

부유분진측정기(PM₁₀) 관측 자료 실시간 품질관리 알고리즘 개발 및 평가

김선영 · 이희춘* · 류상범
국립기상과학원 환경기상연구과

(접수일: 2016년 6월 24일, 수정일: 2016년 8월 9일, 게재확정일: 2016년 9월 13일)

Development and Assessment of Real-Time Quality Control Algorithm for PM₁₀ Data Observed by Continuous Ambient Particulate Monitor

Sunyoung Kim, Hee Choon Lee*, and Sang-Boom Ryoo

Environmental Meteorology Research Division, National Institute of Meteorological Sciences, Jeju, Korea

(Manuscript received 24 June 2016; revised 9 August 2016; accepted 13 September 2016)

Abstract A real-time quality control algorithm for PM₁₀ concentration measured by Continuous Ambient Particulate Monitor (FH62C14, Thermo Fisher Scientific Inc.) has been developed. The quality control algorithm for PM₁₀ data consists of five main procedures. The first step is valid value check. The values should be within the acceptable range limit. Upper (5,000 $\mu\text{g m}^{-3}$) and lower (0 $\mu\text{g m}^{-3}$) values of instrument detectable limit have to be eliminated as being unrealistic. The second step is valid error check. Whenever unusual condition occurs, the instrument will save error code. Value having an error code is eliminated. The third step is persistence check. This step checks on a minimum required variability of data during a certain period. If the PM₁₀ data do not vary over the past 60 minutes by more than the specific limit (0 $\mu\text{g m}^{-3}$) then the current 5-minute value fails the check. The fourth step is time continuity check, which is checked to eliminate gross outlier. The last step is spike check. The spikes in the time series are checked. The outlier detection is based on the double-difference time series, using the median. Flags indicating normal and abnormal are added to the raw data after quality control procedure. The quality control algorithm is applied to PM₁₀ data for Asian dust and non-Asian dust case at Seoul site and dataset for the period 2013~2014 at 26 sites in Korea.

Key words: Quality control algorithm, PM₁₀, error, spike, flag

1. 서 론

기상관측자료는 지역 및 전구 기후의 변동과 변화를 이해하고 확인하는데 중요하게 사용된다. 그러나 관측지점의 이전, 관측장비의 교체, 자료처리 방법의 변화, 도시화 등과 같은 관측 환경의 변화로 관측자

료의 비균질성(inhomogeneity)이 발생할 수 있다 (Easterling and Peterson, 1995). 관측자료의 비균질성은 관측장비가 실제의 값을 측정하는 과정에서 다양한 오차를 발생하게 한다. 이와 같은 측정오차가 발생된 관측자료를 이용하기 위해서는 품질관리 과정이 반드시 필요하다. 품질관리를 수행하지 않은 자료를 사용하여 연구 및 현업에 이용한다면 부정확한 결과를 생산할 수 있다(Ryoo et al., 2006).

품질관리(Quality Control)는 자료의 신뢰성 확보를 위해서 수행하는 품질검사를 포함하는 일련의 행위이다. 사용자가 관측자료를 최적으로 사용할 수 있도록

*Corresponding Author: Hee Choon Lee, Environmental Meteorology Research Division, National Institute of Meteorological Sciences, 33, Seohobuk-ro, Seogwipo-si, Jeju 63568, Korea.
Phone: +82-64-780-6564, Fax: +82-64-738-6512
E-mail: lee.heechoon@korea.kr

Table 1. Summary for quality control algorithm of observation data of reference.

Reference	Range	Step	Consistency	Persistence	Spatial
Fiebrich and Crawford (2001)	○	○	×	○	○
Feng et al. (2004)	○	×	○	×	○
Lanci and Lussana (2009)	○	×	○	×	○
Liljegren et al. (2009)	○	×	○	×	○
Lee et al. (2010)	○	×	○	○	○
KMA (2011)	○	×	○	○	○
Nie et al. (2012)	○	×	○	×	○
Jiang et al. (2013)	○	×	○	○	×
Chae et al. (2014)	○	○	○	○	×

품질검사를 통해 오류자료 검출 및 오차 보정의 과정을 통하여 관측자료의 정확도를 최대한 높일 수 있다 (WMO, 2013; KMA, 2015).

고품질 자료의 확보를 위해 다양한 기상관측자료에 대한 품질관리 연구가 수행되었다(Table 1). 오클라호마 주와 오클라호마 대학이 함께 구축한 Mesonet에서는 기후범위검사, 단계검사, 지속성검사, 공간검사, 동일장비 비교검사 과정이 포함된 5단계 품질관리 시스템을 개발하였다(Fiebrich and Crawford, 2001). Feng et al. (2004)은 중국 전역 726개 기상관측지점의 10개 기상요소를 대상으로 5단계의 품질관리를 실시하였다. Lanci and Lussana (2009)는 이탈리아의 중규모 기상 관측망에 적용할 3단계의 자동화된 자료 품질관리 절차를 개발하였다. Liljegren et al. (2009)은 미국에서 화학물질 유출 시 비상대응 프로그램을 위한 기상자료 품질관리를 실시하였다. 이때 풍속, 풍향, 기온 등 8가지 기상요소에 대해 5단계의 품질관리 알고리즘을 적용하였다. Lee et al. (2010)은 극한값 검사, 내적일치성 검사, 기후범위 검사, 공간분포 검사 등의 4단계로 구성된 품질관리 알고리즘을 개발하고, 부산 기온 자료를 이용하여 검사를 실시하였다. Nie et al. (2012)은 Global Telecommunication System (GTS)에 보고된 강수량 자료를 대상으로 5단계의 품질관리를 실시하여 NOAA CMAP (CPC Merged Analysis of Precipitation) 및 GPCP (Global Precipitation Climatology Project) 자료와 높은 일치도를 보이는 New Global Daily Precipitation (NGDP) 자료를 생산하였다. Chae et al. (2014)은 WISE 복합기상센서에서 관측·수집되는 자료에 적용할 수 있는 6단계의 품질관리시스템을 개발하였다. 기상청에서는 2005년부터 다양한 기상관측자료를 대상으로 19가지 방법의 실시간, 준실시간, 비실시간 품질관리를 실시하고 있다(KMA, 2011). 위에 언급된 일반적인 기상요소를 대상으로 하는 품질검사뿐만 아니라 대기 오염물질 자료를 대상으로 하는 품질검사에 대한 연구도 수행되었다. Jiang et al. (2013)은 시간평균 지상 PM₁₀ 자료를 3차원 변분자료동화에 사

용하기 위해서 범위검사와 시간연속성검사로 이루어진 두 단계의 품질관리를 수행하였다. Shin et al. (2014)은 PM₁₀, O₃, NO₂, CO, SO₂ 등과 같은 대기오염물질 자료에 적용할 수 있는 5단계의 품질검사 방법을 개발하였다.

황사 및 대기환경에 대한 국민의 관심이 증대됨에 따라 기상청은 웹사이트를 통해 시간평균 PM₁₀ 농도 자료를 제공하고 있다. PM₁₀ 자료는 일반적인 기상자료와 달리 급격한 값의 변동이 발생할 수 있기 때문에 다양한 기상관측자료에 적용하고 있는 기상청의 관측자료 품질관리 방법의 적용에는 한계가 있다. 따라서 PM₁₀ 관측자료는 국민에게 제공되기 전에 품질관리를 통한 신뢰성 확보가 필수적으로 수행되어야 한다. Shin et al. (2014)은 PM₁₀을 비롯한 대기오염물질 자료에 적용할 수 있는 5단계의 품질검사 방법을 개발하였다. 이 방법은 관측 값이 결측 데이터인지 여부, 관측 값이 최소/최대 범위 내에 있는지 여부, 관측 값이 최대변화율 범위 내에 있는지 여부, 관측 값이 연속적으로 지속되는지 여부, 관측 값이 잔차 범위 이내인지 여부의 총 5단계로 이루어져 있다. Shin et al. (2014)이 개발한 방법은 수집된 과거 자료가 일정기간 존재할 경우에 적용할 수 있는 것이다. 실시간으로 수집되어 예보 및 자료 제공 등 대국민 서비스를 위한 기상청 PM₁₀ 자료의 품질검사에 적용하기에는 어려움이 있다. 또한 PM₁₀ 자료의 품질검사에서는 일반적인 기상자료와는 다르게 연속적이지 않은 급격한 값의 변동이 발생한다. 이와 같은 PM₁₀ 자료의 특성이 반드시 고려된 품질검사 방법이 개발되어야 한다.

따라서 본 연구에서는 PM₁₀ 자료의 특성을 고려한 기존의 품질관리 기법을 선택적으로 적용, 새로운 기법의 개발을 통해 기상청의 부유분진측정기 관측 자료에 적용할 수 있는 실시간 품질관리 알고리즘을 개발하고자 한다. 또한 개발된 품질관리 알고리즘을 황사와 비황사 사례 및 전국 26개 지점의 2년(2013~2014년) 동안의 자료에 적용한 후 품질관리 결과를 평가하고자 한다.

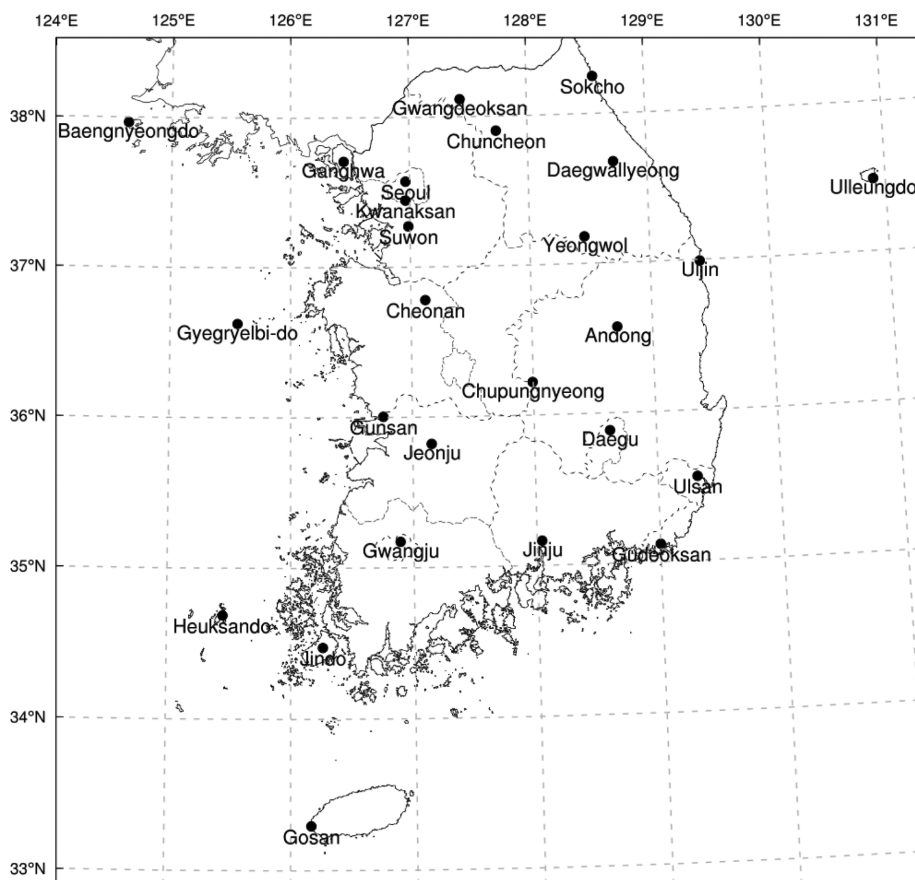


Fig. 1. The location of selected PM₁₀ station for this study.

2. 연구자료

기상청은 2003년 7개 지점을 시작으로 현재까지 전국 27개 지점에 부유분진측정기(FH62C14, Thermo Fisher Scientific Inc.)를 설치하여 매 5분마다 대기 중 PM₁₀ 농도를 관측하고 있다. PM₁₀은 입자의 지름이 10 μm 이하인 입자상 물질의 질량농도를 일컫는 것으로 현재 대부분의 국가에서 PM₁₀의 질량농도를 기초로 입자상 물질의 대기환경기준을 설정하고 있다 (Kim et al., 2014). 기상청에서 운용중인 부유분진측정기의 측정원리는 다음과 같다. 부유분진측정기의 인렛을 통해 대기 중의 공기가 흡입되어 기기 내에 장착된 필터에 직경이 10 μm 이하인 에어로졸 입자가 침적된다. 에어로졸이 포집된 필터 여지에 탄소동위원소 ¹⁴C에서 방출되는 베타선을 쏘아 감쇄된 베타선을 검출기로 측정하여 단위 체적 당 PM₁₀ 농도(μg m⁻³)를 산출한다(National Institute of Meteorological Sciences, 2014).

본 연구에서는 2013~2014년의 2년간 부유분진측

정기 관측자료를 보유하고 있는 전국 26개 지점을 대상으로 품질관리를 실시하였다(Fig. 1, Table 2). 또한 서울지점의 PM₁₀ 농도가 높았던 황사와 비황사 사례를 각각 선정한 후 품질관리를 실시하였다. 선정된 황사 사례는 2014년 3월 18~19일까지의 2일간이며, 황사가 발생하지 않았지만 PM₁₀ 농도가 높았던 비황사 사례는 2008년 9월 8일과 2009년 5월 15일이다.

품질검사 결과 평가를 위해 수동품질관리(Manual Quality Control: MQC) 자료를 이용하였다. 부유분진측정기의 기기이상 또는 원인 불명의 이상 값이 발생한 경우에는 관측자에 의해 실시간으로 수동품질관리가 실시된다. 원인 불명의 이상 값이 발생하였으며, 인근 환경부 관측 자료보다 급격한 농도 상승이 확인된 경우, 눈이 오거나 그친 직후인 경우, 강수 현상이 있는 경우, 황사 예보가 있었으나 황사 영향이 없음에도 높은 농도가 나타난 경우에 수동품질관리가 실시된다. 수동품질관리를 수행한 자료에는 품질정보가 추가되어 사용자에게 제공된다.

Table 2. Information for the observation by stations (as of the end of 2014).

Station	Altitude (m)	Time for beginning of observation (year.month.day)	Time for instrument replacement ¹ (year.month.day)	Time for station relocation ² (year.month.day)
Sokcho	18.06	2006.12.21	-	-
Gwangdeoksan	1,066	2004.05.29	2014.12.29	-
Daegwallyeong	772.57	2005.06.14	-	-
Chuncheon	77.71	2006.12.21	-	-
Baengnyeongdo	144.86	2003.04.16	2013.12.06	-
Seoul	85.8	2008.05.13	-	-
Ulleungdo	222.8	2005.06.14	-	-
Gwanaksan	622.38	2003.04.16	2013.12.26	-
Suwon	34.06	2008.05.13	-	-
Yeongwol	240.6	2007.02.26	-	2011.09.15
Uljin	50	2005.06.29	2005.11.20	-
Chupungnyeong	244.73	2004.07.15	2014.12.29	-
Andong	140.1	2007.02.26	-	-
Gunsan	23.2	2003.04.16	2013.12.06	-
Daegu	64.08	2007.02.26	-	2014.12.05
Jeonju	53.4	2008.05.13	-	-
Ulsan	34.57	2008.05.13	-	-
Gwangju	72.38	2004.07.15	2014.12.29	-
Gudeoksan	518.46	2005.06.29	-	-
Heuksando	76.49	2003.04.16	2013.12.06	-
Jindo	476.47	2008.05.13	-	-
Gosan	74.29	2003.04.16	2013.12.26	-
Jinju	30.21	2007.02.26	-	-
Ganghwa	47.01	2003.04.16	2013.12.06	-
Gyegryelbi-do	103.56	2004.12.24	2014.12.29	-
Cheonan	81.5	2004.07.15	2014.12.29	-
Yeonpyeong-do	20	2014.12.29	-	-

¹Replacement of same type instrument.²Relocated within observation field of same station.

3. 품질관리 절차(QC procedure)

부유분진측정기 관측자료 실시간 품질관리 절차는 총 5단계로 관측범위한계 검사(Valid value check), 기기에러 검사(Error code check), 지속성 검사(Persistence check), 시간 연속성 검사(Time continuity check), 튀는 값 검사(Spike check)의 절차로 구성된다(Fig. 2). 모든 단계를 통과한 경우 정상 값으로 판단하며, 한 단계라도 통과하지 못하는 경우 오류 가능성이 있는 값이라고 판단하였다. 품질관리 단계 중 어느 단계에서든 관측 값이 오류 판정을 받으면 그 이후 과정에서는 검사를 중단하고 해당 시각의 관측 값에 대해서는 품질관리를 종료하였다.

첫 번째 단계인 관측범위한계 검사는 관측기기의 관측범위를 벗어나는지 여부를 검사하는 것이다. 부유분진측정기는 0.1~5,000 µg m⁻³ 범위까지 관측할 수

있다(Thermo Fisher Scientific, 2010). 이 범위를 벗어나는 값이 관측되는 경우 오류가 있다고 판단하였다.

두 번째 단계인 기기에러 검사는 관측기기가 기록하는 에러 및 상태코드가 정상인지 여부를 검사하는 것이다. 부유분진측정기의 에러 및 상태코드는 총 6자리로 구성되어있다. Table 3에 제시된 것과 같이 앞 두 자리는 error status, 가운데 두 자리는 warning status, 마지막 두 자리는 operating status이다. 각 상태 별로 기록된 값이 '00'인 경우는 정상이며, 그 외의 값을 기록하는 경우는 기기에 문제가 있는 경우이다. 따라서 세 가지 상태를 종합하여 나타낸 6자리 코드가 '000000'이면 그 값은 정상으로 판단한다. 그러나 Table 3에서 볼 수 있는 것과 같이 각 상태 별로 사용되지 않는 코드가 있다. error status 코드 중 '40', '80', warning status 코드 중 '04', '10', '80', operating status 코드 중 '04'는 사용되지 않는다. 따라서 위에

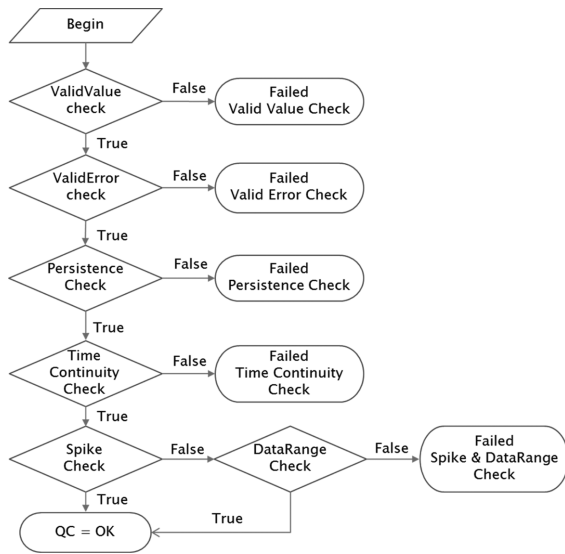


Fig. 2. Flowchart of quality control for PM₁₀ data.

언급된 코드의 조합으로 이루어진 경우도 정상적인 값으로 판단하였다. 에러 및 상태코드가 위의 예외적인 상태를 제외하고 나타난 경우는 비정상적인 값이라고 판단하여 의심자료로 분류하였다.

세 번째 단계인 지속성 검사는 일정기간 동안 값의 변화가 있는지 여부를 검사하는 것이다. Zahumenský (2004)는 기상 관측요소 별로 과거 기준시간에 대하

여 최소변동량을 임계치로 설정하여 품질검사의 판단 기준으로 사용하도록 권고하고 있다. 본 연구에서도 Zahumenský (2004)의 권고를 준용하여 과거 1시간 동안 PM₁₀ 농도의 변동량이 0인 경우 오류가 있다고 판단하여 의심자료로 분류하였다.

네 번째 단계인 시간 연속성 검사는 관측 자료의 이례치를 제거하기 위해 시간의 연속성을 검사하는 것이다. 이 단계에서는 Jiang et al. (2013)에서 사용한 품질관리 방법을 적용하였다[식(1)~(2)]. A가 식(2)에서 구해진 $m(t)$ 보다 작거나 같으면 그 값은 정상이라고 판단하였다. 자료의 정상여부를 판단하는 기준은 Jiang et al. (2013)이 제시한 것을 사용하였다.

$$A = \min|PM_{10}(t) - PM_{10}(t \pm 1)| \quad (1)$$

$$m(t) = 50 + 0.15PM_{10}(t) \quad (2)$$

자료의 불확실성을 증가시키는 중요한 원인 중의 하나는 여러가지 원인으로 인해 발생하는 오류 값이다. 다섯 번째 단계에서는 Papale et al. (2006)가 사용한 방법과 같이 중위절대편차(Median Of Absolute Deviation; MAD)를 이용하여 뒤는 값을 검사하였다. 3개의 연속적인 PM₁₀ 자료의 차이를 이용하여 식(3)과 같이 d 를 구한 후 식(4)의 조건을 만족하면 정상 값으로 판단하였다. 여기서 Md 는 식(3)에서 구한 d 의 중간 값이며, MAD는 식(5)의 방법으로 정의된다.

$$d = (PM_{10}(t) - PM_{10}(t - 1)) - (PM_{10}(t + 1) - PM_{10}(t)) \quad (3)$$

Table 3. Status code of PM₁₀ instrument.

Error status		Warning status		Operating status	
00	Normal operation	00	Normal operation	00	Normal operation
01	Sum status of the data and program memory	01	Maximum pump power	01	Mechanical filter change
02	Sum status of sampling and measuring system	02	Mass limit is exceeded; filter change will follow	02	Plateau check and zero adjustment of the mass (R ⁰)
04	Sum status of pressure sensors	04	Not used	04	Not used
08	Sum status of air flow regulation	08	Relative Humidity > (RH Set point + 10%)	08	Pump switched off
10	Sum status of the sample heater	10	Not used	10	Manual operation (off-line, keyboard enabled)
20	Sum status of temperature measurement	20	Compensation of the mass signal $\geq \pm 5\%$	20	Calibration enabled
40	Reserved	40	Two consecutive plateau checks are wrong during filter change: Rref/R1 < 30% or > 70%	80	Power supply switched on
80	Not used	80	Not used		

Table 4. 2 × 2 contingency table.

QC \ MQC	Abnormal	Normal
Abnormal	Hit (a)	False alarm (b)
Normal	Miss (c)	Correct negative (d)

$$d \geq Md - (z \times MAD) \text{ 또는 } d \leq Md + (z \times MAD) \quad (4)$$

$$MAD = \text{median}(|d - Md|) \quad (5)$$

Papale et al. (2006)은 자료 자체가 주는 영향과 방법의 민감도를 추정하기 위해 세 가지의 서로 다른 z 값(4.0, 5.5, 7.0)을 연구에 적용하였다. 본 연구에서는 z값을 7.0으로 설정하였으며, 이는 수 차례의 실험을 통해 경험적으로 산출한 것이다.

위의 다섯 번째 단계는 검사를 하고자 하는 시각의 PM₁₀ 농도가 낮을 경우 전후 값에서 약간의 변화만 있어도 오류로 판단할 수 있다. 이러한 문제점을 보완하기 위하여 다섯 번째 단계에서 오류 값으로 판정한 경우 자료 범위 검사를 실시하였다. Jiang et al. (2013)의 연구결과에서 제시한 시간 연속성 기준을 참고하였으며, 검사를 하고자 하는 시각의 값이 이전 1 시간 평균 PM₁₀ 농도의 ±15% 이내에 포함되면 그 값은 정상으로 판단하였다.

위의 5단계의 품질관리 절차를 수행하여 산출된 결과의 정확도를 평가하기 위해 MQC 자료와 비교하였다. 품질관리 결과가 오류 값이고 MQC 자료가 비정상 값인 경우는 Hit (a), 품질관리 결과는 오류 값이었지만, MQC 자료가 정상 값인 경우는 False alarm (b)으로 구분하였다. 품질관리 결과는 정상 값이었으나, MQC 자료가 비정상 값인 경우는 Miss (c), 품질관리 결과와 MQC 자료가 모두 정상 값인 경우는 Correct negative (d)로 구분하였다. Table 4과 같이

2 × 2 분할표를 작성한 후, 식(6)~(9)의 방법으로 Skill score를 구하여 품질관리 결과의 정확도를 평가하였다. ACC는 예측의 정확도, POD는 탐지확률, FAR은 허위경보율, CSI는 사건 발생의 정확도를 의미한다. ACC가 높고 FAR이 낮을수록 본 연구에서 제시한 품질관리 방법이 적절하다고 할 수 있다.

$$ACC = \frac{a + b}{a + b + c + d} \quad (6)$$

$$POD = \frac{a}{a + c} \quad (7)$$

$$CSI = \frac{a}{a + b + c} \quad (8)$$

$$FAR = \frac{b}{a + b} \quad (9)$$

4. 품질관리 결과

서울 지점을 대상으로 황사 사례와 황사가 관측되지 않는 않았지만 비정상적으로 높은 PM₁₀ 농도가 나타나는 비황사 사례를 선정하여 품질관리를 실시하였다. 품질관리에 적용한 황사 사례는 2014년 3월 18~19일의 2일간이며, 비황사 사례는 2008년 9월 8일과 2009년 5월 15일이다. 또한 전국 27개 지점 중 최근 2년(2013~2014년) 간의 자료를 보유하고 있는 26개 지점을 대상으로 품질관리를 적용한 후 그 결과를 평가하였다.

4.1 황사 사례(2014년 3월 18~19일)

서울에서 관측된 황사사례를 선정하여 품질 검사를 실시하였다. 고비/내몽골에서 발원한 황사는 2014년 3월 18~19일의 이틀에 걸쳐 전국적으로 영향을 주었다. 2014년 3월 18일 10시 15분에 서울에서 황사가 관측되

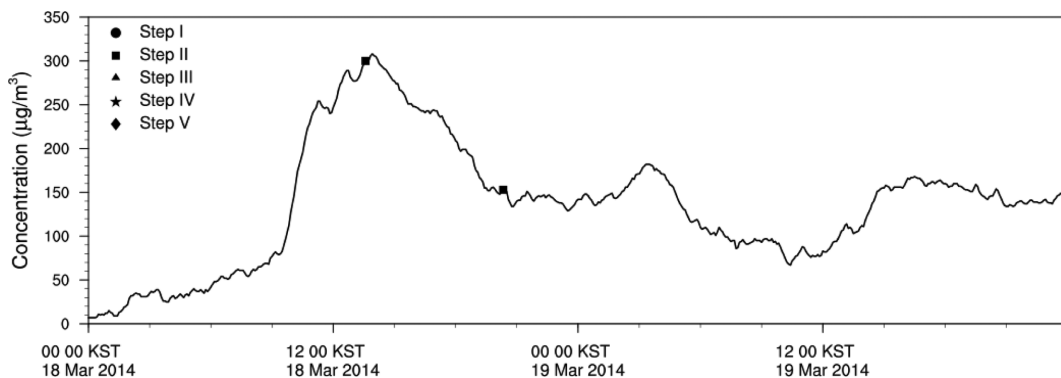


Fig. 3. Change of concentration for PM₁₀ during Asian Dust event period in 18~19 March 2014. Line indicates normal data and markers indicate flagged data.

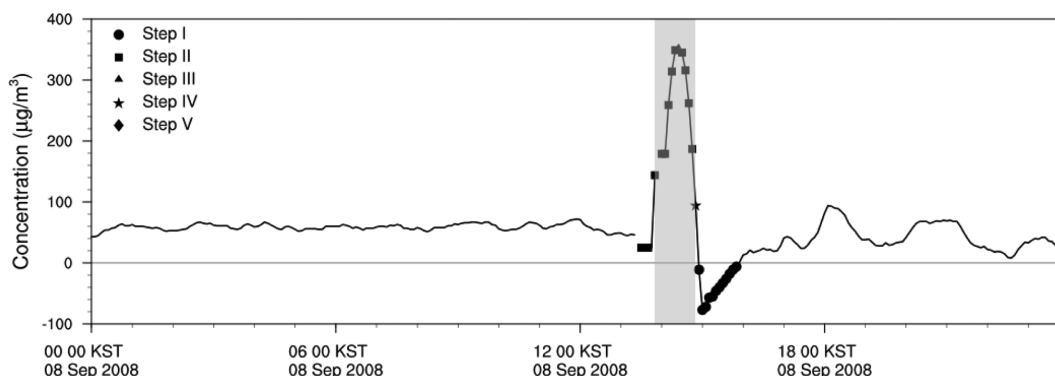


Fig. 4. Change of concentration for PM₁₀ during non-Asian Dust event period on 08 September 2008. Line indicates normal data and markers indicate flagged data. MQC time indicate shaded.

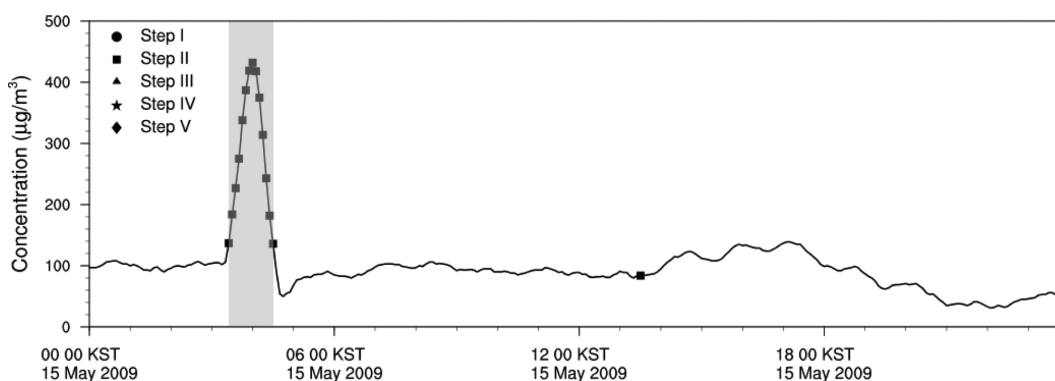


Fig. 5. Change of concentration for PM₁₀ during non-Asian Dust event period on 15 May 2009. Line indicates normal data and markers indicate flagged data. MQC time indicate shaded.

었으며, 2014년 3월 19일 05시 30분까지 지속되었다. 3월 18일 00시~10시까지의 시간평균 PM₁₀ 농도는 10~96 µg m⁻³이었으나 11시 평균 PM₁₀ 농도는 200 µg m⁻³로 급격하게 증가하여 18시까지 200 µg m⁻³ 이상을 상회하는 값을 기록하였다(Fig. 3). MQC 자료를 검토한 결과 PM₁₀ 농도가 이전시간에 비해 급격하게 증가하였다. 이 경우는 황사 유입으로 인한 고농도의 PM₁₀이 관측된 것이기 때문에 MQC가 이루어지지 않았다.

이 사례를 대상으로 5단계의 알고리즘을 적용하여 품질검사를 실시하였다. 품질검사 결과 정상적인 값은 검은 실선으로 오류가 있는 값은 오류 종류에 따라 모양을 달리하여 나타내었다(Fig. 3). 사례 기간의 자료 864회 중 정상 값으로 걸러진 것은 862회로 품질관리 결과 대부분의 값이 정상 값으로 판정되었다. 비정상 값으로 걸러진 값은 두 번째 단계인 기기예러 검사에서 걸러졌다. 기기예러 검사 결과 오류 값으로 판정한 13시 35분과 20시 20분의 경우 황사가 관측된 시간에 기록되었으며, 각각 '000049', '000042'였

다. 그러나 이는 실제로는 나타날 수 없는 코드로 이때의 값은 정상이라고 할 수 있다.

4.2 비황사 사례

4.2.1 2008년 9월 8일

2008년 9월 8일 00시~14시까지의 시간평균 PM₁₀ 농도는 53~65 µg m⁻³를 유지하고 있었으나, 13시 50분~14시 50분까지의 PM₁₀ 농도는 95~358 µg m⁻³로 이전시간의 평균값에 비해 높게 나타났다. 또한 2008년 9월 8일 14시 55분~15시 50분까지 1시간 동안 0 이하의 PM₁₀ 값이 기록되었다(Fig. 4). 2008년 9월 8일의 경우 서울에서는 황사가 관측되지 않았지만 이때의 MQC 자료를 검토한 결과 13시 50분~14시 45분까지 약 1시간 동안 MQC가 수행되었음을 확인할 수 있었다.

Figure 4는 2008년 9월 8일 비황사 사례에 대해 품질검사 알고리즘을 적용한 결과이다. 전체 288개의 PM₁₀ 값 중 오류로 걸러진 자료는 총 28개였으며, 첫 번째 단계인 관측범위 한계 검사에서 15개, 기기예러

Table 5. Frequency and ratio of flagged data for PM₁₀ by step.

Case	18~19 March 2014		8 September 2008		15 May 2009	
	Frequency	Ratio (%)	Frequency	Ratio (%)	Frequency	Ratio (%)
I	0	0.0	15	5.2	0	0.0
II	2	0.2	12	4.2	15	5.2
III	0	0.0	0	0.0	0	0.0
IV	0	0.0	1	0.4	0	0.0
V	0	0.0	0	0.0	0	0.0

Table 6. Skill score of quality control by observation stations.

Station	Hit	False alarm	Miss	Correct negative	ACC	POD	FAR	CSI
Sokcho	1,919	1,120	57	201,673	0.99	0.97	0.37	0.62
Gwangdeoksan	16,901	6,301	135	181,911	0.97	0.99	0.27	0.72
Daegwallyeong	6,998	238	99	195,557	1.00	0.99	0.03	0.95
Chuncheon	3,900	8,633	0	184,364	0.96	1.00	0.69	0.31
Baengnyeongdo	5,977	1,646	62	194,041	0.99	0.99	0.22	0.78
Seoul	3,023	832	40	200,208	1.00	0.99	0.22	0.78
Ulleungdo	18,525	2,708	47	179,630	0.99	1.00	0.13	0.87
Gwanaksan	9,454	27,661	164	166,138	0.86	0.98	0.75	0.25
Suwon	1,917	1,661	33	203,409	0.99	0.98	0.46	0.53
Yeongwol	2,688	604	59	190,700	1.00	0.98	0.18	0.80
Ulsan	2,980	238	76	204,312	1.00	0.98	0.07	0.90
Chupungnyeong	5,355	669	31	197,652	1.00	0.99	0.11	0.88
Andong	1,613	325	93	205,500	1.00	0.95	0.17	0.79
Gunsan	1,675	388	199	202,061	1.00	0.89	0.19	0.74
Daegu	3,499	14,734	136	188,111	0.93	0.96	0.81	0.19
Jeonju	2,327	9,754	28	182,187	0.95	0.99	0.81	0.19
Ulsan	2,192	2,313	27	195,418	0.99	0.99	0.51	0.48
Gwangju	7,574	1,469	29	188,473	0.99	1.00	0.16	0.83
Gudeoksan	17,405	2,173	307	175,772	0.99	0.98	0.11	0.88
Heuksando	3,731	713	33	198,628	1.00	0.99	0.16	0.83
Gosan	8,574	1,113	153	190,880	0.99	0.98	0.11	0.87
Jinju	2,955	1,170	63	201,874	0.99	0.98	0.28	0.71
Ganghwa	2,415	983	246	201,144	0.99	0.91	0.29	0.66
Gyegryelbi-do	6,648	5,809	61	184,508	0.97	0.99	0.47	0.53
Cheonan	4,428	12,365	30	184,120	0.94	0.99	0.74	0.26
Jindo	7,122	6,239	24	186,522	0.97	1.00	0.47	0.53
Average	5,838	4,302	86	191,723	0.98	0.98	0.34	0.65

검사에서 12개, 시간 연속성 검사에서 1개가 검출되었다. 부유분진측정기로는 관측이 불가능한 $0 \mu\text{g m}^{-3}$ 이하의 값을 기록한 14시 55분~15시 50분까지와 관측이 이루어지지 않아 값이 기록되지 않은 경우 관측 범위한계 검사에서 오류 값으로 검출되었다. 이전 시간에 비해 높은 PM₁₀ 농도를 보이는 13시 50분~14시 50분까지의 기간 동안 기기예러 검사에서 오류 값이 검출되었다. 14시 50분의 경우 이전시각의 PM₁₀ 농도에 비해 $92 \mu\text{g m}^{-3}$ 이상 증가하여 4번째 단계인 시간 연속성 검사에서 오류 값으로 검출되었다.

4.2.2 2009년 5월 15일

앞의 사례와 마찬가지로 황사가 관측되지 않았던 2009년 5월 15일의 경우 03시 30분~04시 25분까지 약 1시간에 걸쳐 $184\sim432 \mu\text{g m}^{-3}$ 의 높은 PM₁₀ 값이 기록되었다. 2009년 5월 15일 00시~03시까지의 시간 평균 PM₁₀ 농도는 $95\sim103 \mu\text{g m}^{-3}$ 를 유지하고 있었으나, 03시 평균 PM₁₀ 농도는 $235 \mu\text{g m}^{-3}$ 로 이전 시간에 비해 2배 정도 높은 값을 보이고 있다(Fig. 5). 이때의 MQC 자료를 검토한 결과 03시 25분~04시 30분까지 약 1시간 동안 MQC가 수행되었음을 확인할

Table 7. Ratio of flag in case of false alarm by observation stations.

Station	Step I	Step II	Step III	Step IV	Step V
Sokcho	0.1	95.5	0.0	0.3	4.1
Gwangdeoksan	0.0	98.3	0.0	0.0	1.7
Daegwallyeong	0.0	68.5	0.0	0.0	31.5
Chuncheon	0.0	99.4	0.0	0.0	0.5
Baengnyeongdo	0.0	61.9	0.0	0.4	37.7
Seoul	0.0	91.8	0.0	0.1	8.1
Ulleungdo	0.1	91.7	0.0	0.3	7.9
Gwanaksan	0.0	99.6	0.0	0.1	0.3
Suwon	0.0	94.3	0.0	0.2	5.5
Yeongwol	0.0	74.3	0.0	0.0	25.7
Ulsan	0.8	56.3	0.0	0.0	42.9
Chupungnyeong	0.0	92.2	0.0	0.0	7.8
Andong	0.0	84.6	0.0	0.0	15.4
Gunsan	0.0	84.5	0.0	0.0	15.5
Daegu	0.0	99.4	0.0	0.0	0.6
Jeonju	0.0	99.5	0.0	0.0	0.5
Ulsan	0.0	97.8	0.0	0.1	2.1
Gwangju	0.0	94.3	0.0	0.1	5.6
Gudeoksan	0.0	93.0	0.0	0.5	6.5
Heuksando	0.0	84.0	0.0	0.4	15.6
Gosan	0.0	99.2	0.0	0.0	0.8
Jinju	0.0	88.0	0.0	0.3	11.8
Ganghwa	0.1	63.4	0.0	0.1	36.4
Gyegryelbi-do	0.0	94.9	0.0	0.0	5.1
Cheonan	0.0	99.7	0.0	0.0	0.3
Jindo	0.0	99.2	0.0	0.0	0.8

수 있었다. 5단계의 알고리즘을 적용하여 품질검사를 실시하였으며, 품질검사 결과 정상적인 PM₁₀ 값은 검은 실선으로, 오류가 있는 값은 오류 종류에 따라 모양을 달리하여 나타내었다(Fig. 5).

오류로 걸러진 자료는 전체 288개 중 총 15개로 5.2%를 차지한다. 오류로 걸러진 값을 검사 단계별로 구분하여 Table 5에 나타내었다. 이전 시각에 비해 비정상적으로 높은 PM₁₀ 농도를 보이는 03시 25분~04시 30분까지의 약 1시간동안 기기예러 검사에서 오류 값으로 검출되었다. 또한 전후 시각의 값과 비교하여 정상적인 값으로 판단되는 13시 30분에도 기기예러 검사에서 오류 값으로 검출하였다.

4.3 2013~2014년 PM10 자료에 대한 품질관리

전국 26개 지점의 2013~2014년의 2년의 기간을 대상으로 5단계의 알고리즘을 적용하여 품질검사를 실시하였다. 품질검사 결과를 MQC 자료와 비교하여 skill score를 구한 후 Table 6에 제시하였다.

검사를 실시한 26개 지점마다 결측값 등의 차이는 있지만 대략 약 20만개의 관측자료에 대해 검사가 이

루어졌다. 품질검사 결과 오류로 판단하였으나, MQC 자료에서는 정상인 경우를 나타내는 False alarm의 횟수는 238~27,661회로 지점마다 차이가 있었다. False alarm의 대부분은 품질관리 결과 기기예러 검사에서 오류로 걸러졌으나 MQC 자료에서는 정상인 경우였다. 기기예러 검사로 인한 오류로 False alarm을 기록한 비율은 전국 26개 지점 평균 88.7%이다. False alarm 빈도가 27,661회로 가장 높았던 관악산의 경우 99.6%인 27,550회가 기기예러로 인한 것이었다(Table 7). 이런 경우의 PM₁₀ 농도는 검사 시간 전후의 값과 큰 차이가 없었다.

품질검사 결과 정상으로 판단하였지만 MQC 자료에서는 오류인 경우를 나타내는 miss의 횟수가 0~307회로 각 지점마다 차이가 있었다. miss 횟수가 307회로 가장 많은 구덕산의 MQC 자료를 확인한 결과 관측장비 고장, 원인 불명 등으로 MQC가 이루어진 경우가 대부분이었다. 이 때의 PM₁₀ 농도는 검사 전후 시각의 값과 큰 차이는 없었으나 MQC가 수행되어 오류로 기록된 경우였다.

ACC는 검사 결과의 정확도를 의미하는 것으로 검

사를 실시한 전국 26개 지점에서 0.86~1.00의 값을 나타내어, 높은 정확도를 보이고 있다. 이는 품질검사 결과 오류로 판단한 값이 MQC 자료에서도 오류 값으로 나타난 경우로 품질검사 알고리즘의 결과가 비교적 정확하다고 할 수 있다.

CSI는 품질검사 결과 오류로 판단한 값을 오류 값과 관련된 총 수로 나눈 것이다. 오류 값과 관련 없는 것(d)를 ACC에서 제외하고 계산한 것이다(Table 4 참조). CSI 또한 0과 1 사이의 값을 가지며 그 값이 1에 가까울수록 정확도가 높다. CSI와 ACC를 동시에 고려하여 품질검사 결과의 정확도를 판단할 수 있다. CSI는 0.19~0.95의 값을 가지며 전 지점 평균 0.56의 값을 보이고 있다.

POD는 실제 오류 값을 품질검사 결과 오류 값으로 판정한 것이 어느 정도인지를 나타내는 것을 의미한다. POD 또한 0과 1 사이의 값을 가지며 그 값이 1에 가까울수록 정확도가 높다. PM₁₀ 자료 품질검사 결과 전 지점에서 0.89 이상의 높은 POD 값을 보이고 있다.

FAR은 허위경보율로 품질검사 결과 PM₁₀ 값이 오류라고 판단하였지만 실제로는 정상 값인 비율을 나타낸다. FAR은 0과 1 사이의 값을 가지며 작을수록 높은 정확도를 보인다. FAR은 지점에 따라 0.03~0.81을 나타내며 전 지점 평균 0.34이다. 대관령의 FAR은 0.03으로 26개 지점 중 가장 낮고, 대구와 전주에서 0.81로 가장 높다. 기기에러로 인해 높은 빈도의 False alarm이 기록된 경우 FAR도 높게 나타났다. 실제로 대구의 경우 오류로 걸러진 자료의 97.8%가 기기에러로 인한 오류였으며, 전주의 경우도 97%를 차지하였다. FAR 값이 크게 나타나는 대구와 전주의 경우 기기에러 검사를 제외한 4단계의 알고리즘을 적용하여 품질관리를 실시하였다. 그 결과 대구와 전주의 FAR은 각각 0.09, 0.12로 크게 감소하였다. 관측자료의 품질에 영향을 미칠 수 있는 관측장비의 상태를 점검할 수 있는 기기에러 검사는 품질관리 과정에서 필수적이다. 그러나 PM₁₀ 값은 정상범위에 속하지만 이때의 기기에러 코드는 비정상인 경우가 있다. 이 때 기기에러 검사를 거치면 PM₁₀은 오류 값으로 판정되며 그 비율은 평균 88.7%로 적지 않은 수치이다. 따라서 위의 사례와 같이 품질관리 결과의 정확도에 영향을 미칠 수 있는 기기에러 검사에 대한 추가적인 논의가 필요하다고 판단된다.

5. 요약 및 결론

본 연구에서는 기상청에서 관측하고 있는 PM₁₀ 자료에 적용할 수 있는 품질관리 알고리즘을 개발하였다. 품질관리 알고리즘은 관측범위한계 검사, 기기에러 검사, 지속성 검사, 시간 연속성 검사, 튀는 값 검

사의 총 5단계로 구성되어 있다. 품질관리 단계 중 어느 단계에서든 관측 값이 오류 판정을 받으면 그 이후 과정에서는 검사를 중단하며 해당 관측 값에 대해서는 품질관리를 종료하고 오류로 판단하였다.

5단계의 품질관리 알고리즘을 서울지점의 황사 및 비황사 사례를 선정하여 적용하였다. 또한 품질관리 알고리즘의 정확도를 판단하기 위하여 2013~2014년의 2년간의 부유분진측정기 관측자료를 보유하고 있는 전국 26개 지점을 대상으로 품질검사를 실시하였다. 그 결과는 다음과 같다. 서울 지점의 황사 사례 기간 동안의 높은 PM₁₀ 농도는 정상 값으로 판단하였으나, 비황사 사례 기간의 비정상적으로 높은 PM₁₀ 농도는 오류 값으로 판단하였다. 전국 26개 지점의 ACC는 평균 0.98로 품질관리의 정확도가 비교적 높았다. FAR은 지점에 따라 0.03~0.81로 차이가 있었지만 전 지점 평균 0.34의 값을 나타내고 있다. 기기에러로 인해 높은 빈도의 False alarm이 기록된 경우 FAR도 높게 나타났다. 따라서 품질관리 결과의 정확도에 영향을 미칠 수 있는 기기에러 검사에 대한 추가적인 논의가 필요하다고 판단된다.

본 연구 결과는 고품질의 PM₁₀ 자료 생산을 위한 기초자료로 활용될 수 있을 것이다. 본 연구를 통해 PM₁₀ 자료의 오류 및 의심자료를 검출하고 오류 자료 판정을 위한 효율적이고 객관적인 방법을 제시할 수 있다. 또한 관측자의 수동품질관리 수행과정에서 참고할 수 있는 가이드를 제공할 수 있을 것으로 판단된다.

감사의 글

본 연구는 국립기상과학원 주요사업 '기상업무지원 기술개발연구(NIMS-2016-3100)'의 지원으로 수행되었습니다.

REFERENCES

- Chae, J.-H., M.-S. Park, and Y.-J. Choi, 2014: The WISE quality control system for integrated meteorological sensor data. *Atmosphere*, **24**, 445-456 (in Korean with English abstract).
- Easterling, D. R., and T. C. Peterson, 1995: A new method for detecting undocumented discontinuities in climatological time series. *Int. J. Climatol.*, **15**, 369-377.
- Feng, S., Q. Hu, and W. Qian, 2004: Quality control of daily meteorological data in China, 1951-2000: A new dataset. *Int. J. Climatol.*, **24**, 853-870.
- Fiebrich, C. A., and K. C. Crawford, 2001: The impact of unique meteorological phenomena detected by Oklahoma Mesonet and ARS micronet on automated qual-

- ity control. *Bull. Amer. Meteor. Soc.*, **82**, 2173-2187.
- Jiang, Z., Z. Liu, T. Wang, C. S. Schwartz, H.-C. Lin, and F. Jiang, 2013: Probing into the impact of 3DVAR assimilation of surface PM10 observations over China using process analysis. *J. Geophys. Res.*, **118**, 6738-6749.
- Kim, K.-J., S.-H. Lee, D.-R. Hyeon, H.-J. Ko, W.-H. Kim, and C. H. Kang, 2014: Composition comparison of PM10 and PM2.5 fine particulate matter for Asian dust and haze events of 2010~2011 at Gosan site in Jeju Island. *Anal. Sci. Tech.*, **17**, 1-10 (in Korean with English abstract).
- KMA, 2011: *Manual for Quality Control of National Climate Data*. KMA, 57 pp.
- _____, 2015: *The Guidelines for Integrated Data Quality Management of National Climate Data*. KMA, 83 pp.
- Lanci, M., and C. Lussana, 2009: An automated data quality control procedure applied to a mesoscale meteorological network. *Proc., 9th EMS annual meeting/9th European Conf. Applications of Meteorol.*, Toulouse, 335 pp.
- Lee, J.-D., Y. Choi, and C. Park, 2010: Developments of quality control algorithms for Korean temperature data. *J. Climate Res.*, **5**, 162-174 (in Korean with English abstract).
- Liljegren, J. C., S. Tschopp, K. Rogers, F. Wasmer, and L. Liljegren, 2009: Quality control of meteorological data for the chemical stockpile emergency preparedness program. *J. Atmos. Oceanic Technol.*, **26**, 1510-1526.
- National Institute of Meteorological Science, 2014: *Handbook of Asian Dust Observation Environment*. KMA, 114 pp.
- Nie, S.-P., Y. Luo, W.-P. Li, T.-W. Wu, X.-L. Shi, and Z.-Z. Wang, 2012: Quality control and analysis of global gauge-based daily precipitation dataset from 1980 to 2009. *Adv. Climate Change Res.*, **3**, 45-53.
- Papale, D., and Coauthors, 2006: Towards a standardized processing of Net Ecosystem Exchange measured with eddy covariance technique: algorithm and uncertainty estimation. *Biogeosciences*, **3**, 71-583.
- Ryoo, S.-B., Y.-H. Kim, T.-H. Kwon, and I.-S. Park, 2006: Inhomogeneities in Korean climate data (I): Due to site relocation. *Atmosphere*, **16**, 215-223 (in Korean with English abstract).
- Shin, H.-J., J.-S. Shon, S.-A. Noh, J.-H. Park, M.-D. Lee, J.-C. Kim, and S.-J. Lee, 2014: Air quality data analysis system and air quality data analysis method. KR-A-1020140082564, 22 pp.
- Thermo Fisher Scientific, 2010: FH62C14 Continuous Ambient Particulate Monitor Instruction Manual.
- WMO, 2013: *Guide to the Implementation of a Quality Management System for National Meteorological and Hydrological Services*. WMO-No.1100, 85 pp.
- Zahumenský, I., 2004: Guidelines on Quality Control Procedures for Data from Automatic Weather Stations. WMO CIMO/OPAG-SURFACE/ET-ST&MT-1/Doc., 6.1(2), 10 pp.