

ANN(Artificial Neural Network)를 사용한 수원시 미세먼지 농도 예측

Using ANN(Artificial Neural Network) to Predict Particulate Matter Concentration in Suwon

강은하

연구진

연구책임자 강은하 (수원시정연구원 연구위원)
참여연구원 오미현 (수원시정연구원 과제연구원)

연구 자문위원

권순박 (한국철도기술연구원 책임연구원)

© 2017 수원시정연구원

발행인 이재은

발행처 수원시정연구원

경기도 수원시 권선구 수인로 126

(우편번호) 16429

전화 031-220-8001 팩스 031-220-8000

<http://www.suwon.re.kr>

인쇄 2017년 10월 31일

발행 2017년 10월 31일

I S B N 979-11-87778-73-8 (93530)

이 보고서를 인용 및 활용 시 아래와 같이 출처 표시해 주십시오.

강은하. 2017. 「ANN(Artificial Neural Network)를 사용한 수원시 미세먼지 농도 예측」.
수원시정연구원.

비매품

국문 요약

본 연구에서는 수원시 국지배출원(예: 차량 통행량 등)과 대기오염물질이 미세먼지 농도 증가에의 얼마나 상관성이 있는지를 알아보기 위한 목적으로 대기오염물질 간의, 그리고 미세먼지 증가 영향요인 간의 상관관계를 알아보는 연구를 수행하였다. 복잡한 대기오염물질의 생성, 이동, 소멸 등을 다 고려하는 기존의 메카니즘을 사용한 모델링이 아닌 통계학적 접근 방법과 인공신경망(ANN: Artificial Neural Network) 기법으로 공공데이터로 구축되어 있는 자료를 활용하여 미세먼지 농도 예측을 시도하였다. 다만, 제한된 연구기간 및 예산으로 인해 2016년 당해 연도에 대한 수원시 내 기 측정되고 있는 대기오염물질 측정 자료와 교통량, 기상 자료를 제한적으로 수집하였고, 인공신경망 기법을 도시 미세먼지 농도 예측 모델에 적용하고자 하는 기초 연구적 성격으로 추진되었다.

PM₁₀, PM_{2.5} 농도에 영향을 주는 다른 대기환경기준 오염물질, 기상요소, 인근 지점으로 교통량과의 관계를 다중선형회귀분석을 통해 알아보고, 통계적으로 유의한 자료를 사용하여 인공신경망기법 중 랜덤포레스트 방식을 적용한 PM₁₀ 농도 예측 모델을 구축하였다. 그 결과 예측값과 실측값이 $R^2=0.77$ 로 잘 예측하는 결과를 얻었다. 그러나 인근 지점 도로교통량은 분석결과 통계적으로 유의한 영향을 주지 못했다. 그 원인으로서는 PM_{2.5} 결측으로 모델에 PM₁₀만을 사용할 수밖에 없었고, 측정소 인근의 차량 통행량의 1년 동안 시간대별 패턴 및 교통량의 변화가 적어 모델에 변수로 작용하기 어려웠던 점, 미세먼지 배출원과 이차 생성 기작으로 생성된 PM_{2.5}, PM₁₀과의 시간차가 반영되지 못하는 점 등을 들 수 있다. 미세먼지는 배출원에서 직접 배출되기 보다는 가스상 오염물질로부터 오랜 시간에 걸쳐 생성되고 변환되는 최종산물이기 때문에 도로의 자동차 배출가스가 대기 중에 배출된 후 1~2일의 시간을 거쳐 미세먼지로 변환되는 것과 같은 오염물질간의 시간차가 이 모델에 반영되지 못하는 것도 원인이 될 수 있을 것이다.

PM₁₀ 예측에서 가중치가 가장 높았던 CO를 활용하여 CO를 10% 낮출 경우 PM₁₀이 4.5% 저감된다는 결과를 통해 PM₁₀ 저감을 위해서는 일차 배출원 지시자로서의 CO 농도를 저감하는 방식의 접근이 필요할 것이다. 후속 연구를 통해 CO와 수원시 도로 자동차 통행량과의 관계, 나아가 인근 고속도로 통행량과의 관계를 찾아서 뒷받침하는 것이 필요할 것이다.

따라서 본 연구에서 구축된 모델에 PM_{2.5} 자료를 더 추가하여 분석한다면, 측정소 운영 중 생긴 결측치에 대해 예측값으로 임시 대치가 가능할 수 있을 것으로 보인다.

또한, 향후 모델의 예측 정확도를 향상하기 위해서는 효과적인 데이터 전처리 법이 추가로 연구되어야 할 것이다. 랜덤 포레스트를 이용한 미세먼지 예측 모델은 과거 자료 학습을 통해 예측을 하기 때문에 자료의 양이 많을수록 예측 정확도가 높아지게 되므로 후속 연구에서는 보다 많은 데이터를 수집하여 모델에 적용해야 한다.

후속연구로 대기 오염물질이 대기 중에 잘 혼합되는 특성을 반영하여 모델 입력 자료의 공간적 규모를 확대하여 수원시나 경기 남부권 도로교통량을 대표할 수 있는 교통량 자료와 수원시 또는 수도권 남부 대표 대기오염도 자료를 활용한 모델링도 필요할 것이다. 또는 수원시와 안산, 성남, 화성, 서울 등 인접 도시들, 백령도와 같은 배경대기 측정소의 대표 측정 자료를 사용한 모델링을 통해, 외부 유입 오염원으로 인한 수원시 미세먼지 농도 증가 기여도를 판단하는 연구도 필요할 것이다.

주제어 : 미세먼지, 인공신경망(ANN), 미세먼지 농도 예측, 대기질 개선

| 차례 |

제1장 서론	1
제1절 연구의 배경 및 목적	1
1. 연구의 배경 및 목적	1
제2절 연구의 범위 및 방법	3
1. 연구범위	3
2. 연구방법	3
제2장 ANN 모델 선행연구 고찰	5
제1절 ANN 모델 개념 고찰	5
1. 대기오염도에 영향을 주는 인자	5
2. 대기오염모델링을 통한 대기질 예측	6
3. 인공신경망(Artificial Neural Network, 이하 ANN) 개념	11
제2절 ANN을 이용한 대기질 예측 선행연구	15
1. 대기과학에서의 인공신경망(ANN) 적용	15
2. 공기이동궤적을 이용한 PM2.5의 ANN예측	16
3. 미세먼지 예측을 위한 통계적 방법의 개발 사례	18
4. 인공지능 기반 수도권 지하역사 실내공기질 예측 사례 고찰	19
5. 인공신경망 기법을 이용한 태풍 강도 및 진로 예측	25
제3절 시사점	29
제3장 수원시 대기질 현황	31
제1절 수원시 대기오염농도	31
1. 수원시 주변지역과의 비교	31
2. 수원시 국가대기오염측정소 현황	33
3. 수원시 대기오염농도	34
4. 수원시 대기오염물질 상관성 분석	41
5. 수원시 미세먼지 배출량	44
제2절 시사점	46

제4장 ANN을 활용한 수원시 대기질 분석	47
제1절 ANN을 활용한 수원시 미세먼지 농도 예측	47
1. ANN 기법	47
2. 분석을 위한 자료구축 및 유사 활용 사례	47
제2절 대기질 영향요인과 대기오염물질에 따른 미세먼지 농도 예측	55
1. 분석 자료 및 방법	55
2. 분석결과	57
제3절 결과 고찰	70
제5장 결론	71
제1절 연구의 의의	71
제2절 정책제언	73
참고문헌	75
Abstract	77

| 표 차 례 |

<표 3-1> 수원시 국가대기오염측정소 현황	33
<표 4-1> 도로상 이동측정 결과	51
<표 4-2> 도로 이격거리별 측정결과	52
<표 4-3> 교통량과 대기오염물질 간의 상관성 분석	53
<표 4-4> 자동차 규모별 세부기준	57
<표 4-5> 변수들 간의 상관계수	62
<표 4-6> 회귀분석 결과(PM_{10})	63
<표 4-7> 회귀분석 결과($PM_{2.5}$)	63
<표 4-8> 변수들의 기본적인 통계적 특성	64
<표 4-9> PM_{10} 과 변수들과의 관계	67
<표 4-10> 모델링 결과	68

| 그림 차례 |

<그림 1-1> 수도권 대기오염농도 추이(미세먼지, 오존)	1
<그림 2-1> 미세먼지의 발생원 및 생성과정	6
<그림 2-2> Interrelated origins of ozone, fine particulate matter (PM _{2.5}), and haze ..	7
<그림 2-3> Chemical links between the ozone and PM formation processes	8
<그림 2-4> 이차생성 유기 에어로솔의 모델과 실측값 비교	8
<그림 2-5> 미세먼지 예보시스템	9
<그림 2-6> Source contributions to PM _{2.5} measured for 3 years in Seoul	11
<그림 2-7> 인공신경망 구조	12
<그림 2-8> 인공지능과 머신 러닝, 딥러닝의 차이	13
<그림 2-9> 딥러닝이 문제를 해결하는 방식	14
<그림 2-10> SO ₂ emissions with SOFM	15
<그림 2-11> 인근지역의 PM _{2.5} 농도 분포	16
<그림 2-12> MLP 타입의 인공신경망 구조	16
<그림 2-13> 예측모델 결과(붉은점:학습자료, 검은색:모델결과)	17
<그림 2-14> 다중선형회귀모형을 통해 모의된 PM ₁₀ 의 농도와 관측농도의 상관성 ..	18
<그림 2-15> ANN통해 모의된 PM ₁₀ 의 농도와 관측농도의 상관성	18
<그림 2-16> 지하역사 공조관리시스템	19
<그림 2-17> 외기, 급기구, 대합실, 승강장 데이터간의 관계 모식도	20
<그림 2-18> 환기실/대합실/승강장 PM ₁₀ 예측값과 실측값 비교(11-12월), 주황색 선: 1 시간 후 실측값, 푸른색 선: 1시간 후 예측값	21
<그림 2-19> 환기실/대합실/승강장 PM ₁₀ 예측값과 실측값 산점도(11-12월)	22
<그림 2-20> 활용된 입력변수(input)와 인공신경망 모식도	23
<그림 2-21> 인공신경망을 활용한 대합실/승강장 PM ₁₀ 연산값과 실측값 산점도(11-12월)	23
<그림 2-22> 예측된 PM ₁₀ 과 실측된 PM ₁₀ 비교	24
<그림 2-23> 공조시스템 급배기 변동제어 알고리즘	25
<그림 2-24> 5일 동안 12시간 간격으로 7개의 인공신경망 모델(ANNM)과 7개의 다중선형 회귀모델(MLRM)에 의해 예측된 태풍 강도변화의 평균절대오차(MAE) 비교	27
<그림 2-25> 인공신경망 모델과 다중선형회귀모델의 진로예측 비교	28
<그림 3-1> 수원시 주변 지역	31

<그림 3-2> 2016년 수원 바람장미	31
<그림 3-3> 수원시 주변지역 SO ₂ 와 NO ₂ 농도 비교	32
<그림 3-4> 수원시 주변지역 O ₃ 와 CO 농도 비교	32
<그림 3-5> 수원시 주변지역과의 미세먼지 농도 비교	32
<그림 3-6> 수원시 대기오염측정소 위치	33
<그림 3-7> 수원시 PM ₁₀ 농도 변화 추이(2010.01~2016.12)	34
<그림 3-8> 수원시 PM ₁₀ 연도별 변화	35
<그림 3-9> 수원시 PM _{2.5} 농도 변화 추이(2015.01~2016.12)	35
<그림 3-10> 수원시 SO ₂ 농도 변화 추이(2010.01~2016.12)	36
<그림 3-11> 수원시 SO ₂ 연도별 변화	36
<그림 3-12> 수원시 NO ₂ 농도 변화 추이(2010.01~2016.12)	37
<그림 3-13> 수원시 NO ₂ 연도별 변화	37
<그림 3-14> O ₃ -NO _x 광분해 cycle	38
<그림 3-15> 수원시 O ₃ 농도 변화 추이(2010.01~2016.12)	39
<그림 3-16> 수원시 O ₃ 연도별 변화	39
<그림 3-17> 수원시 CO 농도 변화 추이(2010.01~2016.12)	40
<그림 3-18> 수원시 CO 연도별 변화	40
<그림 3-19> PM _{2.5} 와 PM ₁₀ 상관성 분석(2013~2016)	41
<그림 3-20> PM ₁₀ 과 가스상 오염물질 간의 상관성 분석(2016)	42
<그림 3-21> PM _{2.5} 와 가스상 오염물질 간의 상관성 분석(2016)	42
<그림 3-22> 미세먼지(PM ₁₀ , PM _{2.5})와 온·습도 간의 상관성 분석(2016)	43
<그림 3-23> CO와 NO ₂ , O ₃ 와 NO ₂ 의 상관성 분석(2016)	43
<그림 3-24> 수원시 미세먼지 배출량 추세	44
<그림 3-25> 수원시 부문별 PM ₁₀ 배출비율	44
<그림 4-1> 랜덤포레스트 시각화	47
<그림 4-2> 기온역전현상	48
<그림 4-3> 2005년 WS & PM ₁₀	49
<그림 4-4> 서울 미세먼지 농도와 풍속의 연변화	49
<그림 4-5> PM ₁₀ 농도의 시계열(전국, 수도권, 서울)	50
<그림 4-6> 연간 평균PM ₁₀ (%)	50
<그림 4-7> 10m 풍속 변화	50

<그림 4-8> PM농도와 풍속의 회귀분석	50
<그림 4-9> 분석 기간 동안의 시간대별 평균 교통량 및 시간평균 PM ₁₀ 농도(좌), NO ₂ 농도(우)	52
<그림 4-10> 서울역 지점에서의 일일 교통량과 일평균 PM ₁₀ (좌), NO ₂ (우) 농도의 상관관계	53
<그림 4-11> 교통량 자료 위치	54
<그림 4-12> 대기오염측정망과 교통량 자료 측정 위치	55
<그림 4-13> 활용된 변수들의 Raw Data	56
<그림 4-14> PM ₁₀ · PM _{2.5} 농도와 풍속의 변화 추이(2016년 8월~12월)	58
<그림 4-15> 풍속과 미세먼지농도의 상관성	58
<그림 4-16> 습도와 미세먼지농도의 상관성	58
<그림 4-17> 교통량과 미세먼지농도(PM ₁₀ , PM _{2.5})의 상관성	59
<그림 4-18> 대형차 교통량과 미세먼지농도(PM ₁₀ , PM _{2.5})의 상관성	59
<그림 4-19> 교통량과 NO ₂ , O ₃ 농도의 상관성	60
<그림 4-19> 교통량과 NO ₂ +O ₃ 농도의 상관성	60
<그림 4-20> 2016년 수원시의 시간대별 교통량 평균 추이	61
<그림 4-22> 랜덤 포레스트 함수를 이용한 모델 구축 과정	65
<그림 4-23> 예측값과 실측값의 분포도	66
<그림 4-24> 예측값과 실제값의 산점도	66
<그림 4-25> 예측값과 실제값의 비교	67
<그림 4-26> 1년간 CO 농도 변화 패턴	68
<그림 4-27> 2가지 분석 경우 비교 그래프, 검은색: CO 변화 없는 예측값, 주황색: CO를 50% 줄인 후 예측값	69

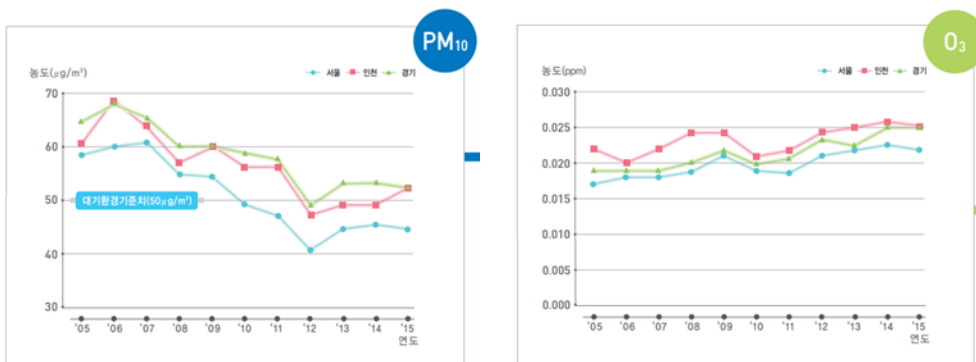
제1장 서론

제1절 연구의 배경 및 목적

1. 연구의 배경 및 목적

미세먼지는 1급 발암물질로 도시 시민 건강을 악화시키는 대기오염물질로 미세먼지 발생원과 메카니즘이 매우 복잡하여 많은 연구가 있음에도 불구하고 효과적인 개선책을 찾지 못하고 있다. 장거리 이동 및 대기 중 체류시간이 긴 미세먼지에 대해서는 한반도 전역, 또는 수도권 내에 걸친 광역적 규모의 연구가 많이 진행되고 있다. 그러나 시민이 체감하는 미세먼지 농도는 집 앞, 근무지 앞, 학교 앞 등의 농도로 국지 배출원(주변 도로의 차량 통행량, 주변 사업체 등)의 영향을 크게 받고 있다.

국내 미세먼지 오염도는 PM₁₀의 변동추이로 볼 때 2000년대 이후 지속적으로 개선되는 추세이나, 2013년부터 정체되고 국민들이 실생활에서 체감하는 오염도는 높아지는 상황이다. 오존은 2005년 이후 지속적인 증가추세를 보여주는데, 가스상 오염물질로부터 광화학 반응에 의해 생성된다는 점에서 PM_{2.5} 중 이차생성 에어러솔(Secondary aerosol)과 발생원이 유사하다는 점에서 주목할 만하다.



자료: 수도권대기환경정보고서, 우리동네대기질2015

<그림 1-1> 수도권 대기오염농도 추이(미세먼지, 오존)

따라서 수원시 내 국지배출원이 수원시 미세먼지 농도에 어떻게 기여하는지를 밝힐 수 있는가에 대한 연구를 통해 자체적으로 저감할 수 있는 방안은 무엇인지를 알아보고자 한다. 이를 위해 수원시 국지 배출원(예: 차량 통행량 등)과 대기오염물질이 미세먼지 농도 증가에

의 얼마나 상관성이 있는지를 알아보기 위해 기존의 메카니즘을 사용한 모델링이 아닌 통계학적 접근 방법과 인공신경망(ANN: Artificial Neural Network) 기법을 사용해 보고자 한다.

제2절 연구의 범위 및 방법

1. 연구범위

- 공간적 범위
 - 수원시 관내
- 시간적 범위
 - 2014~2016년 확정통계자료 보유기간
- 내용적 범위
 - ANN 모델 사용 대기오염물질 상관성 분석
 - 수원시 영향요인 평가

2. 연구방법

- 2016년도 수원시 대기오염물질 측정자료 수집 및 분석
- ANN 모델 사용 대기오염물질간의 상관성 및 수원시 영향요인(예: 도로 및 차량 이용현황 등)과 미세먼지 농도와의 상관성 분석
- 대기오염물질 저감, 도로 차량 통행량 저감을 통한 미세먼지 농도 저감 효과 예측
- 문헌고찰: 문헌고찰을 통한 ANN 모델결과 해석

제2장 ANN 모델 선행연구 고찰

제1절 ANN 모델 개념 고찰

1. 대기오염도에 영향을 주는 인자

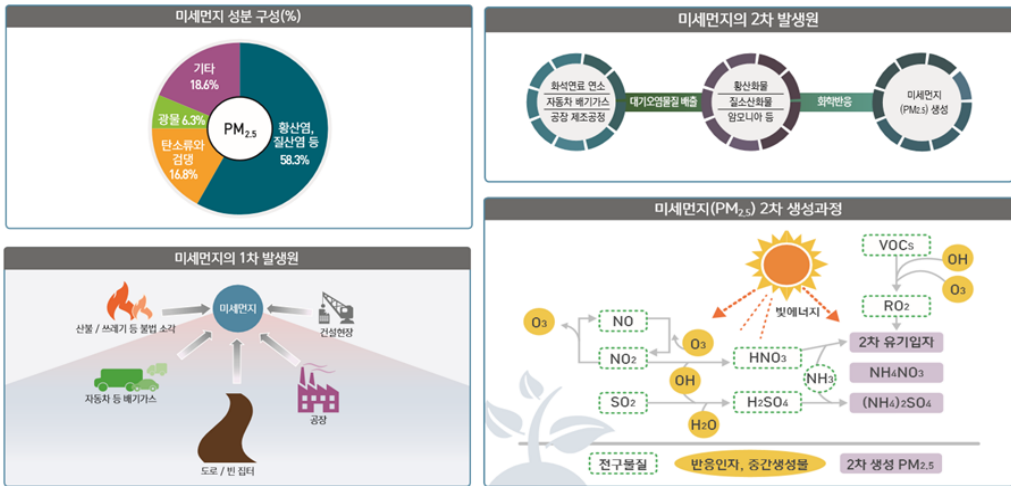
미세먼지를 이루는 성분은 그 미세먼지가 발생한 지역이나 계절, 기상조건 등에 따라 달라질 수 있다. 일반적으로는 대기오염물질이 공기 중에서 반응하여 형성된 덩어리(황산염, 질산염 등)와 석탄·석유 등 화석연료를 태우는 과정에서 발생하는 탄소류와 검댕, 지표면 흙먼지 등에서 생기는 광물 등으로 구성된다.

미세먼지는 굴뚝 등 발생원으로부터 고체 상태의 미세먼지로 나오는 경우(1차적 발생)와 발생원에서는 가스 상태로 나온 물질이 공기 중의 다른 물질과 화학반응을 일으켜 미세먼지가 되는 경우(2차적 발생)로 나누어 질 수 있다.

석탄·석유 등 화석연료가 연소되는 과정에서 배출되는 황산화물이 대기 중의 수증기, 암모니아와 결합하거나, 자동차 배기가스에서 나오는 질소산화물이 대기 중의 수증기, 오존, 암모니아 등과 결합하는 화학반응을 통해 미세먼지가 생성되기도 하는데 이것이 2차적 발생에 속한다. 2차적 발생이 중요한 이유는 수도권만 하더라도 화학반응에 의한 2차 생성 비중이 전체 미세먼지(PM_{2.5}) 발생량의 약 2/3를 차지할 만큼 매우 높기 때문이다.

대기오염물질인 휘발성 유기화합물, 질소산화물, 황산화물 등이 미세먼지로 전환되는 과정은 다음과 같다. 우선 자동차 배기가스, 주유소 유증기 등에 많이 포함된 휘발성 유기화합물(VOCs)은 반응성이 강한 물질(OH, O₃ 등)과 화학반응을 일으켜 2차 유기입자(Secondary Organic Particles)가 된다(환경부, 2016).

미세먼지(PM_{2.5}, PM₁₀ 등)는 직접 발생되거나, 광화학반응 및 물리적 기작에 의해 생성되고 변환되는 등 매우 복잡한 기작을 갖고 있다. 따라서 기존 모델링 예측값은 측정값보다 최대 1/100이하까지 과소평가하기도 한다(Volkamer et al., 2007).



자료: 환경부(2016), 바로알면 보인다. 미세먼지, 도대체 뭘까?

<그림 2-1> 미세먼지의 발생원 및 생성과정

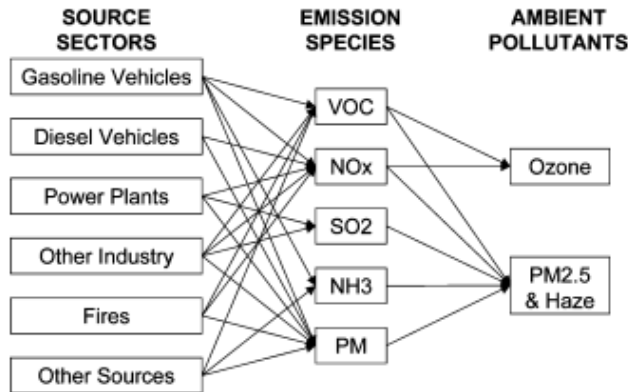
선행연구에 따르면 우리나라의 농도에 영향을 끼치는 주요 배출원으로는 국내에서 배출되는 지역 배출원과 국외에서 배출되는 외부 배출원이 있으며(Fast et al., 2007; Jones, Harrison, & Baker, 2010) 특히 외부 배출원은 한반도의 풍상측에 위치한 중국의 배출원으로 이 바람을 타고 장거리 수송되기 때문에 여러 기상조건의 영향을 받는다(Lee et al., 2001; Yi et al., 2001).

2. 대기오염모델링을 통한 대기질 예측

배출원-배출물질-대기 중 오염물질 사이 비선형(non-linear)관계가 있으며, PM_{2.5} 생성 반응의 복잡성 때문에 메커니즘을 통한 대기질 예측에는 많은 가정이 필요하며, 불확실도 또한 여전히 높다.

하지만 오존, PM_{2.5}, 연무(haze; 주로 미세먼지로 기인)는 공통적인 배출원과 전구물질(그림 2-2)을 공유하지만, 서로 다른 시기에 최종 상태(예: 오존, PM_{2.5}, 연무)에 도달하며, 지역적 경계와 함께 각 오염물질에 대한 모델링 프로토콜이 다르다. 대기오염물질은 복합혼합물에서 발생하지만 그동안의 대기질 규제 정책은 개별 오염물질만을 대상으로 제한되었다. 따라서 상대적으로 전체 대기질 규제정책은 오존과 미세먼지는 인체에의 노출이 동시에 발생하고 전구물질도 동일함에도 불구하고 생성 프로세스와 생성시간(예: 오존은 수시간 내 생성 및 24시간 내 소멸, 이차생성 미세먼지는 1~2일 내 생성, 7~10일 정도 대기 중에 존재)이 다르기 때문에, 별도의 프로세스와 별도의 시간계획을 수립하여 추진해야 한다.

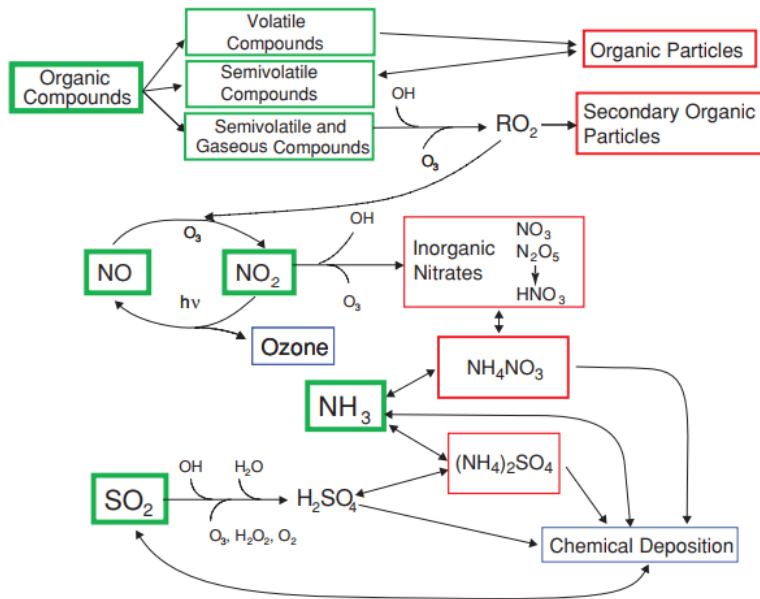
이러한 점 또한 대기질 예측의 어려움을 시사한다.



자료: Cohan et al., 2007. An integrated framework for multipollutant air quality management and its application in Georgia, Environ. Manage. 40, 545-554.

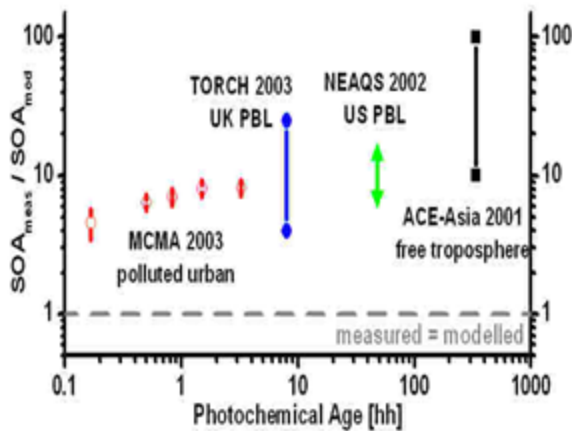
<그림 2-2> Interrelated origins of ozone, fine particulate matter(PM_{2.5}), and haze

이차생성 에어러솔의 전구물질은 <그림 2-3>의 녹색상자로 표시되어 있다. 전구물질 중 유기물질은 가스상(volatile)으로 존재하기도 하지만, 응축된 액상(non-volatile)으로 존재하기도 하며, 조건에 따라 가스상이나 액상으로 그 상태가 변화되어(semi-volatile) 존재하기도 한다. 그리고 대기 중 광화학반응이 일어날 수 있는 조건과 상대 반응물질이 존재하면 광화학반응(photo-oxidation) 또는 물리적 기작(condensation)에 의해 에어러솔 형태로 존재하게 되며, 광화학반응메카니즘 모델링을 통해 전구물질로부터의 생성되는 에어러솔 농도를 계산하는 연구가 이루어지고 있다. 하지만 광화학반응에 의한 미세먼지의 생성 및 소멸이 물리적인 기작에 의한 생성(예: nucleation)이나 응축, 휘발 과정과 결합되고, 이러한 반응은 오픈된 공간인 공기궤(air mass)에서 일어나며, 이 공기궤는 또한 기상현상에 의해 이동되기 때문에 정확한 메카니즘을 규명한다는 것 자체가 매우 불확실하다고 할 수 있다. 이러한 불확실성은 특히 미세먼지(PM_{2.5})의 이차 생성과정에서 두드러지며, 기존 모델링 예측값은 측정값보다 최대 1/100이하까지 과소평가하기도 하였다(Volkamer et al., 2007).



자료: McMurry, 2003. Particulate matter science for policy makers: a NARSTO assessment.

<그림 2-3> Chemical links between the ozone and PM formation processes



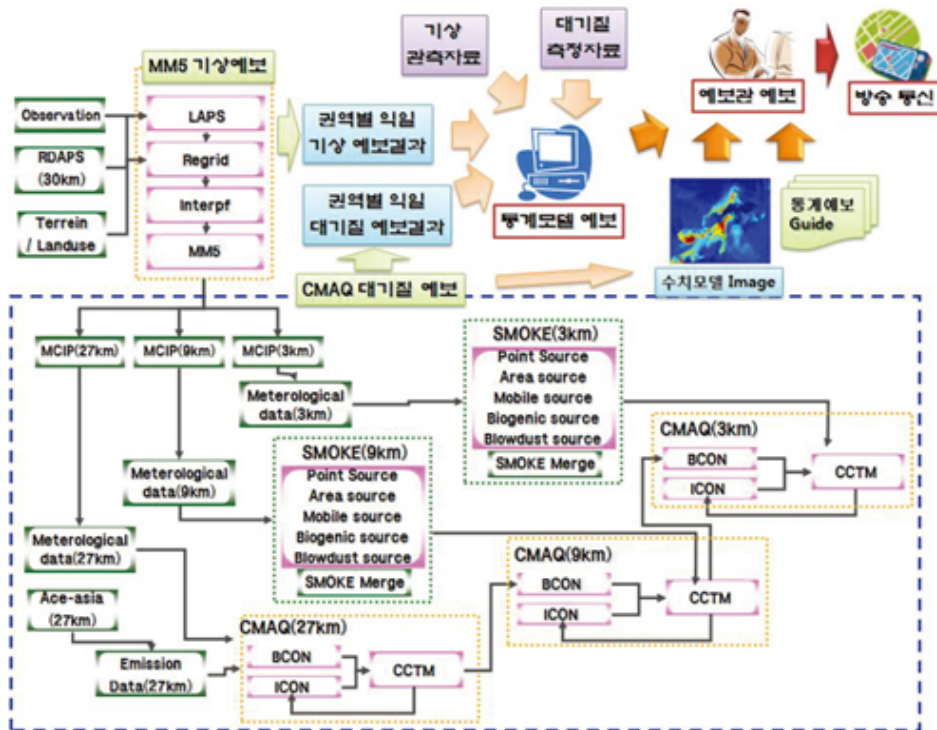
자료: Volkamer et al., 2007. Secondary organic aerosol formation from anthropogenic air pollution: Rapid and higher than expected. GEOPHYSICAL RESEARCH LETTERS, VOL. 33, L17811, doi:10.1029/2006GL026899

<그림 2-4> 이차생성 유기 에어러솔의 모델과 실측값 비교

1) 대기질 예보 모델링

대기질 예보는 대기화학모델링(CAMx, CMAQ, UAM-V 등), 기상모델(MM5, RAMS, WRF 등), 배출원 모델(SMOKE 등)을 사용하여 대기질을 예측한다.

현재 운영 중인 미세먼지 예보시스템은 통계 예보모델을 이용하여 회귀모델, 신경망모델, 의사결정모델을 이용하여 예보관이나 시스템에서 종합하여 일평균 미세먼지를 지수로 발표하고 있다. 그러나 월경성(long range transport) 오염원과 같은 급격한 대기질 변화에 대하여 예측을 할 수 없다는 단점이 있다. 이를 보완하기 위해 한국 대기질 예보시스템과 같은 수치모델과 통계모델을 통합하여 미세먼지를 예보하는 통합대기질 모델링시스템을 제작하여 운영 중이다.



자료: 2009녹색기술 리더 50선, 수도권대기환경청.

<그림 2-5> 미세먼지 예보시스템

2) 배출원 수용 모델링

인위적 오염원이 대기 중 입자상 물질에 미치는 영향을 파악하기 위해서 통계분석법을 이용한 다양한 발생원 추정방법이 개발되어왔다. 대표적으로 화학적 질량 보존법(Chemical mass balance), 인자 분석법(Factor analysis), 다변량 분석법(Multiple linear regression) 등이 있다. 이 중 에어로졸의 성분조성자료를 사용하는 CMB와 같은 수용모델법이 많이 이용되어 왔다.

CMB, PMF 모델링 등 알고 있는 대기오염원으로부터 미세먼지의 배출원 기여도 평가) 관련 연구에서는 서울에서 2008년부터 2010년까지 장기간에 걸쳐 중량농도법으로 PM_{2.5}의 질량농도를 측정하고 종합적인 성분분석(이온, 탄소, 금속)을 실시한 후, 성분분석 자료를 CMB와 PMF의 입력 자료로 활용하여 PM_{2.5}의 주요 배출원별 기여율을 산정하였다. 이때 측정결과와 불확도는 각 성분별 검출한계로부터 아래의 식으로 산출하여 PMF에 적용하였다.

$$\circ \text{Uncertainty} = 2 \times \text{MDL, if concentration} \leq \text{MDL} = \text{MDL, if concentration} > \text{MDL}$$

PMF에서 자료 행렬 X는 n개의 행과 m개의 열로 구성된다. 이때 n과 m은 각각 시료와 화학종의 수를 의미한다. 행렬 X는 분석을 통해 G(n × p)과 F(p × m)의 두 행렬로 인수 분해 되고, 잔여성분 값들은 행렬 E로 표현됨. 이때 p는 산출된 인자의 수를 나타낸다.

$$\circ X = GF + E$$

G는 p개의 발생원들의 발생원 기여농도를 의미하며, F는 발생원별 성분조성을 나타낸다. PMF는 각 측정 자료의 불확도를 기반으로 다음의 식으로 정의된 함수 Q를 최소화하는 해를 산출한다.

$$Q = \sum_{i=1}^n \sum_{j=1}^m \left(\frac{e_{ij}}{h_{ij}s_{ij}} \right)^2$$

$$e_{ij} = x_{ij} - \sum_{k=1}^p g_{ik} f_{kj}$$

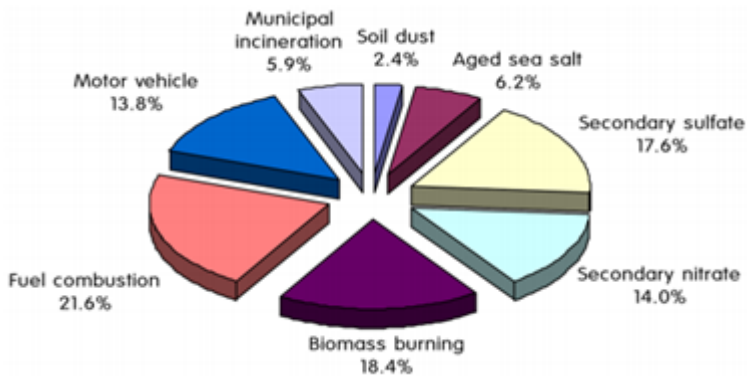
$$h_{ij} = \begin{cases} 1 & \text{if } |e_{ij}/s_{ij}| \leq \alpha \\ |e_{ij}/s_{ij}|/\alpha & \text{otherwise.} \end{cases}$$

1) 전하은 외, 2015, 서울지역 PM_{2.5}의 배출원별 기여도 추정 연구, 한국도시환경학회지 제15권 1호 7-16.

이때 S_{ij} 는 측정자료 X_{ij} 의 불확도이다. PMF는 각 발생원의 화학종 농도와 각 시료들에 대한 발생원 기여농도가 음의 값이 될 수 없다($f_{kj} \geq 0, g_{kj} \geq 0$)는 논리를 기반으로 하여 최소자승법으로 인자분석을 수행한다.

CMB 모델에 사용되는 source profile은 실제 서울지역에서의 source profile과 다를 수 있으므로 CMB 모델의 단점을 보완하여 조금 더 현실성 있는 인자를 산출하기 위하여 PMF(EPA PMF ver. 3.0)를 적용시켜 보았으며, 대기 중 $PM_{2.5}$ 농도에 대한 배출원별 기여도를 추정된 결과는 다음 그림과 같다.

PMF 분석결과, 서울지역 $PM_{2.5}$ 에 대해 자연적으로 배출되는 토양 및 해염입자 기원, 대기 중 화학반응에 의해 2차적으로 생성되는 황산암모늄이나 질산암모늄 형태의 이차 에어로졸, 그리고 연료연소, 생체연소, 자동차 배출원, 폐기물 소각을 기원으로 하는 총 8개의 배출원 인자가 산출되었다.



자료: 전하은 외, 한국도시환경학회지, 2015.

<그림 2-6> Source contributions to $PM_{2.5}$ measured for 3 years in Seoul

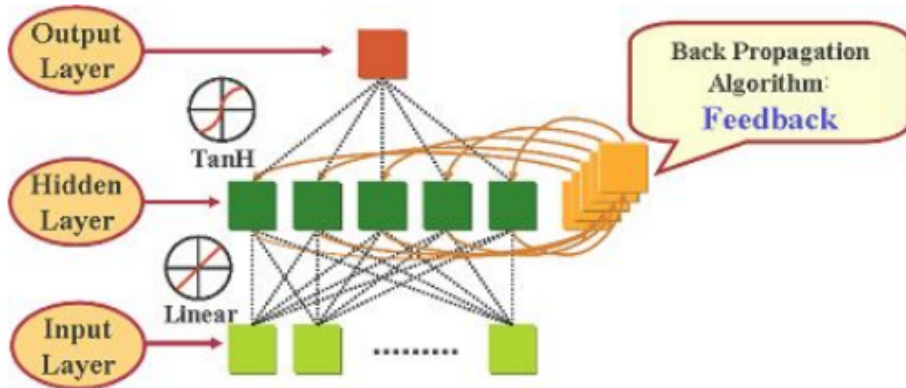
3. 인공신경망(Artificial Neural Network, 이하 ANN) 개념

인공신경망(이하 ANN)은 최근에 들어 기상학 분야에 사용되는 통계적인 기법으로 비선형 함수의 근사, 또는 복잡하고 비선형적인 역학적 현상의 모델링에 유용하게 사용된다.²⁾

ANN은 인간 두뇌에서 일어나는 반응 결정체를 자연계 현상의 예측(또는 추정) 과정에 도입한 알고리즘으로, 반복 학습을 통해서 스스로 새로운 환경에 적응하고, 일반화하기 때문에 모형식의 형태를 알 수 없거나 비선형모형으로 예측의 과정이 매우 복잡한 경우에 적용할 수 있다.

2) 부산대학교 기후예측연구실, (http://www.climate-prediction.org/res_sta.php)

신경망 모형은 학습이 이루어진 후, 신경망이 경험했던 것들과는 다른 형태의 입력에 대하여도 적당히 반응할 수 있다. 그림에서 보듯이 ANN 모형의 구성은 입력 변수들이 주어지는 입력층, 최종적인 결과가 주어지는 출력층 그리고 입력층과 출력층 사이에 1개 이상의 은닉층이 존재한다.



자료: http://www.climate-prediction.org/res_sta.php,

<그림 2-7> 인공신경망 구조

1) 인공지능 · 머신러닝 · 딥러닝의 차이³⁾

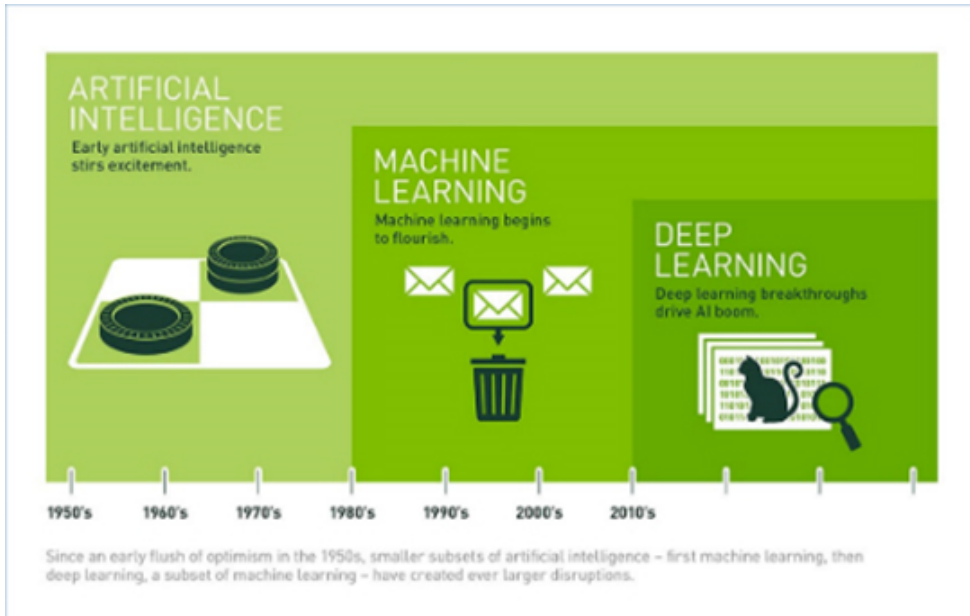
인공지능은 인간의 지능을 기계로 구현한 것으로 인간의 감각, 사고력을 지닌 채 인간처럼 생각하는 것을 ‘일반 AI(General AI)’ 라고 한다. 그러나 현재의 기술 발전 수준에서 만들 수 있는 인공지능은 ‘좁은 AI(Narrow AI)’ 의 개념에 포함된다(아이티비즈, 2016). 좁은 AI는 소셜 미디어의 이미지 분류 서비스나 얼굴 인식 기능 등과 같이 특정 작업을 인간 이상의 능력으로 해낼 수 있는 것이 특징이다.

머신러닝은 인공지능을 구현하는 구체적 접근 방식이다. 기본적으로 알고리즘을 이용하여 데이터를 분석하고, 분석을 통해 학습하며, 학습한 내용을 기반으로 판단이나 예측을 한다. 따라서 궁극적으로는 의사 결정 기준에 대한 구체적인 지침을 소프트웨어에 직접 코딩해 넣는 것이 아닌, 대량의 데이터와 알고리즘을 통해 컴퓨터 그 자체를 ‘학습’ 시켜 작업 수행 방법을 익히는 것을 목표로 한다.

딥러닝은 완전한 머신러닝을 실현하는 기술이라고 할 수 있다. 초기 머신 러닝 연구자들이 만들어 낸 또 다른 알고리즘인 인공신경망에 영감을 준 것은 인간의 뇌가 지닌 생물학

3) 아이티비즈, 2016. (<http://www.it-b.co.kr/news/articleView.html?idxno=12442>.)

적 특성, 특히 뉴런의 연결 구조였다. 딥러닝은 인공신경망에서 발전한 형태의 인공지능으로, 뇌의 뉴런과 유사한 정보 입출력 계층을 활용해 데이터를 학습한다.



자료: 아이티비즈, 2016. (<http://www.it-b.co.kr/news/articleView.html?idxno=12442>.)

<그림 2-8> 인공지능과 머신 러닝, 딥러닝의 차이

2) 딥러닝이 문제를 해결하는 방식

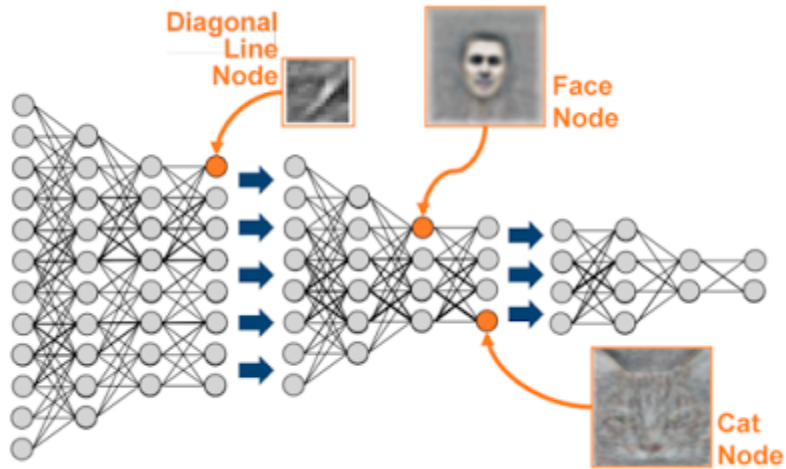
딥러닝의 핵심은 분류를 통한 예측이다. 수많은 데이터 속에서 패턴을 발견해 인간이 사물을 구분하듯 컴퓨터가 객체를 분별한다. 이 같은 분별 방식은 지도학습(supervised learning)과 비지도 학습(unsupervised learning)으로 나뉜다.⁴⁾

기존의 기계학습 알고리즘들은 대부분 지도 학습에 기초하고 있다. 지도 학습 방식은 컴퓨터에 먼저 ‘이런 이미지가 고양이야’ 라고 학습을 시켜주면, 학습된 결과를 바탕으로 고양이 사진을 판별하게 되는 것이다. 사전에 반드시 학습 데이터가 제공돼야만 한다. 사전 학습 데이터가 적으면 오류가 커지므로 데이터양도 충분해야 한다.

반면 비지도 학습은 이 과정이 생략되며 ‘이런 이미지가 고양이야’ 라고 학습시키지 않아도 자율적으로 컴퓨터가 ‘이런 이미지가 고양이군’ 이라고 학습하게 된다. 지도 학습 방식보다 진일보한 방식이지만 고도의 연산 능력이 요구돼 웬만한 컴퓨팅 능력으로는 시

4) 인공지능의 적용 사례, <http://blog.naver.com/guitar0318/220680384227>

도하기 쉽지 않다. 구글은 비 지도학습 방식으로 2012년 1,000대의 컴퓨터로 1,000만개의 유튜브 이미지를 딥러닝으로 분석해 <그림 2-9> 사람과 고양이를 구분하는 딥러닝 기술을 개발하였다.



자료: 쉽게 풀어쓴 딥러닝의 거의 모든 것, 2015, <http://t-robotics.blogspot.kr/2015/05/deep-learning.html>

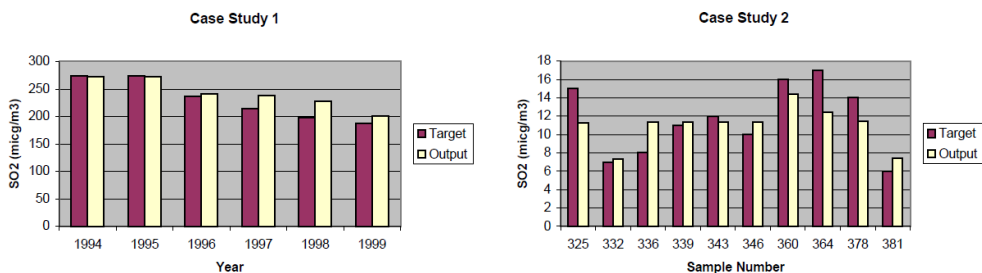
<그림 2-9> 딥러닝이 문제를 해결하는 방식

제2절 ANN을 이용한 대기질 예측 선행연구

1. 대기과학에서의 인공신경망(ANN) 적용

Gardner and Dorling(1998)은 인공 신경망(ANN)을 사용하면 대기과학 분야에서 이론적인 모델을 구축하기 어려운 비선형 시스템에서의 예측이 가능하다고 기술하였다. Kolehmainen et al.(2000)은 하이브리드 인공신경망 모델을 사용해 1995~1997년간 측정된 오염도와 기상 자료(시간별 자료)를 사용하여 모델링한 결과, 가스상 오염물질 예측에 유용하였음을 기술하는 등 21세기에 들어오면서 대기과학 분야에서 ANN을 활용하고자 하는 많은 시도와 발전이 있었다.

Barai et al.(2007)은 여러 종류의 인공신경망사용 모델 중 대기질 예측에 가장 효과적인 모델(Self Organizing Feature Maps: SOFM)을 사용하여 VOCs, SO₂, PM₁₀의 배출량을 예측하였다. 연구에서는 자료 수가 제한되어 있고 노이즈와 오차를 갖고 있는 대기질 자료에 대해 예측을 할 수 있는 인공신경망 모델을 찾고자 하였다. 다양한 인공신경망 모델인 Recurrent Network Model(RNM), Change Point detection Model with RNM(CPDM), Sequential Network Construction Model(SNCM), Self Organizing Feature Maps(SOFM)을 대기질 예측에 적용하여 보았고 그 결과 SOFM을 사용하여 가장 효과적으로 예측할 수 있었다. 개발된 모델은 연평균과 같은 장기자료와 일평균같은 단기자료를 사용하여 농도를 재현하고 예측하는 데 사용하였다. <그림 2-10>은 장기 및 단기 자료에서의 SO₂ 배출량 예측 결과를 보여주었다.

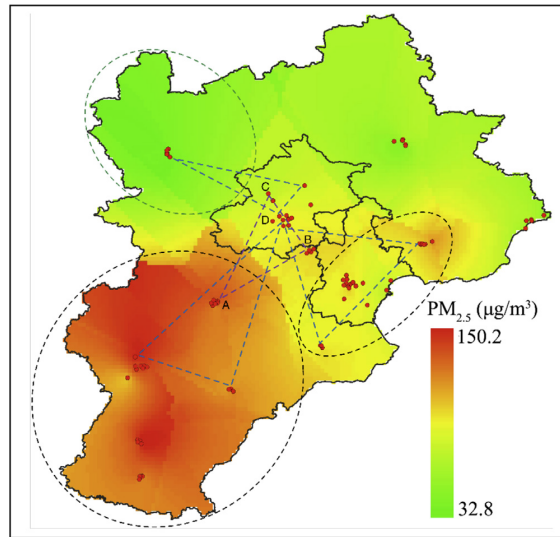


자료: Barai et al. 2007, Neural Network Models for Air Quality Prediction: A Comparative Study.

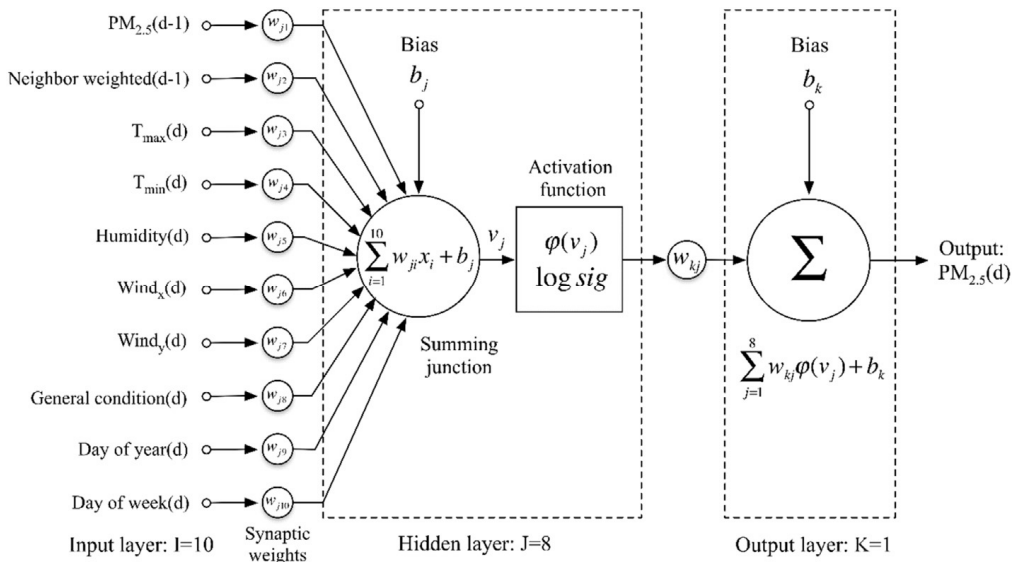
<그림 2-10> SO₂ emissions with SOFM

2. 공기이동궤적을 이용한 PM_{2.5}의 ANN예측

Feng et al.(2015)는 공기이동궤적으로부터 PM_{2.5}의 일평균 농도를 2일 전에 예측하는 데 ANN 기법 [MLP(Multi Layer Prediction) type]을 사용하였다. 2013년 9월~2014년 10월까지의 베이징, 톈진, 허베이 성 등의 13개 측정소 자료를 활용(〈그림 2-11〉), 공기이동궤적과 기상 자료(풍향, 풍속 등)로부터 유의한 예측 결과를 도출하였다. 입력자료 및 구체적 모델 구조는 〈그림 2-12〉와 같다.

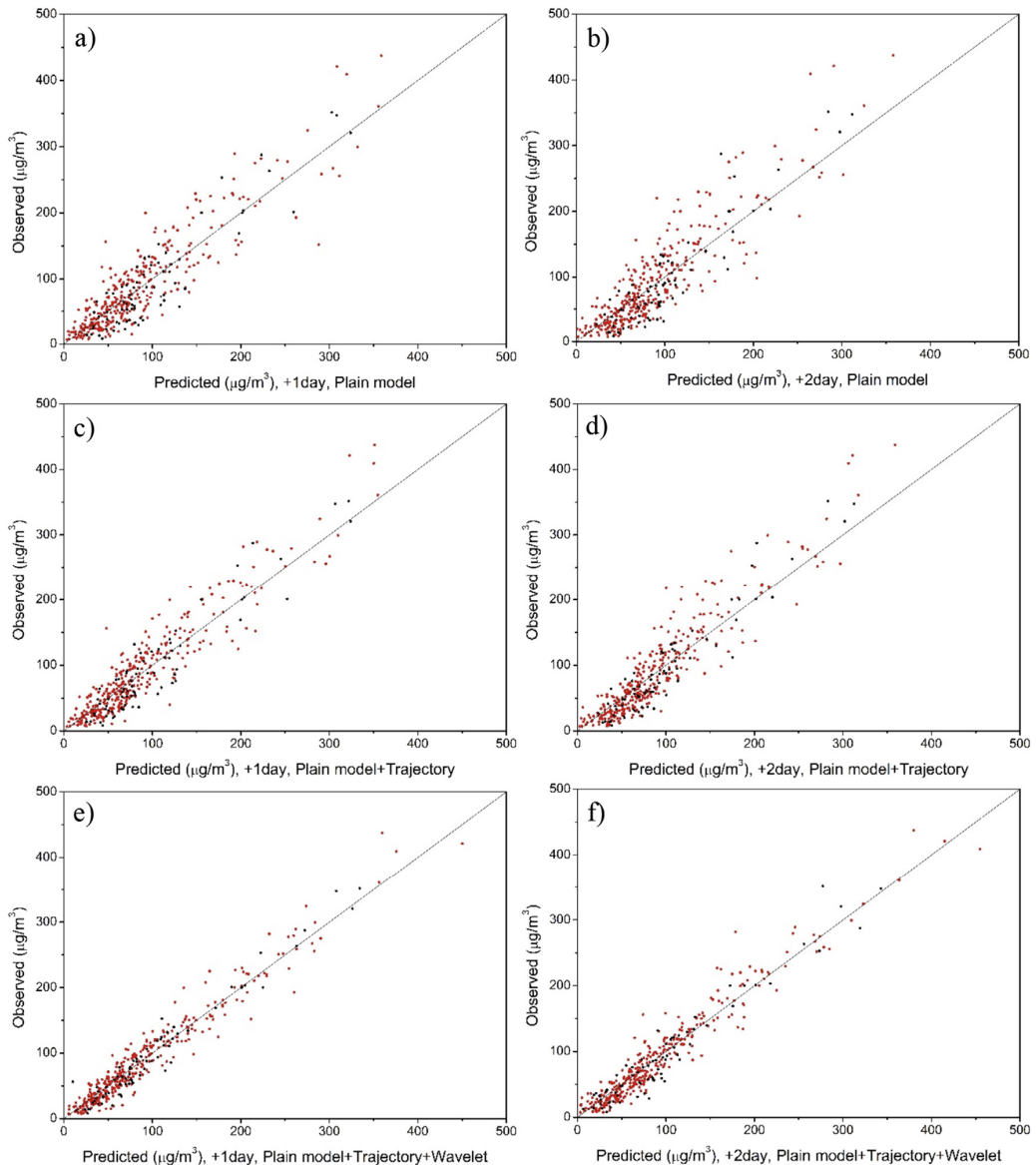


〈그림 2-11〉 인근지역의 PM_{2.5} 농도 분포



〈그림 2-12〉 MLP 타입의 인공신경망 구조

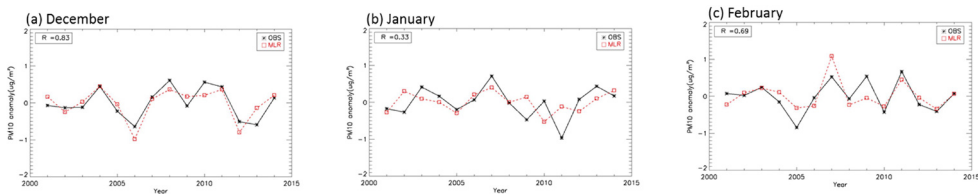
학습된 값(붉은색 점)과 모델링 예측값(검은색 점)과의 상관성은 <그림 2-13>에 나와 있는 바와 같이, 모델이 매우 잘 예측함을 보여준다. 또한 기상궤적과 wavelet transformation을 적용했을 때 정확도가 가장 높은 것을 보여주었다. 즉, 대상 지점의 PM_{2.5} 농도는 주변 지역에서의 공기이동에 의해 영향을 매우 많이 받고 있다는 것을 알 수 있었다.



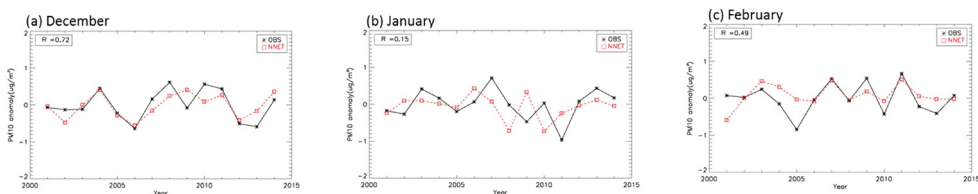
<그림 2-13> 예측모델 결과(붉은점:학습자료, 검은색:모델결과)

3. 미세먼지 예측을 위한 통계적 방법의 개발 사례⁵⁾

이한솔(2016)은 최근에 증가하는 겨울철 미세먼지 농도의 원인을 밝혀보고자 하였다. 이를 위해 과거 14년 간 한반도 중부 지방에서 관측된 지상 월 평균 농도와 Modern-Era Retrospective analysis for Research and Applications(MERRA) 재분석 중관 자료와의 상관성 분석을 통해서 겨울철 미세먼지 농도에 가장 중요하게 영향을 미치는 기상 인자를 밝혀냈다(이한솔, 2016). 또한 분석된 기상 변수와 겨울철 한반도 농도의 상관성을 토대로, 다중선형회귀 분석과 인공신경망을 이용하여 미래의 겨울철 미세 먼지의 연간 변동성을 예측할 수 있는 간단한 통계적 방법을 개발하고자 하였다. 개발된 다중선형회귀 모형과 인공신경망 모형의 정확도를 평가해 본 결과 다중선형회귀 모형을 통해 모의된 PM₁₀의 농도와 관측의 상관계수는 12월, 1월, 2월에 대해 각각 0.82, 0.33, 0.69 이었고, 인공신경망의 경우 각각 0.72, 0.15, 0.49 이었다. 다중선형회귀 모형이 인공신경망 모형보다 한반도 겨울철의 연간 변동성을 좀 더 잘 예측하는 것으로 나타났다(이한솔, 2016).



<그림 2-14> 다중선형회귀모형을 통해 모의된 PM₁₀의 농도와 관측농도의 상관성



<그림 2-15> ANN통해 모의된 PM₁₀의 농도와 관측농도의 상관성

5) 이한솔, 2016, 한국의 겨울철 PM 예측을 위한 간단한 통계적 방법의 개발, 서울대학교 석사학위논문.

4. 인공지능 기반 수도권 지하역사 실내공기질 예측⁶⁾ 사례 고찰

이 장에서는 권순박(2017)의 ‘인공지능 기반 수도권 지하역사 실내공기질 예측기술’ 연구 결과를 제시함으로 실내공기질 제어에 있어서의 ANN를 활용 필요성 및 실내공기질 예측 사례를 고찰하였다.

1) 공조환경관리시스템 한계

기본적으로 대합실, 승강장의 공조환경은 계절별, 시간대별로 단순 on-off 작동(타이머작동)이나 최근 역사에 있어서는 엔텔피 자동제어로 관리되고 있다. 그 중 승강장은 특히 전체 또는 국소 배기 방식으로 가동되어 냉방에너지 등의 과다소비가 발생한다. 또한 실외 미세먼지 고농도와 같은 외기오염 발생 시에는 외기를 전면 차단하도록 운영되고 있으나 승강장 내부 오염상황을 개선시킬 수 있는 방안은 사실상 부재하였다. 일부 역사에 대해서는 고가의 환경모니터링시스템(TMS)을 구축하여 실내대기질을 모니터링하고는 있으나, 이 활용한 연동제어까지는 현재 하지 못하고 있는 것이 실정이다.



자료: 권순박(2017), 인공지능 기반 수도권 지하역사 실내공기질 예측기술, 한국철도기술연구원.

<그림 2-16> 지하역사 공조관리시스템

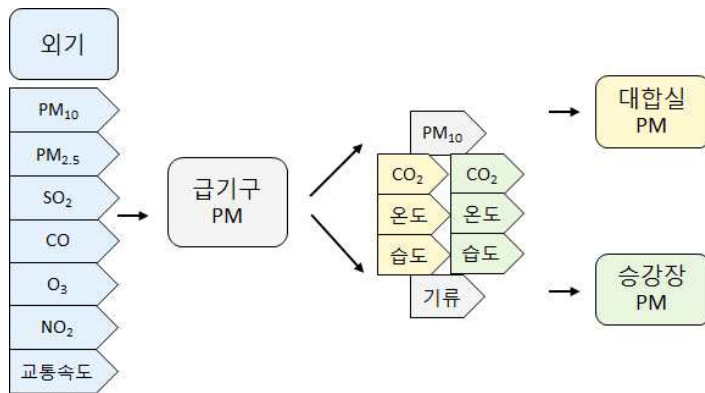
기존 공조설비 제어는 수동제어·자동제어가 되는 것으로 현재값을 기반으로 설정값을 맞추어가는 방식이다(측정-비교-판단-수정). 그러나, 최적 공조설비 제어는 인공지능기반의 복잡하고 다양한 현재 상태를 기반으로 예측값과 설정값을 맞추어가는 예측제어 방식이여

6) 권순박, 2017, ‘인공지능 기반 수도권 지하역사 실내공기질 예측기술’, 한국철도기술연구원.

야 한다(데이터수집-예측-판단-수정). 이를 위해서는 저가의 비용으로 양질의 빅데이터 수집이 필요하다.

2) 공공데이터를 활용한 급기구 미세먼지 예측 후, 대합실과 승강장 미세먼지 농도 예측(다중회귀분석)

권순박(2017)에서는 아래 모식도와 같이, 현재의 외기의 공공데이터로 1시간 후 급기구의 PM₁₀을 예측할 수 있을 것으로 가정하고 급기구 PM₁₀이 예측 가능하다면, 그 예측 값과 내부데이터로 대합실과 승강장의 PM₁₀도 예측 가능할 것이라고 가정한 분석을 수행하였다.



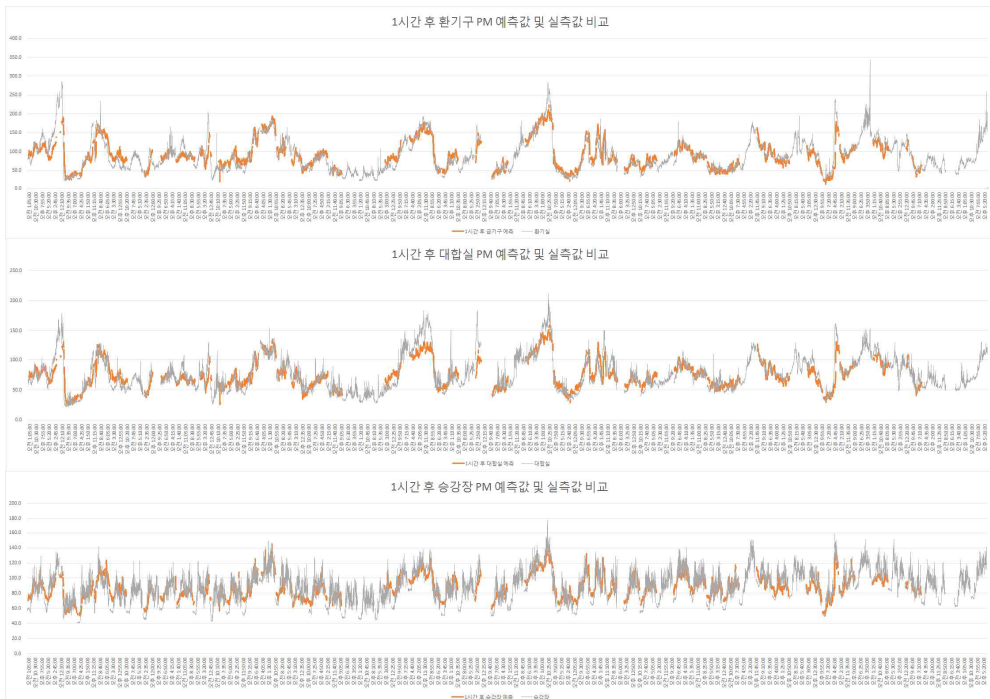
자료: 권순박(2017), 인공지능 기반 수도권 지하역사 실내공기질 예측기술, 한국철도기술연구원.

<그림 2-17> 외기, 급기구, 대합실, 승강장 데이터간의 관계 모식도

각 데이터간의 상관분석 및 예측식 도출을 위한 다중선형회귀분석은 IBM SPSS Statistics 22. 프로그램을 사용하였다. 1시간 후 대합실 및 승강장의 PM₁₀ 회귀식(예측식)은 아래와 같다.

- 지하역사 유입 PM₁₀ = 20.2 + (-0.3 × PM₁₀) + (2.7 × PM_{2.5}) + (97.1 × NO₂) + (-0.3 × T.Speed*)
- 1시간 후 대합실 PM₁₀ = 28.1 + (0.6 × 유입PM₁₀) + (0.15 × C.CO₂) + (-0.006 × P.CO₂) + (-1.3 × C.T.) + (0.66 × C.H.) + (-0.57 × P.H.)
- 1시간 후 승강장 PM₁₀ = 55.6 + (0.36 × 유입PM₁₀) + (-0.03 × C.CO₂) + (0.06 × P.CO₂) + (-1.7 × C.T.) + (-0.9 × C.H.) + (0.2 × P.T.) + (0.98 × P.H.)

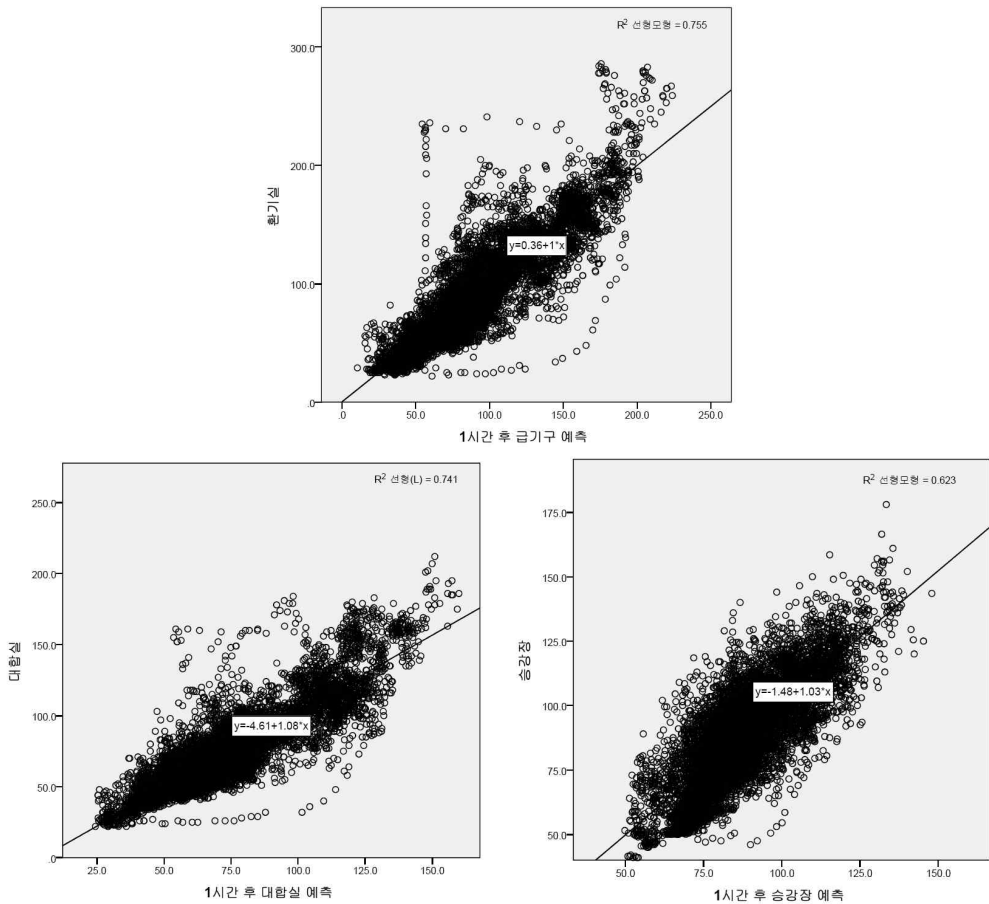
11월, 12월 두달간의 데이터로 예측식 검증을 수행하였고, <그림 2-18>의 예측값(푸른색 선)은 실측값(주황색 선)과 경향이 잘 일치함을 보여주었다.



자료: 권순박, 2017, 인공지능 기반 수도권 지하역사 실내공기질 예측기술, 한국철도기술연구원.

<그림 2-18> 환기실/대합실/승강장 PM₁₀ 예측값과 실측값 비교(11-12월), 주황색 선: 1시간 후 실측값, 푸른색 선: 1시간 후 예측값

환기실의 실측값과 예측값을 나타낸 산점도의 R^2 는 0.755, 기울기는 1 이었다. 대합실의 경우에는 R^2 는 0.741, 기울기는 1.08이고 승강장의 R^2 는 0.623, 기울기는 1.03으로 분석하였다. 두달 간의 PM₁₀ 평균 값은, 환기실 예측값은 92.0 $\mu\text{g}/\text{m}^3$, 실측값은 89.9 $\mu\text{g}/\text{m}^3$ 였으며, 대합실 예측값은 76.5 $\mu\text{g}/\text{m}^3$ 실측값은 76.4 $\mu\text{g}/\text{m}^3$, 승강장 예측값과 실측값은 모두 88.3 $\mu\text{g}/\text{m}^3$ 으로 모두 오차범위 내에서 일치하는 결과를 얻었다. 권순박(2017)은 다중회귀분석을 통해 실외 미세먼지 농도와 물리적인 실내외 환경조건을 가지고 실내 미세먼지 농도를 예측할 수 있었다.

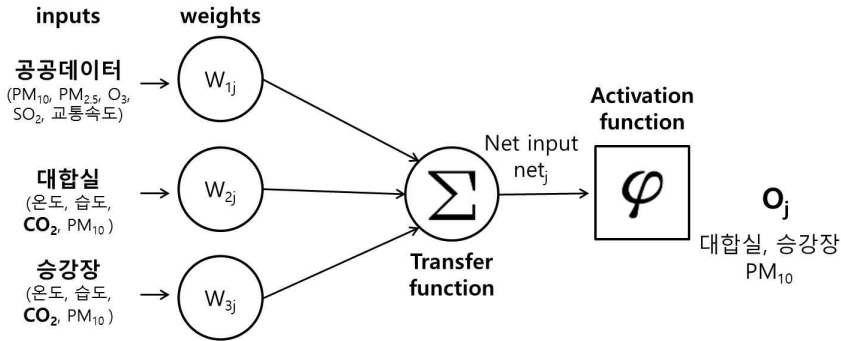


자료: 권순박, 2017, 인공지능 기반 수도권 지하역사 실내공기질 예측기술, 한국철도기술연구원.

<그림 2-19> 환기실/대합실/승강장 PM₁₀ 예측값과 실측값 산점도(11-12월)

3) 인공지능망을 활용한 지하역사 미세먼지 예측

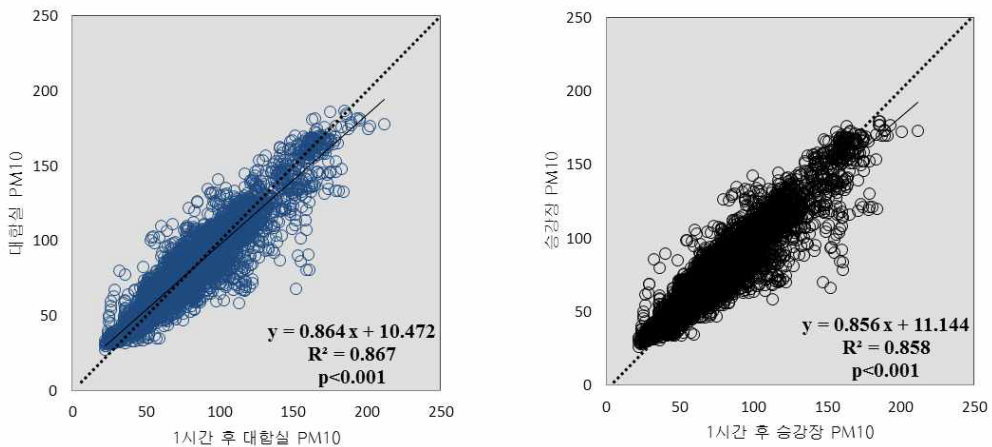
권순박(2017)은 동일한 입력변수를 인공신경망을 활용하여 대합실과 승강장의 PM₁₀을 계산하고, 계산된 각 지하역사의 PM₁₀을 실제 PM₁₀ 측정값과 비교하였다. 각 입력변수는 각자의 가중치를 학습하여 변환함수(transfer function)에 적용하였고, 이를 활용하여 최적의 단위 입력값을 만들고 이 값을 활용하여 대합실과 승강장의 PM₁₀이 연산되는 방식이 사용되었다.



자료: 권순박, 2017, 인공지능 기반 수도권 지하역사 실내공기질 예측기술, 한국철도기술연구원.

<그림 2-20> 활용된 입력변수(input)와 인공신경망 모식도

인공신경망을 활용하여 예측값 PM_{10} 과 실측값 PM_{10} 를 비교하여 1시간 후의 미세먼지 농도를 예측한 결과, 대합실의 R^2 는 0.864, 승강장은 0.858이며 유의수준 0.001 이하였다. 대합실에서 평균 예측값은 $76.83 \mu\text{g}/\text{m}^3$, 실측값은 $76.79 \mu\text{g}/\text{m}^3$ 를 나타내었으며 승강장에서 평균 예측값은 $77.20 \mu\text{g}/\text{m}^3$, 실측값은 $77.19 \mu\text{g}/\text{m}^3$ 로 나타내어 대합실과 승강장 모두 실측값의 0.01% 이하의 오차범위 안에 들었다.

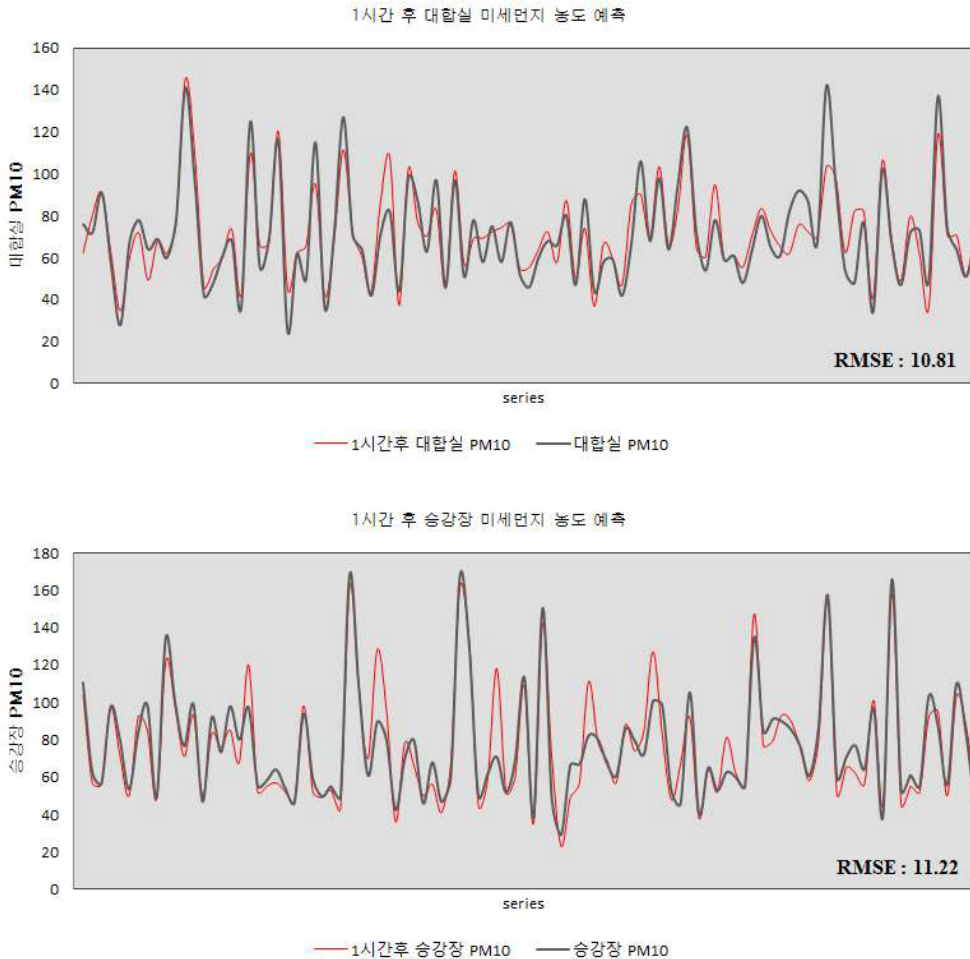


자료: 권순박, 2017, 인공지능 기반 수도권 지하역사 실내공기질 예측기술, 한국철도기술연구원.

<그림 2-21> 인공신경망을 활용한 대합실/승강장 PM_{10} 연산값과 실측값 산점도(11-12월)

대합실과 승강장에서 PM_{10} 예측값과 실측값을 비교할 결과 <그림 2-22>과 같이 나타났으며, 예측값 PM_{10} 과 실측값 PM_{10} 의 평균제곱근오차(RMSE)는 대합실에서 $10.81 \mu\text{g}/\text{m}^3$, 승강장에서 $11.22 \mu\text{g}/\text{m}^3$ 를 나타내었다. 즉, 다중 회귀분석과 동일한 입력값을 적용한 ANN 활용

인공신경망 기법으로 1시간 후 미세먼지농도를 오차범위 $\pm 11 \mu\text{g}/\text{m}^3$ 수준으로 예측가능하다는 결과를 제시하였다.



자료: 권순박, 2017, 인공지능 기반 수도권 지하역사 실내공기질 예측기술, 한국철도기술연구원.

<그림 2-22> 예측된 PM₁₀과 실측된 PM10 비교

4) 공조제어 알고리즘

ANN 모델 활용 결과로, 권순박(2017)은 다음 <그림 2-23>에서 제안하는 바와 같이 1시간 후 승강장의 미세먼지 농도를 예측해서 실내기준치 이하일 경우에는 에너지 절약 모드 (Energy saving mode)로, 기준치 이상일 경우에는 공기청정모드(Air cleaning mode)로 운전하는 방식으로 공조시스템 변동제어 알고리즘으로 활용 가능하다는 내용을 제시하였다.



자료: 권순박, 2017, 인공지능 기반 수도권 지하역사 실내공기질 예측기술, 한국철도기술연구원

〈그림 2-23〉 공조시스템 급배기 변동제어 알고리즘

즉, 권순박(2017) 사례는 상시 측정 자료인 실외 미세먼지 농도 자료를 활용하여 실내 미세먼지 농도를 예측하고 이 값을 활용하여 실내공기질 관리가 가능한 미세먼지 공조시스템을 운영할 수 있다는 방법을 제시한 것이었다.

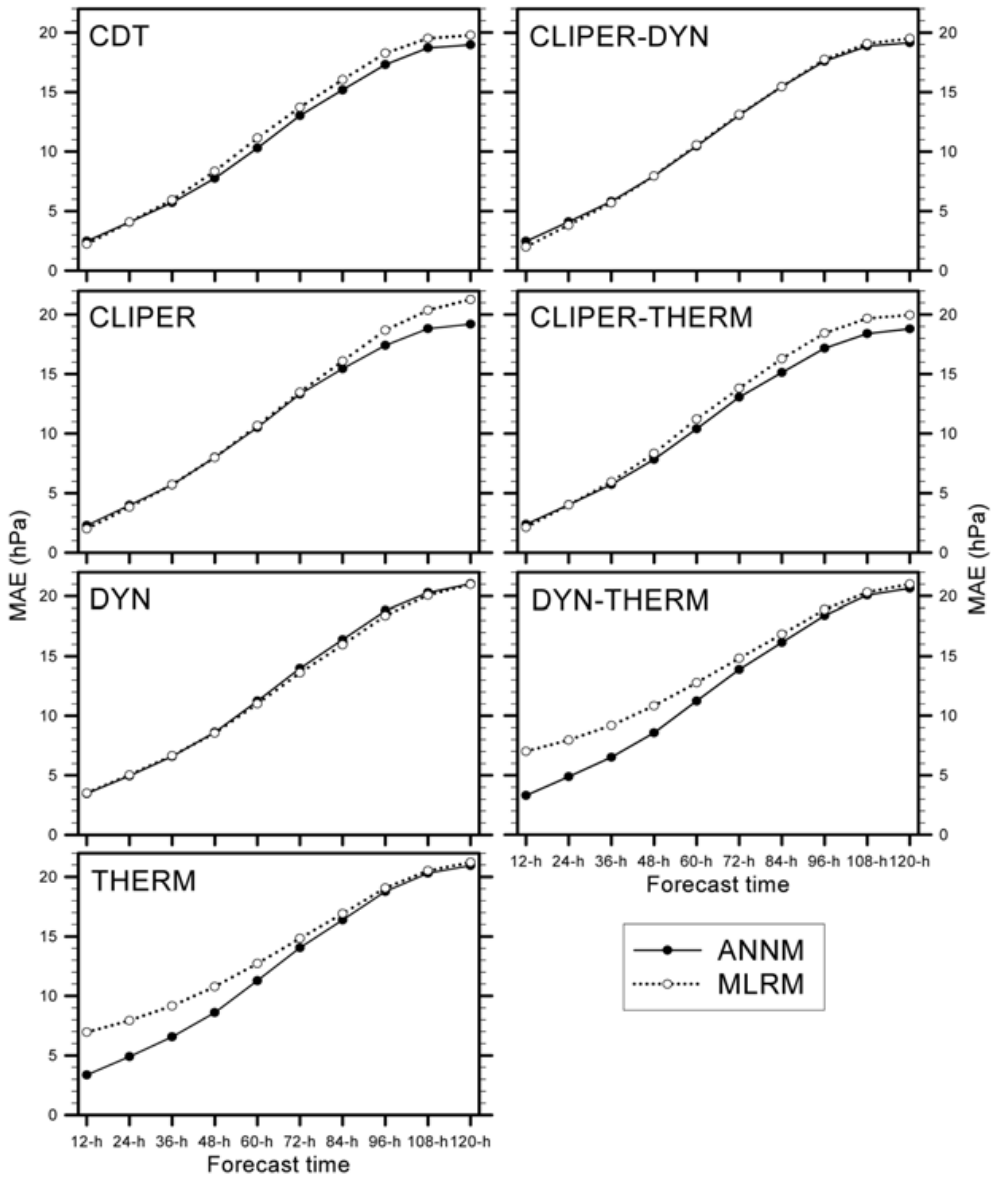
5. 인공지능경망 기법을 이용한 태풍 강도 및 진로 예측⁷⁾

최기선 외(2009)의 ‘인공신경망 기법을 이용한 태풍 강도 및 진로 예측’ 연구에서는 북서태평양에서 발생한 태풍에 대해 발생 12시간 후부터 5일 동안 12시간 간격으로 태풍의 강도 및 진로를 예측하기 위한 인공신경망 모델을 개발하였다. 사용되어진 11개의 예측인자는 특성에 따라 세 개의 군집[CLIPER(Climatology and Persistent model), dynamic(DYN), thermal(THERM)]으로 분류되었으며, 분류된 예측인자를 이용하여 일곱개의 인공신경망 모델들[CDT(CLIPER-dynamic-thermal), CLIPER, DYN, THERM, CLIPER-DYN, CLIPER-THERM, DYN THERM 모델]로 나누었다(최기선 외, 2009). 이 일곱개의 인공신경망 모델들의 오차분석을 통해, 강도와 진로예측 모두에서 CLIPER-THERM 모델이 가장 좋은

7) 최기선 외, 2009, 인공지능경망 기법을 이용한 태풍 강도 및 진로 예측, 한국지구과학회, v. 30, no. 3, p. 294 -304, June 2009

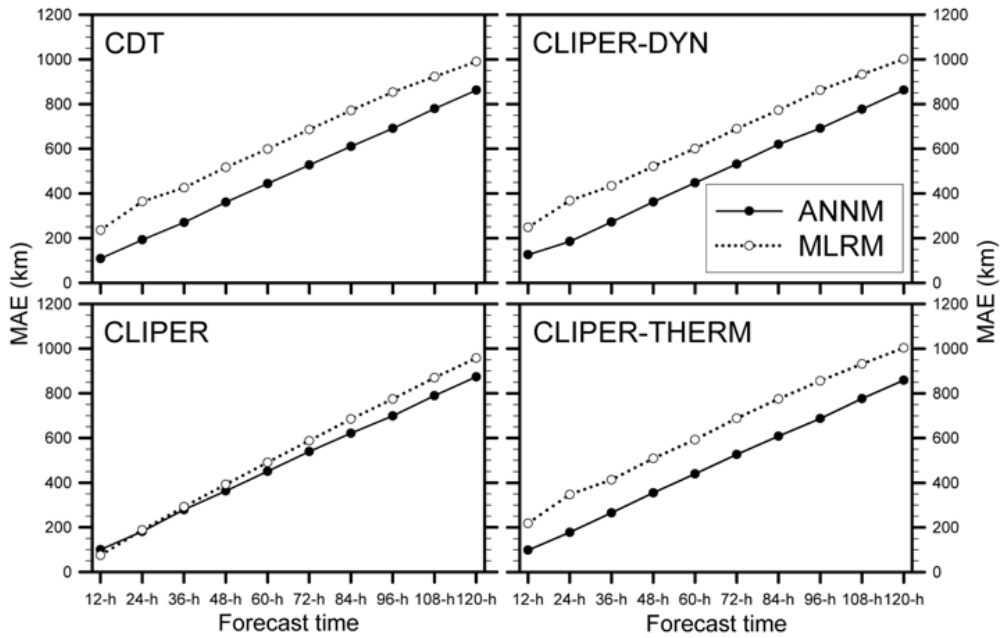
예측성능을 가졌다는 것을 기술하였다. 또한 이 연구결과를 통해 태풍의 강도 및 진로예측 시에 태풍의 초기치가 반드시 필요하며 특히, 태풍의 열적 상태가 예측성능에 중요한 위치를 차지함을 기술하였다.

또한, 인공신경망 모델에서 사용된 예측인자들을 사용하여 다중선형회귀모델을 만들어 그 예측 결과를 비교한 결과 태풍의 강도 및 진로 모두 인공신경망 모델이 더 나은 예측성을 보여 주었다(최기선 외, 2009). 그러나 인공신경망 모델은 결과의 도출과정이 명확하게 설명될 수 없다는 큰 취약점을 가지고 있으며 분류나 예측 결과만을 제공할 뿐이며 어떻게 그러한 결과가 산출되었는가에 대한 명확한 근거를 설명하지 못한다는 단점이 있다는 것도 한계점으로 나타났다.



자료: 최기선 외, 2009, 인공지능기법 이용 태풍 강도 및 진로 예측, 한국지구과학회.

<그림 2-24> 5일 동안 12시간 간격으로 7개의 인공지능기법 모델(ANNM)과 7개의 다중선형회귀모델(MLRM)에 의해 예측된 태풍 강도변화의 평균절대오차(MAE) 비교



자료: 최기선 외, 2009, 인공지능경망 기법을 이용한 태풍 강도 및 진로 예측, 한국지구과학회.

<그림 2-25> 인공지능경망 모델과 다중선형회귀모델의 진로예측 비교

제3절 시사점

ANN을 사용한 상관성 분석 및 예측 모델은 분명한 메커니즘 없이 관측 결과를 직접 사용하여 학습함으로써 인과관계가 있는 영향요인을 추출해 내는 방식이다. 따라서 메커니즘 규명이 어려운 경우, 물리적인 영향요인(예: 도로 교통량, 실내-실외 환기 비율 등)을 미세먼지 농도 기여 메커니즘에 대입하기 어려운 경우, 그 원인을 알아보고자 할 때 시도되고 있었다.

문헌 고찰을 통해서 ANN을 사용하여 주변 지역 미세먼지 농도로부터 한 지점의 미세먼지 농도를 예측하거나, 실외 미세먼지 농도와 환기상태를 통해 실내 미세먼지 농도를 예측하거나, 또는 태풍 발생 후 진로 및 강도를 예측하는 식의 연구가 진행되었음을 알 수 있었다. ANN 모델은 다중선형회귀분석과 비교하였을 때, 예측을 잘 하는 것으로 평가되었으며, ANN의 여러 모델링 방식 중 해당 분야에 잘 맞는 방법을 선정하여 적용하는 것이 예측 정확성을 높이는 데 중요하다는 것도 시사하였다.

다만, 메커니즘 입력 없이 과거 자료 학습을 통해 예측하기 때문에 입력자료(대기질 농도, 기상, 교통량, 환기 등)에 대한 정확성이 담보되어야 한다. 또한 제공한 입력 자료에 한해서 상관성을 찾아내기 때문에, 가능한 이론적으로 타당한 상관성이 있는 입력 자료를 제공하여 모델링을 해야 한다는 것도 중요한 고려사항임을 알 수 있었다.

따라서 다음 장에서 수원시 대기질 예측을 위해 ANN 기법을 적용할 때, 문헌고찰 시사점에서 도출된 바와 같이 정확한 입력자료의 사용, 수원시 대기질의 전반적인 특징 분석 및 미세먼지 농도 증가에의 기여원인 조사 등을 수반하고자 한다.

제3장 수원시 대기질 현황

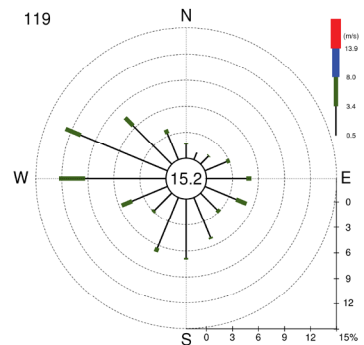
제1절 수원시 대기오염농도

1. 수원시 주변지역과의 비교

2016년 대기환경월보(www.airkorea.or.kr)의 PM₁₀, PM_{2.5}, SO₂, NO₂, O₃, CO 농도를 활용하여 수원시와 주변지역과의 대기오염도를 비교하였다. 대기질 측정 자료는 국가대기오염측정망(국립환경과학원 QA/QC 검증 후 공개자료)에서 측정되는 자료를 사용하였고 백령도, 안산시와 비교하였다. 백령도 측정소는 우리나라 최서북단에 위치하여 한반도에 유입되는 장거리 이동된 오염물질을 측정하는 국가 배경농도 측정소이며, 안산시는 서풍이 불 때 수원시 대기질에 영향을 줄 수 있는 인접 도시이다. 안산시와 수원시 측정자료는 각 도시별 다수 측정지점에서 측정된 자료를 평균한 값을 사용하였다. <그림 3-1>은 비교대상 측정지점의 위치, <그림 3-2>는 2016년 수원시의 주풍향이 서풍계열임을 알려주는 바람장미이다.



<그림 3-1> 수원시 주변 지역

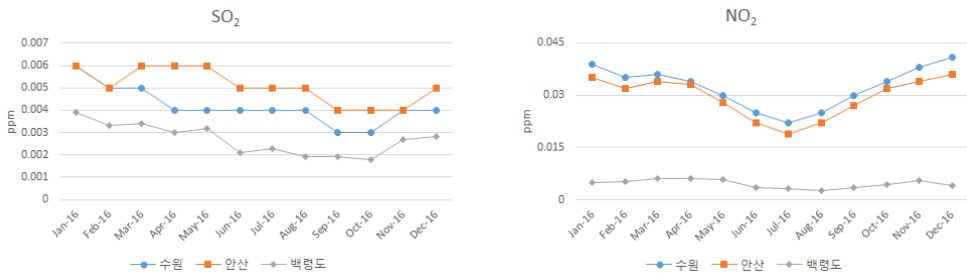


<그림 3-2> 2016년 수원 바람장미

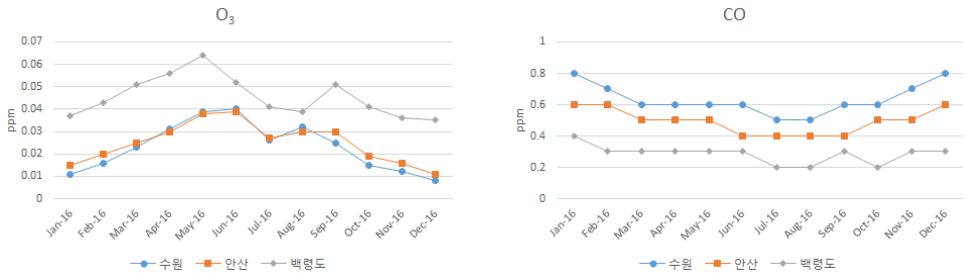
대기오염물질 월 평균값을 비교한 결과, NO₂와 CO의 경우 백령도보다 도시지역인 수원시와 안산시에서 높게 나타났다. 하지만, 체류시간이 수시간 이내로 짧은 NO₂의 경우 도시지역인 수원시와 안산시의 농도가 배경농도 지점인 백령도보다 많게는 10배 이상 높았다. O₃의 경우 도시지역인 수원, 안산보다 백령도 농도가 더 높았는데 그 원인은 광화학반응에 의해 질소산화물이 오존으로 변환되었기 때문으로 사료된다. NO₂와 미세먼지의 경우 월별 추이는 각 안산시, 수원시의 경우 매우 유사한 패턴으로 변화였다. 특히 미세먼지가 다른 오염물질에 비해 더 비슷한 경향을 보여 미세먼지 농도에는 지역 배출원 뿐만 아니라 기

상조건과 장거리 수송된 외부 배출원의 영향이 큰 것으로 판단된다. PM₁₀은 특히 안산시의 수원시의 농도가 백령도보다는 높았지만, 수원, 안산 두 지역의 농도차이가 거의 보이지 않아 지역적 차이를 뚜렷히 볼 수 없었다. 다만 PM_{2.5}가 SO₂와 유사하게 안산시의 농도가 수원시보다 높아 지역적인 차이를 볼 수 있었다.

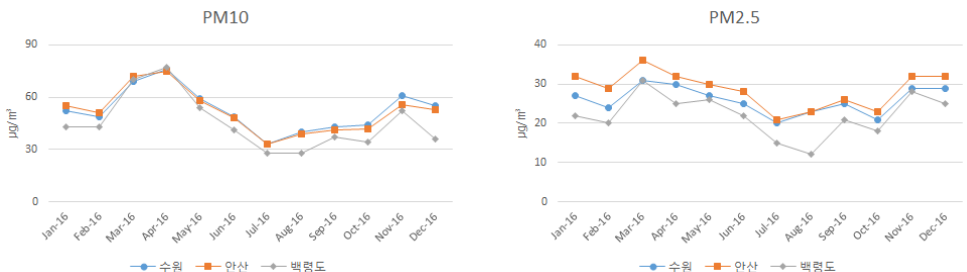
백령도 측정값은 단일 지점 결과이지만, 수원시와 안산시는 각각 7개소의 측정소 값을 평균한 결과이므로 인접 지역의 평균경향을 나타내어, 측정지점간 차이가 두드러지지 않을 수도 있다. 따라서 다음 장에서는 수원시 내 측정소별 농도를 비교하였다.



<그림 3-3> 수원시 주변지역 SO₂와 NO₂ 농도 비교



<그림 3-4> 수원시 주변지역 O₃와 CO 농도 비교



<그림 3-5> 수원시 주변지역과의 미세먼지 농도 비교

2. 수원시 국가대기오염측정소 현황

수원시에는 7개의 측정망을 설치·운영하고 있으며, 측정항목은 SO₂, CO(일부 CO₂), O₃, NO₂, PM₁₀, PM_{2.5}, 기상정보 등이다. 도시대기측정망은 도시지역의 평균대기질 농도를 파악하기 위해 설치·운영되며, 도로변 대기 측정망은 자동차통행량과 유동인구가 많은 도로변 대기질 파악을 위해 설치·운영된다. 측정자료는 매 시간 농도 평균이며, 경기도 보건환경연구원의 QA/QC를 거친 확정자료를 사용하였다.

<표 3-1> 수원시 국가대기오염측정소 현황

구분	측정소명	주소	설치 년도	측정항목
도시 대기	1.신평동	팔달구 신평로 23번길 68(선경도서관)	1986	SO ₂ , CO, O ₃ , NO ₂ , PM ₁₀ , PM _{2.5} , 기상정보 등 (고색동은 PM _{2.5} 측정하지 않음)
	2.인계동	팔달구 효원로241(수원시청)	1987	
	3.영통동	영통구 영통로 217번길 12(영통2동주민센터)	2001	
	4.천천동	장안구 서부로 2066(성균관대학교 제2공학관)	2003	
	5.고색동	권선구 서부로 1600(차량등록사업소)	2006	
	6.광고동	영통구 범조로 129(이의중학교)	2013	
도로변 대기	7.동수원	팔달구 경수대로 596(동수원사거리, SK생명)	2005	



<그림 3-6> 수원시 대기오염측정소 위치

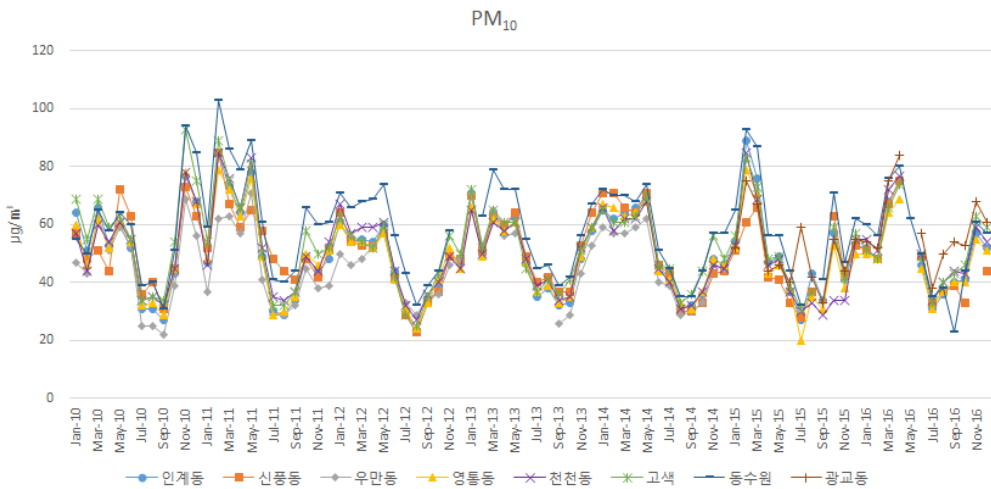
3. 수원시 대기오염농도

수원시 대기오염측정망 7곳의 데이터(2010.01~2016.12)를 다음과 같이 대기오염물질별로 분석하였다.

1) PM₁₀

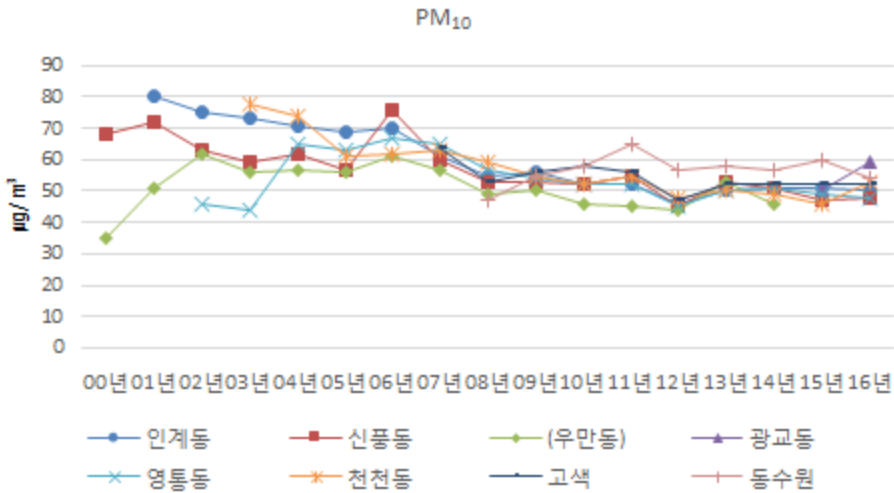
2010년부터 2016년까지의 월별 미세먼지 농도 변화추이를 살펴본 결과 도로변(동수원)에서 PM₁₀ 농도 높게 나타났으며, 특히 봄과 겨울철에 다소 높게 나타났다. 도로변(동수원)측정소를 제외하고 나머지 측정소에서는 비슷한 경향이 나타났다. 광교동의 경우 2015년부터 다른 도시대기 측정소보다 다소 농도가 높은 경향을 보였다.

※ 우만동 측정소는 2014년까지만 측정, 광교동 측정소는 2015년부터 측정됨.



<그림 3-7> 수원시 PM₁₀ 농도 변화 추이(2010.01~2016.12)

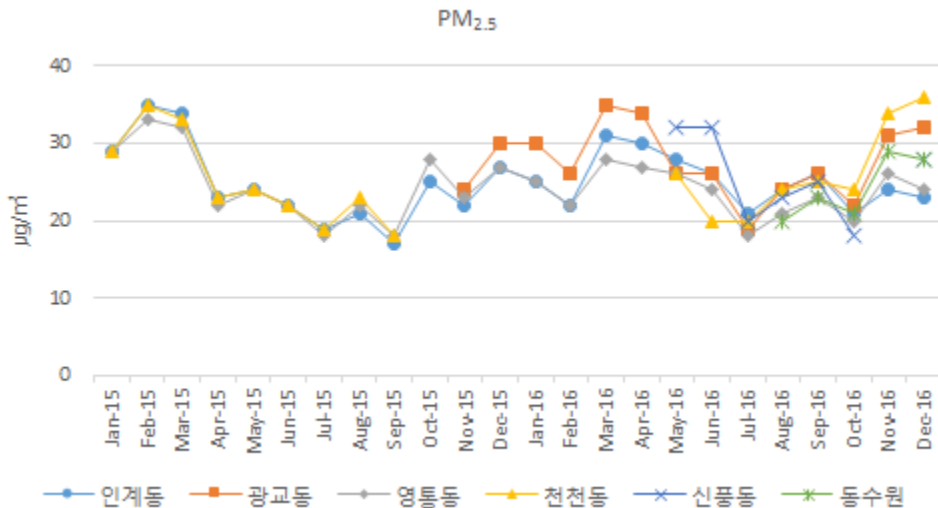
PM₁₀의 연도별 변화 추이를 보면 2000년부터 점차 감소되는 추세를 보이나 2010년부터는 크게 감소하지 않고 일정한 오염도를 보인다.



<그림 3-8> 수원시 PM₁₀ 연도별 변화

2) PM_{2.5}

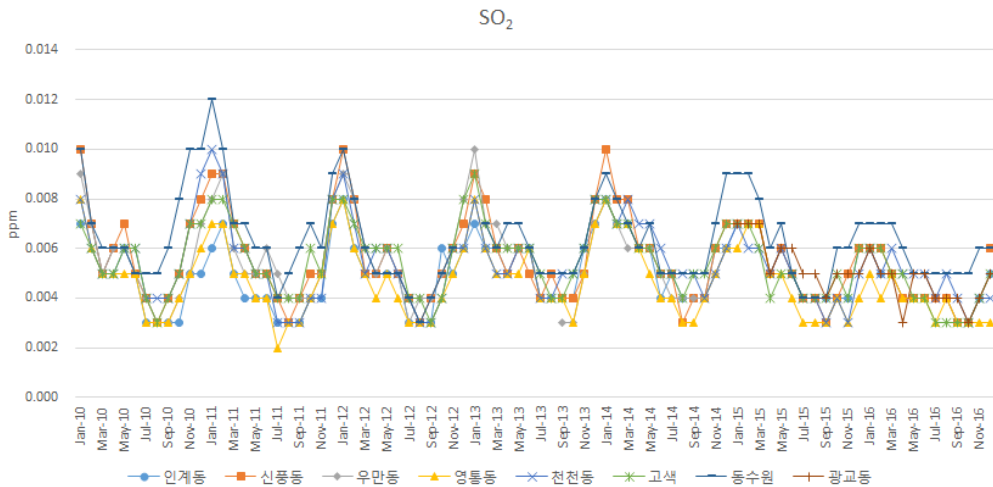
PM_{2.5}는 2015년부터 공식적으로 측정되어 농도추이를 분석한 결과 6곳의 측정소의 변화 추이는 비슷한 양상을 보이며 봄과 겨울철에 높은 것을 알 수 있다. 광고동의 경우 측정시기부터 다른 지점보다 농도가 높았는데 이러한 현상은 PM10에서도 유사하였다. 또한 PM10의 경우와는 다르게 PM_{2.5}는 도로변이라고 해서 높게 나타나지는 않았다.



<그림 3-9> 수원시 PM_{2.5} 농도 변화 추이(2015.01~2016.12)

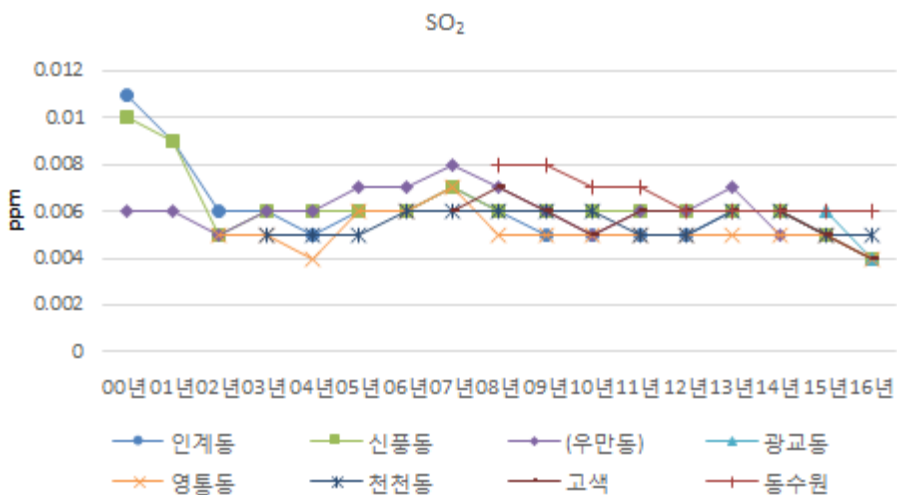
3) 이산화황(SO₂)

이산화황(SO₂)의 오염도는 점차 감소하고 있는 추세이며, 2010년과 2015년 겨울철에 도로변 측정소에서 이산화황 농도가 높게 나타났으며, 국지적 오염원 때문인지 확산이 되지 않아서인지 원인은 불확실하다.



<그림 3-10> 수원시 SO₂ 농도 변화 추이(2010.01~2016.12)

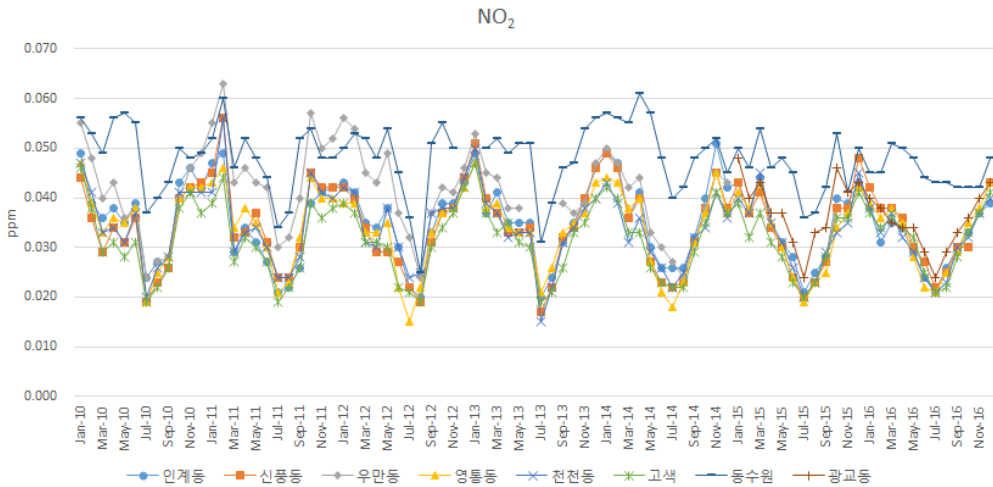
이산화황의 연도별 추세를 보면 2000년부터 감소하는 추세를 보이며, 동수원 측정소는 일정한 추세를 보인다.



<그림 3-11> 수원시 SO₂ 연도별 변화

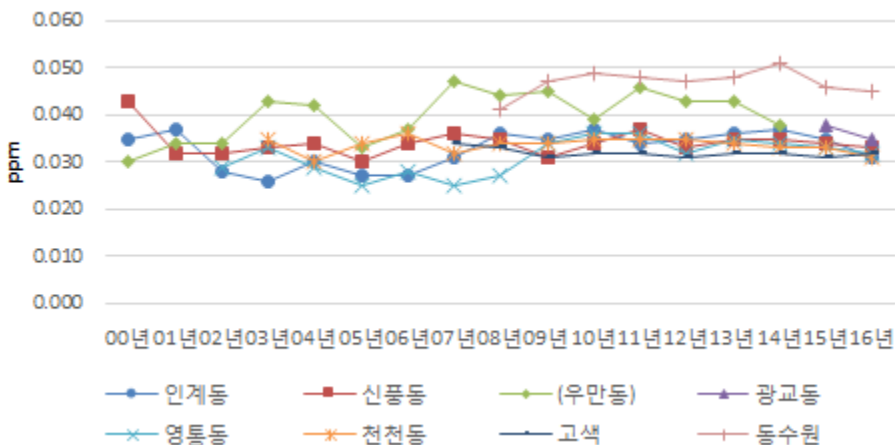
4) 이산화질소(NO₂)

이산화질소(NO₂) 농도는 여름철에 낮게 나타나는 경향을 보이며, 도로변에서 이산화질소 농도는 다른 측정소에서보다 항상 높게 나타나고 있다. 이산화질소의 주요배출원인인 자동차의 영향으로 보인다. 광고동의 경우 다른 도시대기 측정소보다는 다소 높은 경향을 보였으나, 2015년부터 측정했기 때문에 일관성있게 보이는 결과라고 해석할 수는 없었다.



<그림 3-12> 수원시 NO₂ 농도 변화 추이(2010.01~2016.12)

이산화질소의 2000년부터 연도별 추세를 보면 크게 감소하거나 증가하지 않고 있으며 도로변 측정소인 동수원에서 다른 측정소에 비해 다소 높게 나타나고 있다.

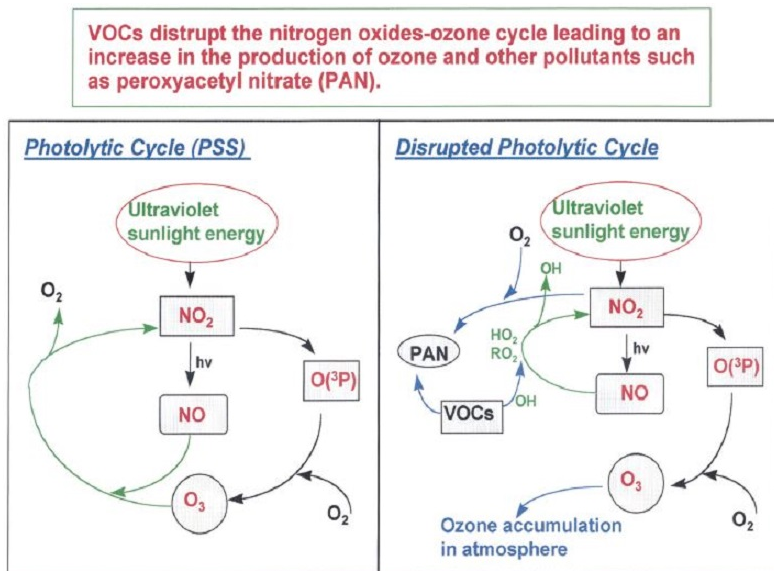


<그림 3-13> 수원시 NO₂ 연도별 변화

5) 오존(O₃)

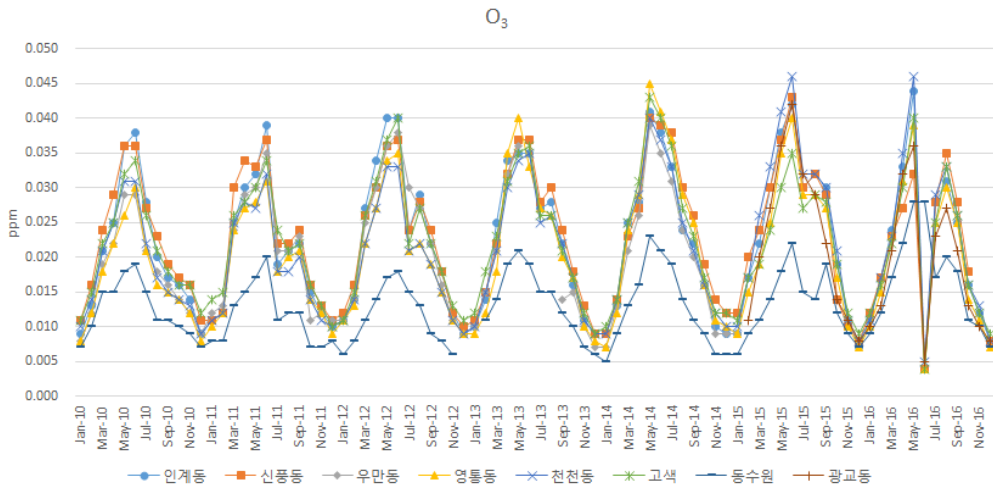
도시지역에서 오존은 자동차의 배기가스에서 배출되는 NO_x (NO+NO₂)와 CO, 그리고 VOCs에 의한 광화학반응을 통해 생성된다(Jacob, 1999). 대기 중 질소산화물 NO_x는 1차 오염물질일 뿐만 아니라 VOCs와 함께 O₃과 같은 광화학 2차 오염물질의 전구물질로 주로 화석연료의 연소와 자동차 배기가스로부터 배출된다. 우리나라는 도시화 및 산업화에 따라 그 배출량이 지속적으로 증가하였다

오존 농도는 여름철(6-8월)에 높게 나타나며, 이러한 계절적 특성은 서울과 같은 대도시에서도 나타나는데 이는 공기 이동경로와 강수 등 기상요소로 인한 전구물질의 농도와 광화학 활동도의 차이에 기인한다. 우리나라는 대부분 봄에는 북서풍으로 중국 대륙에서 이동하는 공기의 영향을 주로 받아 전구물질의 농도가 높지만 광화학 활동도가 여름에 비해 낮다. 반면 늦여름에는 태평양을 거쳐 오는 공기와 집중 호우로 인해 전구물질의 농도가 1년 중 가장 낮지만 온도와 수증기 농도가 가장 높고 광화학 활동도가 높다. 이산화질소 농도가 높은 곳에서는 오존 농도가 낮다. 도로변(동수원) 데이터를 보면 알 수 있듯이 오존 농도가 다른 지점에 비해 낮게 나타났다. 이는 다음 그림과 같이 O₃-NO_x 광분해 cycle 때문이다.



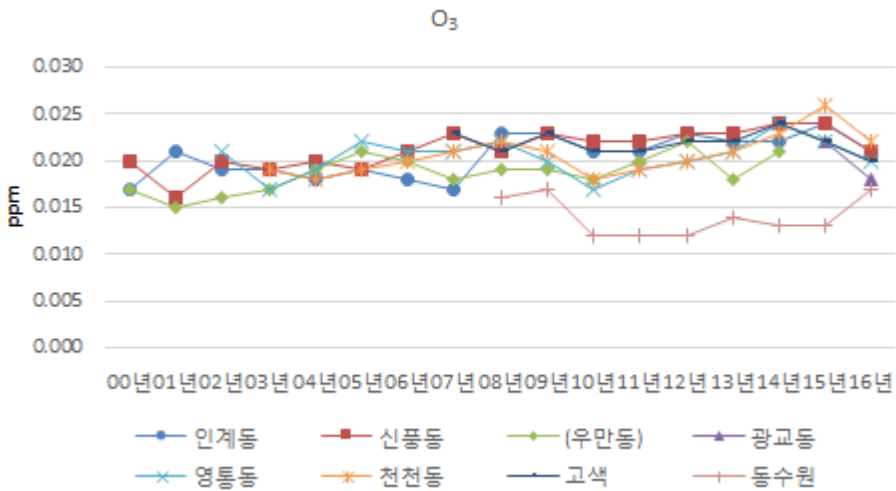
자료: 김득수, 2016. 서울 도심대기의 NO, NO₂와 O₃ 사이의 대기화학적 특성 연구, 한국대기환경학회지 제32권 제4호

<그림 3-14> O₃-NO_x 광분해 cycle



<그림 3-15> 수원시 O₃ 농도 변화 추이(2010.01~2016.12)

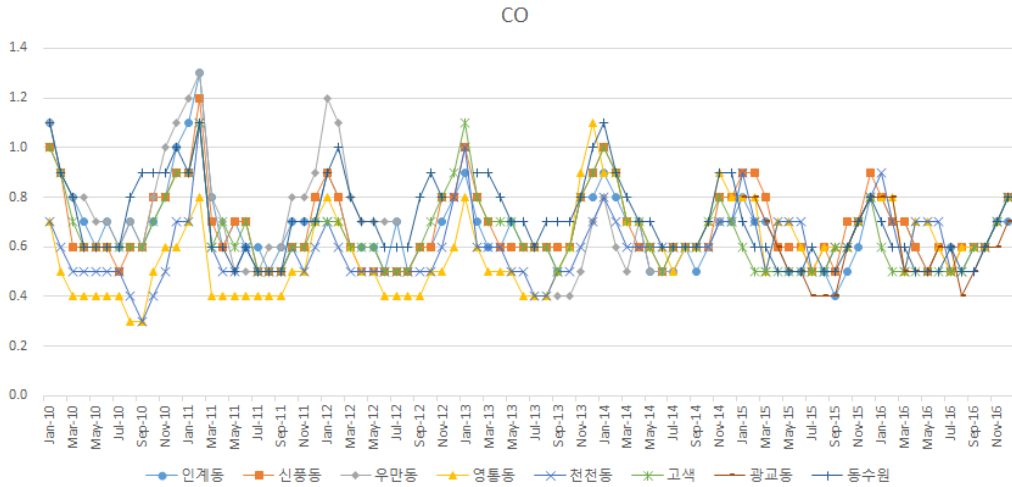
오존의 연도별 변화 추이를 보면 2000년부터 다소 증가하는 추세를 보이고 있으며, 질소 산화물 농도가 상대적으로 높았던 동수원 측정소에서 현저히 낮음을 확인 할 수 있다.



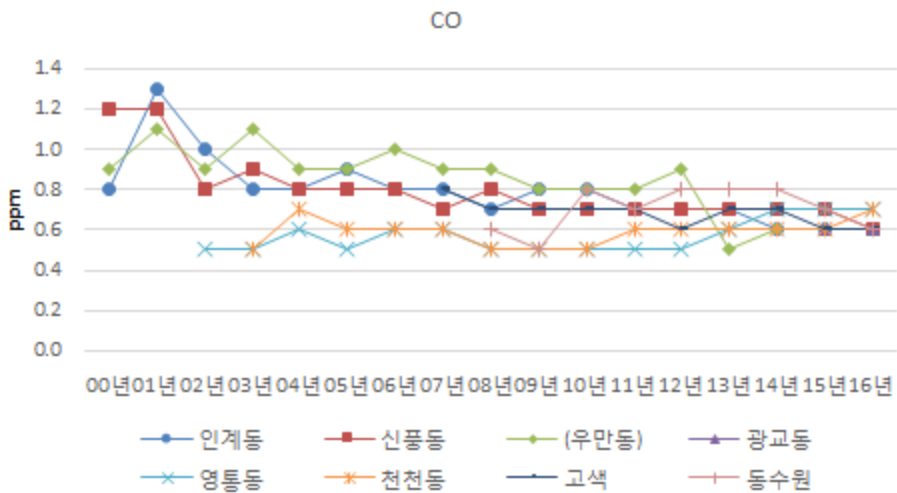
<그림 3-16> 수원시 O₃ 연도별 변화

6) 일산화탄소(CO)

일산화탄소 농도는 모든 측정소에서 비슷하게 나타나는 경향이 보인다. 연도별 농도 변화 추이를 보면 2000년부터 점차 감소되고 있다.



<그림 3-17> 수원시 CO 농도 변화 추이(2010.01~2016.12)

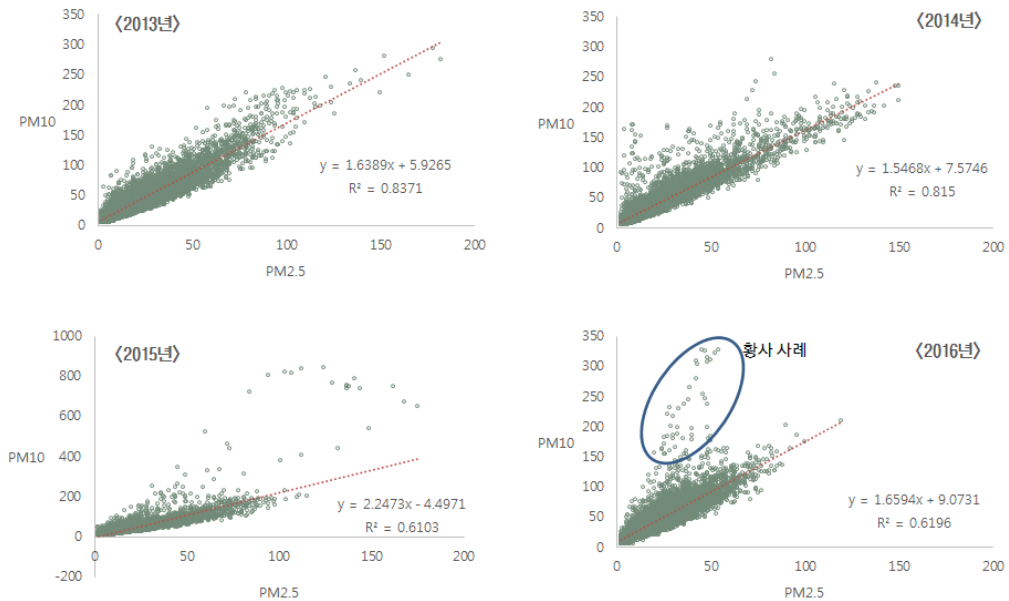


<그림 3-18> 수원시 CO 연도별 변화

4. 수원시 대기오염물질 상관성 분석

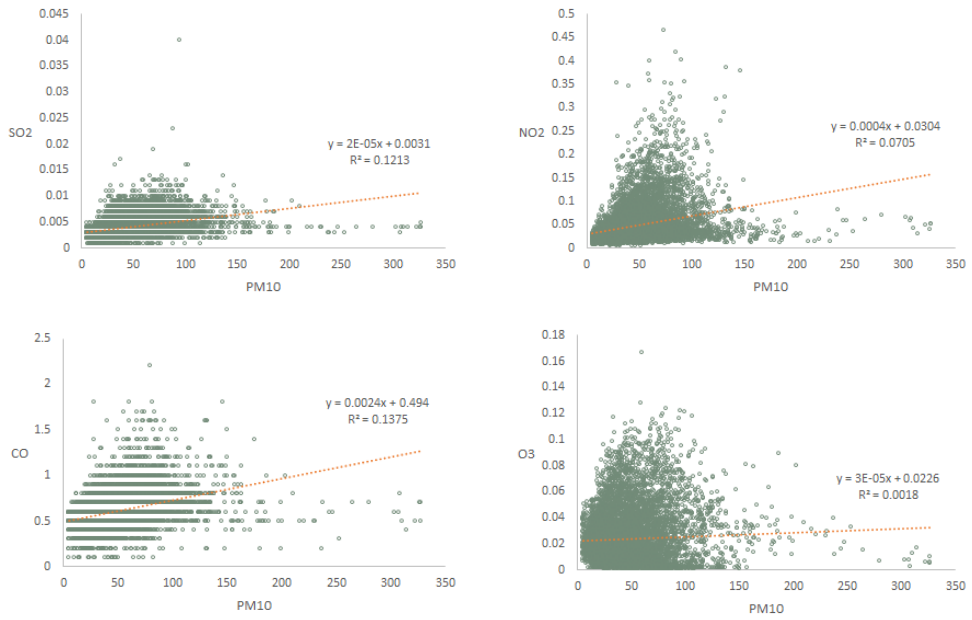
앞서 수원시 농도 변화를 살펴본 결과 도로변(동수원)측정소를 제외하고 다른 측정소간의 농도가 비슷하여 대표측정소로 인계동(수원시청사)을 선정·분석하였다.

2013년부터 2016까지 PM_{2.5}와 PM₁₀ 간의 상관성을 분석해 본 결과, 2013년과 2014년에 R² 값은 0.8로 상관성이 높게 나타났다. 그러나 2015년, 2016년의 경우 PM_{2.5}는 높지 않으나 PM₁₀만 높은 외부 황사 유입 사례 등이 있어 R²값이 다소 낮게 나왔다.

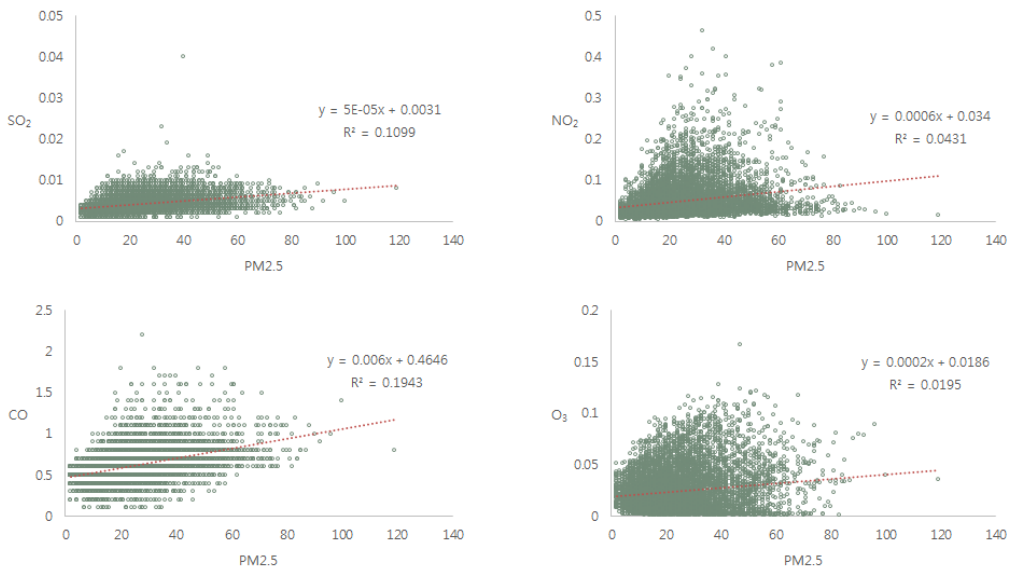


<그림 3-19> PM_{2.5}와 PM₁₀ 상관성 분석(2013-2016)

PM₁₀·PM_{2.5}와 가스상 오염물질 간의 상관성 분석은 다음과 같다. R²값이 약 0.002~0.2로 미세먼지와 가스상 오염물질 간의 상관성은 낮은 것으로 판단된다. 미세먼지 농도 증가에는 CO, SO₂ 등 양(positive)의 상관관계를 보이지만 단순한 상관성 분석으로는 오염물질 간 상관관계를 보기는 어려웠다. 그 중 CO가 높아질수록 PM₁₀과 PM_{2.5}가 높아지는 것은 볼 수 있었다.

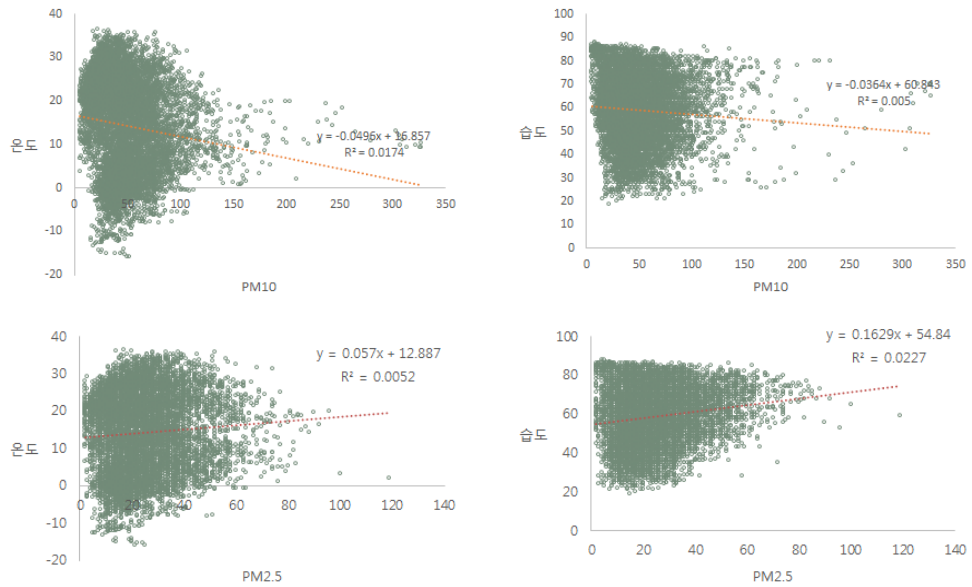


<그림 3-20> PM₁₀과 가스상 오염물질 간의 상관성 분석(2016)



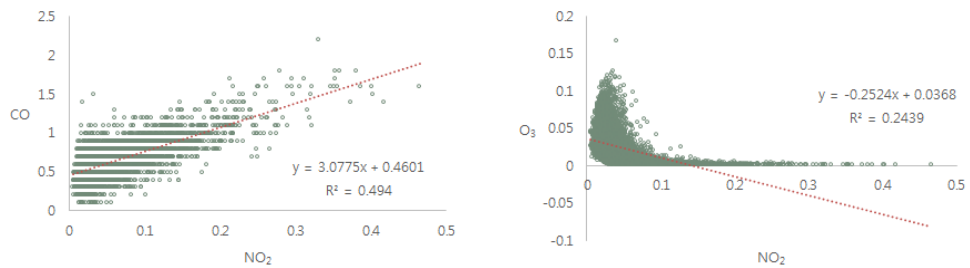
<그림 3-21> PM_{2.5}와 가스상 오염물질 간의 상관성 분석(2016)

PM₁₀과 온·습도 간의 상관성 분석 결과 음(-)의 상관관계를 보이는 반면, PM_{2.5}의 경우 온·습도와 양(+)의 상관관계를 보인다. 하지만 R²값이 너무 낮아 유의미한 상관관계라고 판단하기에는 어려웠다.



<그림 3-22> 미세먼지(PM₁₀, PM_{2.5})와 온·습도 간의 상관성 분석(2016)

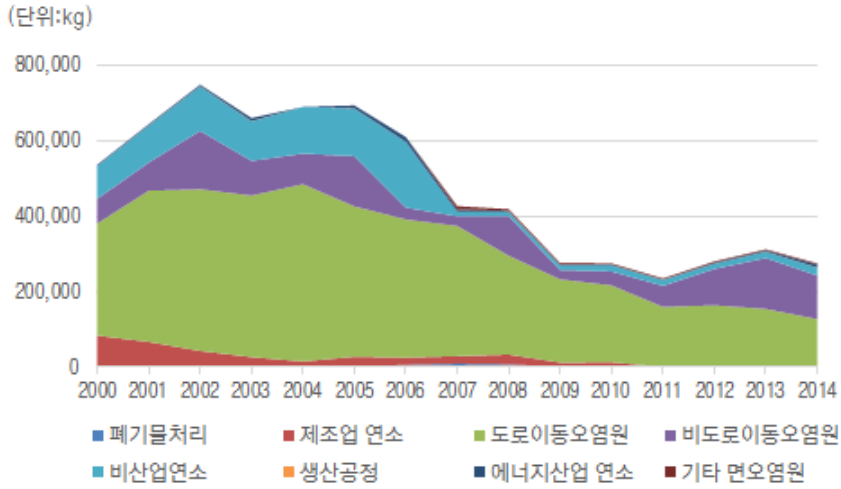
CO와 NO₂, O₃와 NO₂의 상관성 분석 결과 CO-NO₂는 양(+)의 상관관계를 보였으며, O₃-NO₂는 음(-)의 상관관계를 보였다. 이 중 CO와 NO₂와의 관계에서 R²값이 0.49로 유의미한 상관성을 띄었다.



<그림 3-23> CO와 NO₂, O₃와 NO₂의 상관성 분석(2016)

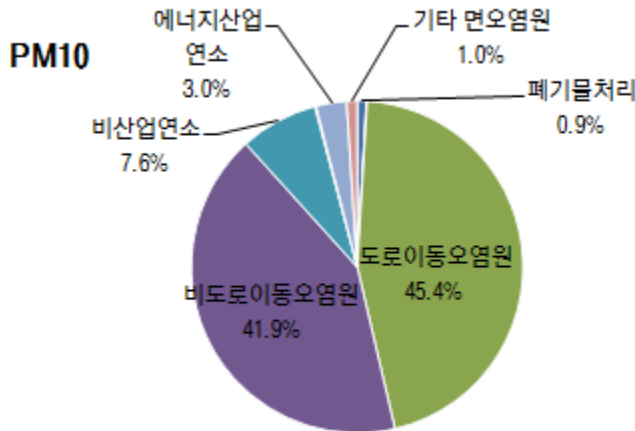
5. 수원시 미세먼지 배출량

수원시의 2000년에서 2014년까지 PM₁₀ 배출량 추이를 살펴보면, 증감을 반복하지만 배출량이 꾸준히 감소하는 추이를 보인다. 특히 도로이동오염원에서 많이 줄어든 것으로 나타났다.



<그림 3-24> 수원시 미세먼지 배출량 추세

2014년 PM₁₀ 배출원별 배출비율을 살펴보면 도로이동오염원이 전체배출량의 45.4%로 가장 많은 비중을 차지하고 있으며, 다음으로는 비도로이동오염원이 41.9%을 차지하고 있다.



<그림 3-25> 수원시 부문별 PM₁₀ 배출비율

미세먼지의 가장 많은 배출비율을 차지하는 도로이동오염원 배출원으로는 버스, RV, 승용차, 승합차, 이륜차, 택시, 특수차, 화물차가 포함되며, 비도로이동오염원 배출원은 건설장비, 농업기계, 철도가 포함된다.

제2절 시사점

수원시는 수도권 남부에 위치한 도심지역이기에 국가배경대기 측정소인 백령도보다 PM₁₀, PM_{2.5}, SO₂, NO₂, O₃, CO 의 농도가 모두 높았다. 이는 수원시 또는 수원시 인근, 나아가 대한민국 내 대기오염물질 배출에 의해 수원시 대기오염도가 높아졌음을 시사하였다. 수원시 서쪽에 인접해 있는 안산시와의 농도 비교에 있어서는 SO₂, CO의 농도 차이가 두드러졌으며, 오히려 체류시간이 긴 PM_{2.5}, PM₁₀의 경우 농도 차이가 크지 않았다. 다만 PM_{2.5}의 경우 SO₂와 유사하게 안산시 > 수원시 > 백령도의 순으로 농도가 높게 나타나 산업시설과 같은 SO₂ 배출이 높은 경우 PM_{2.5}도 높게 관측된다는 것을 알 수 있었다.

수원시 관내 7개 측정소의 대기오염도 현황을 보면, 도로변 측정소에서 PM₁₀, NO₂의 농도가 다른 도시대기 측정소보다 높아서 도로오염원 영향이 큼을 알 수 있었다. 그 외 6개 도시대기 측정소의 경우 인근 건설공사가 최근 잦은 광교동 측정소를 제외하고는 전반적으로 대기오염도가 비슷한 추세와 농도 분포를 보였다.

그러나 대기오염물질 간 선형회귀분석 결과, PM_{2.5}와 PM₁₀, CO와 NO₂ 짝을 제외하고는 다른 물질 간 상관성에서는 유의성을 찾을 수 없었다. PM_{2.5}와 PM₁₀, CO와 NO₂는 각각 R²=0.62, R²=0.49로 매우 유의한 양(+)의 상관성을 보였다. CO와 NO₂는 주요 배출원이 자동차 배기가스와 에너지 연소로 동일하기에 양의 상관성을 보였을 것이라 사료된다. 그러나, PM_{2.5}와 PM₁₀간의 상관성은 물질은 다른데 배출원이 동일하여서, 상대 오염물질의 배출원 추적에 활용될 수 있는 CO와 NO₂와의 관계 때문이 아니라, 물리적으로 입경이 10 μm보다 작은 에어러솔을 의미하는 PM₁₀에 PM_{2.5}(입경이 2.5μm보다 작은 에어러솔)가 포함되어 있는 이유가 주 원인일 것이라 사료된다. 그러나 PM₁₀이나 PM_{2.5}의 배출원 중 40%이상이 도로이동오염원임에도 불구하고 NO₂, CO나 PM_{2.5}, PM₁₀와는 단순 선형회귀분석 결과에서는 상관성을 찾을 수 없었다. 그 이유로는 광화학적 반응, 공기체의 이동 및 기상 조건의 변동 등이 반영되지 않았기 때문일 것이다.

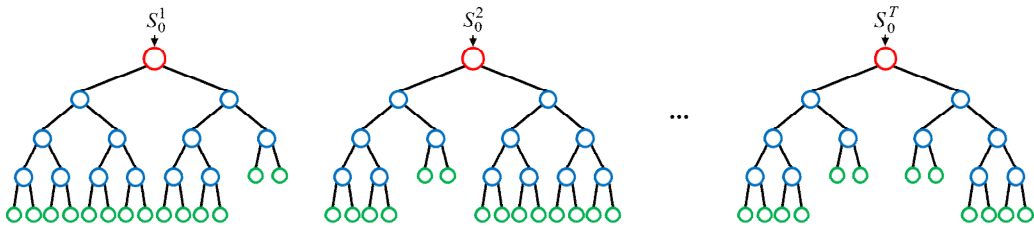
따라서 다음 장에서는 다중회귀분석이나 ANN(인공신경망) 기법을 사용하여 단순 선형회귀분석으로 찾을 수 없었던 대기오염물질 간 상관관계, 그리고 미세먼지의 증가에 기여하는 예상 요인과 미세먼지 농도 증가와의 상관성을 찾을 수 있는지를 알아보려고 한다.

제4장 ANN을 활용한 수원시 대기질 분석

제1절 ANN을 활용한 수원시 미세먼지 농도 예측

1. ANN 기법

랜덤 포레스트는 기계 학습(Machine learning)⁸⁾ 중 하나로서 부트스트랩 방식을 이용하여 다수의 표본을 생성하고 여러 개의 의사결정나무(decision tree) 모형을 적용하여 그 결과를 종합하는 방법이다. 랜덤포레스트 모형은 무작위성을 최대로 주기 위해 부트스트랩 표본을 생성할 뿐만 아니라 각 의사결정나무에서의 설명변수에 대해 무작위로 표본을 수집하는 방식이다. 이러한 무작위성이 최대로 되면 의사결정나무들 간에 상관관계가 줄어들게 되고 이로 인해 예측오차가 줄어든다. 특히 설명변수가 다수일 때 예측력이 매우 뛰어나며 안정적인 모형을 제공한다. 이러한 특징을 가진 랜덤포레스트는 분류, 회귀 분석 등에 주로 사용된다. 본 연구에서는 수원시 공공데이터를 활용하여 수원시 내의 미세먼지(PM_{10})를 예측하고자 랜덤포레스트를 이용한 예측 모델을 구축하였다. 본 연구에서는 오픈 소스 프로그래밍 언어 중 하나인 python을 사용하였고 모델 구축을 위해 RandomForestRegressor 함수를 사용하였다.



<그림 4-1> 랜덤포레스트 시각화

2. 분석을 위한 자료구축 및 유사 활용 사례

1) 대기오염농도 자료

수원시 7개소 대기오염측정망 PM_{10} , $PM_{2.5}$, SO_2 , NO_2 , O_3 , CO 의 시간대별 데이터(2016년)를 구축하였다. 자료에 대한 전반적인 분석은 3장에서 기술하였다.

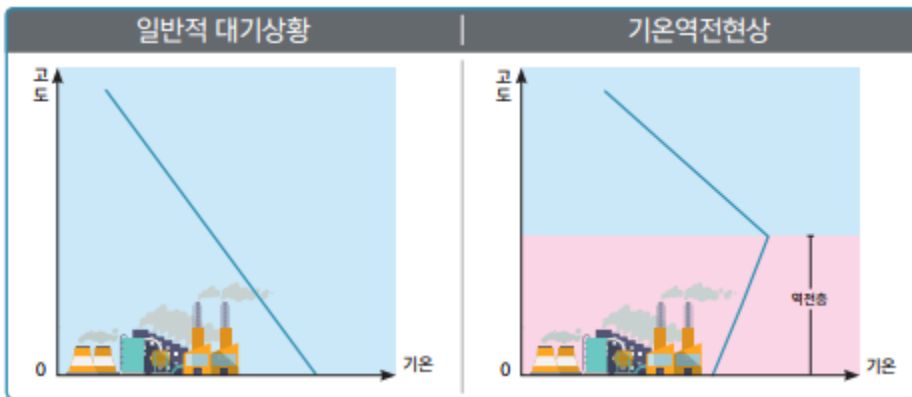
8) 머신러닝은 인공지능을 구현하는 구체적 접근 방식이다. 기본적으로 알고리즘을 이용해 데이터를 분석하고, 분석을 통해 학습하며, 학습한 내용을 기반으로 판단이나 예측을 한다.

2) 기상자료

(1) 미세먼지와 기상과의 관계

대기 중 기온은 일반적으로 고도가 100 m 높아질 때마다 약 0.6 °C씩 낮아진다. 그런데 거꾸로 고도가 높아질수록 기온이 올라가는 현상이 나타나기도 하는데 이를 ‘기온역전’이라 한다. 기온역전은 일교차가 큰 계절이나 산간분지지역에서 자주 발생하곤 한다.

대개 공기는 더울수록 밀도가 낮아져 더운 공기는 위로, 차가운 공기는 아래로 이동한다. 그러나 기온역전이 발생하면 고도가 낮은 쪽에 무거운 공기가, 높은 쪽에 가벼운 공기가 위치해 무게 차에 의한 공기의 상하이동이 일어나지 않는다. 그러면 지상에서 주로 발생하는 미세먼지 등 대기오염물질이 지상층에 머무르게 되고 계속하여 쌓이면서 그 농도가 높아지게 된다(환경부, 2016).



자료: 환경부, 2016, 바로 알면 보인다. 미세먼지, 도대체 뭘까?

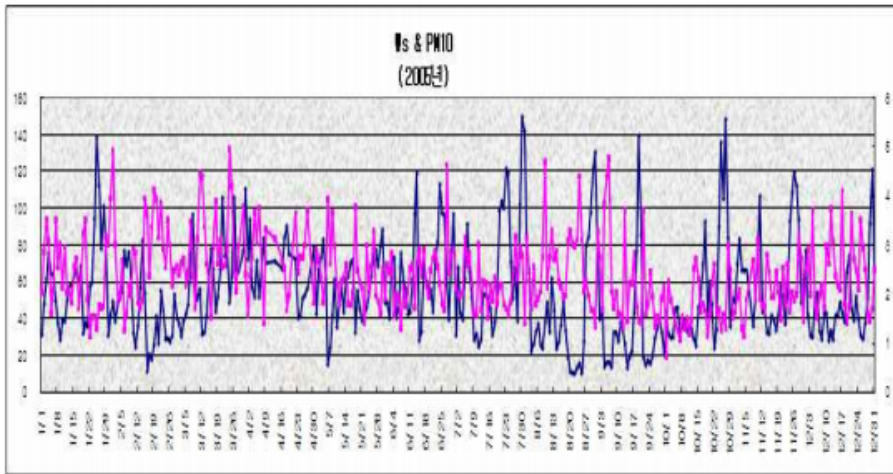
<그림 4-2> 기온역전현상

기온 역전현상 뿐만 아니라 낮밤 지면 기온차에 의해 대기 혼합 경계층(mixing boundary layer)은 밤에 주로 낮고 낮에 높아지게 되는데, 미세먼지의 경우 국지배출이나 생성 없이 밤에 높고 낮에 낮은 상황이 발생하기도 한다.

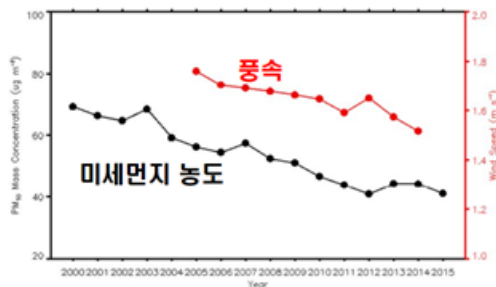
(2) 기상자료와 미세먼지 상관관계 연구 사례

대기질 모델링 시 화학물질 조성, 수송, 제거 과정 등을 고려하기 위해서 시간과 공간에 대한 기상자료는 필수적인 입력 자료이다. 경미한 기상자료의 변화도 대기질 모델링 입력 자료로 사용될 경우 막대한 영향을 가져오게 된다(Dolwick, 2002).

풍속이 미세먼지에 미치는 영향을 분석한 사례⁹⁾(채희정, 2009)에서는 서울시에서 비행사일의 평균 풍속과 PM10의 변화추이는 계절과 상관없이 풍속이 강하면 PM10이 낮아지는 경향을 보였다. PM10 이 높게 유지되던 날도 풍속이 강해지면 PM10은 낮아졌다. 김상우(2016)도 ‘미세먼지와 기후변화’ 연구에서 과거 15년 자료를 통해 PM10 농도가 낮아지는 경향과 풍속이 낮아지는 경향이 유사한 것을 보았다.



<그림 4-3> 2005년 WS & PM₁₀



출처: 김상우(2016) '미세먼지와 기후변화'

<그림 4-4> 서울 미세먼지 농도와 풍속의 연변화

김순태(2016)의 ‘국내 on-road에서의 미세먼지 발생 및 현상규명’ 사례¹⁰⁾에서는 미세먼지 농도는 지역 풍속과 강한 음의 상관관계를 보이며 지역 환기 감소는 한국의 심각한 미

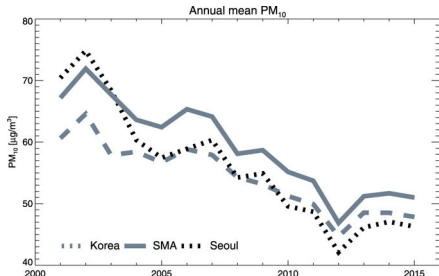
9) 채희정, 2009, 풍속과 풍향이 미세먼지에 미치는 영향, 서울시

10) 김순태, 2016, ‘국내 on-road에서의 미세먼지 발생 및 현상규명’, 한국대기환경학회 2017공동학술심포지엄

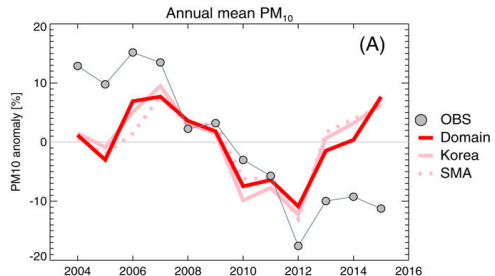
세먼지를 유발하는 정체된 조건과 관련이 있음을 기술하였다.

앞 부분에서 살펴본 바와 같이 기상 조건은 오염 물질 및 그 전구물질의 방출량의 변화뿐만 아니라 지역 헤이즈 및 대기 오염의 주요 원인이 되기도 한다. 바람은 수직 혼합 및 지역 환기를 제어 할 수 있기 때문에. 잔잔한 바람에 의한 지속적 정체 상태는 지역 오염 증가의 발달로 이어지는 치명적인 조건을 제공한다. 역사적인 대기오염사건인 런던스모그도 정체된 공기로 인해 배출된 대기오염물질이 확산되지 못하고 지속적으로 쌓이면서 대규모 인명피해가 발생하였다.

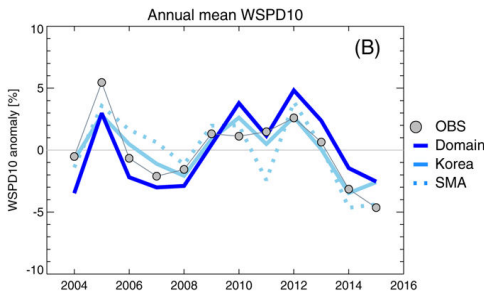
김순태(2016)의 연구에서는 실제로 여러 기상 변수를 테스트 한 결과, 풍속과 PM 농도의 연관성을 발견하였다. PM 농도와 평균화 된 10 m 풍속의 정규화 된 예외(연간 평균을 빼 다음 2004-2015 평균으로 나눈 값)를 비교한 결과, PM 농도의 경년 변동은 모든 지역에서 매우 유사한 패턴을 나타내었다(그림 4-6, 4-7). 모델로부터의 풍속과 표면 PM 농도 사이의 정규화 된 변칙의 산포 도표 <그림 4-8>에서는 최소 제곱 회귀선 적합도는 $PM [\%] = -2.1 WS [\%]$ 이며 피어슨 상관 계수 $R = -0.86$ 으로 상관성이 매우 높았다.



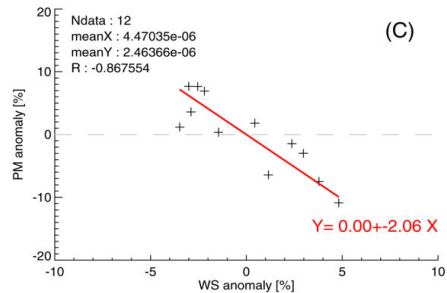
<그림 4-5> PM₁₀ 농도의 시계열(전국, 수도권, 서울)



<그림 4-6> 연간 평균 PM₁₀(%)



<그림 4-7> 10m 풍속 변화



<그림 4-8> PM농도와 풍속의 회귀분석

(3) 모델에 사용한 기상자료

수원시 기상을 대표할 수 있는 수원기상대와 대기오염측정소에 같이 설치되어 있는 AWS의 온도, 습도, 풍향, 풍속 자료를 구축하였다.

3) 도로교통량(국지오염원)

(1) 미세먼지와 교통량과의 관계 연구 사례

임영욱(2013) ‘자동차의 미세먼지의 건강영향 및 관리방안’¹¹⁾에서는 실제 도로에서 미세먼지 농도를 측정하여 분석하였다.

외부 대기질의 노출수준에 따른 오염물질의 유입정도 및 교통량 변화에 따른 유입 수준을 평가하기 위해 도로상 이동측정(5회 반복측정)을 하였으며 오전 9시에서 10시 사이에 미세먼지가 고농도로 측정되었다(〈표 4-1〉).

〈표 4-1〉 도로상 이동측정 결과

측정시간	PM ₁₀ ($\mu\text{g}/\text{m}^3$)	PM _{2.5} ($\mu\text{g}/\text{m}^3$)	PM _{1.0} ($\mu\text{g}/\text{m}^3$)
오전9시~오전10시	108±189	26±19	6±4
오전12시~오후1시	44±50	12±7	3±2
오후3시~오후4시	42±28	11±6	3±1

*2013년, 강남 르네상스호텔 사거리
자료: 임영욱, 2017. ‘자동차의 미세먼지의 건강영향 및 관리방안’, 연세대학교 환경공해연구소. 한국대기환경학회, 2017 공동학술심포지엄

교통 밀집지역에서 도로이격거리별 미세먼지 농도를 측정한 결과는 〈표 4-2〉와 같다. 서울 강남구 르네상스호텔 사거리를 중심으로 100 m별 이격거리를 두어 중심가 지점과 외곽 지역 지점을 설정하여 한 노출평가에서는 중심가 지역의 미세먼지농도가 높게 나타났다.

11) 임영욱, 2017. ‘자동차의 미세먼지의 건강영향 및 관리방안’, 연세대학교 환경공해연구소. 한국대기환경학회 2017 공동학술심포지엄

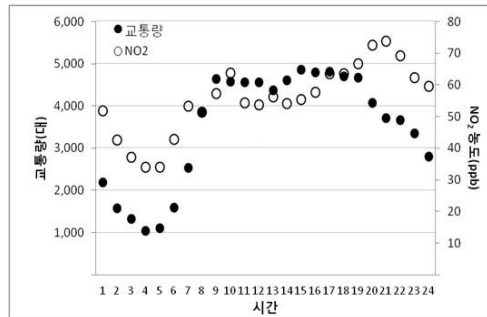
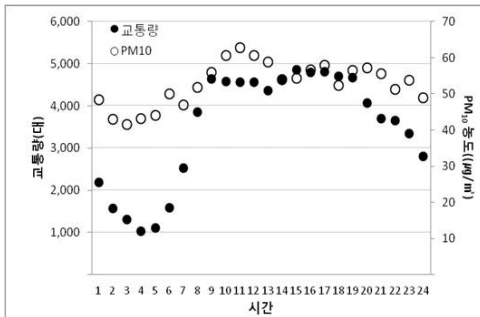
<표 4-2> 도로 이격거리별 측정결과

이격거리	PM ₁₀ ($\mu\text{g}/\text{m}^3$)	PM _{2.5} ($\mu\text{g}/\text{m}^3$)	PM _{1.0} ($\mu\text{g}/\text{m}^3$)
100m 이내	127	69	25
100m < 지점 < 200m	79	45	17
200m 이상	71	37	15

자료: 임영욱, 2017. ‘자동차의 미세먼지의 건강영향 및 관리방안’, 연세대학교 환경공해연구소. 한국대기환경학회, 2017 공동학술심포지엄

국립환경과학원(2012)에서 수행한 ‘교통밀집지역 자동차 통행량 증감에 따른 농도변화 평가’에서는 2010년 10월 11일 ~ 10월 17일 일주일 동안 서울역에 위치한 도로변 대기측정소와 회현역 주변의 교통량 조사지점에서 조사된 교통량을 활용하여 차량통행량과 대기 중 오염물질 농도사이의 관계를 평가하였다.

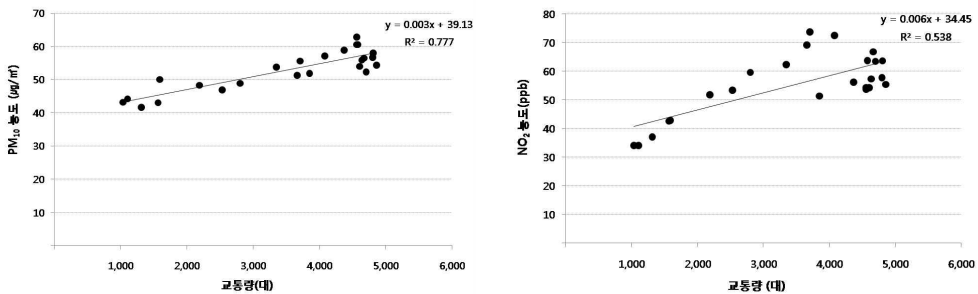
교통량은 일반적인 출근시간대인 오전 7시부터 3,000대 이상으로 20시까지 지속적으로 증가한 후 20시 이후 서서히 감소하는 경향을 나타내었다. <그림 4-9>에 나타난 바와 같이 분석 대상기간동안의 시간대별 PM₁₀의 평균 농도 변화 추이는 오전 7시 이후 교통량이 증가함에 따라 증가하다가 오전 11시 ~ 12시 사이에 대기 중 농도가 가장 높았으며, 이후에는 53 $\mu\text{g}/\text{m}^3$ ~ 60 $\mu\text{g}/\text{m}^3$ 범위 내에서 증감을 반복하였다. 한편 분석 대상기간 동안의 시간대별 NO₂의 평균 농도 변화 추이는 PM₁₀과 마찬가지로 교통량에 따라 증가하는 경향을 나타내었다. 낮 시간에는 교통량의 증감이 둔화되고, NO₂의 농도는 크게 변화하지 않았다. 반면 하루 일과가 종료되는 20시 이후 교통량 감소와 함께 밤시간 대 오존분해반응이 시작되면서 에 의해 NO₂의 농도가 증가하였다가 교통량 감소와 함께 다시 낮아지는 것을 보였다.



자료: 국립환경과학원, 2012, 교통밀집지역 자동차 통행량 증감에 따른 농도변화 평가

<그림 4-9> 분석 기간 동안의 시간대별 평균 교통량 및 시간평균 PM₁₀ 농도(좌), NO₂농도(우)

상관성 분석을 실시한 결과 교통량 증가에 따른 PM₁₀ 농도는 선형적인 비례관계를 나타내었다. 한편 NO₂ 농도의 경우 다음 그림에 나타난 바와 같이 시간대별 평균 교통량과 NO₂ 평균 농도 사이에 R²값이 0.538로 나타나 이 두 인자 사이에 선형적인 상관관계가 있는 것으로 판단되나 PM₁₀의 경우에 비해 그 상관성은 낮았다. 이 연구는 일주일이라는 제한된 시간 동안, 강수나 바람의 급변화 같은 특정 기상요인의 영향이 없는 상황에서 교통량과 대기오염물질간의 상관성을 분석한 것이라 다른 연구결과들과 비교하였을 때 상대적으로 뚜렷한 상관성을 얻을 수 있었을 것이다.



자료: 국립환경과학원, 2012. 교통밀집지역 자동차 통행량 증가에 따른 농도변화 평가

<그림 4-10> 서울역 지점에서의 일일 교통량과 일평균 PM₁₀(좌), NO₂(우) 농도의 상관관계

조수석 등(2011)이 수행한 교통량에 따른 도로변 대기질 영향에서 24시간 동안의 교통량과 대기오염물질 상관관계 분석결과 CO > NO_x > SO₂ > NO₂ > O₃ > PM₁₀ > PM_{2.5} 순으로 감소하였다. 교통량과 대기오염물질의 상관관계는 입자상 오염물질보다는 가스상 오염물질에서 더 높은 상관성을 나타냈다.¹²⁾

<표 4-3> 교통량과 대기오염물질 간의 상관성 분석

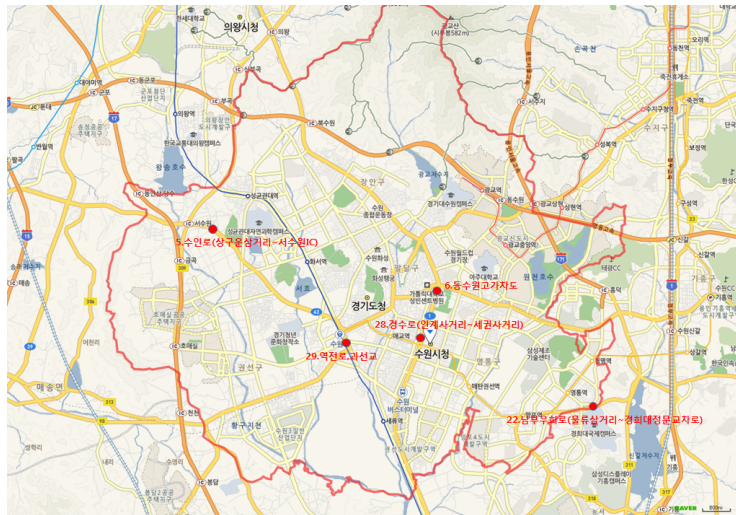
	SO ₂	NO ₂	NO _x	CO	O ₃	PM ₁₀	PM _{2.5}	Traffic
SO ₂	1	0.464	0.423	0.487	0.002	0.532	0.474	0.345
NO ₂		1	0.848	0.710	-0.277	0.447	0.425	0.303
NO _x			1	0.709	-0.433	0.287	0.276	0.355
CO				1	-0.310	0.468	0.492	0.380
O ₃					1	0.017	-0.128	0.252
PM ₁₀						1	0.781	0.168
PM _{2.5}							1	0.052
Traffic								1

자료 : 조수석 외, 2011. 기상조건 및 교통량에 따른 도로변 대기질 영향, 서울특별시보건환경연구원보 제47호

12) 조수석 외, 2011. 기상조건 및 교통량에 따른 도로변 대기질 영향, 서울특별시보건환경연구원보 제47호

(2) 교통량 자료 구축

미세먼지와 교통량과의 관계를 알아보기 위해 국가대기오염측정망 주변의 시간대별 교통량 자료(수원시 도시안전통합센터)를 수집하였다. 대기오염측정소 주변과 가까우면서 교통 혼잡지역 등을 고려하여 교통량 자료를 구할 수 있는 지점을 선정하였으며 다음 그림에 위치를 나타내었다.



<그림 4-11> 교통량 자료 위치

제2절 대기질 영향요인과 대기오염물질에 따른 미세먼지 농도 예측

1. 분석 자료 및 방법

1) 분석 자료

수원시 도로 교통량 5지점과 수원시 대기오염 측정소 7지점의 미세먼지 농도, 기상대 자료간의 통계분석을 실시한 결과 상관성을 찾아볼 수 없었다. 측정소와 교통량 자료의 위치가 근접한 지점이 아닌 곳도 있고, 미세먼지 유발요인은 기상조건 뿐만 아니라 지역 배출원과 장거리 수송된 외부 배출원의 영향 등 많기 때문에 교통량의 영향이 너무 작게 차지하는 것으로 나타나는 것이 원인일 수도 있었다. 따라서 주요 결과로는 측정소와 교통량 자료의 위치가 제일 근접하며 도로변 대기를 측정하는 동수원 지점을 선정하여 국지적 요인이 미세먼지 농도에 기여하는 바를 알아보하고자 하였다.



<그림 4-12> 대기오염측정망과 교통량 자료 측정 위치

모델링을 위해 동수원에 위치해있는 도로변 대기측정망에서 추출된 PM₁₀, SO₂, NO₂, CO, O₃, 온도, 습도, 풍속, 풍향 데이터와 그 주변 동수원고가차도의 교통량 데이터(소형, 중형,

대형 차량 통과대수를 변수로 이용하였다(자동차 규모별 세부기준은 <표 4-4>을 참고). 자료는 수원시에서 공공데이터로 제공하는 2016년 1월 1일부터 12월 31일까지 1시간 단위의 데이터를 이용하였다. 초기 변수와 전체 관찰 값의 개수는 각각 13개와 8,784개였으며, 데이터 전처리 후 최종적으로 모델에 사용된 변수와 관찰 값의 개수는 각각 30개와 7,077개로 예측 모델 구축을 위해 Training Data와 Test Data를 각각 70%, 30%로 구분하였다.

No	datetime	PM10	SO2	NOX	CO	O3	Temp	Humi	car_count	car_small	car_medium	car_large	windspeed	winddirection
1	2016.1.1 0:00	92	0.01	0.276	1.6	0.004	0	73	1698	1243	418	37	0.5	NW
2	2016.1.1 1:00	102	0.009	0.25	1.8	0.004	-0.3	75	1632	1219	395	18	0.2	NNE
3	2016.1.1 2:00	93	0.009	0.212	1.6	0.004	-0.7	75	1085	884	190	11	0.1	NNE
4	2016.1.1 3:00	91	0.008	0.186	1.3	0.004	-1.1	76	748	607	139	2	0.1	NNW
5	2016.1.1 4:00	90	0.008	0.197	1.4	0.004	-0.9	77	694	578	114	2	0.2	NNE
6	2016.1.1 5:00	96	0.009	0.216	1.4	0.004	-1.2	78	847	719	123	5	0.3	NNE
7	2016.1.1 6:00	94	0.009	0.263	1.3	0.004	-1.1	79	1076	869	148	9	0.4	NNE
8	2016.1.1 7:00	87	0.008	0.201	1.2	0.004	-0.9	77	907	558	335	14	0.6	NNW
9	2016.1.1 8:00	89	0.008	0.196	1.1	0.004	-0.5	74	1218	509	693	16	0.5	NNW
10	2016.1.1 9:00	94	0.008	0.185	1	0.004	0	72	1588	694	873	21	1	NNE
11	2016.1.1 10:00	92	0.009	0.198	1	0.004	1	68	1794	813	971	10	0.4	NNW
12	2016.1.1 11:00	97	0.008	0.174	1.1	0.005	2.5	65	2170	1027	1125	18	0.7	WSW
13	2016.1.1 12:00	95	0.008	0.113	0.9	0.006	3.8	61	2362	1161	1181	20	1	SW
14	2016.1.1 13:00	85	0.006	0.068	0.6	0.01	4.2	58	2262	1045	1203	14	1.1	SW

<그림 4-13> 활용된 변수들의 Raw Data

〈표 4-4〉 자동차 규모별 세부기준

종류	경형	소형	중형	대형
승용 자동차	배기량이 1000cc 미만 으로서 길이 3.6m · 너 비 1.6m · 높이 2.0m 이 하인 것	배기량이 1,600cc 미만 인 것으로서 길이 4.7m · 너비 1.7m · 높 이 2.0m 이하인 것	배기량이 1,600cc 이상 2,000cc 미만이거나 길 이 · 너비 · 높이 중 어 느 하나라도 소형을 초 과하는 것	배기량이 2,000cc 이상 이거나, 길이 · 너비 · 높이 모두 소형을 초과 하는 것
승합 자동차	배기량이 1000cc 미만 으로서 길이 3.6m · 너 비 1.6m · 높이 2.0m 이 하인 것	승차정원이 15인 이하 인 것으로서 길이 4.7m · 너비 1.7m · 높 이 2.0m 이하인 것	승차정원이 16인 이상 35인 이하이거나, 길 이 · 너비 · 높이 중 어느 하나라도 소형을 초과하 여 길이가 9m 미만인 것	승차정원이 36인 이상 이거나, 길이 · 너비 · 높이 모두가 소형을 초 과하여 길이가 9m 이상 인 것
화물 자동차	배기량이 1000cc 미만 으로서 길이 3.6m · 너 비 1.6m · 높이 2.0m 이 하인 것	최대적재량이 1ton 이 하인 것으로서, 총중량 이 3.5ton 이하인 것	최대적재량이 1ton 초 과 5ton 미만이거나, 총 중량이 3.5ton 초과 10ton 미만인 것	최대적재량이 5ton 이 상이거나, 총중량이 10ton 이상인 것
특수 자동차	배기량이 1,000cc 미만 으로서 길이 3.6m · 너 비 1.6m · 높이 2.0m 이 하인 것	총중량이 3.5ton 이하인 것	총중량이 3.5ton 초과 10ton 미만인 것	총중량이 10ton 이상인 것
이륜 자동차	배기량이 50cc 미만(최 고정격출력 4kW 이하) 인 것	배기량이 100cc 이하 (최고정격출력 11kW 이 하)인 것으로 최대적재 량(기타형에만 해당한 다)이 60kg 이하인 것	배기량이 100cc 초과 260cc 이하(최고정격 출력 11kW 초과 15kW 이 하)인 것으로 최대적재 량이 60kg 초과 100kg 이하인 것	배기량이 260cc (최고 정격출력 15kW)를 초과 하는 것

자료: 「자동차관리법」 시행규칙 별표1

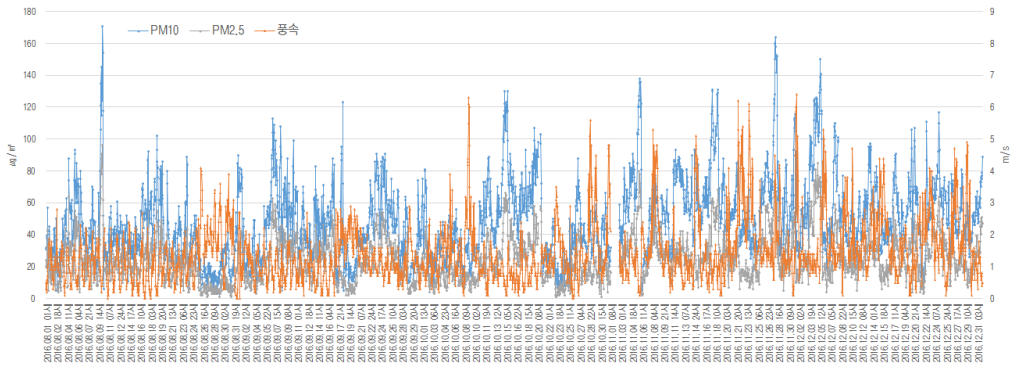
2) 분석 방법

미세먼지에의 영향요인을 찾기 위해 대기오염물질 및 영향요인(기상, 교통량)간 선형회귀분석, 다중 요인으로부터의 영향력을 찾기 위한 다중선형회귀분석, 그리고 ANN 기법(랜덤포레스트 방법)을 통한 영향요인 분석을 하였다.

2. 분석결과

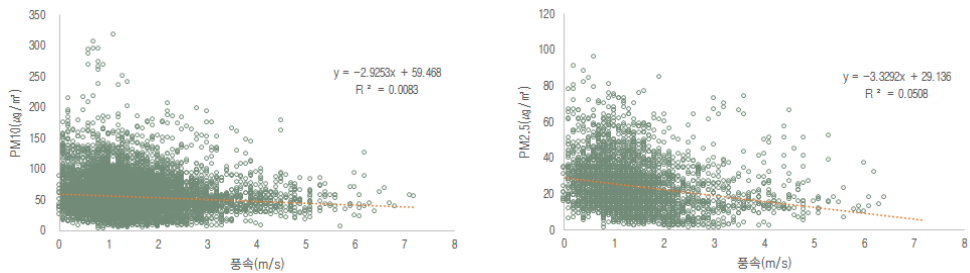
1) 단순 상관관계: 기상과 미세먼지농도

동수원 측정소의 PM₁₀, PM_{2.5} 농도와 풍속을 시계열로 나타내었다. 2016년의 PM_{2.5} 농도는 1월부터 7월까지의 기기교정으로 인한 결측치가 있어 8월부터 12월까지 비교하였다. 항상 그런 것은 아니지만 풍속이 높을 때 미세먼지농도가 낮은 경우를 확인 할 수 있었다.



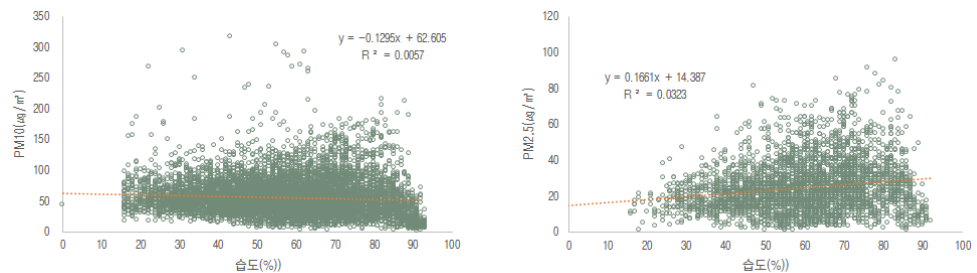
<그림 4-14> PM₁₀ · PM_{2.5} 농도와 풍속의 변화 추이(2016년 8월~12월)

풍속과 PM₁₀, PM_{2.5}의 상관성을 분석한 결과 음의 관계를 보인다. R²값이 높지 않기 때문에 상관성이 높지는 않지만 미세먼지 농도가 유난히 높을 때 풍속이 낮을 때가 있었다. PM_{2.5}가 풍속의 영향을 더 많이 받는 것으로 나타났다.



<그림 4-15> 풍속과 미세먼지농도의 상관성

습도에 따른 PM₁₀, PM_{2.5}의 변화를 보면 상관성이 높지는 않으나 PM_{2.5}농도는 습도가 높아질수록 높아지는 경향이 보이며, PM₁₀은 낮아지는 경향이 보인다.

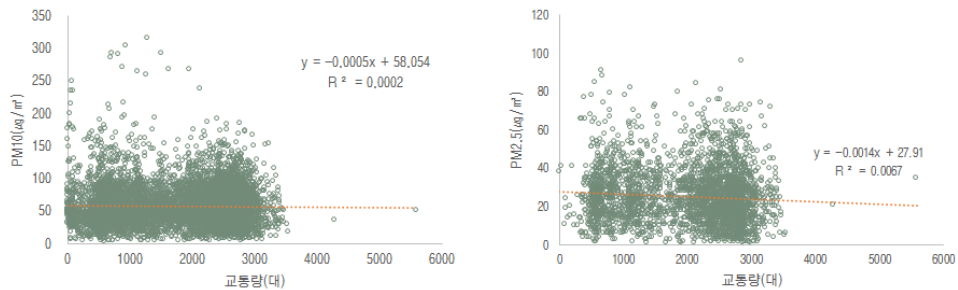


<그림 4-16> 습도와 미세먼지농도의 상관성

2) 단순 상관관계: 교통량과 대기오염

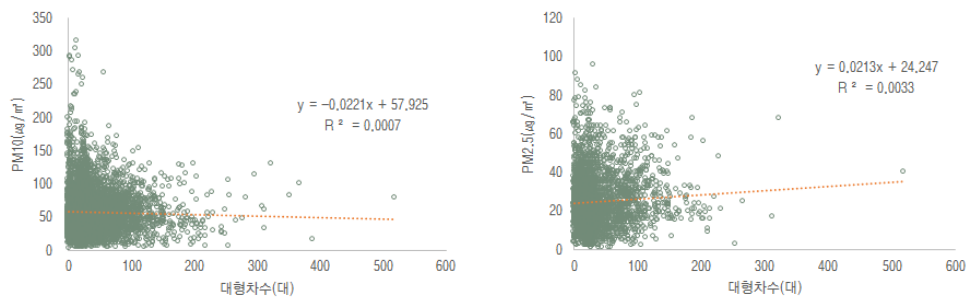
교통량과 미세먼지 농도의 상관성을 보면 예상과는 다르게 상관성이 높지는 않으나 $PM_{2.5}$ 가 PM_{10} 보다 교통량에 따른 변화가 있는 것으로 보인다.

교통량은 시간대별 다른 추이를 보이며 대기오염농도는 기상조건 등에 의해서도 변화되고 한 가지 요인으로만 설명되기에는 한계점이 있기 때문에 상관성이 낮은 것으로 판단된다.



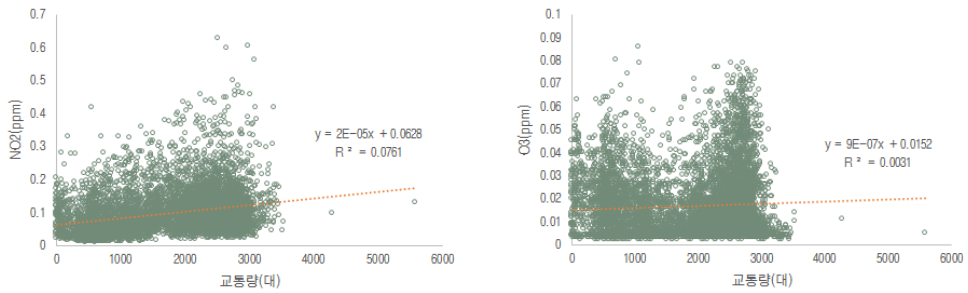
<그림 4-17> 교통량과 미세먼지농도(PM_{10} , $PM_{2.5}$)의 상관성

대형차량의 교통량과 미세먼지 농도 상관성을 보면 PM_{10} 보다 $PM_{2.5}$ 가 대형차량의 영향을 더 받는 것으로 보인다.

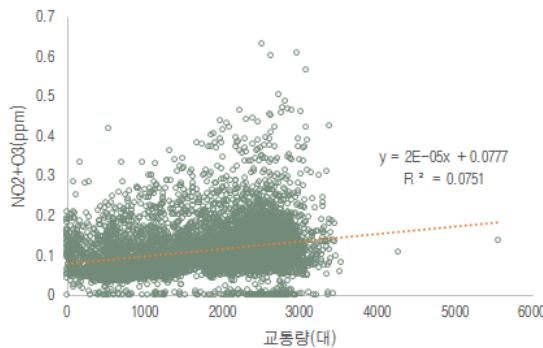


<그림 4-18> 대형차 교통량과 미세먼지농도(PM_{10} , $PM_{2.5}$)의 상관성

교통량과 NO_2 , O_3 의 농도 상관성을 보면 차량이 증가하면 NO_2 가 증가하는 경향이 보이지만 상관성이 높지는 않다.



<그림 4-19> 교통량과 NO₂, O₃농도의 상관성



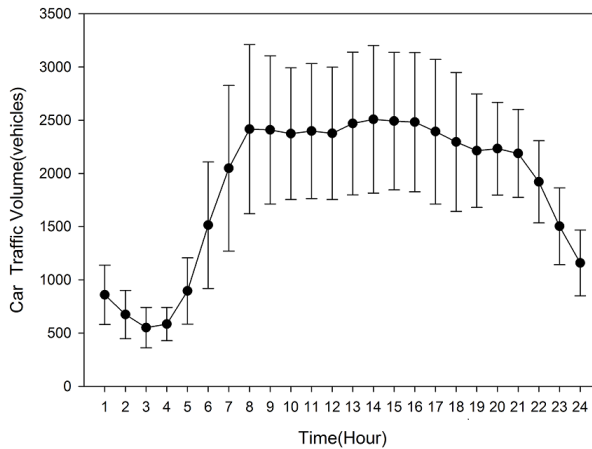
<그림 4-19> 교통량과 NO₂+O₃농도의 상관성

영향요인과 미세먼지와와의 단순 선형회귀분석만으로는 통계적으로 유의한 상관성을 찾을 수는 없었다. 하지만, 풍속이 낮거나 습도가 높으면 PM_{2.5}가 높아지는 경향이 보였는데 이는 풍속이 낮을수록 안정한 대기층으로 인해 대기오염물질이 확산되지 못하고 쌓이게 되는 물리적인 특성으로 설명될 수 있을 것이다. 또한 습도가 높을 때 PM_{2.5}가 높아진 것은 PM_{2.5}의 주성분이 이차생성 에어로졸(주로 황산화물, 유기화합물, 질산화물 등)로 이들 성분이 흡습성이 높은 특성을 갖고 있는 것으로 설명될 수 있다.

교통량과 미세먼지, 질소산화물 농도와의 상관성은 특히 보기 어려웠으나, 대형차량 수가 많을 경우 PM_{2.5}가 높아지는 현상, 교통량이 늘어날수록 NO₂가 높아지는 경향 등은 볼 수 있었다. 이는 도로교통량이 광화학반응이 연계된 대기오염도에 기여를 하고 있음을, 그러나 배출원과 반응산물 사이에 다양한 반응 및 기작이 있기 때문에 선형관계가 나타나지 않음을 시사한다.

동수원 고가차도 교통량 시간대별 패턴을 분석해 보면 다음 <그림 4-20>과 같았다. 즉, 통량은 요일별 시간대별 패턴이 비슷하였다. 수원시 동수원 고가차도의 2016년 시간대별

통행량 그래프를 보면 출근 시간대인 오전에 차량이 급격히 증가하다가 오후 8시 이후로 감소하는 경향으로 항상 비슷한 패턴임을 확인 할 수 있다. 그러나 대기오염물질 농도는 그날의 기상 조건, 계절적 요인 등 증감요인이 다양하기 때문에 단순히 교통량이 많으면 미세먼지 농도가 무조건 높아진다고 판단하기에는 어려움이 있음은 이 자료를 통해서도 알 수 있었다.



<그림 4-20> 2016년 수원시의 시간대별 교통량 평균 추이

3) 다중회귀분석

먼저 각 변수들 간의 상관계수와 유의확률은 다음과 같다. 모든 변수들 간의 상관계수에 대한 유의확률이 유의한 것으로 확인되고 있으며, PM_{10} -교통량의 유의확률은 $p > 0.05$ 로 다른 변수들에 비해 유의하지 않았다.

PM_{10} - $PM_{2.5}$ 의 상관계수는 0.898로 뚜렷한 유의성을 보였으며, 두 번째로 유의성이 높은 것은 NO_2 -CO로 상관계수가 0.686으로 나타났다.

〈표 4-5〉 변수들 간의 상관계수

		PM ₁₀	PM _{2.5}	SO ₂	NO ₂	CO	O ₃	온도	습도	풍속	교통량
상관 성	PM ₁₀	1.000	.898	.446	.421	.592	-.129	-.204	.065	-.104	.021
	PM _{2.5}	.898	1.000	.363	.332	.554	-.096	-.119	.190	-.210	-.081
	SO ₂	.446	.363	1.000	.686	.616	-.287	-.253	-.104	-.141	.115
	NO ₂	.421	.332	.686	1.000	.825	-.386	-.328	-.104	-.196	.264
	CO	.592	.554	.616	.825	1.000	-.297	-.238	.108	-.292	.050
	O ₃	-.129	-.096	-.287	-.386	-.297	1.000	.507	-.125	.163	.150
	온도	-.204	-.119	-.253	-.328	-.238	.507	1.000	.303	-.343	.133
	습도	.065	.190	-.104	-.104	.108	-.125	.303	1.000	-.458	-.240
	풍속	-.104	-.210	-.141	-.196	-.292	.163	-.343	-.458	1.000	.210
	교통량	.021	-.081	.115	.264	.050	.150	.133	-.240	.210	1.000
유의 확률	PM ₁₀	.	.000	.000	.000	.000	.000	.000	.000	.000	.131
	PM _{2.5}	.000	.	.000	.000	.000	.000	.000	.000	.000	.000
	SO ₂	.000	.000	.	.000	.000	.000	.000	.000	.000	.000
	NO ₂	.000	.000	.000	.	.000	.000	.000	.000	.000	.000
	CO	.000	.000	.000	.000	.	.000	.000	.000	.000	.004
	O ₃	.000	.000	.000	.000	.000	.	.000	.000	.000	.000
	온도	.000	.000	.000	.000	.000	.000	.	.000	.000	.000
	습도	.000	.000	.000	.000	.000	.000	.000	.	.000	.000
	풍속	.000	.000	.000	.000	.000	.000	.000	.000	.	.000
	교통량	.131	.000	.000	.000	.004	.000	.000	.000	.000	.

미세먼지(PM₁₀, PM_{2.5}) 농도에 영향을 끼치는 인자를 알아보기 위해 다중회귀분석을 실시하였다. 회귀분석에 투입되는 독립변수들 가운데 설명력이 가장 높은 변수들로 회귀모델을 구성하는 방법으로 하였다. 첫 번째 단계에서는 종속변수 간 상관관계가 가장 높은 변수를 투입하고, 두 번째 단계에서는 종속변수 간 편상관계가 있는 변수들을 투입하였다. 다음으로 각 단계별로 설명력이 높은 변수에 대한 유의성 검증을 실시하며, 유의하지 않은 변수는 제거하였다. 교통량은 상관성이 너무 낮은 관계로 다중회귀분석 과정 중 영향력 평가 중 자동 제외되었다. 즉, 다중회귀분석에서 교통량과의 관계를 찾을 수는 없었다.

PM10 회귀식의 유의성은 0.000이고, 회귀모형의 설명량도 0.232의 설명력(R²)이 있다. 회귀식에서 독립변수인 CO에 해당하는 계수는 47.039, SO₂는 3413.403이며 이에 대한 영향력을 비교하면 CO가 0.440으로 영향력(표준화 계수 β)이 제일 높게 나타났으며 그 다음으로 SO₂가 0.252로 확인되었다. 다른 변수들의 계수와 영향력은 다음 표에 나타내었다.

$$\begin{aligned} \blacksquare \text{ PM}_{10} = & 31.427 + (47.039 \times \text{CO}) + (3413.403 \times \text{SO}_2) + (358.269 \times \text{O}_3) + (-97.125 \\ & \times \text{NO}_2) + (-0.323 \times \text{온도}) + (-3.456 \times \text{풍속}) + (-1.48 \times \text{습도}) \end{aligned}$$

〈표 4-6〉 회귀분석 결과(PM₁₀)

종속 변수	독립 변수	비표준화 계수(B)	표준오차 (SE)	표준화 계수(β)	t	유의 확률	Durbin-Watson	R ²
PM ₁₀	(상수)	31.427	2.205		14.254	0.000	0.209	0.232
	CO	47.039	1.813	0.440	25.951	0.000		
	SO ₂	3413.403	183.273	0.252	18.625	0.000		
	O ₃	358.269	29.907	0.171	11.979	0.000		
	NO ₂	-97.125	8.941	-0.208	-10.863	0.000		
	온도	-0.323	0.040	-0.115	-8.132	0.000		
	풍속	-3.455	0.418	-0.110	-8.267	0.000		
	습도	-1.48	0.022	-0.087	-6.776	0.000		

PM_{2.5}도 위와 같이 분석을 실시하였으며 회귀식의 유의성은 0.000 이고, 회귀모형의 설명량도 0.388의 설명력이 있다. 회귀식에서 독립변수인 CO에 해당하는 계수는 35.706, NO₂는 -86.796이며 이에 대한 영향력을 비교하면 CO가 0.760으로 영향력이 제일 높게 나타났으며 그 다음으로 NO₂가 -0.426으로 확인되었다. 다른 변수들의 계수와 영향력은 다음 표에 나타내었다.

$$\blacksquare \text{ PM}_{2.5} = 1.525 + (35.706 \times \text{CO}) + (-86.796 \times \text{NO}_2) + (1288.651 \times \text{SO}_2) + (0.104 \times \text{습도}) + (-0.189 \times \text{온도}) + (124.883 \times \text{O}_3) + (-0.915 \times \text{풍속})$$

〈표 4-7〉 회귀분석 결과(PM_{2.5})

종속 변수	독립 변수	비표준화 계수(B)	표준오차 (SE)	표준화 계수(β)	t	유의 확률	Durbin-Watson	R ²
PM _{2.5}	(상수)	1.525	1.552		0.983	0.326	0.385	0.388
	CO	35.706	1.359	0.760	26.282	0.000		
	NO ₂	-86.796	6.352	-0.426	-13.665	0.000		
	SO ₂	1288.651	142.468	0.187	9.045	0.000		
	습도	0.104	0.017	0.113	6.036	0.000		
	온도	-0.189	0.028	-0.146	-6.813	0.000		
	O ₃	124.883	21.464	0.117	5.818	0.000		
	풍속	-0.915	0.282	-0.063	-3.244	0.001		

PM₁₀과 PM_{2.5}의 표준화계수를 통한 영향력을 각각 살펴보면, PM₁₀은 CO와 SO₂가 가장 영향력이 높았고 NO₂의 영향력은 4번째였다. 그러나 PM_{2.5}는 CO 다음으로 NO₂가 영향력이 높았다. NO₂는 대기 중 수 시간 이내의 짧은 체류시간으로 인해 주로 국지오염원이 있을

때 농도가 높아진다. 그리고 SO₂는 대기 중 체류시간이 길며, 배출특성 상 우리나라보다는 중국과 같은 외부 유입 기여도가 높다(이태형 외, 2016). 따라서 수원시에서 측정되는 PM₁₀은 외부 유입영향이 상대적으로 PM_{2.5}보다 크다는 것을 유추할 수 있다. CO는 외부 유입되거나 국지 오염원이거나 대기오염물질이 배출될 때 같이 배출되며 체류시간이 매우 길기 때문에 국지/외부 오염원이 구분되지 않는다.

4) 랜덤포레스트 방법 미세먼지 예측 모델링

(1) 기초 통계 및 데이터 전처리(preprocessing)

본 연구에서 사용된 변수들의 기본적인 통계적 특성을 아래 표에 나타내었다. 총 raw data 수는 8,784개이지만 장비 문제 등으로 인한 결측값으로 인해 유효 데이터 수는 각각 다르게 나타났다. 모델링에서는 PM_{2.5}가 1월부터 7월까지 결측인 관계로 보다 많은 학습량을 확보하기 위하여 미세먼지 두 항목 중 데이터가 많은 PM₁₀만을 사용하였다.

<표 4-8> 변수들의 기본적인 통계적 특성

구분	N	평균	표준편차	최소값	최대값
PM10	8651	55.1	30	1	316
SO ₂	8478	0.006	0.002	0.000	0.051
CO	8599	0.58	0.27	0.20	2.60
NO ₂	8280	0.099	0.062	0.008	0.627
O ₃	8652	0.017	0.014	0.002	0.091
Temp	8769	14.2	11	-16	35
Humi	8769	57.5	17	0	93
car(Total)	7504	1872	878	0	5580
car(small)	7504	866	356	0	1808
car(medium)	7504	969	614	0	3900
car(large)	7504	36	36	0	519
windspeed	8768	1.5	1	0	7

모델의 정확도를 높이기 위해 올바른 데이터를 준비하는 것은 매우 중요하기 때문에 데이터 전처리 과정은 필수적이다. 데이터 전처리란 주제와 일관성이 맞지 않는 데이터 또는 잘못된 값이 포함된 데이터가 있거나 데이터가 누락되었을 때 이를 제거하거나 수정하여 최대한 데이터의 정확성을 높이는 것을 말한다. 뿐만 아니라 중요성이 현저히 낮은 데이터를 과도하게 많이 포함한 경우 이들을 적절히 축소 조절하여 사용에 용이한 형태로 변경시켜주기도 한다. 본 연구에서는 첫 번째로 누락된 데이터를 파악하고 처리하고자 하였다.

변수별로 데이터가 누락된 정도와 시점이 제각각이었다. 누락된 데이터를 예측하거나 평균 값으로 채우는 방법도 있으나 본 연구에서는 누락된 값들을 모두 제거하는 방식을 채택하였다. 변수마다 효과적인 전처리 방식이 다르기 때문에 여러 가지 전처리 방식을 분석하여 전처리를 수행하였다.

(2) 예측모형 모델링(Modeling) 및 검증

전처리 후 최종 사용할 변수는 33개이며 각각 7,077개가 존재한다. 데이터를 랜덤을 섞은 후 그 중 70%인 4,953개는 모델 구축을 위한 Training data로 사용하고 30%인 2,124개는 모델의 검증을 위한 Test data로 사용하였다.

```
In [13]: from sklearn.ensemble import RandomForestRegressor
seed = 37
model = RandomForestRegressor(random_state=seed)
model

Out[13]: RandomForestRegressor(bootstrap=True, criterion='mse', max_depth=None,
max_features='auto', max_leaf_nodes=None,
min_impurity_split=1e-07, min_samples_leaf=1,
min_samples_split=2, min_weight_fraction_leaf=0.0,
n_estimators=10, n_jobs=1, oob_score=False, random_state=37,
verbose=0, warm_start=False)

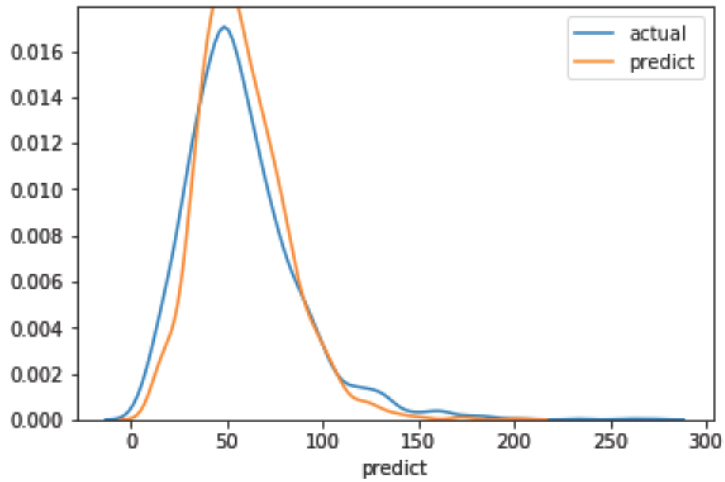
In [14]: model.fit(X_train, y_train)

/Users/Min/anaconda/envs/tensorflow/lib/python3.5/site-packages/ipykernel_launcher.py:1: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example using ravel().
"""Entry point for launching an IPython kernel.

Out[14]: RandomForestRegressor(bootstrap=True, criterion='mse', max_depth=None,
max_features='auto', max_leaf_nodes=None,
min_impurity_split=1e-07, min_samples_leaf=1,
min_samples_split=2, min_weight_fraction_leaf=0.0,
n_estimators=10, n_jobs=1, oob_score=False, random_state=37,
verbose=0, warm_start=False)
```

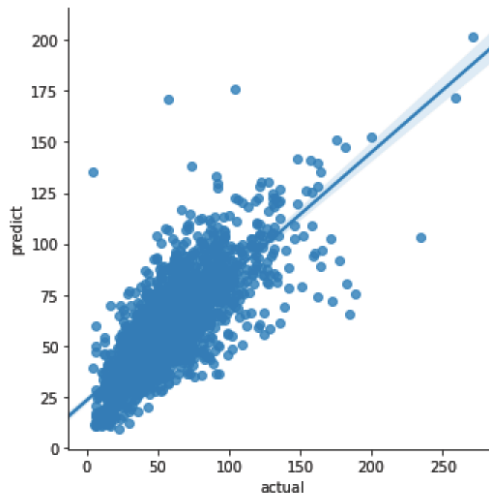
<그림 4-22> 랜덤 포레스트 함수를 이용한 모델 구축 과정

Training data로 구축된 모델의 R^2 는 0.9284로 나타났다. 이 모델의 신뢰성을 검증하기 위해 Test data를 모델에 대입하여 미세먼지 값을 예측하고 그 결과를 나타내었다. 실제 측정된 값과 모델로 예측된 값의 분포도는 아래 그림과 같다.



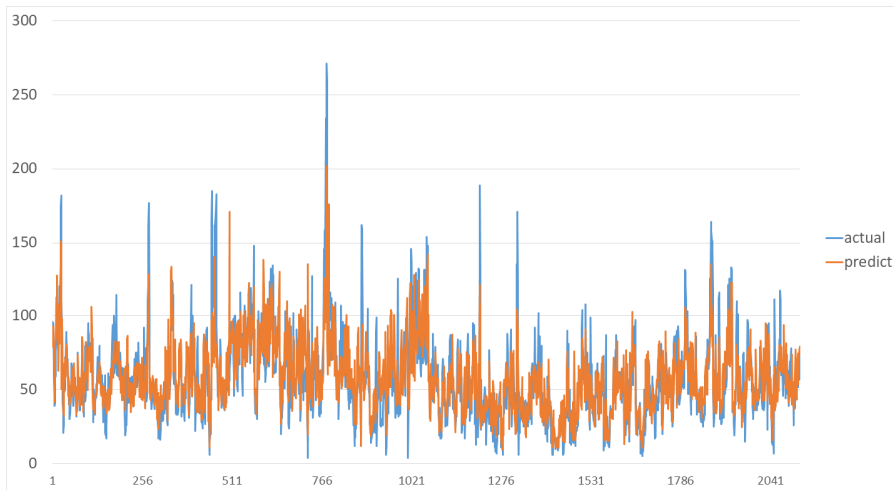
<그림 4-23> 예측값과 실측값의 분포도

산점도를 아래 그림에 나타내어 두 변수간의 상관관계를 나타내었다. R^2 는 0.7735로 모델의 R^2 보다 조금 낮게 나타났다.



<그림 4-24> 예측값과 실제값의 산점도

시간의 흐름에 따라 실측값과 예측 값을 그래프에 나타내었다. 높거나 낮게 튀는 몇 개의 값을 제외하고 대부분 잘 예측하는 것으로 나타났다. 즉, 랜덤 포레스트 기법 활용 예측 모델링을 통해 오차 범위를 인정하는 범위 내에서 독립변수들의 농도만을 가지고 미세먼지(PM_{10}) 농도를 예측하는 데 사용할 수 있을 것이다.



〈그림 4-25〉 예측값과 실제값의 비교

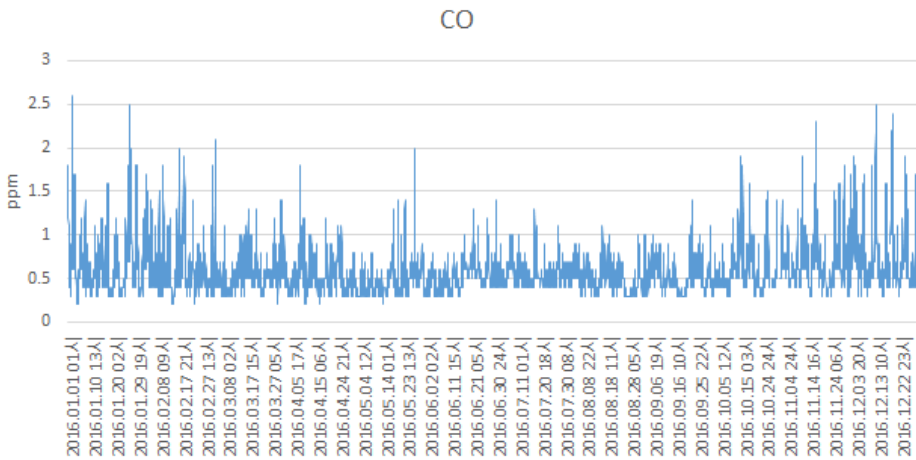
(3) 미세먼지의 영향요인

앞서 구축한 모델에서 사용한 종속변수(PM_{10})와 독립변수들의 관계를 파악해보면 CO가 관계도가 가장 높은 편이었다. 따라서 PM_{10} 농도를 예측함에 있어서 독립변수 중 어떤 변수가 가장 중요하며, 미세먼지 농도 증가에 영향을 줄 수 있는 요인을 저감하면 얼마나 미세먼지 농도가 저감될 수 있는 가를 알아보기 위한 분석을 수행하였다. 다음 표에서 weight의 값이 높을수록 관계도가 높은 것으로 CO 다음으로는 온·습도가 영향요인으로 나타났다. 인근 지역 교통량은 상대적인 가중치가 가장 낮게 나타났는데, 이는 다중회귀분석에서도 동일한 결과가 도출되었던 바, 인근지역 교통량이 직접적인 미세먼지 농도 증가에 영향을 주지는 않은 것으로 사료된다.

〈표 4-9〉 PM_{10} 과 변수들과의 관계

변수	weight
SO ₂	0.066897
NO ₂	0.056698
CO	0.221075
O ₃	0.054004
온도	0.104497
습도	0.108345
교통량	0.032595

하지만, CO는 자동차와 같은 도로오염원, 산업시설, 난방 등 미세먼지 배출원에서 직접 배출되는 오염물질이다. 따라서 미세먼지 농도 증가에 가장 큰 영향을 주는 CO 농도를 감소시키에 따라 미세먼지 예측 농도가 어떻게 달라지는 지를 구현해 보았다. 2016년 1년동안 1시간 간격의 CO 농도의 최소값은 0.2 ppm, 최대값은 2.6 ppm이고 평균은 0.58 ± 0.28 ppm으로 측정되었고, 다음과 같은 분포를 갖고 있다.



<그림 4-26> 1년간 CO 농도 변화 패턴

CO 감소율에 따라 미세먼지 예측농도 변화를 확인하고자 CO 농도를 ① 감소시키지 않았을 경우와 ② 10%, ③ 20%, ④ 50% 감소시키는 경우를 비교분석하였다. 모델링 결과는 아래표와 같다.

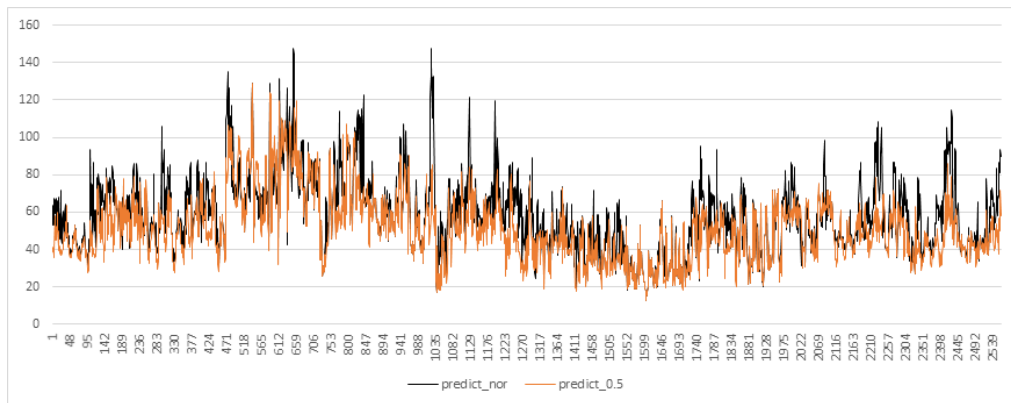
<표 4-10> 모델링 결과

		N	평균	표준편차	최소값	최대값
	actual	2561	59.5	35.6	3	316
1	predict_nor	2561	58.1	20.1	12.4	147.8
2	predict_0.9	2561	55.4	19.2	12.4	144.3
3	predict_0.8	2561	54.3	19.0	12.4	140.2
4	predict_0.5	2561	51.0	17.6	12.4	129.3

Test 데이터의 개수는 2,561개이며 이 기간 동안 PM₁₀ 실측값의 평균은 59.5 $\mu\text{g}/\text{m}^3$ 이며 CO 농도에 변화를 주지 않고 모델링했을 때 예측 농도는 58.1 $\mu\text{g}/\text{m}^3$ 으로 나타났다. CO를

감소 10%로 감소시켰을 때 PM_{10} 예측농도는 4.5%로 감소하는 것으로 나타났다. PM_{10} 예측농도는 CO 20% 감소 시 6.5%, 50%로 감소시켰을 때에는 12.2% 감소하는 것으로 나타났다. 미세먼지의 최대값 역시 CO 감소율에 따라 확연히 감소하는 추세를 보였다. 아래 그림은 CO 농도 변화 없이 PM_{10} 농도를 예측하였을 경우와 CO 농도를 50%로 감소시킨 후 예측하였을 때의 두 가지 경우를 나타낸 그래프이다.

즉, CO 농도를 50% 감소시킨 후 예측한 농도는 특히 PM_{10} 의 피크 농도가 낮아졌다.



<그림 4-27> 2가지 분석 경우 비교 그래프, 검은색: CO 변화 없는 예측값, 주황색: CO를 50% 줄인 후 예측값

제3절 결과 고찰

이 장에서는 PM_{10} , $PM_{2.5}$ 농도에 영향을 주는 다른 대기환경기준 오염물질, 기상요소, 인근 지점 도로 교통량과의 관계를 단순 상관분석, 다중선형회귀분석을 통해 알아보고, 이들 자료를 통해 미세먼지 농도를 예측하는 예측 모델을 구축하였다. 예측모델은 ANN 기법으로 랜덤포레스트 방식을 사용한 모델링으로 구축되었고, 예측값과 실측값은 $R^2=0.77$ 로 높거나 낮은 튀는 값을 제외하고는 잘 일치하였다. 즉, 각각의 대기오염물질 간 단순 상관관계는 낮았지만, ANN기법 예측 모델링으로 PM_{10} 외 대기환경기준 오염물질, 기상요소 값 등을 활용하여 PM_{10} 농도를 예측할 수 있었다.

그러나 국지오염원으로 미세먼지 농도에 크게 영향을 줄 것이라 예상했던 인근 지점 도로교통량은 분석결과 통계적으로 유의한 영향을 주지 못했다. 그 원인으로는 $PM_{2.5}$ 결측치로 인해 모델에 $PM_{2.5}$ 가 아닌 PM_{10} 을 사용할 수밖에 없었고, PM_{10} 의 경우 CO나 SO_2 와 같이 체류시간이 길고(적어도 1일 이상) 외부 유입 오염원이 기여를 많이 하는 오염물질과 상관성이 높았기 때문으로 생각된다. 또한 측정소 인근의 차량 통행량은 1년 동안 시간대별 패턴 및 교통량이 크게 변하지 않았으나, PM_{10} 은 외부 유입 영향, 도로 외 다른 배출원의 영향, 대기 확산 및 바람과 같은 기상 영향 등 매우 복잡한 원인에 의해 농도가 변하기 때문이기도 할 것이다. 또한 $PM_{2.5}$, PM_{10} 과 같은 미세먼지는 배출원에서 직접 배출되기 보다는 가스상 오염물질로부터 오랜 시간에 걸쳐 생성되고 변환되는 최종산물이기 때문에 도로의 자동차 배출가스가 대기 중에 배출된 후 1~2일의 시간을 거쳐 미세먼지로 변환되는 것과 같은 오염물질간의 시간차가 이 모델에 반영되지 못하는 것도 원인이 될 수 있다.

PM_{10} 예측에의 가중치가 가장 높았던 CO를 활용하여 CO를 10%, 20%, 50% 낮췄을 경우 PM_{10} 이 얼마나 감소될 것인가를 예측해 본 결과에서는, CO를 10% 낮출 경우 PM_{10} 은 4.5%가 낮아졌는데, 주로 피크농도 값이 낮아지는 결과를 얻었다. 즉, PM_{10} 농도를 저감하기 위해서 CO 농도를 지시자로 간주하고 CO를 저감하는 방식의 접근이 실질적으로 미세먼지 농도를 낮추는 데에도 적용될 수 있다는 시사점을 얻을 수 있었다.

제5장 결론

제1절 연구의 의의

본 연구에서는 수원시 국지배출원(예: 차량 통행량 등)과 대기오염물질이 미세먼지 농도 증가에의 얼마나 상관성이 있는지를 알아보기 위한 목적으로 대기오염물질 간의, 그리고 미세먼지 증가 영향요인간의 상관관계를 알아보는 연구를 수행하였다. 복잡한 대기오염물질의 생성, 이동, 소멸 등을 다 고려하는 기존의 메카니즘을 사용한 모델링이 아닌 통계학적 접근 방법과 인공신경망(ANN:Artificial Neural Network) 기법을 사용하여 공공데이터로 구축되어 있는 자료를 통한 미세먼지 농도 예측을 시도하고자 한 첫 연구이다. 다만, 제한된 연구기간 및 예산으로 인해 2016년 당해연도에 대한 수원시 내 기 측정되고 있는 대기오염물질 측정 자료와 교통량, 기상 자료를 제한적으로 수집하였고, 인공신경망 기법을 도시 미세먼지 농도 예측 모델에 적용하고자 하는 기초 연구(Preliminary study)적 성격으로 추진되었다. 기초 연구적 성격을 보충하기 위하여 ANN 모델링을 구현함과 함께 다중회귀분석도 같이 수행하여 연구결과로 제시하였다.

주요 연구 결과는 다음과 같다.

수원시는 도심지역으로 한반도 내 대기오염물질 배출로 인해 배경대기 지역보다 대기오염도가 높았다. 수원시 서쪽에 인접해 있는 안산시와의 농도 비교에 있어서는 SO₂, CO의 농도 차이가 두드러졌으며, 오히려 체류시간이 긴 PM_{2.5}, PM₁₀의 경우 농도 차이가 크지 않았다. 수원시 관내에서는 다른 도시대기 측정소보다 도로변 측정소에서 PM₁₀, NO₂의 농도가 높아서 도로오염원 영향이 큼을 간접적으로 알 수 있었다. 그러나 그 외 6개 도시대기 측정소의 경우 인근 건설공사가 최근 잦은 광교동 측정소를 제외하고는 전반적으로 대기오염도가 비슷한 추세와 농도 분포를 보였다.

수원시 도로변 측정소를 대상으로 대기오염물질 간 선형회귀분석 결과, PM_{2.5}와 PM₁₀, CO와 NO₂ 짝을 제외하고는 다른 물질 간 상관성에서는 유의성을 찾을 수 없었다. PM₁₀이나 PM_{2.5}의 배출원 중 40%이상이 도로이동오염원임에도 불구하고 NO₂, CO나 PM_{2.5}, PM₁₀와는 단순 선형회귀분석 결과에서는 상관성을 찾을 수 없었는데 그 이유로는 광화학적 반응, 공기계의 이동 및 기상 조건의 변동 등이 반영되지 않았기 때문일 것이다.

따라서 PM₁₀, PM_{2.5} 농도에 영향을 주는 다른 대기환경기준 오염물질, 기상요소, 인근 지점 도로 교통량과의 관계를 다중선형회귀분석을 통해 알아보고, 통계적으로 유의한 자료를 사용하여 인공신경망기법 중 랜덤포레스트 방식을 적용한 PM₁₀ 농도 예측 모델을 구축하

였다. 그 결과 예측값과 실측값이 $R^2=0.77$ 로 잘 예측하는 결과를 얻었다. 그러나 인근 지점 도로교통량은 분석결과 통계적으로 유의한 영향을 주지 못했다. 그 원인으로는 $PM_{2.5}$ 예측치로 인해 모델에 $PM_{2.5}$ 가 아닌 PM_{10} 을 사용할 수밖에 없었고, PM_{10} 의 경우 CO나 SO_2 와 같이 체류시간이 길고(적어도 1일 이상) 외부 유입 오염원이 기여를 많이 하는 오염물질과 상관성이 높았기 때문으로 생각된다. 또한 측정소 인근의 차량 통행량은 1년 동안 시간대별 패턴 및 교통량이 크게 변하지 않았으나, PM_{10} 은 외부 유입 영향, 도로 외 다른 배출원의 영향, 대기 확산 및 바람과 같은 기상 영향 등 매우 복잡한 원인에 의해 농도가 변하기 때문이기도 할 것이다. 또한 $PM_{2.5}$, PM_{10} 과 같은 미세먼지는 배출원에서 직접 배출되기 보다는 가스상 오염물질로부터 오랜 시간에 걸쳐 생성되고 변환되는 최종산물이기 때문에 도로의 자동차 배출가스가 대기 중에 배출된 후 1~2일의 시간을 거쳐 미세먼지로 변환되는 것과 같은 오염물질간의 시간차가 이 모델에 반영되지 못하는 것도 원인이 될 수 있을 것이다.

PM_{10} 예측에의 가중치가 가장 높았던 CO를 활용하여 CO를 10% 낮출 경우 PM_{10} 이 4.5% 저감된다는 결과를 얻을 수 있었다. 즉, PM_{10} 농도를 저감하기 위해서 CO 농도를 지시자로 간주하고 CO를 저감하는 방식의 접근이 실질적으로 미세먼지 농도를 낮추는 데에도 적용될 수 있을 가능성도 있다. 후속 연구를 통해 CO와 수원시 도로 자동차 통행량과의 관계, 나아가 인근 고속도로 통행량과의 관계를 찾아서 뒷받침하는 것이 필요할 것이다.

또한, 향후 모델의 예측 정확도를 향상하기 위해서는, 효과적인 데이터 전처리 법이 추가로 연구되어야 할 것이다. 또한, 랜덤 포레스트를 이용한 미세먼지 예측 모델은 과거 자료 학습을 통해 예측을 하기 때문에 자료의 양이 많을수록 예측 정확도가 높아지게 되므로 후속 연구에서는 보다 많은 데이터를 수집하여 모델에 적용해 보고자 한다.

후속연구로 대기 오염물질이 대기 중에 잘 혼합되는 특성을 반영하여 모델 입력 자료의 공간적 규모를 확대하여 수원시나 경기 남부권 도로교통량을 대표할 수 있는 교통량 자료와 수원시 또는 수도권 남부 대표 대기오염도 자료를 활용한 모델링도 필요할 것이다. 또는 수원시와 안산, 성남, 화성, 서울 등 인접 도시들, 백령도와 같은 배경대기 측정소의 대표 측정 자료를 사용한 모델링을 통해, 외부 유입 오염원으로 인한 수원시 미세먼지 농도 증가 기여도를 판단하는 연구도 필요할 것이다.

제2절 정책제언

수원시에서는 현재 7개소의 대기오염측정소를 운영하고 있으나 그 중 미세먼지, 특히 $PM_{2.5}$ 는 측정기기교정 등으로 인해 결측치가 다른 오염물질들에 비해 많다. 따라서 본 연구에서 구축된 모델에 $PM_{2.5}$ 자료를 더 추가하여 분석한다면, 측정소 운영 중 생긴 결측치에 대해 예측값으로 임시 대처가 가능할 수 있을 것으로 보인다. 미세먼지 (특히 $PM_{2.5}$)는 시민 건강에 직접적인 영향을 주는 관심물질로 미세먼지 오염도가 높을 경우 시민 행동요령을 보급하는 판단 자료로 사용하고 예측 농도를 시민에게 제공해 주는 것도 필요할 것이라 제안한다.

또한 앞 절에서 제시한 후속연구를 수행한다면, 수원시 내에서 자체 배출원을 저감함으로써 줄일 수 있는 대기오염물질의 농도를 예측하여 수원시 대기오염물질 배출 저감 목표 설정의 근거로 활용할 수도 있을 것이다.

나아가 기상요소와 대기오염물질간의 보다 명확한 상관관계를 파악하게 된다면 도심 내 대기환경 취약지역 개선을 위한 공간가이드라인을 작성하는 데 활용될 수 있을 것이다. 또한 바람장 예측을 통해 계획지구에서 풍속이 약화되거나 정체되는 지역 및 풍향의 변화가 크게 발생하는 지역을 파악하여 토지이용계획의 수립 시 저 풍속 지역으로 설정하고 도시미기후 관리계획을 수립하는 데에도 활용할 수 있을 것이다.

단, 이러한 모델링은 많은 양의 데이터를 필요로 하기 때문에 정확도를 높이기 위해서는 측정데이터의 관리가 매우 중요하다. 수원시에서 관리하고 있는 대기오염측정소의 자료에 대한 QA/QC, 측정기기 관리, 교통량 자료 DB 구축이 필요할 것이다.

| 참고 문헌 |

<국문 자료>

- 김득수, (2016). 서울 도심대기의 NO, NO₂와 O₃ 사이의 대기화학적 특성 연구, 한국대기환경학회 지 제32권 제4호
- 김순태, (2016). ‘국내 on-road에서의 미세먼지 발생 및 현상규명’, 한국대기환경학회 2017 공동학술심포지엄
- 국립환경과학원, (2012). 교통량 밀집지역의 자동차 대기오염 영향연구
- 권순박 외, (2017). 인공지능기반 수도권 지하역사 실내공기질 예측기술, 한국철도기술연구원 수도권 대기환경청, (2009). 녹색기술 리더 50선
- 수도권 대기환경청, (2015). 우리동네 대기질
- 이한솔, (2016). 한국의 겨울철 PM예측을 위한 간단한 통계적 방법의 개발, 서울대학교
- 안현진 외, (2015). 수원시 대기 중 오존과 질소산화물의 장기변동 특성 (1991~2012), 한국대기환경학회지 제31권 제4호
- 임영욱, (2017). ‘자동차의 미세먼지의 건강영향 및 관리방안’, 연세대학교 환경공해연구소. 한국대기환경학회 2017 공동학술심포지엄
- 전하은 외, (2015). 서울지역 PM_{2.5}의 배출원별 기여도 추정 연구, 한국도시환경학회지, 15(1), 7-16
- 조수석 외(2011). 기상조건 및 교통량에 따른 도로변 대기질 영향, 서울특별시보건환경연구원보 제47호
- 채희정, (2009). 풍속과 풍향이 미세먼지에 미치는 영향, 서울시
- 최기선 외, (2009). 인공신경망 기법을 이용한 태풍 강도 및 진로 예측, 한국지구과학회
- 환경부, (2016). 바로 알면 보인다. 미세먼지, 도대체 뭘까?

<영문 자료>

- Barai et al. (2007), Neural Network Models for Air Quality Prediction: A Comparative Study, Soft Computing in Industrial Applications, 290-305
- Cohan et al. (2007), An integrated framework for multipollutant air quality management and its application in Georgia, Environ. Manage. 40, 545-554
- Fast, J., B. Foy, F. Rosas, E. Caetano, G. Carmichael, L. Emmons et al. (2007) A meteorological overview of the MILAGRO field campaigns, Atmospheric Chemistry and Physics, 7, 2233-2257.

- Feng et al. (2015), Artificial neural networks forecasting of PM2.5 pollution using air mass trajectory based geographic model and wavelet transformation, *Atmos. Environ.*, 107, 118-128
- Gardner and Dorling (1998), Artificial neural networks (the multilayer perceptron) - a review of applications in the atmospheric sciences. *Atmos. Environ.* 32, 2627-2636
- Lee, J. H., Y. P. Kim, K. C. Moon, H. K. Kim, C. B. Lee (2001) Fine particle measurements at two background sites in Korea between 1996 and 1997, *Atmospheric Environment*, 35, 635-643
- McMurry, (2003), Particulate matter science for policy makers: a NARSTO assessment.
- Volkamer et al., 2007. Secondary organic aerosol formation from anthropogenic air pollution: Rapid and higher than expected. *GEOPHYSICAL RESEARCH LETTERS*, VOL. 33, L17811, doi:10.1029/2006GL026899

Abstract

Using ANN(Artificial Neural Network) to Predict Particulate Matter Concentration in Suwon

This study aims to understand how the local emissions(e.g. road traffic by cars) and standard air pollutants affect the Particulate Matter(PM₁₀ and/or PM_{2.5}) concentration increase in Suwon-si, Kyonggi-do in Republic of Korea. We used simple statistical approaches such as single and multiple regression analysis and the Artificial Neural Network(ANN) forecasting model rather than mechanism-based forecasting model such as CMAQ as an effort to use public open data for PM pollution prediction. We used air pollutants measurement data (PM₁₀, SO₂, NO₂, CO, O₃) in air pollution monitoring sites in Suwon, road traffic data on the Dongsuwon-ro Road and meteorological data monitored in Suwon meteorological monitoring station in 2016. This study was performed as a preliminary study by covering literature reviews about ANN model application to air quality prediction, basically analyzing air quality in Suwon and introducing ANN forecasting model to predict PM₁₀ concentration from other standard air pollutants concentration and road traffic volume and meteorological data. The random forest forecasting model results showed that predicted PM₁₀ was matched to measured PM₁₀ ($R^2=0.77$). The major factors to affect PM₁₀ increase were air pollutants with long life time such as CO and SO₂. In addition, the road traffic volume did not significantly affect the PM₁₀ prediction. We observed a similar correlation and trend from the multiple regression analysis using the same data as ANN model used. Possible reasons of low correlation between road traffic and PM could be the missing PM_{2.5} data in model, the appearance time gap between emission from road traffic and observation of PM due to the photochemical secondary formation of PM and the contribution of regionally transported PM to local PM concentrations. If we add PM_{2.5} analysis to this model for further

study, the predicted PM_{10} and $PM_{2.5}$ concentration could be used as temporal substitute information in the air quality monitoring station in case of data missing. In addition, we suggest to expand the spatial dimension from Suwon to southern Kyonggi-do to consider the air masses' mixing characteristics. The prediction of PM in Suwon from PM concentrations measured in neighbor cities such as Ansan, Seongnam, Hwaseong, Seoul, and Baengnyeong-do, the remote air quality monitoring station using ANN prediction model is also suggested to understand the contribution of the air pollution transported from out of Suwon to PM increase in Suwon as a future study.

keyword : particulate matter, Artificial Neural Network(ANN), Particulate Matter Concentration prediction, Air quality improvement

| 저자 약력 |

강은하

이학박사

수원시정연구원 안전환경연구실 연구위원(현)

E-mail : ehkang09@suwon.re.kr

주요 논문 및 보고서

「수원시 실내공기질 관리계획 수립 연구」(2016, 수원시정연구원)

「수원시 대기 중 오존과 질소산화물의 장기변동 특성(1991~2012)」(2015, 한국대기환경학회)

「수원시 (초)미세먼지(PM10, PM2.5) 특성평가」(2014, 수원시정연구원)

「수원시 대기오염 변화추이 및 오염특성 분석」(2013, 수원시정연구원)

