

기술자료

국내 기준성 대기오염물질의 고농도발생 현황과 권역간의 연관성분석: PM_{2.5}를 중심으로

Analytical Study on the Trends of the High PM_{2.5} Concentrations and Interregional Association

김태희*, 명지수, 정미숙, 이용희, 서인석

국립환경과학원 대기질통합예보센터

Taehee Kim*, Jisu Myoung, Misuk Jung, Yonghee Lee, Insuk Suh

Air Quality Forecasting Center, National Institute of Environmental Research, Incheon, Republic of Korea

접수일 2020년 1월 21일
수정일 2020년 2월 28일
채택일 2020년 4월 8일

Received 21 January 2020
Revised 28 February 2020
Accepted 8 April 2020

*Corresponding author
Tel : +82-(0)32-560-7257
E-mail : kimth0806@korea.kr

Abstract The standards for PM_{2.5} forecast grade were tightened after March 27, 2018 (bad grade: daily average more than [51 µg/m³] → [31 µg/m³]). In this study, The characteristics of high concentration according to the tightened standards were analyzed from two perspectives. The daily mean data for each forecast areas over 3 years (December 2015 to November 2018) was used for analysis. First, The association between the forecast areas of the same time zone was analyzed using the conditional probability. On the whole season, Gangwon-east and Jeju had the lowest association with other regions. On the contrary, Seoul metropolitan area and Chungbuk were highly interconnectivity in spring, and Gangwon-west, Chungbuk and Gyeonggi northern were the most important areas that affected other regions in winter. Second, The association between the forecast areas considering time continuity using high concentration prediction models for each areas that was developed by logistic regression analysis. The high temporal association was found in most of the area except Gyeonggi southern, Daejeon and Gyeongbuk that means if high concentration occur in A area in today, high concentration will occur in A area in tomorrow. Based on the odds ratio of the prediction model, Gyeonggi northern was selected as the most influential area impact on the probability of occurring high concentration of the other areas, and Busan, Daegu and Gyeongnam were selected as the least influential region.

Key words: PM_{2.5}, High concentration, Conditional probability, Predictive model, Logistic regression analysis

1. 서 론

경제협력개발기구(Organization for Economic Cooperation and Development, OECD)는 초미세먼지(Particulate Matter less than 2.5 µm, PM_{2.5})로 인해 연간 18,200명이 사망할 정도로 한국의 대기오염이 심각한 수준이라 경고하였다(Zhang *et al.*, 2017). 또한 대책을 마련하지 않는다면 2060년 OECD 회원국 중 한국은 조기사망률 및 경제적 피해가 가장 클 것이라

발표한 바 있다(Zhang *et al.*, 2017). 환경부는 대기오염으로 인한 국민 건강 피해를 최소화하기 위해 2013년 8월 미세먼지(Particulate Matter less than 10 µm, PM₁₀)에 대한 시범예보를 시작으로 2015년에는 PM_{2.5}와 오존(Ozone, O₃)을 추가하여 매일 4회(5시, 11시, 17시, 23시), 19개 권역(그림 1)에 대해 예보하고 있다. 국가대기질예보는 다양한 수치모델 결과와 관측자료를 기반으로 한 예보관의 종합적인 판단에 의해 결정된다. 이에 수치모델의 정확도 향상과 더불어

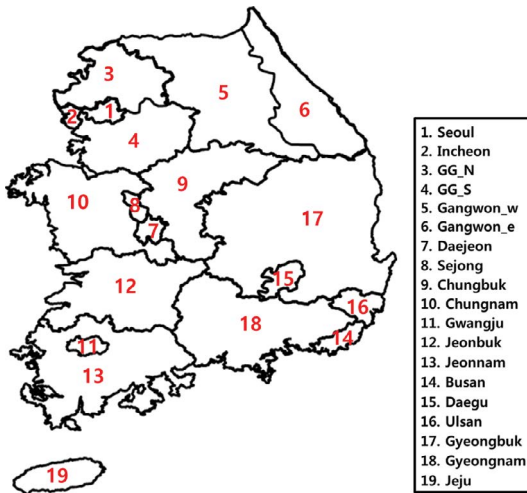


Fig. 1. Information on 19 target regions of National Air Quality Forecast.

어 관측자료의 시공간적 특성을 파악하는 것은 예보 결정에 있어 중요한 부분이다. 특히 2018년 3월 27일부터 $PM_{2.5}$ 의 대기환경기준과 ‘나쁨’ 예보등급기준이 일평균 $51 \mu\text{g}/\text{m}^3$ 이상에서 주요 선진국 수준인 $36 \mu\text{g}/\text{m}^3$ 이상으로 변경됨에 따라 과거대비 고농도발생일수(‘17년 서울 기준 20일 → 64일)와 장기지속사례(‘17년 서울기준 3일 이상 지속 2회 → 12회)의 빈도가 증가되면서, 변경된 기준에 따른 $PM_{2.5}$ 시공간적 고농도발생특성을 파악할 필요성이 대두되고 있다.

과거 선행연구들은 군집분석, 상관성분석 등, 다양한 통계방법을 이용하여 PM_{10} 과 $PM_{2.5}$ 의 시공간적 특성을 파악하였다. Do and Jung (2017)은 부산의 도시대기측정소 최적화작업의 사전연구로 군집분석을 활용하여 측정소 간의 PM_{10} 과 O_3 농도 유사성을 분석한 바 있으며, Oh and Lim (2017)은 우리나라 지역별 PM_{10} 과 $PM_{2.5}$ 의 위험도를 분석하여 우리나라 동부지역(경북, 울산 등)과 서부지역(광주 등)은 서로 상반된 위험도를 지니고 있는 것을 확인하였고, 개선이 가장 시급한 지역으로 수도권을 제시한 바 있다. 이후 Yeo and Kim (2019)은 2001년부터 2017년까지의 전국 PM_{10} 농도를 분석하여 국민들의 체감만큼 고농도발생빈도와 연속일수가 증가하였는지 확인한 바

있으며, Lee *et al.* (2019)은 서울지역의 시간별, 일별, 요일별, 주간별, 월별, 계절별 $PM_{2.5}$ 시계열적 오염특성을 파악하기 위하여 14개의 오염원과 기상변수간의 통계분석을 실시한 바 있다.

이러한 선행연구를 기반으로 더 나아가 예측에 초점을 맞춘 연구 또한 수행되어 왔다. Lee *et al.* (2006)은 회귀분석과 신경망분석 그리고 SVFR (Support Vector Regression)분석을 활용해 포항시 PM_{10} 예측모형을 개발하였고, Koo *et al.* (2010)은 신경망방법과 고농도회귀·신경망방법을 이용하여 수도권지역의 PM_{10} 예측모형을 개발한 바 있다. 이후 Lee *et al.* (2018)은 CMIP5에 참여한 모델을 이용하여 대기패턴 변화에 따른 PM_{10} 정체 여부를 분석하여 예측모형을 개발하였고, 이를 이용하여 대기패턴변화와 미래의 고농도발생의 연관성을 파악한 바 있다. 이처럼 PM_{10} 에 대한 과거와 미래의 연관성을 분석한 연구는 선행된 바 있으나, $PM_{2.5}$ 에 대한 연구는 미비한 실정이다.

본 연구에서는 과거 3년간의 전국 도시대기측정망 $PM_{2.5}$ 자료를 이용하여 변경된 나쁨 기준($36 \mu\text{g}/\text{m}^3$ 이상)에 따른 예보권역별 동일시간대의 발생확률과 시간적 연속성이 고려된 연관성을 파악하고자 한다. 3.1장에서는 예보권역·계절별 고농도발생현황을 먼저 파악하고, 3.2장에서는 동일시간대에 대한 권역 간의 고농도발생확률을 분석하였다. 마지막으로 3.3장에서는 로지스틱 회귀분석을 이용하여 내일 고농도발생 예측모형을 개발하여, 권역 간의 연관성을 분석하였다.

2. 연구 방법

변경된 $PM_{2.5}$ ‘나쁨’ 예보기준(일평균농도 $36 \mu\text{g}/\text{m}^3$ 이상)에 따른 전국 19개 예보권역의 시공간적 고농도발생특성을 파악하고자 조건부확률과 로지스틱 회귀분석을 이용하여 동시간대의 권역 간 ‘조건부 고농도발생확률’을 산정하고 시간적 연속성을 고려한 ‘고농도발생예측모형’을 개발하였다. 조건부확률

은 어떤 사건이 일어났다는 조건 하에 해당 사건이 일어날 확률이다. 이에 조건부 고농도발생확률 ($P(R_B|R_A)$)은 오늘 A 권역에서 고농도가 발생하였을 때, 오늘 B 권역에서도 고농도가 발생할 확률을 나타낸다. 이는 식 1과 같이 A와 B 권역에서 동시에 고농도가 발생한 일수 ($n(R_B \cap R_A)$)와 A 권역에서 고농도가 발생한 일수 ($n(R_A)$)의 비로 산정되며, 1 혹은 100%에 가까울수록 동시간대 권역 간 고농도발생연관성이 크음을 의미한다.

$$P(R_B|R_A) = \frac{P(R_B \cap R_A)}{P(R_A)} = \frac{n(R_B \cap R_A)}{n(R_A)} \quad (1)$$

“고농도발생예측모형”에 사용된 로지스틱 회귀분석(식 2)은 일반적인 회귀분석의 목표와 동일하게 설명변수와 반응변수 간의 관계를 구체적인 함수로 나타내나, 범주형 데이터를 대상으로 하는 차이점이 있다. 예측모형은 통계패키지 R (<https://www.rstudio.com>)을 이용하여 각 권역별로 개발되었다. 내일 C 권역의 예측모형 개발에는 전체 권역의 오늘 고농도발생 유무를 사용하여, 오늘의 고농도발생분포(설명변수)가 내일 C 권역의 고농도발생(반응변수)에 미치는 영향을 종합적으로 분석하였다. 이때 모델의 적합성 판단을 위하여 R에서 제공하는 ‘backward selection’ 방법을 사용하여, 유의수준 0.01에서 유의하지 않은 권역을 하나씩 줄여가며 모델의 성능을 향상시켰으며, 2018년 12월부터 2019년 4월까지에 대해 예측모형의 적중률(Hit rate)을 산정하여 정확성을 검증하였다. 개발된 예측모형에서 각 설명변수의 계수인 편회귀계수(α_n)와 오즈비(e^{α_n})는 설명변수와 반응변수 간의 연관성을 나타낸다. 편회귀계수는 다른 설명변수들이 모두 일정한 상태에서 하나의 설명변수 대비 반응변수가 변하는 정도를 나타내는 지수이다. 편회귀계수가 양수이면 설명변수와 반응변수가 참일 확률(고농도가 발생할 확률)은 비례하고, 반대로 음수이면 설명변수와 반응변수가 참일 확률은 반비례한다. 따라서 동일 예측모델 안에서 편회귀계수가 클수록 설명변수와 반응변수 간의 연관성이 더 크음을 의미

하며, 이는 편회귀계수가 가장 큰 권역의 오늘 고농도발생 유무가 내일 C 권역의 고농도발생에 가장 큰 영향을 준다는 것을 의미한다. 오즈비는 설명변수에 따른 반응변수가 참일 확률과 거짓일 확률의 비로서, 편회귀계수와 마찬가지로 설명변수가 반응변수에 미치는 영향의 크기와 방향을 파악할 수 있다. 참일 확률과 비교하여 거짓일 확률이 작으면 오즈비는 1보다 큰 값을 가지며, 거짓일 확률이 크면 1보다 작은 값을 가진다. 이때 참일 확률은 고농도가 발생할 확률이며, 거짓일 확률을 고농도가 발생하지 않을 확률이다.

$$\log \frac{p}{1-p} = y_0 + \alpha_1 R_1 + \alpha_2 R_2 + \dots + \alpha_n R_n \quad (2)$$

p : probability that the response variable is true

R_n : n-th explanatory variable

α_n : partial regression coefficient of R_n

y_0 : intercept

본 연구에서는 한 권역 이상에서 일평균농도가 36 $\mu\text{g}/\text{m}^3$ 이상인 날을 “고농도일”이라 정의하였고, 고농도일이 연속해서 나타난 사례를 “고농도지속사례”로 선정하였다. 분석에는 2015년 12월부터 2018년 11월까지의 권역별 일평균농도자료가 사용되었고, 권역별 일평균농도자료는 에어코리아 (www.airkorea.or.kr)에서 제공하는 도시대기측정망자료를 이용하여 각 예보권역에 해당하는 관측소의 1시~24시까지의 데이터를 평균하여 산정하였다. 고농도사례일 중 결측이 존재하는 날은 총 73일이었으며 고농도예측모형 개발에는 제외되었다.

3. 연구 결과

3.1 고농도발생 현황

총 3년 동안 고농도일은 총 527일로 전체의 48.1%를 차지하였다. 그림 2는 권역별, 계절별 고농도일의 발생빈도를 나타낸다. 계절별 발생빈도는 겨울, 봄,

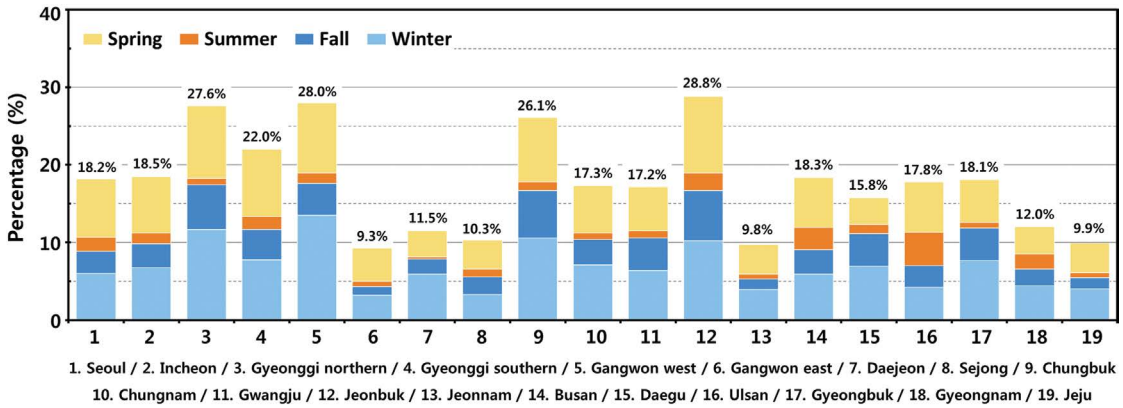


Fig. 2. The seasonal percentage of high concentration episode by each region. The value represents the total percentage for three years.

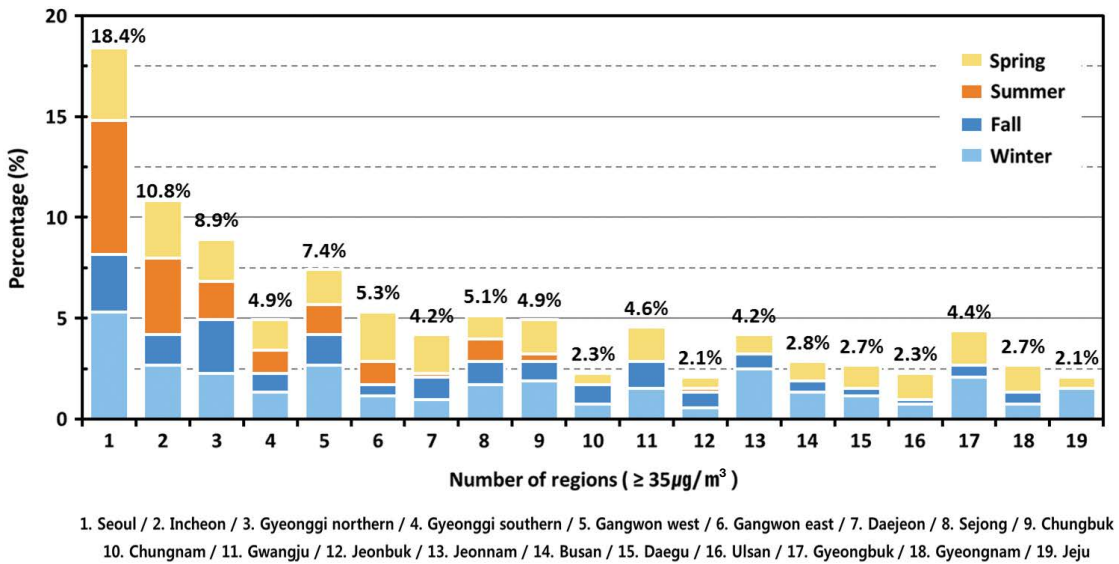


Fig. 3. The seasonal percentage of the number of areas in which high concentrations occurred at the same day. The value represents the total percentage for three years.

가을, 여름 순으로 높았으며, 각각 173일 (15.8%), 157일 (14.3%), 102일 (9.3%), 95일 (8.7%)에 해당하였다. 이러한 계절별 특징은 권역별로 동일하였다. 권역별 발생빈도는 전북에서 가장 높았으며, 영동에서 가장 낮았다. 발생빈도 상위 3개의 권역에는 전북 (316일, 28.8%), 영서 (307일, 28.0%), 경기북부 (303일, 27.6%)가 해당하였으며, 하위 3개의 권역에는 영동 (102일,

9.3%), 전남 (107일, 9.8%), 제주 (109일, 9.9%)가 해당하였다.

그림 3은 계절별 고농도 발생권역 수의 분포를 나타낸다. 전반적으로 고농도일 발생빈도와 발생권역 수는 반비례하였다. 과반수 (10개 이상) 권역에서 고농도가 발생한 날은 총 158일로서 전체의 30.0%를 차지한 반면, 1~3개의 권역에서 고농도가 발생한 날

은 총 201일로 38.1%를 차지하였다. 특히 여름의 경우, 과반수 권역에서 고농도가 발생한 날은 1일인 반면, 1~3개의 권역에서 고농도가 발생한 날은 65일로, 여름에 발생한 고농도일은 상대적으로 발생 권역 수가 적음을 알 수 있다. 고농도일이 2일 이상 지속된 사례(고농도지속사례)는 겨울(36개), 봄(25개), 가을(21개), 여름(16개) 순으로 많이 발생하였으며, 이는 앞선 고농도발생빈도 순서와 동일하다. 계절별 평균 지속기간은 봄이 6일, 여름이 5일, 가을과 겨울이 4일로, 지속기간에 따른 빈도 분포는 동일하게 왼쪽으로 치우쳐진 양의 왜도분포를 보였다. 봄, 여름, 가을, 겨울의 최대 지속기간은 27일, 9일, 8일, 10일로 지속강도는 봄에 가장 강하게 나타났으며, 단일 권역의 최대 지속시간 또한 19일(전북), 7일(울산), 7일(경기북부), 10일(영서, 충북)으로 고농도 최대 발생빈도 권역과 동일하였다.

3.2 고농도발생 확률

그림 4는 A권역과 B권역 사이의 고농도발생 조건부확률($P(R_B|R_A)$)을 나타낸다. 이는 오늘 A권역에서 고농도가 발생하였을 때 B권역에서도 고농도가 발생

할 확률이다. 예를 들어 서울 권역(A)에서 고농도가 발생하였을 때 경기북부 권역(B)에서도 고농도가 발생할 확률은 80% 이상이나, 경기북부 권역(A)에서 고농도가 발생하였을 때 서울 권역(B)에서도 고농도가 발생할 확률은 80% 미만이다. 이때 A권역과 B권역 간에 $P(R_B|R_A)$ 와 $P(R_A|R_B)$ 모두 80% 이상인 경우, '상호조건부확률이 80% 이상이다.'라고 한다. 이는 두 권역 중 하나의 권역에서 고농도가 발생하였을 때, 80% 이상의 확률로 나머지 권역에서 고농도가 발생할 수 있음을 의미하며, 나아가 두 권역 간의 고농도발생연관성이 높음을 의미한다.

전체기간에 대한 분석 결과, 서울과 인천, 경기북부와 영서, 부산과 울산은 상호조건부확률이 80% 이상으로 고농도발생연관성이 높았다. 타 권역과의 영동, 대전, 세종, 전남, 제주의 조건부확률은 평균 34%, 48%, 38%, 42%로 상대적으로 낮은 특징을 보였다. 특히 영동과 영서, 대전과 충북은 인접한 권역이지만 발생기준권역(A)에 따라 조건부확률이 상이하였다. 영동과 영서의 고농도조건부확률($P(R_{영서}|R_{영동})$)은 90%인 반면, 영서와 영동($P(R_{영동}|R_{영서})$)의 조건부확률은 30%로 낮았으며, 마찬가지로 대전과 충북의 조

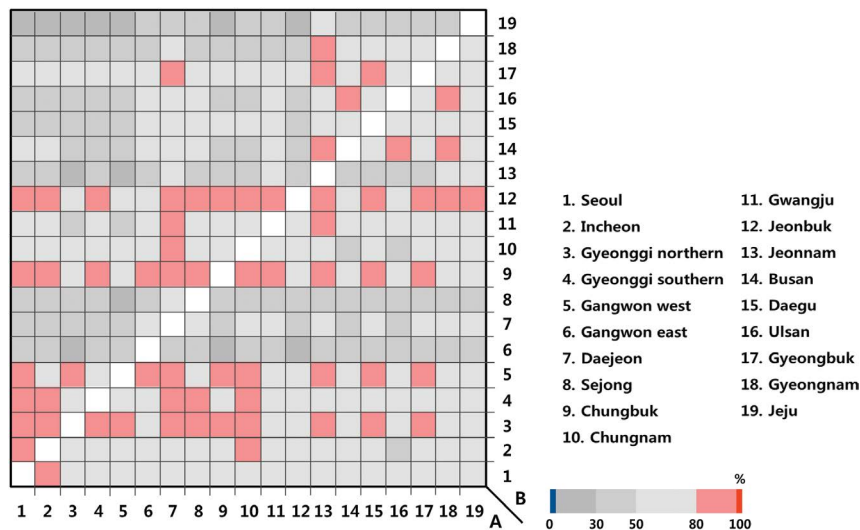


Fig. 4. The conditional probability for A and B region for total period. It means the probability that the high concentrations occur in B region at the same time, when high concentrations occur in A region. This result is only established when the width is A region and the length is B region. It doesn't establish in the opposite case.

Table 1. The information of regions that is high interrelationship about occurrence of high concentrations in Summer. These include regions where the mutual conditional probability is 0% ($P(R_B|R_A)$ and $P(R_A|R_B) \geq 0\%$).

Region			Region				
1	Seoul	↔	Jeju	16	Gangwon east	↔	Daejeon
2	Incheon	↔	Daejeon	17	Gangwon east	↔	Jeju
3	Incheon	↔	Chungbuk	18	Daejeon	↔	Busan
4	Incheon	↔	Gwangju	19	Daejeon	↔	Ulsan
5	Incheon	↔	Jeonnam	20	Daejeon	↔	Gyeongbuk
6	Incheon	↔	Gyeongbuk	21	Daejeon	↔	Jeju
7	Incheon	↔	Jeju	22	Sejong	↔	Chungnam
8	Gyeonggi northern	↔	Gwangju	23	Sejong	↔	Jeju
9	Gyeonggi northern	↔	Jeonnam	24	Chungnam	↔	Jeonnam
10	Gyeonggi northern	↔	Daegu	25	Chungnam	↔	Gyeongbuk
11	Gyeonggi northern	↔	Gyeongbuk	26	Chungnam	↔	Jeju
12	Gyeonggi northern	↔	Jeju	27	Jeonnam	↔	Jeju
13	Gyeonggi southern	↔	Jeonnam	28	Daegu	↔	Jeju
14	Gyeonggi southern	↔	Jeju	29	Gyeongbuk	↔	Jeju
15	Gangwon west	↔	Jeju				

건부확률은 99%인 반면 충북과 대전의 조건부확률은 44%로 낮은 확률을 보였다. 이처럼 지리적으로 인접한 권역이더라도 고농도발생연관성이 상이한 권역이 존재하였다.

계절별 분석 결과, 봄과 가을 그리고 겨울의 조건부확률은 전체기간과 유사하였으나, 여름에는 조건부확률이 0%로 연관성이 매우 낮은 권역들이 다수 발견되어 권역 간의 연관성이 낮은 계절에 속하였다. 표 1은 여름에 상호조건부확률이 0%인 권역들로 나타낸 것으로 총 29쌍이 해당하였다. 이 중 대전과 충남은 지리적으로 인접해 있음에도 불구하고 상호조건부확률이 0%로 나타났으며, 상호연관성이 가장 낮은 특징을 보였다. 이 외에도 연관성이 낮은 권역에는 제주가 해당하였고, 충북, 광주, 전북, 부산, 울산, 경남을 제외한 나머지 12개의 권역과의 연관성이 모두 낮게 나타났다. 표 2는 고농도발생연관성이 높은 권역들로서 A 권역과 B 권역의 조건부확률이 100%이거나 상호조건부확률이 80% 이상이다. 먼저 봄의 경우, 대전과 충북 그리고 대전과 전북의 조건부확률은 100%로, 대전에서 고농도가 발생하였을 때 반드시 충북과 전북에서 고농도가 발생하였으나, 그 반대의 경우는 성립하지 않았다. 상호조건부확률이 80% 이상인 권역들은 총 9

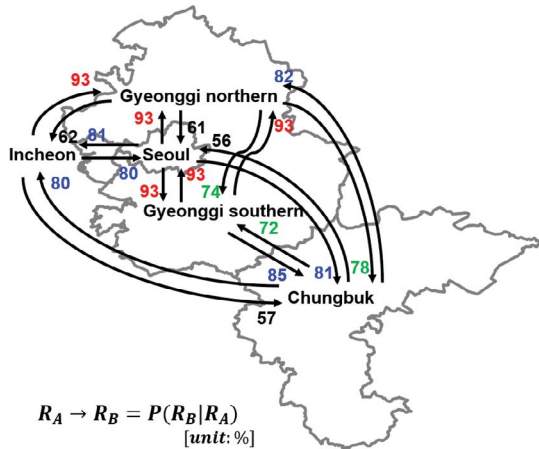


Fig. 5. The distribution of conditional probability in the Capital region (Seoul, Incheon, Gyeonggi northern, Gyeonggi southern) and Chungbuk in spring.

쌍으로 서울과 인천, 서울과 경기남부, 경기북부와 경기남부, 경기북부와 영서, 경기남부와 충북, 부산과 울산이 이에 해당하였다. 이 중 인접한 권역인 서울, 인천, 경기북부, 경기남부 그리고 충북은 전반적으로 고농도발생 상호연관성이 큰 것으로 분석되었다(그림 5). 여름의 경우, 상호조건부확률이 80% 이상인 권역은 존재하지 않았으나, A 권역과 B 권역의 조건부확률

Table 2. The information of regions that is high interrelationship about occurrence of high concentrations. These include regions where the conditional probability is 100% or where the mutual conditional probability is more than 80%.

		Region	
		R_A	R_B
Conditional probability = 100% [$P(R_B R_A) = 100\%$]	Spring	Daejeon	Chungbuk, Jeonbuk
	Summer	Gyeonggi northern	Seoul, Gyeonggi southern
		Daejeon	Sejong, Jeonbuk
		Gwangju	Jeonbuk
		Gyeongbuk	Busan, Ulsan, Gyeongnam
	Fall	Seoul	Gyeonggi northern
		Daejeon	Chungbuk
		Jeonnam	Chungbuk, Gwangju, Jeonbuk
	Winter	Seoul	Gyeonggi northern
		Incheon	Gyeonggi northern
Gyeonggi southern		Gyeonggi northern	
Daejeon		Gangwon west, Chungbuk	
Sejong		Gyeonggi northern, Gangwon west, Chungbuk	
Jeonnam		Gwangju, Jeonbuk	
Ulsan	Busan		
Mutual conditional probability $\geq 80\%$ [$P(R_B R_A)$ and $P(R_A R_B) \geq 80\%$]	Spring	Seoul	Incheon
		Seoul	Gyeonggi southern
		Gyeonggi northern	Gyeonggi southern
		Gyeonggi northern	Gangwon west
		Gyeonggi southern	Chungbuk
		Busan	Ulsan
	Summer	-	-
	Fall	Chungbuk	Jeonbuk
	Winter	Seoul	Incheon
		Incheon	Chungnam
Gyeonggi northern		Gangwon west	
Gyeonggi northern		Chungbuk	
Chungbuk		Jeonbuk	
Jeonnam		Gyeongnam	
Daegu		Gyeongbuk	
Ulsan	Gyeongnam		

이 100%인 고농도발생연관성이 높은 권역이 다수 존재하였다. 이러한 권역에는 경기북부와 서울, 경기북부와 경기남부, 대전과 세종, 대전과 전북, 광주와 전북, 경북과 부산, 경북과 울산 그리고 경북과 경남이 해당하였다. 가을의 경우, 서울과 경기북부, 대전과 충북, 전남과 충북, 전남과 광주 그리고 전남과 전북의 조건부확률이 100%로 고농도발생연관성이 높았으며, 충북과 전북의 상호조건부확률은 80% 이상으로 상호

연관성이 있는 것으로 판단되었다. 마지막으로 겨울의 경우, 조건부확률이 100%인 권역이 총 11쌍으로 권역간의 고농도발생연관성이 4계절 중 가장 높았으며, 상호조건부확률이 80% 이상인 권역 또한 8쌍으로 가장 높은 상호연관성을 보였다. 특히 영서와 충북 그리고 경기북부는 자신을 제외한 나머지 18개 권역과의 조건부확률이 각각 평균 94.9%, 91.8%, 91.2%로 고농도발생연관성이 가장 높은 권역으로 분석되었다.

Table 3. The equation and hit rate of regional prediction model that is calculate the probability of high concentration occurrence.

Region (R_n)	Model $\left(\log \frac{p}{1-p}\right)$	Hit rate (%)
1 Seoul	$-2.4715182 + 1.0068315R_1 + 0.8378894R_3 + 0.8623243R_{12}$	77.3
2 Incheon	$-2.4590160 + 0.9336457R_2 + 0.9694621R_3 + 0.7842804R_9$	81.3
3 GG_N	$-1.8691161 + 0.8768208R_2 + 1.1493144R_3 + 0.8837247R_9$	76.0
4 GG_S	$-2.1898867 + 0.9897730R_2 + 0.8398026R_3 + 0.9360731R_9$	74.7
5 Gangwon_w	$-2.1166174 + 1.0368692R_3 + 1.5105476R_5 + 0.6749599R_{12}$	77.3
6 Gangwon_e	$-3.5035104 + 1.7356617R_5 + 0.8546809R_6 + 0.745371R_{11}$	88.0
7 Daejeon	$-3.3972597 + 1.0371367R_5 + 1.3231637R_{10} + 0.8898005R_{12}$	77.3
8 Sejong	$-3.3981446 + 0.9883847R_3 + 1.6477709R_8 + 0.8948534R_{12}$	72.7
9 Chungbuk	$-2.1710161 + 0.8554309R_5 + 1.1616964R_6 + 1.0621046R_{12}$	74.0
10 Chungnam	$-2.7249851 + 1.0789502R_3 + 1.1825875R_{10} + 0.8420969R_{12}$	77.3
11 Gwangju	$-2.8244785 + 0.8882846R_3 + 1.3431075R_9 + 1.2869415R_{11} + 0.6739227R_{16} - 0.8043783R_9$	80.7
12 Jeonbuk	$-1.9300311 + 1.0396093R_3 + 0.7354586R_{11} + 1.3969940R_{12}$	78.0
13 Jeonnam	$-3.467713 + 1.596993R_3 + 2.139282R_{13}$	84.0
14 Busan	$-2.467215 + 1.394062R_9 + 1.638123R_{16}$	78.7
15 Daegu	$-2.435694 + 2.150270R_7 + 1.258757R_{16}$	77.3
16 Ulsan	$2.261763 + 2.510460R_{16}$	81.3
17 Gyeongbuk	$-2.8654735 + 1.1499992R_5 + 1.5773343R_9 + 0.7226943R_{16}$	80.0
18 Gyeongnam	$-3.2363821 + 1.1814663R_5 + 0.7936621R_{11} + 0.8832434R_{13} + 1.0015342R_{16}$	85.3
19 Jeju	$-3.0862625 + 1.1411116R_4 + 0.8619163R_{11} + 1.0836660R_{19}$	84.7

3.3 로지스틱 회귀분석을 통한 권역별 연관성 분석

표 3은 로지스틱 회귀분석을 이용하여 개발한 각 권역별 고농도발생 예측모형식과 예측모형의 적중률(Hit rate)이다. 이때 적중률은 각 권역별 예측모형에서 예측한 확률이 0.7 이상인 날을 고농도일로 가정하여 산정한 결과이다. 개발된 예측모형은 평균 적중률이 79.3%로 우수한 예측성능을 보였다. 이때 가장 높은 적중률을 보인 권역은 영동(88.0%)이었으며, 가장 낮은 적중률을 보인 권역은 세종(72.7%)이었다.

가장 우수한 예측성능을 보인 영동(반응변수)의 예측모형에는 영서와 영동 그리고 광주의 고농도발생 유무가 설명변수로 사용되었다. 편회귀계수는 모두 양수로서 영서와 영동 그리고 광주에서 오늘 고농도가 발생할 경우, 내일 영동의 고농도발생확률이 증가하였다. 이처럼 대부분의 예측모형에서 설명변수들의 편회귀계수는 양수로 설명변수권역에서 오늘 고농도가 발생할 경우, 내일 반응변수권역의 고농도발생확률이 증가하는 것으로 분석되었다. 예외적으로 광주의 예측모형에서 경북은 편회귀계수가 음수

(-0.8043783)로 경북에서 오늘 고농도가 발생할 경우, 내일 광주의 고농도발생확률은 감소하는 특징을 보였다.

설명변수권역들의 오늘 고농도발생이 내일 반응변수권역의 고농도발생에 미치는 영향의 크기를 정량적으로 비교하고자 오즈비를 산정하였다(그림 6). 앞선 편회귀계수와 동일하게 광주의 예측모형에서 경북의 오즈비는 0.4로 오늘 경북의 고농도발생과 내일 광주의 고농도발생확률은 반비례관계에 있음을 뜻한다. 또한 이는 경북에서 고농도가 발생하였을 때, 내일 광주에서 고농도가 발생할 확률은 발생하지 않을 확률의 0.4배임을 의미한다. 대부분의 고농도예측모형에서 오즈비는 4 이하의 값을 가지는 것으로 나타났다. 오즈비가 4 초과인 반응변수(내일)와 설명변수(오늘)는 총 9개로 영동과 영서(5.7), 세종과 세종(5.2), 전북과 전남(8.5), 전남과 경기남부(4.9), 부산과 울산(5.1), 울산과 울산(12.3), 경북과 충북(4.8), 대구와 대전(8.6)이 해당하였다. 특히 울산의 고농도 예측모형에는 오늘의 울산의 고농도발생 유무만 설

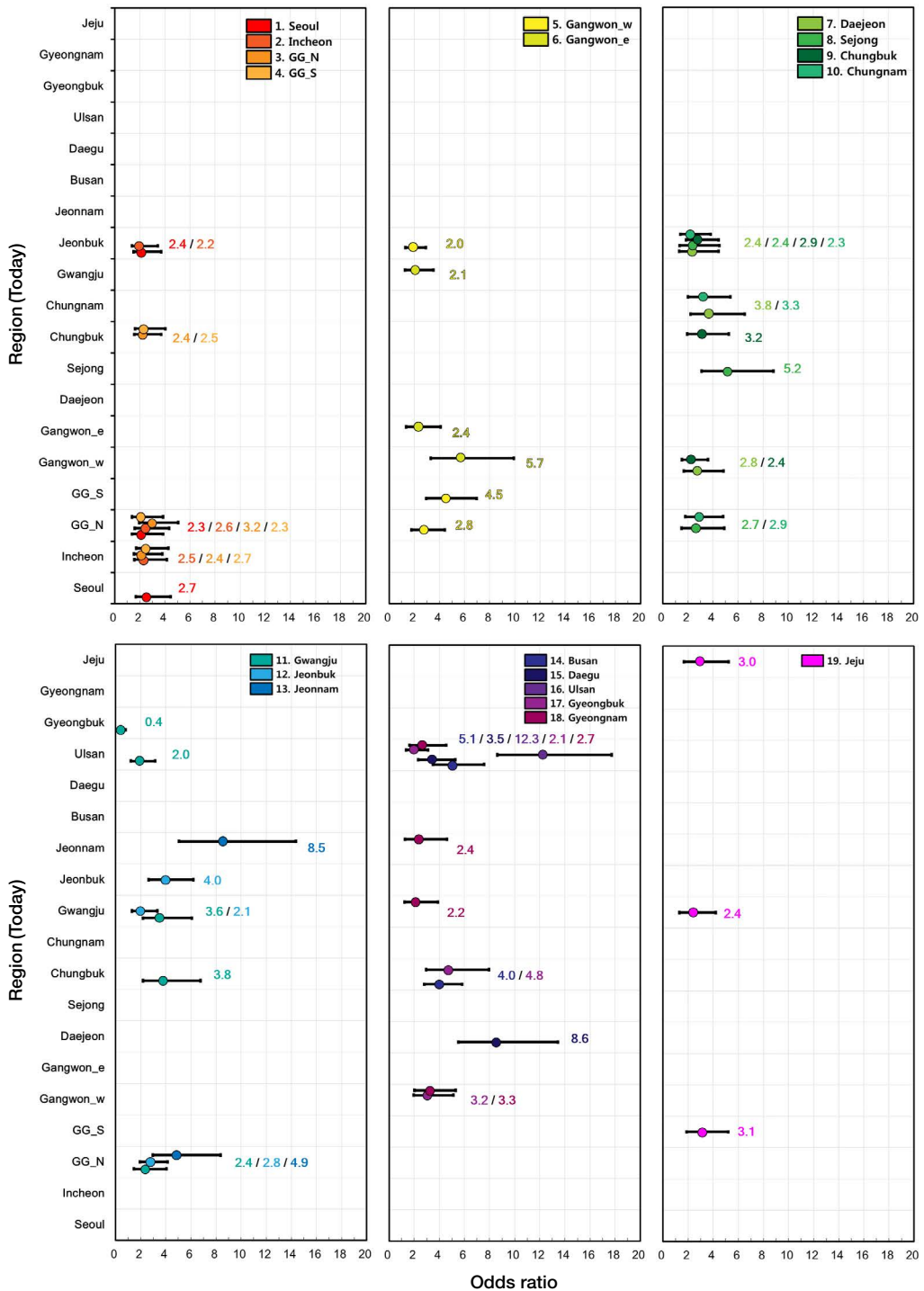


Fig. 6. The distribution of odds ratio of regional prediction model. The confidence intervals (95% CI) were adjusted for multiple comparisons by using the Bonferroni correction. If the probability of true is greater than false, odds ratio has a value greater than 1. In this study, the probability of true means the probability of high concentration occurring.

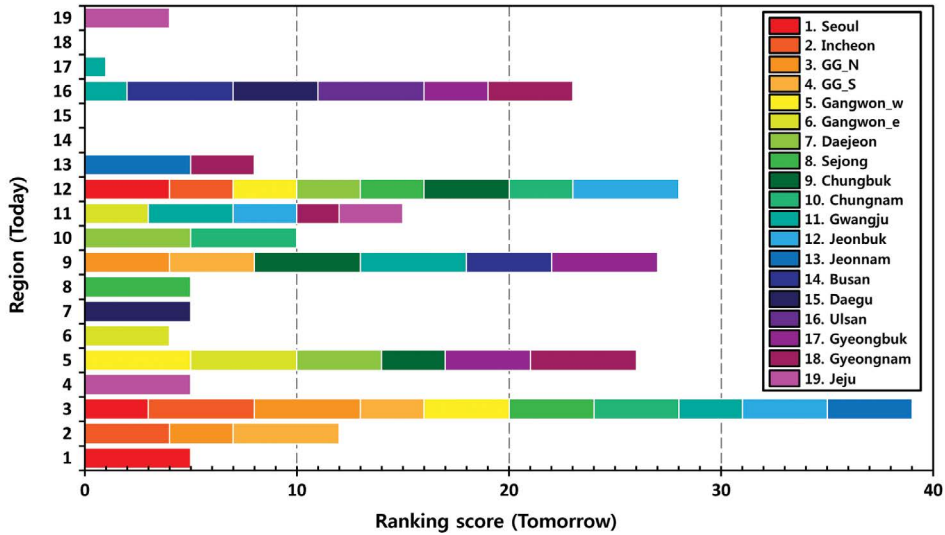


Fig. 7. The results of ranking score about dependent variable (tomorrow) based on the odds ratio strength of the explanatory variable (today). The top ranking for each regional prediction model is five points, and the lower the ranking, the smaller the one score will be.

명변수로 사용되는 특징을 보였으며, 오늘 고농도 발생 시 내일 고농도발생확률은 고농도가 발생하지 않을 확률의 12.3배로 외부의 영향보다 자체 오염물질 배출의 영향을 더 많이 받음을 알 수 있었다. 그림 7은 각 고농도예측모형별로 오즈비 강도를 기준으로 1순위 권역부터 5순위 권역까지 5점에서 1점의 점수를 매긴 후, 설명변수(y 축)를 기준으로 점수의 합계(x 축)와 반응변수별 점수 분포(범례)를 나타낸 결과이다. 예를 들어, 오늘 인천의 고농도발생 유무는 내일 인천과 경기북부 그리고 경기남부의 고농도발생 확률에 영향을 주며, 권역별 점수는 4점, 3점, 5점으로 경기남부와 인천 그리고 경기북부 순으로 연관성이 크게 나타났다. 대부분의 권역들은 A권역의 내일 고농도예측모형에 A권역의 오늘 고농도발생 유무가 사용되어 연관성이 높은 것으로 나타났으나, 경기남부와 대전 그리고 경북은 오늘과 내일의 고농도발생 연관성이 낮은 특징을 보였고, 이는 자체오염물질 배출보다 외부의 영향을 더 많이 받는 것을 의미한다. 경기북부는 총 39점으로 연관성이 가장 높은 권역으로 선정되었으며, 경기북부의 오늘 고농도발생 유무

는 내일 10개 권역의 고농도발생확률에 영향을 주는 것으로 분석되었다. 반면, 부산과 대구 그리고 경남은 총 0점으로 오늘 고농도발생 유무가 영향을 주는 권역은 없는 것으로 분석되었다.

4. 요약 및 결론

본 연구에서는 2015년 12월부터 2018년 11월까지를 대상으로 변경된 $PM_{2.5}$ 나쁨 기준에 따른 예보권역별 고농도발생특성을 분석하였다. 이를 위해 조건부확률을 이용하여 동일시간대의 A권역과 B권역의 고농도발생확률을 파악하였고, 로지스틱 회귀분석을 이용한 고농도예측모형을 개발하여 오늘의 A권역과 내일의 B권역의 고농도발생 연관성을 분석하였다.

대상기간 동안 계절별 고농도발생빈도는 겨울, 봄, 가을, 여름 순으로 높았으며, 권역별 고농도발생빈도는 전북에서 가장 높게 나타났다. 마찬가지로 고농도 지속사례 또한 겨울, 봄, 가을, 여름 순으로 많이 발생하였으나, 지속강도는 봄에 가장 강한 것으로 분석되

었다. 조건부확률을 이용한 동일시간대의 고농도발생 확률은 봄과 가을 그리고 겨울에는 유사한 특징을 보였으나, 여름에는 조건부확률이 0%인 권역이 다수 발견되었다. 특히 여름에는 상호조건부확률이 0%인 권역이 총 29쌍으로 권역 간의 연관성이 가장 낮은 계절에 속하였다. 이러한 특징은 고농도 발생권역 수 분포에서도 동일하게 나타나, 고농도일(95일) 중 1~3개의 권역에서 고농도가 발생한 날은 65일(68.4%)로 발생 권역 수가 적은 특징을 보였다. 권역별 분석결과, 수도권과 충청은 봄에 고농도발생 상호연관성이 크게 나타났다으며, 영서와 충청 그리고 경기북부는 타 권역과의 조건부확률이 평균 94.9%, 91.8%, 91.2%로 겨울철 고농도발생에 있어 중요한 권역임을 알 수 있었다. 반면 지리적인 요인에 의해 강원영동과 제주는 전반적으로 다른 권역과의 조건부확률이 상대적으로 낮았으며, 특히 강원영동은 태백산맥에 의해 강원영서와 인접한 권역이지만 발생기준권역에 따라 조건부확률이 상이하였다($P(R_{영서}|R_{영동}) > P(R_{영동}|R_{영서})$). 이처럼 인접한 권역이지만 연관성이 상대적으로 낮은 권역에는 대전과 충북·충남이 해당하였다. 시간적 연속성을 고려하였을 때, 대부분의 권역들은 A권역의 오늘 고농도발생 유무가 A권역의 내일 고농도발생확률에 미치는 영향이 큰 것으로 나타났으나, 경기남부와 대전 그리고 경북은 낮은 특징을 보였다. 특히 울산은 유일하게 다른 권역의 고농도발생 유무가 예측모형에 사용되지 않는 특징을 보이며 외부의 영향을 적게 받는 것으로 분석되었다. 종합적으로 오늘 고농도발생 유무가 내일의 고농도발생확률에 미치는 영향이 가장 높은 권역으로는 경기북부가 선정되었으며, 부산과 대구 그리고 경남은 가장 적은 것으로 분석되었다.

본 연구에서는 변경된 PM_{2.5} 나쁨 기준에 따른 권역별 고농도발생 연관성을 동시간대와 시간적 연속성을 고려하여 분석하였다. 다양한 통계방법이나 인공지능을 이용한 예측모형 개발에 있어 이러한 권역별 고농도발생 연관성을 고려한다면, 보다 향상된 예측 정확도를 기대할 수 있을 것으로 보이며, 국가대기질예보의 고농도발생 유무 판단근거로 활용할 수

있을 것으로 판단된다. 다만, 현재 국가대기질예보의 고농도등급(나쁨과 매우나쁨)에 맞추어 세분화된 예보등급에 따른 추가 분석이 필요하며, 다양한 기상상황 및 배출분포 등이 추가적으로 고려된다면 보다 종합적인 이해에 바탕을 둔 자료로 활용될 수 있을 것이라 사료된다.

감사의 글

이 논문은 환경부의 재원으로 국립환경과학원의 지원을 받아 수행되었습니다(NIER-RP2019-174).

References

- Do, W.-G., Jung, W.-S. (2017) An Analysis of Similarity between Air Quality Monitoring Stations in Busan using Cluster Analysis, *Journal of Environmental Science International*, 26(8), 927-938, (in Korea with English abstract), <https://doi.org/10.5322/JESI.2017.26.8.927>
- Koo, Y.-S., Yun, H.-Y., Kwon, H.-Y., Yu, S.-H. (2010) A Development of PM₁₀ Forecasting System, *Journal of Korean Society for Atmospheric Environment*, 26(6), 666-682, (in Korea with English abstract), <https://doi.org/10.5572/KOSAE.2010.26.6.666>
- Lee, H.-J., Jeong, Y.M., Kim, S.-T., Lee, W.-S. (2018) Atmospheric Circulation Patterns Associated with Particulate Matter over South Korea and Their Future Projection, *Journal of Climate Change Research*, 9(4), 423-433, (in Korean with English abstract), <https://doi.org/10.15531/kscrcr.2018.9.4.423>
- Lee, T.-J., Park, M.-B., Kim, D.-S. (2019) Time Series Assessment of PM_{2.5} Source Contributions and Classification of Haze Patterns in Seoul, *Journal of Korean Society for Atmospheric Environment*, 35(1), 97-124, (in Korea with English abstract), <https://doi.org/10.5572/KOSAE.2019.35.1.097>
- Lee, Y.-S., Kimm, H.-G., Pakr, J.-S., Kim, H.-K. (2006) Study on Statistical Forecasting Models of PM₁₀ in Pohang Region by the Variable Transformation, *Journal of Korea Society for Atmospheric Environment*, 22(5), 614-626, (in Korean with English abstract).
- Oh, J.W., Lim, T.J. (2017) Regional Analysis of Particulate Matter

Concentration Risk in South Korea, *Journal of the Korean Society of Safety*, 32(5), 157-167, (in Korea with English abstract), <https://doi.org/10.14346/JKO-SOS.2017.32.5.157>

Yeo, M.J., Kim, Y.P. (2019) Trend of the PM₁₀ concentrations and High PM₁₀ Concentrations Cases in Korea, *Journal of Korean Society for Atmospheric Environment*, 35(2), 249-264, (in Korea with English abstract), <https://doi.org/10.5572/KOSAE.2019.35.2.249>

Zhang, Q., Jiang, X., Tong, D., Davis, S.J., Zhao, H., Geng, G., Feng, T., Zheng, B., Lu, Z., Streets, D.G., Ni, R., Brauer, M., van Donkelaar, A., Martin, R.V., Huo, H., Liu, Z., Pan, D., Kan, H., Yan, Y., Lin, J., He, K., Guan, D. (2017) Transboundary Health Impacts of Transported Global Air

Pollution and International Trade, *Nature*, 543, 705-719, <https://doi.org/10.1038/NATURE21712>

Authors Information

김태희 (국립환경과학원 대기질통합예보센터 연구사)
 명지수 (국립환경과학원 대기질통합예보센터 전문위원)
 정미숙 (국립환경과학원 대기질통합예보센터 전문위원)
 이용희 (국립환경과학원 대기질통합예보센터 연구사)
 서인석 (국립환경과학원 대기질통합예보센터 연구관)