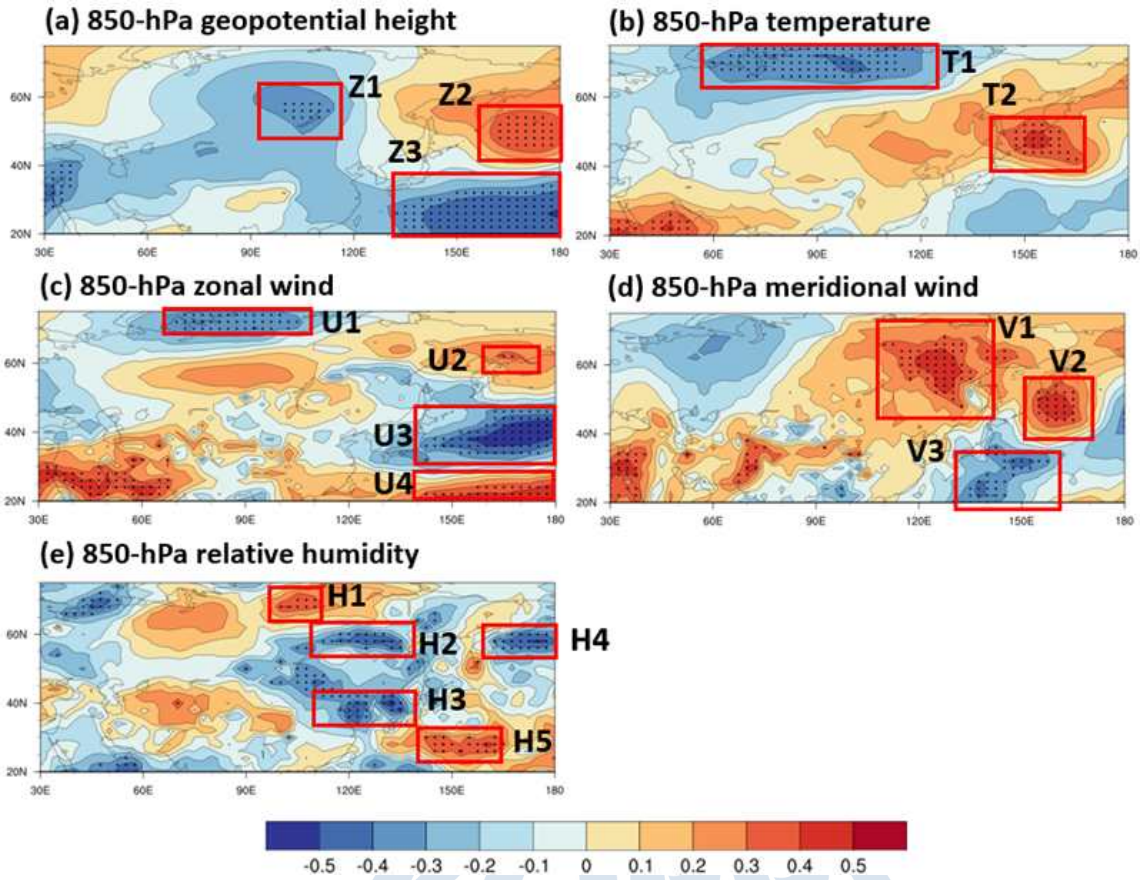


그림 22. 월평균 서울  $PM_{10}$  농도와 ERA5 재분석 자료의 (a) 850-hPa 지위고도, (b) 850-hPa 온도, (c) 850-hPa 동서바람, (d) 850-hPa 남북바람, (e) 850-hPa 습도 간 상관계수의 공간장 (점은 95% 신뢰수준에서 유의함). 빨간 사각형 안 유의한 지역의 평균값은 랜덤 포레스트 기반  $PM_{10}$  농도 예측 모델의 입력 자료로 사용됨.

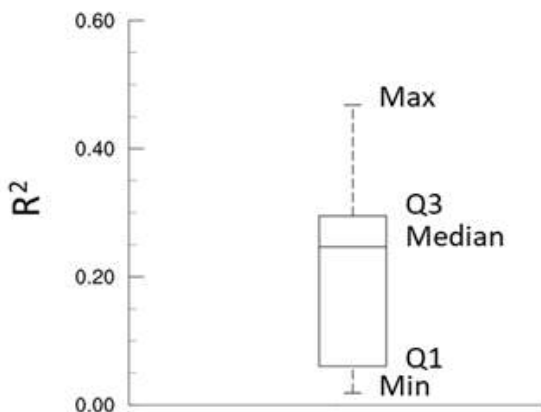


그림 23. 랜덤 포레스트 기반  $PM_{10}$  농도 예측 모델(26개)의 결정계수 분포.



예측 성능이 낮은 모델의 입력 변수별 중요도는 큰 의미가 없어서 결정계수 기준 상위 25% (결정계수 > 0.3)의 7개 모델로 입력 변수의 중요도를 그림 24로 나타내었다. 중요도가 가장 높은 변수는 Z2로 그 값이 7% 정도 되고 H3까지는 중요도가 6% 이상이다. 하지만 H3 다음인 Z1부터는 중요도가 크게 감소한다. 대체로 한반도에 가까울수록 중요도가 높게 나타났다. 하지만 어느 한 입력 변수가 압도적으로 중요하진 않아서 특정 변수가 모델의 예측 성능을 크게 좌우한다고 보기 어렵다. 특정 지역의 대기 변수를 평균해 모델의 입력 자료로 사용하기보다 넓은 지역의 공간 특성을 모두 고려할 수 있는 합성곱신경망 기법을 사용하여 모델을 구축하는 합당한 것으로 보인다.

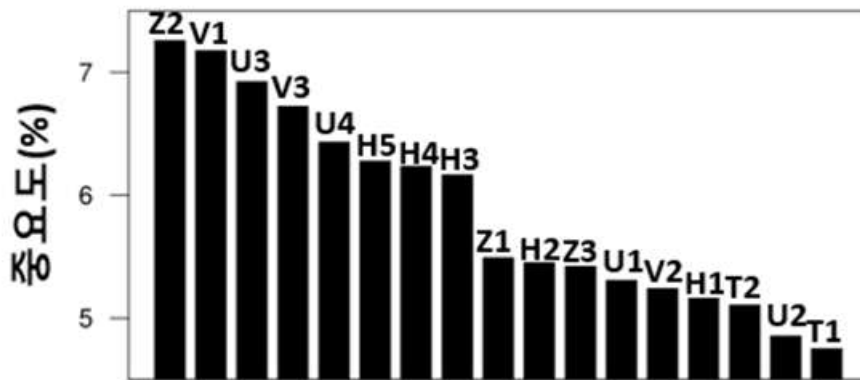


그림 24. 상위 25%( $R^2 > 0.3$ )의 성능을 보이는 랜덤 포레스트 기반 PM10 농도 예측 모델의 각 입력 변수별 중요도.

## 2. 합성곱신경망 모델

본 과제의 최종목표인 중장기 미세먼지 농도 예측 모델 프로토타입의 개발을 위해 사용한 머신러닝 기법은 합성곱신경망(CNN, convolutional neural network)이다. 예측하고자 한 변수는 서울의 월평균  $PM_{2.5}$  농도이며 입력 자료로는 ERA5 재분석 자료와 KPOPS-Earth 과거 모의 자료가 쓰였다. ERA5 자료는 모델의 학습 과정과 예측에 모두 사용되었으며, KPOPS-Earth 자료는 예측에만 사용되었다. 학습 과정에서 ERA5 자료만 사용한 이유는 관측 기반으로 학습이 되어야 나중에 KPOPS-Earth 모델이 업데이트되었을 때 새로 학습을 하지 않더라도 예측이 가능하기 때문이다.

CNN 모델 개발에 사용되는 자료는 가장 먼저 학습, 검증, 평가 기간에 따라 분류된다. 학습 기간에는 모델의 예측 오차를 줄이는 방향으로 weighting parameter가 조정되고, 검증 기간에는 모델의 오차를 줄이는 방향으로 hyper parameter가 조정된다. 이후 평가 기간에는 학습, 검증과정을 거친 모델로 예측을 수행하여 모델의 성능을 평가한다.

모델의 학습 과정에는 2011년부터 2020년까지 1, 2, 11, 12월 자료를 사용했다. 학습, 검증, 평가 기간에 따른 자료의 분류에는 K-fold와 bagging 앙상블 방법을 사용했다. 월평균 자료를 이용하여 모델을 구축하기 때문에 자료의 수가 적은 문제가 있는데 K-fold, bagging 앙상블 방법을 사용하면 자료의 수를 늘릴 수 있다. K-fold 방법은 잭나이프 방법과 유사한 방법으로 이를 이용하면 평가 기간을 연별로 달리하여 10개의 앙상블 멤버가 생긴다. Bagging 앙상블

방법은 평가 기간 외 나머지 기간의 자료를 8:2의 비율로 학습 자료, 검증 자료로 랜덤 샘플링 하는 것으로 이 방법을 이용해 총 9개의 앙상블 멤버를 만들었다. K-fold 방법과 Bagging 앙상블 방법을 이용한 자료 분류의 모식도는 그림 25와 같다.



그림 25. K-fold와 bagging 앙상블 방법을 이용한 자료 분류. 여기서 K는 10, N은 9에 해당

학습 과정에서 CNN 모델의 입력 자료는 9개의 기상 변수의 아노말리가 이용되며 이때 9개의 기상 변수 아노말리에는 각각 격자별로 월평균 서울  $PM_{2.5}$  농도 아노말리와 ERA5 자료의 기상 변수 아노말리 간 상관계수의 세제곱이 곱해지는 작업이 수행된다. 비록 상관계수는 선형적인 상관성을 나타내는 값이지만 이처럼 입력 자료에 인위적으로 개입하는 것이 모델 예측 성능에 긍정적인 영향을 끼칠 수 있다. 또한, 입력 자료인 기상 변수의 아노말리는 -1에서 1 사이, 예측 변수인  $PM_{2.5}$  농도는 0에서 1 사이의 값을 갖도록 정규화했다. 그림 26은 CNN의 기본 구조를 모식도로 나타낸 것으로 입력 자료는 처음에 3개의 convolution, max pooling layer를 거친 뒤 fully connected layer, absolute function에 의해 마지막으로 출력값이 나온다. 이 출력값은 입력 자료별로 계산되는 값이며 이들의 합이 최종 출력값으로 학습 과정에서 손실 함수 계산에 사용되었다.

CNN 모델의 예측 성능을 평가에는 ERA5 자료로 서울의 월평균  $PM_{2.5}$  농도를 예측한 모델과 KPOPS-Earth로 농도를 예측한 모델을 사용했다. 두 모델 모두 학습 과정에서는 ERA5 자료가 사용되었다. RMSE, 상관계수(R), 정확도(ACC), 고농도 감지 확률(POD), 오경보율(FAR)을 비교하는 방식으로 모델의 예측 성능을 평가했다. ACC, POD, FAR는 등급 기준이 있어야 계산할 수 있다. AirKorea에서는 일평균  $PM_{2.5}$  농도의 등급 기준을 정하였지만, 월평균  $PM_{2.5}$  농도의 등급 기준을 별도로 정의하지 않는다. 모델의 예측 성능 평가를 위해 일평균 농도의 등급 기준에 대한 퍼센타일을 구하여 월평균  $PM_{2.5}$  농도의 같은 퍼센타일 값으로 등급 기준을 정했다. 일평균  $PM_{2.5}$  농도와 월평균  $PM_{2.5}$  농도의 등급 기준은 표 4과 같다.

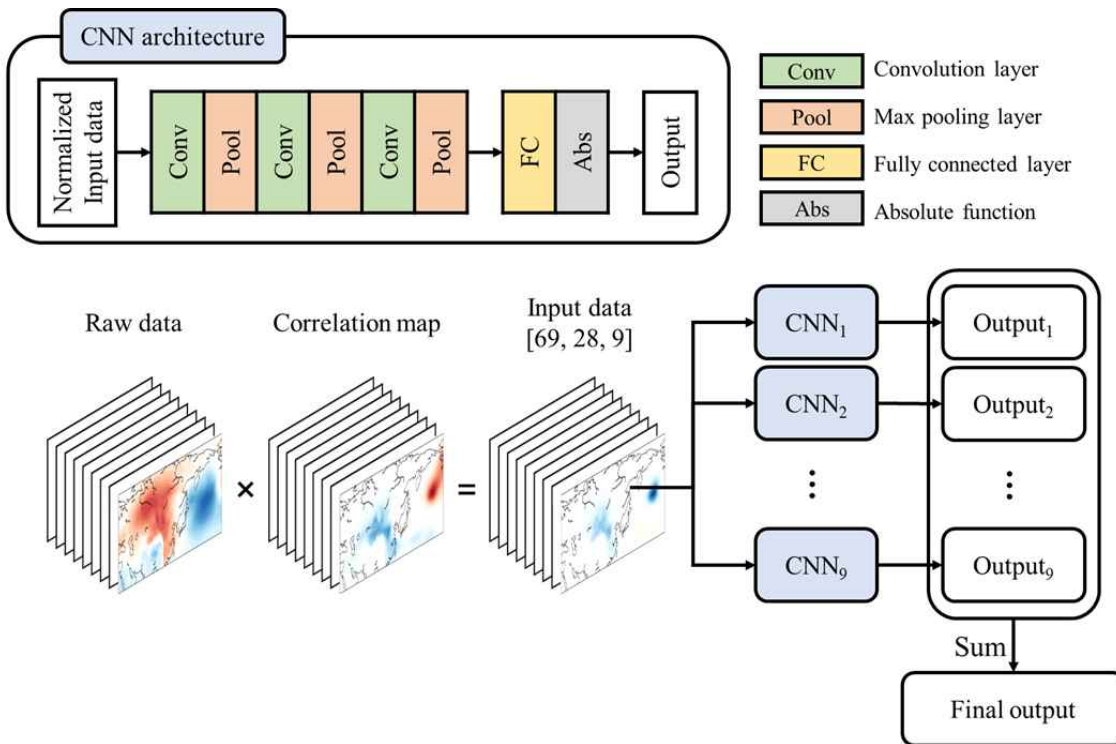


그림 26. CNN 모델의 모식도. Raw\_data는 ERA5 재분석장 혹은 KPOPS 예측장, Correlation map은 ERA5 재분석장과 서울의 관측 PM<sub>2.5</sub> 농도와 상관계수장으로 사용된 변수는 총 9개

표 4. 각 등급별 일 평균 PM<sub>2.5</sub> 농도 기준, 이에 대한 퍼센타일과 월 평균 PM<sub>2.5</sub> 농도 기준 (\*는 일의 자리에서 반올림 했음을 의미)

Level	Daily PM <sub>2.5</sub> conc. ( $\mu\text{g m}^{-3}$ )	Percentile of daily PM <sub>2.5</sub> conc.	Monthly PM <sub>2.5</sub> conc. ( $\mu\text{g m}^{-3}$ )
Low	PM <sub>2.5</sub> <16	PM <sub>2.5</sub> <20 <sup>th</sup>	PM <sub>2.5</sub> <20*
Moderate	16≤PM <sub>2.5</sub> <36	20 <sup>th</sup> ≤PM <sub>2.5</sub> <77 <sup>th</sup>	20*≤PM <sub>2.5</sub> <30*
High	36≤PM <sub>2.5</sub> <76	77 <sup>th</sup> <PM <sub>2.5</sub> <99 <sup>th</sup>	30*≤2PM <sub>2.5</sub> <40*
Very high	76≤PM <sub>2.5</sub>	99 <sup>th</sup> ≤PM <sub>2.5</sub>	40*≤PM <sub>2.5</sub>

CNN 모델의 입력 자료를 처리하는 작업에서 PM<sub>2.5</sub> 농도와 대기 변수 간 상관계수의 세제곱을 곱하는 과정이 있었다. 이는 다소 인위적인 과정으로 아무것도 곱하지 않은 경우(\_R0), 상관계수의 1승(\_R1), 2승(\_R2), 3승(\_R3)을 곱한 네 가지 경우 CNN 모델의 성능을 비교했다(표 5). ERA5 자료를 사용한 예측 모델에서는 ERA5\_R3 모델이 가장 우수한 성능을 보여주었다(RMSE, R, ACC, POD, FAR은 각각  $3.37 \mu\text{g m}^{-3}$ , 0.73, 80.0%, 66.7%, 20.0%). KPOPS-Earth 자료를 사용한 예측 모델에서는 RMSE, R, ACC가 가장 우수한 모델은 KPOPS\_R2 모델이지만 고농도를 맞춘 경우는 단 한 사례로 POD가 8.3%에 불과하다. KPOPS\_R3 모델은 RMSE, ACC가 다소 낮지만, POD가 가장 높았다. 예측을 ERA5 자료로 하든 KPOPS-Earth 자료로 하든 세제곱을 곱했을 때 가장 좋은 성능을 보여주므로 입력 자료에 세제곱을 곱하여 PM<sub>2.5</sub>

농도를 예측하는 것이 가장 합리적이다.

표 5. CNN 모델의 예측 성능(R0은 상관계수를 곱하지 않은 것, R1, R2, R3은 각각 상관계수의 1승, 2승, 3승을 의미)

CNN Model	RMSE ( $\mu\text{g m}^{-3}$ )	R	ACC (%)	POD (%)	FAR (%)
ERA_R0	4.42	0.53	65.0	0.0	100.0
ERA_R1	4.72	0.60	67.5	0.0	0.0
ERA_R2	4.51	0.50	65.0	8.3	66.7
ERA_R3	3.37	0.73	80.0	66.7	20.0
KPOPS_R0	4.39	0.47	69.6	0.0	100.0
KPOPS_R1	4.73	0.50	69.6	0.0	0.0
KPOPS_R2	4.29	0.50	71.7	8.3	50.0
KPOPS_R3	4.44	0.46	63.0	41.7	58.3

그림 27은 서울의 관측 PM2.5 농도, CNN-ERA5 모델의 예측 PM2.5 농도와 CNN-KPOPS 모델의 예측 PM2.5 농도 시계열, 그림 28은 같은 결과를 산포도로 나타낸 것이다. CNN-ERA5 모델의 예측 성능은 CNN-KPOPS 모델보다 대체로 우수한다. 이는 CNN-KPOPS가 고농도 구간에서 과소 모의를 하기 때문으로 보인다. 이번 과제에서 개발한 CNN 모델은 프로토 타입이므로 추후 연구를 진행하여 성능을 개선할 필요가 있다.

극지연구소

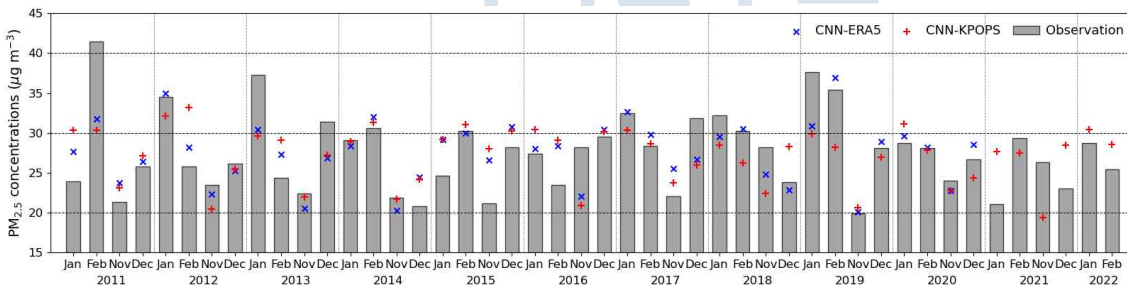


그림 27. 관측 및 예측의 월 평균 PM2.5 농도. 회색은 관측 PM2.5 농도, 파란색 ×는 CNN-ERA5 모델의 예측, 빨간색 +은 CNN-KPOPS의 예측이다.

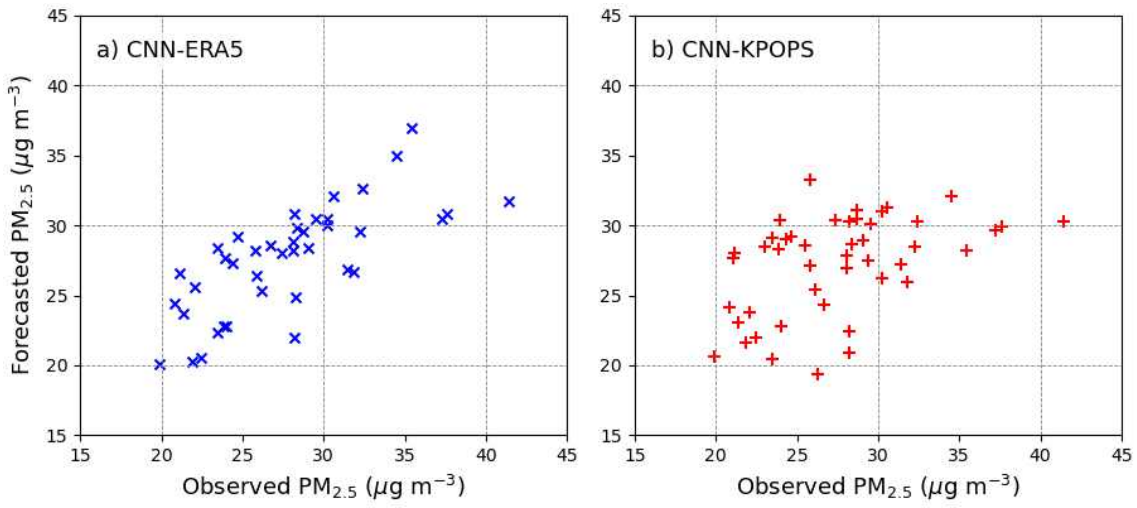


그림 28. CNN-ERA5, KPOPS 모델의 예측 농도와 관측 PM<sub>2.5</sub> 농도의 산포도.



## 제 4 장 연구개발목표 달성도

### 제 1 절 연구목표 달성도

#### 1. 연구목표 및 연구 내용

본 연구개발의 최종목표는 ‘한반도 미세먼지 농도의 중장기 예측을 위한 기후 인자 선정, 확보 및 머신러닝 기반 예측 모델 프로토타입 구축이다. 연차별 연구개발 목표 및 내용은 표 6와 같다.

표 6. 본 연구개발의 연차별 연구개발 목표 및 내용

연차	연구개발 목표	연구개발 내용
1차년도 (2020)	관측/재분석 자료를 분석해서 한반도 미세먼지 농도에 영향을 끼치는 고위도 기후 인자를 분석	<ul style="list-style-type: none"> <li>- 한국, 중국 관측소에서 측정한 미세먼지 농도 자료를 수집하고 DB 정리</li> <li>- 미세먼지 예측에 필요한 북극 고위도 기후 인자 DB 정리</li> <li>- DB 구축한 북극 기후 인자와 한반도 미세먼지 간의 중장기 시간 규모에서의 상관성을 분석</li> </ul>
2차년도 (2021)	기후모델을 이용해서 한반도 미세먼지 농도에 영향을 끼치는 고위도 기후 인자를 평가	<ul style="list-style-type: none"> <li>- 기후모델링 시스템의 과거 모의 자료에서 고위도 기후 인자 추출하여 DB 확대</li> <li>- 과거 모의자료 기반 고위도 기후 인자와 한반도 미세먼지 간의 중장기 시간 규모 상관성을 분석</li> </ul>
3차년도 (2022)	추출된 개별 고위도 예측 인자의 잠재적 중요도 판단 및 미세먼지 예측에 머신러닝 기법 활용	<ul style="list-style-type: none"> <li>- Random forest 기법 등을 이용하여 각 잠재적 예측 인자의 미세먼지 예측에의 중요도를 분석</li> <li>- 신경망 모델을 이용하여 북극 및 고위도 기후 기반 시험적 중장기 미세먼지 예측 시도</li> </ul>

## 2. 연구개발 성과

연차별 연구개발목표, 정성적 성과와 달성도는 표 5와 같다.

표 7. 본 연구개발의 정성적 성과

연차	연구개발 목표	주요 연구개발 실적	달성도
1차년도 (2020)	관측/재분석 자료를 분석해서 한반도 미세먼지 농도에 영향을 끼치는 고위도 기후 인자를 분석	<ul style="list-style-type: none"> <li>- 한국, 중국 관측소에서 측정한 미세먼지 농도 자료를 자동으로 수집하는 시스템 개발 및 데이터베이스로 구축 완료</li> <li>- 미세먼지 예측에 필요한 북극, 고위도 지역의 ERA5 재분석 자료와 북극, 고위도 지역의 각종 기후 인자의 인덱스를 데이터베이스로 정리 완료</li> <li>- 데이터베이스로 구축한 북극, 고위도 기후 자료와 한반도 미세먼지 간의 중장기 시간 규모에서 상관성을 확인</li> </ul>	100%
2차년도 (2021)	기후모델을 이용해서 한반도 미세먼지 농도에 영향을 끼치는 고위도 기후 인자를 평가	<ul style="list-style-type: none"> <li>- 기후모델링 시스템으로 생산한 과거 모의 자료에서 고위도 기후 인자 추출하여 데이터베이스 확대</li> <li>- 과거 모의 자료 기반 고위도 기후 인자와 한반도 미세먼지 간의 중장기 시간 규모 상관성을 확인</li> </ul>	100%
3차년도 (2022)	추출된 개별 고위도 예측 인자의 잠재적 중요도 판단 및 미세먼지 예측에 머신러닝 기법 활용	<ul style="list-style-type: none"> <li>- Random forest 기법을 이용하여 서울의 월평균 PM<sub>2.5</sub> 농도 예측 모델 개발 및 각 잠재적 예측 인자의 중요도를 확인</li> <li>- 합성곱신경망 기반으로 북극 및 고위도 기후 자료를 입력 자료로 활용하여 서울의 월평균 PM<sub>2.5</sub> 농도 예측 모델의 프로토타입을 개발 완료</li> </ul>	100%

## 제 5 장 연구개발결과의 활용계획

- KPOPS-Earth 모델로 미래 기후를 모의하여 실제 중장기 서울  $PM_{2.5}$  농도를 예측해 현업 예보를 제공
- 머신러닝 기법을 사용하는 예측 모델의 성능을 향상시킬 수 있는 다양한 방법으로 모델 예측 성능 개선과 이를 중요한 연구결과로 제공
- 중장기 시간 규모의 머신러닝 기반 미세먼지 농도 예측 모델 개발 분야에서 국제적으로 선도하는 지위로 도약
- 국가 차원에서 미세먼지 피해 예방책과 대응책 수립의 판단 지표 제공
- 미세먼지 관련 산업의 의사결정을 위한 가이드라인 제공
- 미래 기후 변화가 한반도 미세먼지 농도에 끼치는 영향을 정량적으로 제시
- 미세먼지 예측이 필요한 다양한 학문 분야에서 공동 연구 계기 마련





## 제 6 장 참고문헌

- Kim, H. S., Park, I., Song, C. H., Lee, K., Yun, J. W., Kim, H. K., ... & Han, K. M. (2019). Development of a daily PM 10 and PM 2.5 prediction system using a deep long short-term memory neural network model. *Atmospheric Chemistry and Physics*, 19(20), 12935-12951.
- Kim, B. Y., Lim, Y. K., & Cha, J. W. (2022). Short-term prediction of particulate matter (PM10 and PM2.5) in Seoul, South Korea using tree-based machine learning algorithms. *Atmospheric Pollution Research*, 13(10), 101547.
- Yang, G., Lee, H., & Lee, G. (2020). A hybrid deep learning model to forecast particulate matter concentration levels in Seoul, South Korea. *Atmosphere*, 11(4), 348.
- Park, S., Kim, M., Kim, M., Namgung, H. G., Kim, K. T., Cho, K. H., & Kwon, S. B. (2018). Predicting PM10 concentration in Seoul metropolitan subway stations using artificial neural network (ANN). *Journal of hazardous materials*, 341, 75-82.
- Sayeed, A., Lops, Y., Choi, Y., Jung, J., & Salman, A. K. (2021). Bias correcting and extending the PM forecast by CMAQ up to 7 days using deep convolutional neural networks. *Atmospheric Environment*, 253, 118376.
- Zhao, J., Deng, F., Cai, Y., & Chen, J. (2019). Long short-term memory-Fully connected (LSTM-FC) neural network for PM2.5 concentration prediction. *Chemosphere*, 220, 486-492.
- Doreswamy, Harishkumar, K. S., Yogesh, K. M., & Gad, I. (2020). Forecasting air pollution particulate matter (PM2.5) using machine learning regression models. *Procedia Computer Science*, 171, 2057-2066.
- Li, X., Peng, L., Yao, X., Cui, S., Hu, Y., You, C., & Chi, T. (2017). Long short-term memory neural network for air pollutant concentration predictions: Method development and evaluation. *Environmental pollution*, 231, 997-1004.
- Hui-Jun, W., Huo-Po, C., & Jiping, L. (2015). Arctic sea ice decline intensified haze pollution in eastern China. *Atmospheric and oceanic science letters*, 8(1), 1-9.
- Kim, J. H., Kim, M. K., Ho, C. H., Park, R. J., Kim, M. J., Lim, J., ... & Song, C. K. (2019). Possible link between arctic sea ice and January PM10 concentrations in South Korea. *Atmosphere*, 10(10), 619.
- Zou, Y., Wang, Y., Zhang, Y., & Koo, J. H. (2017). Arctic sea ice, Eurasia snow, and extreme winter haze in China. *Science Advances*, 3(3), e1602751.
- Oh, H. R., Ho, C. H., Park, D. S. R., Kim, J., Song, C. K., & Hur, S. K. (2018). Possible relationship of weakened Aleutian low with air quality improvement in Seoul, South Korea. *Journal of Applied Meteorology and Climatology*, 57(10), 2363-2373.
- Jeong, Jaemin, et al. "Statistical predictability of wintertime PM2.5 concentrations over East Asia using simple linear regression." *Science of The Total Environment* 776 (2021): 146059.
- Hur, S. K., Oh, H. R., Ho, C. H., Kim, J., Song, C. K., Chang, L. S., & Lee, J. B. (2016). Evaluating the predictability of PM10 grades in Seoul, Korea using a neural network model based on synoptic patterns. *Environmental Pollution*, 218, 1324-1333.
- Chang-Hoi, H., Park, I., Oh, H. R., Gim, H. J., Hur, S. K., Kim, J., & Choi, D. R. (2021a). Development of a PM2.5 prediction model using a recurrent neural network algorithm for the

- Seoul metropolitan area, Republic of Korea. *Atmospheric Environment*, 245, 118021.
- Ho, C. H., Heo, J. W., Chang, M., Choi, W., Kim, J., Kim, S. W., & Oh, H. R. (2021b). Regulatory measures significantly reduced air-pollutant concentrations in Seoul, Korea. *Atmospheric Pollution Research*, 12(7), 101098.
- Lee, S., Ho, C. H., & Choi, Y. S. (2011). High-PM10 concentration episodes in Seoul, Korea: Background sources and related meteorological conditions. *Atmospheric Environment*, 45(39), 7240-7247.
- Lee, S., Ho, C. H., Lee, Y. G., Choi, H. J., & Song, C. K. (2013). Influence of transboundary air pollutants from China on the high-PM10 episode in Seoul, Korea for the period October 16-20, 2008. *Atmospheric Environment*, 77, 430-439.
- Oh, H. R., Ho, C. H., Kim, J., Chen, D., Lee, S., Choi, Y. S., ... & Song, C. K. (2015). Long-range transport of air pollutants originating in China: A possible major cause of multi-day high-PM10 episodes during cold season in Seoul, Korea. *Atmospheric Environment*, 109, 23-30.
- Lee, G., Ho, C. H., Chang, L. S., Kim, J., Kim, M. K., & Kim, S. J. (2020). Dominance of large-scale atmospheric circulations in long-term variations of winter PM10 concentrations over East Asia. *Atmospheric Research*, 238, 104871.
- Jia, B., Wang, Y., Yao, Y., & Xie, Y. (2015). A new indicator on the impact of large-scale circulation on wintertime particulate matter pollution over China. *Atmospheric Chemistry and Physics*, 15(20), 11919-11929.
- Jeong, J. I., & Park, R. J. (2017). Winter monsoon variability and its impact on aerosol concentrations in East Asia. *Environmental Pollution*, 221, 285-292.
- Yun, S. G., & Yoo, C. (2019). The effects of spring and winter blocking on PM10 concentration in Korea. *Atmosphere*, 10(7), 410.



주 의

1. 이 보고서는 극지연구소 위탁과제 연구결과보고서입니다.
2. 이 보고서 내용을 발표할 때에는 반드시 극지연구소에서 위탁연구과제로 수행한 연구결과임을 밝혀야 합니다.
3. 국가과학기술 기밀유지에 필요한 내용은 대외적으로 발표 또는 공개하여서는 안됩니다.