Journal of Industrial Technology, Kangwon Natl. Univ., Korea, No. 44:1 (2024) 1~7
ISSN 1229-9588
eISSN 1598-1371
https://doi.org/10.22805/JIT.2024.44.1.001

다중선형회귀와 인공신경망을 이용한 2015년 이전 $PM_{2.5}$ 일일 평균 수치 추정 방법론 제안

Daily $PM_{2.5}$ Estimation using Multiple Linear Regression and Artificial Neural Networks Before 2015

허진우[a], 박세준[b*]
Jin-Woo Huh[a], SeJoon Park[b*]

[a] IDeA Ocean Inc., Senior Researcher, 535 Bongcheon-ro, Gwanak-gu, Seoul 08789, Republic of Korea
[b] Division of Energy Resources and Industrial Engineering, Kangwon National University, Associate professor, Chuncheon 24341, Republic of Korea

Received 21 December 2023; Revised 22 December 2023; Accepted 28 December 2023

Abstract

Since 2015, the $PM_{2.5}$ measurement data has been publicly available nationwide in South Korea, but its use is restricted to after 2015, unlike other air pollutants. To overcome this limitation, multiple linear regression and artificial neural network models were developed to predict the daily average $PM_{2.5}$ values in South Korea before 2015. The daily data of air pollution measurement(SO_2, CO, O_3, NO_2, PM_{10}) and meteorological observation data (temperature, humidity, wind speed, atmospheric pressure, precipitation, snowfall) were used as input variables to develop regional prediction models for five regions(Seoul, Incheon, Gwangju, Daejeon, Ulsan) and a national prediction model. The models were developed and validated using the air pollution measurement data after 2015, and applied to predict $PM_{2.5}$ values before 2015. The multiple linear regression model showed R^2 values of 0.80 nationwide, 0.73 in Seoul, and 0.67 in Incheon, which enabled estimation of daily average $PM_{2.5}$ values before 2015. The artificial neural network model showed good prediction power with R2 values of 0.79 in Gwangju, 0.81 in Daejeon, and 0.72 in Ulsan. The regional prediction models showed good prediction power in most regions, and both the multiple linear regression and artificial neural network models showed good prediction power.

Keywords: Air pollutant, Fine particulate matter, Artificial neural networks, Multiple linear regression, Prediction

1. 서 론

$PM_{2.5}$는 WHO 산하 국제 암연구소에서 1군 발암물질로 지정되었으며 대한민국의 대기환경보전법에서 정의한 환경성 질환의 원인인 입자성 대기오염물질로 내연기관 작동, 산업 활동, 2차 생성 등의 요인으로 발생한다. 인체가 $PM_{2.5}$에 노출되는 경우 호흡기에 염증반응, 산화손상을 유발하고 심혈관계 질환 발병에 영향을 미치거나 암 발병 위험이 증가하며 안질환을 유발하는 등, 유병률과 사망률을 증가시킨다[1-2]. 이러한 유해성에 주목하여 많은 연구진들이 $PM_{2.5}$ 노출에 따른 유병률, 사망률, 상관관계 등의 주제로 연구를 진행하고 있다. 하지만 다른 대기오염물질과 달리 $PM_{2.5}$는 2015년 1월 1일부터 전국단위 측정 자료가 공개되었으며 질환 발병 자료인 국민건강보험공단 자료는 이보다 앞서 기록되었다. 따라서 측정 자료가 없는 2015년 이전의 국민건강보험공단 자료를 활용하는데 한계점이 있다. 이는 국민건강보험공단 자료뿐만 아니라 $PM_{2.5}$ 연관 자료를 사용하는 연구의 제한점이기도 하다. 이를 보완하기 위해 시정(visibility)을 이용한 서울과 춘천의 과거 연평균 $PM_{2.5}$ 농도 추이를 예측하는 연구가 진행되었으며, 통계적인 방법을 이용하여 부산의 2015년 이전 연평균 농도를 최대 R^2 0.84로 예측하는 연구가 진행되었다[3-4]. 하지만 이러한 노력에도 불구하고 대한민국의 기상과 대기환경에 적합한 과거 예측 방법론의 개발이 미흡한 실정이며 일간, 주간 단위의 과거 $PM_{2.5}$ 예측 연구가 미흡하다. 따라서 본 연구에서는 2015년 이후의 대기오염측정자료와 종관기상관측자료를 이용하여 2015년 이전 전국, 서울, 인천, 광주, 대전, 울산의 $PM_{2.5}$ 일일 평균을 예측하는 다중선형회귀, multiple linear regression(MLR) 모델과 인공신경망, artificial neural networks(ANN) 모델을 개발하고

* Corresponding author. Tel.: +82-33-259-5544
fax: +82-33-259-5544
E-mail address: sonmupsj@kangwon.ac.kr (SeJoon Park).

해당 모델을 검증한다. 검증된 모델에서, 우수한 모델을 선정하여 실제 2015년 이전 $PM_{2.5}$ 일일 평균을 얼마나 정확히 예측하는지 집중측정망에서 측정된 2012~2014년 $PM_{2.5}$ 실측치와 비교했다.

2. 연구 방법

2.1 대기오염 측정자료

본 연구는 대기오염측정자료로 일반대기오염측정망과 집중측정망의 자료를 사용했다. 일반대기오염측정망은 5종의 세부 측정망으로 나뉜다. 도시대기 측정망은 도시지역의 평균 대기질 농도를 파악, 도로변대기 측정망은 차량과 유동인구 통행이 많은 도로변 대기질을 파악하고 국가배경농도 측정망은 국가적배경농도와 타국으로부터의 오염물질유출과 유입을 파악하며 교외대기 측정망은 도시 주변 교외의 배경농도를 파악하는데 목적이 있다[5].

일반대기오염측정자료는 2012년 1월 1일부터 2021년 12월 31일까지 Air Korea에서 획득한 그림 1의 지역인 서울, 인천, 광주, 대전, 울산지역 모든 일반대기오염측정소의 시간단위 $PM_{2.5}$, SO_2, CO, O_3, NO_2, PM_{10} 측정치를 사용했다. 집중측정망은 배경지역과 주요권역별 대기질을 확인하고 오염물질의 유입과 유출을 파악, 분석하고 오염원인 규명을 위해 설치한 측정망이다. 집중측정망의 자료는 2012년 1월 1일부터 2014년 12월 31일까지의 자료로, 국립환경과학원에서 제공한 그림 1에 위치한 수도권(서울), 백령도(인천), 호남권(광주), 중부권(대전), 영남권(울산) 각 1개소의 시간단위 $PM_{2.5}$, PM_{10}측정치를 사용했다. 일반대기오염측정망과 집중측정망 자료는 결측치를 제외한 후, 지역별 일일 평균(mean), 분산(var), 표준편차(SD), 최대 & 최소를 산출했으며 각 지역별 측정치를 합산하여 전국의 일일 평균, 분산, 표준편차를 산출했다. 2015년 이후의 일반대기오염측정자료는 MLR과 ANN 모델 학습(In sample)사용한다. 2015년 이전 자료는 학습된 모델에 입력하여 PM2.5일일 평균 예측치를 산출한다. 집중측정자료는 일반대기오염측정자료를 이용하여 도출된 예측치의 검증(Out of sample)에 사용했다. 이때 집중측정망에서 측정된 PM_{10}은 $PM_{2.5}$예측치를 구할 때, 2015년 이전의 PM_{10}을 대신하여 입력되며, 측정된 $PM_{2.5}$는 모델에서 예측치와 비교하여 실제 성능을 확인한다.

2.2 종관기상관측자료

종관기상관측자료는 100~10,000km의 공간규모에서 발생하는 대기 순환 규모의 기상현상을 파악하기 위해, 지정된 관측소에서 관측된 기상관측자료이다.

기상현상은 대기오염물질의 측정치를 결정하는 요인 중 하나이다. 특히 $PM_{2.5}$의 거동과 수치를 예측하는 여러 연구에서는 기온, 습도, 풍속, 대기압, 강수, 적설을 주요한 변수로 선정했다[6-8]. 따라서 본 연구는 2012년 1월 1일부터 2021년 12월 31일까지 기상청 기상자료개방포털에서 획득한 그림 1에 위치한 서울, 인천, 광주, 대전, 울산지역 모든 관측소의 시간단위 기온, 습도, 풍속, 대기압, 강수, 적설 관측 자료를 사용했다. 획득한 기상관측자료는 결측치를 제외하고 평균(mean), 분산(var), 표준편차(SD), 최대 & 최소를 산출했으며 각 지역별 측정치를 합산하여 전국의 일일 평균, 분산, 표준편차를 산출했다. 2015년 이후의 기상관측자료는 MLR과 ANN 모델의 학습(In sample)에 사용되며, 2015년 이전 자료는 학습된 모델에 입력하여 모델의 실제 성능을 검증(Out of sample)한다.

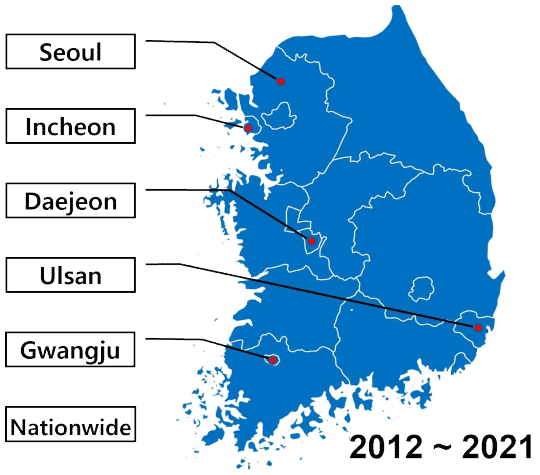

Fig. 1 Location of air pollutant and meteorological factor sampling stations.

2.3 $PM_{2.5}$ 일일 평균 예측 모델 개요

본 연구에서 전국과 서울, 인천, 광주, 대전, 울산의 2015년 이전 일일 평균 $PM_{2.5}$수치 예측 모델로 MLR과 ANN 모델을 사용했다. ANN 모델은 Multilayer Perceptron방식의 인공신경망을 이용하여 개발했다. 모델 개발에는 Microsoft R Open 4.0.2 환경에서, R의 내장 선형회귀함수인 lm 함수로 MLR 개발을, ANN 개발에는 R의 neuralnet library를 사용했다. 모델의 입력변수는 대기오염물질측정자료와 종관기상관측자료의 일일 평균, 분산, 표준편차이다. 반응변수는 해당 입력변수와 같은 날의 $PM_{2.5}$ 일일 평균 수치이다. 서울, 인천 광주, 대전, 울산에서 각각 도출된 일일 평균, 분산, 표준편차는 지역단위 자료를 구성하고, 위의 5곳의 자료를 종합하여 도출된 일일 평균, 분산, 표준편차는 전국단위 자료를 구성한다.

위의 전국단위와 지역단위 자료 중, 2015년 이후의 자료를 이용하여 각각 MLR과 ANN 모델을 학습시킨다. 이때 입력변수는, 반응변수인 $PM_{2.5}$ 일일 평균수치와 동일한 날의 자료를 이용하여 모델을 학습시킨다. 이러한 학습의 결과로, 2015년 이전 일반대기오염측정망에서 측정된 SO_2, CO, O_3, NO_2, PM_{10}과 종관기상관측소에서 측정된 기온, 습도, 풍속, 대기압, 강수, 적설의 일일 통계량을 입력하면, 추정된 $PM_{2.5}$ 일일 평균 수치가 도출된다. 추정된 $PM_{2.5}$ 일일 평균 수치는 2015년 이전 전국과 서울, 인천, 광주, 대전, 울산을 대표하는, 집중 측정망에서 측정된 $PM_{2.5}$ 측정치와 비교하여 실제 추정 성능을 확인한다. 이때 MLR 모델은 입력변수 가공이 요구되지 않으나 ANN 모델 특성상 0과 1사이의 입력을 요구하기 때문에 식(1)의 min-max scaling으로 변환했다.

$$x_{scaled} = \frac{x - min(x)}{max(x) - min(x)} \tag{1}$$

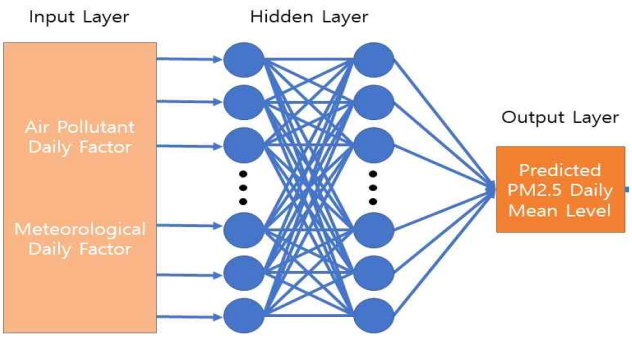

Fig. 2 Artificial Neural Networks Model Diagram

MLR 모델은 반응변수와 입력변수간의 관계를 선형방정식으로 표현한 회귀모델로 식(2) 같이 반응변수와 입력변수의 선형방정식 결과 오차제곱합이 최소가 되는 회귀계수와 오차를 구한다.

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \epsilon \tag{2}$$

ANN 모델은 그림 2와 같이 입력변수 자체인 input layer, 숨겨진 hidden layer, 예측치를 출력하는 output layer로 구성된다. hidden layer를 구성하는 node는 이전 layer의 출력을 받아 Activation function인 식(3)을 통해 출력한다.

$$f(x) = \frac{1}{1 + \exp(x)} \tag{3}$$

이때 반응변수와 ANN 모델에서 출력된 예측치 간, 오차제곱합의 편미분 절댓값과 비교한다. 모델 개발 과정에서 layer의 node간 결합의 연결강도를 조정하여, 오차제곱합의 편미분 절댓값이 threshold보다 작게 개발되며, threshold에 도달하거나 지정한 학습 횟수에 도달하면 종료된다[9]. ANN 모델은 MLR 모델보다 많은 연산이 요구되며 hidden layer의 개수, layer의 node 개수, threshold와 같은 모델 변수를 결정해야 하지만 성능이 우수하며 오차에 강건하여 시계열 예측에 널리 사용된다.

MLR과 ANN 모델은 편향된 예측을 하는 과대적합을 방지하고 모델의 일반화를 위해 필수적이다. 두 모델은 공통적으로 입력변수의 조합을 구성하는데 5가지의 일반대기오염측정자료와 6가지 종관기상관측자료의 일일 통계량에서 평균, 평균과 표준편차, 평균과 분산을 조합한 3가지 입력변수 조합을 구성했다. 이때 MLR의 경우, 표 1과 같이 입력변수를 선택하는 것 이외에 모델의 변수 설정이 필요 없으나 ANN은 모델의 변수 설정이 필요하다. 이를 위해 표 2와 같이 입력변수, hidden layer 내부의 nodes와 hidden layer의 구성 형태, 학습 종료의 기준인 threshold를 모델 변수의 요인으로 설정했다. 이러한 60(3 x 4 x 5)가지, 요인의 수준 조합을 실험하여 전국과 지역 모두 일반화가 가능한 모델을 도출한다. layer 개수는 사전 실험에서 3개 이상은 유의미한 성능향상이 없고, 1개는 저조한 성능의 모델로 개발되어 2개로 고정한다. layer의 nodes는 일반적으로 입력변수의 개수만큼 설정하는 것이 적절할 수 있으나 시행착오가 요구된다[10]. 또한 대한민국의 $PM_{2.5}$ 일일 평균 수치를 예측한 선행연구에서도 동일한 방식으로 layer와 nodes를 설정한 사례가 있다[11]. 사전실험에서 50개 이상의 nodes는 정상적인 모델로 개발되지 않았기 때문에 44개를 최대로 설정했으며 최소 입력변수 개수인 11의 배수로 수준을 지정했다. threshold는 작을수록 과대적합이, 높을수록 실측치와 예측치간 오차가 커지는 경향을 보인다. 따라서 적절한 threshold를 지정하기 위해 표 2의 threshold는 설정 가능한 최소인 0.01부터 0.025, 0.05, 0.075, 0.1로 설정했다.

Table 1 Levels of factors for MLR Model

Factor	Input Variable
Level	Mean Mean, Var Mean, SD

Table 2 Levels of factors for ANN parameters

Factor	Input Variable	Nodes	Threshold
Level	Mean Mean, Var Mean, SD	11 22 33 44	0.01 0.025 0.05 0.075 0.1

2.4 실험 자료 분할 및 평가 지표

표 3은 모델 개발에 사용된 입력변수 자료의 분할 비율을 나타낸 것이다. 학습된 모델 성능을 객관적으로 평가하기 위해서는 모델 개발에 사용하지 않은 자료로 검증해야 한다. 따라서 2015년 이후의 입력변수는 모델의 개발에 쓰이는 In sample로 선정, Out of sample은 2012~2014년의 자료로 모델의 실제 과거 예측 성능을 확인할 수 있다. 훈련에 사용된 자료는 2015년 1월 1일에서 2021년 12월 31일까지(2557)의 일일 단위 5가지의 일반대기오염측정자료와 6가지 종관기상관측자료의 일일 통계량(평균, 분산, 표준편차)이며 검증에 사용된 자료는 2012년에서 2014년까지(1095일)의 5가지 일반대기오염측정자료와 6가지 종관기상관측자료 일일 통계량이다. 즉 훈련자료와 테스트 자료는 2557일과 1095일로 70%와 30%의 비율을 가진다. 본 연구는 표 1과 표 2의 각각 모델의 요인 수준에 따른 모델의 Out of sample 결과로 상위 모델을 선발하고, 선발된 모델의 2015년 이전 PM2.5 일일 평균 수치 추정 성능과 모델에서 도출된 과거 일일 평균 추정치의 추세를 확인한다.

모델 성능을 평가하기 위해 평가 지표로 R squared(R^2)와 absolute error in nσ을 사용했다. R^2는 반응변수(y)와 예측치($\hat{y}$)간의 편차제곱합에서 반응변수와 반응변수 평균($\bar{y}$)간의 편차제곱합을 나눠, 1에서 차감한 것으로 성능평가에 주로 사용되는 지표이다(식 4). absolute error in nσ는 반응변수와 예측치간 절대차가 PM2.5 일일 평균 수치의 분산 근사치 12(1σ)와 약수(0.5σ, 0.25σ)보다 작은 경우가 전체에서 차지하는 비율을 나타내며 1에 가까울수록 높은 성능의 모델이다(식 5).

$$R^2 = 1 - \frac{(\hat{y}-y)^2}{(y-\bar{y})^2} \quad (4)$$

$$Absolute\ Error\ in\ n\sigma = \frac{\sum_{i=1}^{k} a_i}{k} \quad (5)$$

$$a_i = \begin{cases} | \ \hat{y} - yvert < n\sigma \cdots 1 \\ | \ \hat{y} - yvert > n\sigma \cdots 0 \end{cases}$$

$$(n = 1, 0.5, 0.25 \ ; \ \sigma = 12)$$

3. 결 과

본 연구에서는 $PM_{2.5}$의 2015년 이전 일일 평균 수치를 도출하기 위해 2015년 이후 일반대기오염측정망과 종관기상관측소에서 측정된 자료를 이용하여 MLR과 ANN 모델을 개발했다. 그 결과로 2015년 이전의 $PM_{2.5}$를 제외한 일반대기오염측정 자료와 종관기상관측자료를 입력하면 추정된 $PM_{2.5}$ 일일 평균 수치가 도출되는 모델이 개발되었다. 표 3은 요인과 수준의 조합에 따른 MLR모델 중 전국과 개별 지역에서 가장 높은 가장 높은 Out of sample R^2을 달성한 모델을, 표 4는 마찬가지로 ANN 모델을 나타낸 것이다. 표 4에서 전국 단위 MLR 모델은 다른 지역단위 모델보다 높은 0.8 이상 R^2를 보였으며 광주, 대전, 서울이 뒤를 이었다. 표 4의 ANN 모델은 전국 단위가 0.79이나 그 수치는 MLR 모델보다 낮다. MLR과 ANN 모델을 비교했을 때 전국 단위, 서울, 인천은 MLR 모델이 높으나 광주, 대전,울산은 ANN 모델의 예측력이 높은 것을 확인할 수 있다.

Table 3 Table of the best-performing individual regional MLR models

Area	Input Variable	R^2	
		In sample	Out of sample
Nationwide	Mean, SD	0.8898	0.8043
Seoul	Mean	0.7862	0.7303
Incheon	Mean	0.8112	0.6744
Gwangju	Mean, SD	0.8526	0.7886
Daejeon	Mean, SD	0.8243	0.7851
Ulsan	Mean, SD	0.8709	0.5917

Table 4 Table of the best-performing individual regional ANN models

Area	Input Variable (Nodes; Threshold)	R^2	
		In sample	Out of sample
Nationwide	Mean (33; 0.025)	0.9518	0.7978
Seoul	Mean (44; 0.1)	0.8382	0.7142
Incheon	Mean, SD (11; 0.025)	0.9386	0.1607
Gwangju	Mean, SD (33; 0.01)	0.8946	0.7968
Daejeon	Mean, SD (33; 0.05)	0.8807	0.8192
Ulsan	Mean, SD (33; 0.075)	0.9062	0.7296

여기서 인천은 다른 지역에 비해 예측력이 낮은 것을 알 수 있는데, 인천에 해당하는 집중측정망은 백령도에 위치한다. 이는 내륙과 도서지역을 포괄하는 2015년 이후 $PM_{2.5}$ 측정치와 비교했을 때 2014년 이전, 백령도에 위치한 집중측정망의 지역적 특성이 다르기 때문에 비교적 다른 내륙 지역과 비교했을 때 편차가 발생하는 것이다. ANN 모델을 사용한 대전의 결과가 가장 높은 예측력을 보이며, 그 다음으로 전국 단위 ANN, MLR 모델이 잘 예측했다는 것을 알 수 있다. 인천을 제외한 대부분의 지역이 0.72 이상, 최대 0.81의 높은 정확도로 2015년 이전 $PM_{2.5}$ 일평균 수치를 추정했다는 것을 알 수 있다. MLR과 ANN 모델 중, 표 3와 표 4의 결과 중, 개별 지역에서 가장 높은 Out of sample R^2를 보이는 모델을 선정하면 다음과 같다. MLR 모델은 광주, 대전, 울산, ANN 모델은 전국, 서울 인천을 높은 정확도로 2015년 이전의 $PM_{2.5}$ 일일 평균 수치를 추정했다. 그에 대한 실측치와 예측치의 절대차가 얼마나 작은지 확인할 수 있는 결과는 표 5와 같다. 우선 인천을 제외한 모든 모델에서 도출된 추정치가 실측치인 집중측정망의 일일 평균 농도와의 편차가 90% 이상 12 이내이며 70% 내외가 6 이내의 절대차로 추정했음을 알 수 있다.

특히 대전의 경우 46%의 추정치와 실측치 차이가 3 이내의 매우 정확한 결과를 도출했음을 알 수 있다. 사용 모델의 입력변수, 그리고 모델에서 조정할 변수를 적절히 조절하면 MLR과 ANN 모델은 예측에 적합한 모델이며 높은 예측성능을 달성하여 2015년 이전의 $PM_{2.5}$ 일일 평균 수치를 추정할 수 있다. 전반적으로 실측치와 예측치가 작은 오차 내에서 분포하는 것을 그림 3에서 보는 것과 같이 알 수 있다. 우선 2015년 이전의 $PM_{2.5}$를 R^2 0.80으로 예측한 전국 모델은 실측치와 예측치 산점도가 직선에 근접하게 분포함을 확인할 수 있다. 서울은 0.80대의 전국과 대전보다 낮은 0.73을 기록했고 그림 3의 산포도 개형도 비교적 타원에 가까우나 직선에 근접함을 알 수 있다. 인천은 개중 가장 낮은 성능으로 분포가 가장 넓다. 광주는 0.79의 R^2로 서울과 전국, 대전의 중간 정도의 분포를 띈다. 대전은 가장 높은 예측력을 보이며 0.81의 성능을 달성했고 6개의 그래프 중 가장 직선에 근접한 분포를 띈다. 마지막으로 울산은 서울과 유사한 성능 지표를 보였으나 Out of sample이 다른 그래프보다 희박하다. 이는 실측치가 되는 집중측정망자료에서 결측값이 다른 지역보다 많아 발생한 현상이다.

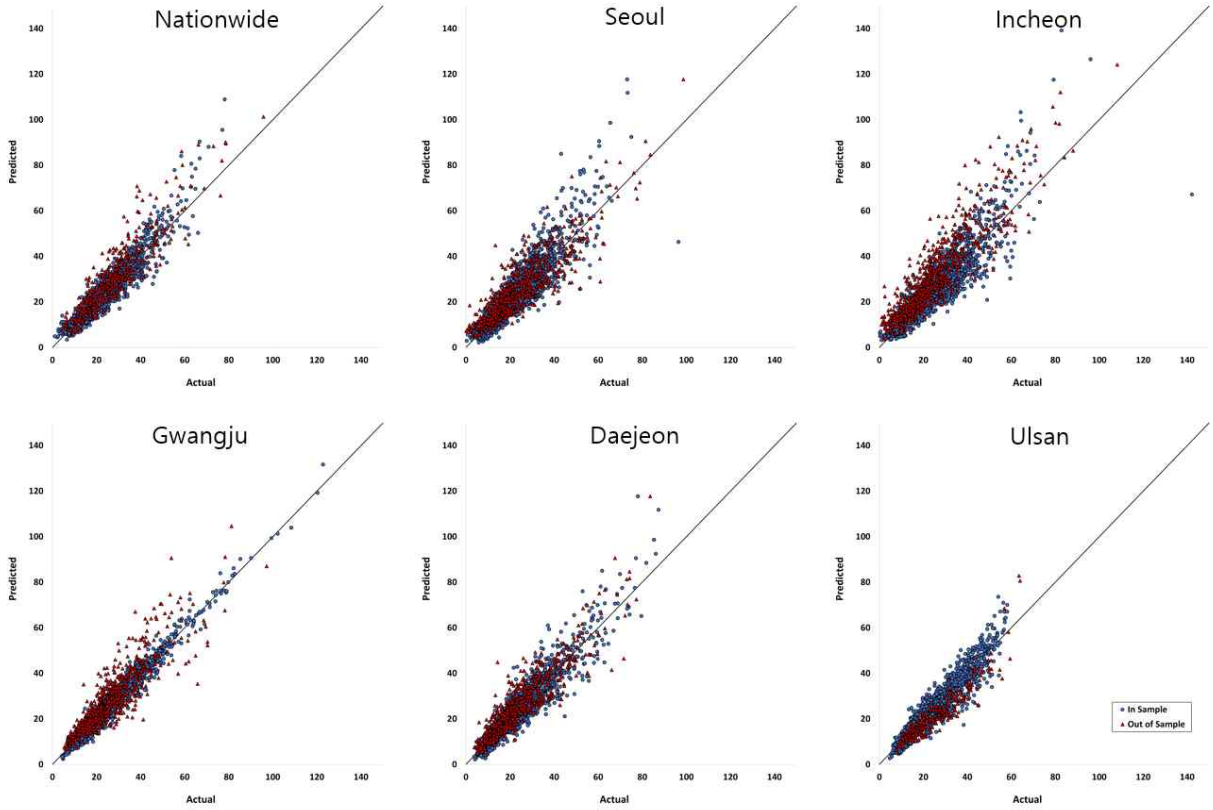

Fig. 3 Scatter plot of predicted and actual values derived from the best performing model in each region

Table 5 The absolute error in nσ of the highest performing model

Area(Model)	In sample			Out of sample		
	Absolute error in			Absolute error in		
	1σ	0.5σ	0.25σ	1σ	0.5σ	0.25σ
Nationwide (MLR)	0.9878	0.8912	0.6619	0.9107	0.6784	0.4006
Seoul (MLR)	0.9593	0.7926	0.4937	0.9041	0.6463	0.3670
Incheon (MLR)	0.9495	0.7844	0.5093	0.7595	0.5010	0.2696
Gwangju (ANN)	0.9827	0.8583	0.5874	0.9202	0.7087	0.4318
Daejeon (ANN)	0.9780	0.8693	0.6013	0.9484	0.7525	0.4649
Ulsan (ANN)	0.9913	0.9158	0.6701	0.9566	0.7244	0.4409

4. 결 론

본 연구는 2015년 이후의 종관기상관측자료와 일반대기오염물질 측정자료를 이용하여 일일 통계량을 산출했고 그 중 중심의 대푯값인 평균과 산포의 대푯값인 분산, 표준편차의 조합으로 입력변수를 구성했다. 구성한 입력변수를 이용하여 다중선형회귀 모델과 인공신경망을 이용하여 2015년 이전, $PM_{2.5}$ 일일 평균 수치를 예측하는 모델을 개발했고 그 결과를 서울, 인천, 광주, 대전, 울산에서 측정된 집중측정망 $PM_{2.5}$ 측정치와 비교하여 실제 2015년 이전일일 평균 수치를 추정하는 모델을 평가 했다. 그 결과 입력변수는 평균, 평균과 표준편차 일 때, 전국과 모든 지역에서 높은 성능을 보였다. MLR 모델은 표 5에서와 같이 전국(R^2= 0.80), 서울(R^2= 0.73), 인천(R^2= 0.367)에서 가장 높은 성능으로 2015년 이전 $PM_{2.5}$ 일일 평균 수치를 추정했으며 ANN 모델은 광주(R^2= 0.79), 대전(R^2= 0.81), 울산(R^2= 0.72)에서 가장 높은 성능으로 추정했다.

인천은 도서지역에 위치한 집중측정망의 특성상 인천 지역 전체를 대표하는 2015년 이전 $PM_{2.5}$ 일일 평균 수치를 추정하는데 어려움이 있어 다른 지역보다 다소 낮은 성능을 보였다. 본 연구는 연간 단위의 선행연구보다 작은 일일 단위의 $PM_{2.5}$를 높은 정확도로 추정할 수 있으며 지상관측 자료만을 이용하여 과거의 추세를 추정할 수 있는 방법론을 제안한다. 이러한 결과로 2015년 이후로 제한되어 있는 대한민국의 $PM_{2.5}$ 유관 연구와 관측 개시 이전 $PM_{2.5}$ 농도 수치 변화 추정 연구에 도움이 될 것이다.

5. 감사의 글

본 논문은 환경부의 재원으로 국립환경과학원의 지원을 받아 수행하였습니다(NIER-2022-03-00-008).

References

[1] Jang, AS., 2014, Impact of particulate matter on health, Journal of the Korean Medical Association 57:9 763-768.

[2] Kim, Y., Choi, Y. H., 2020, Different adverse effects of air pollutants on dry eye disease: Ozone, PM2.5, and PM10, Environmental Pollution 265 115039.

[3] Lee, Y. H., Kwak, K.H., 2018, Using visibility to estimate PM2.5 concentration trends in Seoul and Chuncheon from 1982 to 2014, Journal of Korea Society for Atmospheric Enviroment 34:1 156-165.

[4] Lee, S. H., Kim, S. Y., 2018, An approach to estimating national-scale annual-average concentrations of PM2.5 before 2015 when national air quality monitoring data are available in South Korea, Journal of Korea Society for Atmospheric Enviroment 34:6 806-821.

[5] Ministry of Environment, 2020, ANNUAL REPORT OF AIR QUALITY IN KOREA, Republic of Korea Ministry of Environment 8p.

[6] Jee, J. B., Cho, C. R., 2022, An Analysis of Meteorological Characteristics for Fine Particles on the Korean Peninsula during Wintertime, 2015-2021, Journal of Korea Society for Atmospheric Enviroment 38:3 394-413.

[7] Zheng, M., Salmon, L. G., 2005, Seasonal trends in PM2.5 source contributions in Beijing, China, Atmospheric Environment 39:22 3967-3976.

[8] Huang, G., Li, X., 2021, PM2.5 concentration forecasting at surface monitoring sites using GRU neural network based on empirical

mode decomposition, Sicience of The Total Environment 768 144516.

[9] Gunther, F., Fritsch, S., 2010, Neuralnet: training of neural networks, The R Journal 2:1 30.

[10] Shmueli, G., Bruce, P. C., 2017, Data mining for business analytics: concepts techniques and applications in R, John Wiley & Sons, Seoul, Korea.

[11] Huh, J. W., Youn, J. S., 2023, Development of a Prediction Model for Daily PM2.5 in Republic of Korea by Using an Artificial Neural Network, Applied Sciences 13:6 3575.