Thesis for a Ph. D. Degree

Development of Regionally Focused Algorithm for AIRS Temperature and Humidity Retrievals Using a Moving-Window Technique

이동창 기법을 이용한 지역 중심 AIRS 온도/습도 연직 분포 산출 알고리즘 개발

August 2017

School of Earth and Environmental Sciences
Graduate School

Seoul National University

Hyun-Sung Jang

Development of Regionally Focused Algorithm for AIRS Temperature and Humidity Retrievals Using a Moving-Window Technique

By

Hyun-Sung Jang

A Dissertation submitted to the Faculty of the
Graduate School of the Seoul National University
in partial fulfillment of the requirements
for the Degree of Doctor of Philosophy

Degree Awarded:
August 2017

Advisory Committee:

Professor Sang-Woo Kim, Chair
Professor Byung-Ju Sohn, Advisor
Professor Kwang-Mog Lee
Professor Myoung-Hwan Ahn
Doctor Yoonjae Kim

이학박사학위논문

이동창 기법을 이용한 지역 중심 AIRS 온도/습도 연직 분포 산출 알고리즘 개발

Development of Regionally Focused Algorithm for AIRS Temperature and Humidity Retrievals Using a Moving-Window Technique

2017년 8월

서울대학교 대학원
지구환경과학부
장 현 성

Abstract

Regionally focused algorithm for Atmospheric Infrared Sounder (AIRS) temperature and humidity retrievals was developed. We first employed regression model with a moving window technique. This is done by relating the AIRS measurements to temperature and humidity profiles with consideration of regionally and seasonally changing local climatology. Regression coefficients were obtained from four-year (2006-2009) of ECMWF interim data over East Asia and simulated AIRS radiances. Result showing a notable improvement of mean biases, compared to the regression retrieval which does not consider local features, suggests that the moving-window technique can produce better regression retrievals by including the local climatology in the regression model.

For further improvement of the regression retrieval, one dimensional variational (1DVAR) physical model was also included in our algorithm. Error covariance matrix for the moving-window regression was obtained by using pre-developed regression retrieval and its error covariance. To assess the performance of 1DVAR using the mowing-window regression as a priori, error statistics of the physical retrievals from clear-sky AIRS measurements during four months of observation (March, June, September, and December of 2010) were compared; the results obtained using new a priori information were compared with those using a priori information from a global set of training data which are classified into six classes of infrared (IR) window channel

brightness temperature. This comparison demonstrated that the physical retrieval from the moving-window regression shows better result in terms of the root mean square error (RMSE) improvement. For temperature, RMSE improvements of 0.1 – 0.2 K and 0.25 – 0.5 K were achieved over the 150 – 300 hPa and 900 – 1000 hPa layers, respectively. For water vapor given as relative humidity, the RMSE was reduced by 1.5 – 3.5% above the 300 hPa level and by 0.5 – 1% within the 700 – 950 hPa layer.

As most of improvements due to use of the moving-window technique were shown in situations in which the relationship between measured radiances and atmospheric state is not clear, we investigated a possible use of surface data for further improving AIRS temperature and humidity retrievals over the boundary layer. Surface data were statistically and physically used for our AIRS retrieval algorithm. Results showing reduced RMSEs at both the surface level and the boundary layer, suggest that the use of surface data can help better resolve vertical structure of temperature and moisture near the surface layer by alleviating the influences of incomplete channel weighting function near the surface on the retrieval.

In conclusion, developing regionally focused algorithm, the inclusion of climate features in the AIRS retrieval algorithm can result in better temperature and humidity retrievals. Further improvement was also demonstrated by adding surface station data to the channel radiances as pseudo channels. Since the hyperspectral sounder is available on the geostationary platform, the development of regionally focused algorithm could

enhance its applicability to enhance our ability to monitor and forecast severe weather.

Keyword: Temperature profile, Humidity profile, IR hyperspectral sounder, AIRS, Linear regression model, 1DVAR

Student Number: 2008-20416

Table of Contents

List of Figures

1. Introduction

Temperature and moisture profiles are fundamental data for describing thermodynamic features of the atmosphere and thus accurate measurements of those parameters are prerequisite for better understanding and better prediction of the atmospheric phenomena. With the advent of satellite-based hyperspectral infrared measurement technologies such as Atmospheric Infrared Sounder (AIRS; Chahine et al. 2006), global pictures of three-dimensional temperature and moisture became available, recently yielding a high vertical resolution of ~1 – 2 km. One dimensional variational (1DVAR) based physical methods have been developed for retrieving relevant parameters in a manner consistent with both satellite measurements and a priori condition (Eyre 1989; Li and Huang 1999; Li et al. 2000; Susskind et al. 2003; Carissimo et al. 2005; Gambacorta 2013; Masiello and Serio 2013).

Retrievals from hyperspectral sounder have been widely used for achieving a variety of objectives. One of the major achievements from the satellite-based sounding is to extend our understanding of atmospheric phenomena from the weather to the climate (e.g. Dessler et al. 2008; Tian et al. 2010; Kahn et al. 2011 among many others). In particular, better understanding of the role of water vapor in the global climate system can be brought with these retrievals. Moreover, the retrievals are utilized for monitoring severe weather (i.e., Weisz et al. 2015) as well as for improving numerical weather prediction (Le Marshall et al. 2006; Hilton et al. 2012; Zheng et al. 2015 among

many others). In line with these, many attempts have been made to improve the physical methods (Li et al. 2007; Kwon et al. 2012; Smith et al. 2012; Bisht et al. 2015), but solving the retrieval problem remains a challenge (Grieco et al. 2011; Serio et al. 2015). Nonetheless, the knowledge obtained from these previous studies should have been useful for developing better data assimilation methods for numerical weather forecasting.

It has been long recognized that a priori information for constraining the physical method is essential for solving ill-posed inverse problems (Bouttier and Courtier 1999; Prunet et al. 2001; Zhang et al. 2014). One can use regression-based retrievals from satellite measurements for providing a priori information (Li et al. 2000; Kwon et al. 2012). In this case, regression retrievals are used as a priori but also used as a first guess in the physical method. Therefore, improving the regression method is considered to yield more accurate retrieval.

Different from this approach, regression retrievals can be used as first guess only while a priori information is obtained from climatology or numerical model outputs. For those, better regression is also thought to be an important way to improve the physical retrievals. Recent updates of AIRS and Infrared Atmospheric Sounding Interferometer (IASI; Blumstein et al. 2004) operational products (i.e., version 6 of AIRS L2 retrieval from National Aeronautics and Space Administration (NASA), and version 6 of IASI L2 retrieval from European Organisation for the Exploitation of Meteorological Satellites

(EUMETSAT)) reflect such efforts. In the case of AIRS version 6, training data sets were classified in terms of surface type and pressure, latitude, season, and day/night (Olsen 2013). This effort was based on the assumption that well-categorized regression coefficients can reduce the nonlinearity between observed radiance and underlying atmospheric state. Similarly, IASI version 6 introduced a piecewise regression method for generating better first guess (EUMETSAT[a] 2016). In the EUMETSAT method for IASI, training segments are stratified into 480 classes, such that each retrieval case is examined to determine the most appropriate class out of 480 classes using scan angle, solar zenith angle (i.e., day or night), surface altitude, radiances from IASI, and collocated radiances from Advanced Microwave Sounding Unit-A (AMSU-A) and Microwave Humidity Sounder (MHS).

If well-classified training data set is a key to improving regression algorithm, then a question may arise about whether regression methods based on regionally and seasonally varying regression coefficients can yield a better a priori information or first guess, compared to globally derived coefficients. Along this line of reasoning, a method (called “the moving-window technique” in this study) is introduced here to yield a better, regionally focused a priori information as well as first guess, for improving the 1DVAR based physical retrievals of clear-sky temperature and moisture profiles from AIRS infrared hyperspectral measurements. This idea is not completely new, and in fact, as shown in Thompson et al. (1985) and Chedin et al. (1985), the idea has been tested by

partitioning the global ensemble into various classes for improving the first guess fields.

This study first compares two types of regression retrievals from (1) regionally and seasonally driven regression coefficients obtained for the moving-window technique, and from (2) Cooperative Institute for Meteorological Satellite Studies (CIMSS) regression coefficients derived from a global training data set (Weisz et al. 2007; Weisz et al. 2013). The CIMSS regression uses the clear-sky training data set (Borbas et al. 2005) classified with scanning angles and six classes of the window channel brightness temperature (TB) (referred to as the TB-based classification technique) (Weisz et al. 2003). Comparison is made over the East Asian region, for four months spanning a year (i.e., March, June, September, and December of 2010). Then, regression retrievals obtained from the two methods are used as a priori information constraining the CIMSS 1DVAR based physical retrieval model (Li et al. 2000). The effect of different types of a priori information on the accuracy of physical retrieval is studied by comparing the error statistics of physical retrievals using the two sets of regression data against reference profiles.

Once we confirm that the moving-window technique is helpful for improving temperature and humidity soundings, we examine the impact of use of surface station data in the satellite-based retrieval for further improving regionally focused AIRS algorithm. In inclusion of the surface station data in the retrieval is interesting because those data should have much more information on in particular the surface than the

climatology provided by moving-window technique. As in the possible improvement of satellite retrieval from use of climatology information, the radiative signal delineating the boundary layer would be helpful if surface station data are used. Such effort may be helpful to improve the retrieval performance not only at surface level where surface station is located, but also over the boundary layer above the surface station.

In this study, two approaches are considered for the synergetic use of surface data in AIRS temperature and humidity retrievals. The impact of surface data is first evaluated through including those data in a regression model. In doing that, a new regression model is developed by using the surface data as additional predictors. Obtained results will be compared with results from the regression model without surface data. This comparison provides an opportunity to examine the vertical extent which the surface data might influence as well as the magnitude of the improvement in a statistical perspective. As the second approach, direct use of surface data for a 1DVAR is considered. In this approach, surface data are used as if additional channels. Note that the Jacobians linking the variation of surface data to the variation of temperature and moisture profiles should be defined, along with introduction of the error covariance matrix for surface data. The proposed two approaches will be tested in this study over the Korea peninsula using surface observations at Automatic Weather Station (AWS) stations maintained by Korea Metrological Administration.

In summary, the main objective of this research is to develop regionally focused

algorithm for better retrieving temperature and moisture profiles. In doing so, a synergetic use of climatology features and surface station data is proposed. The procedures include the development of a new regression model and construction of the error covariance and Jacobian for 1DVAR. Results will be validated against retrievals from pre-developed algorithm as well as against re-analysis data. We hope that this study leads to better ability of monitoring and forecasting severe weather phenomena through the improved data assimilation.

The dissertation is organized as follows. In section 2, we review basic understanding of temperature and moisture sounding, including the regression method and the physical method (i.e., 1DVAR). Here, various documented previous studies are examined. Then detailed information on AIRS measurement as well as other infrared hyperspectral sounders is provided in section 3. In section 4, a method (called "a moving window technique" in this study) which uses regional and seasonal climatology for temperature and moisture soundings is introduced and its benefits are discussed by comparing with pre-developed algorithm and re-analysis data. In section 5, the effects of moving-window technique on the physical retrieval are described. Then, in section 6, possible use of surface data for better retrieval of temperature and moisture profiles from AIRS radiance measurement will be examined. Summary and our conclusion follow in section 7.

2. Review of previous satellite-based temperature and humidity soundings

There has been a huge amount of studies for temperature and humidity soundings from satellite measurements. Here we review documented satellite-based temperature and humidity retrieval methods as well as general understanding of atmospheric sounding from IR measurements.

In a nonscattering atmosphere, observed radiance R_v for a given wavenumber v is

$$R_v = \varepsilon_{s,v} B_{s,v} \tau_{s,v} + \int_{p_s}^{0} B_v d\tau_v + (1-\varepsilon_{s,v}) \int_{0}^{p_s} B_v d\tau_v^* + R' \quad (1)$$

where subscript s denotes surface, B is the Plank radiance, ε is the surface emissivity, τ is the upward transmittance, p is the surface pressure, $\tau^* = \tau_s^2/\tau$ is the downwelling transmittance, and R' is the reflected solar radiation. Four terms on right hand side are surface contribution term, upwelling component of atmospheric contribution term, reflected downwelling component of atmospheric contribution term, and reflected solar energy contribution term on observed radiance, respectively. Last term, R', is generally hard to be simulated by using radiative transfer model, thus we avoid the spectral regions where R' term is dominant. Since equation (1) is a function of temperature and τ, which is a function of absorption coefficient and absorber amount, principle of temperature and moisture sounding can be explained from equation (1). In case of CO_2 channels, for example, absorber amount and absorption coefficient are well known, so

temperature information can be retrieved. In addition, if we use various CO_2 channels whose absorption coefficients are different, we can retrieve temperature profile also. This is because lower and higher atmospheric information are included in weak and strong absorption channels, respectively; weak absorption channels detect radiation coming from lower atmosphere, whereas strong absorption channels only detect radiation coming from higher atmosphere. Figure 1 provides examples of thermal emission spectra from Hanel et al. (1971) to explain how to change the shape of the infrared spectra for various atmospheric situations. It can be seen that temperature structure is well reflected in CO_2 band area. Similarly, when temperature and absorption coefficient are known, absorber amount can be extracted. From this principal, it is possible to retrieve amount of atmospheric water vapor. Besides those, in case of window channels, terms related to gas absorption (e.g., second and third terms) can be neglected and $\tau_{s,v}$ is close to unity, so those channels have ability to retrieve surface information. These are the theoretical reasons why hyper-spectral sounder, which observes almost whole thermal spectra, can resolve high resolution temperature and moisture profiles.

From a practical point of view, solving the equation (1) is an ill-posed problem. One could solve this problem by a regression method. The merits of regression method are physically straightforward to understand principle of method and very quickly retrieve the atmospheric state once well-trained **C** matrix is obtained.

As very early work, Smith et al. (1970) has investigated a least squares regression method for obtaining global temperature profile from satellite radiation measurements, particularly those obtained by the Satellite Infra-Red Spectrometer (SIRS) aboard the Numbus 3 satellite. This study is followed by previous works such as Wark and Fleming (1966), Rodgers (1966), Westwater and Strand (1968), which indicate that maximum information about the atmosphere's thermal structure may be derived from satellite radiation observations through the use of statistical relationship. In the regression method, atmospheric state vector **X** is retrieved from $\mathbf{CY^T}$, where **C** is regression model and **Y** is predictors (e.g., observed radiances). Best fitting solution of $\mathbf{X = CY^T}$ is the one that minimizes $\sum\mathbf{(X-CT)^2}$, and minimization yields $\mathbf{C = XY(Y^TY)^{-1}}$. In order to obtain **C,** various pairs of **X** and **Y** which cover possible relations between atmospheric states and satellite observations are needed. Basic concept of regression method is quite straightforward, but lots of effort to improve regression model is tried.

In Smith and Woolf (1976), new technique is formulated for using eigenvectors of covariance matrices to retrieve temperature and moisture profiles for 17-channels of High-resolution Infra-Red Sounder (HIRS) measurements and 5-channels of SCAnning Microwave Spectroradiometer (SCAMS) measurements. Since eigenvectors of covariance matrices provide the most economical representation of any variable, computational power requirements are much less than those of other regression methods. Moreover, the effects of random errors of observation on the analytical solutions are

greatly suppressed since the system of equations to be solved is heavily over-determined.

With advent of satellite measurements, the number of channels sufficiently increase. Goldberg et al. (2003) noted that principal component analysis (PCA) is an effective way for thinning AIRS hyperspectral measurements. PCA is also used for the quality control. Similarly, the approach was evaluated for AIRS in Weisz et al. (2007), but in this study, Moderate Resolution Imaging Spectroradiometer (MODIS) derived cloud amounts are additionally used for the cloud screening. Not only temperature and moisture profiles, but also ozone profile, surface temperature and surface emissivity are simultaneously retrieved in Weisz et al. (2007). In order to obtain the better regression relationship, a scene classification method by using brightness temperature from window channels and satellite viewing angle was applied.

Recent update of operational products (i.e. NASA product version 6 for AIRS, or EUMETSAT product version 6 for IASI) showed efforts for improving regression retrievals. For the AIRS operational product, an artificial neural network (ANN) technique was adopted for minimizing a non-linearity between the radiance and the atmospheric state. Also, training segments were stratified in terms of surface type and pressure, latitude, season, and day/night (Olsen 2013). For the IASI operational product, training segments were stratified into 480 classes, such that each retrieval case is examined to determine the most appropriate class out of 480 classes using scan angle, solar zenith angle (i.e., day or night), surface altitude, radiances from IASI, and

collocated radiances from Advanced Microwave Sounding Unit-A (AMSU-A) and Microwave Humidity Sounder (MHS). Moreover, quality flag which delivers, to certain extent, the accuracy of regression product was simultaneously retrieved with atmospheric parameters such as temperature, moisture, and ozone profiles, surface pressure, and surface temperature.

Because of the high non-linearity in the relationship between water vapor amount and satellite measurements, a method for describing water vapor amount was important issue for improving regression method. Weisz et al. (2007), Smith et al. (2012), and Weisz et al. (2013) used a natural logarithm of mixing ratio for describing the water vapor amount. Since Li (1994) showed that change of water vapor in natural logarithm of mixing radio is proportional to change of brightness temperature from satellite measurement, those approaches are quite reasonable. However, the operational product of IASI uses dew point temperature for describing the water vapor amount, based on the empirical sensitivity test result. In case of Thapliyal et al. (2012), hybrid-regression retrieval was proposed. This study demonstrated that the use of the natural logarithm of humidity as a predictnand gives better accuracy for the drier atmosphere, whereas direct humidity as a predictand provides better accuracy for the wetter atmosphere. In addition to the unit issue, the vertical resolution aliasing issue caused by layer observation problem of satellite measurement was newly studied in Smith et al. (2015). Besides this, a new effort by August (2015), who uses the 3-dimensional atmospheric structure in

regression procedure, also indicated that improvement of regression method is on-going problem.

The other method to solve the ill-posed problem is physical method which finds a solution from minimizing the following cost function

$$J(\mathbf{X}) = [\mathbf{Y^m} - \mathbf{Y}(\mathbf{X})]^{\mathrm{T}} \mathbf{R}^{-1} [\mathbf{Y^m} - \mathbf{Y}(\mathbf{X})] + (\mathbf{X} - \mathbf{X}_0)^{\mathrm{T}} \mathbf{B}^{-1} (\mathbf{X} - \mathbf{X}_0) \qquad (2)$$

where superscript T denotes the transpose, **X** is the candidate of solution, $\mathbf{Y^m}$ is the observed radiances, **Y**(**X**) is the simulated radiances of candidate of solution, $\mathbf{X_0}$ is the background that constrains the solution, **B** is the error covariance of background, and **R** is the error covariance of observed radiances that includes instrument noise and forward model error. Theoretically, **X** and **Y** do not have bias and **R** and **B** are uncorrelated in order to obtain the optimal solution. Otherwise, the solution will be suboptimal. Often $\mathbf{X_0}$ is obtained from numerical weather model, a climate mean, retrievals from regression model. On the right hand side, first term and second term indicates that how far the current solution, **X**, is away from observed radiances and background, respectively, and each term is normalized by **R** and **B**. When $J(\mathbf{X^*})$ is a minimum of the function *J*, we assume **X*** as a solution that is consistent with both observed radiances and background. This physical approach for the satellite retrieval was firstly proposed by Rodgers (1976).

In Eyre (1989), a new approach was proposed for inverting satellite sounding data for use in numerical weather prediction based on Rodgers's study. In this study, the

application of the scheme to the data from the TIROS Operational Vertical Sounder (TOVS) was discussed. According to Eyre (1989), by using the following Newtonian iteration

$$\mathbf{X}_{n+1} = \mathbf{X}_n - J''(\mathbf{X}_n) \cdot J'(\mathbf{X}_n)^{-1} \tag{3}$$

the following quasi-nonlinear iteration form could be obtained,

$$\delta\mathbf{X}_{n+1} = (\mathbf{F}_n^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{F}_n + \mathbf{B}^{-1})^{-1}\mathbf{F}_n^{\mathrm{T}}\mathbf{R}^{-1}(\delta\mathbf{Y}_n + \mathbf{F}_n\delta\mathbf{X}_n) \tag{4}$$

where $\delta\mathbf{X_n} = \mathbf{X_n} - \mathbf{X_0}$, $\delta\mathbf{Y_n} = \mathbf{Y^m} - \mathbf{Y}(\mathbf{X_n})$, $\mathbf{F}$ is the linearized Jacobean matrix which satisfies $\delta\mathbf{Y} = \mathbf{F}\delta\mathbf{X}$.

Although the nonlinear problem could be partly solved by 1DVAR and physical understanding of retrieval procedure could be possible, a speed of the 1DVAR is one of the issues for the evaluation. Different from Rodgers (1976) and Eyre (1989) approaches (so-called inverse Hessian method), which uses a Laplacian of the cost function as a slope of the iteration, the steepest decent method was also reported in Levenberg (1944) and Marquardt (1963). In these methods, a smoothing factor was used. Generally, this smoothing factor was chosen empirically (Twomey, 1963). In Susskind et al. (1984), 0.0005 was chosen, while smoothing factor of 0.1 was chosen in Smith et al. (1985) and Hayden (1988). Although there is no clear and definite way to select the smoothing factor, empirically obtained constraint successfully controlled the 1DVAR system. To keep the benefit of both Hessian and steepest decent method, combination of Hessian method and steepest decent method was also tried (Ma et al., 1999). In the

combined algorithm, the solution tends to an inverse Hessian method if the smoothing factor approaches to 0, while the solution tends to the steepest decent method with a smaller step size if the smoothing factor approaches to infinite.

Since a major reason of adopting the smoothing factor is uncertainty of the background error, those previous studies also indirectly implied importance of the background error in 1DVAR system. In case of Li and Huang (1999), they developed the discrepancy principle to find a smoothing factor physically. They proved that the smoothing factor is unique for given observation, and it could be estimated by comparing magnitude of background departure and error of observation which includes measurement noise error and radiative transfer model error. Based on the discrepancy principle, 1DVAR algorithm for TOVS was reported in Li et al. (2000). It is interesting to note that an efficient form of radiative transfer equation introduced in Li (1994) was successfully applied in this TOVS retrieval algorithm. It suggests that not only numerical iteration and convergence technique, but also radiative transfer theory are simultaneously advanced with the history of 1DVAR evaluation.

Although there have been lots of effort for the evaluation of 1DVAR, solving retrieval problem is still challenging and many advanced methodologies are tried for improving 1DVAR. In Grieco et al. (2011), correlation interferometry, which is a particular application of Fourier transform spectroscopy with partially scanned interferograms, was analyzed for deeper understanding of the radiative signal. They

separated the atmospheric emission from the background by using partially scanned interferograms and they found that the vertical spatial resolution of the retrieved water vapor profile is only determined by the broad feature in the spectrum (Figure 2). Considering that Jacobian in 1DVAR is the core of cost minimization, such effort could be very helpful for explaining the atmospheric signal in hyper-spectral measurements. Use of cloud-cleared radiance for removing cloud contaminated signal in brightness temperature could also be considered as advanced 1DVAR (Susskind et al. 2003; Gambacorta 2013). In those algorithms, a portion of radiative signal caused by cloud is simultaneously analyzed with retrieving atmospheric parameters. Besides those, use of surface station data (Liu et al. 2014) or use of numerical weather prediction (Jin et al. 2008; Schmit et al. 2008; Li et al. 2009) could also be considered as advanced efforts for improving 1DVAR.

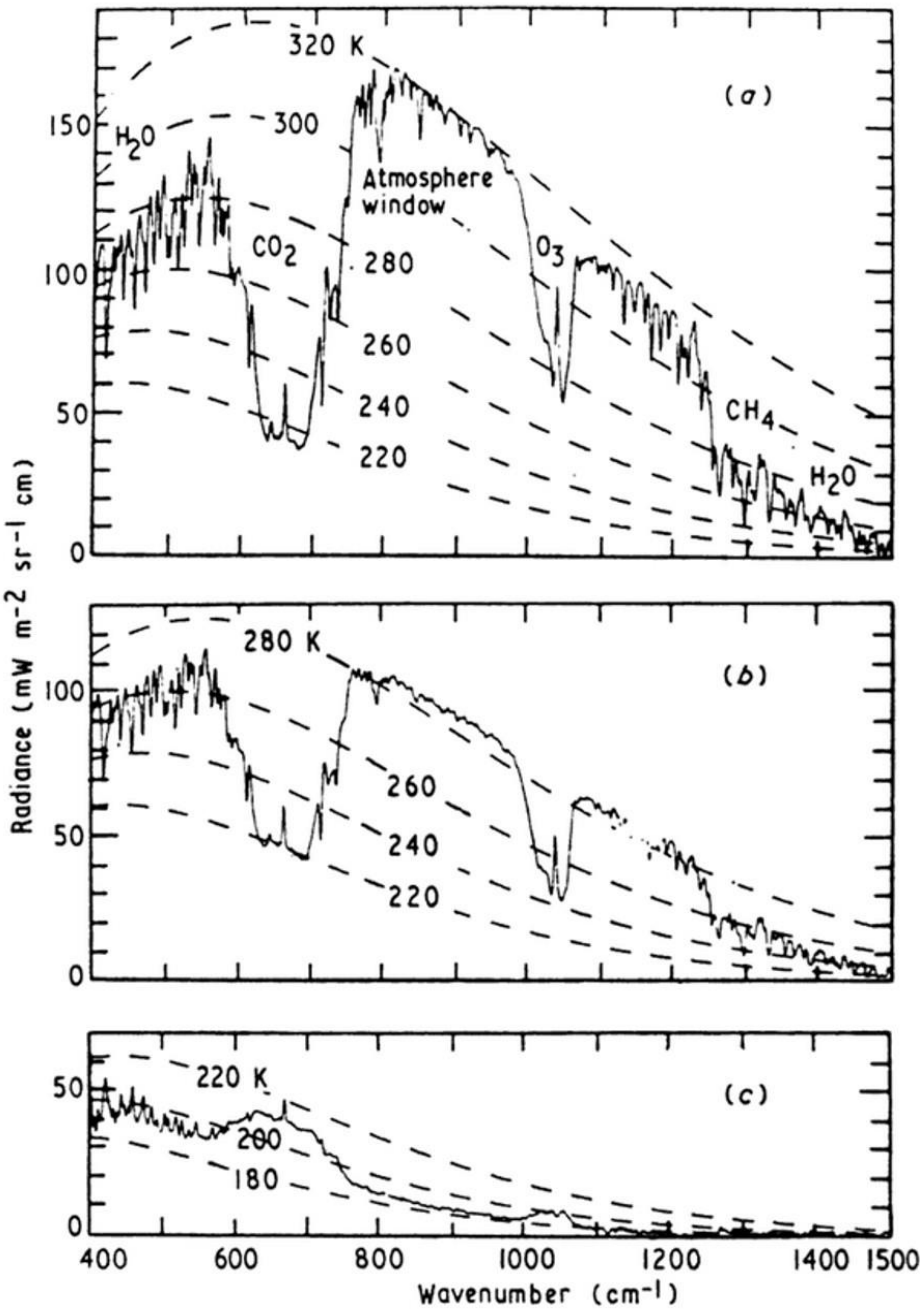

Figure 1. Examples of thermal emission spectra recorded by IRIS D on Nimbus 4. Radiances of black bodies at several temperatures are superimposed. (a) Sahara; (b) Mediterranean; (c) Antarctic; (all apodized) (Hanel et al. 1971).

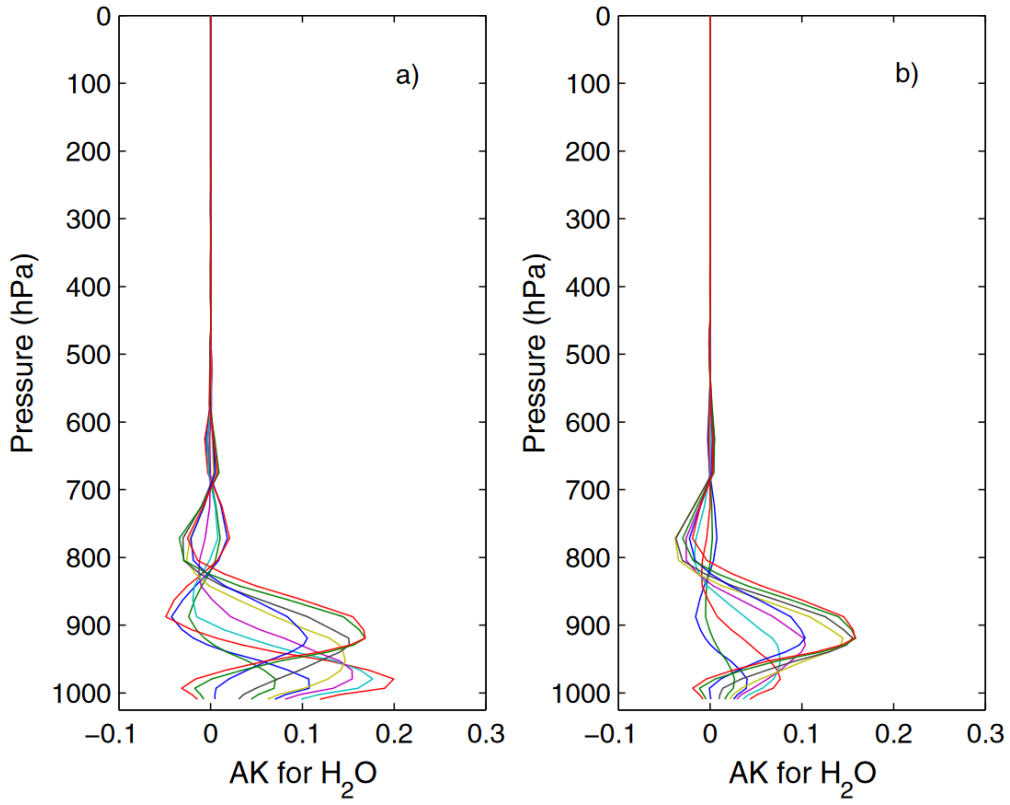

Figure 2. Example of IASI averaging kernels for H_2O for a tropical atmospheric model for the pressure range 1025 to 900 hPa. a) Averaging kernels computed considering the full IASI spectral coverage and channels. Panel b) is the same as a), but now the averaging kernels have been computed with a partial interferogram extending in the range of [0.0090, 2] cm. The results have been obtained with the Masuda emissivity for the sea surface (Grieco et al. 2011).

3. Infrared hyperspectral measurements

There has been a huge amount of studies for temperature and humidity soundings from satellite measurements. Here we review documented satellite-based temperature and humidity retrieval methods as well as general understanding of atmospheric sounding from IR measurements. Here, we provide detailed information of AIRS used in this study. AIRS is a first hyper-spectral IR instrument on board the Aqua satellite, launched on May 4, 2002 (Aumann et al. 2003; Chahine et al. 2006). It was built by NASA Jet Propulsion Laboratory, and is a cross-track scanning instrument. Its scan mirror rotates around an axis along the line of flight and detects IR energy from the Earth. As the spacecraft moves along, this mirror sweeps the ground creating a scan swath that extends roughly 1650 km. The scan geometries of AIRS and its partner microwave instruments, AMSU-A, and the Humidity Sounder for Brazil (HSB), are illustrated in Figure 3. Aqua is in a sun-synchronous orbit with an ascending node at 1:30 pm local time. It covers the globe two times (daytime and nighttime) a day, but some points near the equator are missed since the swaths do not overlap at low latitudes. The spatial resolution is 13.5 km at nadir and 41 × 21.4 km at the scan extremes. Within the AIRS instrument the IR energy is separated into wavelengths, similar to rain drops splitting sunlight into a rainbow. By having multiple IR detectors, each senses radiance at a particular wavelength with 2378 detectors covering the IR spectral range of 3.8–4.6 μm, 6.2–8.2 μm, and 8.8–15.4 μm at a nominal spectral resolution of $\lambda/\Delta\lambda = 1200$.

Radiometric sensitivity expressed as the noise equivalent differential temperature (NedT) for a scene temperature of 250 K is 0.07–0.4 K for 3.75–11 μm, and 0.27–0.68 K for 11.75–15.4 μm.

Absorption by gases in the atmosphere is mainly a feature of the IR bands, and thus each IR wavelength is quite sensitive to temperature and other gases. The AIRS upwelling spectrum suggests that an AIRS spectrum contains gaseous absorption bands by water vapor, CO_2, O_3, and other trace gases. In addition to the main absorption bands, it is noticed that the atmospheric gases are nearby transparent in the vicinity of 11 μm, referred to as an 'IR window region' which mainly reflects surface conditions. Using information on the window and absorption regions over an AIRS spectrum, temperature and water vapor profiles have been officially produced and distributed with the purpose of improvement of weather forecasts and climate researches. Not only the atmospheric profile retrieval, the extremely narrow spectral resolution allows us to resolve vertical distribution of other parameters such as cloud and dust.

After the AIRS mission, various IR hyperspectral missions have been continued. The Infrared Atmospheric Sounding Interferometer (IASI) is another success of IR hyperspectral mission (Blumstein et al. 2004). IASI on board the Metop-A satellite, launched on 19 October 2006, is the first polar-orbiting interferometer and the second IASI on Metop-B is also successfully launched on 17 September 2012. It measures infrared part of the electromagnetic spectrum at a horizontal resolution of 12 km over a

swath width of about 2200 km. Global observations of 3.62 to 15.5 μm with 8461 channels can be provided twice a day with 14 orbits in a sun-synchronous mid-morning orbit (9:30 equatorial passing time, descending node). The third instrument will be mounted on the Metop-C and launched in 2018.

The Cross-track Infrared Sounder (CrIS), on aboard the Suomi National Polar-orbiting Partnership (NPP) spacecraft launched on October 28, 2011, is also a recent success of IR hyperspectral mission (Bloom 2001; Han et al. 2013). It is a Fourier transform spectrometer with 1305 spectral channels and produces high-resolution, three-dimensional temperature, pressure, and moisture profiles like AIRS and IASI. 1305 spectral channels have 3 groups: LWIR group (4.14 – 15.38 μm); MWIR group (5.71 – 8.26 μm); SWIR group (3.92 – 4.64 μm) group. Compared to previous IR hyperspectral sounders such as AIRS and IASI, much better instrumental noise level is reported (Figure 4). In order to maintain the continuity of observations from CrIS, Joint Polar Satellite System-1 (JPSS-1) will follow the Suomi NPP satellite, planned to launch in 2017.

With the advent of satellite measurement technique, geostationary hyperspectral sounding missions are proceeding. The Geostationary Interferometric Infrared Sounder (GIIRS) on Feng-Yun – 4 (FY4) satellite, launched by China Meteorological Administration, is a first geostationary hyperspectral sounder. As other hyperspectral sounders, main mission of the GIIRS is temperature and humidity profiling with high-

resolution. However, benefit of continuous observation drastically increases with increased temporal resolution. Moreover, new product such as wind profile by tracking water vapor features could be available from GIIRS. The GIIRS covers two spectral ranges with 913 channels: 8.85 – 14.3 μm with 0.8 cm^{-1} spectral resolution; 4.44 – 6.06 μm with 1.6 cm^{-1} spectral resolution. NEdR of the former is 0.5 mW m^{-2} sr^{-1} cm while that of the latter is around 0.1 mW m^{-2} sr^{-1} cm. In order to detect cloud in daytime, GIIRS also carries visible channels (0.55 – 0.75 μm). At the first version of FY4 (i.e., FY-4A), spatial resolution is around 16 km but it will be improved to 8 km in next series of FY4.

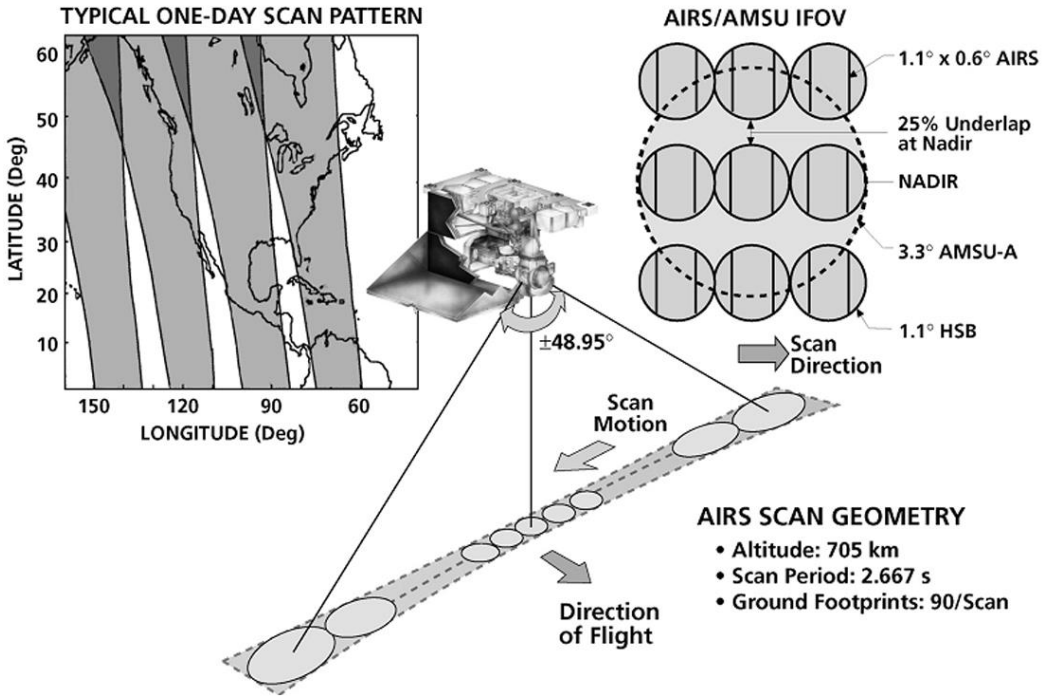

Figure 3. AIRS scan geometry and the relative alignment of the AIRS/AMSU/HSB footprints (Aumann et al., 2003).

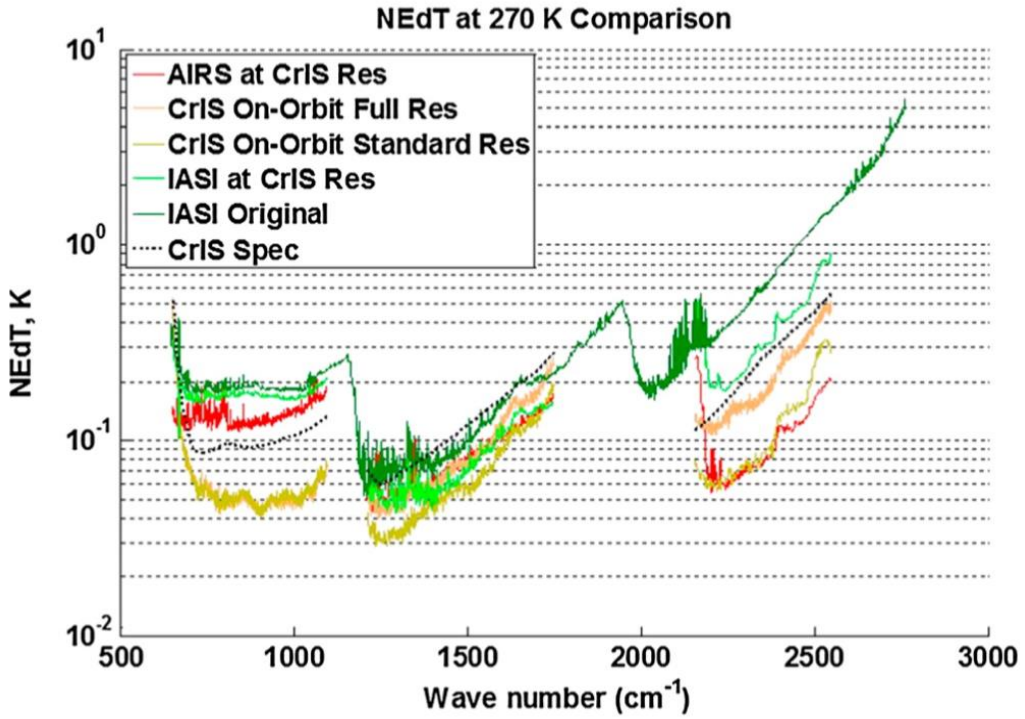

Figure 4. CrIS on-orbit nominal NEdT is compared with AIRS (at CrIS spectral resolution) and IASI NEdT estimated at IASI original and CrIS-like spectral resolution. The black curve depicts CrIS spec requirements. NEdT was estimated for scene temperature of 270 K (Zavyalov et al. 2013).

4. Development of regionally focused regression model

We first try to develop an improved regression model by considering seasonal and regional climatology features in the regression procedure. General steps for developing regression model are described in Figure 5. Starting from collecting data for simulation, radiance simulation, noise adding procedure, radiance compression, training of regression coefficients, and testing of regression model are included for regression model development.

4.1. Construction of training data

In developing a statistical regression model, a set of training data is necessary, from which predictand parameters (here temperature and moisture profiles) can be related to predictor variables (i.e., AIRS radiances and surface pressure). In addition, surface pressure is used as a predictor variable for maximizing the retrieval performance (Weisz et al. 2007; Thapliyal et al. 2012; Olsen 2013).

Here we simulate expected AIRS radiances using four-year-long (2006–2009) European Center for Medium-Range Weather Forecasts (ECMWF) Interim reanalysis data (referred to as ERA-I; ECMWF 2009) as inputs to the radiative transfer model; these inputs are the temperature, moisture, ozone profiles, surface temperature, and surface pressure. Since ECMWF interim data are one of the reanalysis data with well-

defined physical consistency between atmospheric parameters, and have very high accuracy about 1K for temperature and 0.5 g kg^{-1} for humidity, we could assume that simulation dataset based on ERA-I are able to reflect atmospheric variability realistically. Temporal resolution is six hour (00, 06, 12, 18 UTC), and spatial resolution is 1.5 ° with fixed 37 pressure levels (1000, 975, 950, 925, 900, 875, 850, 825, 800, 775, 750, 700, 650, 600, 550, 500, 450, 400, 350, 300, 250, 225, 200, 175, 150, 125, 100, 70, 50, 30, 20, 10, 7, 5, 3, 2 and 1 hPa). ERA-I data are publicly available on the ECMWF Data Server (http://data-portal.ecmwf.int/data/d/interim_daily/). Details of ERA-I data are described at Dee et al. (2011).

The regression method is developed over the East Asian analysis domain (15°N – 55°N, 90°E – 150°E). Considering horizontal resolution and temporal resolution with research domain, about 4500 samples are available for one day. However, all data are not required because lots of samples are duplicated. Therefore, we try to avoid unnecessary duplication by using random selection. Surface emissivity is also used to enhance the simulation accuracy and the data are obtained from the global infrared surface emissivity database produced by CIMSS (Borbas et al. 2007; Seemann et al. 2008). One may construct the training data for the regression from satellite measurements collocated with reference data. However, this method is prone to errors related to not well-known sensor characteristics and collocation mismatch. On the other hand, although the simulated brightness temperatures are not true, due to inaccurate

radiative transfer modeling and specification of measurement errors, the simulation approach has an advantage of avoiding the collocation problem and thus being easily equipped with known reference data for constructing a priori as well as for validation.

For the clear sky radiance simulation, Stand-Alone Radiative Transfer Algorithm (SARTA; Strow et al. 2003) is used, which ensures high-accuracy fast simulations of AIRS radiance. It was reported that uncertainty of the SARTA method results is ~0.2 K for CO_2 absorption channels. For H_2O absorption channels, uncertainty is reported to be mostly less than 1 K. All 2378 AIRS channels are simultaneously simulated using the SARTA model, except the channels that exhibit noisy behavior, e.g., owing to significant instrumental noise, non-Gaussian noise distribution, or poor spectral response function (Weisz et al. 2007). After subjectively identifying and removing these noisy channels, 1435 channels are retained for channel simulations. Figure 6 shows one example of brightness temperature spectrum and instrumental noise for all AIRS channels and selected channels. Although hundreds of channels are removed, selected channels cover whole spectral regions related to CO_2, window, ozone and H_2O. Twenty-eight shortwave infrared (SWIR) channels are also retained, but are only used for nighttime retrieval analysis; thus, regression analysis is performed separately for daytime and nighttime.

Random instrumental noise is added to the simulated radiances, based on the assumption that instrumental noise is Gaussian-distributed, with a standard deviation of

noise-equivalent differential temperature. Figure 7 shows error statistics of noisy radiances which follow our assumption; mean bias is close to zero and root mean square of differences is close to NedT.

Although the number channels are sufficiently reduced from channel selection, information between channels may be still correlated. Therefore, the principal component analysis (PCA) has been applied to compress the 1435 radiances (Smith and Woolf, 1976). In order to apply the PCA on radiances, we need to define one subset which reflects almost all of variability of radiances. Then, we subtract the mean radiances from each individual radiance, and calculate their covariance. After applying eigenvector analysis on the covariance matrix, 1435 eigenvectors and corresponding eigenvalues are calculated. When we sort eigenvalues in descending order, values are very quickly decreasing; starting from 110000, second to sixth eigenvalues are 2600, 1800, 500, 380, 100 for one subset. Since eigenvalue reveals that how much covariance can be explained by its associated eigenvector, we could expect that small portions of eigenvectors are good enough to explain variability of radiances. In other words, the small number of projection coefficients (PCs), which is the value when we project radiances on the eigenvectors, instead of the large number of radiances could be substituted for retrievals problem. To determine the number of eigenvectors for retrievals, we examine the sensitivity between errors of regression retrievals and the number of eigenvectors (Figure 8). Here, errors are defined by vertically cumulative

RMS errors. Generally, both RMS errors of temperature and moisture decrease as the number of eigenvectors increases, but the slope decreases. As a result, when the number of eigenvectors exceeds forty, accuracy is almost constant. Therefore, we choose forty as the number of eigenvectors for radiance compression. In addition, we perform same test but with noise-free radiances to examine impact of instrumental noise on retrievals. For noise-free radiances, errors sharply decrease until first forty eigenvectors. However, errors continuously decrease as the number of eigenvectors increases, whereas errors of retrievals with noisy radiances slightly increases; for example, temperature RMSE difference between use of 100 eigenvectors and 40 eigenvectors is 0.0033 (=0.9335–0.9302) for experiment with noisy radiances while –0.0068 (=0.8193–0.8261) for experiment with noise-free radiances. As a results, we expect that instrumental noises are included after around fortieth eigenvector. It is reconfirming that use of PCA on radiances could reduce instrumental noise signal in noisy radiance (Huang and Antonelli, 2001).

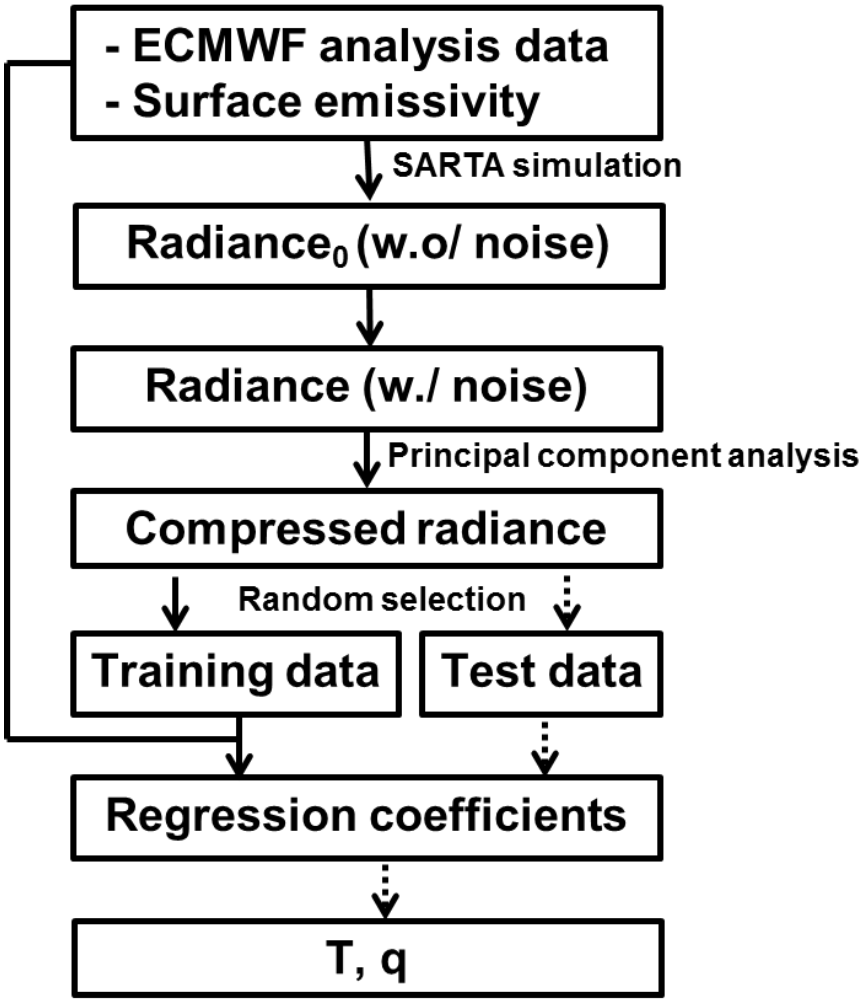

Figure 5. Schematic diagram showing procedures of developing a regression method

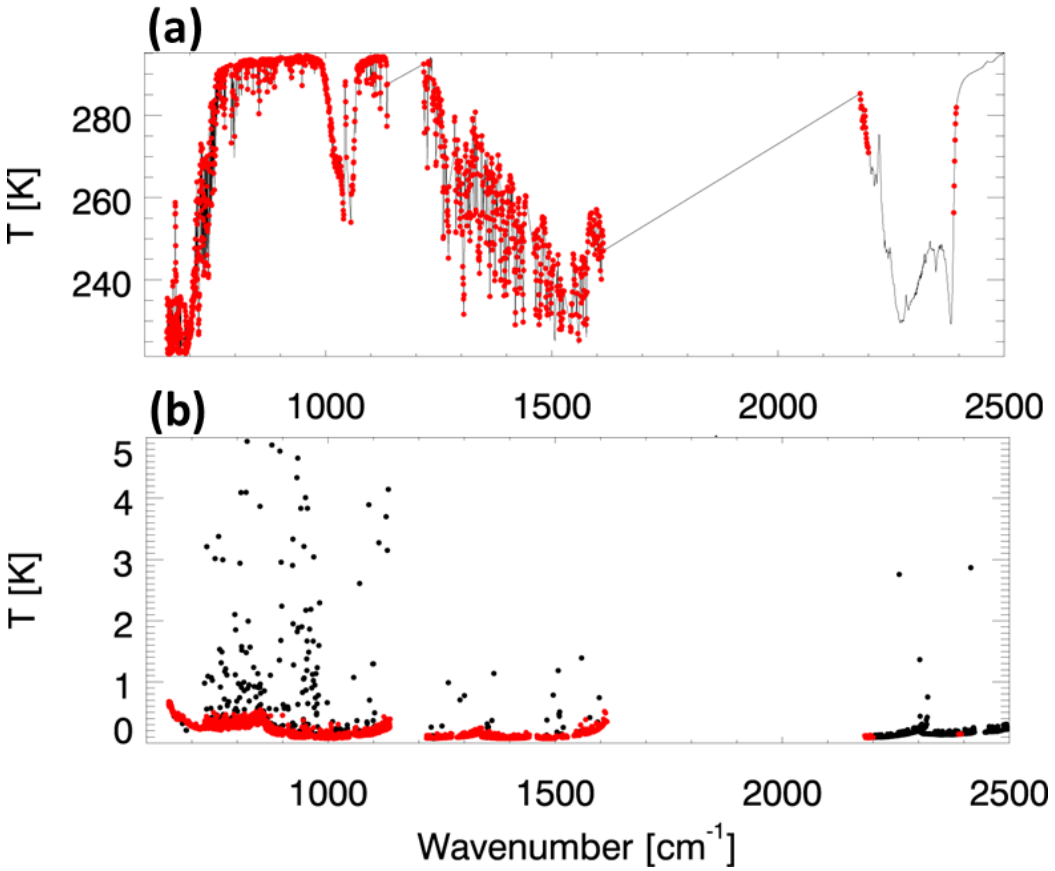

Figure 6. One case of (a) AIRS spectrum and (b) instrument error. Black dots indicate original 2378 AIRS channels while red dots indicate 1435 selected channels based on instrument noise.

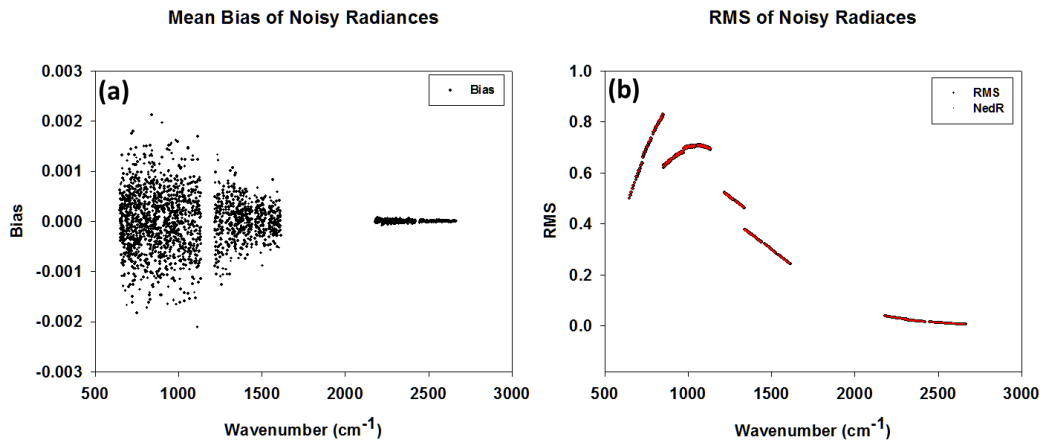

Figure 7. (a) Mean bias and (b) RMS of noisy radiances. Mean bias and RMS are defined by an average of noisy radiances minus noise-free radiances and root mean square of noisy radiances minus noise-free radiances, respectively. In RMS plot, red dots are NedR which corresponds to given NedT

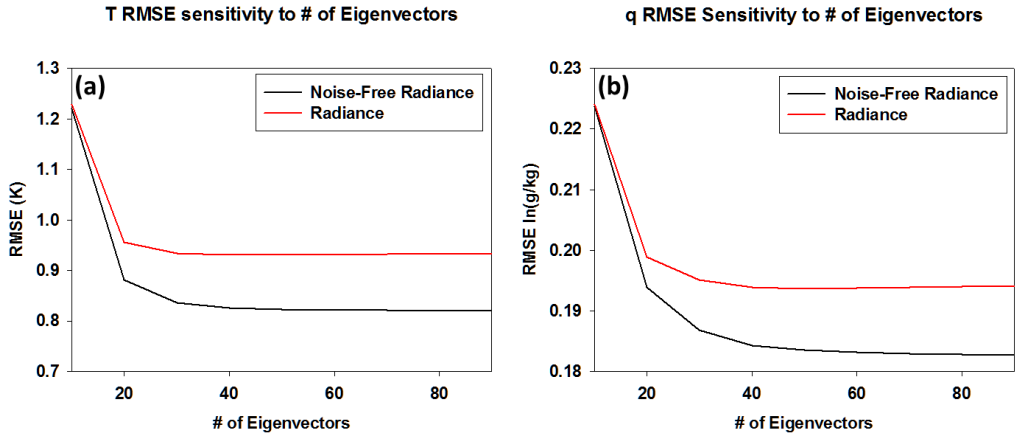

Figure 8 Sensitivity between the number of eigenvectors for compressing the radiances and RMSE of regression retrievals for (a) temperature and (b) moisture. In sensitivity test, black and red lines applied noise-free radiances and noisy radiances, respectively.

4.2. Moving-window regression model

Adopting a linear regression model (Smith and Woolf 1976; Weisz et al. 2007), we apply the moving-window technique for classifying the training data set in terms of their region and season. As mentioned in section 3.1, for a given set of predictor variables (here a vector $\mathbf{Y_R}$ of 41 parameters, containing data for 40 compressed radiances plus one surface pressure), the predictand $\mathbf{X_R}$ can be retrieved as follows:

$$\mathbf{X_R} = \mathbf{C}\mathbf{Y_R}^{\mathrm{T}} \tag{5}$$

The regression model finds the solution by minimizing $\sum(\mathbf{X_R} - \mathbf{C}\mathbf{Y_R}^{\mathrm{T}})^2$ where the regression coefficient **C** is defined as follows:

$$\mathbf{C} = \mathbf{X_R}\mathbf{Y_R}(\mathbf{Y_R}^{\mathrm{T}}\mathbf{Y_R})^{-1} \tag{6}$$

To obtain regionally-based regression coefficients for the temperature and moisture retrievals, we define a moving window as a 10° × 10° box. Figure 9 exemplifies the retrieval procedure by showing the analysis domain, two adjacent windows for the retrieval, and associated two boxes for training. For retrieval for Box 1, the regression relationship is trained on a larger 20° × 20° training Box 1, concentric with the retrieval Box 1. After completing the retrieval for Box 1, the regression window is moved to the retrieval Box 2, where retrieval is performed with another regression relationship obtained from the 20° × 20° training Box 2. In this way, spatial discontinuity between neighboring windows is minimized. This procedure is repeated until the retrieval is

completed over the entire analysis domain. Seasonal variations are incorporated into the regression by training on different seasons (here, the four seasons of March–May, June–August, September–November, and December–February). To reduce the temporal discontinuity between two adjacent seasons, the regression for each season is conducted over a period of five months by adding one month before the season and another month after the season in a target; for example, regression coefficients for spring (March–May) are obtained from the February–June period. Thus, regression coefficients relevant to each AIRS pixel are obtained by considering the pixel location and the associated season.

In addition, the effect of the scan angle (θ) is also included in the regression model; 11 scanning angles from 0° to 49° are considered, following Weisz et al. (2007):

$$\sec(\theta) = 1 + \Delta \cdot j \tag{7}$$

where $\Delta = 0.0524$, and $j = 0, 1, 2, \ldots, 10$. Applying this method to AIRS measurements, retrieval at a specific θ is obtained by linearly interpolating retrievals at two adjacent viewing angles, with respect to the relative air mass associated with the viewing angles. The retrievals obtained using the moving-window regression method are henceforth denoted by $\mathbf{X}_{\mathbf{R}}^{\mathbf{W}}$.

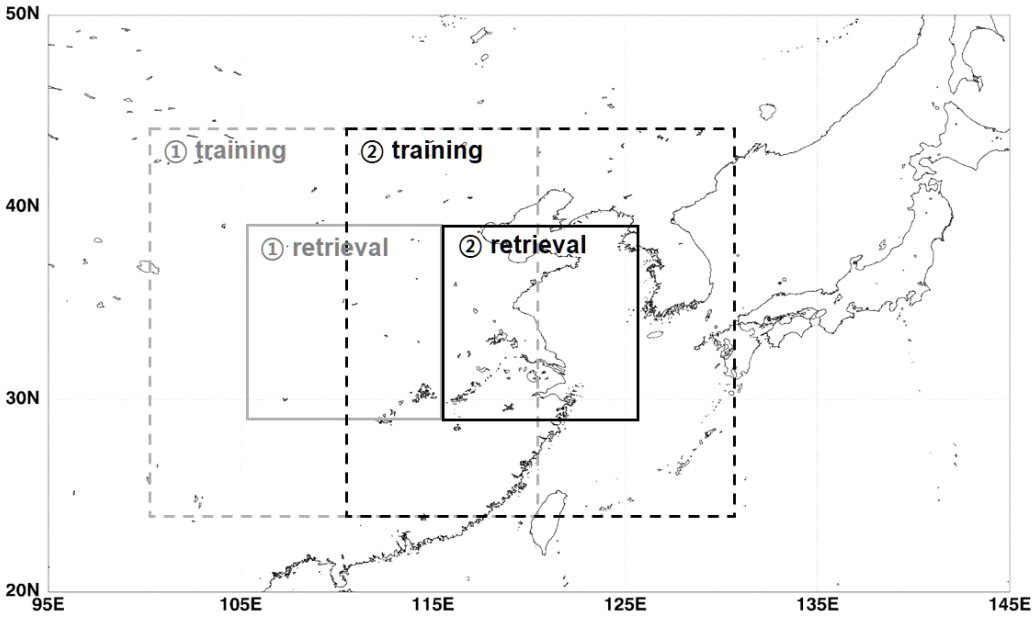

Figure 9. East Asian analysis domain used in this study and an example of conducting the regressions in two neighboring regression domains. Dashed and solid boxes indicate the areas that were used for training and retrieval, respectively.

4.3. Detecting clear-sky FOVS from MODIS measurements

Since infrared measurements are significantly affected by cloud, effort for screening cloud signal in hyper-spectral measurements should be necessary. In this study, retrievals are performed only over clear sky regions to avoid the problems caused by cloud contamination. One of the effective ways to find contaminated FOVs is the use of other high-resolution image channel measurements. Determining the clear-sky pixel, we use cloud information obtained from the IMAPP MODIS-AIRS collocation package (https://cimss.ssec.wisc.edu/imapp/uwairs_utils_v1.0.shtml), which provides cloud amount at each AIRS pixel (Li et al. 2004). Since the MODIS cloud mask has 1 km of spatial resolution, it is possible to estimate the cloud amount inside 14 km resolution of AIRS FOV. Figure 10 shows an example showing MODIS-AIRS collocation result from the IMAPP package. It is noted that change of footprint size with respect to satellite viewing angle is considered in the collocation procedure. Details of collocation method are described in Nagle and Holz (2009). In this study, the AIRS pixel is considered to be clear when the obtained cloud amount is zero.

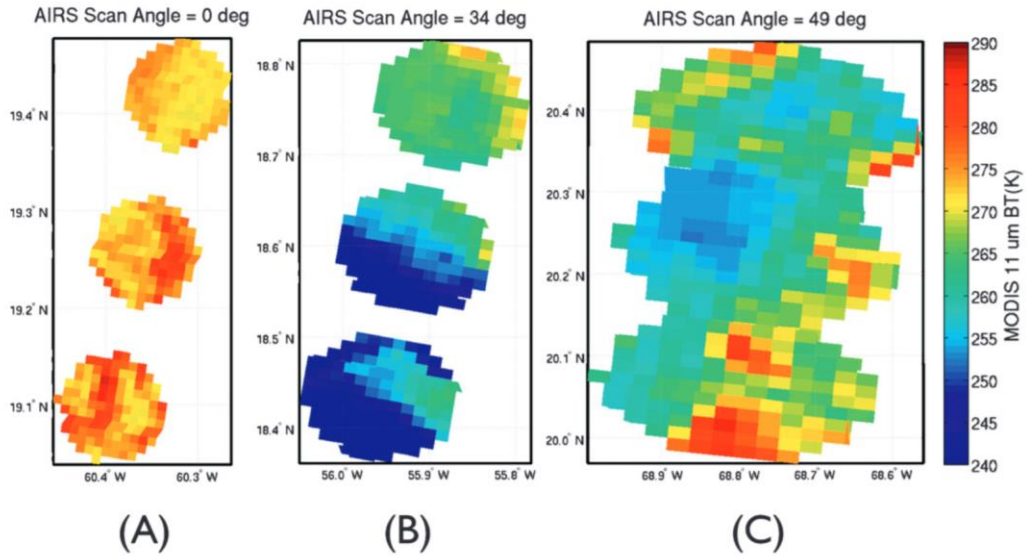

Figure 10. An example of applying the collocation to an imager (MODIS) and infrared sounder (AIRS) measurements is presented for three different AIRS scan angles. The images present the collocation applied to the MODIS 1 km resolution 11 μm BT mearuements. (Nagle and Holz, 2009)

4.4. Error analysis

4.4.1. Validation by using independent simulation dataset

To validate the developed model could cover various atmospheric states, independent cases of about 45,000 (2.6% of one-year ECMWF data over East Asia) are randomly selected. Conceptually, validation by using the simulation dataset highly depends on the accuracy of radiative transfer modeling. Consideration of trace gas is one of the problems. Simulation was conducted by using only temperature, moisture, and ozone. Even though we omitted the channels affected by strong trace gases such as N_2O, CH_4 and CO, possible contamination in the real observation may still exist; it is not shown in validation by using the simulation dataset. Another problem is about the noise on observations. Noises on the real sensor cannot be reproduced perfectly, though we minimize the problem from the channel selection and noise adding procedures in the radiance simulation. Similarly, sensor calibration problem can affect the validation results. Therefore, possible unexpected spoiling by the sensor noise on the performance of the retrieval cannot be shown in the validation based on the simulation dataset. Nevertheless, the validation based on the simulation dataset has a great advantage, because it is free from the collocation problem. We assume that the ‘true’ state is known. In order to validate retrievals from observations, we need another data set as a reference. Generally, reference data have a different temporal and spatial resolution, so we need a

collocation procedure which could introduce artificial errors in the validation. On the other hand, the reference data have their own errors because they do not represent absolute true state. To sum up, we should note that the validation results from the simulation dataset reveal a theoretical performance of our regression model; but it should be useful to diagnose the regression model.

In this study, two results are compared. One is retrievals based on the moving-window regression model, and the other is retrievals from the CIMSS regression model (Weisz et al. 2007; 2013). There are two major differences between two regression models. First is the training database. As mentioned above, our data base only uses atmospheric states over East Asia, while the pre-developed regression model is designed for retrieving global atmospheric states from the tropics to the polar regions. Second thing is the classification procedure. Our regression coefficients are seasonally and regionally obtained, while the pre-developed regression model classifies the training set into 6-classes from mean brightness temperature of 11 channels around wavenumber 910 cm^{-1} (window channel) which mainly contains information on the surface temperature.

Mean biases and root mean square error (RMSE) are given in temperature-pressure or moisture-pressure diagram (Figure 11). For temperature retrievals from moving-window regression method, mean biases are close to 0 K for whole troposphere. RMSE in troposphere above 700 hPa are between 0.8 K and 1.2 K and below 700 hPa, RMSE

increase from 1.2 K to 3 K. For moisture, mean biases are also close to 0 for whole troposphere, and RMSE of whole troposphere are around 0.4. CIMSS regression retrievals also show reasonable results but retrievals from moving-window regression model show better results in terms of mean bias and RMSE in the given test dataset.

The performance of regression model is generally decided by how well regression coefficients could reproduce relations between atmospheric states and radiances. In other words, when the relations are not easy to define because of nonlinearity or spoiling from other environmental factors, RMSE will increase. Also, if the new relations which never experienced in the training procedure are included in the test dataset, RMSE will increase; this is a theoretical limitation of the regression method and there is no way to overcome without extending the training dataset. In this study, we assume that 270,000 cases obtained during four-year ERA-I data could cover almost all possible variability of relations between atmospheric states and radiances.

In case of error statistics for the temperature, interesting features are shown; RMSE increases near the surface level. One of explanations is that temperature over the lower atmosphere changes much dynamically compared to the higher atmosphere, so necessity of large coverage including more atmospheric variability degrades the performance of regression model. Considering that radiances observed from satellite are not for a single level but for a layer mean value, surface temperature which is much dynamically changing than that of atmosphere will be able to affect channels influenced by lower

level temperature. In other words, the way of satellite measurements could bring larger variability of lower atmospheric temperature.

Moreover, RMSE increasing around the boundary layer can be related to the surface emissivity. When atmospheric composition is well-known (this is a reasonable assumption for CO_2 channels), atmospheric term in Equation (1) is only a function of temperature, whereas surface term in Equation (1) is a function of surface emissivity as well as surface temperature. Since the observed radiance is affected by surface emissivity, surface affected channels need to consider both surface temperature and surface emissivity; it causes more complex relations between lower level channels and corresponding atmospheric states. Therefore, accuracy for the lower atmosphere is in general lower than that for the higher atmosphere.

In case of moisture, retrievals are more difficult than the temperature because relations between radiances and water vapor amount is relatively more complex; relationship between water vapor and radiance is known for highly nonlinear (Schmetz et al. 1995; Picon et al. 2003). In this connection, CO_2 channels and window channels are very accurately simulated using the fast radiative transfer model, while water vapor channels, especially strong water vapor channels, are relatively hard to simulate (Strow et al. 2003). Therefore, the water vapor retrieval algorithm based on the simulation dataset may show slightly degraded results when real radiance observations are applied. To minimize such a nonlinearity, logarithm of mixing ratio is used instead of mixing

ratio for the regression model development because longwave radiation (e.g., radiance of water vapor channel) is roughly proportional to the logarithm of its concentration.

In order to examine regional or seasonal differences of error statistics, we further analyze the RMSE on longitude-latitude frame at 300 hPa and 850 hPa in summer and winter season. Temperature and moisture RMSEs are shown in Figure 12 and Figure 13, respectively. For temperature, errors in winter are slightly higher than those in summer at both higher and lower troposphere. Temperature RMSEs at the higher troposphere are between 0.8 K and 1.2 K, and significant regional differences are not shown though errors over lower latitudes show slightly smaller than higher latitudes in summer season (Figure 12a–b). In contrast to the higher troposphere, errors at the lower troposphere show larger regional differences (Figure 12c–d). Errors in the north-west of the analysis domain show much larger RMSEs around 4 – 5 K, while those over the south-east show smaller RMSEs around 1.5 – 2.5 K. Such patterns look like a land-sea contrast. Because regions showing larger errors are generally in high elevated regions, the number of data used for the regression is relatively small which may create generally more uncertain nature of error statistics. Such high errors over the high elevated regions may be due to the increasing nonlinearity between the dynamical temperature change in the lower atmosphere and surface's irregular elevations which cannot be resolved in the regression model.

On the other hand, moisture RMSEs do not show significant regional differences in

both higher and lower troposphere, but errors are higher in winter than in summer (Figure 13). Conceptually, RMSEs of the water vapor amount will be correlated with RMSEs of temperature; because atmospheric contribution to the observed radiation is a function of temperature as well as absorbers amount, as described in Equation (1), inappropriate temperature information could be directly linked to misleading of absorber amount. One reason why such misleading does not appear in Figure 13c–d may be due to the use of logarithm of mixing ratio during the regression procedure, which could suppress the large moisture change at lower troposphere (e.g., smoothing effects). In other words, the regression model could capture only smoothly varying water vapor features in particular in the lower troposphere. In fact, RMSEs in the lower troposphere during the summer show similar error characteristics in temperature but magnitudes of errors are very smoothed. However, RMSE in the lower atmosphere in winter do not show such patterns, so additional explanations are needed. One of the possibilities is that regression coefficients are more relying on absorber amount than the temperature. It does not mean that temperature information in water vapor channel is not important in retrieval. Errors from unsuitable relations between radiances and water vapor amounts, nonlinearity, could be more significant in the accuracy of moisture retrievals. In Figure 13d, highest RMSEs are shown in coastal regions but not land or ocean region, leading to an expectation that the nonlinearity of regression model might cause such result.

In Figures 14–15, regional or seasonal improvement of the regression model at 300 hPa and 850 hPa are shown. In this analysis, improvement is defined by the differences of RMSE between moving-window regression retrievals and CIMSS regression retrievals, and negative values imply that our regression retrievals have less RMSE than CIMSS regression retrievals. In case of temperature, similar improvement of about 0.3 K is shown in the higher troposphere in both summer and winter (Figure 12a–b), while in the lower troposphere regional and seasonal variation of improvement is shown (Figure 12c–d). For the lower troposphere, CIMSS regression retrievals show better results than our regression retrieval in west part of research domain where surface elevation is high. However, in the other regions moving-window regression retrievals show better results; especially, RMSEs are reduced by up to 1 K in the boxes around the Korean peninsula in winter.

In case of moisture, all results show optimistic improvement (Figure 15). For higher troposphere, much improvement over the dry regions of summer season is shown than the other regions, while weak regional dependency of improvement is shown in winter (Figure 15a–b). For lower troposphere, much improvement is shown over humid regions of winter season than other regions, and another big improvement is shown over the west-south region in summer (Figure 15c–d)). Including the temperature results, full physical explanation about how differences of two regression models such as classification methods or training dataset could cause such regional and seasonal

differences of improvement is not clear, but at least we could conclude that improvement shown in Figure 11 is relevant for nearly all the East Asia regions.

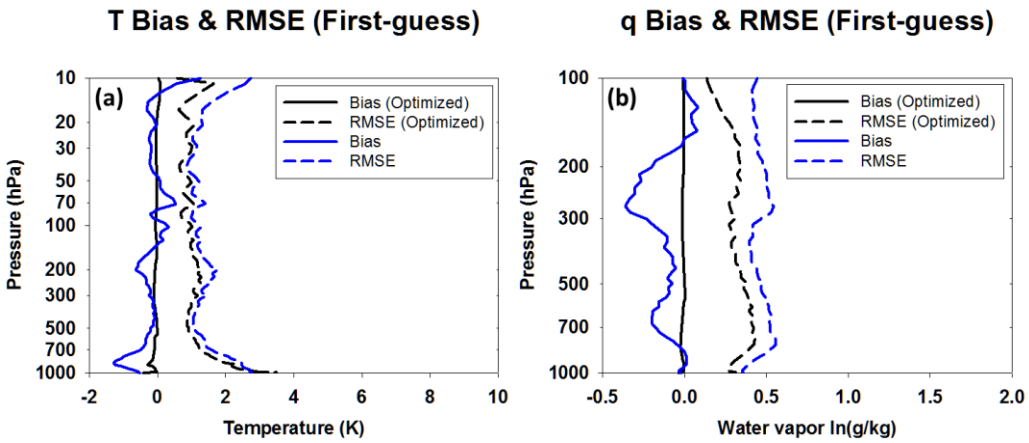

Figure 11. Comparison between moving-window regression retrievals (black) and CIMSS regression retrievals (blue) for (a) temperature and (b) moisture. Solid and dashed lines indicate mean bias and RMSE, respectively.

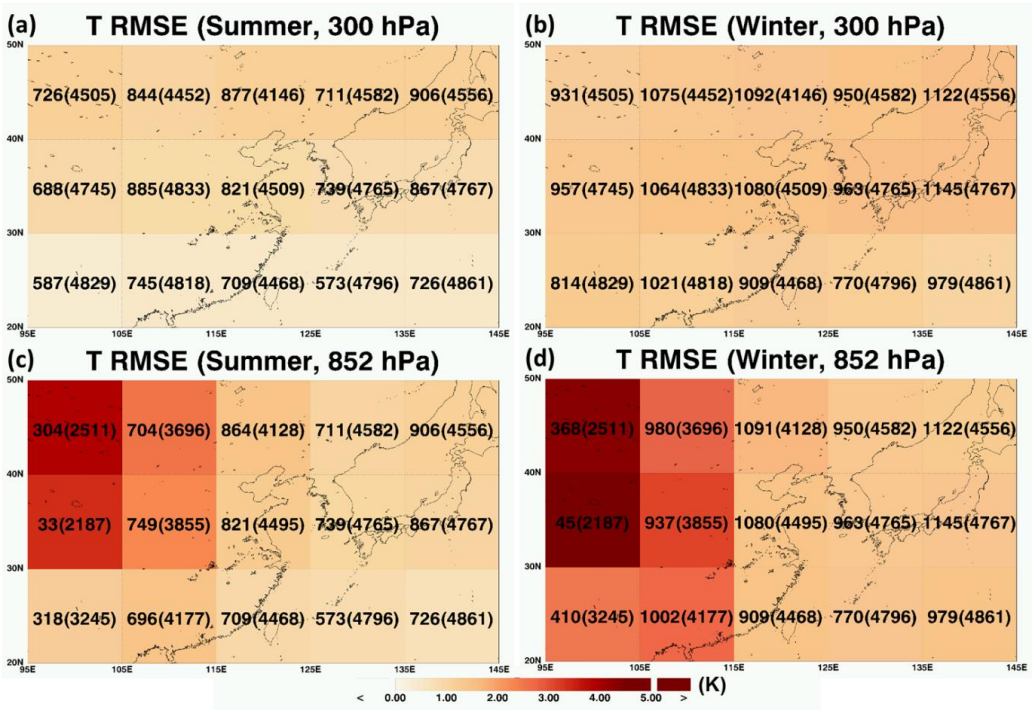

Figure 12. RMSE map for regression retrievals of temperature in summer (a, c) and winter (b, d). Upper (a, b) and bottom (c, d) panels show errors of 300 hPa and 850 hPa, respectively. The numbers inside boxes indicate the number of test samples and training samples. The latter is written in parentheses.

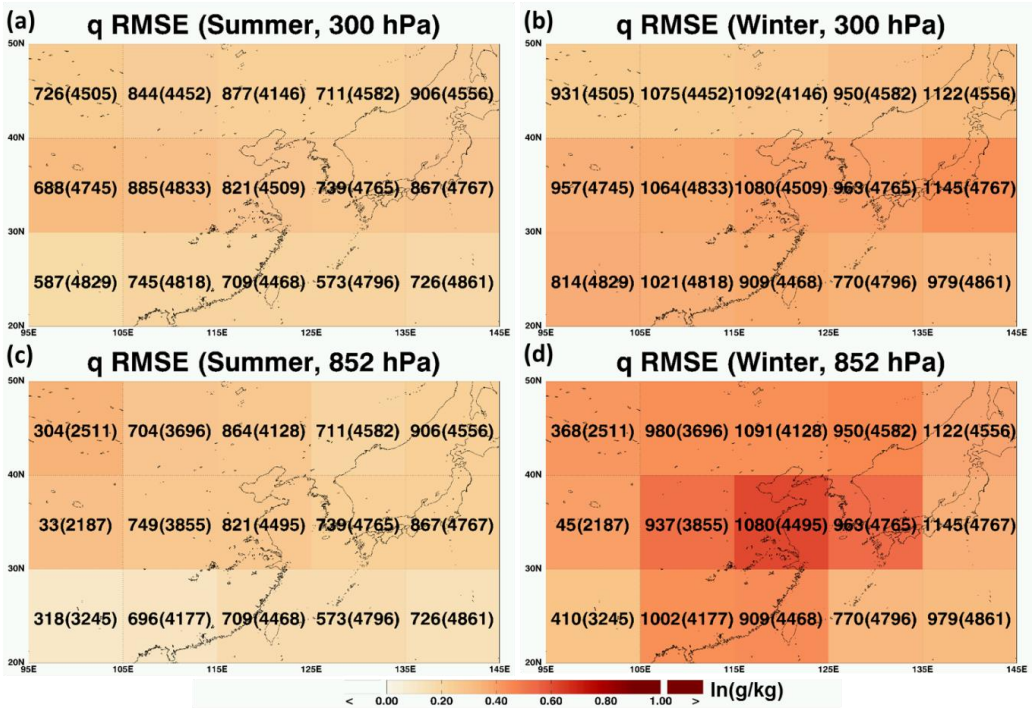

Figure 13. Same as Figure 12, but for moisture.

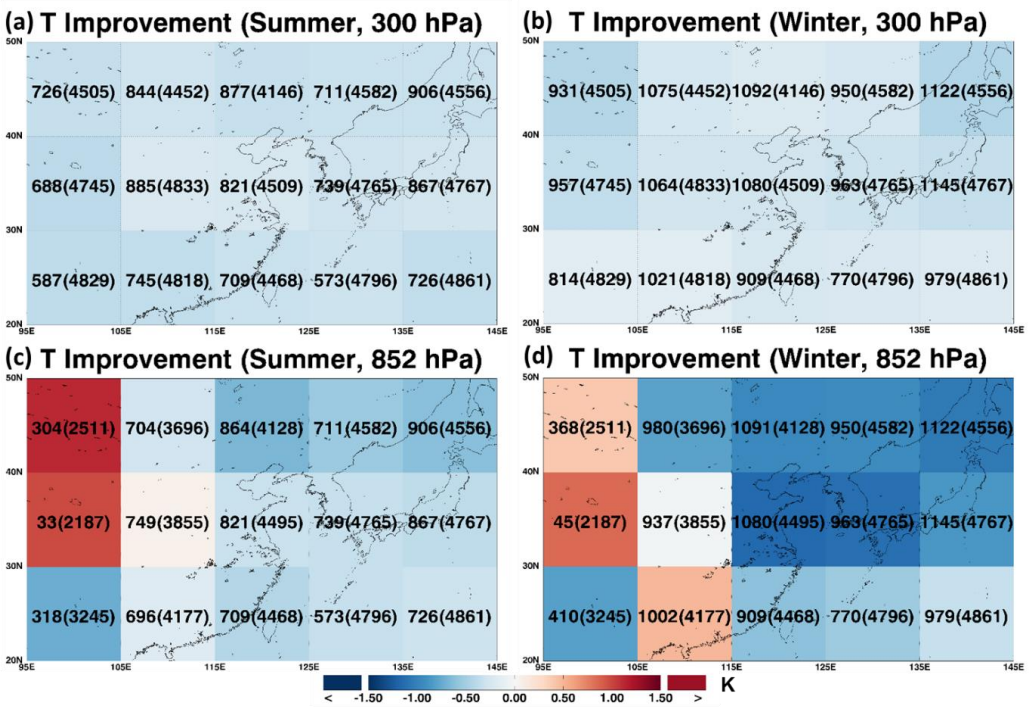

Figure 14. Same as Figure 12, but for the improvement of temperature RMSE. Here, improvement is defined by RMSE of moving-window regression retrievals minus RMSE of CIMSS regression retrievals.

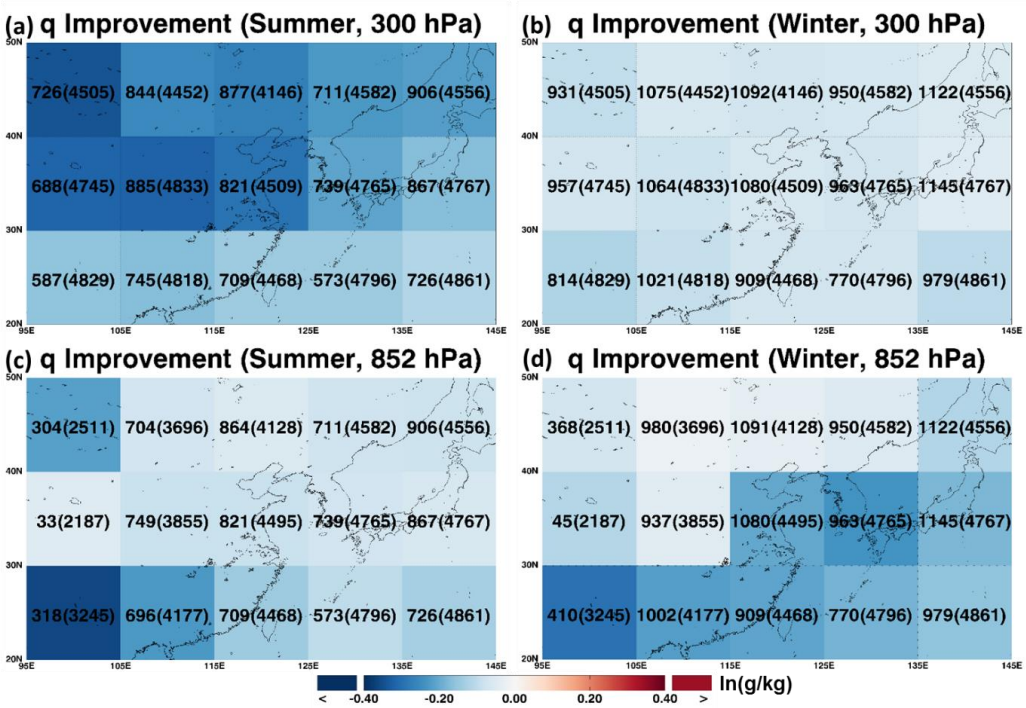

Figure 15. Same as Figure 14, but for the improvement of moisture.

4.4.2. Case study

To validate the developed model could cover various atmospheric states, independent Here, we test the developed regression model for one randomly selected case, and compare the results with CIMSS regression retrievals. AIRS level 1B radiances for 0435 UTC 24 September 2012 are selected, and temporally closest ECMWF interim data for 06 UTH reanalysis are used as a reference. To remove cloud-contaminated AIRS pixels, MODIS cloud mask products are collocated. Figure 16 provides reference for temperature and moisture which are spatially collocated with AIRS granule. In this observation, the AIRS granule covers the area around the Korean peninsula. In the temperature field (Figure 16a), a cold temperature area showing around 270 K is located over the north-east region of the granule, while high temperature area around 280 K is shown over the south region of the granule. It is noted that a thermal tongue is developed along with Shandong peninsula and Liaodong peninsula, and a thermal core is shown in Liaodong peninsula. On the other hand, for moisture, dry regions are widely shown over Liaodong peninsula and middle of Manchuria. In contrast humidity over the north-west and the south-west shows relatively higher values than the other regions. From the south-west of granule, a humid band is extended from Shanghai area to middle Korea, and to East Sea.

Retrieval results from the moving-window regression are shown in Figure 17a along with CIMSS regression retrievals for temperature (Figure 17c). Figure 17b and

17d show the differences from reference. In this study, differences are defined by retrievals minus reference. General features of both regression retrievals are in good agreement with the reference, but two products are slightly different. Over the Yellow Sea, CIMSS regression retrievals show gradually decreasing patterns from south to north, while moving-window regression retrievals show slightly increasing patterns around Liaodong peninsula like reference. Around the Yellow Sea and the Korean peninsula, both products overestimate temperature and the overestimation is slightly larger for the moving-window regression retrievals (Figure 17b, d). However, in Manchuria area, underestimation by CIMSS regression retrievals is larger than that by moving-window regression retrievals. In case of moisture, two products also show features similar to the reference, but different bias trends are shown (Figure 18). Moving-window regression retrievals show an overestimation over the Manchuria area, while CIMSS regression retrievals show an underestimation over the Yellow Sea and the Korean peninsula. From this study, results may not be generalized, however, the test results suggest that moving-window regression retrievals demonstrated that results are comparable to the reference data, as for the CIMSS regression retrievals.

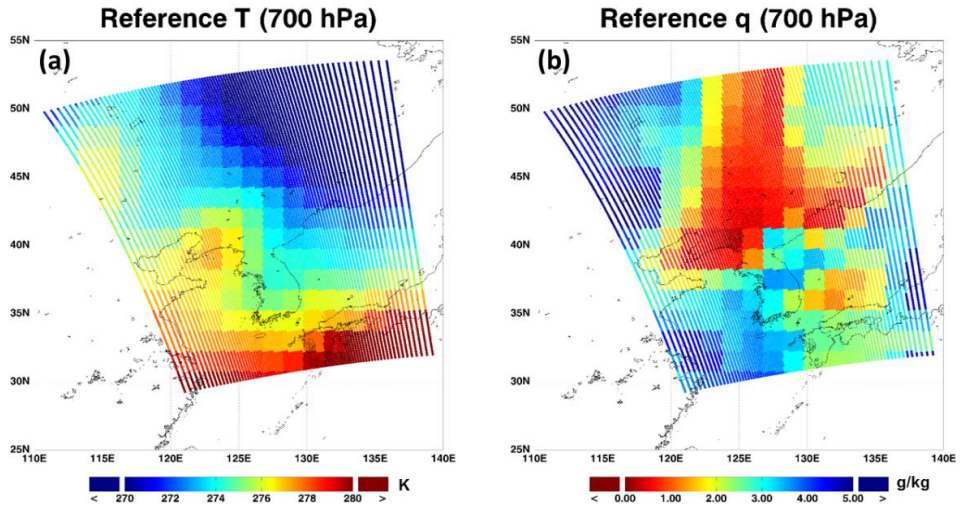

Figure 16. The spatial distribution of (a) temperature and (b) moisture at 700-hPa from ECMWF reanalysis for 06 UTC 24 September 2012. The data are spatially collocated with AIRS granule for 0435 UTH 24 September 2012.

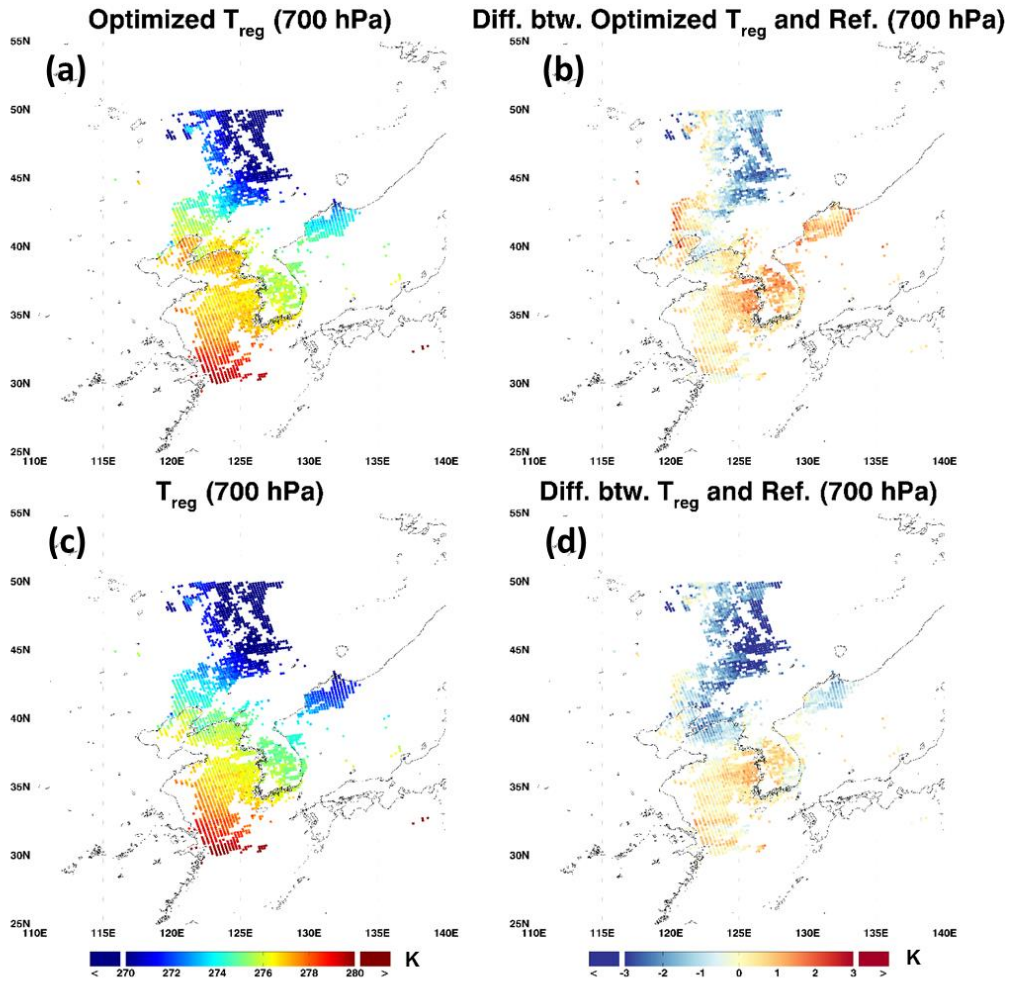

Figure 17. Comparison between (a–b) moving-window regression retrievals and (c–d) CIMSS regression retrievals of temperature at 700 hPa. Left and right panels are regression retrievals and their differences from reference, respectively. Here, ECMWF reanalysis data are used as reference.

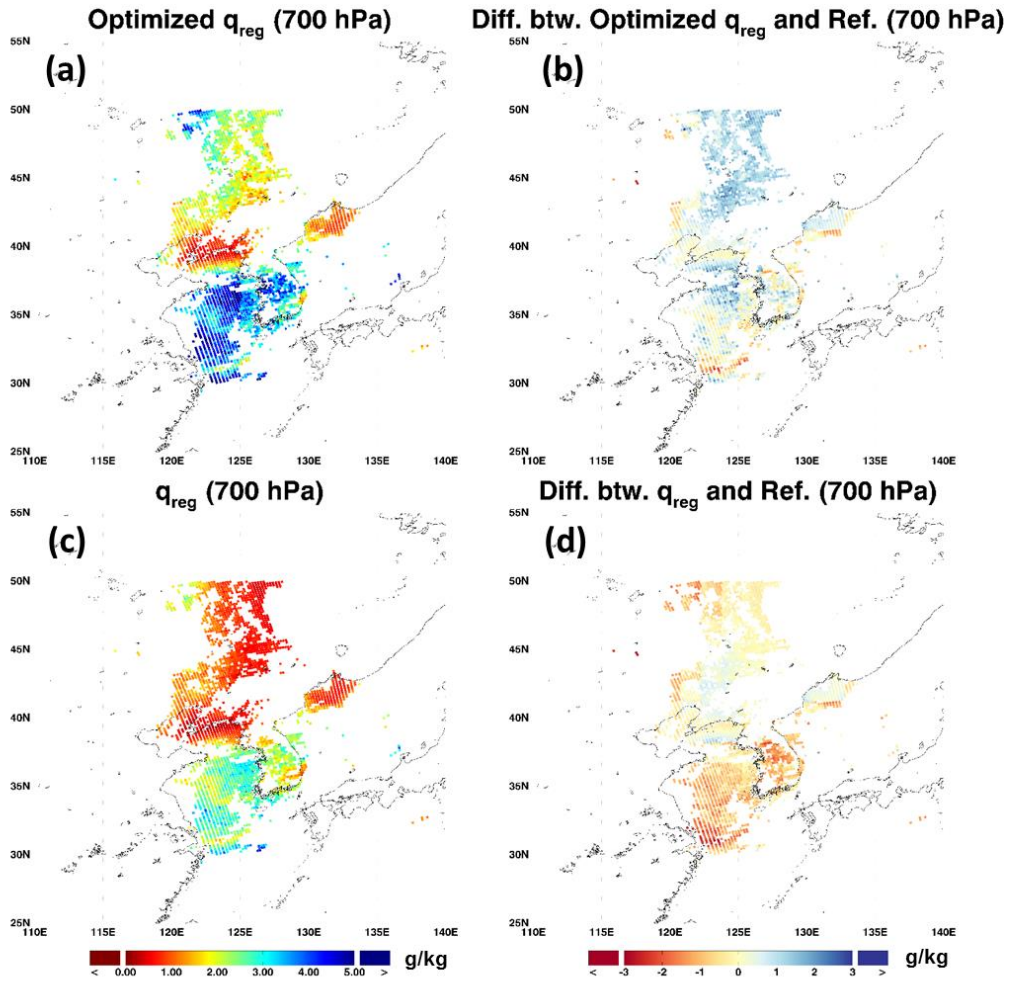

Figure 18. Same as Figure 17, but for moisture at 700 hPa.

4.4.3. Comparison retrievals from real observation with reanalysis data

In order to quantitatively evaluate retrieval performance, more AIRS measurements are collected. In this comparison, all AIRS granules over the East Asian domain for four months (March, June, September, and December of 2010) are used as inputs. As an additional predictor, we use surface pressure from collocated ERA-I data. Retrieval accuracy is examined in terms of error statistics for ERA-I data (used as reference) and collocated AIRS retrievals are averaged over the 0.75° × 0.75° ERA-I grid box. Linear time interpolation of ERA-I data is also considered, for determining the AIRS observation time. In examining the retrieval accuracy, however, a caution should be exercised because both ERA-I and AIRS retrievals are subject to uncertainties and thus the difference between two fields should not be a direct measure of AIRS errors. Furthermore, here, spatial and temporal interpolation may induce another source of error. Nevertheless, in this study, the collocation-caused error is considered to be negligible, and errors in ERA-I is considered to be minor although the surface layers in model outputs tend to be more uncertain.

Here we provide mean temperature and moisture profiles to help understand the error statistics. The mean temperature profile in Figure 19 shows a monotonic decrease from 290 K to 210 K, from the surface to the level of 100 hPa. The mean relative humidity exhibits a pattern of reduction from ~70% at the surface to ~35% at 500 hPa,

followed by an increase up to the level of 250 hPa. This peak may be caused by a relatively lower ice saturation vapor pressure that was used for calculating the relative humidity. Above the level of 250 hPa, the humidity drops to 20% at 150 hPa. It is also interesting to note that the reduction in the relative humidity is highest in layers below 850 hPa.

The accuracies of moving-window regression ($\mathbf{X}_{\mathbf{R}}^{\mathbf{W}}$) and CIMSS regression ($\mathbf{X}_{\mathbf{R}}^{\mathbf{C}}$) are presented in Figure 20. In the case of water vapor comparison, the retrieved water vapor mixing ratio is converted into relative humidity using the saturation vapor pressure estimated from the collocated ERA-I temperature. Nearly 150,000 matched samples are used for this comparison, but the number of data used for the comparison is smaller near the surface owing to the surface topography.

Regarding temperature, the mean bias of $\mathbf{X}_{\mathbf{R}}^{\mathbf{W}}$ is under 0.6 K and the RMSE is in the 1.5 – 3 K range. The bias appears to be smaller compared with the bias suggested by the CIMSS regression retrieval, except for the layer higher than the pressure level of 150 hPa. On the other hand, the RMSEs of $\mathbf{X}_{\mathbf{R}}^{\mathbf{W}}$ and $\mathbf{X}_{\mathbf{R}}^{\mathbf{C}}$ show nearly the same patterns and have nearly the same magnitudes, although the moving-window technique yields slightly smaller values near the surface and for the upper troposphere, above the 150 – 300 hPa layer. It is of interest to note that both retrievals show largest RMSEs near the surface layer (below ~800 hPa), which may be due to the effects of surface

parameters (i.e., skin temperature and surface emissivity) on the temperature channels. The larger RMSEs may also stem from more uncertain features of model outputs (here ERA-I) near the surface. Distinct surface features such as diurnal variations of temperature and associated humidity field variations can also induce collocation errors, causing larger RMSE errors in the surface layers.

The fact that the major improvement of the current moving-window technique, in comparison to $\mathbf{X}_{\mathbf{R}}^{\mathbf{C}}$, is with respect to the mean bias, suggests that the major benefit introduced by the moving-window technique appears to be the removal of the mean bias. It is because the current regression method may fit the retrievals better into the local climatology; yet, the proposed method seems to have a limited impact on explaining the variance of the parameter from the mean climatology. In particular, mid-tropospheric temperature around the 350 – 850 hPa layer shows no improvement in terms of RMSE. In this layer, we expect the regression performance to be relatively independent of the classification method. In this regime, a relatively straightforward relationship between temperature and observed radiance can be proposed. One possible reason for this is that mid-tropospheric temperature channels are nearly independent of surface parameters. That is, radiance observed from mid-tropospheric temperature channels is a function only of atmospheric temperature. A relatively simple temperature profile found in the mid-troposphere can also explain good regression performance. The other possible reason is that the classification approach employed in this study may not be fully

optimal. If it is classified against atmospheric situations, better regression can be expected over the mid-troposphere. For example, the recent validation study for IASI, based on the classification with respect to the atmospheric situations (so-called piecewise regression), reported that the precision in the mid-troposphere is closer to 1 K (EUMETSAT[b] 2016).

One of the interesting things in error statistics are error peaks in the upper troposphere (~200 – 300 hPa), suggesting less confidence for retrievals over that layer. Increasing complexity of the temperature field may be one reason for this. This layer is a layer in which the stratospheric influence starts to appear. It is noted that the lapse rate also changes around this layer. A detailed discussion of this phenomenon is beyond the scope of the present study, but we expect this phenomenon to be also related to seasons and geographical location, because the moving-window technique shows a significant improvement over the 200 – 300 hPa layer.

The water vapor retrievals based on the moving-window technique are compared with the CIMSS regression results (Figure 20b). The mean bias ranges from −12% to 2% and the RMSEs are in the 8 – 22% range. Compared with the CIMSS regression retrievals, the moving-window technique yields significantly smaller mean bias, except for the surface layer, in which the mean bias increases toward the negative value. The RMSE becomes smaller over the whole troposphere, but the major improvement appears to occur over the 200 – 400 hPa atmospheric layer.

Similar to the temperature retrieval, the main improvement for the water vapor retrieval is also in the bias removal, probably because this model more precisely describes the local climatology. In addition, smaller variances of regression coefficients again suggest that regionally and seasonally focused training data can help enhance the regression performance.

Similar to the results obtained for temperature (Figure 20a), peaks of RMSEs of relative humidity are found over the upper troposphere of 200 – 300 hPa (Figure 20b). It may be partly due to the amplified error scale caused by the lower saturation vapor pressure of ice.

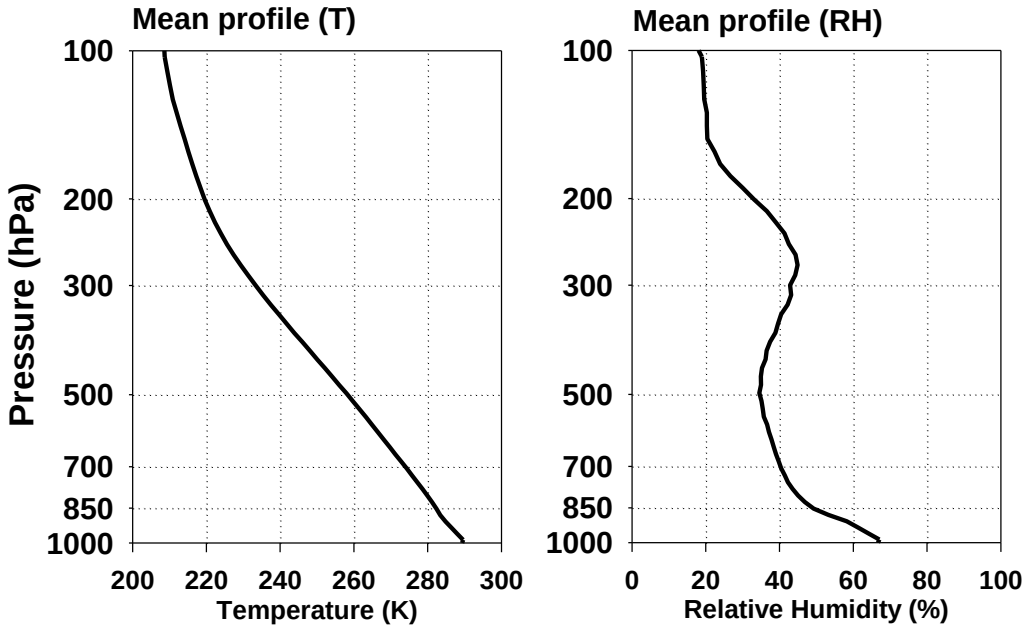

Figure 19. Mean profiles of collocated ERA-I temperature and relative humidity data used for comparison. Left and right panels represent temperature and relative humidity, respectively.

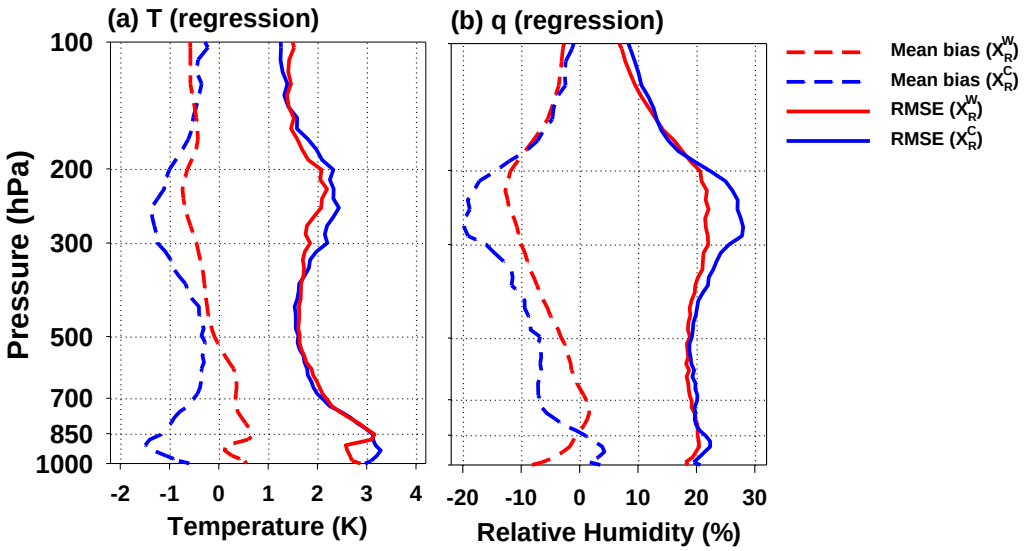

Figure 20. Error statistics for (a) retrieved temperature and (b) relative humidity. For error statistics, mean biases (dashed lines) and RMSEs (solid lines) between regression retrievals obtained using the moving-window technique (red lines) and those obtained using the CIMSS regression retrievals (blue lines) are shown in (a) and (b).

5. Impact of a priori information improvement on accuracy of 1DVAR

So far we showed that the moving-window regression method yields better temperature and moisture profiles than those obtained using the CIMSS regression method with training on global data. Because, in general, physical methods with better initial data will yield better retrievals, it is of interest to examine to what extent the proposed method improves retrieval, compared with the retrieval results obtained using the CIMSS regression method.

5.1. 1DVAR model

In this study, we adopt CIMSS 1DVAR model (Li et al. 2000), which uses an analytical form of the radiative transfer equation (Li 1994). Some salient features of this CIMSS model include the use of an error balancing factor (γ) to yield more accurate error covariance. That is $\mathbf{B}^{-1} = \gamma\mathbf{I}^{-1}$ is applied in Equation (4), where γ is the smoothing factor. The value of γ is dependent on the observations, the observation error, and the first-guess of solution; often it is chosen empirically (Susskind, 1984; Smith et al., 1985; Hayden, 1988). The smoothing factor plays a critical role in the solution; if γ is too large, then the solution could be over-constrained and largely biased, while if γ is too small, the solution could be under-constrained and unstable. In the CIMSS 1DVAR model, a

simple numerical approach is adopted to obtain γ. Starting with $\gamma_0 = 1$, γ is updated every iteration by the following procedure:

- If $\|F(\boldsymbol{X}_n) - \boldsymbol{Y}^m\|$ decreases, then decrease γ by a factor of 0.8.
- If $\|F(\boldsymbol{X}_n) - \boldsymbol{Y}^m\|$ increases, then increase γ by a factor of 1.8 and keep the $\mathbf{X_{n-1}}$.

Since the variability of the atmosphere could be explained by a limited eigenvector, δ**X** in Equation (4) is substituted by **A** which satisfies $\mathbf{X} - \mathbf{X_0} = \mathbf{\Phi A}$ where **Φ** is the matrix of EOFs and **A** is the corresponding projection coefficient score. It is obvious that $\mathbf{\Phi}^T\mathbf{\Phi} = \mathbf{I}$. Defining $\mathbf{F'}=\mathbf{F\Phi}$ and $\mathbf{B}^{-1} = \gamma\mathbf{I}^{-1}$, Equation (4) becomes

$$\mathbf{A}_{n+1} = (\mathbf{F}_n'^T\mathbf{R}^{-1}\mathbf{F}_n' + \gamma\mathbf{I}^{-1})^{-1}\mathbf{F}_n'^T\mathbf{R}^{-1}(\delta\mathbf{Y}_n + \mathbf{F}_n'\mathbf{A}_n) \tag{8}$$

where $\mathbf{A_0} = 0$. Equation (5) is applied to retrieve the solution from AIRS observation. More details can be found in Li et al. (2000).

5.1.1. Background error covariance

To use the regression retrievals $\mathbf{X_R^W}$ as a priori information for the physical retrieval, its error covariance ($\mathbf{B^W}$) should be defined, in order to determine a relative importance to the observation, as expressed in Eq. (2). According to Bouttier and Courtier (1999), the errors in the background are modelled as follows,

$$\mathrm{B} = \overline{(\varepsilon_b - \overline{\varepsilon_b})(\varepsilon_b - \overline{\varepsilon_b})^T} \tag{9}$$

where $\varepsilon_b = \mathbf{X_b} - \mathbf{X_t}$, $\mathbf{X_b}$ is background field, and $\mathbf{X}_t$ is true state. The average of errors,

$\overline{\varepsilon_b}$, are called biases, and they are the sign of a systematic problem in the regression model. Biases are vectors of the same kind of the atmospheric state, so their interpretation is straightforward. Since physical method, discussed in section 3, assumed no biased state, it is noted that the calculation procedure for **B** removes such bias term.

Black lines in Figure 21 indicate error statistics of physical retrievals by using moving-window retrievals as background and above derived $\mathbf{B}^{\mathbf{W}}$ as error covariance of background in the simulation verification dataset. Results show that errors of physical retrievals are higher than those of regression retrievals (gray) for both temperature and moisture. This is may be due to the imbalance between $\mathbf{B}^{\mathbf{W}}$ and **R**. Theoretically, $\mathbf{B}^{\mathbf{W}}$ is independent of **R,** but empirically, balance between $\mathbf{B}^{\mathbf{W}}$ and **R** plays an importance role in the physical method. Therefore, we empirically adjust $\mathbf{B}^{\mathbf{W}}$ to satisfy the balance between background and observation. In this study based on the simulation dataset, a background error covariance divided by 10 is used, and such adjustment corresponds to use a tenth part of smoothing factor, γ_0, in equation (8). Results of the balanced physical method are shown in red lines of Figure 21. In both temperature and moisture, RMSEs of physical retrievals are evidently reduced. It should be noted that those results are obtained from the simulation dataset.

For more physically reasonable evaluation of $\mathbf{B}^{\mathbf{W}}$ for real AIRS measurements, one may directly construct $\mathbf{B}^{\mathbf{W}}$ from ERA-I data, but uncertainties of the background covariance should be associated with errors in ERA-I data. Instead of directly

calculating $\mathbf{B^W}$ from ERA-I data, $\mathbf{B^W}$ is assumed to be similar to the one that is used for other regression methods; here, we utilize the pre-developed diagonal error covariance matrix ($\mathbf{B^C}$) that was used in the CIMSS regression retrieval ($\mathbf{X_R^C}$). Here, the error ratio between $\mathbf{X_R^W}$ and $\mathbf{X_R^C}$ is assumed to be proportional to the ratio between the diagonal component of $\mathbf{B^W}$ and that of $\mathbf{B^C}$:

$$diag(\mathbf{B^W}) = \boldsymbol{\alpha} \cdot diag(\mathbf{B^C}) \tag{10}$$

where *diag*(·) denotes the diagonal component and $\boldsymbol{\alpha}$ is the ratio of the error variance between $\mathbf{X_R^W}$ and $\mathbf{X_R^C}$. ERA-I is used as a reference for estimating the ratio of the error variance between $\mathbf{X_R^W}$ and $\mathbf{X_R^C}$. Since standard deviation between two products are very similar to each other as shown in Figure 22, similar error covariance would be obtained. Final results of $\mathbf{B^W}$ for temperature and moisture are shown in Figure 23a-b, respectively.

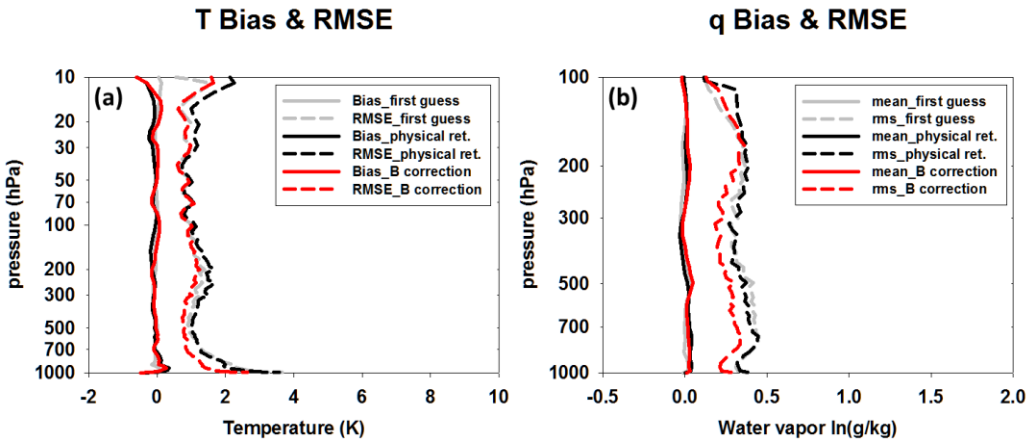

Figure 21. Comparison between regression retrievals (gray), physical retrievals (black) and physical retrievals with empirically adjusted **B** (red) for (a) temperature and (b) moisture. Solid and dashed lines indicate mean bias and RMSE, respectively.

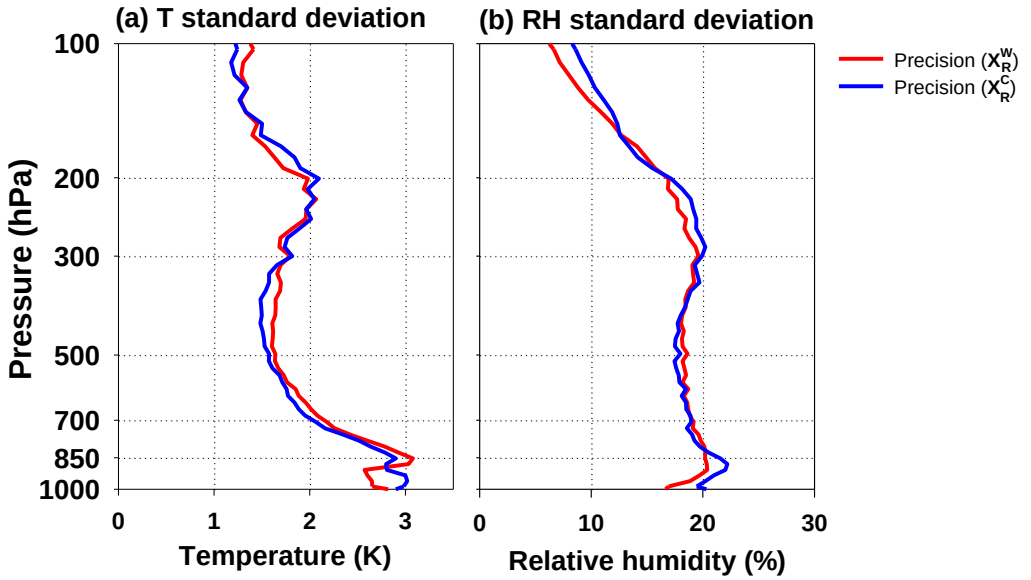

Figure 22. Standard deviation of moving-window regression (red) and CIMSS regression (blue) for (a) temperature and (b) moisture.

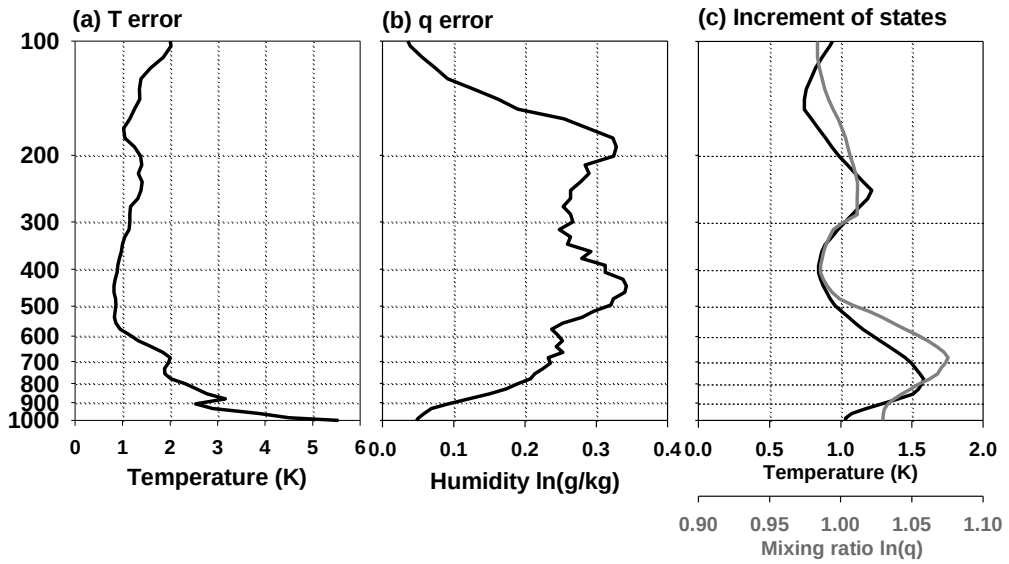

Figure 23. Amplitudes of $\mathbf{B}^{\mathbf{W}}$ for (a) temperature and (b) moisture, and (c) corresponding increment of states for the given 1 K TB perturbations in 1DVAR.

5.1.2. Averaging kernel

For a given **B** (here $\mathbf{B} = \mathbf{B}^{W}$), kernel (**A**) explaining the relative importance between **B** and **R** can be expressed as follows:

$$\mathbf{A} = (\mathbf{K}^{T}\mathbf{R}^{-1}\mathbf{K} + \mathbf{B}^{-1})^{-1}\mathbf{K}^{T}\mathbf{R}^{-1} \quad (11)$$

where **K** denotes a Jacobian. In this study, the observation error **R** is defined by AIRS instrumental noise plus radiative transfer model error (0.2 K). No observation error correlation is assumed. In Eq. (11), the kernel **A** is used to estimate a best state for given observations and a priori information (Rogers, 2000), and the best estimate of **X** (i.e. $\mathbf{X}_{best}$) for given observations is defined as follows:

$$\mathbf{X}_{best} = \mathbf{X}_{0} + \mathbf{A}(\mathbf{Y} - \mathbf{K}\mathbf{X}_{0}) \quad (12)$$

Since the kernel **A** is a matrix, we provide increments of temperature and moisture from $\mathbf{X}_0$ to $\mathbf{X}_{best}$ by assuming all TB perturbations are 1 K (i.e. $\mathbf{Y} - \mathbf{K}\mathbf{X}_0 = 1$ for all channels) -- Figure 23c. For temperature, relatively larger increments are shown in 200 – 300 hPa layer and below 500 hPa level, while in moisture most of increments are made below the 500 hPa level. Thus, we expect relatively larger improvements over those layers when 1DVAR is applied.

5.1.3. Residual analysis for convergence criteria and quality control

For a given **B** (here $\mathbf{B} = \mathbf{B}^{W}$), kernel (**A**) explaining the relative importance In this

study, to determine whether the physical retrieval is converged to an acceptable solution, the following brightness temperature residual (Res) is used:

$$\mathrm{Res} = \sqrt{\frac{1}{\mathrm{M}} \sum (\mathbf{Y} - \mathbf{Y}(\mathbf{X_p}))^2} \tag{13}$$

where M is the number of used channels. During the minimization of the cost function in Eq. (8), Res is calculated and is examined whether the convergence criterion, Res < 0.1 K, is met for the ideal solution. But, even after the minimization is completed, all cases may not meet Res < 0.1 K condition. In this study, these cases are also considered to be the loosely optimized solutions if they satisfy the quality control criterion, Res < 1 K, as in Kwon et al. (2012).

Actually, the gradient of the cost function should reach close to zero when the global minimum is achieved in 1DVAR and thus the gradient of the cost function can be used as a criterion. However, in many cases, the global minimum is not available, resulting in relatively larger residuals even if the norm of the gradient of the cost function becomes close to zero. That is why we allowed 1 K residual as another criterion (or quality flag). As expected, the majority of retrievals are less than 2% of the norm of the gradient; cumulative distributions function indicates 75%, 90%, 93% of the population for 0%, 1%, 2% of the norm of the gradient, respectively -- Fig. 24.

Here, to facilitate our understanding of the algorithm, a flowchart of summarizing the 1DVAR-based physical algorithm is provided in Figure 20. With given AIRS TBs

and a priori information (i.e. regression retrieval), first updated state ($\mathbf{X_2}$) is estimated from the guess field ($\mathbf{X_1}$) using an iterative method of 1DVAR, which starts with a regression retrieval. From now, at each iteration step, residuals of both guess and updated fields are calculated. When the residual of updated field becomes smaller and is less than 0.1 K, the iteration stops and the updated state is set to a final solution (here, best solution). If Res ≥ 0.1 K the algorithm keeps updating from the previous estimate. At this time, the speed of iteration is enhanced by reducing the weight of a priori information with the use of error balancing factor (i.e. $\mathbf{B}^{-1} = \gamma\mathbf{B}^{-1}$, $\gamma = 0.8$). This iteration process for updating is allowed up to 6 times. After 6 times, the last updated state is considered to be a final solution. It should be noted that original CIMSS 1DVAR in Li et al. (2000) uses $\gamma\mathbf{I}^{-1}$.

During the minimization process, the residual of updated state may become larger than the previous state. If this is the case, the increased weight of a priori information is used (i.e. $\mathbf{B}^{-1} = \gamma\mathbf{B}^{-1}$, $\gamma = 1.8$) to allow a more stable iteration in the next step. This stabilizing process is allowed up to 3 times. If the residual is not further minimized in 3 times, the iteration stops and the state having the lowest residual is set to a final solution. In this algorithm, iteration up to 9 times is allowed, and the mean iteration number in this study is found to be about 6.4.

As shown in Figure 25, thus, the final solution is one of following three; (1) best solution satisfying the preset convergence criterion (Res < 0.1 K), (2) loosely optimized

solution (0.1 K ≤ Res < 1 K), and (3) not-converged solution (Res ≥ 1 K). In this study, first two classes (i.e. Res < 1 K) outlined as a dotted box in Figure 20 are considered to be accepted physical solutions.

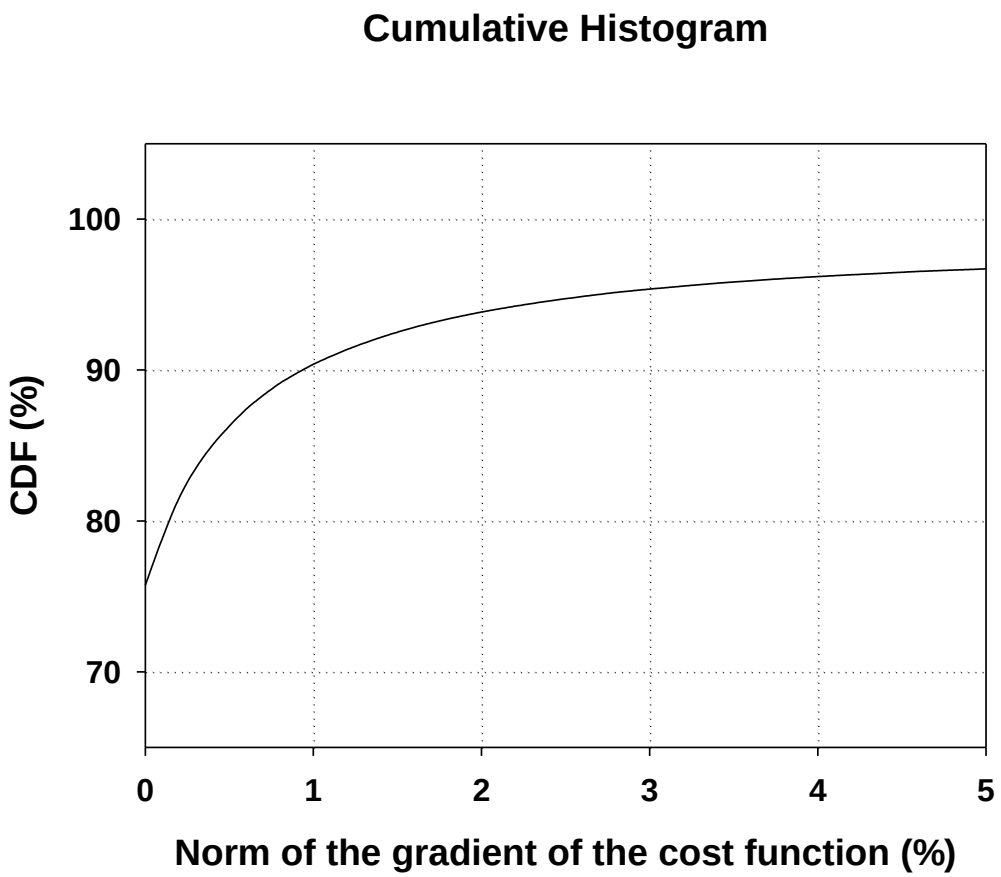

Figure 24. Cumulative histogram for the change of the cost. Only successful retrievals, whose residual is less than 1 K, are used for calculating histogram.

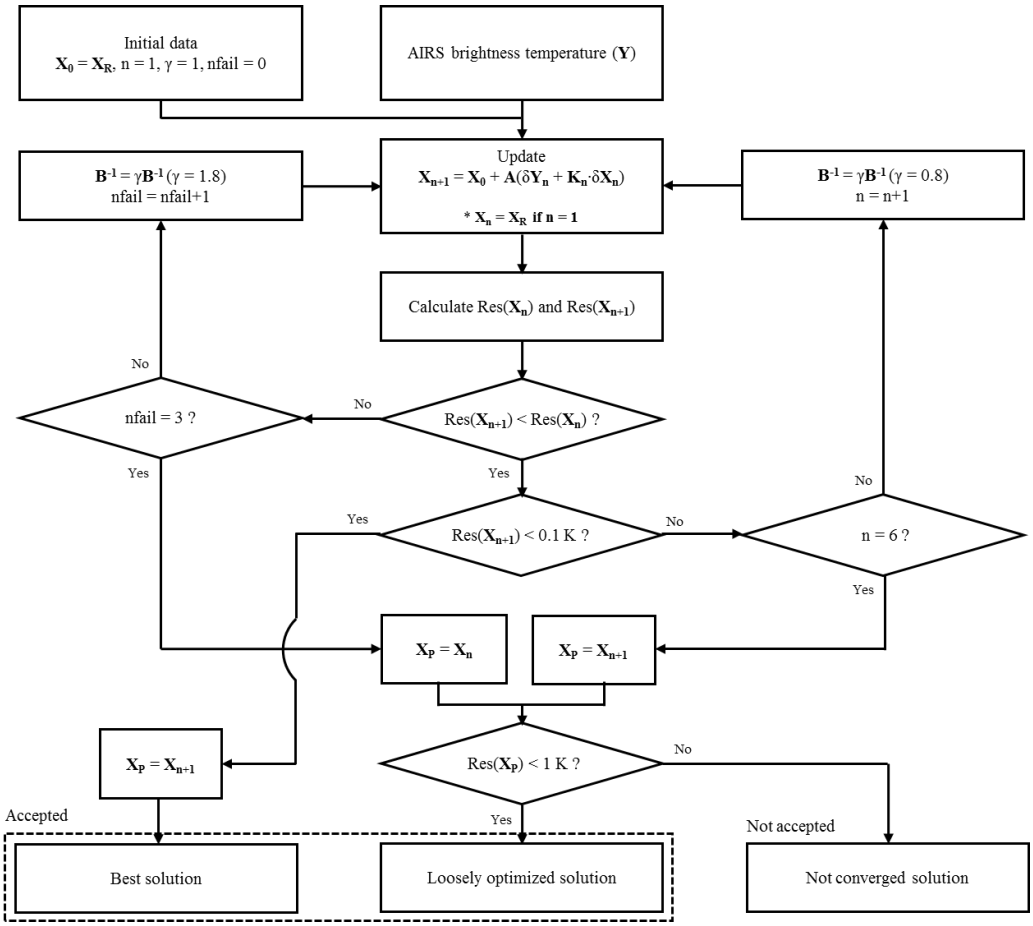

Figure 25. Flowchart of the 1DVAR employed in this study for AIRS temperature and moisture retrievals.

5.2. Error analysis

5.2.1. Validation by using independent simulation dataset

In order to show the impact of providing high quality background on physical method for improving its accuracy, we compared two physical retrievals whose background is coming from moving-window regression retrievals and CIMSS regression retrievals. The evaluation is carried out on the CIMSS physical retrieval algorithm developed by Li et al. (2000) which has been discussed in section 5.1. To check the best performance of physical method, only accepted solutions whose residual is less than 1 K are used both moving-window physical retrieval and CIMSS physical retrieval. If residual increases with iteration (e.g., divergent), regression retrieval is chosen as final solution.

Figure 26 shows impact of background optimization both temperature and moisture. Before comparing moving-window physical retrievals with CIMSS physical retrieval, error statistics for our physical retrievals are analyzed. For temperature, statistics of moving-window physical retrievals are almost same as moving-window regression retrievals in free troposphere. This indicates that regression retrievals are physically well balanced with satellite radiances in free troposphere. In lower troposphere below 700 hPa, RMSE are significantly reduce about 0.5 K compare to regression retrievals, so we could expect that information of radiance on physical method could be helpful to

make better retrievals. In case of moisture, moving-window physical retrievals also show better results than moving-window regression retrievals for whole troposphere about 0.2. From such results, we could expect that moisture retrievals from regression method are not balanced with satellite radiances. Also, such imbalanced retrievals could be improved from physical method, and well balanced final moisture retrievals show much lower RMSE than first-guess.

Regional or seasonal error statistics of physical retrievals are given in Figure 27–28. This is the same as Figure 12–13 but for the physical retrievals. As mentioned above, RMSE of temperature at higher troposphere do not much changed from Figure 12a–b. However, when we compare Figure 12c–d and Figure 27c–d, RMSE of lower troposphere are sufficiently reduced over high elevated region where larger errors were shown. Considering that the number of survived data are much smaller than first-guess data, such results come from filtering of imbalanced regression retrievals as well as error reduction by physical method. As shown in Figure 12, RMSE in winter are still higher than those in summer. For moisture, there are very small regional and seasonal variability of RMSE at both higher and lower troposphere like Figure 13, but generally RMSE are reduced in Figure 27. Also, it is noted that the regions showing high regression retrieval error show high physical retrieval error.

In order to show improvement between moving-window physical retrievals and CIMSS physical retrievals in terms of RMSE reduction, similar map which is shown in

Figure 14–15 but for the physical retrievals are given in Figure 29–30. Generally, when moving-window retrievals are used as background, we could obtain more accurate temperature or moisture in a given simulation dataset. For temperature, impact of background optimization is very significant at higher troposphere, especially over land region in summer (Figure 27a–b). Since there is very small change of error statistics between moving-window physical retrievals and moving-window regression retrievals, most of improvement is coming from degradation of CIMSS physical retrieval. Therefore, we carefully doubt the failure of CIMSS physical retrieval due to imbalance of error covariance at higher troposphere in a given dataset; it does not necessarily mean that CIMSS physical retrieval will show such results globally. On the other hands, improvement at lower troposphere is similar or larger than those of regression retrievals. Considering that moving-window physical retrievals show reduced errors compared to moving-window regression retrievals, CIMSS physical retrievals also show similar scale of error reduction from CIMSS regression retrievals. Large improvement is shown over high elevated region in summer, while in winter large improvement is shown over the high latitudes. However, the number of cases where large improvement is shown is much smaller than the other region, so we could expect that such large improvement cannot critical on general error statistics. In addition, difference between moving-window physical retrieval and CIMSS physical retrieval could be reduced in comparison with difference between moving-window regression retrieval and CIMSS

regression retrieval because role of observation terms on physical method is very critical. Such impact is shown in Figure 30a, and 30d. As mentioned above, error of physical retrieval for moisture is reduced over whole troposphere, so we could expect that error reduction of CIMSS physical retrieval is relatively larger than moving-window physical retrieval. Contrary to Fig. 30a and 30d, large improvement is shown over the north-east region in winter at higher troposphere (Figure 30b) and the west region in summer at lower atmosphere (Figure 30c); this is suggesting that error reduction of moving-window physical retrievals over those regions are relatively larger than CIMSS physical retrieval.

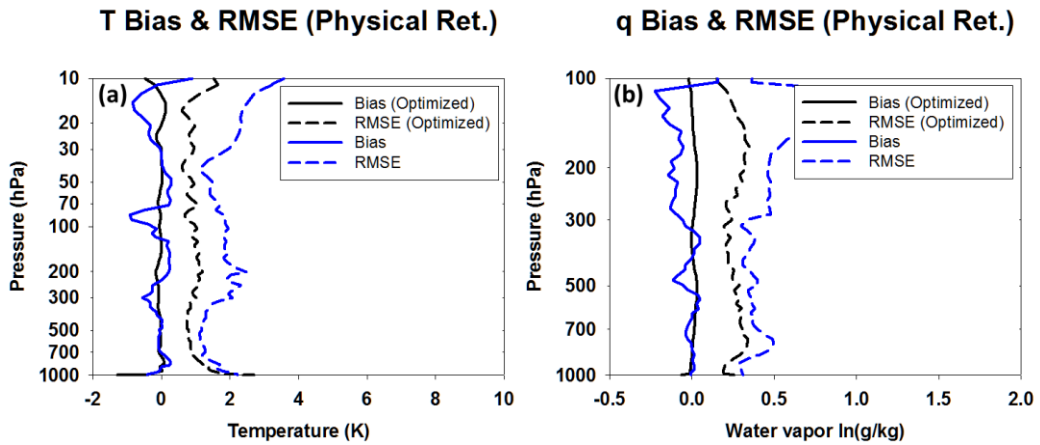

Figure 26. Comparison between moving-window physical retrievals (black) and CIMSS physical retrievals (blue) for (a) temperature and (b) moisture. Moving-window physical retrievals are the solution of physical method when moving-window regression retrievals are used as background. Solid and dashed lines indicate mean bias and RMSE, respectively.

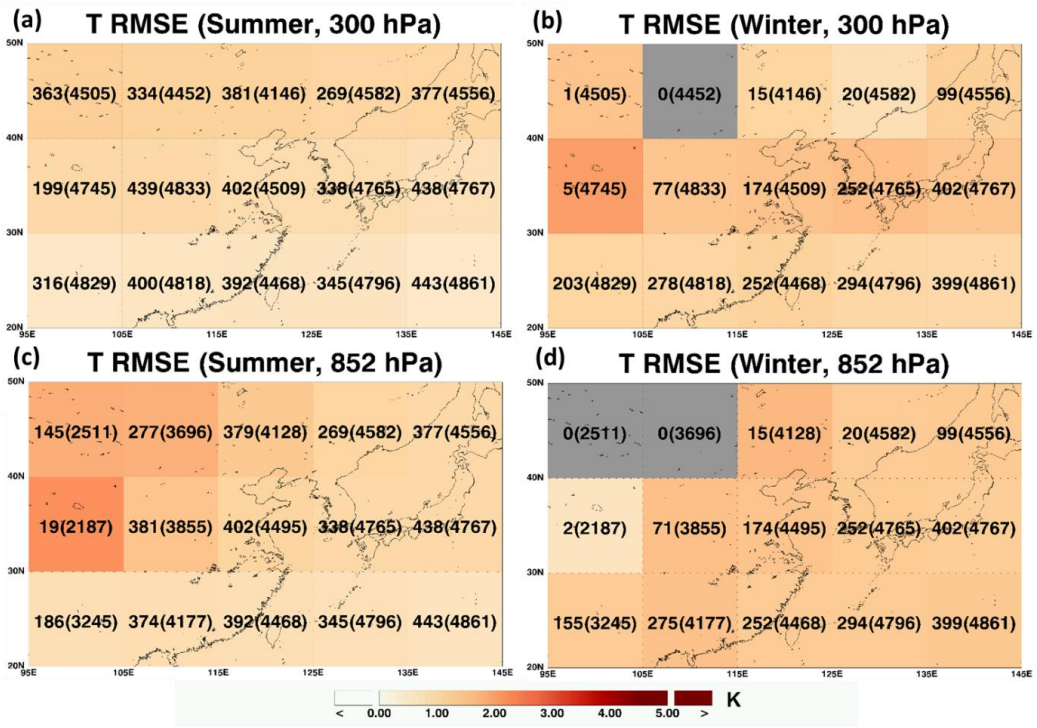

Figure 27. Same as Figure 12, but for physical method. Only accepted solutions whose residual is less than 1 K are used.

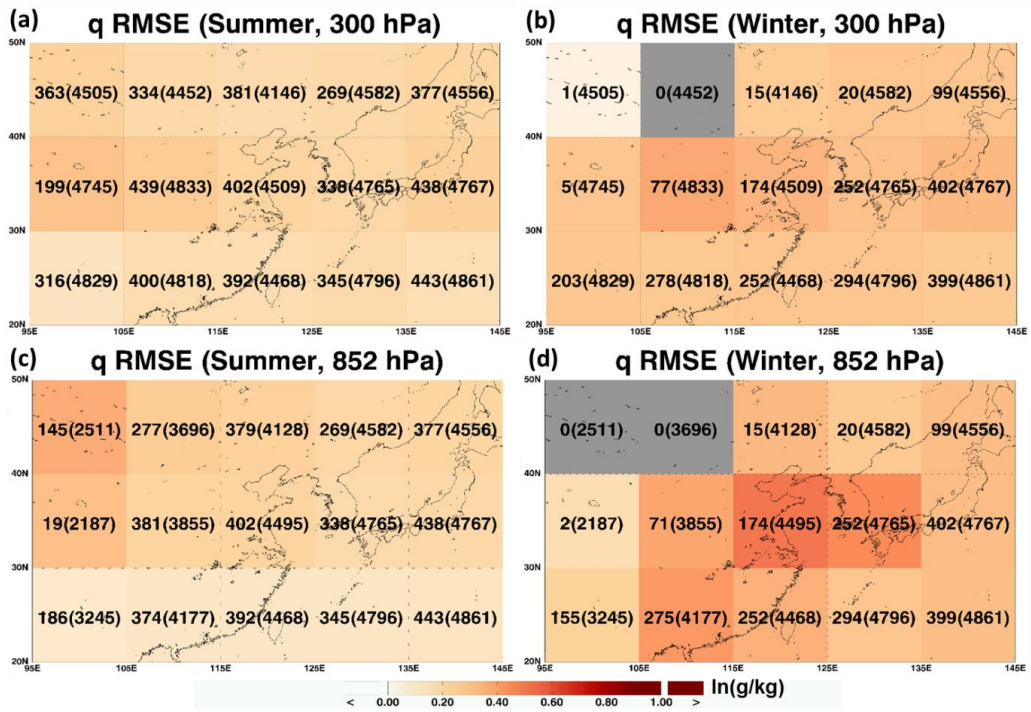

Figure 28. Same as Fig. 27, but for moisture.

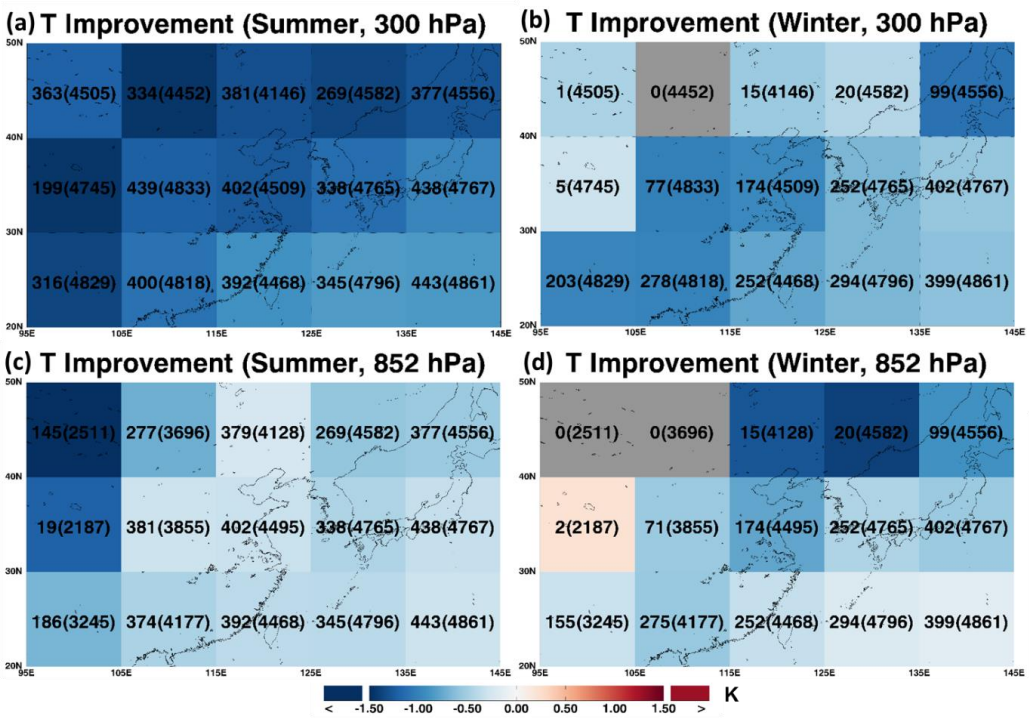

Figure 29. Same as in Figure 14, but the improvement is defined by RMSE of moving-window physical retrievals minus RMSE of CIMSS physical retrievals.

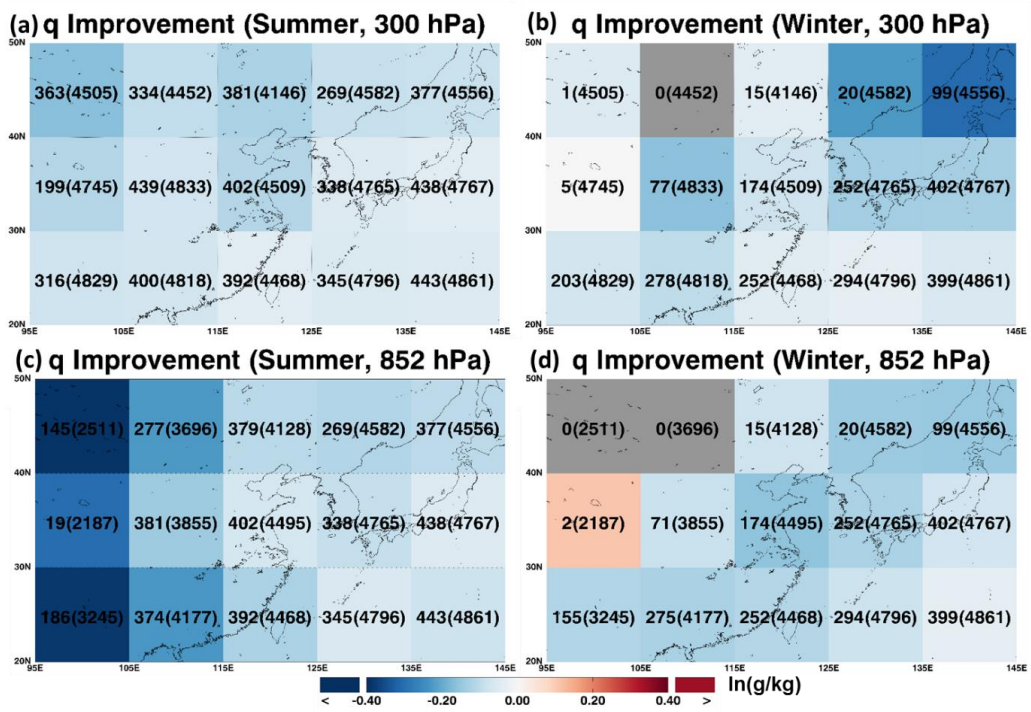

Figure 30. Same as Figure 29, but for moisture.

5.2.2. Case study

In order to show the impact of providing high quality background on physical method We further analyze the impact of background on the real case discussed in section 4.4.2. Because we cannot conclude moving-window regression retrievals are quantitatively better than the CIMSS regression retrieval in the real observation case, the result of impact study also cannot be concluded. Therefore, we only show differences between two products, moving-window physical retrievals and CIMSS physical retrievals in this step. The results of temperature are shown in Figure 31. Like general features of both regression retrievals are very similar, both physical retrievals also show similar results (Figures 31a and c), but gradually decreasing patterns over the Yellow sea in CIMSS regression retrievals does not appear in CIMSS physical retrievals (Figure 31c). As a results, difference patterns compared to reference are more similar each other, though CIMSS physical retrievals show larger underestimation over the Manchu and smaller overestimation over the Korean peninsula, the Yellow sea and Liaodong peninsula. Also, from the results, we could carefully expect that moving-window regression retrievals are relatively close to reference because difference patterns of moving-window regression retrieval are much similar to difference patterns of both physical retrievals. In case of moisture, both physical retrievals show very similar results (Figure 32a, c). Considering that both regression retrievals had different directed error patterns, the results of physical retrievals of moisture highly relies on

observed radiances than background. As a results, different directed biases of two products are changed into very similar patterns (Figure 32b, d). Also, magnitude of differences sufficiently reduced, and generally moving-window physical retrievals show relatively smaller biases than CIMSS physical retrievals. Therefore, we carefully expect moving-window physical retrievals of moisture are close to reference.

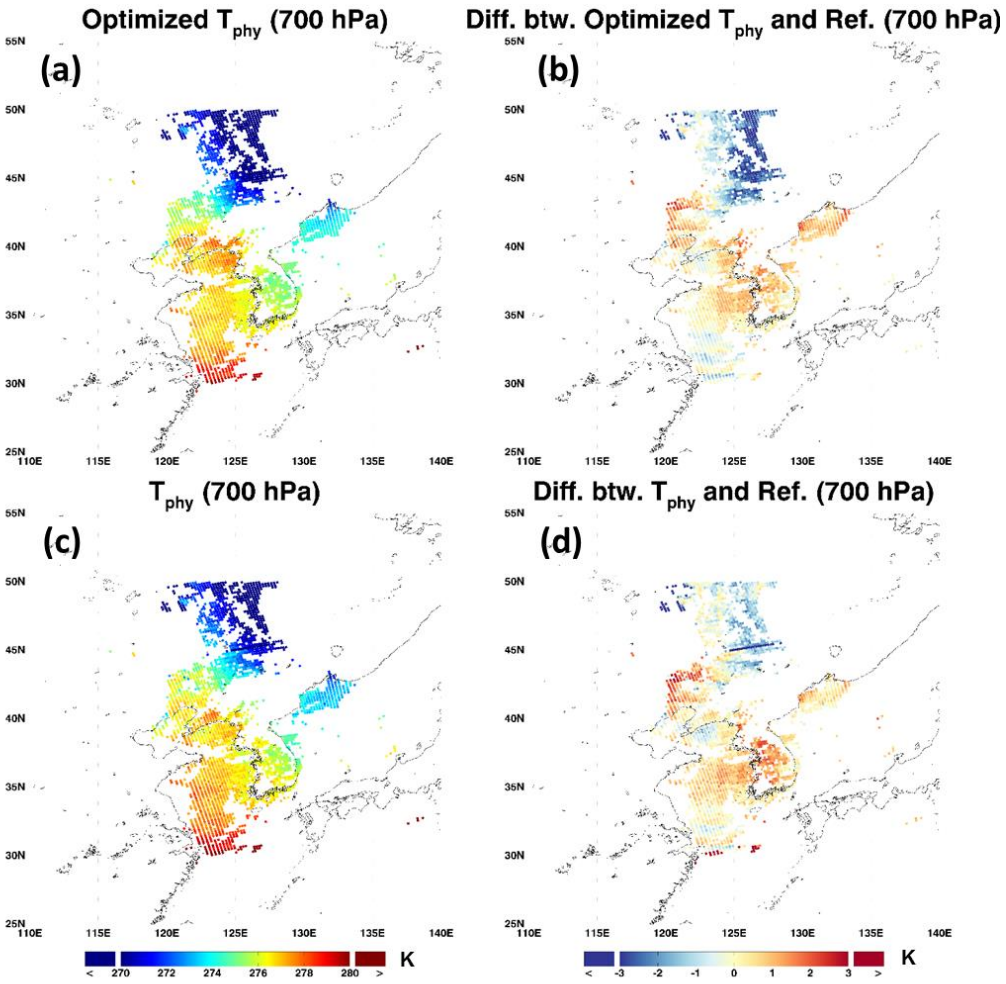

Figure 31. Same as Fig. 17, but for physical retrievals of temperature at 700 hPa.

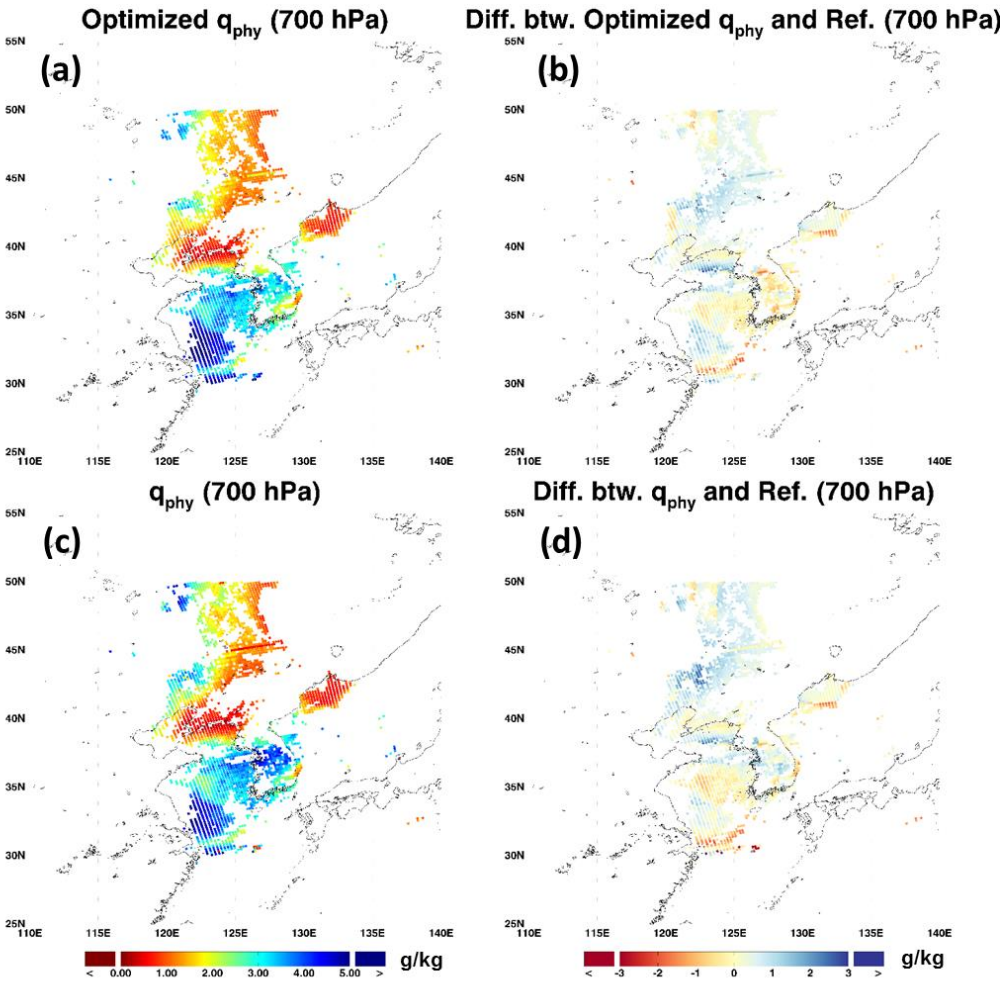

Figure 32. Same as Fig. 18, but for physical retrievals of moisture at 700 hPa.

5.2.3. Comparison retrievals from real observation with reanalysis data

Here, comparison results between moving-window physical retrieval ($\mathbf{X}_{\mathbf{P}}^{\mathbf{W}}$) and CIMSS physical retrieval ($\mathbf{X}_{\mathbf{P}}^{\mathbf{C}}$) by using real AIRS observation are provided. For temperature, the mean bias of $\mathbf{X}_{\mathbf{P}}^{\mathbf{W}}$ ranges from −0.5 K to 0.3 K for the entire troposphere, and the RMSE is in the 1.2 – 2.5 K range (Figure 33a). It is noted that these results are significantly improved, in terms of both the mean bias and RMSE, compared with the regression results in Figure 20a. In particular, much improvement is evident for the lower layer near the surface. The better performance demonstrated by this physical algorithm seems to be largely attributed to improving the radiative consistency between the observed radiance and the calculated retrieval radiance. Improved performance of the physical method using CIMSS regression retrievals as inputs is also clear. Although the bias values in the two cases may be similar to each other, $\mathbf{X}_{\mathbf{P}}^{\mathbf{W}}$ seems to yield a smaller RMSE than $\mathbf{X}_{\mathbf{P}}^{\mathbf{C}}$, likely owing to a smaller error in $\mathbf{X}_{\mathbf{R}}^{\mathbf{W}}$, as shown in Figure 20b. It is interesting to note that the RMSEs in the two cases are nearly the same over the 350 – 700 hPa layer and over the layer above the 150 hPa layer, although that of $\mathbf{X}_{\mathbf{R}}^{\mathbf{W}}$ is slightly larger than the CIMSS regression results for that layer, suggesting that the CIMSS physical method is less dependent on the input data

over the 350 – 700 hPa layer.

With regard to the moisture retrieval (Figure 33b), both methods also yield significant improvement, compared with regression retrievals, especially over the surface – 400 hPa layer. Although better results can be expected using the moving-window regression retrieval, which shows a slightly smaller RMSE for the surface – 400 hPa layer, $\mathbf{X}_{\mathbf{P}}^{\mathbf{W}}$ and $\mathbf{X}_{\mathbf{P}}^{\mathbf{C}}$ show nearly the same RMSEs. It is considered that the physical retrieval of water vapor tends to be dominantly determined by radiance rather than by a priori information. In other words, the error associated with the a priori information is relatively larger than that associated with AIRS radiances.

It is worthwhile to note that both water vapor retrievals show consistent dry biases over the upper troposphere and these dry biases may cause large RMSEs over the 200 – 300 hPa layer. Considering that the retrievals have residuals smaller than 1 K, and thus the radiance consistency is strong, we suspect the humid bias of the ERA-I reference data in the upper troposphere. Such humid bias of ERA-I in the upper troposphere against radiosonde observations has been reported (Noh et al. 2016).

In this study, retrieval outcomes having Res ≥ 1 K are considered as not-converged solutions. At this point, it is quite interesting to examine error characteristics of cases not meeting the quality control criterion. Results showing Res ≥ 1 K for both T and q exhibit biases and RMSEs similar to those obtained for Res < 1 K, albeit larger in magnitude, in particular over the lower atmosphere below 700 hPa (Figure 34a and 34b).

Considering similar statistics above 700 hPa level, it is suggested that the lower atmosphere could cause such a larger residual owing to the imbalance between the observed and calculated radiances. Compared with $\mathbf{X}_{\mathbf{P}}^{\mathbf{C}}$, the RMSEs of $\mathbf{X}_{\mathbf{P}}^{\mathbf{W}}$ are smaller, both for temperature and water vapor. For temperature, the RMSEs over the 150 – 300 hPa and 850 – 1000 hPa layers are improved, while for water vapor, the RMSEs over the 200 – 400 hPa and 850 – 1000 hPa layers are improved. In addition, the biases of $\mathbf{X}_{\mathbf{P}}^{\mathbf{W}}$ below the 150 hPa layer are smaller than those of $\mathbf{X}_{\mathbf{P}}^{\mathbf{C}}$.

In conclusion, $\mathbf{X}_{\mathbf{P}}^{\mathbf{W}}$ can be considered more accurate than $\mathbf{X}_{\mathbf{P}}^{\mathbf{C}}$, in terms of both temperature and water vapor retrievals, as shown in Figure 33–34. This conclusion can also be drawn by analyzing the final results of the physical retrieval (Figure 35), which is in fact an average over the results shown in Figure 33–34. In other words, even for optimization procedures that fail to meet the quality control criteria, the physical retrieval model yields a final solution that is at least equivalent to the a priori inputs. It is shown that the mean biases and RMSEs are nearly equivalent to the results in Figure 33a and 33b, suggesting that the majority of physical retrievals satisfy Res < 1 K. Again, it is clearly shown that the regression retrieval with the moving-window technique improves the physical retrievals of temperature and water vapor profiles. For temperature, improved RMSEs are evident in the 150 – 350 hPa layer and below the 700 hPa level. On the other hand, the RMSEs of the humidity profiles are clearly

improved in the layer above ~300 hPa. The RMSEs in the 700 – 950 hPa layer indicate reduced magnitudes associated with the physical retrievals. If we focus on the comparison of accepted solutions, improved RMSEs of temperature are evident in the 150 – 350 hPa layer and below the 900 hPa level. On the other hand, the RMSEs of the humidity profiles are clearly improved in the layer above ~300 hPa. The RMSEs in the 750 – 900 hPa layer indicate reduced magnitudes associated with the physical retrievals. However, it should be noted that the temperature retrieval bias becomes large above the 200 hPa level, and the water vapor retrieval bias is higher at the surface.

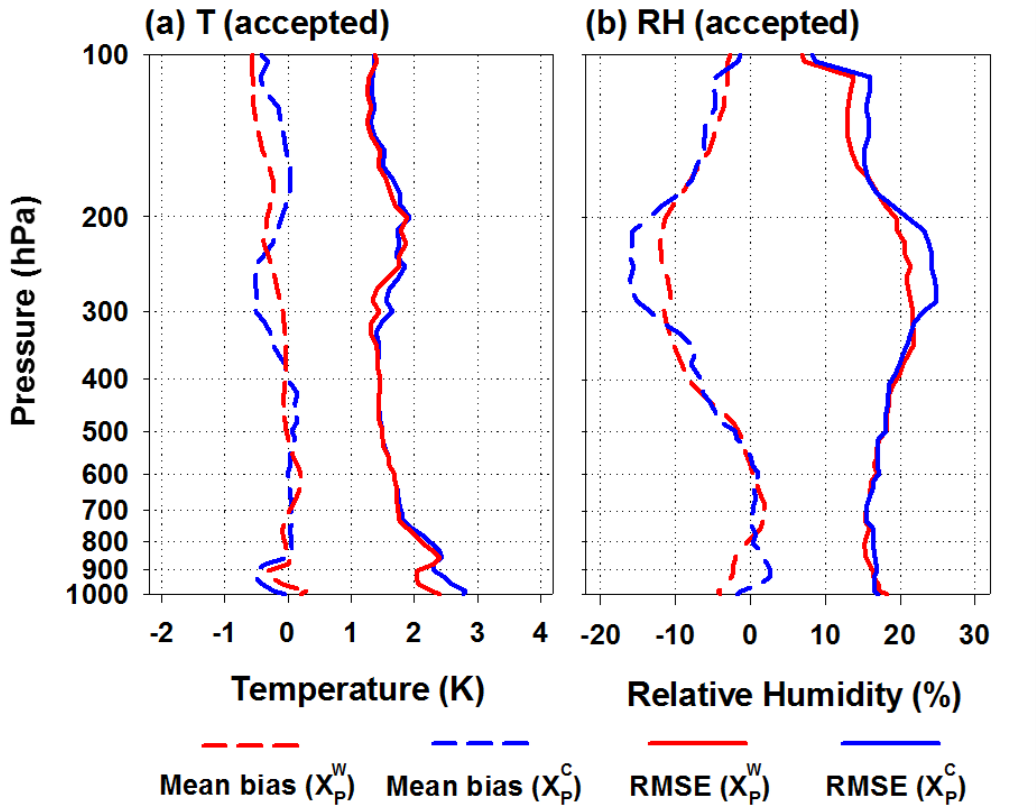

Figure 33. Comparison of the mean biases (dashed lines) and RMSEs (solid lines) between physical retrievals from the moving-window regression (red lines) and CIMSS regression (blue line). Only accepted solutions are plotted.

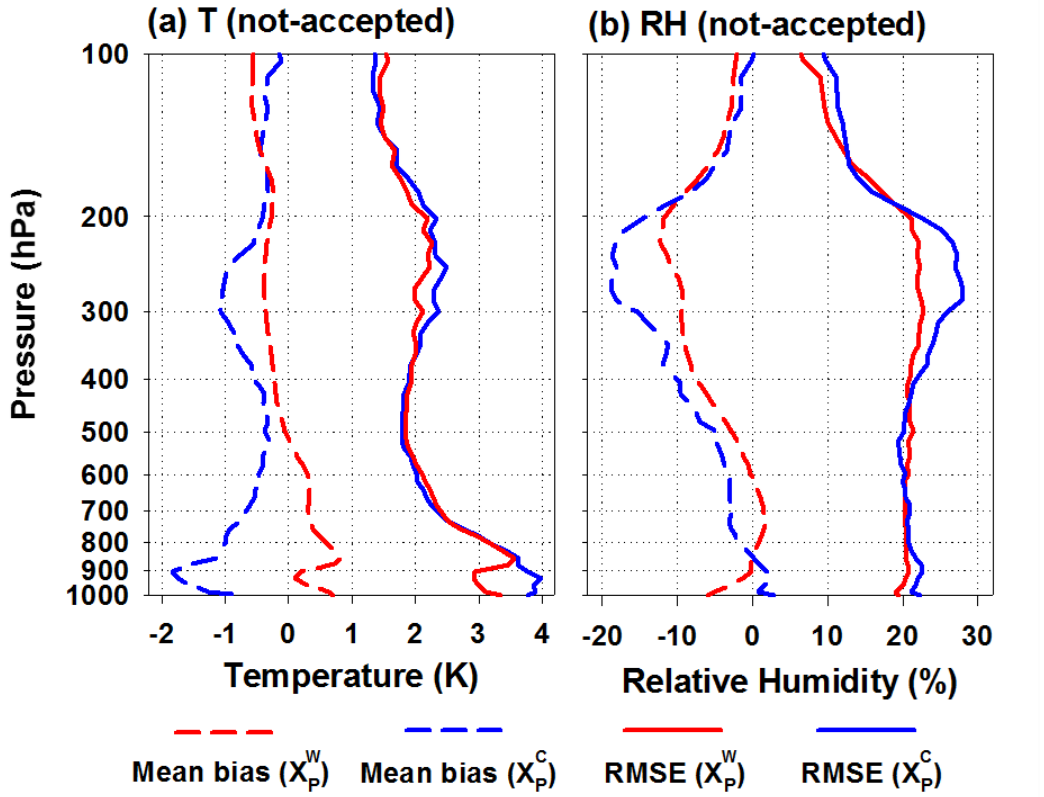

Figure 34. Same as Figure 33, but for not-accepted solutions.

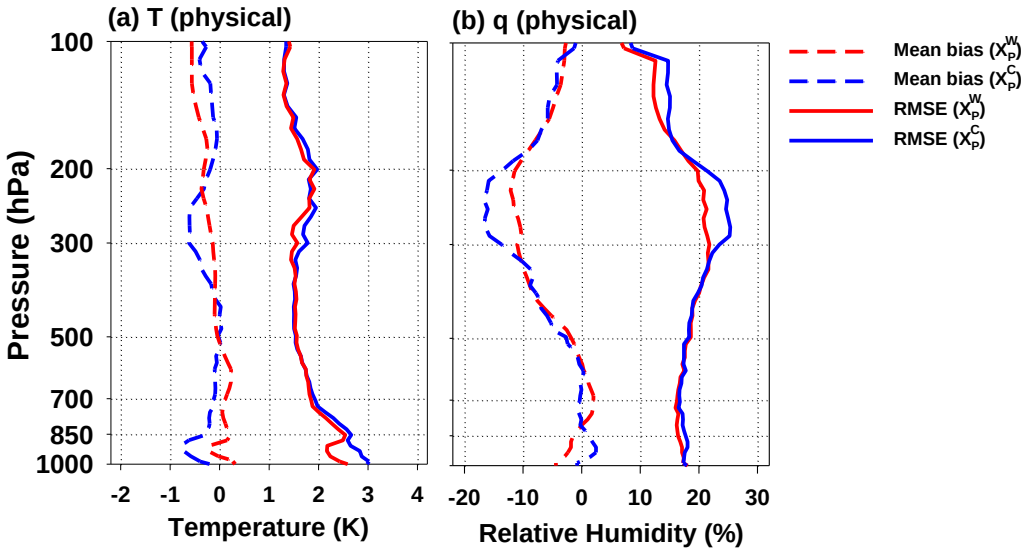

Figure 35. Same as Figure 33, except that all physical retrievals are considered here.

6. Synergetic use of AWS data for AIRS T/q retrievals

Through previous sections, it was noted that climatological features of temperature and moisture profiles successfully enhanced the performance of retrieval. Most of improvement is especially evident over the boundary layer where radiative signal is not easily interpreted. Along this line, it is quite interesting to examine the impact of surface observation data for better retrieval over boundary.

In this study, possible use of Automatic Weather Station (AWS) data operated by Korea Meteorological Administration (KMA) in AIRS temperature and moisture soundings is investigated. Currently, research area is limited over the Korea Peninsula but it can be extended if globally distributed surface observations or buoy data are available. Figure 36 exemplifies the spatial distribution of AWS data for temperature and moisture over the Korean peninsula on September 1, 2010. As shown in the distribution of observed temperature, AWS data in Korea have a very fine spatial resolution. However, due to data quality control problem, moisture data are missing at some stations. Available number of stations increases with time. The nominal accuracy of AWS data is about 0.5 K for temperature and 5 % for relative humidity.

Here, two approaches (i.e., statistical approach and physical approach) are considered to evaluate the role of AWS data in the temperature and moisture retrievals from satellite measurements. In the statistical approach, AWS data are used as additional predictors. All experiment design and used data are the same as in the moving-window

regression development except for the window location. The regression window is centered over the Korean peninsula and then regression coefficient is derived again. The regression coefficients also have 41 predictors (i.e. 40 compressed radiances + surface pressure). On the other hand, new regression coefficient is derived from 43 predictors including 2 more from AWS temperature observation and AWS moisture observation. By comparing two retrieval results, impact of the AWS data can be evaluated. Figure 37 shows the vertical distribution of RMSE in the simulation dataset with no AWS data included. By considering the RMSE after the AWS data are included, the degree of error improvement in % is defined as follows:

$$\text{Error improvement (\%)} = \frac{\text{RMSE}_{\text{w/o AWS}} - \text{RMSE}_{\text{w/ AWS}}}{\text{RMSE}_{\text{w/o AWS}}} \times 100 \qquad (14)$$

where $\text{RMSE}_{\text{w/oAWS}}$ ($\text{RMSE}_{\text{w/AWS}}$) represent RMSE from the retrievals when AWS data are not included (included) as predictors. Positive improvement percentage implies a reduced RMSE by adding two parameters from surface AWS observations.

Since the statistical approach could bring indirect information about the linkage between AWS surface data and atmospheric profile, additional experiment with a physical method may bring in better understating of the impact of AWS data. That is, it is interesting to examine the role of AWS data beyond the correlation between surface observation and atmospheric profile. In doing so, the 1DVAR is modified to use AWS data as additional super channels. Not only error for defining the contribution of AWS data in 1DVAR, but also corresponding Jacobian should be defined. In the statistical

experiment, noises or errors in the experiment may be partly included in the regression coefficient. By contrast, in the physical experiment, misunderstanding of uncertainty directly is attributed to the experiment result. Since the empirical error balancing of 1DVAR (i.e., tuning process) is beyond this study, physical approach is evaluated using the simulation dataset, whose errors or Jacobians are completely controlled. As a priori information, moving-window regression retrievals are used and experiment region is confined within the Korean peninsula. Two retrieval results are compared: one with only AIRS channels, and the other with AWS data plus AIRS channel radiances. Error improvement given in Eq. (5), is also used for examining the error improvement due to the addition of AWS data.

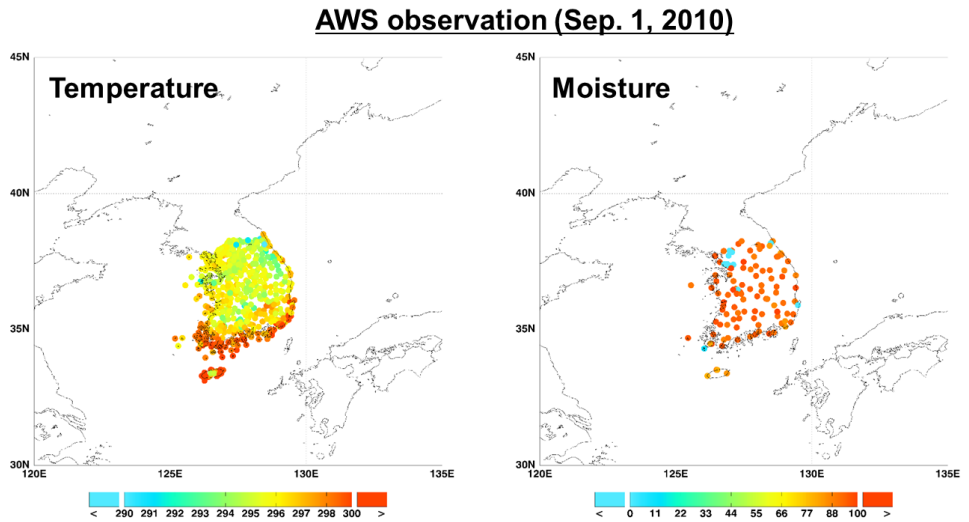

Figure 36. Spatial distribution of AWS data over the Korea peninsula at September 1, 2010. Left and right panels indicate temperature and relative humidity, respectively.

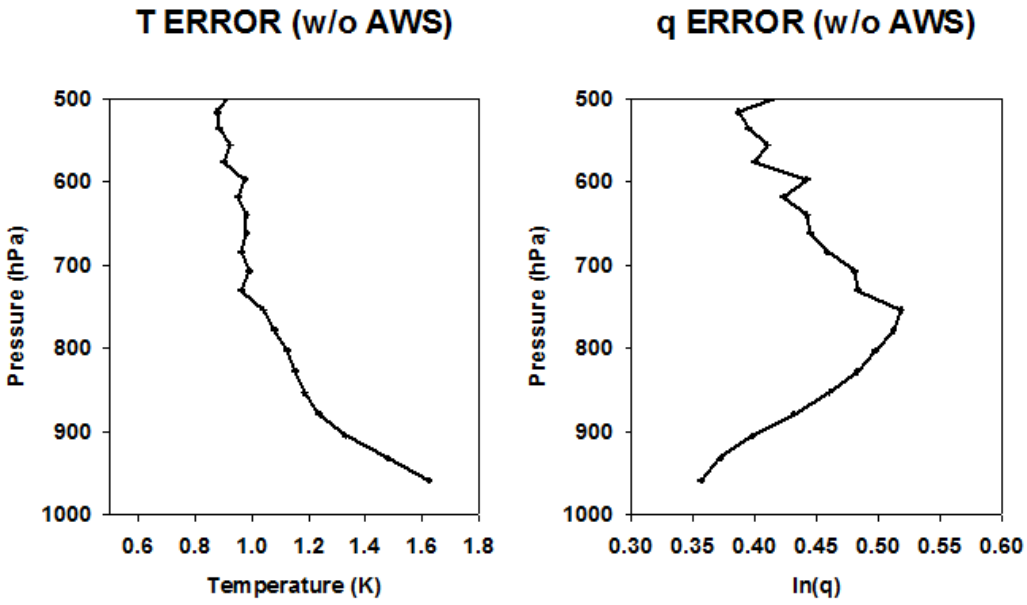

Figure 37. Error statistics for (left) temperature and (right) moisture profile over the Korean peninsula. AWS data are not used as predictors.

6.1. Impact of AWS data on AIRS T/q soundings: Statistical perspective

6.1.1. Pseudo-AWS data for training

Since the current study uses a simulation database, real AWS observations cannot be directly used for the regression procedure. Thus pseudo-AWS data are defined. Four types of pseudo-AWS data are defined and their respective results are shown in Figure 38. Variable names for four types of experiments using various pseudo-AWS data are as follows:

(1) 2m: temperature and moisture at 2m

(2) extp: extrapolated temperature and moisture at surface level from profile data

(3) intp: interpolated temperature and moisture for the surface level from profile data and land model data

(4) lowest: lowest level temperature and moisture from profile data

Figure 38 shows error improvements in the retrieved temperature and moisture profiles when four types of pseudo-AWS are used. For temperature, lowest experiment shows a largest improvement while, extp experiment shows a smallest improvement. Results from 2m and intp experiments are comparable. On the other hand, for moisture, the lowest experiment shows the best improvement, while others show comparable results. For both temperature and moisture retrievals, lowest experiment shows the best

performance since the retrieval solution of lowest level is given as a predictor. More important thing is that such improvement is not limited at lowest level; the information appears to propagate upward up to 700 hPa level. Based on this, it is expected that satellite observations tend to provide information better solving temperature in the 700 hPa - surface layer. Nevertheless, it may not be used because of the ill-posed problem; the information in a satellite channel does not represent a level but a weighted layer. In other words, even if very accurate lowest information is given as a layer information, improvement is made in spite of partly solving the ill-posed problem. Such impact is quite strong especially for the moisture. On the other hand, extp experiment shows a worst result. It can be expected because extrapolation could bring in erroneous results.

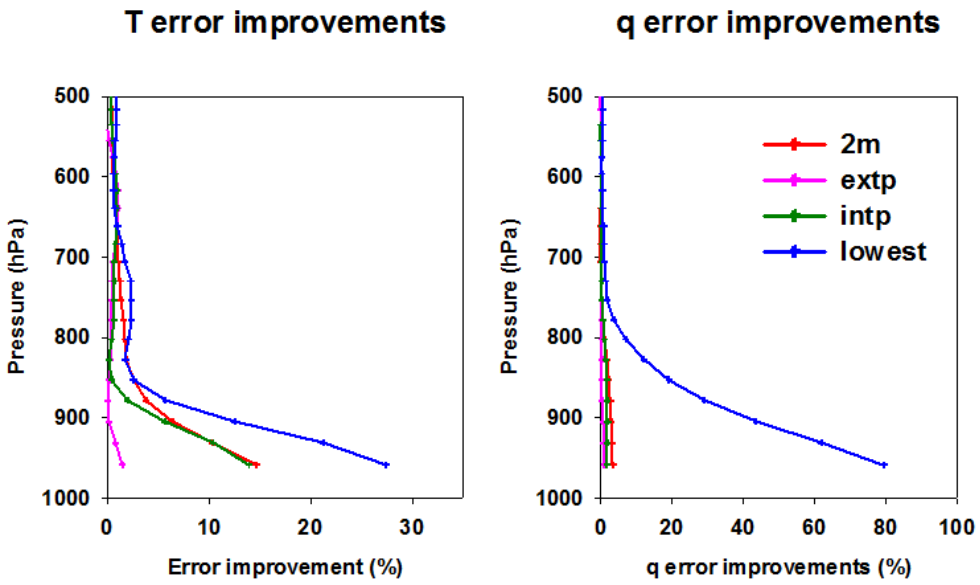

Figure 38. Error improvement for (left) temperature and (right) moisture. Four different types of pseudo-AWS data are assumed. (2m: temperature and moisture at 2m, extp : extrapolated temperature and moisture at surface level from profile data, intp : interpolated temperature and moisture at surface level from profile data and skin data, and lowest : lowest level temperature and moisture from profile data)

6.1.2. Retrieval sensitivity related to error of AWS data

As discussed in section 6.1, the main purpose of conducting the error analysis for the use of pseudo-AWS data is to understand behaviors of regression the coefficient with respect to errors in pseudo-AWS data. This is because the difference between real AWS data and pseudo-AWS data can induce erroneous results. By this error analysis, uncertainty of AWS data itself is also considered.

In this study, as errors the bias and standard deviation are defined. The bias indicates a systematical error, while the standard deviation indicates a random error. In order to simulate an error, the normal distribution is used. For the given normal distribution, five levels of biases and five levels of standard deviations are employed. Considering the nominal temperature error of AWS is to be 0.5 K, various errors are assumed by scaling the nominal error in ways to have 0.05 K, 0.1 K, 0.5 K, 1 K, and 5 K errors. Similarly, for moisture, 0.5 %, 1%, 5%, 10%, and 50 % of relative humidity errors are assumed. Negative relative humidity and super saturated relative humidity are adjusted to have 0 % and 100 %, respectively.

At first, sensitivity of retrieval performance with respect to bias is examined and results are presented in Figure 39. Five different levels of biases for temperature and moisture are separately added to pseudo-AWS data. For example, Figures at left panels show the sensitivity when pseudo-AWS temperature data have biases, while Figures at right panels show the sensitivity when pseudo-AWS moisture data have biases. Upper

and lower panels are for temperature and moisture, respectively. Different colors indicate different amount of errors. As shown in Figure 38, all experiments have positive impact on the error improvement and the improvement reaches until around 700 hPa. There are no negative values although AWS data have significant bias. More important thing is that the bias cannot affect the retrieval performance. This is because the regression model is a linear model. Since the atmospheric states are statistically estimated, regression coefficient successfully absorbs the systematic bias. In case of biased pseudo-AWS moisture data, improvements of larger and largest experiments show different patterns because of the adjusted values. If negative relative humidity or super saturated relative humidity are permitted, those differences are removed (not shown).

In conclusion, the bias of AWS data cannot affect the retrieval performance once the bias is estimated during the regression procedure. In order to estimate the bias of pseudo-AWS data, one of the possible ways is to repeat the training with respect to various assumed biases and then test the coefficient for real AWS data.

By contrast, random error is relatively hard to estimate through the repetition of training. This is because regression coefficient cannot resolve why the regression performance decreases. Not only misunderstanding of AWS data error but also regression procedure brings in RMSE. Figure 40 shows the sensitivity of the retrieval performance with respect to the random error of pseudo-AWS temperature data. Upper

and lower panels are results of temperature and moisture improvements, respectively. Here, three cases can be assumed as follows:

(1) Pseudo-AWS data have random error and it is perfectly known in regression procedure and application

(2) Pseudo-AWS data do not have error (or smaller error than AWS data) but real AWS data of application procedure have unexpected random error

(3) Pseudo-AWS data have random error but real AWS data of application procedure do not have error (or reduced error)

Results of three cases are shown in left, middle and right panels of Figure 40, respectively. Case (1) is an ideal case. Error improvement decreases with increasing AWS data error. This is reasonable because the weight of predictors for AWS data should be reduced when the data are spoiled by random error. There are no negative values in left panels of Figure 40, indicating that larger random error cannot spoil the original retrieval performance (i.e., regression retrievals without AWS data) even if it cannot help for improving the retrieval performance.

Interesting features are found in the case (2). When real AWS data have unexpected error, which is not counted in the regression procedure, use of AWS data spoils regression performance. In other words, retrieval without AWS data can be much better. In this case, regression coefficient provides very high weight to AWS predictors because AWS data have gold information to retrieve atmospheric states. Although real AWS data

have large error, regression highly relies on this bad information in this regression model. In order to avoid case (2), one possible way is that providing artificially larger random error on pseudo-AWS data like case (3). Because this way is very conservative, AWS data cannot fully provide their whole gold information into regression retrieval. Nevertheless, it does not spoil the retrieval algorithm. Note that there are no negative values in case of (3). Almost the same results are shown in Figure 41, which represent the sensitivity of retrieval performance with respect to error of pseudo-AWS moisture data. In case of moisture, unexpected AWS error is very critical compared to temperature case. From this result, it can be expected that role of AWS data of moisture could be much significant than that of temperature.

To sum up, error of AWS predictors has significant impact on regression model. Since regression model is linear model, systematic error can be solved through a repeat of training with respect to various assumed bias. For random error, if amount of random error is not exactly known, conservative approach, which providing slightly larger error than expected error, would be better. This is because overestimation of the weight for AWS predictors may spoil regression model when unexpected random error occurs.

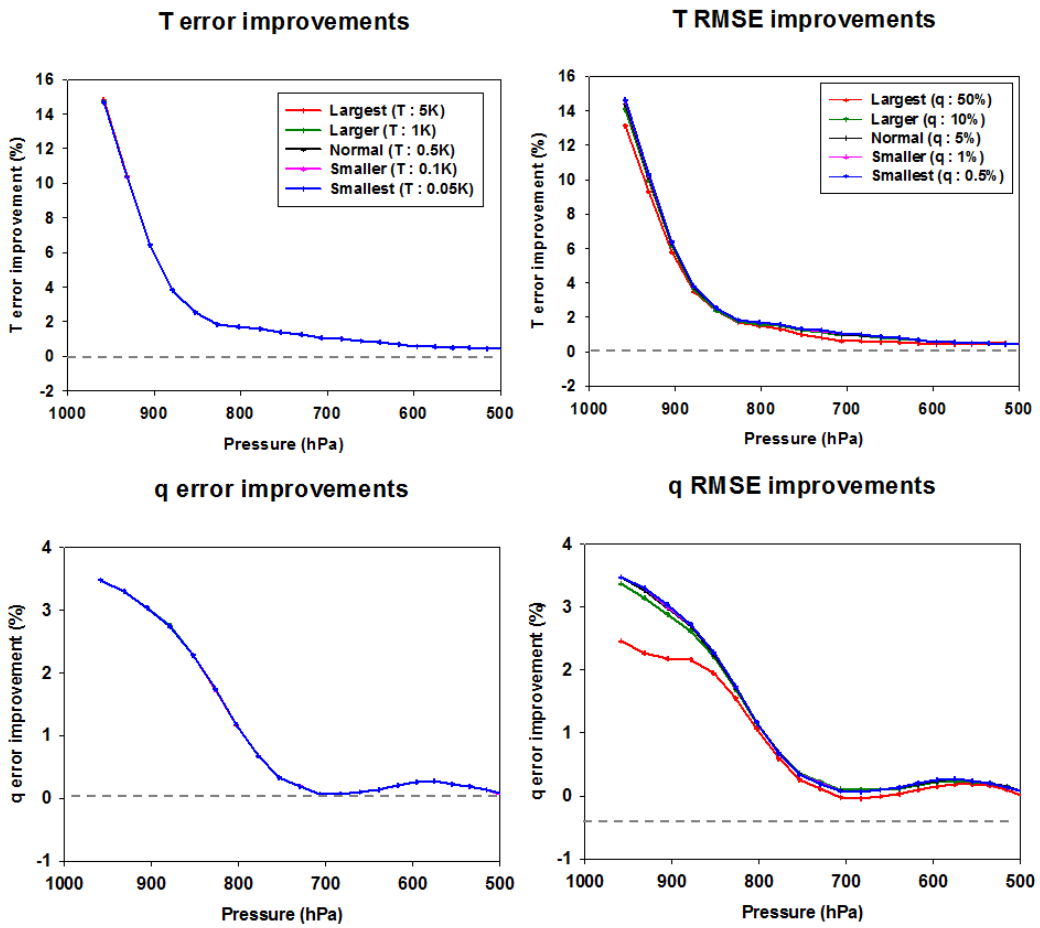

Figure 39. Sensitivity of temperature and moisture retrievals with respect to biased temperature predictor (left) and biased moisture predictor (right). Upper and lower panels indicate results of temperature and moisture retrievals, respectively. Different colors indicate differently assumed biases.

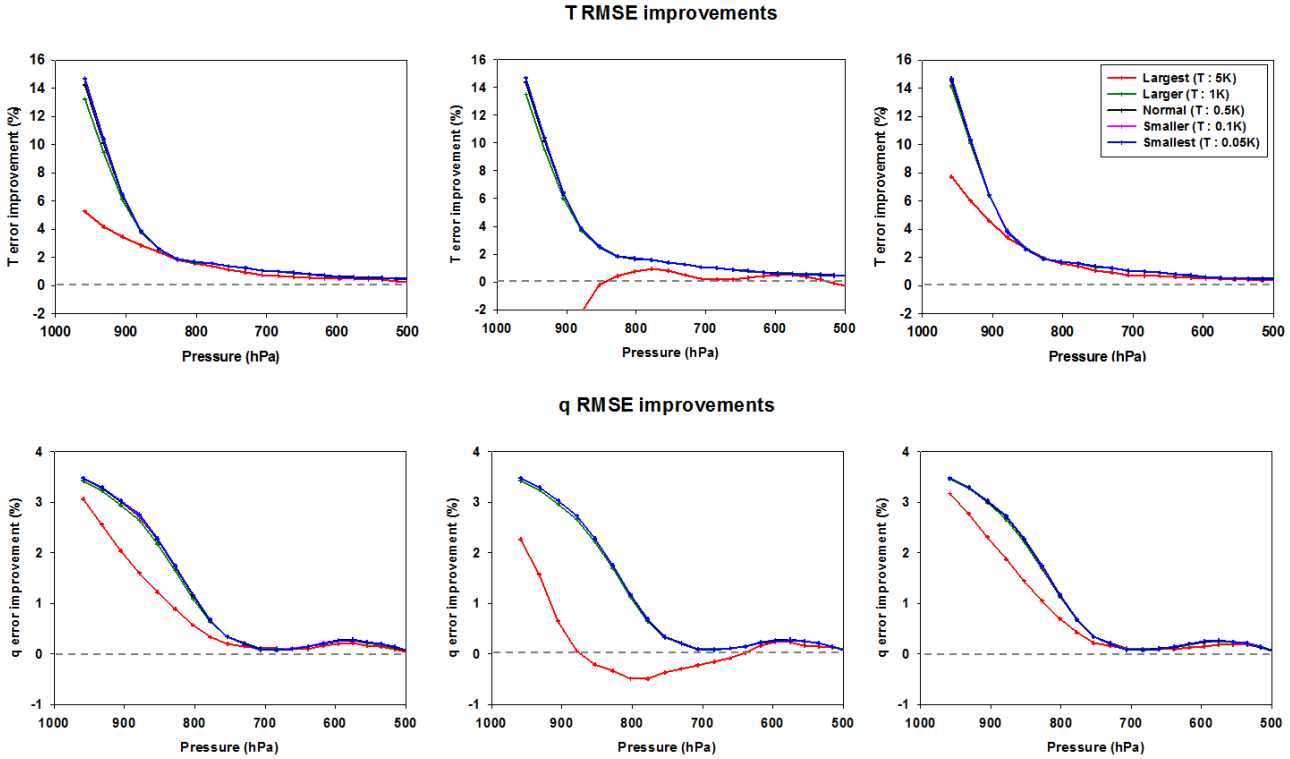

Figure 40. Sensitivity of temperature and moisture retrievals when temperature predictor. Three cases are considered as follows: (left) AWS has random error and it is perfectly known, (middle) AWS has unexpected error, (right) artificial random error is considered for noise-free AWS data. Upper and lower panels indicate results of temperature and moisture retrievals, respectively. Different colors indicate differently assumed random errors.

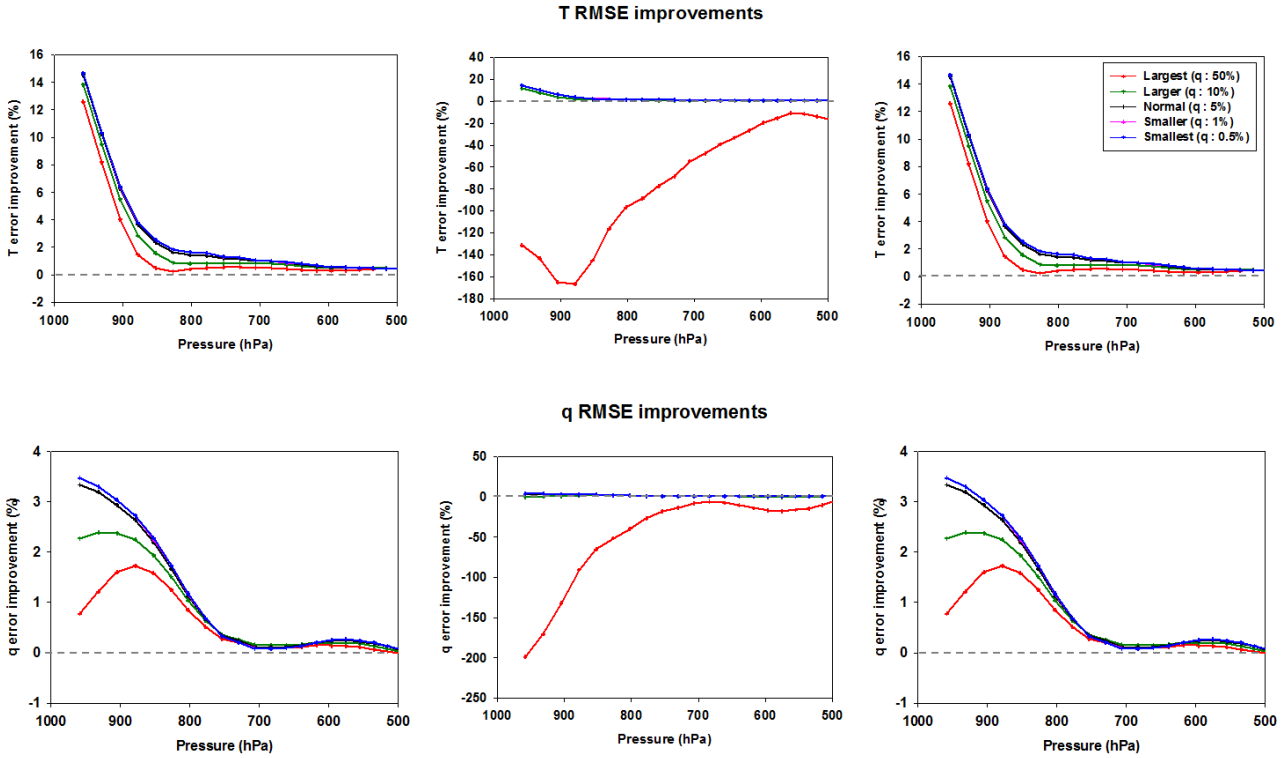

Figure 41. Same as Figure 35 but moisture predictor has various random errors.

6.1.3. Change of regression coefficient due to use of AWS data

For more deep understanding of role of AWS predictors, changes of regression coefficient for lowest level temperature and moisture are analyzed. Before doing that, brief understanding of regression coefficient used in this study is discussed. In the regression model, predictors are simply multiplied to regression coefficient, which links measurements and atmospheric state. That is, amplitude of regression coefficient indicates how much of measurements is converted into physical space. If a given channel has very noisy signal, a regression coefficient related to this channel should be closed to zero because signal from this channel is physically meaningless. By contrast, a given channel has gold information, amplitude of regression coefficient related to this gold channel should be very large. However, if information of gold channel is highly correlated with the other channels, amplitude of regression coefficient could be smaller. This is because atmospheric state can be estimated without this correlated channel. Therefore, it can be expected that change of regression coefficient for AIRS channels could bring information about impact of AWS data for retrieval problem. It should be noted that scale of input signal is also important for deciding amplitude of regression coefficient. It means that even though change of regression coefficient is large, this change cannot affect final retrieval results.

Figure 42 show change of regression coefficient for lowest temperature retrieval. If impact of use of AWS data is significant, change of magnitude is possibly significant.

Similarly, impact is not significant, the magnitude is closed to zero. Red and blue lines indicate coefficient difference between coefficient with and without AWS data. Since blue line uses noisy AWS data as predictors, magnitude of difference is relatively smaller than red line. In order to know which channels are related to which atmospheric parameters mean radiances are plotted also. From the mean spectrum brief channels locations are shown as follows:

- 0 ~ 380: CO_2 channel channels (i.e., temperature channels)
- 380 ~ 650: window channels
- 650 ~ 790: ozone channels
- 790 ~ 900: window channels
- 900 ~ 1430: water vapor channels

As shown in Figure 42, use of AWS data affect whole AIRS spectrum. This indicates that whole AIRS spectrum could directly or indirectly affect lowest temperature retrieval. Considering magnitude of input data, which is shown in green color, relative importance is highest in window channels as expected. Interesting thing is that not only lower temperature sounding channels, but also lower water vapor sounding channels significantly affect lowest temperature retrieval. In other words, gold information of AWS data is highly related to lower level water vapor channels; if AWS data is not available, lower level water vapor channels explain large portion of the lower level temperature.

Impact of AWS data for lowest level humidity is also quite similar to temperature case (Figure 43). As shown in temperature case, whole AIRS spectrum affect lower level humidity retrieval. However, difference of regression coefficient between noisy and noise-free data looks to be much larger than temperature case. It is expected that very accurate AWS data is required to improve lowest level moisture data. Otherwise, improvement of regression retrieval is not that significant. Compared to temperature case, contribution of lower temperature channels and window channels are slightly reduced but role of lower humidity channels is still notable.

To sum up, use of AWS data mostly affects role of window, lower temperature and lower water vapor channels. Among them, role of lower level water vapor channels is quite notable. It can be seen that information of lower level water vapor channels is relatively unstable, so AWS data positively affect retrieval performance through fixing contribution of these channels.

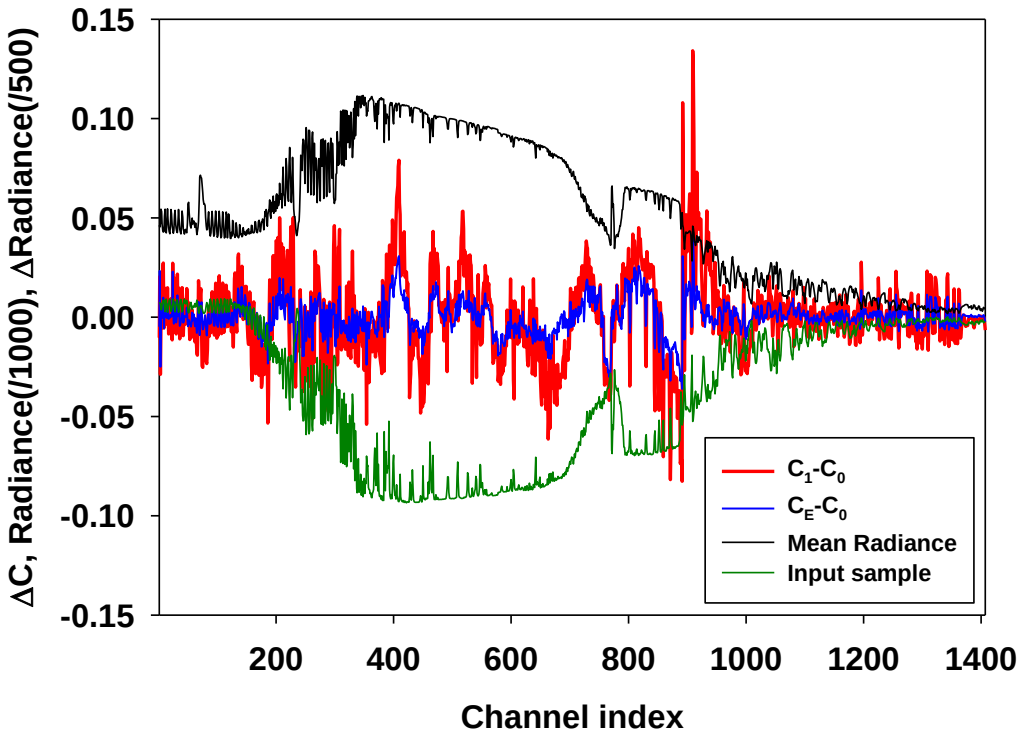

Figure 42. Differences between regression coefficient with and without AWS (red: noise-free AWS, blue: noisy AWS) for lowest level temperature retrieval. Black and green lines indicate mean radiance and magnitude of sample, respectively.

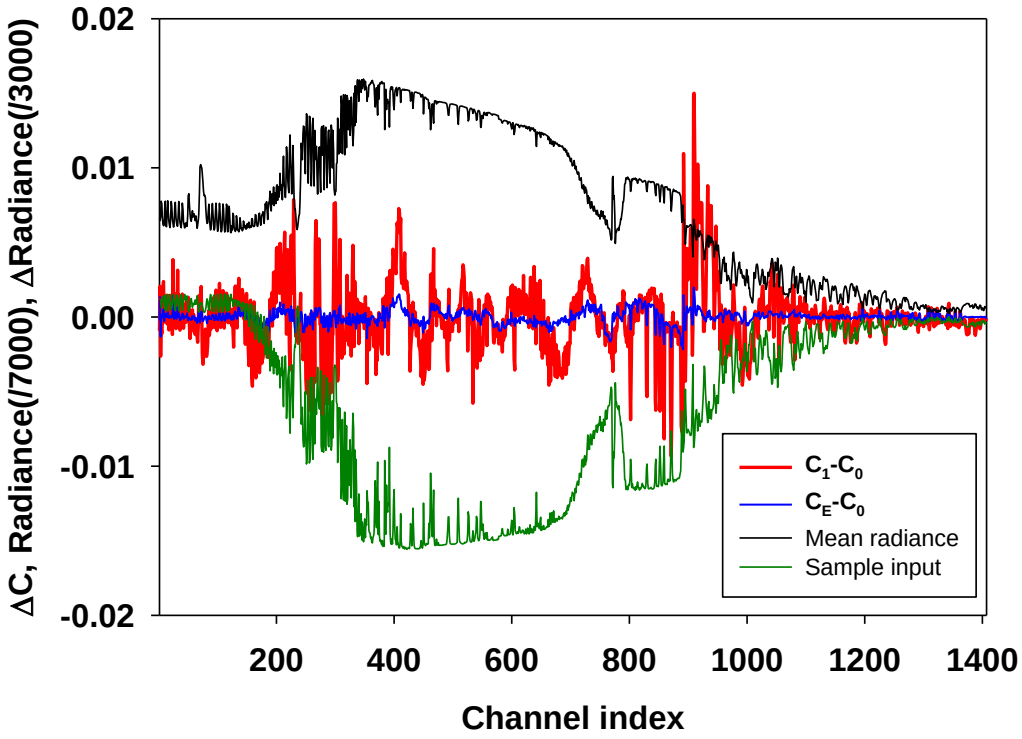

Figure 43. Same as Figure 42 but for lowest level water vapor retrieval.

6.1.4. Application

Now real AWS data is applied on regression coefficient with real AIRS measurements. Details of used data is described in section 4.4.3 but data over the Korean peninsula are used for application. In order to use AWS data as predictors for AIRS retrieval, collocation should be done. For temporal collocation, AWS data observed at AIRS passing time are used, while for spatial collocation, AWS data located in 10 km from center of AIRS field of view are used. It should be noted that unexpected AWS error also includes collocation error.

Comparison results between retrievals with and without AWS are shown in Figure 44. As shown in simulation result in section 6.1, retrievals over boundary layer are improved when AWS data are used as predictors for both temperature and water vapor. For temperature, RMSE is reduced up to 1 K, while for moisture, RMSE is reduced up to 8 % point. As expected, amount of improvement decreases with increasing height. For temperature, impact of AWS reaches around 800 hPa level, while for moisture, the impact reaches around 750 hPa.

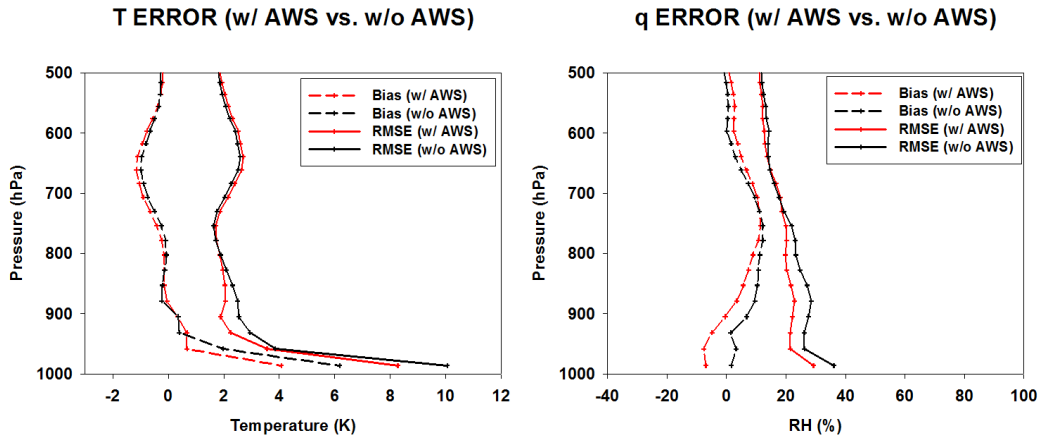

Figure 44. Comparison of the mean biases (dashed lines) and RMSEs (solid lines) between regression retrievals with AWS data (red lines) and without AWS data (blue line).

6.2. Impact of AWS data on AIRS T/q soundings: Physical perspective

6.2.1. 1DVAR with AWS observation

Result shown in section 6.1 suggests that AWS data have enough information for improving boundary layer temperature and moisture retrieval beyond information in AIRS measurements. Since improvement is not limited over AWS data layer, we carefully concluded that improvement could be possible from information synergy between AWS data and AIRS measurements. That is, better explanation for complex radiative information, which is affected by surface, could be possible by adding AWS data in retrieval algorithm. However, one may argue that only correlation between surface states and boundary layer states also brings error improvement. At this point, it is quite interesting to explain this improvement in physical perspective.

Same 1DVAR described in section 5.1 is used. For the experiment, atmospheric state and corresponding AIRS observation are obtained from SARTA simulation result. Details of this simulation data are described in section 4.1. As AWS data, 2 m temperature and moisture data are used. Here, errors of AWS data are assumed that they have relatively very smaller errors than observation and a priori information: 0.5 and 0.1 are assumed for temperature in kelvin unit and moisture in logarithm of mixing ratio unit, respectively.

In order to apply AWS data in 1DVAR, AWS data are considered as AIRS measurements but they do not need radiative transfer calculation. Here, we provide detail step for applying AWS data as satellite measurement by comparing with use of AIRS observation in 1DVAR. In 1DVAR procedure, increments of each channels are calculated at first. In other words, difference between observed AIRS radiance and calculated AIRS radiance of current guess is calculated. Similar difference for AWS data should also be obtained at first. Here, temperature and moisture at lowest level from current guess is considered as calculated AWS data. That is, difference between this lowest value of current guess and AWS data is calculated. After that, this difference is normalized by square root of observation error covariance. For the next step, background increment is calculated. Although this increment is independent with observation, converted value in observation dimension is dependent with characteristics of observation. For AIRS measurements, state increment is converted to observation increment by using Jacobian from radiative transfer calculation. Therefore, we should define Jacobian for AWS data. Considering the observation area of AWS data, delta function type of Jacobian whose peak locates at surface level is assumed. It means that background increment at lowest level is considered as background increment of AWS observation. This increment is also normalized by square root of observation error. Now we have new normalized increments; observation increment of AWS data and background increment in AWS observation dimension. After calculating averaging

kernel by using background error covariance and Jacobian of AIRS and AWS, discussed in 5.1.2, modified state is estimated by multiplying the averaging kernel and the sum of the increment (i.e., normalized observation increment plus normalized background increment). If residual of modified state decreases but it is not less than convergent threshold, one more iteration is proceeded by using modified state as current guess. Details of minimization procedure is shown in Figure 25 in section 5.1.3.

6.2.2. Result

Figure 45 shows error improvement defined in Eq. (14). For both temperature and moisture soundings, error improvements show positive values over boundary layer. Magnitude of improvement is around 4 %. As shown in regression result, improvement is not limited at lowest level where AWS observation is located. Since delta function type of Jacobian is used, AWS observation cannot directly affect boundary layer improvement. The only way to improve boundary layer information from AWS data is that improved near surface state due to AWS observation is used as constraint when AIRS observation improves boundary layer state during minimization procedure.

Compared to regression results, degree of improvement is not that large. The one of the possible reasons is that AIRS measurements and background (i.e., regression retrieval) are too accurate in this experiment. In other words, it is very hard to improve atmospheric state by using single AWS observation for given error states. The other

possible reason is that statistical correlation between surface layer and boundary layer cannot be used for solving AIRS radiative signal. In the regression model, this correlation can be used as an evidence to estimate boundary layer condition. Nevertheless, above result clearly indicates that use of AWS data can improve retrieval over boundary layer in physical perspective. Once well-balanced error value for AWS data is obtained, use of AWS data should be helpful for explaining radiative signal of AIRS measurement.

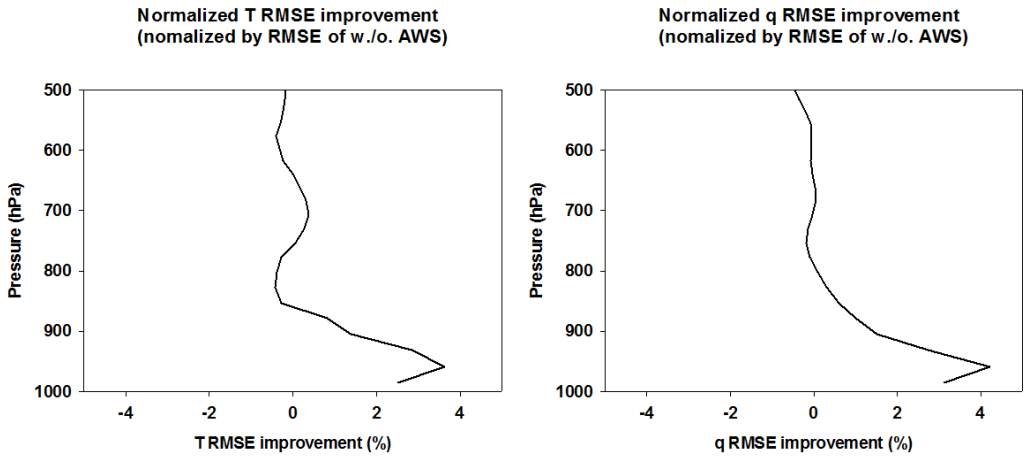

Figure 45. Error improvement for (left) temperature and (right) moisture when AWS data are used as super-channels which have delta function type of Jacobian.

7. Summary and discussion

To improve the temperature and moisture retrievals from hyperspectral AIRS measurements, we attempted first to improve the regression retrievals that can be used as a priori information for the physical retrieval model. In doing so, a moving-window technique was developed based on the assumption that the inclusion of local climate features in the regression procedures will yield better regression results compared with the CIMSS regression method that does not consider local features. The moving-window technique performs the regression at a given location (a 10° × 10° grid box) and at a given time (any season). For regional and global applications, the regression box can be continuously moved to the adjacent location (i.e., another 10° × 10° box) for performing regression on that particular box area until the area in interest is covered. The obtained a priori information was then used to constrain the CIMSS physical model, and the results were compared with those obtained using the a priori information obtained from the CIMSS regression method. In this study, the developed moving-window technique was applied to four months (March, June, September, and December of 2010) of AIRS clear-sky measurements over East Asia.

Notably, regression retrievals based on the moving-window technique yielded smaller mean biases and RMSEs, compared with the CIMSS regression retrievals. Because regional and seasonal climate variations were accounted for by the moving-window technique, the mean bias, implying the accuracy of the estimated climatology,

was substantially improved. The assumption of linearity between the atmospheric state and radiance variables must be more valid when using the moving-window technique, because the atmospheric state in a 10° × 10° regression box is one of many realizations yielding the box climatology. In other words, a relatively smaller deviation from the mean state can be expected, compared with the method in which training data set is collected over a much larger area. By the same token, by regressing over a narrower time window (here, season) a better linear relationship can be expected. In the end, regionally and seasonally varying regression coefficients may result in better performance, compared with performance that is obtained when training data are collected over larger regions and longer times.

The effect of localization of regression coefficients derived from the moving-window technique appears to be more significant in situations in which the relationship between measured radiances and atmospheric state is not clear (e.g., the relationship between the measured radiances and lower boundary layer states). The moving-window technique is particularly advantageous in this situation, compared with the TB-based classification technique used with the CIMSS regression method. The TB-based classification technique may be useful for reflecting the TB dependency on the statistical relationship between the atmospheric state and radiance. However, it is thought that the moving-window technique may result in a less accurate performance in extreme cases, compared with the CIMSS TB-based classification technique, because,

compared with the latter method, the moving-window technique is more susceptible to the variation from the local climatology.

The use of a priori information from the moving-window regression method improved the physical retrieval. For temperature, RMSE improvements of 0.1 – 0.2 K and 0.4 – 0.6 K were demonstrated over the 150 – 300 hPa and 850 – 1000 hPa layers, respectively. For water vapor expressed as relative humidity, RMSEs were reduced by 1.5 – 4.5% above the 300 hPa atmospheric pressure level and by 0.5 – 1.5% over the 700 – 950 hPa layer, compared with the results obtained when CIMSS regression retrievals were used as initial guess fields. However, no effect was observed for the mid-troposphere around 350 – 650 hPa. Water vapor retrievals were not much improved in the lower boundary layer. These benign results suggest that the regression-based improvement of the initial field data may be limited by only using climatological features. Other methods can be devised. For example, including also the numerical weather prediction (NWP) model background in regression procedure can improve the a prior information for 1DVAR physical retrieval (Jin et al. 2008; Schmit et al. 2008; Li et al. 2009), adding different types of observations, such as surface observations, to the retrieval procedures, as suggested by Liu et al. (2015), may be one way of improving the retrievals.

Along this line, possible inclusion of observations of meteorological variables at the surface as inputs to the retrieval model has been examined. In this study, AWS data

over the Korean peninsula are employed as surface observations, further improving the boundary layer temperature and moisture retrievals. In doing so we first conduced the error analysis using the linear regression model. Five levels of biases and random errors are assumed and sensitivity test is performed assessing the retrieval performance. Sensitivity test results suggest that the systematic error of AWS data is not a problem since the regression model is a linear model. However, for the random error, an overestimation of the weight for AWS-related predictors may spoil the regression algorithm if unexpected random errors occur. If the amount of random error is not exactly known, the conservative approach providing a slightly larger error than an expected error, may be better for the stable retrieval. For deeper understanding of the impact of AWS data in the regression model, we examined how regression coefficients are changed with the use of AWS data. Analysis results suggest that use of AWS data mostly affects window, lower temperature and lower water vapor channels. Among them, change of lower level water vapor channels due to AWS data is quite notable. It can be interpreted that the information of lower level water vapor channels is relatively unstable and AWS data positively affect the retrieval performance through altering contribution of these channels.

Results from the modified regression model in which AWS data are used as additional predictors, suggest that AWS data add information for improving the boundary layer temperature and moisture retrieval to the information of AIRS

measurements alone. Since the correlation in the state between the surface air and the boundary layer can enhance the error improvement in the regression approach, the impact of AWS data is also examined in terms of physical perspectives. The 1DVAR algorithm is modified by updating the Jacobians for AWS data. In addition, relative contribution of AWS data compared to a priori information and AIRS measurements is defined in the modified 1DVAR model through the error covariance matrix. For both AIRS temperature and moisture retrievals, RMSEs are improved up to 4% over the boundary layer. Since the delta function type of Jacobian is used, AWS observations cannot directly affect the boundary layer improvement; improved description of the atmospheric condition near the surface due to AWS data may lead to the improvement retrieval in the boundary layer.

It suffices to say that the improvement of temperature and humidity retrievals from the regionally focused algorithm can surely enhance our ability to monitor the severe weather. By the same token, the knowledge obtained from this study would be useful for developing better data assimilation methods for numerical weather forecasting. Future work can be directed at how this algorithm can be applied for future IR hyperspectral sounder, especially for geostationary hyperspectral sounder such as GIIRS on FY-4.

References

Aires, F., A. Chédin, N. A. Scott, 2002: A regularized neural net approach for retrieval of atmospheric and surface temperatures with the IASI instrument. *J. Appl. Meteor*. **41**, 144–159.

August T. 2015: Evolutions and self-organisation of the piece-wise linear regression for IASI, *20th Int. TOVS Study Conf.*, Lake Gevena, WI, USA, University of Wisconsin.

Aumann, H. H., M. T. Chahine, C. Gautier, M. D. Goldberg, E. kalnay, L. M. McMillin, H. Revercomb, P. W. Rosenkranz, W. L. Smith, D. H. Staelin, L. L. Strow, and J. Susskind, 2003: AIRS/AMSU/HSB on the Aqua mission: Design, science objectives, data products, and processing systems. *IEEE Geosci. Remote Sens. Lett.*, **41(2)**, 253–264.

Bisht, J. S. H., P. K. Thapliyal, M. V. Shukla, and R. Kumar, 2015: Improvement in humidity profile retrieval for hyperspectral sounder using principal component-based regression algorithm. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, **99**, doi:10.1109/JSTARS.2014.2366376.

Bloom, M. T., 2001: The Cross-track Infrared Sounder (CrIS): a sensor for operational meteorological remote sensing. *Proc. IGARSS*, **3**, 1341–1343.

Blumstein, D., G. Chalon, T. Carlier, C. Buil, P. Hebert, T. Maciaszek, G. Ponce and T. Phulpin, 2004: IASI instrument: Technical overview and measured performances. *Proc. SPIE*, **5543**, 196–207.

Borbas, E. E., S. W. Seemann, H.-L. Huang, J. Li, and W. P. Menzel, 2005: Global profile training database for satellite regression retrievals with estimates of skin temperature and emissivity. Proc. 14th Int. ATOVS Study Conf., Beijing, China, CIMSS/University of Wisconsin—Madison, 763–770. [Available online at

http://cimss.sec.wisc.edu/itwg/itsc/itsc14/proceedings/B32_Borbas.pdf.]

Borbas, E. E., R. O. Knuteson, S. W. Seemann, E. Weisz, L. Moy, and H.-L. Huang, 2007: A high spectral resolution global land surface infrared emissivity database. *Joint 2007 EUMETSAT Meteorological Satellite Conference and the 15th Satellite Meteorology and Oceanography Conference of the American Meteorological Society*, Amsterdam, Netherlands, Amer. Meteor. Soc., [Available online at http://www.eumetsat.int/website/wcm/idc/idcplg?IdcService=GET_FILE&dDocName=PDF_CONF_P50_S10_03_BORBAS_P&RevisionSelectionMethod=LatestReleased&Rendition=Web.]

Bouttier, F., and P. Courtier, 1999: Data assimilation concepts and methods. Presented at *the Meteorological Training Course Lecture Series.* European Centre for Medium-Range Weather Forecasts, Reading, England.

Carissimo, A., I. De Feis, and C. Serio, 2005: The physical retrieval methodology for IASI: The δ-IASI code. *Environ. Modell. Softw.*, **20**, 1111–1126.

Chahine, M. T., T. S. Pagano, and H. H. Aumann, 2006: AIRS: Improving weather forecasting and providing new data on greenhouse gases. *Bull. Amer. Meteorol. Soc.*, **87(7)**, 911–926.

Chedin, A., N. A. Scott, C. Wahiche, and P. Moulinier, 1985: The Improved Initialization Inversion method: A high resolution physical method for temperature retrievals from satellites of the TIROS-N series. *J. Climate Appl. Meteor.*, **24**, 128–143.Dessler, A. E., P. Yang, and Z. Zhang, 2008: Water-vapor climate feedback inferred from climate fluctuations, 2003–2008. *Geophys. Res. Lett.*, **35**, L20704. doi:10.1029/2008GL035333.

Crosby, D., and M. Weinreb, 1974: Effect of incorrect atmospheric statistics on the accuracy of temperature profiles derived from satellite measurements. *Environ. Modell. Softw.*, **3**, 41–51.

Dee, D. P., S. M. Uppala, A. J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. A. Balmaseda, G. Balsamo, P. Bauer, P. Bechtold, A. C. M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, C. Delsol, R. Dragani, M. Fuentes, A. J. Geer, L. Haimberger, S. B. Healy, H. Hersbach, E. V. Hólm, L. Isaksen, P. Kållberg, M. Köhler, M. Matricardi, A. P. McNally, B. M. Monge-Sanz , J.-J. Morcrette, B.-K. Park, C. Peubey, P. de Rosnay, C. Tavolato, J.-N. Thépaut, and F. Vitart, 2011: The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.*, **137**, 553–597.

Dessler, A. E., P. Yang, and Z. Zhang, 2008: Water-vapor climate feedback inferred from climate fluctuations, 2003–2008. *Geophys. Res. Lett.*, **35**, L20704. doi:10.1029/2008GL035333.

ECMWF, 2009: ERA-Interim project. European Centre for Medium Range Weather Forecasts Public Datasets web interface, accessed 14 November 2012.

EUMETSAT[a], 2016: IASI Level 2: Product Generation Specification. EPS.SYS.SPE.990013, 124 PP, [Available online at http://www.eumetsat.int/website/wcm/idc/idcplg?IdcService=GET_FILE&dDocName=PDF_DMT_743506&RevisionSelectionMethod=LatestReleased&Rendition=Web.]

EUMETSAT[b], 2016: IASI L2 PPF v6.2 – Validation Report. EUM.RSP.REP.16/857500, 74 PP, [Available online at http://www.eumetsat.int/website/wcm/idc/idcplg?IdcService=GET_FILE&dDocName=PDF_DMT_857500&RevisionSelectionMethod=LatestReleased&Rendition=Web.]

Eyre, J. R., 1989: Inversion of cloudy satellite sounding radiances by nonlinear optimal estimation. I: Theory and simulation for TOVS. *Q. J. R. Meteorol. Soc.*, **115**, 1001–1026.

Fourrié, N., and Thépaut J.-N., 2003. Evaluation of the AIRS near-real-time channel selection for application to numerical weather prediction. *Q. J. R. Meteorol.*

Soc., **129**, 2425–2439.

Gambacorta, A., 2013: The NOAA Unique CrIS/ATMS processing system (NUCAPS): Algorithm theoretical basis documentation, 78 PP, [Available online at http://www.ospo.noaa.gov/Products/atmosphere/soundings/nucaps/docs/NUCAPS_ATBD_20130821.pdf.]

Grieco, G., G. Masiello, C. Serio, R. L. Jones, and M. I. Mead, 2011: Infrared Atmospheric Sounding Interferometer correlation interferometry for the retrieval of atmospheric gases: the case of H_2O and CO_2. *Appl. Opt.*, **50**(22), 4516–4528.

Goldberg, M. D., Y. Qu, L. M. McMillin, W. Wolf, L. Zhou, M. Divakarla, 2003: AIRS near-real-time products and algorithms in support of operational numerical weather prediction. *IEEE Trans. Geosci. Remote Sens.*, **41**, 379–389.

Han, Y. and Coauthors, 2013: Suomi NPP CrIS measurements, sensor data record algorithm, calibration and validation activities, and record data quality, *J. Geophys. Res. Atmos.*, **118**, 12734-12748, doi:10.1002/2013JD020344.

Hanel, R. A., B. Schlachman, D. Rogers, and D. Vanous, 1971: Nimbus 4 Michelson Interferometer. *Appl. Opt.*, **10**, 1376-1382.

Hayden, C. M., 1988: GOES-VAS simultaneous temperature-moisture retrieval algorithm. *J. Appl. Meteor.*, **27**, 705–733.

Hilton, F., and Coauthors, 2012: Hyperspectral Earth observations from IASI: Five years of accomplishments. *Bull. Amer. Meteor. Soc.*, **93**, 347–370.

Ho, S.-P., J. Gille, D. Edwards, J. Attie, M. Deeter, J. Warner, G. Francis, and D. Ziskin, 2001: The role of a priori information in the retrieval of CO profiles from Terra-MOPITT measurements. *11th Conference on Satellite Meteorology and Oceanography*, Madison, WI, USA, P5.53.

Huang, H.-L., and P. Antonelli, 2001: Application of principal component analysis to

high- resolution infrared measurements compression and retrieval. *J. Appl. Meteor.*, **40**, 365–388.

Jin, X., J. Li, T. J. Schmit, J. Li, M. D. Goldberg, and J. J. Gurka, 2008: Retrieving clear-sky atmospheric parameters from SEVIRI and ABI infrared radiances. *J. Geophys. Res.*, **113**, D15310. doi:10.1029/2008JD010040.

Kahn, B. H., and Coauthors, 2011: Temperature and water vapor variance scaling in global models: Comparisons to satellite and aircraft data. *J. Atmos. Sci.*, **68**, 2156–2168, doi:10.1175/2011JAS3737.1.

Kwon, E.-H., J. Li, J. Li, B. J. Sohn and E. Weisz, 2012a: Use of total precipitable water classification of a priori error and quality control in atmospheric temperature and water vapor sounding retrieval. *Adv. Atmos. Sci.*, **29**, 263–273.

Kwon, E.-H., B. J. Sohn, W. L. Smith, and J. Li, 2012b: Validating IASI temperature and moisture sounding retrievals over East Asia using radiosonde observations. *J. Atmos. Ocean. Tech.*, **29**, 1250–1262.

Luo, M., C. P. Rinsland, C. D. Rodgers, J. A. Logan, H. Worden, S. Kulawik, A. Eldering, A. Goldman, M. W. Shephard, M. Gunson, and M. Lampel, 2007: Comparison of carbon monoxide measurements by TES and MOPITT: Influence of a priori data and instrument characteristics on nadir atmospheric species retrievals. *J. Geophys. Res.*, **112**, D09303, doi: 10.1029/2006JD007663.

Le Marshall, J., and Coauthors, 2006: Improving global analysis and forecasting with AIRS. *Bull. Amer. Meteor. Soc.*, **87**, 891–894.

Levenberg, K., 1944: A method for the solution of certain non-linear problems in least squares. *Quar. Appl. Math.*, **2**, 164-168.

Li, J., 1994: Temperature and water vapor weighting functions from radiative transfer equation with surface emissivity and solar reflectivity. *Adv. Atmos. Sci.*, **11**, 421–426.

Li, J., and H.-L. Huang, 1999: Retrieval of atmospheric profiles from satellite sounder measurements by use of the discrepancy principle. *Appl. Opt.*, **38**, 916–923.

Li, J., W. Wolf, W. Menzel, W. Zhang, H. Huang, and T. Achtor, 2000: Global soundings of the atmosphere from ATOVS measurements: The algorithm and validation. *J. Appl. Meteor.*, **39**, 1248–1268.

Li, J., W. P. Menzel, F. Sun, T. J. Schmit, and J. Gurka, 2004: AIRS subpixel cloud characterization using MODIS cloud products. *J. Appl. Meteor.*, **43**, 1083–1094.

Li, J., J. Li, E. Weisz, and D. K. Zhou, 2007: Physical retrieval of surface emissivity spectrum from hyperspectral infrared radiances. *Geophys. Res. Lett.*, **34**, L16812, doi:10.1029/2007GL030543.

Li, Z., J. Li, W. P. Menzel, J. P. Nelson III, T. J. Schmit, E. Weisz, and S. A. Ackerman, 2009: Forecasting and nowcasting improvement in cloudy regions with high temporal GOES sounder infrared radiance measurements. *J. Geophys. Res. Atmos.*, **114**, D09216, doi:10.1029/2008JD010596.

Liu, C.-Y., G.-R. Liu, T.-H. Lin, C.-C. Liu, H. Ren and C.-C. Young. 2014: Using surface stations to improve sounding retrievals from hyperspectral infrared instruments. *IEEE Trans. Geosci. Remote Sens.*, **52**(11), 6957–6963.

Ma, X., T. Schmit, and W. Smith, 1999: A nonlinear physical retrieval algorithm—Its application to the GOES-8/9 sounder. *J. Appl. Meteor.*, **38**, 501–513.

Marquardt, D., 1963: An algorithm for least-squares estimation of nonlinear parameters. *J. Appl. Math.*, **11**(2), 431-441.

Masiello, G., and C. Serio, 2013: Simultaneous physical retrieval of surface emissivity spectrum and atmospheric parameters from infrared atmospheric sounder interferometer spectral radiances. *Appl. Opt.*, **52**(11), 2428-2446, doi: 10.1364/AO.52.002428.

Nagle, F. W., and R. E. Holz, 2009: Computationally efficient methods of collocating satellite, aircraft, and ground observations. *J. Atmos. Oceanic Technol.*, **26**, 1585–1595, doi:10.1175/2008JTECHA1189.1

NASA, 2007: AIRS/Aqua L1B Infrared geolocated and calibrated radiances V005, version 005, Goddard Earth Sciences Data and Information Services Center, accessed 27 November 2014.

Noh, Y.-C., B. J. Sohn, Y. Kim, S. Joo, and W. Bell, 2016: Evaluation of Temperature and Humidity Profiles of Unified Model and ECMWF Analyses Using GRUAN Radiosonde Observations. *Atmosphere*, **7**, 94.

Olsen, E. T., 2013: AIRS/AMSU/HSB Version 6 changes from version 5. NASA-JPL Tech. Rep., 28 PP. [Available online at http://disc.sci.gsfc.nasa.gov/AIRS/documentation/v6_docs/v6releasedocs-1/V6_Changes_from_V5.pdf.]

Menke, W., 1984: Geophysical data analysis: Discrete inverse theory, *Academic Press*, New York, USA.

Picon, L., R. Roca, S. Serrar, J. L. Monge, and M. Desbois, 2003: A new METEOSAT ‘‘water vapor’’ archive for climate studies. *J. Geophys. Res.*, **108**, doi:10.1029/2002JD002640.

Prunet, P., J.-N. Thépaut, and V. Cassé, 1998: The information content of clear-sky IASI radiances and their potential for numerial weather prediction. *Q. J. R. Meteorol. Soc.*, **124**, 211–241.

Prunet, P., V. Cassé, and J.-N. Thépaut, 2001: Infrared atmospheric sounding interferometer data information content; instrument characterization and the impact of a priori information. *Tellus*, **53A**, 380–402.

Rodgers, C. D., 1966: Satellite infrared radiometer; A discussion of inversion methods, *University of Oxford Clarendon Laboratory Memorandum* No. 66.13, England, 25 PP.

Rodgers, C. D., 1976: Retrieval of atmospheric temperature and composition from remote measurements of thermal radiation. *Rev. Geophys. Space Phys.*, **14**, 609–624.

Rodgers, C. D., 1996: Information content and optimization of high spectral resolution measurements. Presented at *Optical spectroscopic techniques and instrumentation for atmospheric and space research II. SPIE*, **2830**.

Schmit, T. J., J. Li, J. J. Gurka, M. D. Goldberg, K. Schrab, J. Li, and W. Feltz, 2008: The GOES-R Advanced Baseline Imager and the continuation of GOES-N class sounder products. *J. Appl. Meteor. Climatol.*, **47**, 2696–2711.

Schmetz, J., C. Geijo, W. Menzel, K. Strabala, L. van de Berg, K. Holmlund, and S. Tjemkes, 1995: Satellite observations of upper tropospheric relative humidity, clouds and wind field divergence. *Contrib. Atmos. Phys.*, **68**, 345–357.

Seemann, S. W., E. E. Borbas, R. O. Knuteson, G. R. Stephenson, and H.-L. Huang, 2008: Development of a global infrared land surface emissivity database for application to clear sky sounding retrievals from multispectral satellite radiance measurements. *J. Appl. Meteor. Climatol.*, **47**, 108–123.

Serio, C., and Coauthors, 2015: Infrared atmospheric sounder interferometer radiometric noise assessment from spectral residuals. *Appl. Opt.*, **54**(19), 5924–5936.

Smith, W. L., H. M. Woolf, W. J. Jacob, 1970: A regression method for obtaining real-time temperature and geopotential height profiles from satellite spectrometer measurements and its application to NIMUS 3"SIRS" observations. *Mon. Wea. Rev.*, **98**, 582–603.

Smith, W. L., and H. M. Woolf, 1976: The use of eigenvectors of statistical covariance matrices for interpreting satellite sounding radiometer observations. *J. Atmos. Sci.*, **33,** 1127–1140.

Smith, W. L., H. M. Woolf, C. M. Hayden, and A. J. Schreiner, 1985: The simultaneous

export retrieval package. *Tech. Proc. Second Int. TOVS Study Conf.*, Igis, Austria, CIMSS, 224–253.

Smith, W. L., E. Weisz, S. V. Kireev, D. K. Zhou, Z. Li, E. E. Borbas, 2012: Dual-regression retrieval algorithm for real-time processing of satellite ultraspectral radiances. *J. Appl. Meteorol. Clim.*, **51**, 1455-1476.

Smith, W. L., E. Weisz, H. Revercomb, and N. Smith, 2015: The retrieval of atmospheric profiles from satellite radiances for NWP data assimilation, *20th Int. TOVS Study Conf.*, Lake Gevena, WI, USA, University of Wisconsin, [Available online at https://cimss.ssec.wisc.edu/itwg/itsc/itsc20/papers/4_04_smith_paper.pdf.]

Strow, L. L., S. E. Hannon, S. D. Souza-Machado, H. E. Motteler, and D. Tobin, 2003: An overview of the AIRS radiative transfer model. *IEEE*, **41**, 303–313.

Susskind, J., J. Rosenfield, D. Reuter, and M. T. Chahine, 1984: Remote sensing of weather and climate parameters from HIRS2/MSU on TIROS-N. *J. Geophys. Res.*, **89**, 4677–4697.

Susskind, J., C. Barnet, and J. Blaisdell, 2003: Retrieval of atmospheric and surface parameters from AIRS/AMSU/HSB data in the presence of clouds. *IEEE Trans. Geosci. Remote Sensing*, **41(2)**, 390–409.

Thapliyal, P. K., M. V. Shukla, J. H. Bisht, P. K. Pal and R. R. Navalgund, 2012: Improvement in the retrieval of humidity profiles using hybrid regression technique from infrared sounder data: a simulation study. *Meteorol. Appl.*, doi: 10.1002/met.1330.

Tian, B., D. E. Waliser, E. J. Fetzer, and Y. L. Yung, 2010: Vertical moist thermodynamic structure of the Madden-Julian Oscillation in Atmospheric Infrared Sounder retrievals: An update and a comparison to ECMWF interim re-analysis, *Mon. Wea. Rev.*, **138**(12), 4576-4582, 10.1175/2010mwr3486.1.

Thompson, O. E., M. D. Goldberg and D. A. Dazlich, 1985: Pattern recognition in the

satellite temperature retrieval problem. *J. Climate Appl. Meteor.*, **24**, 30–48.

Twomey, S., 1963: On the numerical solution of the Fredholm integral equations of the first kind by inversion of the linear system produced by quadrature. *J. Assoc. Comput. Mach.*, **10**, 79 –101.

Wark, D. Q., and H. E. Fleming, 1966: Indirect measurements of atmospheric temperature profiles from satellite: I. Introduction. *Mon. Wea. Rev.*, **94**(6), 351-362.

Westwater, E., and O. N. Strand, 1968: Statistcal information content of radiation measurements used in indirect sensing. *J. Atmos. Sci.*, **25**(5), 750-758.

Weisz, E., H.-L. Huang, J. Li, S. W. Seemann, E. Borbas, and L. Gumley, 2003: AIRS real-time sounding profile retrieval for IMAPP users. *13th Int. TOVS Study Conf.*, Ste-Adele, QC, Canada, University of Wisconsin, 323–330, [Available online at http://library.ssec.wisc.edu/research_Resources/publications/pdfs/ITSC13/weisz01_ITSC13_2003.pdf.]

Weisz, E., H.-L. Huang, J. Li, E. Borbas, K. Baggett, P. Thapliyal, L. Guan, 2007: International MODIS and AIRS processing package: AIRS products and applications. *J. Appl. Remote Sens.*, **1**, 0135519, doi:10.1117/1.2766867.

Weisz, E., W. L. Smith, and N. Smith, 2013: Advances in simultaneous atmospheric profile and cloud parameter regression based retrieval from high-spectral resolution radiance measurements. *J. Geophys. Res. Atmos.*, **118**, 6433–6443, doi:10.1002/jgrd.50521.

Zhang, J., Z. Li, J. Li, and J. Li, 2014: Ensemble retrieval of atmospheric temperature profiles from AIRS. *Adv. Atmos. Sci.*, **31**(3), 559–569.

Zheng, J., Jun Li, T. J. Schmit, Jinlong Li, and Z. Liu, 2015: The impact of AIRS atmospheric temperature and moisture profiles on hurricane forecasts: Ike (2008) and Irene (2011). *Adv. Atmos. Sci.*, **32**, 319 - 335.

국문초록

Atmospheric Infrared Sounder (AIRS) 온도/습도 연직 분포 산출을 위한 지역기반 알고리즘 개발 연구를 수행하였다. 첫째로 AIRS로 관측한 적외 초분광 복사 휘도로부터 온도/습도 연직 분포를 산출할 수 있는 통계 모델을 구축하였다. 통계 모델을 구축함에 있어 지역 및 계절에 따라 공통적으로 나타나는 온도/습도 연직 분포의 형태를 산출에 활용할 수 있도록 이동창 방법을 개발, 적용하였다. 통계 모델을 구축하기 위한 훈련 자료로 4년 (2006-2009) 간의 ECMWF 재분석 자료와 이 재분석 자료로부터 이론적으로 계산된 AIRS 복사 휘도가 사용되었다. 이동창 방법이 적용되지 않은 통계적 산출물에 비해 뚜렷한 편차의 개선이 나타났고, 이는 이동창 방법을 이용한 기후 정보의 활용이 보다 정확한 통계 모델 구축함에 있어 도움이 될 수 있음을 의미한다.

물리적 측면에서의 추가적 개선을 위한 1차원 변분법 역시 본 지역기반 온도/습도 연직 분포 산출 알고리즘에 포함되었다. 사전에 잘 구축된 통계적 산출물과 에러공분산행렬 쌍을 이용하여, 간접적인 방법으로 이동창 기반의 통계적 산출물의 에러공분산행렬을 산정한 후 1차원 변분법에 적용하였다. 이동창 방법으로부터 얻은 통계적 산출물을 초기장으로 한 물리적 산출물의 정확도를 평가하기 위하여, 4개월 (2010년 3월, 6월, 9월, 12월) 간의 AIRS 청천 관측 밝기 온도로부터 온도/습도 연직 분포를 산출하여 CIMSS의 통계적 산출물을 초기장으로 사용한 물리적 산출물과 비교 분석을 수행하였다. 비교

대상으로 사용된 CIMSS의 통계 모델은 전구 기반의 훈련 자료로부터 개발된 산출 모델로, 대기 창 영역에서 관측된 밝기 온도를 이용하여 전 지구의 대기 상태를 여섯 종류로 구분하여 AIRS의 온도/습도를 산출한다. 물리적 산출물의 비교 결과, 이동창 방법을 통해 생산한 통계적 산출물을 기반으로 한 물리적 산출물이 평균 제곱근 오차 측면에 있어 보다 개선된 결과를 보였다. 온도 산출물의 경우 150 – 300 hPa 층과 900 – 1000 hPa 층에서 각각 0.1 – 0.2 K, 0.25 – 0.5 K의 평균 제곱근 오차 개선을 보였고, 상대습도 산출물의 경우 300 hPa 이상의 고도에서 1.5 – 3.5%, 700 – 950 hPa 층에서 0.5 – 1 %의 평균 제곱근 오차 개선을 보였다.

이동창 방법을 이용했을 때의 대부분의 오차 개선이 관측된 복사 휘도와 대기상태 간의 관계가 명확하지 않은 곳에서 나타났기 때문에, 유사한 방법으로 추가적인 AIRS 온도/습도 산출 개선을 위한 지상 관측 자료의 활용 가능성을 조사하였다. 지상 관측 자료는 통계적, 물리적 방법으로 개발된 지역기반 AIRS 산출 알고리즘에 적용되었다. 적용 결과 지상 관측 자료가 생산된 지면 고도 뿐 만 아니라 경계층 전체에서 AIRS 온도/습도 산출물의 평균 제곱근 오차가 개선되었다. 이러한 결과는 하층 대기 채널의 산출에 있어서의 불안정성이 지상 관측 자료의 사용으로 인해 해소됨으로써 나타난 것으로 해석된다.

종합하면 본 연구에서는 보다 나은 AIRS 온도/습도 연직 구조 산출을 위한 기후 값이 고려된 지역 기반 알고리즘을 개발하였고, 추가적인 개선을

위한 지면 관측 자료의 활용 가능성을 논의 하였다. 위성 관측 기술이 발전함에 따라 정지 궤도 위성의 적외 초분광 관측이 가능해졌기 때문에, 개발된 지역 기반 알고리즘을 이러한 차세대 위성에 적용시킬 경우 악기상 감시 및 예측 능력의 큰 향상을 기대할 수 있을 것이라 판단된다.

주요어: 온도 연직 분포, 습도 연직 분포, 적외 초분광 위성, AIRS, 통계 산출 모델, 1차원 변분법

학 번: 2008-20416