

논문

머신러닝 클러스터링을 이용한 컬럼 관측에 따른 대기 에어로솔 특성 분석

Characteristics of Atmospheric Aerosols Based on Column Measurements by Using Machine Learning Clustering

이권호*, 이규태

강릉원주대학교 대기환경과학과

Kwon-Ho Lee*, Kyu-Tae Lee

Department of Atmospheric & Environmental Sciences,
Gangneung-Wonju National University, Gangneung, Republic of Korea

접수일 2020년 8월 21일

수정일 2020년 9월 11일

채택일 2020년 9월 16일

Received 21 August 2020

Revised 11 September 2020

Accepted 16 September 2020

*Corresponding author

Tel : +82-(0)33-640-2319

E-mail : kwonho.lee@gmail.com

Abstract In this study, we classify atmospheric aerosols according to column measurement and meteorology patterns using machine learning technique. The observation data is the aerosol optical thickness (AOT), angstrom exponent (AE), precipitable water vapor (PWV), and wind at Daegwallyeong (DGL) and Gangneung-Wonju National University (GWNU) from March to June 2016. As a result of time series analysis and correlation regression analysis for individual data, higher mean AOT of 0.063 (32.0%) and AE ~0.031 in Gangneung compared to Daegwanryeong were found. These results show that the size of aerosols at two locations is similar, but the load of the particles is higher in Gangneung. The unsupervised K-means clustering are used as machine learning techniques in order to classify aerosol distribution pattern in the study area. As a result of the machine learning, aerosol patterns were classified into five groups according to local atmospheric conditions. These results of this study will provide useful information for understanding the relationship between factors related to the current status and pattern analysis of aerosols based on atmospheric observation data.

Key words: Aerosol, Machine learning, Aerosol optical thickness, Classification, Air quality

1. 서론

대기중의 에어로솔의 불확실성은 입자의 광학적 및 물리적 특성의 변화로부터 기인한다. 입자의 광학적 및 물리적 특성은 발생원(자연적 또는 인위적) 및 장거리 수송 과정에서 입자의 응집(coagulation) 또는 응축(condensation)과 같은 동적 과정에 의하여 영향을 받게 된다. 그리고 대기 중의 에어로솔이 수분을 흡수하거나(흡습성장), 또는 수분의 손실로 인한 입자 건조화 조차도 입자의 빛 산란 특성을 변화시키기 때문에 주의 깊게 분석해야 한다. 에어로솔 입자의 이러한 특성은 대기 복사 수지 과정에서 중요

한 역할을 하는 것으로 알려져 있으며, 복사 수지 과정에 미치는 영향에 대한 불확실도가 점차 개선되고 있는 것으로 보고되고 있다(IPCC, 2013). 지상에서 분광복사계(spectro-radiometer)를 이용한 관측 네트워크가 전 세계에 널리 퍼져 있지만(예: AERONET, SKYNET 등), 시공간 차원을 모두 관측할 수 있는 수단이 필요하다. 이로 인하여, 입자상 물질의 시공간적인 불확실성을 정량화하기 위한 노력에 어려움이 발생한다. 대기 에어로솔의 특성을 이해하기 위해서는 특정 지점에서의 관측(point measurement) 뿐만 아니라 대기 칼럼 관측(column measurement), 그리고 실시간 관측(in-situ measurement) 등이 복합적으로 이루어

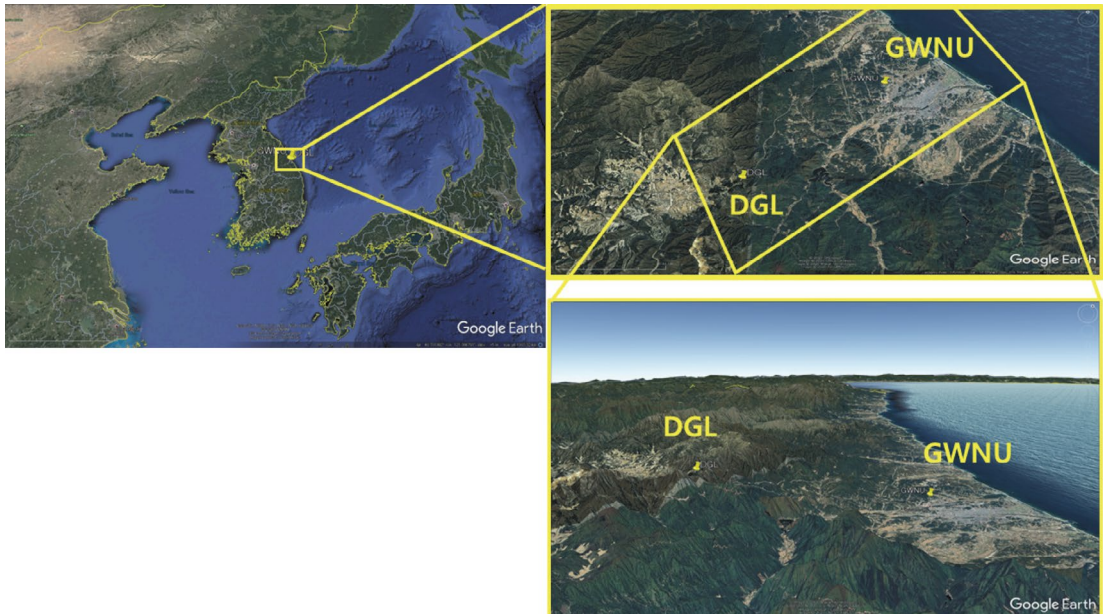


Fig. 1. Region of interest in this study. Yellow pins represent selected ground observation sites at Daegwallyeong (DGL) (longitude = 128.759°E, latitude = 37.687°N, altitude = 837 m a.s.l) and Gangneung-Wonju National University (GWNu) (longitude = 128.867°E, latitude = 37.771°N, altitude = 60 m a.s.l) in Korea.

어려야 하며, 부가적으로 단일 입자에 대한 미세 물리 및 광학 특성과 기상학적 환경 조건까지도 고려되어야 한다. 특히, 지상 관측 네트워크 정보는 에어로솔의 시공간 분포 정보를 반영하는 중요 수단이므로, 현재 대기 중의 에어로솔 입자에 대한 현황 및 변화 추세를 파악할 수 있게 해준다.

최근까지 PM_{10} 농도는 점차 줄어들고 있는 추세이지만, 고농도 사례 일수는 크게 줄어들지 않은 것으로 나타났다 (Lee and Park, 2020; NIER, 2019; Yeo and Kim, 2019). 고농도 오염 사례는 중국에서 발생한 오염물질의 장거리 이동 및 기상학적 조건에 따른 연무 현상에 의하여 흔히 발생하고 있으며 (Shin and Lee, 2016; Lee, 2012), 대도시의 대기오염이 주변 지역에서도 영향을 미치는 결과가 발생하기도 한다 (Park *et al.*, 2013). 중국의 주요 대도시와 인구밀집 지역에서 장거리 이동된 대기 오염물질은 기상조건에 따라 국내 대기질에 영향을 미치게 되며 (Lim *et al.*, 2013), 한반도의 수도권에서 발생하는 대기오염물질

과 혼합되는 경우에는 대기오염 상승효과가 발생하기도 한다. 한반도의 동부 산악지형은 중국이나 수도권 기원의 오염물질이 이동하는 주요 경로이므로 장거리 이동이 지역 대기질에 직접적인 영향을 미칠 수 있다. 특히, 산악 지역이나 해안 지역의 관측 한계 및 동적인 환경 특성으로 인하여 지역 규모의 에어로솔 특성에 관한 연구는 매우 제한적으로 이루어졌다 (Lee *et al.*, 2018; Park *et al.*, 2016).

최근의 대기오염물질 특성에 관한 연구에서는 오염물질의 현황과 예측에 있어서 인공 지능 기법을 활용하는 시도들이 이루어지고 있다. 기존의 관측이나 모델링 체계에서는 현실세계의 복잡한 비선형적인 현상을 이해하는 데 한계가 있으므로, 빅데이터와 컴퓨터 자원을 활용한 방법이 보다 효율적으로 사용가능하다는 것이 증명되고 있다. 예를 들면, 머신러닝 (Machine Learning 또는 기계학습) 알고리즘 중 하나인 신경망 분석기법을 이용한 대기오염물질 농도를 예측한 연구 (Cho *et al.*, 2019b; Jeon and Son, 2018,

Table 1. Specification of measurement instruments used in this study.

Data	Instrument	Measured parameters	Period
Ground station	Sun-sky radiometer (model: CE-318)	Aerosol Optical Thickness, Angstrom Exponent, Precipitable Water Vapor	March~June, 2016
Meteorology	Automatic Weather Stations(AWS)	Wind speed, wind direction	March~June, 2016

Cha and Kim, 2018), 기상 인자로 데이터 마이닝 툴을 이용한 미세먼지 예측 정확도 비교 연구(Oh *et al.*, 2016), 미세먼지 예측을 위한 알고리즘의 적합성을 평가에 관한 연구(Cho *et al.*, 2019a) 등이 보고되고 있다. 그러나 이러한 머신러닝 기반의 대기질 관련 연구는 오염물질의 추세 분석이나 농도 예측분야에 집중되어 있어, 보다 다양한 활용분야에서 사용되고 있지 못하는 실정이다.

따라서, 본 연구는 복잡지역인 대관령과 강릉의 대기 에어로솔의 변화 양상과 기상학적 조건에 관한 요인적 분석 및 주요 현황 특성에 대하여 분석하였다. 또한 연구대상 지역 내 대기 에어로솔 분포 현황에 대한 정보 제공을 위하여 머신러닝 기법 기반의 에어로솔 현황에 관한 유형화를 시도하였다. 본 연구 결과는 대기 관측자료 기반의 에어로솔에 대한 현황과 유형화와 관련된 각 요인과의 관계를 이해하는 데 유용한 정보를 제공할 것이다.

2. 자료 및 방법

2.1 관측 자료

연구대상 지역은 대한민국 영동지역에 위치한 강원도 강릉과 대관령의 관측 지점으로서, 관측 지점 주변은 서쪽의 태백산맥과 동쪽의 동해를 포함하고 있어 다양한 지형 효과의 영향을 받는 지역이다(그림 1 참조). 대관령 관측 지점은 강원도 대관령에 위치한 기상청 구름물리선도센터 구내(동경 128.759°, 북위 37.687°, 해발고도 837 m)이고, 강릉 관측 지점은 국립 강릉원주대학교 교내(동경 128.867°, 북위 37.771°, 해발고도 = 60 m)에서 Cimel사의 Sun-sky

radiometer (모델명 CE-318)를 이용하여 연속 대기 칼럼을 관측하였다. 강릉과 대관령 관측 지점은 직선거리로 약 12 km 떨어져 있으며, 강릉은 대관령에 비하여 약 45도 북동쪽에 위치하고 있다. 두 관측 지점은 근거리에 위치하고 있기 때문에 에어로솔의 분포 변화가 크지 않을 것이나, 대관령이 강릉에 비하여 고지대에 위치하고 있으므로 두 지역의 지형적 특징이 에어로솔 분포 특성에 영향을 미칠 것으로 예상된다.

컬럼 대기 관측 기간은 2016년 3월부터 6월까지 약 4개월 동안 연속 관측하였고, 대기질의 시공간적 분포 및 변화 양상을 정성적, 정량적으로 분석하기 위하여 자동 기상 관측 자료가 보조자료로서 사용되었으며, 각 자료 목록에 대한 상세 설명은 표 1과 같다.

본 연구에서 사용된 복사관측장비인 CE-318은 7개의 중심 파장 대역(340 nm, 380 nm, 440 nm, 500 nm, 675 nm, 870 nm, 939 nm 및 1020 nm)에서 직사광 및 산란광을 측정하는 다중 파장 태양 광도계이다. CE-318은 광도계 센서(파장 분해능은 약 10 nm FWHM, 약 1.2°의 관측 시야각)와 센서의 동적 관측을 위한 로봇, 그리고 자료의 송수신을 위한 통신장치로 구성된다. CE-318의 광도계는 Airmass 값이 7 미만인 낮 동안 직사광 또는 산란광을 여러 번 측정하도록 프로그래밍되어 있다. 직사광 측정은 7개 파장 대역에서 이루어지며, 레일리 산란(Rayleigh Scattering), 오존 및 기타 미량 기체의 흡수에 의한 투과도를 보정한 후 Beer-Bougguer 법칙을 사용하여 에어로솔 광학 두께(Aerosol Optical Thickness)를 산출하게 된다. 그리고 939 nm 채널은 대기 컬럼 내의 수증기에 민감한 파장대 이므로, 수증기를 산출하는 데 사용된다. 또한 이 장비는 태양의 주평면(태양 방위

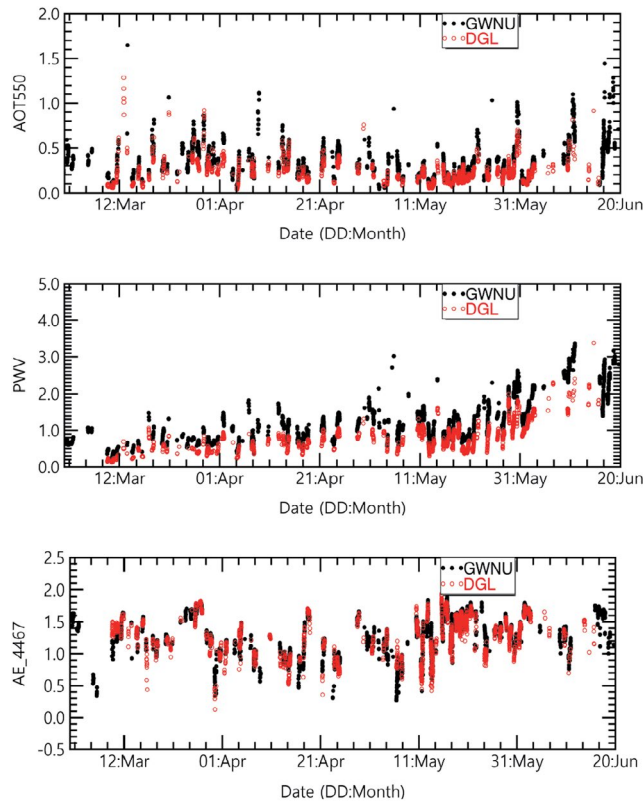


Fig. 2. Time series of aerosol optical thickness at 550 nm (upper), angstrom exponent (middle), and precipitable water vapor (lower) observed at Daegwallyeong (DGL) (red circle) and Gangneung-Wonju National University (GWNU) (black circle) for March~June 2016.

각과 동일한 고정 방위각으로 산란각을 변경하여 관측)을 따라 4개 파장 대역(440 nm, 675 nm, 870 nm 및 1020 nm)의 산란광을 측정함으로써, 에어로솔 크기 분포(입자 크기 범위 0.1~15 μm), 산란 위상 함수 및 굴절률을 역산하여 산출한다(Dubovik and King, 2000).

2.2 머신러닝 클러스터링

연구대상 지역의 에어로솔 분포 특성을 유형화하기 위하여 머신러닝 클러스터링 방법인 비지도 학습(unsupervised learning) 방법 중 하나인 K-mean 방법(MacQueen, 1967)을 사용하였다. K-mean 방법은 지도 학습(supervised learning) 방법에 비해 대량의 자료에서 군집을 발견하기 쉽고 계산 속도가 빠른 것으

로 알려져 있다. K-mean 방법에서는 먼저 입력자료를 무작위로 초기 중심 값을 선택한 후, k개의 중심 값과 각 개별 자료와의 유클리디안(euclidean) 최단 거리(d)를 식(1)과 같이 계산한다.

$$d = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \quad (1)$$

여기서, p_i 는 i번째 입력자료, q_i 는 i번째 자료에 관하여 설정된 중심 값, d는 유클리디안 최단거리 값이다. 그리고 각 클러스터마다 새로운 중심 값을 계산하는 과정을 반복하여 실행한 후, 선택된 중심 값의 변화가 매우 적은 경우에 실행을 멈추게 된다. K-mean 방법은 초기에 중심 값의 개수를 사용자가 지정해야 하므로, 본 연구에서는 군집의 수를 2개 부터 15개까지 순차적으로 늘려가면서 분석 결과를 비

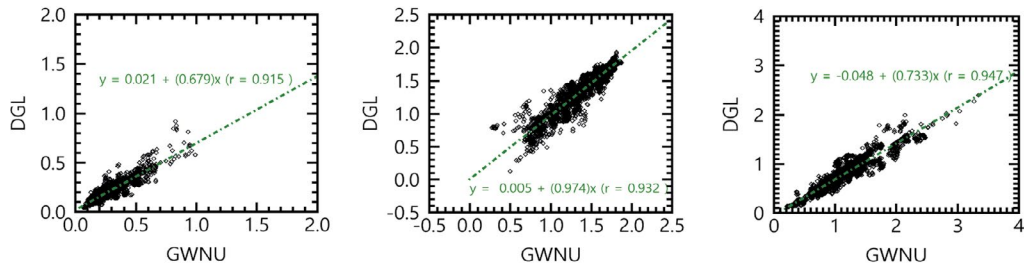


Fig. 3. Scatter plots of aerosol optical thickness at 550 nm (left), angstrom exponent (middle), and precipitable water vapor (right) observed at Daegwallyeong (DGL) (vertical axis) and Gangneung-Wonju National University (GWNu) (horizontal axis) for March~June 2016. Linear regression line and first order linear equation is represented as green color.

교하여 최적의 군집 개수를 정하는 방법을 사용하였다 (Charrad *et al.*, 2014). 머신러닝 클러스터링 기법을 이용하여 결정된 클러스터와 각 클러스터에 해당하는 관측값들의 특성 분석을 통하여 관측지점에서의 에어로솔 특성을 유형화하였다.

3. 결과 및 토의

3.1 대기 칼럼 관측

대기 에어로솔은 대부분 지표 근처에 분포하지만, 지형 및 기상학적 요인에 의하여 고도별 분포 및 이동이 가능하다. 모래폭풍, 대규모 산불과 같은 요인으로 인하여 장거리 이동 특성을 가지는 에어로솔은 지표보다 고도가 높은 지점에서 입자가 많은 반면, 지표에서 배출되는 비산 먼지나 연소성 입자는 지표 근처에서 입자 농도가 증가하게 된다. 이러한 에어로솔 입자의 공간적인 불균일성으로 인하여 지표에서의 미세먼지 농도가 대기 중의 부하량을 모두 반영하지 못하는 주요 원인이 되는 것이다. 그러나, 칼럼 관측은 칼럼 내에 존재하는 입자상 물질의 상대적인 총량에 관한 정보를 제공함으로써, 대기 중의 총 부하량 및 속성 변화를 파악하는 데 도움이 된다. 따라서 칼럼 관측값 중에서 에어로솔 광학두께값과 옴스트롬 지수값(α)의 변화를 비교하였다. 일반적으로 널리 사용되고 있는 550 nm 파장에서의 값(τ_{550})을 사용하기 위하여 CE-318의 7개 파장 중에서 440 nm 에어로솔

광학두께(τ_{440})와 440 nm 및 670 nm에서 산출된 옴스트롬 지수값($\alpha_{440-670}$)을 이용하여 환산하였다(식 2, 3 참조).

$$\alpha_{440-670} = -\frac{\ln(\tau_{670}/\tau_{440})}{\ln(670/440)} \quad (2)$$

$$\tau_{550} = \tau_{440} \times \left(\frac{550}{440}\right)^{-\alpha_{440-670}} \quad (3)$$

그림 2는 대관령과 강릉에서 관측된 에어로솔 광학두께값의 시계열 변화 그래프로써, 관측 기간 평균값은 대관령 (0.197 ± 0.099), 강릉 (0.260 ± 0.134)으로서, 도시 지역인 강릉의 에어로솔 광학두께값이 대관령에 비하여 약 0.063 (32.0%) 높은 수준으로 나타났다. 옴스트롬 지수값은 대관령 (1.362 ± 0.307), 강릉 (1.393 ± 0.293)으로서, 두 지역이 비슷한 범위에 해당하는 결과를 보였으나, 강릉이 ~ 0.031 정도 높은 값을 나타냄으로 인하여 대관령에 비해 크기가 작은 입자가 관측되었음을 알 수 있다. 수증기 값의 경우, 고도가 높은 대관령 (0.792 ± 0.356)이 강릉 (1.145 ± 0.459)에 비하여 낮은 값을 보였다. 이것은 대기중의 수증기가 주로 대기 하층에 존재하는 이유도 있지만, 강릉은 인접한 해안과 호수 등의 영향을 받고 있기 때문이다. 이상의 결과를 종합하면, 강릉에서 대기 에어로솔 입자는 수증기의 영향을 받게 되는 조건이지만, 옴스트롬 지수값에서 차이가 거의 없는 것은 건조 상태의 입자는 강릉의 에어로솔이 보다 적은 크기에서 존재할 수 있음을 증명한다. 즉, 두 관측 지점에

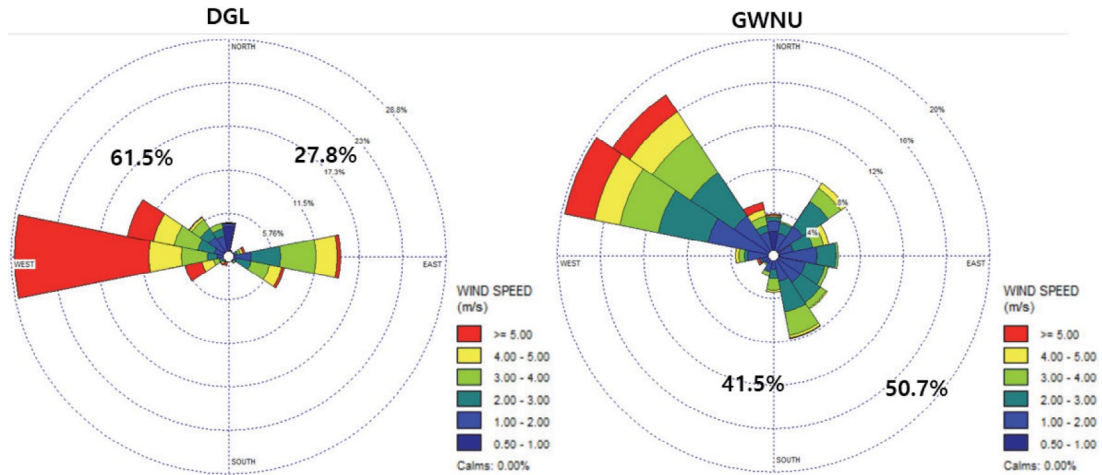


Fig. 4. Wind rose histogram for all wind data obtained during March~June 2016 with the hourly mean wind data observed at DGL station and Gangneung meteorology observatory. Each bin has a width of 22.5° (16 azimuth angles), as well as the corresponding proportion (%) of wind speed data.

서의 대기중 입자상 물질의 크기는 유사하지만, 입자의 총 부하량에 차이가 있는 것을 의미한다. 특히, 인위적인 발생으로 인한 미세 입자의 크기가 흡습 성장하게 되면 가시거리 감소와 함께 인체에 미치는 영향도 중요하게 되므로 향후 추가적인 연구를 통하여 영향을 규명할 필요가 있다.

그림 3은 대관령과 강릉에서 관측된 자료의 관련성을 알아보기 위하여 세 가지 요소(에어로솔 광학두께, 옴스트롬 지수, 수증기)별 산점도와 1차 선형회귀분석 결과이다. 먼저 전체 기간 동안의 에어로솔 광학두께 회귀 분석 결과를 보면, 대관령과 강릉의 선형회귀식은 $y = 0.021 + 0.679x$ ($r = 0.915$)로서 강릉과 대관령의 에어로솔 광학두께값은 상관성이 높으며 강릉이 대관령보다 높은 값을 보였다. 옴스트롬 지수는 $y = 0.005 + 0.974x$ ($r = 0.932$)로서 선형회귀식의 기울기가 거의 1에 근접함을 나타내었다. 수증기는 $y = -0.048 + 0.733x$ ($r = 0.947$)으로 강릉이 더 높은 결과를 보였다. 3가지 요소에 대한 상관회귀 분석 결과는 대관령과 강릉에서 관측 결과값이 강한 상관관계가 있음을 증명하였으며, 강릉에서 관측된 에어로솔 광학두께와 수증기 값이 높은 경향을 나타내고 있다. 이러한 결과는 이전의 시계열 분석 결과에서

유추한 것과 일치한다.

3.2 기상 조건

대기 에어로솔 입자의 농도에 영향을 미치는 요소는 직접적인 배출량과 세정작용이 있지만, 간접적으로는 대기에 의한 확산 요인도 작용을 하게 된다. 일반적으로, 바람이 많이 불고 대기에 의한 확산 및 이류가 많아지게 되며, 그 반대의 경우에는 입자의 농축이 발생할 가능성이 높다. 그러나, 본 연구대상 지역은 산지와 해안 지역이라는 지형적 요인에 의하여 바람의 강약과 이동 방향에 따라 에어로솔 입자의 수송과 축적, 또는 확산이 매우 불확실하게 발생할 가능성이 높다. 따라서 이전의 컬럼 관측자료의 해석을 위하여 기상관측 자료 중 바람 관측자료를 수집하여 분석하였다.

그림 4는 대관령과 강릉에서 관측된 풍속과 풍향 관측 자료를 이용한 바람장미도이다.

대관령의 바람 중 5 m/sec 이상의 풍속을 가진 바람은 주로 서풍 계열의 바람으로 나타났다. 또한, 바람 발생 빈도의 약 61.5%가 서풍 계열이며, 27.8%가 동풍 계열의 바람으로 확인되었다. 즉, 대관령에서 관측되는 에어로솔은 지역 기원이거나, 영서 지방에서

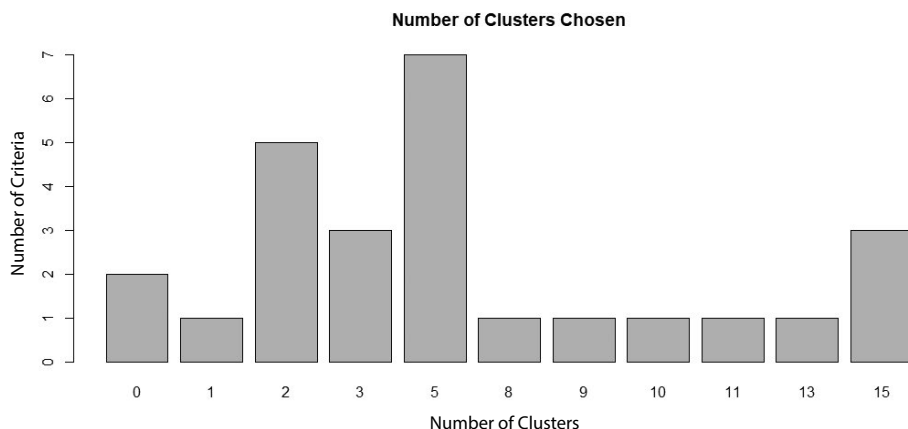


Fig. 5. The best clustering scheme from the different results obtained by varying all combinations of number of clusters, distance measures, and clustering methods using input data (aerosol optical thickness, angstrom exponent, precipitable water vapor, wind speed, wind direction). It is note to worthy that the largest number is best clustering number.

발생한 에어로솔 입자가 서풍 계열의 바람과 함께 이동한 경우가 과반수 이상일 것으로 판단된다. 강릉에서의 바람은 대관령과 마찬가지로 5 m/sec 이상의 풍속을 가진 바람이 서풍 계열이 대부분이며, 전체 바람의 약 41.5%가 서풍 계열, 50.7%가 동풍 계열의 바람이 불었다. 강릉은 대관령 산지에서 불어오는 바람의 풍속이 강하게 나타나고, 해양에서 불어오는 바람은 특정 방향에 상관없이 북동쪽에서 남동쪽에 걸친 모든 방향에서 불어오는 것을 알수 있다. 이러한 바람의 분포는 종관기상 관측이나 지역규모의 기상 조건에 따라 영향을 받을 수 있다. 특히, 산지에서 불어오는 바람이 강하게 발생하면 도시지역에서 발생한 에어로솔 입자의 환기에 영향을 미칠 것이지만, 반대로 해양에서 불어오는 바람은 느린 속도로 오염물질을 이동시켜서 산악 지형에 의하여 차폐될 수 있으므로 오히려 농축이 될 수도 있을 것이다. 두 지역의 바람 특성의 차이는 이전의 3.1장에서 논의되었던 관측 지점별 에어로솔 광학두께 및 옴스트롬 지수의 차이와 일치한다.

그림 4의 결과와 같이 관측 지점에 부는 바람의 특성은 장거리 이동성 에어로솔뿐 아니라 지역 내에서 배출된 대기오염물의 확산과 이동에 영향을 미치는 주요 인자로 간주될 수 있다. 그러나, 관측 지점을 포

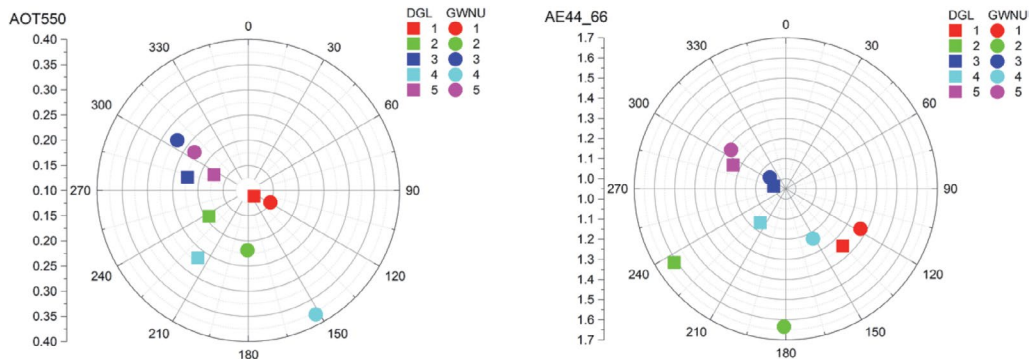
함하는 지역 내 배출원에 대한 정확한 위치 및 배출량 정보가 부족한 상황에서 외부에서 유입과 지역 배출에 대한 기여도 산정은 불가능하며, 본 연구의 목적에서 벗어나게 된다. 또한, 강릉 및 동해상에서 발생한 에어로솔이 정체성 기류로 인하여 누적되는 경우에도 대기오염 현상에 관한 해석에 어려움이 발생하므로, 대기질 및 기상 환경에 관한 종합 관측을 통하여 미세규모의 대기오염 현상에 관한 추가 분석이 필요할 것이다.

3.3 시나리오 분류를 위한 머신러닝 결과

일반적으로 K-mean 방법의 입력 변수의 단위가 서로 다르거나, 동일한 단위를 가지는 변수라도 상당히 다른 분산을 보이는 경우에는 변수값을 표준화하는 과정이 필요하다. K-mean 클러스터링은 공간의 모든 방향에서 등방성이기 때문에, 변수의 분산이 동일하지 않은 경우에는 분산이 작은 변수에 더 많은 가중치를 두는 것과 같으므로 클러스터는 분산이 큰 변수와 함께 분리되는 경향이 나타나게 된다. 따라서, 본 연구에서는 K-mean 클러스터링의 입력자료인 에어로솔 광학두께, 옴스트롬 지수, 수증기, 풍향, 풍속에 대하여 표준화(Normalization)를 적용하여 각 관측 자료의 단위에 따른 분산 문제를 해결하였다. 자료의

Table 2. Mean values of input data classified by K-mean clustering used in this study.

Cluster	1	2	3	4	5
DGL_aot550	0.117	0.193	0.223	0.267	0.175
DGL_ae4466	1.350	1.612	1.011	1.160	1.236
DGL_pwv	0.659	1.315	0.794	0.842	0.454
DGL_wd	135.202	236.478	282.330	216.678	294.396
DGL_ws	3.471	2.840	6.871	4.197	2.110
GWNU_aot550	0.150	0.219	0.272	0.380	0.231
GWNU_ae4466	1.369	1.635	1.047	1.232	1.282
GWNU_pwv	0.978	1.753	1.162	1.311	0.698
GWNU_wd	118.333	180.764	305.398	151.453	305.507
GWNU_ws	2.403	1.981	4.680	2.299	2.728
Number of samples	198	301	176	289	207

**Fig. 6.** Aerosol optical thickness (left) and angstrom exponent (right) as a function of wind direction for 5 clusters.

표준화는 식 (4)와 같은 Z-Score 방식으로 환산된 표준값을 사용하였다.

$$Z = \frac{X - \mu}{\sigma} \quad (4)$$

K-mean 클러스터링의 실행에 앞서 설정된 클러스터의 개수는 Charrad *et al.* (2014)에서 사용된 30가지의 인덱싱을 이용하여 계산된 최적 클러스터 개수 중 가장 많은 스코어 넘버로 제시된 결과값을 클러스터의 개수로 선정하였다. 그림 5에서는 위에서 표준화된 자료(에어로솔 광학두께, 옴스트롬 지수, 수증기, 풍향, 풍속)를 이용하여 계산된 결과로서, 최적 클러스터의 개수는 7개의 스코어 넘버를 획득한 5로 결정되었다.

이상의 혼합으로 생성된 12차원의 데이터를 입력으로 K-mean 클러스터링을 이용하여 비지도 학습을 수행하였다. 표 2는 K-mean 클러스터링 수행결과로 분류된 모든 관측 사례에 대하여 5개의 각 클러스터별 평균값을 나열하였다. K-mean 클러스터링 분석 결과, 5개의 클러스터별 평균값에 대한 특징이 나타났다. 1번 클러스터의 풍향은 대관령과 강릉의 풍향이 각각 135.202°와 118.333°로 강릉에서 대관령 방향으로 매우 건조한 남동풍이 불어오는 경우로서 약한 에어로솔 플룸이 관측된 경우로 판단된다. 2번 클러스터의 풍향은 대관령과 강릉의 풍향이 각각 236.478°와 180.764°로 대관령에서는 서풍계열이 불고 강릉에서는 남풍이 불어오는 경우로서, 대관령과

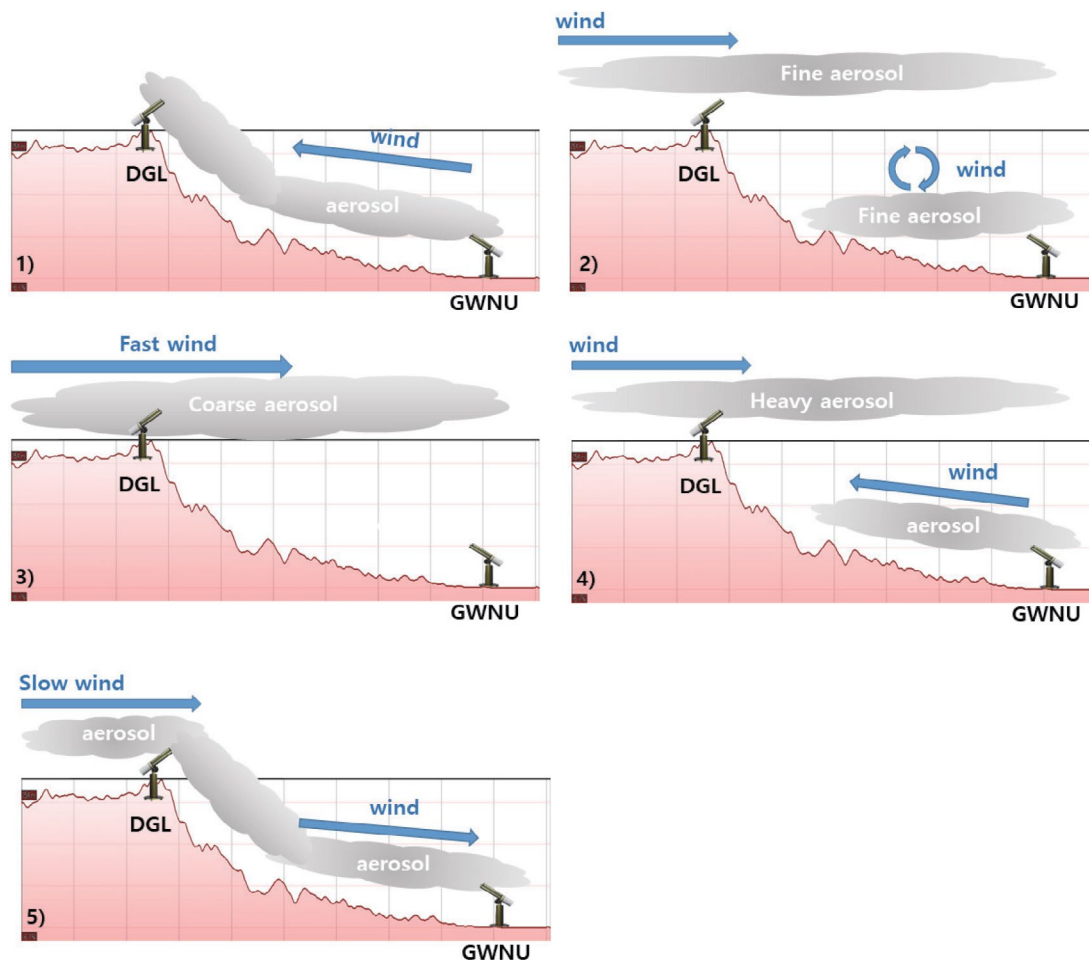


Fig. 7. Aerosol distribution cases in the study area based on 5 clusters.

강릉의 에어로솔의 기원이 다른 경우로 판단된다. 3번 클러스터는 대관령과 강릉이 서풍 계열의 바람이 불고 상대적으로 조대입자가 빠른 풍속과 함께 관측되고 있는 경우이다. 4번 클러스터는 대관령과 강릉의 풍향이 각각 216.678° 와 151.453° 로 대관령에서는 남서풍이 불고 강릉에서는 남동풍이 불면서 고농도의 에어로솔이 관측된 경우이다. 특히, 강릉에서는 하층의 오염과 상층에서는 대관령에서 넘어오는 오염물질이 영향을 미칠 수 있는 경우로, 다른 클러스터의 경우보다 높은 에어로솔 광학두께 값을 나타낸다. 마지막으로 5번 클러스터의 풍향은 대관령과 강

릉에서 모두 서풍 계열이 불지만 대관령과 강릉의 풍향이 각각 294.396° 와 305.507° 로 대관령에서 강릉 방향으로 서풍 계열의 바람이 불어오는 경우로서 약한 에어로솔 플룸이 관측된 경우로 분류된다.

그림 6은 5개 클러스터 결과 중 대관령과 강릉에서 관측된 에어로솔 광학두께와 움스트롬 지수를 풍향에 따른 극좌표로 나타내었다. 각 클러스터는 모두 풍향이 90° 와 330° 사이에 분포하고 있어 북동풍에 의한 영향은 희박한 것을 알 수 있다. 표 2의 결과 분석에서 언급하였지만, 강릉과 대관령의 풍향에 따른 에어로솔 광학두께와 움스트롬 지수의 분포가 나

타나는 특성이 분류되고 있다. 특히, 에어로솔 광학두께값이 낮을수록 두 지점의 풍향은 같아지게 되는 경향이 보인다. 그리고 옴스트롬 지수값도 적어질수록 두 지점의 풍향은 같아지게 되는 경향이 보인다.

이상의 결과를 토대로 연구대상 지역에서의 대기 에어로솔의 분포 특성을 그림 7과 같이 5개의 유형으로 분류하여 묘사하였다. 1번 사례는 강릉지역의 에어로솔이 대관령쪽으로 이동하는 경우로서 주로 지역 내 배경 대기 수준의 에어로솔에 관한 사례이다. 2번 사례는 미세 입자가 서풍과 함께 대관령을 넘어 강릉 쪽으로 이동하고 있고, 강릉에서는 느린 남풍과 함께 지역 내 에어로솔이 정체되어 다소 높은 수준의 에어로솔이 존재하는 사례이다. 3번 사례는 강한 서풍과 함께 황사와 같은 초대입자가 대관령을 넘어 강릉에도 동시에 영향을 미치는 사례이다. 4번의 사례는 대관령에 마찬가지로 고농도 에어로솔이 빠르게 이동하고 있지만, 강릉에서는 남동풍과 함께 지역 내 에어로솔이 대관령쪽으로 이동하면서 상승효과를 나타내는 경우이다. 마지막으로, 5번 사례는 대관령쪽의 약한 에어로솔 플룸이 강릉쪽으로 이동하면서 지역 내에서 배출된 에어로솔의 영향을 연속적으로 받는 경우에 해당한다.

4. 결 론

본 연구에서는 대관령과 강릉에서 동시에 관측된 칼럼 에어로솔 관측자료와 기상관측자료를 이용하여 머신러닝 방법 중 하나인 K-mean 클러스터링을 이용하여 연구대상 지역의 에어로솔의 분포 특성에 관한 유형화를 수행하였다. 2016년 3월부터 6월까지 3개월 간 관측된 대관령과 강릉 관측소의 컬럼 에어로솔 광학두께와 옴스트롬 지수, 수증기, 그리고 풍속과 풍향을 머신러닝을 위한 입력 변수로 선정하여 데이터를 수집하였다. 개별 자료에 대한 시계열 분석과 상관회귀분석 결과, 관측 기간 에어로솔 광학두께 평균값은 도시 지역인 강릉 (0.260 ± 0.134)이 대관령 ($0.197 \pm$

0.099)에 비하여 약 0.063 (32.0%) 높은 수준으로 나타났다. 옴스트롬 지수값의 차이는 ~ 0.031 로 강릉이 대관령에 비해 유사하거나 약간 작은 크기의 입자가 관측되었다. 이러한 결과는 두 관측 지점에서의 대기 중 입자상 물질의 크기는 유사하지만, 입자의 총 부하량은 강릉이 좀 더 높은 값으로 나타났다.

대관령과 강릉은 지형적인 차이로 인하여 기상 관측 결과도 다르게 나타났다. 대관령에서는 전체 발생 빈도의 약 61.5%가 서풍 계열이며, 27.8%가 동풍 계열의 바람이 불었으며, 5 m/sec 이상의 빠른 풍속을 가진 바람은 주로 서풍 계열의 바람이었다. 그러나 강릉에서는 전체 바람의 약 41.5%가 서풍 계열, 50.7%가 동풍 계열의 바람이 불었으며, 5 m/sec 이상의 풍속을 가진 바람은 서풍 계열이 대부분이었다. 바람 조건에 따라 에어로솔 플룸이 직간접적으로 영향을 미치는 경우, 대관령에서 관측되는 에어로솔은 지역 기원이거나, 영서 지방에서 발생한 에어로솔 입자가 서풍 계열의 바람과 함께 이동한 경우가 과반수 이상일 것이다. 강릉은 산지와 해양의 영향을 복합적으로 받고 있기 때문에 에어로솔의 환기와 농축이 바람에 영향을 많이 받을 것이다.

컬럼 관측 자료와 기상관측자료를 이용한 머신러닝 분석을 위하여 수집한 데이터에 표준화를 적용하였고 최적의 클러스터 개수를 5개로 결정하였다. 선정한 클러스터 개수를 토대로 K-mean 클러스터링을 수행한 결과 5가지 클러스터에 대한 입력변수에 대한 특징이 명확히 분류되었으며, 연구대상 지역의 에어로솔 사례를 5개의 유형으로 분류하였다. 첫 번째 사례는 지역 내 배경 대기 수준의 에어로솔에 관한 사례로서 강릉지역의 낮은 수준의 에어로솔이 대관령쪽으로 이동하는 경우이다. 두 번째 사례는 미세 입자가 서풍과 함께 대관령을 넘어 강릉 쪽으로 이동하는 동시에 강릉에서는 느린 남풍과 함께 지역내 에어로솔이 정체되어 다소 높은 수준이 에어로솔이 관측되는 경우이다. 세 번째 사례는 강한 서풍과 함께 초대입자가 대관령과 강릉에 동시에 영향을 미치는 경우이다. 네 번째 사례는 대관령에서 고농도 에어로솔

이 빠르게 이동하고, 강릉에서는 남동풍과 함께 지역 내 에어로솔이 대관령쪽으로 이동하면서 상승효과를 나타내는 경우이다. 마지막 사례는 대관령쪽의 약한 에어로솔 플룸이 강릉쪽으로 이동하면서 지역 내에서 배출된 에어로솔의 영향을 연속적으로 받는 경우이다.

보다 정밀한 분석을 위하여 보다 장기간의 관측 자료를 사용하는 것이 바람직하지만, 본 연구에서 사용된 자료는 근거리의 고도가 다른 두 관측지점에서의 자료로부터 머신러닝 기법을 이용하여 분류한 최초의 연구 결과로 의의가 있다. 따라서, 다양한 지점에서 다양한 시점에서의 관측자료가 필요하지만, 보다 정밀한 분석을 통해 결과를 분석할 필요가 있다.

감사의 글

이 논문은 2019년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(NRF-2019R1I1A3A01062804).

References

- Cha, J., Kim, J. (2018) Development of data mining algorithm for implementation of fine dust numerical prediction model, *Journal of the Korea Institute of Information and Communication Engineering*, 22(4), 595-601, (in Korean with English abstract). <https://doi.org/10.6109/jkiice.2018.22.4.595>
- Charrad, M., Ghazzali, N., Boiteau, V., Niknafs, A. (2014) NbClust: An R package for determining the relevant number of clusters in a data set, *Journal of Statistical Software*, 61(6), 1-36, (in Korean with English abstract). <https://doi.org/10.18637/jss.v061.i06>
- Cho, K., Jung, Y., Kang, C., Oh, C. (2019a) Conformity assessment of machine learning algorithm for particulate matter prediction, 23(1), 20-26, (in Korean with English abstract). <https://doi.org/10.6109/jkiice.2019.23.1.20>
- Cho, K., Lee, B.-y., Kwon, M., Kim, S. (2019b) Air quality prediction using a deep neural network model, *Journal of Korean Society for Atmospheric Environment*, 35(2), 214-225, (in Korean with English abstract). <https://doi.org/10.5572/KOSAE.2019.35.2.214>
- Dubovik, O., King, M.D. (2000) A flexible inversion algorithm for retrieval of aerosol optical properties from Sun and sky radiance measurements, *Journal of Geophysical Research*, 105, 20673-20696.
- International Panel on Climate Change (IPCC) (2013) *Climate Change 2013: the Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. https://www.ipcc.ch/site/assets/uploads/2017/09/WG1AR5_Frontmatter_FINAL.pdf
- Jeon, S., Son, Y.S. (2018) Prediction of fine dust PM₁₀ using a deep neural network model, *The Korean Journal of Applied Statistics*, 31(2), 265-285, (in Korean with English abstract). <https://doi.org/10.5351/KJAS.2018.31.2.265>
- Lee, K.H. (2012) Impact of Northeast Asian biomass burning activities on regional atmospheric environment, *Journal of the Korean Association of Geographic Information Studies*, 15(1), 184-196, (in Korean with English abstract).
- Lee, K.H., Lee, K.-T., Kim, J.-H., Mun, G.-H., Ahn, J.-M. (2018) Characteristics of aerosol mass concentrations and size distribution measured at Anheung, Korea, *Journal of Korean Society for Atmospheric Environment*, 34(5), 677-686, (in Korean with English abstract). <https://doi.org/10.5572/KOSAE.2018.34.5.677>
- Lee, K.H., Park, S.S. (2020) Characteristics of haze episode in 2019 December by using Satellite and ground measurements with trajectory model, *Journal of Korean Society for Atmospheric Environment*, 36(1), 128-138, (in Korean with English abstract).
- Lim, J.H., Park, J.S., Ahn, J.Y., Choi, J.S., Oh, J., Moon, K.J., Hong, Y.D., Han, J.S. (2013) The characteristics of the air pollutants at Baengnyeong Island, a west inflow region of the Korean Peninsula, *Journal of Korean Society of Urban Environment*, 13(3), 267-276, (in Korean with English abstract).
- MacQueen, J. (1967) Some methods for classification and analysis of multivariate observations, *Proceedings of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1, 281-297.
- National Institute of Environmental Research (NIER) (2019) Annual report of air quality in Korea, 2018, Ministry of Environment, <http://library.me.go.kr/search/>

- Detail View.ax?sid=1&cid=5683728.
- Oh, B.D., Park, J.H., Kim, Y.S. (2016) Prediction of the concentration of PM₁₀ using machine-learning, *Journal of Korea Information Science Society*, 20(12), 1674-1676, (in Korean with English abstract).
- Park, S.-S., Cho, S.-Y., Jung, C.-H., Lee, K.-H. (2016) Characteristics of water-soluble inorganic species in PM₁₀ and PM_{2.5} at two coastal sites during spring in Korea, *Atmospheric Pollution Research*, 7(2), 370-383, <https://doi.org/10.1016/j.apr.2015.10.018>
- Park, S.-S., Kim, S.-J., Gong, B.-J., Cho, S.-Y., Kim, J.-C., Lee, S.J. (2013) Investigation on a haze episode of fine particulate matter using semi-continuous chemical composition data, *Journal of Korean Society for Atmospheric Environment*, 29(5), 642-655, (in Korean with English abstract).
- Shin, S.K., Lee, K.H. (2016) Trend analysis of wildland fires and their impacts on atmospheric environment over East Asia, *Asian Journal of Atmospheric Environment*, 10(1), 22-31, <https://doi.org/10.5572/ajae.2016.10.1.022>
- Yeo, M.J., Kim, Y.P. (2019) Trends of the PM₁₀ concentrations and high PM₁₀ concentration cases in Korea, *Journal of Korean Society for Atmospheric Environment*, 35(2), 249-264, (in Korean with English abstract). <https://doi.org/10.5572/KOSAE.2019.35.2.249>

Authors Information

이권호 (강릉원주대학교 대기환경과학과 교수)

이규테 (강릉원주대학교 대기환경과학과 교수)