

동아시아 광역 데이터를 활용한 DNN 기반의 서울지역 PM₁₀ 예보모델의 개발

유 숙 현[†]

Development of PM₁₀ Forecasting Model for Seoul Based on DNN Using East Asian Wide Area Data

SukHyun Yu[†]

ABSTRACT

In this paper, PM₁₀ forecast model using DNN(Deep Neural Network) is developed for Seoul region. The previous Julian forecast model has been developed using weather and air quality data of Seoul region only. This model gives excellent results for accuracy and false alarm rates, but poor result for POD(Probability of Detection). To solve this problem, an WA(Wide Area) forecasting model that uses Chinese data is developed. The data is highly correlated with the emergence of high concentrations of PM₁₀ in Korea. As a result, the WA model shows better accuracy, and POD improving of 3%(D+0), 21%(D+1), and 36%(D+2) for each forecast period compared with the Julian model.

Key words: AI, DNN, PM₁₀ Forecasting, Air Quality

1. 서 론

최근 미세먼지 문제는 고농도의 미세먼지가 장기간 정체된다는데 그 심각성이 크다. 고농도 초미세먼지가 발생하는 경우는 3~4일간 혹은 그 이상 지속되는 경향이 있고 습도가 높은 경우에 생성속도가 가속되기 때문에 짙은 스모그 형태로 나타나고 있다[1].

2014년 2월부터 환경부에서 시행하고 있는 미세먼지 예보는 실제 시민이 느끼는 체감 오염도와 차이를 보이고 있고, 특히 고농도 적중률에 있어서는 더욱 그 차이가 크다. 이러한 미세먼지 문제를 예방적 차원에서 해결하기 위해서는 미세먼지 예보의 성능이 향상되어야 한다.

현재 미세먼지 예보는 수치모델 결과를 중심으로 수행되고 있으나 수치예보는 배출량 및 기상자료 등

으로 대표되는 입력 자료의 불확실성과 수치모델 자체가 복잡한 대기현상을 완전하게 반영하지 못하는 근본적인 한계를 가지고 있다.

수치모델 CASE04는 본 논문에서 개발한 예보 모델의 성능 비교 대상으로 CMAQ을 사용하여 PM₁₀ 예측 농도를 생성하는 모델이다. CMAQ은 가스상, 입자상 대기오염물질을 통합하여 'one-atmosphere' 평가를 할 수 있는 오일러리안 화학수송 모델로 기상 모델링 자료, 배출량 모델링 자료, 광분해 모델 자료 등을 입력 자료로 대기 중 오염물질의 균질 및 비균질 화학반응, 이류 및 확산을 계산한다[1]. 현재 CMAQ은 U.S EPA(United States Environmental Protection Agency) 산하 CMAS(Community Modeling and Analysis System) 센터를 중심으로 꾸준히 개발 보완되고 있다. CASE04는 CMAQ 대기질

* Corresponding Author : SukHyun Yu, Address: (14028) 22 Samdeokro 37beon-gil, Manan-gu, Anyang-si, Gyeonggi-do, Korea, TEL : +82-31-467-1272, FAX : +82-31-463-1249, E-mail : shy@anyang.ac.kr

Receipt date : Aug. 30, 2019, Revision date : Oct. 29, 2019
Approval date : Nov. 25, 2019

[†] Dept. of Information & Communication Eng., Anyang University

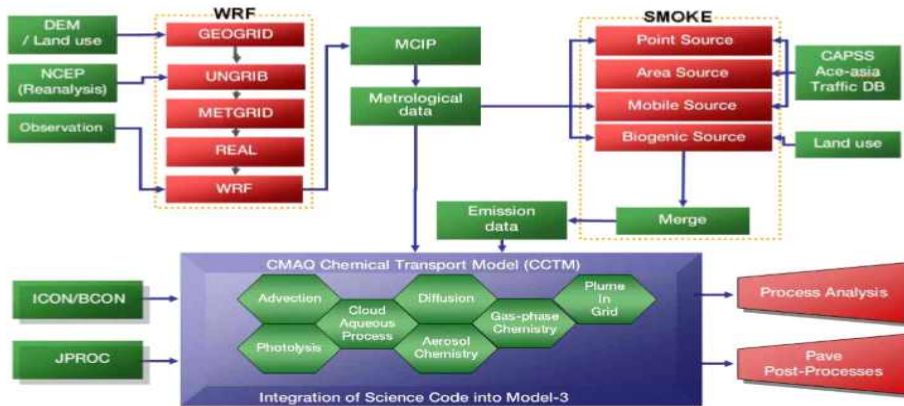


Fig. 1. CMAQ air quality modeling system.

모델링 시스템에 기상 모델인 WRF v3.6.1과 아시아 배출량 자료인 MEIC(2010)과 REAS(2008), 국내 배출량 자료인 2011 CAPSS를 입력하고, 자료동화를 적용하여 모델링하였다.

본 연구에서는 기존 수치모델의 한계를 극복하기 위해서 기상 및 대기질 측정 자료와 예보 자료를 심층신경망(DNN: Deep Neural Network)에 학습시켜 불완전한 수치모델(CASE04)의 결과를 보정하고, 정확도를 향상시킬 수 있는 예보모델을 개발했다. 특히 고농도 적층층을 향상시키기 위해서 예보 대상이 되는 서울 권역의 자료 뿐 아니라 동아시아 지역의 자료도 학습 인자로 추가했다. 이것은 중국 주요 공업 도시지역에서 발생한 고농도 초미세먼지에 의한 스모그가 편서풍 계열의 바람을 타고 국내로 유입되고, 여기에 국내 배출원이 가중되어 고농도 미세먼지가 정체되는 현상을 반영한 것이다[2-4].

미세먼지를 비롯한 대기 오염물질을 예보하는데 인공지능망의 성능은 이미 입증되어 관련 연구들이 진행되고 있다. 관련 연구로는 인공지능망에 날짜 데이터(month, weekly day, julian day), 대기질 데이터, 지형 데이터와 기상 측정 및 예보 데이터를 입력하여 PM₁₀과 PM_{2.5}의 예보와 성능을 평가 분석한 연구[5-13], RNN(Recurrent Neural Network)을 이용하여 O₃ 및 PM₁₀, PM_{2.5}의 농도를 예측한 연구[14-15], RBF(Redial Based Function)를 이용하여 PM₁₀의 농도를 예측한 연구[16]가 있다. 그 밖에 시계열 대기질 자료의 결측치를 처리하기 위해서 DRNN(Deep Recurrent Neural Network)를 기반으로 시공간 예측 프레임워크를 제안한 연구[17], PM₁₀ 예보

성능의 향상을 위해서 민감도 분석을 사용하여 역모델 파라메타를 추정한 연구[18], 대기질 예보의 성능 향상을 위해서 커널 삼중대각 희소행렬을 이용해 고속으로 자료동화 수행한 연구[19] 등이 있다.

본 연구에서는 서울 권역을 대상으로 당일(D+0), 내일(D+1), 모레(D+2)의 PM₁₀ 예보를 위해서 심층신경망을 기반으로 예보 모델을 개발했다. 개발한 예보모델은 동아시아 전체의 기상 및 대기질 측정자료와 예보자료를 이용한 WA 모델이다. WA 모델의 성능은 기존 예보모델인 수치모델 CASE04와 선행 연구에서 개발한 서울권역의 기상 및 대기질 측정자료와 예보자료를 사용한 Julian 모델과 비교 기술했다. 제안한 연구는 다음과 같은 순서로 기술한다. 이어지는 2장에서는 예보에 사용한 입력인자와 심층신경망의 네트워크 구조에 대해서 설명하고, 3장에서는 각 예보모델들을 기술하며, 4장에서는 제안한 모델별 PM₁₀ 예보결과를 나타내고, 5장에서 결론을 맺는다.

2. 입력인자와 네트워크 구조

2.1 입력인자



Fig. 2. East asia region.

본 연구에서는 서울권역의 PM₁₀ 예보를 위해서 한반도 권역을 비롯한 동아시아 권역의 대기질 및 기상 측정 자료와 예보 자료를 학습 인자로 사용했다. 동아시아 권역은 총 20개의 세부권역으로 구분되며 Fig. 2와 Table 1에 구체적으로 기술했다. 이 중 자료를 확보할 수 없는 북한 권역인 R11과 해양 권역인 R20을 제외하고는 모두 학습에 사용했다.

Table 2에는 학습인자로 사용한 대기질 및 기상 측정인자들을 제시했다. 서울 권역을 제외한 나머지 주변 권역은 Table 2에서 제시한 자료 중 PM₁₀과 가장 상관도가 높은 PM₁₀, PM_{2.5}, O₃, NO₂, CO, SO₂,

U, V, PA, HM와 수치모델인 CASE4_PM₁₀을 사용했고, 해당 예보 권역인 서울 권역은 Table 2에 제시한 모든 인자를 사용했다.

2.2 네트워크

본 연구에서는 1개의 입력층과 여러 개의 은닉층, 1개의 출력층을 가지는 DNN을 기반으로 한 예보모델 개발했다. 선행 연구인 Julian 예보모델과 본 논문에서 제안한 WA 예보모델은 입력인자 및 네트워크 파라메타가 서로 다르나, 기본적으로 Fig. 3에 제시한 네트워크 구조를 따른다.

Table 1. Regional code index

Regional code	Area name	Regional code	Area name
R01	Taklamakan Desert	R10	Inland China
R02	Western Mongolia	R12	Seoul
R03	Central Mongolia	R13	Northern Gyeonggi-do
R04	Eastern Mongolia	R14	Southern Gyeonggi-do
R05	Northeast China	R15	Incheon
R06	Hebei, China	R16	Gangwon-do
R07	Shandong, China	R17	Chungcheong-do
R08	Jiangsu, China	R18	Gyeongsang-do
R09	South China	R19	Jeonla-do

Table 2. Input data

Variable	Unit	Min	Max	Variable	Unit	Min	Max
PM ₁₀	µg/m ³	0	180	Solar radiation	0.01MJ/hr-m ²	0	2
PM _{2.5}	µg/m ³	0	110	CASE4_PM ₁₀	µg/m ³	0	180
SO ₂	ppm	0	0.01	Mixed highs	m	100	2000
NO ₂	ppm	0	0.01	Geopotential height_850hpa	m	1250	1500
O ₃	ppm	0	0.08	Pressure_850hpa	Pa	97000	102500
CO	ppm	0	1.4	U_850hpa	m/s	-15	20
U	m/s	-6	6	V_850hpa	m/s	-15	15
V	m/s	-5	5	Relative humidity_850hpa	%	0	50
Pressure	Pa	970	1025	Temperature_850hpa	K	270	310
Temperature	K	260	310	Geopotential height_925hpa	m	710	860
Dew point temperature	K	260	310	U_925hpa	m/s	-15	20
Relative humidity	%	0	100	V_925hpa	m/s	-15	15
Accumulated Precipitation	mm	0	10	Temperature difference_850-925	K	0	8

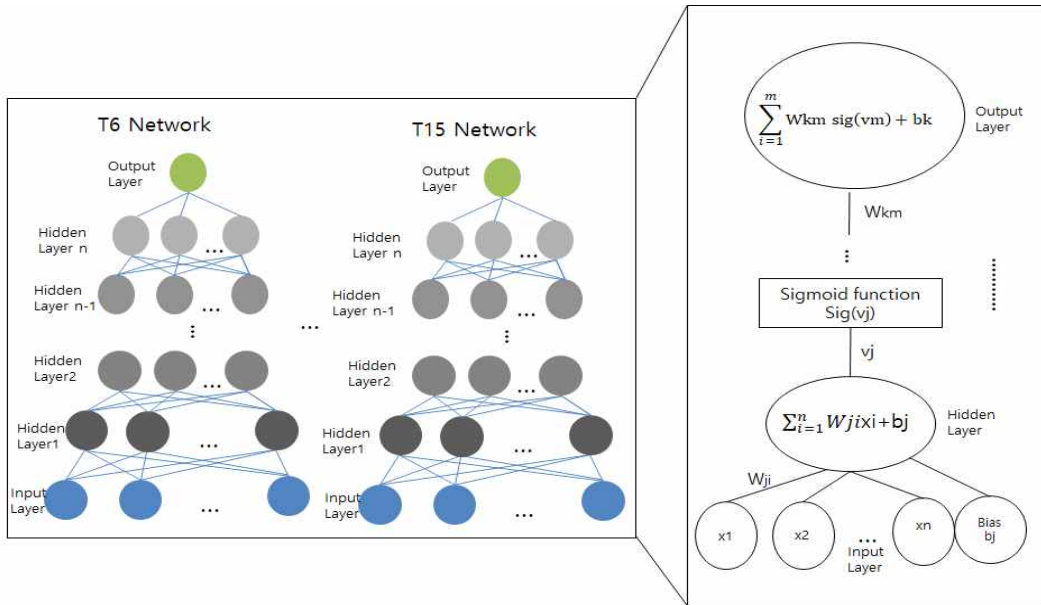


Fig. 3. Proposed DNN architecture.

제안한 예보모델의 네트워크는 총 10개로 구성되는데, 이를 Fig. 4에서는 T6~T15로 기술했다. 10개의 각 네트워크는 해당 시간 구간의 6시간 평균 자료를 입력으로 하여 6시간마다 예보 값을 생성하는데 총 3일(D+0, D+1, D+2) 예보를 수행한다. Fig. 4에 제안한 예보모델에서 사용한 시간 프로세스를 제시하고, 구체적인 시간 프레임은 Table 3에 기술했다. 15시에 예보를 수행하므로 T1~T5 구간은 과거 구간이고, T6~T15는 예보 구간으로 T6~T7이 당일(D+0), T8~T11이 내일(D+1), T12~T15가 모레(D+2)에 해당한다.

각 네트워크의 입력층에는 과거 6시간(T5)의 측정 데이터들과 각 예보구간(T6~T15)의 예보 데이터들이 입력되고, 출력층에는 각 예보구간의 PM₁₀ 측정값이 목표값으로 입력되어 학습이 진행된다. 학습은 경사하강법에 의해 가중치 및 파라미터를 최적화 하는 과정으로 이때, 비용함수는 MSE(Mean Square Error), 활성화함수는 시그모이드를 사용했다.

3. 제안한 예보모델

3.1 예보모델의 개요

본 논문에서는 서울 권역의 3일(D+0, D+1, D+2) 예보를 수행하기 위해서 동아시아 전체의 자료를 사용한 WA 예보 모델을 개발했다.

WA 예보 모델의 근간에 되는 선행 연구 Julian 예보모델은 서울권역의 대기질 및 기상 측정 데이터와 예보 자료, 날짜 데이터를 사용한 모델로 구체적인 학습인자는 Table 4에 제시했다.

날짜 데이터는 미세먼지의 농도와 높은 상관성을 갖는 기본적인 인자로 주어진 날짜를 어떻게 표현하느냐에 따라 미세먼지와 상관성을 더 정확하게 표현할 수 있다. Julian 예보모델에서는 날짜를 제안한 줄리안 멤버십 함수(Fig. 5)에 의해 12개의 인자로 표현한다. 줄리안 멤버십 함수는 시간의 연속성 및 미세먼지의 계절적 특성을 고려하여 날짜를 1월에서 12월을 대표하는 12개의 연속적인 인자로 나타낸다.

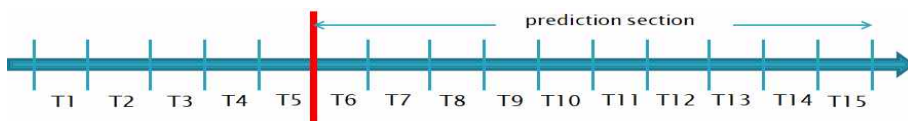


Fig. 4. Time unit concept of forecast model.

Table 3. Time frame of 15 hour forecast model

Time	Time corresponding to section
T1	D-1 07~12 o'clock, Measure at forecast time
T2	D-1 13~18 o'clock, Measure at forecast time
T3	D-1 19~24 o'clock, Measure at forecast time
T4	D+0 01~06 o'clock, Measure at forecast time
T5	D+0 07~12 o'clock, Measure at forecast time
T6	D+0 13~18 o'clock
T7	D+0 19~24 o'clock
T8	D+1 01~06 o'clock
T9	D+1 07~12 o'clock
T10	D+1 13~18 o'clock
T11	D+1 19~24 o'clock
T12	D+2 01~06 o'clock
T13	D+2 07~12 o'clock
T14	D+2 13~18 o'clock
T15	D+2 19~24 o'clock

Table 4. Input data of Julian forecast model

Time	Kind		Data
T5	Observation data(14)	Weather	U, V, Pressure, Temperature, Dew point temperature, Relative humidity, Precipitation, Solar radiation
		Air quality	PM _{2.5} , O ₃ , NO ₂ , CO, SO ₂ , PM ₁₀
T# (#:6~15)	Forecast data(16)	Air quality	CASE4_PM ₁₀
		Weather	Temperature, Pressure, Relative humidity, Mixed highs, U, V, Geopotential height_850hpa, Pressure_850hpa, U_850hpa, V_850hpa, Relative humidity_850hpa, Temperature_850hpa, Geopotential height_925hpa, U_925hpa, V_925hpa, Temperature difference_850-925
T# (#:6~15)	Date(12)		Julian: 12 factor expressed by the Julian membership function

12개의 인자 중 2개만이 활성화되고, 나머지 10개의 인자는 0으로 설정하는데, 활성화되는 2개의 인자는 각각 주어진 날짜의 월(Month)과 인접 월(Adjacent_Month)이다. 예를 들어, 주어진 날짜가 4월 20일 이라면 제안한 줄리안 멤버십 함수에 의해서 인접 월(Adjacent_Month)은 5, 해당 월의 값(Month_value)은 0.8, 인접 월의 값(Adjacent_Month_value)은 0.2가 되어 날짜를 나타내는 12개 인자는 0, 0, 0, 0.8, 0.2, 0, 0, 0, 0, 0, 0가 된다. 이 인자들의 의미는 주어진 날짜(4월 20일)가 4월에 해당되는 날짜이지만, 5월에 인접해있는 날짜이므로, 4월에 해당하는 4번째 인자는 0.8, 5월에 해당하는 5번째 인자는 0.2로 설정하여 날짜의 연속성 및 미세먼지의 계절적

추이를 표현한 것이다.

WA 모델은 중국으로부터 유입되는 미세먼지의 이동성을 고려해서 서울권역의 미세먼지 예보를 수행하는데 동아시아 권역 전체의 자료를 사용한 모델이다. 예보 대상 지역인 서울을 제외한 주변권역인 중국지역(R01~R10)과 국내권역(R13~R19)은 T5의 10개의 측정 자료와 PM₁₀ 예보 값을 사용하고, 서울권역(R12)은 Table 4에서 제시한 인자를 동일하게 사용했다. Table 5에 WA에서 사용한 학습인자를 구체적으로 기술했다.

Table 6에는 예보모델별 네트워크 파라메타를 기술했다. 선행연구인 Julian 예보모델은 측정 자료 14개, 예보 자료 16개, 날짜 자료 12개로 총 42개의 입력

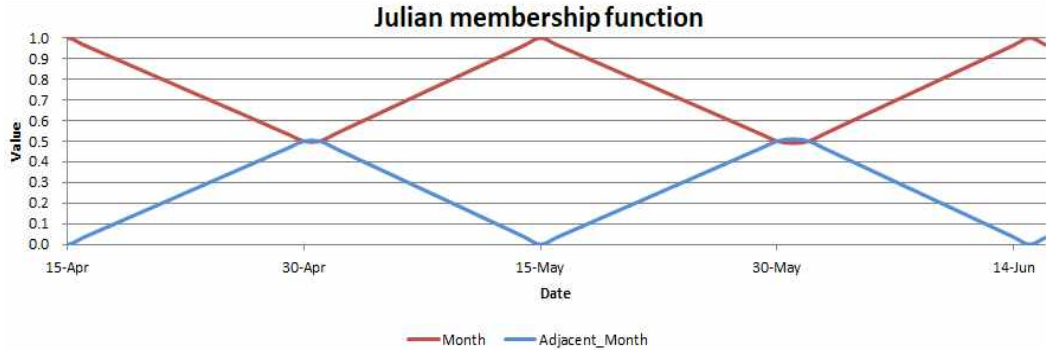


Fig. 5. Julian membership function.

Table 5. WA forecast model

Region	Time	Kind		Data
R12	T5	Observation data(14)	Weather	U, V, Pressure, Temperature, Dew point temperature, Relative humidity, Precipitation, Solar radiation
			Air quality	PM _{2.5} , O ₃ , NO ₂ , CO, SO ₂ , PM ₁₀
	T# (#:6~15)	Forecast data(16)	Air quality	CASE4_PM ₁₀
			Weather	Temperature, Pressure, Relative humidity, Mixed highs, U, V, Geopotential height_850hpa, Pressure_850hpa, U_850hpa, V_850hpa, Relative humidity_850hpa, Temperature_850hpa, Geopotential height_925hpa, U_925hpa, V_925hpa, Temperature difference_850-925
T# (#:6~15)	Date(12)		Julian: 12 factor expressed by the Julian membership function	
R01 ~R10, R13 ~R19	T5	Observation data(10)	Weather	U, V, Pressure, Relative Humidity
			Air quality	PM _{2.5} , O ₃ , NO ₂ , CO, SO ₂ , PM ₁₀
	T# (#:6~15)	Forecast data(1)	Air quality	CASE4_PM ₁₀

Table 6. Network parameter of forecast models

Model	Input layer	Hidden layer1	Hidden layer2	Hidden layer3	Hidden layer4	Output layer	Learning rate
Julian	42	21	10	5	2	1	0.009
WA	229	97	41	7		1	0.009

인자를 사용했고, WA 예보모델은 주변권역의 경우 측정자료 10개, 예보자료 1개로 11개의 입력인자를 사용했고, 이러한 주변권역이 17개이므로 187개의 인자가 입력되고, 예보 대상 권역인 서울은 Julian 모델과 동일하게 42개의 인자를 사용했으므로 총 229개의 인자가 입력되었다.

4. 실험 결과 및 고찰

4.1 성능평가 도구

현재 미세먼지 예보가 지수평가로 시행되고 있으므로 제안한 예보 모델의 성능을 평가하기 위해서 동일하게 지수 평가 도구인 지수적중률, 감지확률, 오경보율 등을 사용했다. 적중률(A: Accuracy)는 전

Level		Prediction value			
		1	2	3	4
Observation value	1	a1	b1	c1	d1
	2	a2	b2	c2	d2
	3	a3	b3	c3	d3
	4	a4	b4	c4	d4

I: , II: , III: , IV:

Evaluation item	Equation
A: Accuracy	$= \frac{(a1 + b2 + c3 + d4)}{N} \times 100(\%)$
POD: Probability of Detection	$= \frac{IV}{(III + IV)} \times 100(\%)$
FAR: False Alarm Rate	$= \frac{(II)}{(II + IV)} \times 100(\%)$

Fig. 6. Items and methods for evaluating the performance of the forecast model.

체 예보 성능을 평가하고, 감지확률(POD: Probability of Detection)은 고농도 성능을 평가하며, 오경보율(FAR: False Alarm Rate)는 잘못 예보된 고농도 비율을 분석하는 도구이다. 자세한 내용은 Fig. 6에 기술했다.

통합 대기환경지수는 대기오염도에 따른 인체 영향 및 체감오염도를 고려하여 개발된 표현방식으로 좋음(1), 보통(2), 나쁨(3), 매우 나쁨(4)로 구분되며, PM₁₀ 지수구간 범위는 Table 7에 나타났다.

4.2 실험 결과

본 연구에서는 서울 권역의 3일(D+0, D+1, D+2) PM₁₀ 예보를 위해서 WA 예보모델을 개발했다. 실험에 사용된 데이터는 서울 권역 및 동아시아 권역의 2015년 01월 01일부터 2019년 03월 12일 기간의 6시간 평균 데이터로 2015년 01월 01일에서 2017년 12월 31일까지는 학습에 사용하고 2018년 01월 01일부터 2019년 03월 12일 기간의 데이터는 평가에 사용했다. Table 8에 4.1절에서 기술한 지수평가 도구에 의해 개발한 예보모델을 평가한 결과를 기술했다. 또한,

Table 7. Integrated Air Quality Index of PM₁₀

level	1		2		3		4
	Min	Max	Min	Max	Min	Max	-
PM ₁₀ (µg/m ³)	0	30	31	80	81	150	151 or more

Table 8. PM10 prediction results of each forecasting model

Model	Day	DNN			CASE04		
		Accuracy	POD	FAR	Accuracy	POD	FAR
Julian	D+0	84	72	11	74	62	26
	D+1	81	51	14	77	66	43
	D+2	76	34	25	76	49	50
WA	D+0	83	75	20	73	59	26
	D+1	83	72	38	78	63	43
	D+2	76	70	48	76	46	52

기존의 예보모델인 수치모델 CASE04와 선행연구인 Julian 모델과 성능을 상호 비교했다. Julian 모델의 경우 지수적증률과 오경보율은 전 예보구간에 걸쳐 우수한 결과를 보이고 수치모델인 CASE04와 비교해도 뛰어난 성능을 보이거나 감지확률의 경우 D+1이상의 예보구간에서 성능이 크게 저하된다. 이것은 예보의 특성상 시간이 멀어질수록 성능이 떨어지는 경향이 있고, 특히 저농도에 비해서 사례가 적은 고농도 예보인 경우에는 그 경향이 더욱 심화된다. 이를 개선하기 위해서 고농도 PM₁₀의 생성에 영향을 미치는 중국 및 서울주변의 자료를 입력에 추가하여 개발한 모델이 WA이다. WA는 모든 기준에서 수치모델인 CASE04보다 좋은 결과를 보이며, 지수적증률은 Julian과 비슷하나 예보의 가장 중요한 구간인 D+1에서 2% 향상된 결과를 나타낸다. POD의 경우 예보 전구간이 70%이상으로 Julian 모델에 비해 예보구간별로 각각 3%, 21%, 36%의 큰 향상을 보인다. 하지만, 과 평가된 사례들로 인해서 Julian 모델에 비해서 오경보율이 다소 증가되었다.

Table 9에는 WA모델에서 크게 개선된 감지확률의 구체적인 사례와 성능을 기술했다. Table 9의 (a)-(c)는 평가기간의 고농도 발생 건 중 2019년에

해당하는 자료를 나타낸 것이다. Date는 고농도 발생 일이고, OBS는 관측 값, Julian과 WA는 각각 Julian 예보 결과, WA 예보결과이고, CASE04는 수치모델의 예보결과이다. T06~T15의 예보구간 중 지면관계상 각 일의 첫 시간 구간인 T06, T08, T12만을 기술했다. (d)-(f)는 모델에 대한 통계 결과로 Hit count는 고농도 적중 개수를 나타내는 것으로 예를 들어, (d)의 WA 열에서 '32/38'의 의미는 38건의 고농도 건수 중 32건을 적중시켰다는 의미이다. 결과적으로 WA 모델이 고농도 적중건수와 RMSE 모두 수치모델이나 Julian에 비해 우수한 결과를 보임을 알 수 있다.

Fig. 7~Fig. 10 에는 평가기간인 2018년 01월 01일부터 2019년 3월 12일까지의 Julian모델과 WA모델의 시계열 그래프와 산포도를 기존 예보모델인 수치모델 CASE04와 상호 비교하여 나타냈다. Fig. 7과 Fig. 8을 보면 Julian모델과 WA모델 모두 수치모델에 비해서 관측 값의 추이를 잘 따라가는 알 수 있다. 그러나 Julian 모델의 경우 고농도 사례에 있어서는 적중률이 떨어지는 경향을 보이는데, WA 모델은 이런 점이 개선된 것을 알 수 있으며 특히, D+1이상의 구간의 고농도 적중률이 향상된 것을 볼 수 있다.

Table 9. Improvement of POD by WA model

T06					T08					T12				
Date	OBS	Julian	WA	CASE 04	Date	OBS	Julian	WA	CASE 04	Date	OBS	Julian	WA	CASE 04
20190111	91	100	99	128	20190104	107	52	88	42	20190103	107	50	70	38
20190112	100	121	108	102	20190111	92	71	101	82	20190110	92	76	83	114
20190113	115	97	106	79	20190112	98	75	125	93	20190111	98	67	119	90
20190114	180	158	148	180	20190113	135	96	146	169	20190112	135	85	133	157
20190115	114	131	133	74	20190114	180	148	162	180	20190113	180	84	149	132
20190119	93	109	111	90	20190118	99	61	86	93	20190117	99	81	130	100
20190128	113	83	88	60	20190119	107	83	105	93	20190118	107	73	77	77
20190205	120	62	91	90	20190122	85	56	85	79	20190121	85	62	76	75
20190206	92	82	110	69	20190203	120	104	87	66	20190202	120	72	119	65
20190227	86	60	74	65	20190205	107	48	72	82	20190204	107	52	58	71
20190228	116	80	102	82	20190206	112	65	109	77	20190205	112	66	132	72
20190301	126	88	106	61	20190224	92	51	85	69	20190223	92	74	73	70
20190302	105	88	115	67	20190228	104	60	118	82	20190227	104	80	107	82
20190305	180	159	145	121	20190301	131	89	132	76	20190228	131	79	119	62
20190306	143	128	136	85	20190302	107	79	103	65	20190301	107	73	62	73

(a) Forecast result(T06)

(b) Forecast result(T08)

(c) Forecast result(T12)

Method	Julian	WA	CASE 04
Hit count	27/38	32/38	22/38
RMSE	29.6	23.2	40.4

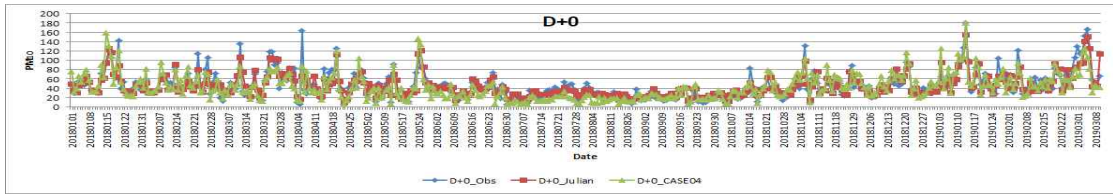
(d) Statistics(T06)

Method	Julian	WA	CASE 04
Hit count	14/37	28/37	25/37
RMSE	32.6	19.9	29.0

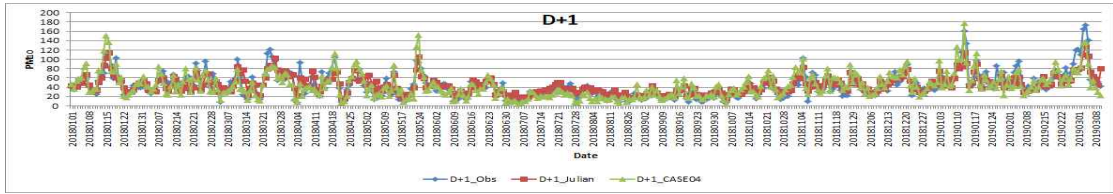
(e) Statistics(T08)

Method	Julian	WA	CASE 04
Hit count	10/35	25/35	18/35
RMSE	40.5	30.2	34.3

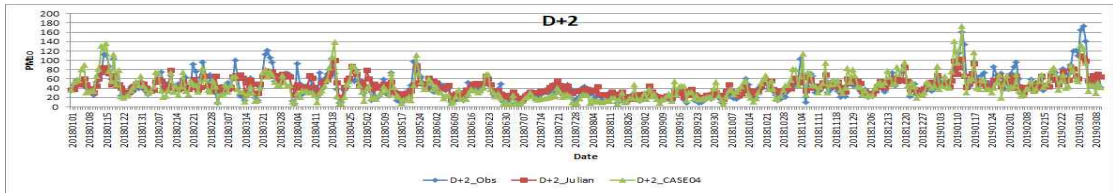
(f) Statistics(T12)



(a) Julian (D+0)

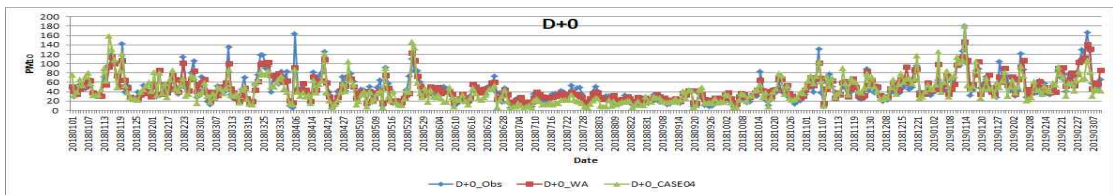


(b) Julian (D+1)

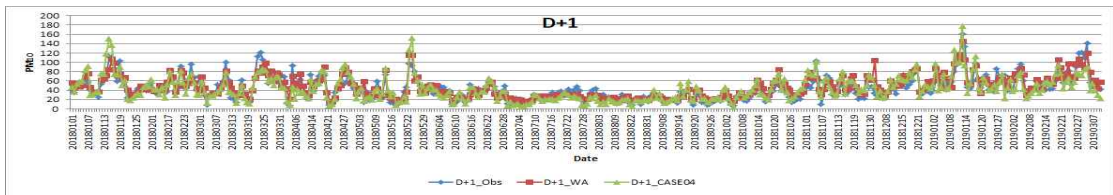


(c) Julian (D+2)

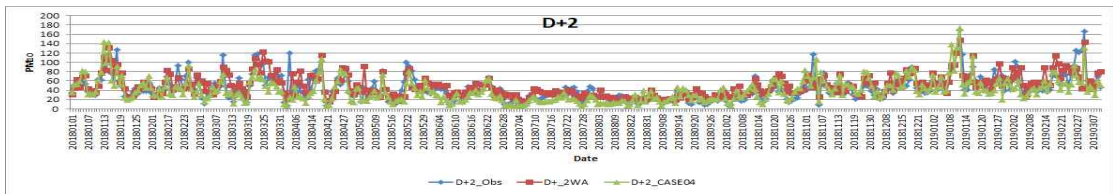
Fig. 7. Time series for Julian (Seoul region).



(a) WA (D+0)

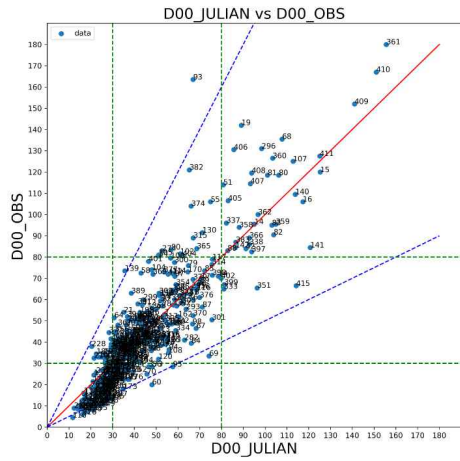


(b) WA (D+1)

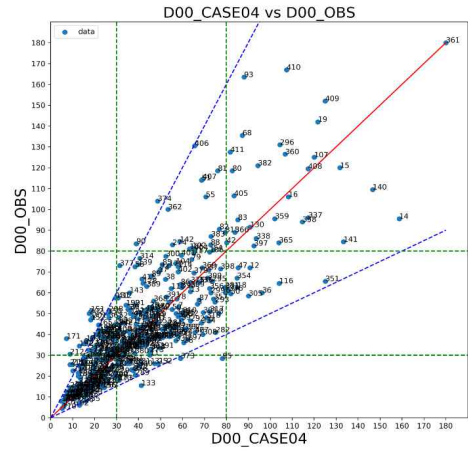


(c) WA (D+2)

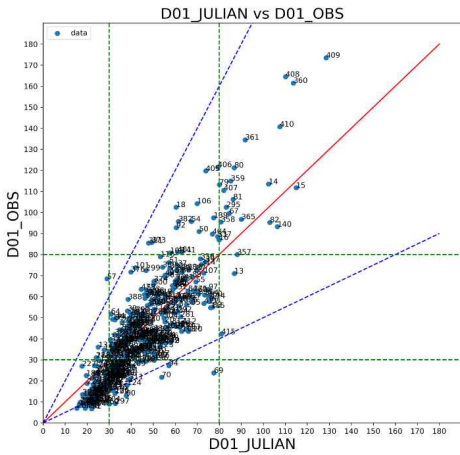
Fig. 8. Time series for WA (Seoul region).



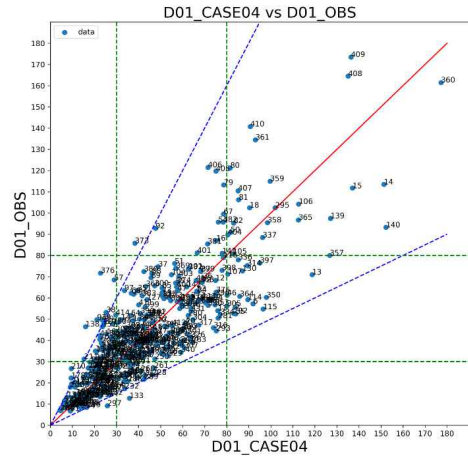
(a) Julian(D+0)



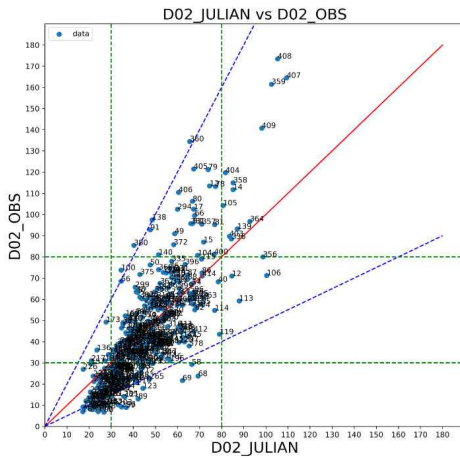
(b) CASE04(D+0)



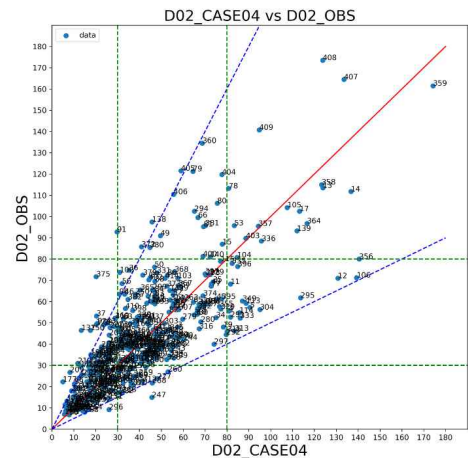
(c) Julian(D+1)



(d) CASE04(D+1)

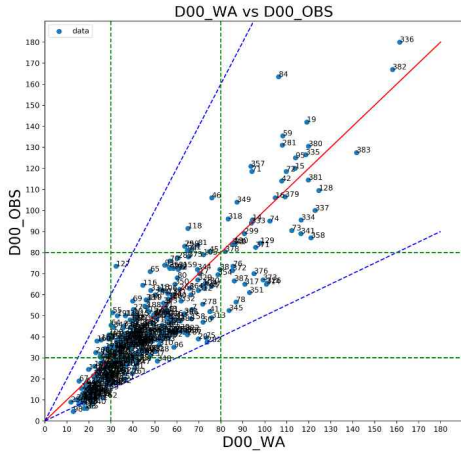


(e) Julian(D+2)

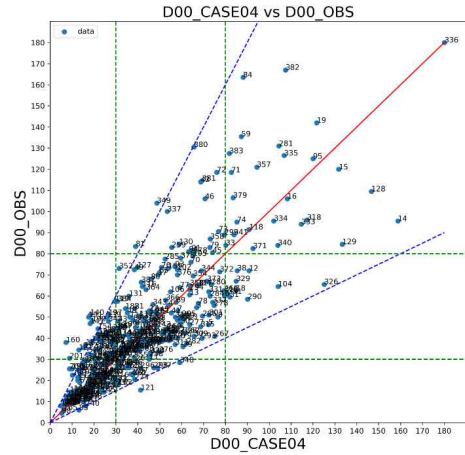


(f) CASE04(D+2)

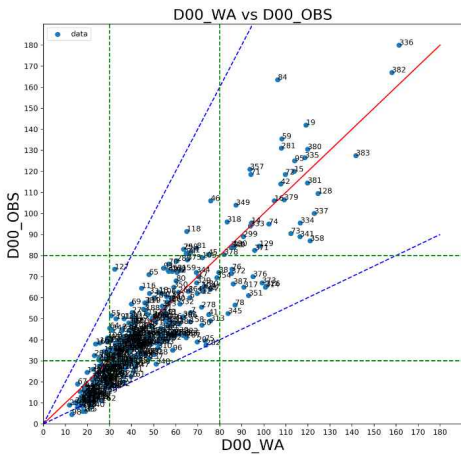
Fig. 9. Scatter of Julian (Seoul region).



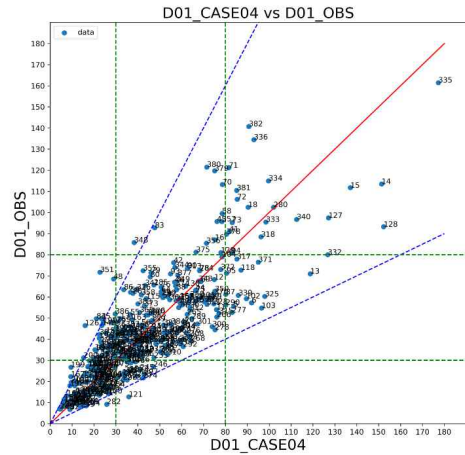
(a) WA(D+0)



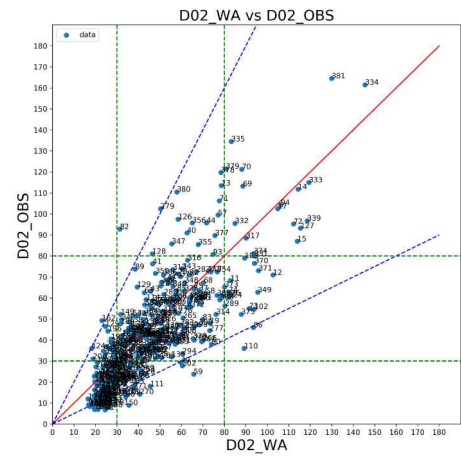
(b) CASE04(D+0)



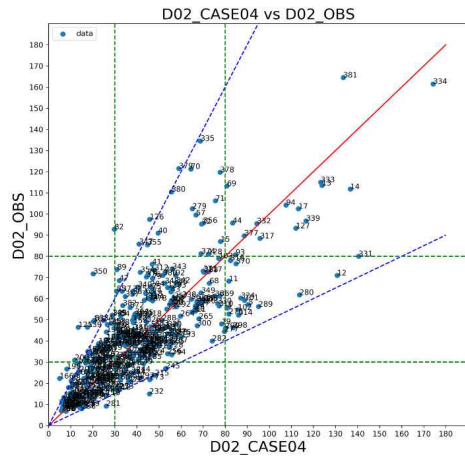
(c) WA(D+1)



(d) CASE04(D+1)



(e) WA(D+2)



(f) CASE04(D+2)

Fig. 10. Scatter of WA (Seoul region).

Fig. 9과 Fig. 10은 관측 값과 예보모델간의 산포도를 나타낸 것으로 x축은 관측 값, y축은 각 예보모델의 예보 값을 나타낸다. 시계열 그래프에서 보인 특징과 유사하게 Julian모델은 지수적증률은 좋으나 고농도 적중률이 떨어지는 모습을 볼 수 있으며, WA 모델에서는 고농도 사례에 대해 잘 적중하는 것을 알 수 있다.

5. 결 론

본 논문에서는 서울 권역 미세먼지의 3일 예보(D+0, D+1, D+2)를 위해서 WA(Wide Area) 예보모델을 개발하고, 기존 예보모델인 수치모델 CASE04와 선행 연구인 Julian 예보 모델과 상호 비교하였다. Julian 예보모델은 서울 권역 예보를 위해서 서울 권역의 자료만을 사용한 모델로 지수적증률과 오경보율은 우수했으나, D+1, D+2 구간에서 감지확률의 성능이 떨어졌다. 이점을 개선하기 위해 본 논문에서는 고농도의 발생 및 생성에 밀접한 영향을 미치는 중국 권역을 비롯한 서울 주변 권역의 자료를 추가적으로 학습에 사용하여 WA 모델을 개발했다. 그 결과 WA의 지수적증률은 83%(D+0), 83%(D+1), 76%(D+2)이고, 감지확률이 75%(D+0), 72%(D+1), 70%(D+2)으로 크게 성능이 향상되었고, 오경보율은 20%(D+0), 38%(D+1), 48%(D+2)로 Julian 모델에 비해서 증가되었다.

이는 학습에 추가한 동아시아 자료들이 고농도 예보에 적합한 학습 인자로 적용되어 고농도 적중률이 향상된 것으로 분석되며, 현재 중국 주요 공업지대에서 발생한 미세먼지의 국내 영향을 고려할 때, WA는 고농도 예보에 적합한 예보 모델임을 알 수 있다.

향후 고농도 예보의 성능은 더욱 향상시키고, 오경보율은 낮추기 위해서 overfitting 개선 연구와 중국과 국내의 미세먼지 이동에 따른 시간 지연성을 고려한 학습 인자 선정이 필요할 것으로 보인다.

REFERENCE

[1] NIER, *A Study of Construction of Air Quality Forecasting System Using Artificial Intelligence(I)*, NIER-SP2017-148, 11-1480523-000 3221-01, 2017.
 [2] S. Lee, C. Ho, and Y. Choi, "Hig-PM10 Con-

centration Episodes in Seoul, Korea: Background Sources and Related Meteorological Conditions," *Atmospheric Environment*, Vol. 45, No. 39, pp. 7240-7247, 2011.
 [3] S. Lee, C. Ho, Y. Lee, H. Choi, and C. Song, "Influence of Transboundary Air Pollutants for China on The High PM10 Episode in Seoul, Korea for the Period October 16-20, 2008," *Atmospheric Environment*, Vol. 77, pp. 430-439, 2013.
 [4] H. Oh, C. Ho, J. Kim, D. Chen, S. Lee, Y. Choi, et al., "Long-range Transport of Air Pollutants Originating in China: Apossible Major Cause of Multi-day High-PM10 Episodes During Cold Season in Seoul, Korea," *Atmospheric Environment*, Vol. 109, pp. 23-30, 2015.
 [5] L.G. McKendry, "Evaluation of Artificial Neural Networks for Fine Particulate Pollution(PM10 and PM2.5) Forecasting," *Journal of the Air and Waste Management Association*, Vol. 52, pp. 1096-1101, 2002.
 [6] D. Voukantsis, K. Karatzas, J. Kukkonen, T. Rasanen, A. Karppinen, and M. Kolehmainen, "Intercomparsion of Air Quality Data Using Principal Component Analysis, and Forecasting of PM10 and PM2.5 Concentrations Using Artificial Neural Network, in Thessaloniki and Helsinki," *Science of the Total Environment*, Vol. 409, Issue 9, pp. 1266-1276, 2011.
 [7] H. Zhang, Y. Liu, R. Shi, and Q. Yao, "Evaluation of PM10 Forecasting Based on th Aritficial Neurla Network Model and Intake Fraction in an Urban Area: A Case Study in Taiyuan City, China," *Journal of the Air and Waste Management Association*, Vol. 63, Issue 7, pp. 755-763, 2013.
 [8] S. Thomas and R.B. Jacko, "Model for Forecasting Expressway Fine ParticulateMatter and Carbon Monoxide Concentration: Application of Regression and Neural Network Models," *Journal of the Air and Waste Management Association*, Vol. 58, Issue 4, pp.

480-488, 2012.

[9] F. Franceschi, M. Cobo, and M. Figueredo, "Discovering Relationships and Forecasting PM10 and PM2.5 Concentrations in Bogota, Colombia, Using Artificial Neural Networks, Principal Component Analysis, and K-mean Clustering," *Atmospheric Pollution Research*, Vol. 9, Issue 5, pp. 912-922, 2018.

[10] S. Park, M. Kim, M. Kim, H. Namgung, K. Kim, K. Cho, et al., "Predicting PM10 Concentration in Seoul Metropolitan Subway Stations Using Artificial Neural Network (ANN)," *Journal of Hazardous Materials*, Vol. 341, pp. 75-82, 2018.

[11] G.D. Gennaro, L. Trizio, A.D. Gilio, J. Pey, N. Perez, M. Cusack, et al., "Neural Network Model for The Prediction of PM10 Daily Concentrations in Two Sites in The Western Mediterranean," *Science of The Total Environment*, Vol. 463-464, pp. 875-883, 2013.

[12] Y. Bai, Y. Li, X. Wang, J. Xie, and C. Li, "Air Pollutants Concentrations Forecasting Using Back Propagation Neural Network Based on Wavelet Decomposition with Meteorological Condition," *Atmospheric Pollution Research*, Vol. 7, Issue 3, pp. 557-566, 2016.

[13] X. Feng, Q. Li, J. Hou, L. Jin, and J. Wang, "Artificial Neural Networks Forecasting of PM2.5 Pollution Using Air Mass Trajectory Based Geographic Model and Wavelet Transformation," *Atmospheric Environment*, Vol. 107, pp. 118-128, 2015.

[14] B.S. Freeman, G. Taylor, B. Gharabaghi, and J. The, "Forecasting Air Quality Time Series Using Deep Learning," *Journal of the Air and Waste Management Association*, Vol. 68, No. 8, pp. 866-886, 2018.

[15] F. Biancofiore, M. Busilacchio, M. Verdecchia, B. Tomassetti, E. Aruffo, et al., "Recursive Neural Network Model for Analysis and

Forecast of PM10 and PM2.5," *Atmospheric Pollution Research*, Vol. 8, Issue 4, pp. 652-659, 2017.

[16] W. Lu, W. Wang, X. Wang, S. Yan, and J.C. Lam, "Potential Assessment of A Neural Network Model with PCA/RBF Approach for Forecasting Pollutant Trends in Mong Kok Urban Air, Hong Kong," *Environmental Research*, Vol. 96, Issue 1, pp. 79-87, 2004.

[17] J. Fan, Q. Li, J. Hou, X. Feng, H. Karimian, and S. Lin, "A Spatiotemporal Prediction Framework for Air Pollution Based on Deep RNN," *Proceeding of ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume IV-4/W2, 2017 2nd International Symposium on Spatiotemporal Computing*, pp. 15-22, 2017.

[18] S. Yu, Y. Koo, and H. Kwon, "Inverse Model Parameter Estimation Based on Sensitivity Analysis for Improvement of PM10 Forecasting," *Journal of Korea Multimedia Society*, Vol. 18, No. 7, pp. 886-894, 2015.

[19] H. Bae, S. Yu, and H. Kwon, "Fast Data Assimilation Using Kernel Tridiagonal Sparse Matrix for Performance Improvement of Air Quality Forecasting," *Journal of Korea Multimedia Society*, Vol. 20, No. 2, pp. 363-370, 2017.



유 숙 현

1999년 안양대학교 컴퓨터공학과
학사
2002년 안양대학교 컴퓨터공학과
석사
2011년 안양대학교 컴퓨터공학과
박사

2012년~현재 안양대학교 정보통신공학과 조교수
관심분야: 패턴인식, 신경망, 영상처리, 병렬처리응용,
딥러닝