

## 통계적 방법에 근거한 AMSU-A 복사자료의 전처리 및 편향보정

이시혜\* · 김상일 · 전형욱 · 김주혜 · 강전호

한국형수치예보모델개발사업단

(접수일: 2014년 8월 29일, 게재확정일: 2014년 11월 20일)

### Pre-processing and Bias Correction for AMSU-A Radiance Data Based on Statistical Methods

Sihye Lee\*, Sangil Kim, Hyoung-Wook Chun, Ju-Hye Kim, and Jeon-Ho Kang

Korea Institute of Atmospheric Prediction Systems (KIAPS), Seoul, Korea

(Manuscript received 29 August 2014; accepted 20 November 2014)

**Abstract** As a part of the KIAPS (Korea Institute of Atmospheric Prediction Systems) Package for Observation Processing (KPOP), we have developed the modules for Advanced Microwave Sounding Unit-A (AMSU-A) pre-processing and its bias correction. The KPOP system calculates the airmass bias correction coefficients via the method of multiple linear regression in which the scan-corrected innovation and the thicknesses of 850~300, 200~50, 50~5, and 10~1 hPa are respectively used for dependent and independent variables. Among the four airmass predictors, the multicollinearity has been shown by the Variance Inflation Factor (VIF) that quantifies the severity of multicollinearity in a least square regression. To resolve the multicollinearity, we adopted simple linear regression and Principal Component Regression (PCR) to calculate the airmass bias correction coefficients and compared the results with those from the multiple linear regression. The analysis shows that the order of performances is multiple linear, principal component, and simple linear regressions. For bias correction for the AMSU-A channel 4 which is the most sensitive to the lower troposphere, the multiple linear regression with all four airmass predictors is superior to the simple linear regression with one airmass predictor of 850~300 hPa. The results of PCR with 95% accumulated variances accounted for eigenvalues showed the similar results of the multiple linear regression.

**Key words:** AMSU-A, bias correction, airmass predictor, multicollinearity, linear regression

## 1. 서 론

수치예보모델에서 전구 분포의 높은 시·공간 해상도를 가진 위성 복사자료(satellite radiance data)의 직접 동화(direct assimilation)가 가능해지면서 초기 입력 자료의 품질이 개선되었고, 이로 인해 선진 현업 예보센터의 예보 성능은 상당히 향상되었다. 관측 값

과 모델 값의 차이를 최소화해 최적의 모델 초기자료를 만드는 자료동화시스템에서 위성자료를 효율적으로 사용하기 위해서는 위성에 특화된 품질검사가 필요하다. 위성자료는 기기나 측정방법의 부정확성, 관측자료의 변수변환을 위해 사용되는 복사전달모델의 불확실성, 모델 배경장이 가진 오차 등으로 인해 계통오차(systematic error)가 발생하므로 자료동화를 위한 관측자료 전처리시스템에서 위성자료의 편향을 제거하기 위한 모듈은 주요 품질검사 과정 중 하나로 알려져 있다(Auligné et al., 2007). 자료동화에서 관측자료의 신뢰도를 높이기 위해서는 정확한 관측오차의 제공과 함께 자료의 분포가 정규분포(Gaussian

\*Corresponding Author: Sihye Lee, Data Assimilation Team, Development Division, Korea Institute of Atmospheric Prediction Systems (KIAPS), 35 Boramae-ro 5-gil, Dongjak-gu, Seoul 156-849, Korea.  
Phone : +82-2-6959-1622, Fax : +82-2-6919-2570  
E-mail : sh.lee@kiaps.org

distribution)를 이루며 편향(bias)이 없다는 기본적인 통계적 가정을 만족해야 한다(Rabier et al., 2000; Rawlins et al., 2007). 편향보정 외에도 구름이나 강수의 영향 제거, 사용불가 채널에 대한 블랙리스트링, 속아내기 등 전처리 정도에 따라 분석장(analysis field)에 미치는 위성의 기여도가 달라질 수 있으므로 자료동화에서 전처리시스템의 개발은 매우 중요하다.

다양한 위성자료 중 Advanced Microwave Sounding Unit-A (AMSU-A)는 세계 최고의 예보효율을 가진 European Center for Medium range Weather Forecasting (ECMWF)에서 수치예보의 성능 향상에 가장 큰 기여를 하고 있는 관측중으로 검증되었다(Cardinali, 2009). AMSU-A는 적외과 위성에 비해 구름과 강수의 영향을 상대적으로 적게 받는 마이크로파 위성으로 산소 흡수 파장대(50~60 GHz)에 있는 12개의 온도 탐측(temperature sounding) 채널과 해양의 구름 및 강수 산출에 유용한 3개의 대기 창 채널을 가지고 있다. AMSU-A와 같은 교차진로 탐측(cross-track scanning) 방식의 위성은 스캔 위치에 따라 지구 대기과 위성 사이의 광학 길이 두께(optical path length)가 달라져 직하점(nadir)과 관측 가장자리(off-nadir)를 관측한 밝기온도(brightness temperature; TB)의 차이가 발생하므로 각 채널별로 스캔 편향(scan bias)이 존재한다(Goldberg et al., 2001). 또한, 지역에 따라 대류권과 성층권의 대기 상태가 다르고, 미량기체 및 에어로졸의 분포가 다르기 때문에 대기질량 편향(airmass bias)이 발생한다(Eyre, 1992). 그러므로 특정 고도의 온도와 밀접한 관련이 있는 각 채널의 특성에 맞추어 편향보정(bias correction)을 수행해 주어야 한다.

대부분의 현업수치예보센터에서는 위성자료의 편향보정계수를 구하기 위해 관측한 복사자료(Observation; O)와 모델 배경장을 이용해 계산한 복사자료(Background; B)의 차이인 관측증분(O-B; innovation)을 이용한다. 영국 기상청에서는 Observation Processing System (OPS)에서 위성자료의 편향보정을 포함한 구름탐지, 채널선택, 1D-Var를 이용한 최종 선택 자료의 점검 등 다양한 전처리를 수행하며, 관측증분을 이용해 계산한 편향보정계수를 사용한다(Hilton et al., 2009). ECMWF에서는 관측자료의 전처리를 위해 Continuous Observation Processing Environment (COPE) 시스템에서 이중자료 제거, 블랙리스트링, 속아내기 등을 수행하며, 변분에 기반 한 품질검사를 위해 관측증분을 사용하고 있다(Isaksen, 2012).

본 연구에서는 Korea Institute of Atmospheric Prediction Systems (KIAPS) 관측자료처리시스템(KIAPS Package for Observation Processing; KPOP)에 포함된 AMSU-A 전처리 과정을 자세히 설명하고, 편향보정계수를 구하기 위한 다양한 선형회귀 방법들을

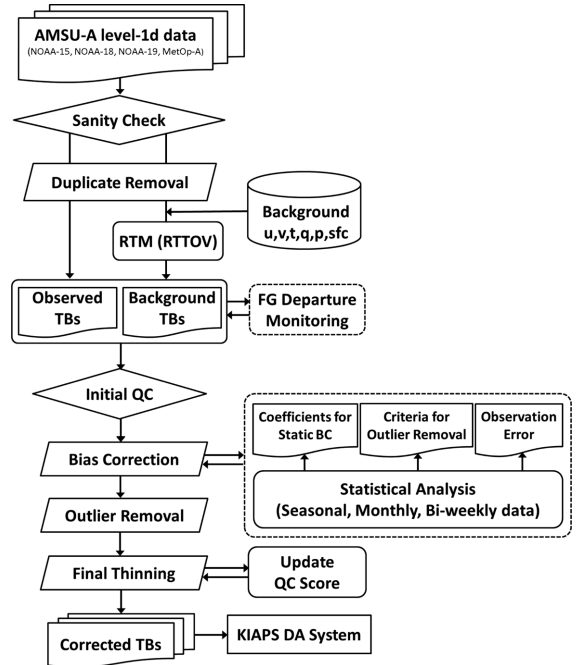


Fig. 1. Flowchart of AMSU-A pre-processing system in KPOP. Solid and dashed lines are on-line and off-line modules, respectively.

비교 분석해 보고자 한다. 일반적으로 편향보정을 수행하기 전에 관측에 문제가 있거나 모델 배경장을 이용해 복사전달모형을 계산하는 과정에서 문제가 되는 자료는 우선적으로 제거하므로 편향보정 이전에 수행하는 전처리 과정에 대해서도 함께 살펴볼 필요가 있다. 본 연구에서는 초기 품질검사를 마친 AMSU-A 자료를 이용해 편향보정계수를 구하기 위한 선형회귀식에서 종속변수와 독립변수들의 관계가 기본적인 통계적 가정을 만족하는지 살펴보았다. 그 후 새롭게 제안된 통계적 방법들과 기존의 방법으로 계산한 편향보정계수를 KPOP AMSU-A 전처리시스템에 각각 적용하였을 때 어떠한 결과적 차이가 있는지 분석하였다.

## 2. 연구 방법

### 2.1 AMSU-A 위성자료 처리시스템

KPOP에서는 편향보정을 포함한 AMSU-A 위성자료의 전처리 모듈을 Fig. 1과 같은 과정으로 개발하였다. 공개 소프트웨어인 ECMWF BUFR (Binary Universal Form for the Representation of meteorological data) decoder를 설치하여 NOAA-15, -18, -19 및 MetOp-A 위성에 탑재된 AMSU-A의 level-1d 자료를 추출하였으며, 추출한 원시 복사자료가 물리적인 실

제 허용 범위에서 관측되었는지 점검하는 정합성 검사(sanity check)를 하였다. AMSU-A는 여러 개의 위성에서 동시에 자료를 생산하기 때문에 자료 처리 시간의 과부하 방지를 위해 전처리 초기 단계에서 동일 지점을 관측한 중복자료를 제거(duplicate removal)하였다. 이 때, AMSU-A의 관측화소(pixel)를 고려하여  $0.5^{\circ} \times 0.5^{\circ}$  내에 들어온 자료들 중 위성의 신뢰도, 지표타입, 구름 오염 유무, 지정된 grid box 중심에서의 거리 등에 가중치를 주어 하나의 값을 선택하였다. 중복자료 제거 모듈을 통과한 자료는 모델 배경장을 이용하여 여러 단계의 품질검사를 거치게 된다.

기상변수로 된 모델자료를 품질검사에 이용하기 위해서는 관측자료와 같은 단위인 밝기온도로의 변수 변환이 필요하다. 본 연구에서는 Numerical Weather Prediction (NWP) Stellite Application Facility (SAF)의 Radiative Transfer for TIROS Operational Vertical Sounder (RTTOV) 버전 10.2를 설치하여 위성자료의 변수 변환을 위한 관측연산자(observation operator)로 활용하였다(Kim et al., 2014). RTTOV의 입력 자료인 모델 연직 층과 2 m 고도에 대한 온도, 습도, 압력, 바람 등의 기상정보는 기상청 현업예보모델에서 관측 자료의 품질검사를 위해 사용하는 배경장 파일에서 추출하였고, 모델 격자에서 정의된 기상 변수들은 수평방향과 연직방향으로 내삽하여 관측과 동일한 지점의 값으로 변환하여 사용하였다.

구름이 없는 맑은 하늘에서의 AMSU-A 자료동화 및 편향보정계수 적용을 위한 초기 품질검사(Initial QC)에서는 Grody et al. (1999, 2001)의 산란지수(scattering index), 해빙지수(sea-ice index), 구름수적(retrieved cloud liquid water) 등을 이용하여 관측오차가 큰 픽셀의 자료를 제거하였다. 또한, 위도별 지형 고도, 지표타입, 4번 채널의 관측증분 등을 이용한 초기 품질검사를 단계별로 수행하였다. 이 때, 지표면 방출률(surface emissivity) 또는 표면 온도의 불확실성이 큰 육지나 해빙이 많은 극 지역 자료는 제외하였다.

오프라인으로 구성된 통계처리 모듈에서는 일정 기간 동안 누적한 관측 자료와 배경장 자료를 이용하여 편향보정계수, 이상치 자료 제거(outlier removal)를 위한 임계값(threshold), 관측오차(observation error) 등 다양한 상수들을 계산한다. KPOP에서는 한 달 간격으로 누적한 위성자료를 이용하여 각 채널들에 대한 편향보정계수를 구한 후 이를 편향보정에 사용하고 있다. 편향보정을 마친 자료는 자료동화시스템에서 받아들이기 힘든 이상치 자료를 제거한 후 각 품질검사 단계에서 할당된 플래그로 각 관측지점에 품질검사 점수(QC score)를 매겨 최종적으로 자료를 슈아낸다(final thinning). 이 때, 자료동화시스템에 최적화된 격자 간격을 고려하여 슈아내기를 수행하며, 모든 품질

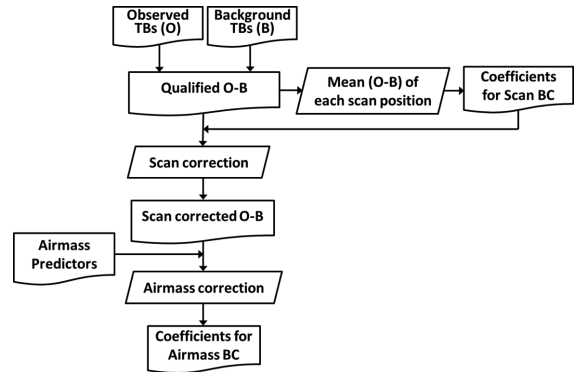


Fig. 2. Flowchart for calculating AMSU-A bias correction coefficients.

검사를 통과한 자료는 편향 없이 정규분포를 이루는 것을 확인한 후 자료동화시스템에 제공된다.

### 2.2 위성 복사자료의 편향보정계수 계산

AMSU-A를 비롯해 위성자료의 편향보정계수를 구하는 방법은 자료동화시스템에서 분석시간에 따라 바뀌는 모델 배경장을 반영하여 매 사이클 마다 업데이트하는 온라인 방법(Dec, 2004, 2005; Auligné et al., 2007)과 일정 기간 자료를 누적하여 통계적 선형회귀식으로 구하는 오프라인 방법이 있다(Atkinson et al., 2005; Baker et al., 2005). 통계적 방법에 근거해 편향보정을 실시하고 있는 대부분의 현업수치예보센터에서는 Harris and Kelly (2001)의 다중선형회귀식(multiple linear regression)을 이용하여 편향보정계수를 구하고 있으며, 기존에 개발된 KPOP에서도 다중선형회귀식을 이용해 AMSU-A의 편향보정계수를 구하고 있다(Lee et al., 2013).

본 연구에서는 편향보정계수를 구하기 위한 다양한 통계적 선형회귀식의 성능을 검증하기 위해 Fig. 1의 KPOP 전처리 과정을 통과한 2012년 11월 AMSU-A 자료에 Fig. 2를 통해 오프라인으로 구한 편향보정계수를 편향보정 단계에 적용하였다. 다양한 통계적 선형회귀식을 구할 때 스캔 편향계수는 모두 동일하게 계산하였고, 대기질량 편향계수만 다르게 계산하였다. 관측자료의 품질검사에 필요한 모델 배경장은 기상청 현업예보모델에서 관측자료의 품질검사를 위해 사용하는 동일한 배경장 자료(e.g., qwqu00.pp\_006)를 기상청에서 제공받았으며, 분석시간을 기준으로  $\pm 3$ 시간 이내에 수집된 관측 값을 배경장과 동일한 시간으로 가정하여 사용하였다.

[1] 편향보정계수를 구하는 첫 번째 방법은 Harris and Kelly (2001)에서 소개한 다중선형회귀식을 이용하는 것이다. Figure 2에서 볼 수 있듯이 스캔 편향을

보정하기 위해 초기 품질검사를 통과한 관측증분(qualified O-B)을 각각의 스캔 위치(scan position)에 대해 구한 후 스캔 가장자리에서 발생하는 편향을 직하점의 값으로 보정하였다(Lee et al., 2013). 스캔 편향보정을 마친 관측 값(scan corrected O)은 대기질량 편향계수를 구하기 위해 모델의 대기장을 대표하는 4개의 대기예측변수들(airmass predictors)과 함께 Eq. (1)과 같은 다중선형회귀식에 이용된다. 이 때, 관측과 배경장과의 차이를 나타내는 관측증분이 종속변수( $Y$ )이고, 대기예측변수로 사용된 Harris and Kelly (2001)의 850~300과 200~50 hPa 층후(thickness)와 ECMWF의 50~5와 10~1 hPa 층후가 독립변수들( $X_1 \sim X_4$ )이다.

$$Y = aX_1 + bX_2 + cX_3 + dX_4 + e. \quad (1)$$

일반적으로 다중선형회귀분석은 독립성을 가진 종속변수가 정규분포를 이루고 있으며, 등분산성(constant variance)을 만족한다는 가정을 전제로 한다. 또한, 독립변수들 간에는 선형관계가 존재하지 않고 각각 독립적이라는 가정을 만족해야 정확한 회귀분석 결과를 얻을 수 있다(Kwon, 2004b). 다중선형회귀분석은 모든 대기예측변수가 고려된다는 장점이 있지만, 회귀분석을 위한 여러 가정이 충족되어야 하고 계산과정이 복잡하다는 단점이 있으므로 대기예측변수들과 관측증분의 특성을 잘 파악하여 사용해야 한다. Equation (1)에서 4개의 대기예측변수들과 관측증분의 다중선형회귀식을 통해 대기질량 편향보정계수인  $a, b, c, d, e$ 를 구한 후 Eq. (2)와 같이 대기질량 편향을 계산하였다. 여기서,  $bias_{j,s}^{air}$ 는 위성  $s$ 의  $j$  채널에서의 대기질량 편향이고,  $Thick_{850}, Thick_{200}, Thick_{50}, Thick_{10}$ 은 모델의 대기장을 대표하는 대기예측변수로 각각 850~300, 200~50, 50~5, 10~1 hPa 층후를 의미한다.

$$bias_{j,s}^{air} = a_{j,s}Thick_{850} + b_{j,s}Thick_{200} + c_{j,s}Thick_{50} + d_{j,s}Thick_{10} + e_{j,s}. \quad (2)$$

[2] 편향보정계수를 구하는 두 번째 방법은 Eq. (3)과 같이 단순선형회귀식(simple linear regression)을 이용하는 것으로 각 채널별로 관측 값과 상관성이 가장 높은 하나의 대기예측변수를 독립변수( $X$ )로 택해 스캔 편향보정된 관측증분(종속변수,  $Y$ )과의 선형회귀계수인  $a, b$ 를 계산하는 것이다.

$$Y = aX + b. \quad (3)$$

채널별로 상관성이 가장 높은 대기예측변수는 각각의 층후와 관측증분 사이의 결정계수( $r^2$ ) 및 평균제곱근오차(root mean square error; RMSE), 적도 표준대기에서 각 채널의 가중함수(weighting function) 최대치를 포함하는 층후를 고려하여 결정하였다. 하층 채

널인 4~6번 채널에서는 850~300 hPa 층후, 대류권 상층과 성층권 하층을 포함하는 7~9번 채널에서는 200~50 hPa 층후, 성층권 중·상층에 해당하는 10~12번 채널에서는 50~5 hPa 층후, 50 km 이상의 중간권 대기 온도를 반영하는 13번 채널에서는 10~1 hPa 층후가 각각 선택되었다.

단순선형회귀식은 다중선형회귀식에 비해 계산과정이 비교적 간단하며, 하나의 독립변수만을 사용하기 때문에 대기예측변수와 종속변수인 관측증분과의 선형관계를 명확히 파악할 수 있는 장점이 있다. 그러나 특정 층을 중심으로 여러 층의 대기 특성을 지닌 채널의 성격을 하나의 층후가 온전히 반영하지 못할 우려가 있다.

[3] 편향보정계수를 구하는 세 번째 방법은 대기예측변수의 주성분(principal component; PC)을 이용한 주성분회귀분석(principal component regression; PCR)이다. 주성분분석(principal component analysis; PCA)은 변수들의 선형 결합을 통해 변수들이 가진 정보를 최대한 설명할 수 있고, 독립적인 새로운 변수들을 유도할 수 있으며, 원래 변수들의 정보를 손실 없이 다른 차원으로 이동시키는 통계적 분석방법이다(Smith, 2002). 만약 독립변수들 사이에 상관성을 나타내는 다중공선성(multicollinearity)이 존재하면 주성분 변수를 이용해 이를 해결할 수 있으나, 주성분 변수에 대한 해석이 쉽지 않다는 단점이 있다. 본 연구에서는 4개 층후들에 대한 고유값(eigenvalue)과 고유벡터(eigenvector)를 구한 후 설명변수인 대기예측변수들이 서로 직교하도록 각각의 층후에 고유벡터를 곱해 대기예측변수들의 주성분  $PC_1 \sim PC_4$ 를 구하였다. 주성분회귀식에서는 Eq. (4)와 같이 여러 조합의 주성분들과 관측증분의 선형회귀계수  $a, b, c, d, e$ 를 구하였다.

$$Y = aPC_1 + bPC_2 + cPC_3 + dPC_4 + e. \quad (4)$$

주성분회귀식을 통해 구해진 계수는 Eq. (5)와 같이 대기질량 편향을 계산하기 위해 각각의 층후에 고유벡터  $eig_{850}, eig_{200}, eig_{50}, eig_{10}$ 를 곱한 값에 이용된다. 여기서,  $bias_{j,s}^{air}, Thick_{850}, Thick_{200}, Thick_{50}, Thick_{10}$  변수는 Eq. (2)에서 설명한 바와 동일하다.

$$bias_{j,s}^{air} = a_{j,s}Thick_{850}eig_{850} + b_{j,s}Thick_{200}eig_{200} + c_{j,s}Thick_{50}eig_{50} + d_{j,s}Thick_{10}eig_{10} = E_{j,s}. \quad (5)$$

이와 같이 [1]~[3]의 다양한 선형회귀식을 이용해 계산한 편향보정계수들의 성능은 6시간 간격으로 묶인 관측자료에 각각의 편향보정계수를 적용한 관측 값(bias-corrected observation; C)과 초기 추정치(first guess; B)와의 차이(즉, 관측증분; C-B)를 한 달 동안 평균한 공간분포 자료(time-averaged geographical mean field of bias-corrected innovation)를 중심으로 분석하였다.

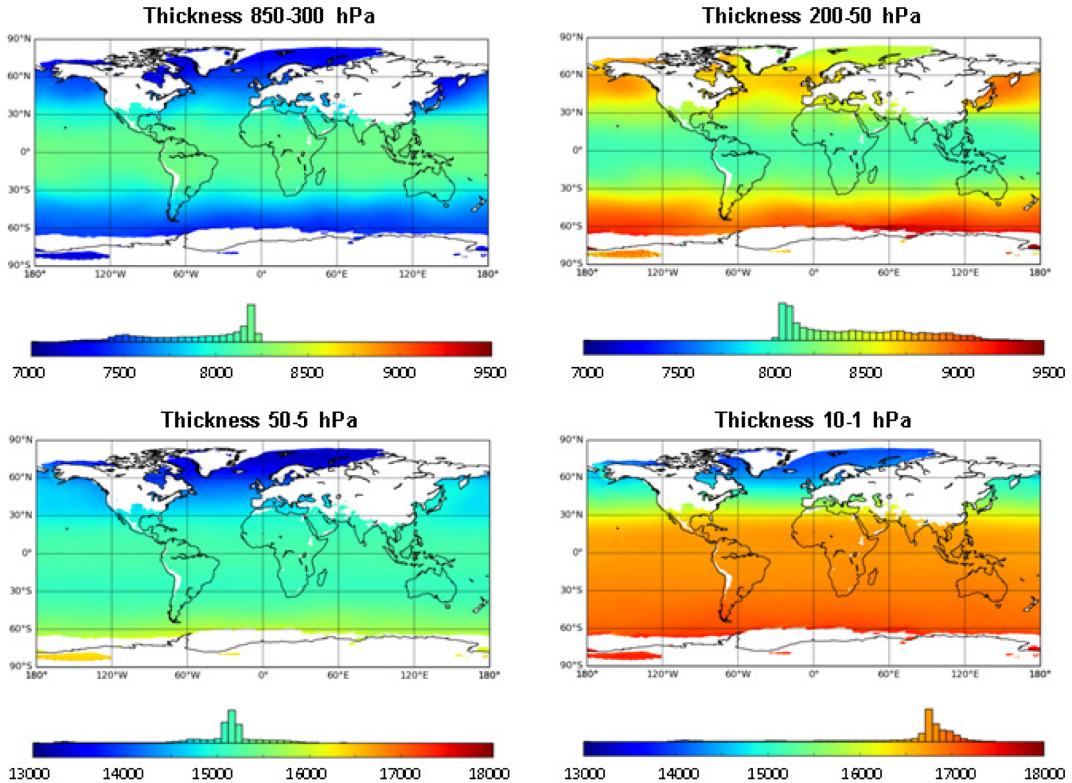


Fig. 3. Spatial distributions and their histograms for airmass predictors for November 2012 (unit: meter).

### 3. 결과 및 토의

#### 3.1 대기예측변수의 독립성

KPOP에서는 위성자료의 대기질량 편향보정을 위해 다중선형회귀식의 계수를 이용하고 있는데, 통계적으로 회귀계수란 선형회귀식에서 독립변수의 값이 한 단위 증가할 때 변화한 종속변수의 양을 의미한다. 다중선형회귀식에는 여러 개의 독립변수가 존재하므로 선형회귀식을 계산할 때 다른 독립변수의 값이 일정하다는 전제가 필요하다. 만약 독립변수들 간에 높은 상관관계가 있으면 종속변수가 동시에 변할 수 있으므로 회귀모형을 분석하기에 앞서 독립변수들 간의 독립성 유무를 조사해 보아야 한다.

Figure 3에서는 독립변수로 사용되는 대기예측변수들의 2012년 11월 평균 공간분포를 보여준다. 각각의 층후는 위도 대에 따라 대기의 특성을 반영하여 두껍거나 얇게 나타났다. 하층 대기장을 대표하는 850~300 hPa 층후는 중층 대기장을 대표하는 200~50 hPa 층후와 강한 음의 상관관계를 보였으며, 상층 대기장을 대표하는 50~5 hPa 층후와 10~1 hPa 층후는 적도를 포함한 중위도에서 강한 상관성을 보였다. 이와 같이 독립

변수들끼리 상관관계를 갖는 현상을 다중공선성이라고 하며, 다중공선성이 존재하면 회귀식의 분산이 매우 커져 정확한 추정 및 검증이 어려울 뿐만 아니라 심한 경우에는 회귀계수의 부호가 바뀌는 기현상이 나타날 수도 있다(Kwon, 2004a). 현실적으로 대기예측변수 간에 완벽한 독립성을 갖기는 매우 어렵기 때문에 선형회귀식을 설명할 수 있는 범위 내에서 KPOP의 편향보정계수가 구해지고 있는지 통계적 유의성에 대한 검정이 필요하다.

다중공선성 진단을 위한 가장 일반적인 방법은 Eq. (6)의 분산팽창지수(variance inflation factor; VIF)를 이용하는 것이다. 여기서,  $r_j^2$ 는 독립변수 중 하나를 종속변수로 하고 나머지 변수를 독립변수로 하는 회귀모형에서의 결정계수이다. VIF는 독립변수 사이에 발생하는 다중공선성으로 인한 분산의 증가를 의미하며, 일반적으로  $k$ 개의  $VIF_j$  중 가장 큰 값이 5~10을 넘으면 다중공선성이 존재한다고 알려져 있다(Jun et al., 2004).

$$VIF_j = \frac{1}{1 - r_j^2}, j = 1, 2, \dots, k. \quad (6)$$

**Table 1.** Monthly averages, variances, coefficients of determination ( $r^2$ ) and variance inflation factors ( $VIF$ ) of air mass predictors.

	Thickness <sub>850-300</sub>	Thickness <sub>200-50</sub>	Thickness <sub>50-5</sub>	Thickness <sub>10-1</sub>
Average (km)	7.87	8.50	15.05	16.45
Variance (km)	0.09	0.12	0.26	0.66
$r^2$	0.93	0.93	0.94	0.93
$VIF$	14.45	14.15	15.42	13.66

Table 1에서는 2012년 11월 한 달 자료를 이용해 계산한 대기예측변수들의 다양한 통계값을 보여준다. 대기의 연직두께를 나타내는 850~300, 200~50, 50~5, 10~1 hPa 층후들의 평균은 각각 7.87, 8.50, 15.05, 16.45 km로 대기 상층으로 갈수록 두꺼운 대기층의 특성을 반영하고 있다. 각 층후의 분산은 하층에서 상층으로 갈수록 0.09, 0.12, 0.26, 0.66 km로 커져 가장 두꺼운 연직 층후인 최상층 대기의 변화량이 가장 크다는 것을 알 수 있다. 이러한 특성을 가진 연직 층후들의 결정계수와 분산팽창지수는 각각 0.93~0.94와 13.66~15.42로 매우 높게 구해졌다. 즉, AMSU-A 편향보정계수 계산에 사용 중인 대기예측변수들은 서로 간에 강한 상관관계를 갖는 다중공선성이 존재함을 알 수 있다.

다중공선성을 해결하기 위한 대안 중 하나는 독립변수 선택 과정에서 상관계수가 높은 하나의 변수를 택하는 방법이 있으며, 이 방법은 종속변수를 설명하는 데 중요한 역할을 하는 독립변수를 제외시키지 않도록 주의를 기울여야 한다. 또 다른 대안은 최소자승법을 이용하는 다중선형회귀식 대신에 능형회귀(ridge regression)나 주성분회귀를 사용하는 것이다 (Marill, 2004). 3.2절에서는 하나의 독립변수를 택하는 단순선형회귀 방법과 주성분을 이용한 주성분회귀 방법을 적용하여 편향보정계수를 구한 후, 다중선형회귀식을 이용해 계산한 기존의 편향보정계수와 새로운 통계적 방법으로 구한 편향보정계수의 성능을 비교해 보았다.

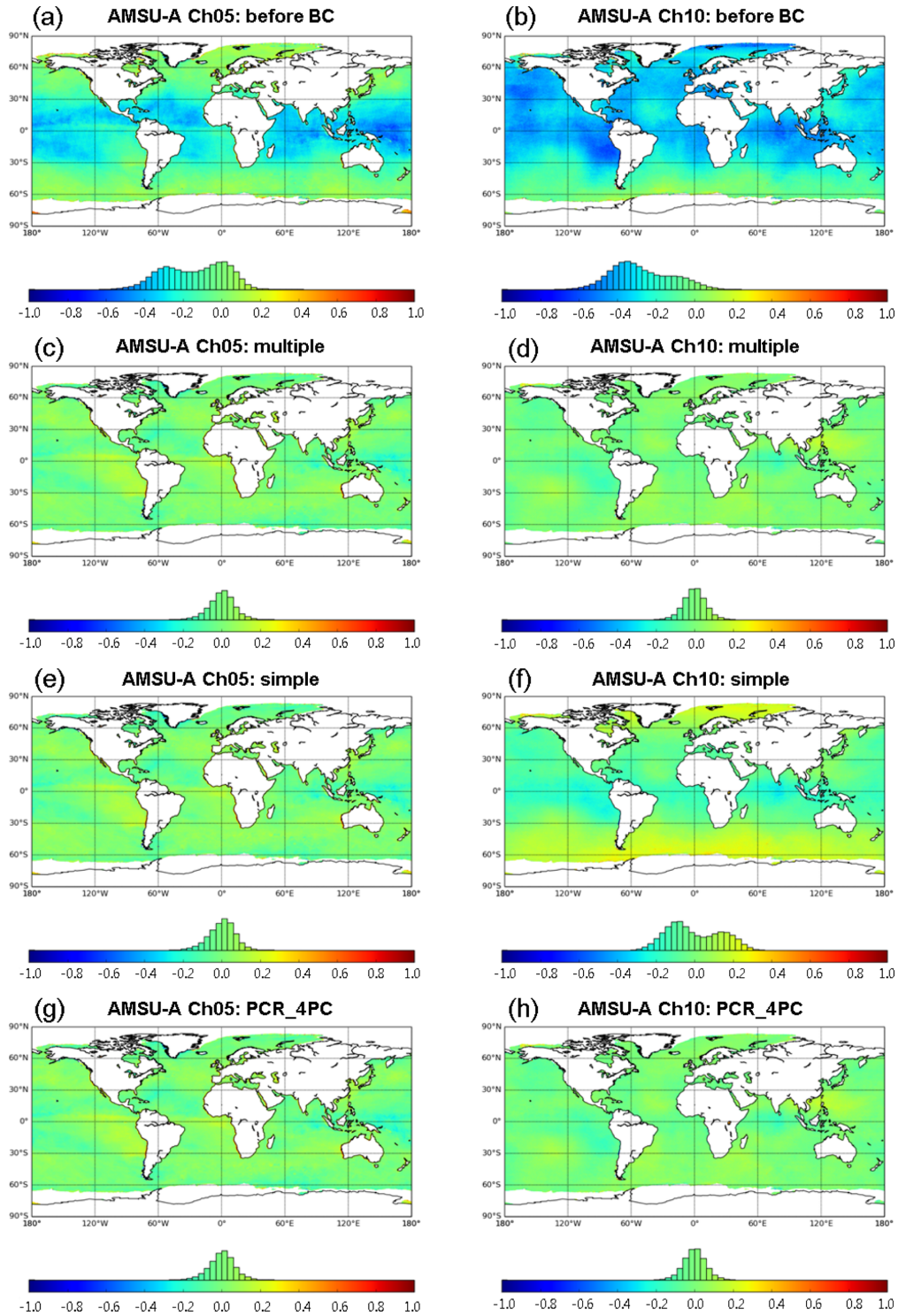
### 3.2 통계적 방법으로 구한 편향보정계수의 비교

Figure 4에서는 10 km 이하의 대류권 온도 특성을 반영하는 AMSU-A 5번 채널과 10~40 km의 성층권 온도 특성을 반영하는 10번 채널의 편향보정 전·후 결과를 보여준다. Figures 4a, b는 Fig. 2의 KPOP 시스템에서 초기 품질검사를 통과하고 편향보정을 수행하기 전의 관측증분이다. 하층 대기 특성을 반영하는 5번 채널은 적도를 포함한 30도 이하의 저위도에서 -0.6 K까지 음의 편향이 나타나고, 상층 대기 특성을 반영하는 10번 채널은 북반구 전체와 남반구 30도 이하의 저위도에 음의 편향이 존재하는 것을 볼 수 있다.

각 채널별로 위도 대에 따라 다르게 나타나는 대기 질량 편향을 보정하기 위해 다중선형회귀식으로 계산한 기존의 편향보정계수를 2012년 11월 한 달 AMSU-A 자료에 적용한 결과, 관측증분은 Figs. 4c, d와 같은 공간분포를 보였다. 5번과 10번 채널의 관측증분 모두 정규분포를 이루고, 대부분의 지역에서 관측과 배경장의 차이가  $\pm 0.2$  K 이내로 나타나 편향보정이 잘 수행된 것을 알 수 있다. 그러나 이 방법은 3.1절에서 검증한 바와 같이 독립변수로 사용된 대기예측변수들 간에 다중공선성이 존재하므로 대안으로 제시된 통계적 방법들을 적용하여 편향보정계수를 재계산한 후 기존의 편향보정 결과를 비교해 보았다.

Figures 4e, f는 단순선형회귀식으로 계산한 편향보정계수를 KPOP 시스템의 편향보정 단계에 적용한 결과이다. 850~300 hPa 층후를 대기예측변수로 선택한 5번 채널의 편향보정계수는 다중선형회귀식 계수를 이용한 Fig. 4c와 유사한 수준의 결과를 보였으나, 50~5 hPa 층후를 선택한 10번 채널의 계수는 편향보정을 제대로 수행하지 못한 것을 알 수 있다. Figure 4f에서 편향보정 이후 상층 대기에는 적도 부근 저위도에 음의 편향이 여전히 강하게 남아있고, 남반구 30도 이상 고위도에는 양의 편향이 새롭게 나타나 관측증분은 정규분포를 이루지 못하였다. 대기 상층 10번 채널에서는 50~5 hPa 층후를 택하기보다 다중공선성이 존재하더라도 4개의 층후를 모두 이용해 편향보정계수를 구하는 것이 더 좋은 결과를 보였다. 이는 독립변수로 사용되는 대기예측변수들의 다중공선성이 선형회귀식의 분석력을 크게 감소시키지 않는 것으로 해석된다.

3.1절에서 독립변수들의 다중공선성 검증에 이용한 분산팽창지수는 독립변수들이 상관관계를 갖지 않을 때에 비해 추정된 회귀계수들의 분산이 얼마나 팽창하는지를 설명해 주는 인자이다. 만약 회귀계수들의 분산이 매우 작다면, 다중공선성으로 인해 회귀계수들의 분산이 10배 이상 증가하여도(즉,  $VIF \approx 10$ ) 대기예측변수들의 예측력은 크게 감소하지 않을 것이다. Equation (1)에서 다중선형회귀계수  $a, b, c, d$ 의 분산은 공분산 행렬을 이용한  $\sigma^2(X^T X)^{-1}$ 로 계산된다(Liao and Valliant, 2012). 10번 채널의 회귀계수 분산들은



**Fig. 4.** (a, b) Spatial distributions and their histograms of innovations before bias correction, and (c, d) bias-corrected innovations based on multiple linear regression, (e, f) simple linear regression and (g, h) principal component regression (PCR) methods for AMSU-A channels 05 (left) and 10 (right) for November 2012 (unit: Kelvin).

각각 1.18E-7, 8.59E-7, 3.93E-8, 1.55E-8로 구해져 예상한 바와 같이 매우 작음을 알 수 있었다. 다중선형 회귀식에서 종속변수의 분산은  $V(Y) = V(aX_1 + bX_2 + cX_3 + dX_4 + e) = |a|^2 V(X_1) + |b|^2 V(X_2) + |c|^2 V(X_3) + |d|^2 V(X_4)$  관계를 만족하므로 종속변수에 해당하는 관측증분의 분산은 기울기와 독립변수들의 분산( $V(X_1) \sim V(X_4)$ )으로 결정된다. 이 때, 다중공선성으로 인해 회귀계수들의 분산도 중요할 수 있으나 한 달 자료를 누적해 통계분석을 하고 있는 본 연구에서는 매우 작은 값으로 계산되었다. 한 달 이상의 위성자료를 누적해 통계적 방법으로 편향보정계수를 구할 때는 다중공선성이 존재하더라도 회귀계수들의 분산이 매우 작게 구해져 관측증분에 미치는 영향이 미미할 수 있지만, 변분에 기반 해 6시간 이하의 자료로 편향보정을 수행한다면 다중선형회귀식에서 다중공선성에 의한 영향도 고려해 주어야 할 것이다.

단순선형회귀식 계수를 이용한 편향보정에서 채널별 결과가 다른 이유를 파악하기 위해 채널에 따라서 다른 대기예측변수를 택한 각각의 회귀모형에서 오차 항에 대한 추정치인 잔차(residual;  $Y-\hat{Y}$ )를 분석하였다. 잔차평균제곱(residual mean square)이라 불리는 평균제곱오차(mean square error; MSE)는 5번 채널에서 0.07, 10번 채널에서 0.12로 구해졌으며, 하층에서 상층으로 갈수록 평균제곱오차가 커졌다(MSE = 0.07~0.54; Table omitted). 오차의 분산 추정치인 평균제곱오차가 작을수록 모형의 적합도가 높기 때문에 (Wackerly and Scheaffer, 2008), 상층 채널에서 사용한 대기예측변수가 대기질량 편향을 보정해 주기에 상대적으로 부족하다는 것을 알 수 있다. 각 채널의 가중합수가 대기의 여러 층에 걸쳐 나타난다는 점을 고려할 때, 상층 채널에서는 단순선형회귀식에서 선택한 대기예측변수가 각 채널의 연직 대기장 특성을 모두 반영하지 못한 것으로 파악된다.

마지막으로 주성분회귀식으로 구한 편향보정계수를 KPOP 시스템에 적용한 결과를 Figs. 4g, h에서 살펴 보았다. 주성분회귀식은 대기예측변수를 독립성을 만족하는 회귀변수인 주성분으로 변환하여 관측증분과의 선형회귀분석을 시도하는 방법이다. 대기예측변수들의 특성을 최대한 포함하기 위해 첫 번째에서 네 번째 주성분들인 Eq. (4)의  $PC_1 \sim PC_4$ 로 회귀계수를 구해 편향보정을 수행했을 때, 그 결과는 다중선형회귀식 계수를 이용한 Figs. 4c, d의 편향보정 결과와 매우 유사하게 나타났다. 이는 새롭게 변환된 주성분들이 상·하층 대기장을 대변하는 대기예측변수들의 특성을 온전히 반영하기 때문이다.

변환된 주성분이 대기예측변수의 특성을 얼마나 설명하고 있는지 살펴보기 위해 주성분을 구할 때 계산되는 대기예측변수의 고유값과 고유벡터를 분석하였

**Table 2.** Eigenvalues of four airmass predictors at each satellite.

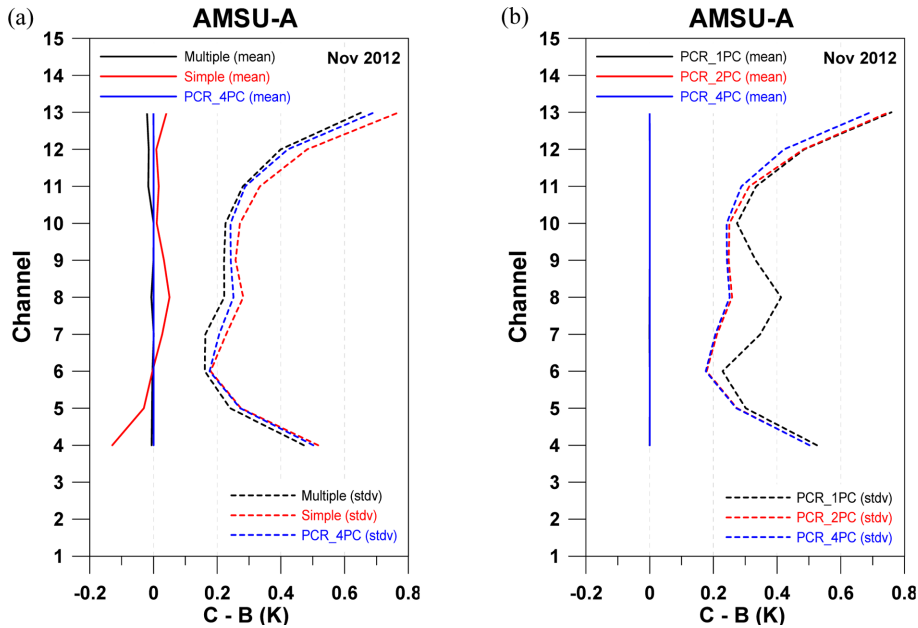
PC	NOAA-15	NOAA-18	NOAA-19	MetOp-A
1	8.78E+3	8.78E+3	8.50E+3	8.35E+3
2	2.31E+4	2.33E+4	2.57E+4	2.25E+4
3	1.99E+5	1.93E+5	1.81E+5	2.01E+5
4	1.06E+6	1.12E+6	1.21E+6	9.54E+5

다. 고유벡터는 주성분을 구하는 과정에서 원래 변수에 선형변환을 가할 때 회전 변환 없이 확대 또는 축소 변환을 수행하는 연산자이고, 확대 또는 축소의 스칼라 양에 해당하는 것이 고유값이다(Jeong et al., 2009). 고유값이 1보다 작으면 주성분은 원래 변수 보다 작은 변수를 가지게 되어 고려할 가치가 없어지게 되는데, 대기예측변수로 쓰인 4개 층후들의 고유값은 모두 1 이상으로 구해졌다. 이 때, 주성분들의 분산은 고유값과 같으므로 각 고유값이 전체 고유값의 합에서 차지하는 비율은 각 주성분이 전체분산을 설명하는 비율이 된다. Table 2에서 위성별로 계산한 대기예측변수들의 고유값들은 각각의 주성분들에 대해 거의 유사한 값을 갖는 것을 알 수 있다. 세 번째와 네 번째 주성분의 위성별 고유값은 각각 1.81E+5~2.01E+5 과 9.54E+5~1.21E+6으로 첫 번째 주성분과 두 번째 주성분의 고유값에 비해 월등히 크게 나타났다. 각 주성분이 전체분산을 설명하는 비율은 세 번째와 네 번째 주성분의 누적비율이 95% 이상으로 높게 구해져 Eq. (4)의 주성분들 중  $PC_3$ 와  $PC_4$ 가 대기예측변수가 지닌 정보를 대부분 설명한다는 것을 알 수 있었다. 3.3절에서는 AMSU-A 각 채널에 대해 다양한 통계적 회귀방법으로 구한 편향보정계수들의 성능을 비교 분석하면서 대기예측변수의 설명력이 높은 주성분만을 이용해 계산한 주성분회귀 계수의 성능도 함께 살펴 보았다.

### 3.3 연직 층에 따른 대기질량 편향보정

위성자료의 편향은 대기의 온도에 따라 다양한 특성을 나타내므로 일반적으로 편향보정계수는 가중합수의 연직분포가 다르게 나타나는 채널별로 각각 계산해서 사용한다. Figure 5a와 Table 3에서는 Eqs. (1)~(5)의 다중선형회귀식, 단순선형회귀식, 주성분선형회귀식으로 구한 편향보정계수를 AMSU-A의 4~13번 채널에 적용한 결과를 보여준다. 위성자료의 편향보정은 관측증분의 공간분포가 0을 중심으로 정규분포를 이루고(즉, 관측증분의 평균이 0에 가깝고), 시간에 따른 관측증분의 표준편차가 작을 때 잘 수행되었다고 할 수 있다. 이러한 기준으로 Fig. 5a와 Table 3의 결과를 분석하면 전 채널에 걸쳐 다중선형회귀식으로





**Fig. 5.** (a) Comparison of bias-corrected innovations (C-B) for different regression methods and (b) PCR method with different number of PCs for AMSU-A each channel. Solid and dotted lines are monthly means and standard deviations of C-B, respectively (unit: Kelvin).

**Table 3.** Monthly means and its standard deviations of bias-corrected innovations by each statistical regression method (unit: Kelvin).

Channel	Multiple linear regression	Simple linear regression	Principal component regression
04	-6.00E-3 ± 0.47	-1.29E-1 ± 0.52	4.55E-5 ± 0.50
05	-4.64E-3 ± 0.24	-3.04E-2 ± 0.27	2.81E-4 ± 0.27
06	-3.30E-3 ± 0.16	-3.17E-3 ± 0.18	1.81E-4 ± 0.18
07	8.27E-4 ± 0.16	2.73E-2 ± 0.23	7.68E-5 ± 0.21
08	-6.57E-3 ± 0.22	5.06E-2 ± 0.28	2.71E-4 ± 0.25
09	-1.64E-4 ± 0.22	3.29E-2 ± 0.26	2.18E-4 ± 0.24
10	-1.61E-5 ± 0.23	1.01E-2 ± 0.27	3.64E-4 ± 0.24
11	-1.61E-2 ± 0.28	1.67E-2 ± 0.33	3.71E-4 ± 0.29
12	-1.49E-2 ± 0.40	8.97E-3 ± 0.48	1.79E-6 ± 0.42
13	-2.01E-2 ± 0.65	4.12E-2 ± 0.77	-1.38E-4 ± 0.69

구한 편향보정계수의 성능이 가장 우수하고, 주성분회귀식 계수와 단순선형회귀식 계수의 성능이 차례로 그 뒤를 따른다. 대기 하층에서는 세 방법으로 구한 편향보정계수들의 성능이 비슷하지만, 상층으로 갈수록 편향보정계수들의 성능 차이가 커지는 것을 알 수 있다.

4-6번 채널은 850~300 hPa 층후의 영향을 가장 많이 받는 채널이며, 단순선형회귀식에서 850~300 hPa 층후를 대기예측변수로 택하고 있다. Figure 5a에서 대기예측변수의 특성이 모두 변환된  $PC_1 \sim PC_4$ 의 주성분을 이용한 선형회귀계수(PCR\_4PC)로 편향보정을 수행했을 때 관측증분의 한 달 평균이 거의 0 (4.55E-

5~2.81E-4; refer to Table 3)에 가까운 값으로 구해져 가장 좋은 정규분포를 이루었다. 반면, Table 3에서 단순선형회귀 계수를 편향보정에 적용했을 때 4번 채널의 관측증분 평균은 -1.29E-1로 다른 두 회귀방법 계수들에 비해 성능이 크게 떨어졌다. 하층 채널들의 편향보정은 다중선형회귀식 계수를 적용할 때 관측증분의 표준편차가 0.16~0.47로 가장 좋은 결과를 보였고, 단순선형회귀식 계수와 주성분회귀식 계수를 적용할 때 관측증분의 표준편차가 각각 0.18~0.52와 0.18~0.50으로 유사한 결과를 나타냈다. 이를 통해 최하층 대기에서도 대기예측변수로 850~300 hPa 층후만이 아

닌 대기 상층 층후까지 고려하는 것이 편향보정에 유리하다는 것을 알 수 있다.

200~50 hPa 층후의 영향이 강한 7~9번 채널은 Fig. 5a의 관측증분 표준편차에서 보여주듯이 다양한 회귀 방법으로 구한 편향보정계수들의 성능 차이가 좀 더 분명하게 나타난다. 하층 채널과 마찬가지로 편향보정계수의 성능은 다중선형회귀식, 주성분회귀식, 단순선형회귀식의 순으로 좋은 결과를 보였다. 주성분회귀계수의 성능을 좀 더 명확히 살펴보기 위해 Fig. 5b에서 대기예측변수의 설명력이 큰 주성분을 선별하여 회귀계수를 구한 후 편향보정에 적용하였다. 대기예측변수의 고유값을 이용해 계산한 주성분의 설명력은 각 위성별로 네 번째 주성분인  $PC_4$ 가 80~85%로 가장 높았고, 세 번째 주성분인  $PC_3$ 가 13~17%로 구해졌다. 대기예측변수의 특성을 95% 이상 설명할 수 있는  $PC_3$ 와  $PC_4$ 를 함께 사용해 구한 주성분회귀계수(PCR\_2PC)의 성능은 기존의 주성분회귀계수(PCR\_4PC)와 비슷한 결과를 보였다. 그러나 약 80%의 대기예측변수 설명력을 가진  $PC_4$ 를 사용하여 산출한 편향보정계수(PCR\_1PC)의 성능은 8번 채널의 관측증분 표준편차가 0.41 K로 크게 나타나 상대적으로 좋지 않음을 알 수 있다. 이는 8번 채널의 가중합수가 대류권계면(tropopause)에 속하는 150 hPa 부근에서 가장 크기 때문에  $PC_4$ 가 이 영역의 대기장 특성을 제대로 반영하지 못했을 가능성이 있으나 현재로서는 이에 대한 명확한 원인을 찾기는 어렵다.

대기 상층의 온도 특성을 반영하는 10~13번 채널의 편향보정 결과는 통계적 회귀방법에 상관없이 대기 최상층으로 갈수록 관측과 배경장의 차이인 C-B의 표준편차가 커지는 일반적인 특성을 나타냈다. 편향보정을 수행한 후에 계산한 관측증분의 표준편차가 커지는 것은 자료가 가진 관측오차가 커지는 것을 의미한다. Figure 5a에서 편향보정계수의 성능이 가장 좋은 다중선형회귀식 방법에서도 4, 12, 13번 채널 등 최하층과 최상층 채널에서 시간에 따른 관측증분의 변화량(즉, 표준편차)이 큰 것을 볼 수 있다. 이 채널들의 관측오차는 편향보정계수를 구하는 방법에 상관없이 클 것으로 예상된다.

#### 4. 결 론

KPOP에서는 AMSU-A의 대기질량 편향보정계수를 구하기 위해 모델의 대기장을 대표하는 850~300, 200~50, 50~5, 10~1 hPa 층후들과 관측증분의 다중선형회귀식을 이용하고 있다. 이 때, 독립변수로 사용된 대기예측변수들 간에는 선형관계가 존재하지 않는다는 통계적 가정을 만족해야 한다. 분산팽창지수를 이용해 대기예측변수인 연직 층후들의 독립성을 검증한

결과, 각 층후들의 분산팽창지수가 모두 10 이상으로 구해져 대기예측변수들 사이에 다중공선성이 존재하였다. 본 연구에서는 대기예측변수들의 다중공선성 문제가 야기되지 않는 단순선형회귀식과 주성분회귀식으로 구한 편향보정계수를 KPOP 시스템에 각각 적용한 후 기존의 다중선형회귀식 계수를 적용한 편향보정 결과와 비교해 보았다. 전 채널에 걸쳐 다중선형회귀식, 주성분회귀식, 단순선형회귀식 계수들의 순으로 편향보정에 우수한 성능을 보임을 알 수 있었다.

주성분회귀식을 이용해 편향보정계수를 구하는 방법은 새롭게 구해진 주성분들이 서로 독립이라는 특성이 있으므로 대기예측변수의 다중공선성 문제를 해결하는 동시에 다중선형회귀식 방법에 버금가는 결과를 보였다. 그러나 대기예측변수의 고유값과 고유벡터를 구해야 하는 번거로움이 있고, 대기예측변수의 특성을 대표할 수 있는 주성분을 선별하여 편향보정계수를 구하는 것도 쉬운 일은 아니다. 본 연구에서는 주성분의 대기예측변수 누적 설명력이 95% 이상이어야 전 채널에 걸쳐 다중선형회귀식 계수에 준하는 성능을 나타낼 수 있었으며, 이를 위해 세 번째와 네 번째 주성분을 함께 사용해 주성분회귀계수를 구해야만 했다. 편향보정계수의 성능 측면에서는 주성분회귀계수의 편향보정 결과가 다중선형회귀계수의 결과를 능가하지 못했지만, 대기예측변수의 독립성 유지 측면에서는 주성분회귀계수가 다중선형회귀계수보다 통계적으로 적합함을 알 수 있었다. 하나의 대기예측변수를 택한 단일선형회귀식 방법은 하층 대기에서는 다중선형회귀계수와 성능 면에서 큰 차이가 없었지만 상층 대기로 갈수록 다른 회귀방법의 계수들과 성능 차이가 크게 나타났다.

만약 선진 현업예보센터에서 다중선형회귀식을 이용해 편향보정계수를 구할 때 본 연구와 동일한 대기예측변수를 이용한다면 독립변수들 간에는 분명 다중공선성이 존재할 것이다. 그럼에도 현재까지 대부분의 현업예보센터에서 이 방법을 그대로 사용하는 것은 다른 대안이 없고, 다중선형회귀식으로 구한 편향보정계수의 성능이 나쁘지 않기 때문이다. 다중선형회귀식은 모델의 대기장을 대표하는 다양한 대기예측변수들을 모두 반영할 수 있다는 장점이 있고, 대기예측변수들의 다중공선성 문제가 없는 주성분회귀식의 계산과정 보다 훨씬 간단하면서도 편향보정에 좋은 성능을 보이기 때문에 현업예보센터에서 여전히 사용 중인 것으로 여겨진다.

본 연구는 현재 KPOP에서 사용 중인 편향보정계수가 다중선형회귀식의 통계적 가정을 만족하지 않고 구해졌음에도 현재 상태로 편향보정에 사용할 수 있는지 살펴보기 위한 선행연구이다. 다양한 선형회귀식을 이용해 계산한 편향보정계수들의 비교는 KPOP

에서 위성자료의 오프라인 편향보정계수를 구하기 위한 회귀방법의 결정을 위한 근거로 활용 가능할 것이다. 향후 자료동화시스템에 좋은 품질의 위성자료를 제공하기 위해서는 보다 개선된 편향보정 모듈을 개발할 필요가 있으며, 이를 위해 다양한 실험들이 추가되어야 할 것이다. 편향보정계수를 구하기 위해 서로 다른 통계적 방법을 적용했을 때 오차의 범위가 어느 정도 되는지, 그 오차를 포함한 자료가 실제 자료동화시스템에 쓰였을 때 영향이 얼마나 되는지에 관한 체계적인 실험을 수행해야 한다. 또한, 편향보정계수의 성능을 평가하기 위해 관측증분이 아닌 관측(O)과 분석(analysis; A)의 차이인 분석증분(increment; O-A)을 비교해야 좀 더 정확한 검증이 이루어질 수 있을 것이다.

### 감사의 글

본 연구는 기상청의 지원을 받는 한국형수치예보모델개발사업단의 연구과제를 통해 수행되었습니다.

### REFERENCES

- Atkinson, A., J. Cameron, B. Candy, and S. English, 2005: Bias correction of satellite data at the Met Office. *ECMWF/EUMETSAT NWP-SAF Workshop on Bias estimation and correction in data assimilation*, 9th November.
- Auligné, T., A. P. McNally, and D. P. Dee, 2007: Adaptive bias correction for satellite data in numerical weather prediction system. *Quart. J. Roy. Meteor. Soc.*, **133**, 631-642.
- Baker, N. L., T. F. Hogan, W. F. Campbell, R. L. Pauley, and S. D. Swadley, 2005: The impact of AMSU-A radiance assimilation in the U.S. Navy's Operational Global Atmospheric Prediction System (NOGAPS). *NRL Memorandum Report (NRL/MR/7530-05-8836)*, Naval Research Laboratory, Monterey, CA.
- Cardinali, C., 2009: Monitoring the observation impact on the short-range forecast. *Quart. J. Roy. Meteor. Soc.*, **135**, 239-250.
- Dee, D. P., 2004: Variational bias correction of radiance data in the ECMWF system. *Proceedings of the ECMWF workshop on assimilation of high spectral resolution sounders in NWP*, Reading, UK, June 28~July 1, 97-112.
- \_\_\_\_\_, 2005: Bias and data assimilation. *Quart. J. Roy. Meteor. Soc.*, **131**, 3323-3343.
- Eyre, J. R., 1992: A bias correction scheme for simulated TOVS brightness temperature. *ECMWF Tech. Memo*, **176**, ECMWF, Reading, UK.
- Goldberg, M. D., D. S. Crosby, and L. Zhou, 2001: The limb adjustment of AMSU-A Observations: Methodology and validation. *J. Appl. Meteorol.*, **40**, 70-83.
- Grody, N., F. Weng, and R. Ferraro, 1999: Application of AMSU for obtaining water vapor, cloud liquid water, precipitation, snow cover and sea ice concentration. *Technical Proceeding of the 10<sup>th</sup> International ATOVS Study Conference*, Boulder, CO, Jan 27~Feb 2, 230-240.
- \_\_\_\_\_, and Coauthors, 2001: Determination of precipitable water and cloud liquid water over oceans from the NOAA 15 advanced microwave sounding unit. *J. Geophys. Res.*, **16**, 2943-2953.
- Harris, B. A., and G. Kelly, 2001: A satellite radiance-bias correction scheme for data assimilation. *Quart. J. Roy. Meteor. Soc.*, **127**, 1453-1468.
- Hilton, F., N. C. Atkinson, S. J. English, and J. R. Eyre, 2009: Assimilation of IASI at the Met Office and assessment of its impact through observing system experiments. *Quart. J. Roy. Meteor. Soc.*, **135**, 495-505.
- Isaksen, L., 2012: The operational data assimilation system. *ECMWF Satellite Data Assimilation Training Course*.
- Jeong, D. H., C. Ziemkiewicz, W. Ribarsky, R. Chang, and C. V. Center, 2009: Understanding Principal Component Analysis Using a Visual Analytics Tool. *Charlotte Visualization Center, UNC Charlotte*.
- Jun, C.-H., M.-G. Jeong, and H.-S. Lee, 2004: *Applied Statistics for Engineers, Chapter 9 Multiple linear regression (Korean edition)*, 165-205.
- Kim, J.-H., J.-H. Kang, and S. Lee, 2014: A comparison of observed and simulated brightness temperatures from two radiative transfer models of RTTOV and CRTM. *J. Korean Earth Sci. Soc.*, **35**, 19-28.
- Kwon, S., 2004a: *Survey analysis (SAS utilizing SPSS), Chapter 6 Multicollinearity (Korean edition)*, 134-154.
- \_\_\_\_\_, 2004b: *Survey analysis (SAS utilizing SPSS), Chapter 10 Regression analysis (Korean edition)*, 211-239.
- Lee, S., J.-H. Kim, J.-H. Kang, and H.-W. Chun, 2013: Development of pre-processing and bias correction modules for AMSU-A satellite in the KIAPS observation processing system. *Atmosphere*, **23**, 453-470.
- Liao, D., and R. Valliant, 2012: Variance inflation factors in the analysis of complex survey data. *Sur. Meth-*

- odol.*, **38**, 53-62.
- Marill, K. A., 2004: Advanced statistics: Linear regression, Part II: Multiple linear regression. *Acad. Emerg. Med.*, **11**, 94-102.
- Rabier, F., H. Jarvinen, E. Klinker, J.-F. Mahfouf, and A. Simmons, 2000: The ECMWF operational implementation of four-dimensional variational data assimilation. I: experimental results and simplified physics. *Quart. J. Roy. Meteor. Soc.*, **126**, 1143-1170.
- Rawlins, F., and Coauthors, 2007: The Met Office global four-dimensional variational data assimilation scheme. *Quart. J. Roy. Meteor. Soc.*, **133**, 347-362.
- Smith, L. I., 2002: A tutorial on principal components analysis. [http://nyx-www.informatik.uni-bremen.de/664/1/smith\\_tr\\_02.pdf](http://nyx-www.informatik.uni-bremen.de/664/1/smith_tr_02.pdf).
- Wackerly, D., and W. Scheaffer, 2008: *Mathematical Statistics with Applications (7 ed.)*. Belmont, CA, USA, Thomson Higher Education. ISBN 0-495-38508-5.