AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Atmospheres

RESEARCH ARTICLE
10.1002/2015JD024008

Key Points:
- We present validation results for neural network retrievals of temperature and water vapor
- We show accuracy and resolution performance for AIRS/AMSU and CrIS/ATMS
- Results demonstrate significant performance improvement over previous AIRS/AMSU retrieval version

Correspondence to:
A. B. Milstein,
milstein@ll.mit.edu

Citation:
Milstein, A. B., and W. J. Blackwell (2016), Neural network temperature and moisture retrieval algorithm validation for AIRS/AMSU and CrIS/ATMS, *J. Geophys. Res. Atmos.*, *121*, 1414–1430, doi:10.1002/2015JD024008.

Received 28 JUL 2015
Accepted 18 DEC 2015
Accepted article online 20 DEC 2015
Published online 18 FEBRUARY 2016

Neural network temperature and moisture retrieval algorithm validation for AIRS/AMSU and CrIS/ATMS

Adam B. Milstein[1] and William J. Blackwell[1]

[1]Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, Massachusetts, USA

Abstract We present comprehensive validation results for the recently introduced neural network technique for retrieving vertical profiles of atmospheric temperature and water vapor from spaceborne microwave and hyperspectral infrared sounding instruments. This technique is currently in operational use as the first guess for the NASA Atmospheric Infrared Sounder (AIRS) Science Team Version 6 retrieval algorithm. The validation incorporates a variety of data sources, independent from the algorithm training set, as ground truth, including global numerical weather analyses generated by the European Center for Medium-Range Weather Forecasts, synoptic radiosonde measurements, and radiosondes dedicated for validation. The results demonstrate significant performance improvements over the previous AIRS/advanced microwave sounding unit (AMSU) operational sounding retrievals in both retrieval error and also show comparable vertical resolution. We also present initial neural network retrieval results using measurements from the Cross-Track Infrared Sounder (CrIS) and Advanced Technology Microwave Sounder (ATMS) currently flying on the Suomi National Polar-orbiting Partnership satellite.

1. Introduction

Spaceborne microwave and hyperspectral infrared (IR) sounding instruments have enabled significant improvements in weather forecast accuracy. From the upwelling thermal emission and scattered radiance observed by these instruments, the three-dimensional distribution of atmospheric temperature and water vapor can be reconstructed [*Rodgers*, 2000]. The IR measurements with high spectral resolution allow for vertical resolution approaching 1 km but are strongly affected by clouds, while the lower resolution microwave measurements have far less sensitivity to clouds. Thus, IR and microwave observations play a complementary role in retrieving the state of the atmosphere in both clear and cloudy conditions.

Recently, we have presented Stochastic Cloud Clearing/Neural Network (SCC/NN) [*Blackwell*, 2012; *Blackwell and Milstein*, 2014; *Tao et al*., 2013], a statistical technique for performing temperature and water vapor retrievals from combined microwave and hyperspectral IR observations. The algorithm combines a statistical method for cloud clearing the radiance spectrum [*Cho and Staelin*, 2006], projected principal components (PPC) for efficient compression of the radiances correlated to the retrieved variables [*Blackwell*, 2005], and a feed-forward neural network [*Blackwell*, 2005; *Blackwell and Chen*, 2009] for performing the retrievals from the cloud-cleared radiance PPCs. SCC/NN was implemented using the AIRS and AMSU instruments and was trained using global training set derived from European Center for Medium-Range Weather Forecasting (ECMWF) fields. Initial validation results were shown, comparing SCC/NN retrievals to collocated ECMWF profiles as ground truth for eight selected days. The results showed significantly improved accuracy and resilience to heavy cloud cover versus Version 5 [*Susskind et al*., 2011] of the AIRS/AMSU Level 2 physical retrieval product. SCC/NN is currently in operational use by NASA, providing the first guess for Version 6 [*Susskind et al*., 2014] of the AIRS/AMSU Level 2 physical retrieval product. The Version 6 algorithm, in part due to the incorporation of the SCC/NN first guess, has shown significantly improved yield in cloud-covered scenes and accuracy over the previous Version 5, which used a linear regression first guess [*Blackwell and Milstein*, 2014; *Susskind et al*., 2014].

Here we present more extensive validation results for SCC/NN, including additional days, additional truth data sources, and additional sensors. We show additional SCC/NN results for the AIRS/AMSU sensor versus ECMWF, including the impact of using only AIRS (and not AMSU). We also show extensive year-round comparisons versus over 400,000 collocated radiosonde measurements spanning three different years, as well as results versus 169 dedicated radiosondes which are distinct from those used in forecasting or reanalysis. We review the algorithm details, emphasizing specifics implemented in the current AIRS Version 6 first guess

that differ or elaborate on previous presentations of SCC/NN. Finally, we present, for the first time, initial neural network retrievals using measurements from the Cross-Track Infrared Sounder (CrIS) and Advanced Technology Microwave Sounder (ATMS).

2. SCC/NN Algorithm

A top-level overview of the SCC/NN algorithm is illustrated in Figure 1. For each 3×3 field of regard (FOR) neighborhood of infrared spectra centered on a microwave measurement field of view (FOV), the stochastic cloud clearing (SCC) algorithm computes a single cloud-cleared infrared radiance spectrum. (SCC/NN performs retrievals on FORs, rather than FOVs.) The cloud-cleared spectrum and microwave measurements are then subsequently transformed using projected principal components which best correlate to the retrieved variable. Neural networks then operate on the transformed, cloud-cleared radiances to compute soundings of temperature and water vapor. Here we review the SCC/NN algorithm for the AIRS/AMSU sensors, emphasizing specifics or slight variations from previous presentations of the algorithm where applicable. We also introduce the new CrIS/ATMS implementation.

2.1. Stochastic Cloud Clearing

SCC [*Cho and Staelin*, 2006], shown in Figure 2, is a statistical algorithm for estimating the cloud-free infrared radiance spectrum from the 3×3 IR FOR, and, when available, the microwave measurements in same position. SCC leverages variable cloud cover and the resulting contrast among neighboring IR FOVs along with the low cloud sensitivity of the microwave data to estimate cleared IR radiances. SCC is trained using measured radiances as input, and targets derived from synthetic clear-sky radiances, calculated from collocated ECMWF fields using the Stand-Alone Radiative Transfer Algorithm (SARTA) [*Strow et al*., 2003]. (No additional bias corrections, beyond the tuning used by the AIRS retrieval algorithm, were applied to the SARTA outputs or observed radiances, as any such systematic effects would be compensated for by a regression-based algorithm.) For surface emissivity, we followed *Cho and Staelin* [2006] in using the emissivity characteristics of ocean over both water and land, as they found that this model introduced little errors due in part to the separate training of land and water regions, and algorithm's ability to compensate for such effects when they appear in both the training and testing sets. We also used a carbon dioxide profile that assumed a concentration of 370 ppm. (No compensation for CO_2 variations by year was applied.)

SCC estimates the cloud-cleared spectrum by performing a series of simple linear and nonlinear operations on the inputs. The nine IR FOVs are sorted by cloudiness, and effective "warm" and "cold" IR spectra are selected and computed and described later in this section. The SCC inputs include seven noise-adjusted principal components (NAPCs) [*Lee et al*., 1990] of the IR spectrum from the warm IR spectrum, three NAPCs for the difference between the warm and cold spectra (based on IR channels that sound low altitudes), a subset of the microwave brightness temperatures that sound tropospheric and lower stratospheric temperatures, the secant of the satellite scan angle θ, and the land fraction. First, a preliminary radiance correction to the warm spectrum is estimated using a regression linear operator **A**. (Note that this correction is an intermediate quantity used to determine subsequent processing steps but is not part of the final correction applied to the radiances.) The scalar output A_0 is first principal component (PC) of the cloud correction spectrum. A_0 is multiplied by sec(θ) and, combined with all the inputs to **A**, is used as inputs to a second linear operator **B**, which outputs four PCs of a cloud correction to the warm spectrum. The PCs are used to compute radiance correction for two channels (4.5117 μm and 4.4813 μm) with weighting function peak heights near 0.47 km and 2.7 km, respectively, which are low altitudes that are useful for determining cloudiness. A threshold test on both channels is used to classify the clearest averaged IR FOV into three categories, "30% least cloudy," "80% least cloudy," and "most cloudy," with thresholds derived from the training set to meet those percentiles. For the 30% least cloudy and 80% least cloudy categories, two additional linear operators, **C** and **D**, respectively, are used with the same inputs used for **B** to compute four PCs of a final cloud correction spectrum. The final correction spectrum is computed using the corresponding four PC vectors. For the most cloudy category, the final cloud correction is the spectrum computed from **B**. The classification step, with different cloud correction applied based on cloudiness, along with the use of A_0 sec(θ) as inputs to subsequent operations, provides simple ways of accommodating nonlinearity.

The warm and cold spectra used in the SCC algorithm inputs are selected and computed as follows. The nine IR FOVs are sorted by cloudiness using the average radiance of channels between 4.16 μm and 5 μm with

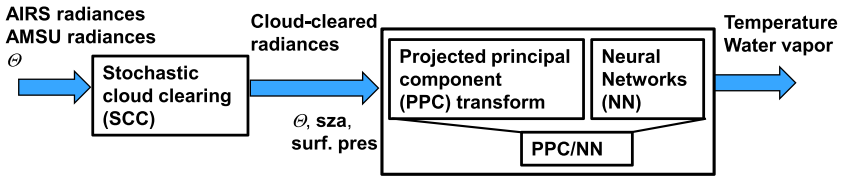

Figure 1. Stochastic Cloud Clearing / Neural Network (SCC/NN).

weighting function peaks at heights between 1 and 3 km, with warmer FOVs assumed to be clearer. The cold spectrum is selected from the cloudiest FOV based on the sorting. Three different average warm spectra are created by selecting the clearest FOV spectrum, the average of the four clearest FOV spectra, and the average of all nine FOV spectra, respectively. For each of these three warm spectra, we compute a cloud correction as described above, with different versions of the operators (**A**–**D**) trained and used for each. Afterward, the final averaged warm spectrum and cloud correction are constructed using the results for the clearest FOV for channels with weighting function peak height between 0 and 5 km, the results for the average of the four clearest FOVs for channels with weighting function peak height between 5 and 10 km, and the results for the average of all nine FOVs for all other channels.

The current SCC algorithm as described above is very similar to the description in *Cho and Staelin* [2006], with a few notable exceptions resulting from subsequent refinements and empirical tradeoffs, found to give slightly improved results: the use of three cloudiness categories with percentile thresholds (30% least cloudy, 80% least cloudy, and most cloudy) rather than two categories with fixed temperature thresholds, calculating the final averaged warm spectrum based on weighting function peak height after the cloud clearing steps rather than before, and the channel pair used to determine cloudiness (4 μm channels rather than 11 and 15 μm channels).

2.2. Projected Principal Components/Neural Network

Neural networks (NNs) are nonlinear function approximators trained to infer a statistical relationship between inputs and outputs from a training ensemble, without requiring direct knowledge of the functional input/output relationship [*Blackwell*, 2005; *Blackwell and Chen*, 2009]. NNs consist of simple computational elements called nodes, which are interconnected. Figure 3 shows the multilayer feed-forward NN structure used in SCC/NN, with an input layer, output layer, and one or two hidden layers. Each hidden layer contains nodes which apply a sigmoidal activation function of the form

$$z_j = \tanh(a_j) \tag{1}$$

where

$$a_j = \sum_{i-1}^{d} w_{ji} x_i + b_j \tag{2}$$

and x_i is the ith input, z_j is the output of the jth node, and w_{ji} and b_j are the weights and biases applied to the inputs. The output layer is linear. The weights and biases are the tuning parameters which are optimized during the training process to minimize a sum-squared error cost function between the inputs and the training targets. The NNs used in SCC/NN were trained using the Levenberg-Marquardt backpropagation algorithm, with early stopping based on network performance on a validation data set [*Blackwell*, 2005].

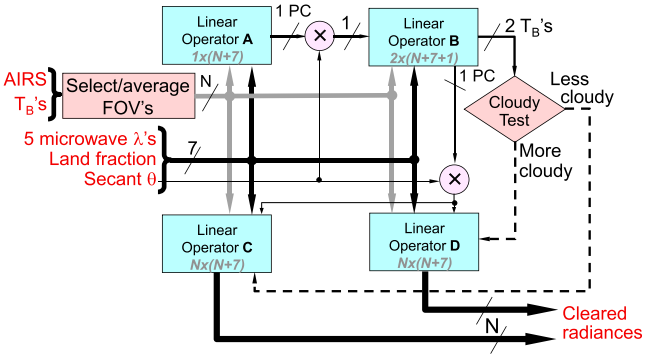

Figure 2. Stochastic Cloud Clearing algorithm. λ is microwave wavelength.

In the SCC/NN algorithm, the NN inputs are compressed microwave and cloud-cleared radiances, along with the secant of scan angle, the secant of solar zenith angle, and the forecast surface pressure

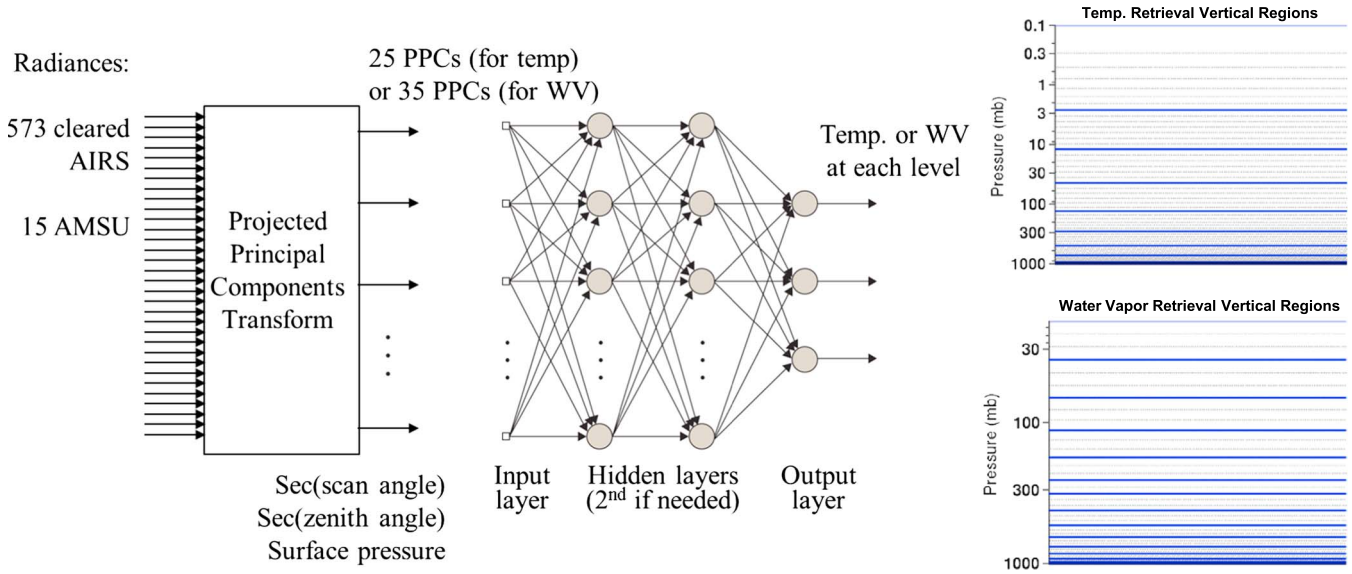

Figure 3. PPC-NN: Combination of Compression and Neural Network.

normalized by 1013 Mb. The NN outputs are the retrieved variables, temperature, and water vapor mixing ratio. The radiances are compressed using the projected principal component (PPC) transform, which has been shown to extract radiance information in each coefficient that is maximally correlated with the retrieved variable [*Blackwell*, 2005]. The training targets are ECMWF reanalysis temperature, skin temperature, and water vapor mixing ratio fields at 60 sigma pressure levels, as described below.

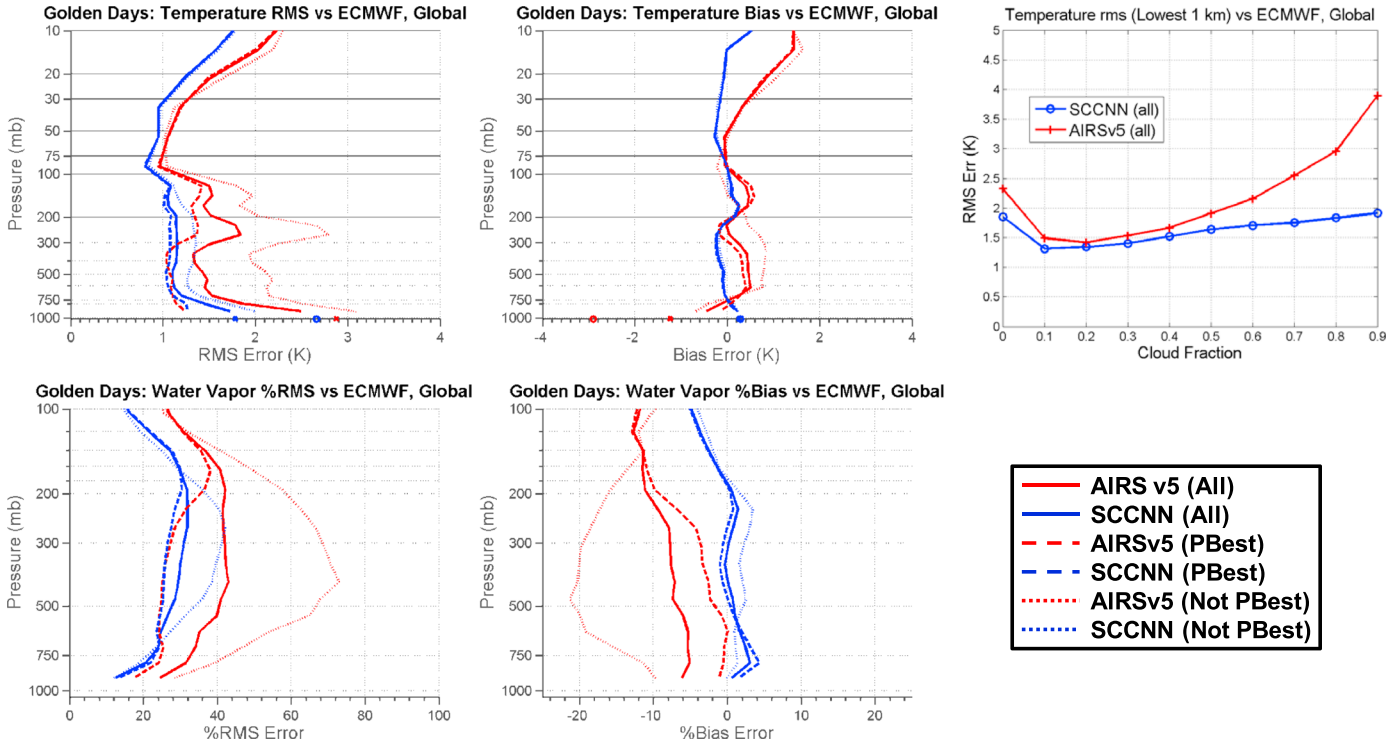

Figure 4. AIRS/AMSU SCC/NN validation result versus ECMWF, with global coverage over eight golden days: Temperature retrieval RMS error and bias, temperature RMS error versus cloud fraction for lowest 1 km, water vapor mixing ratio retrieval %RMS error and bias (as percent of water vapor RMS).

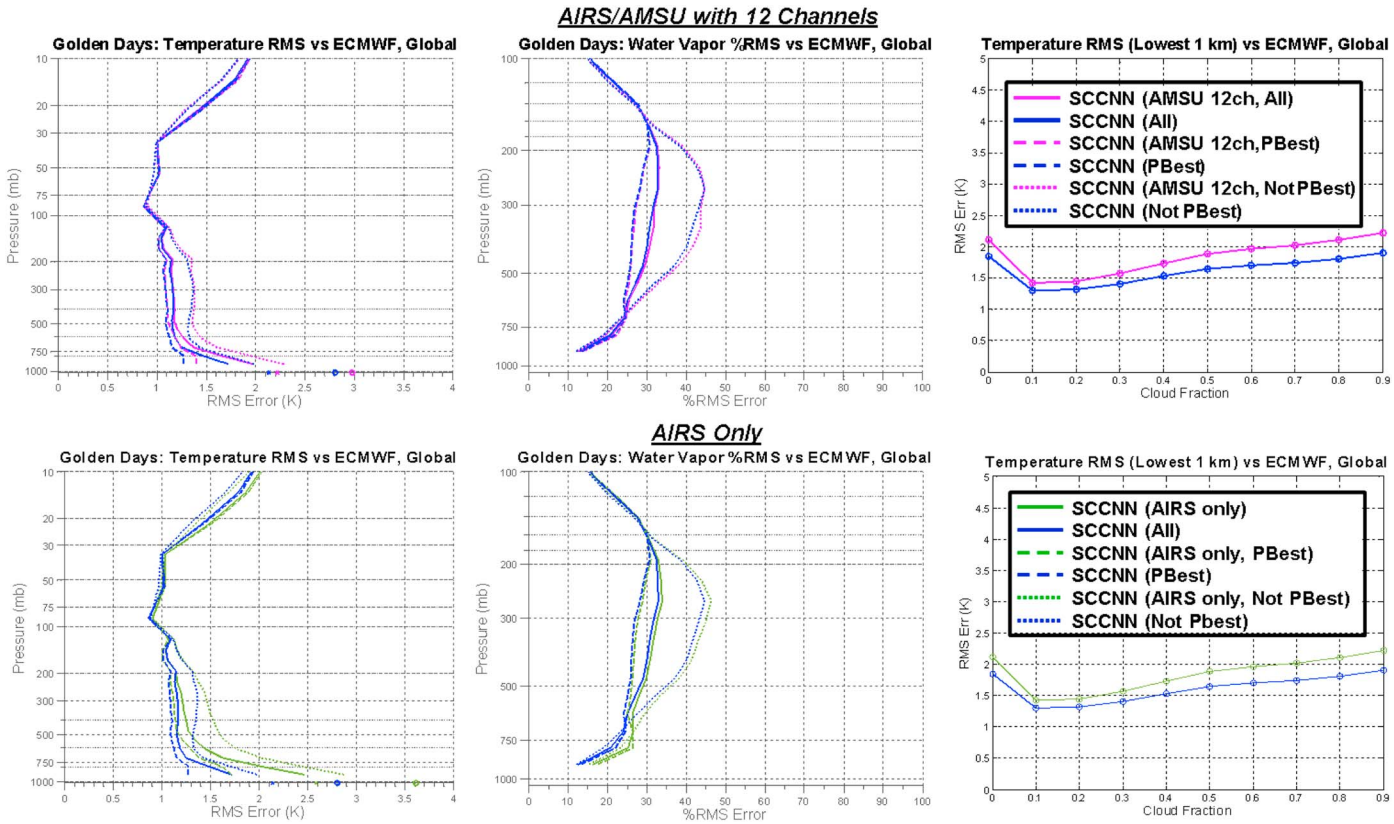

Figure 5. AIRS/AMSU golden days, effect of excluding AMSU channels 4, 5, 7, and effect of AIRS-only. Temperature retrieval RMS error, temperature RMS error versus cloud fraction for lowest 1 km, water vapor mixing ratio retrieval %RMS error.

We optimized the architecture of each NN based on empirical tradeoffs. For the temperature retrievals, we used 25 PPCs for the IR/microwave radiances. The 60-pressure-level profile is subdivided into nine vertical regions, shown in Figure 3, with a NN dedicated to each region. Each NN output corresponds to retrieved temperature at a pressure level. The NN contains a single hidden layer with 20 hidden nodes. The same NN architecture is used to retrieve surface temperature as well. For water vapor, we used more radiance inputs and hidden nodes due to the greater nonlinearity of the retrieval. We used 35 radiance PPCs and two hidden layers, with 25 and 15 hidden nodes. We subdivide each 60-pressure-level profile into 15 vertical regions, shown in Figure 3. The 60-level retrievals computed by the NNs are adapted to the 100-level AIRS support product levels by linearly interpolating, on a logarithm pressure scale, from the 60 sigma pressure level, facilitating their use by the AIRS Version 6 physical retrieval algorithm as a first guess. While the vertical regions used by the NN are not overlapping, significant discontinuities between vertical regions were not generally noted.

2.3. Training Set and Stratification

For both the SCC and NN algorithms, the training set was drawn from ECMWF temperature, skin temperature, and water vapor mixing ratio fields from every fourth day between 1 December 2004 and 31 January 2006. To assist in accommodating nonlinearity in the cloud clearing and NN retrievals, we stratify the radiance data and training targets using five variables: orbit node type (ascending and descending), latitude region (polar region north of 60°N, polar region south of 60°S, and remaining temperate region), season, and surface type. The surface-type categories differ by latitude region, with category types determined in development based on empirical performance tradeoffs. For the temperate region the types are ocean and land, with land binned into eight categories according to surface pressure. These stratifications and categories are similar to those used by *Cho and Staelin* [2006] and were introduced in part to provide additional nonlinear capability for the SCC algorithm.

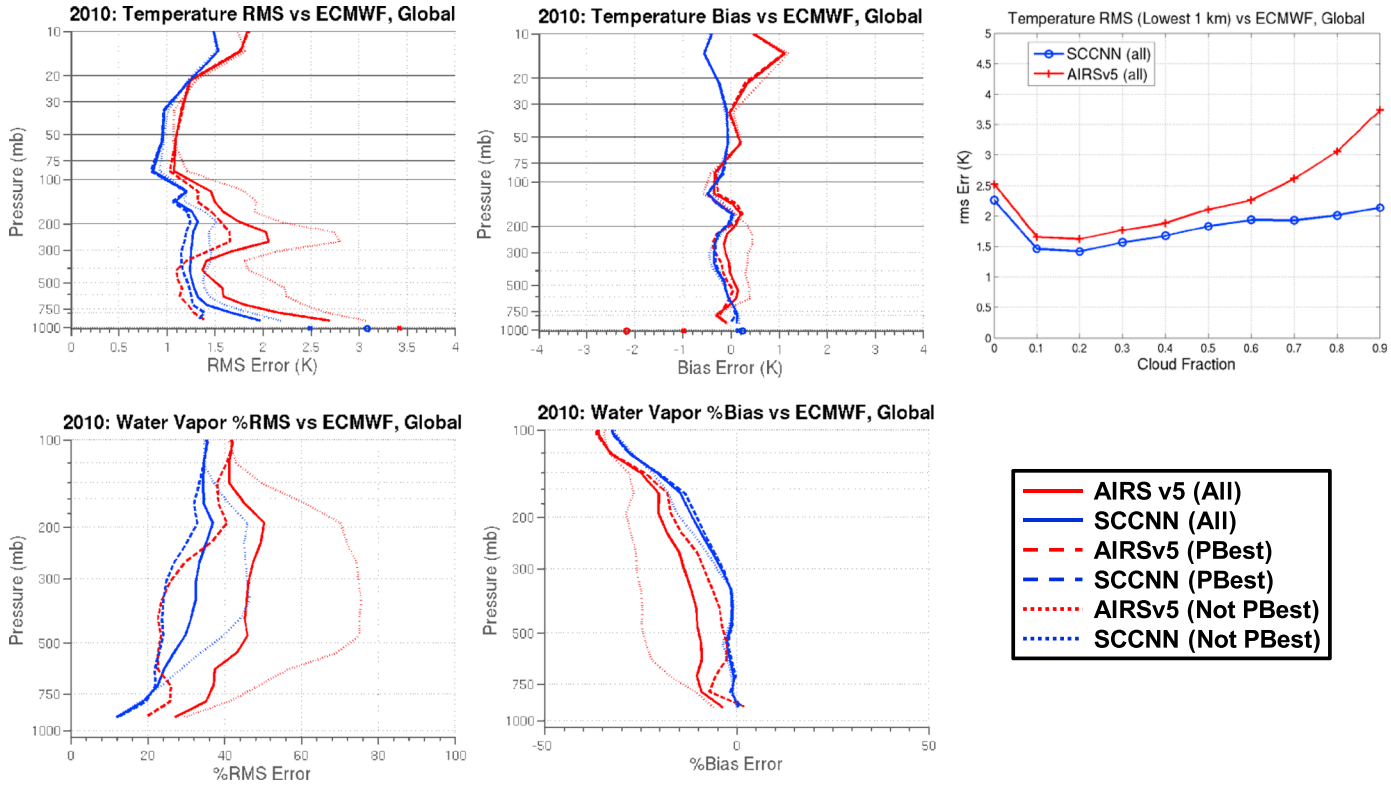

Figure 6. AIRS/AMSU 2010: 11 selected days. Temperature retrieval RMS error and bias, temperature RMS error versus cloud fraction for lowest 1 km, water vapor mixing ratio retrieval %RMS error and bias (as percent of water vapor RMS).

For the North Pole region, the types are ocean, frozen ocean, and five categories of land binned by surface pressure. The South Pole region is similar but with seven land categories. These five variables lead to a total of 200 stratifications. A different set of SCC and NN coefficients are trained and used in each stratification, where each training set contains approximately 30,000 training profiles are used, along with 5000 test profiles used for early stopping of the NN training algorithm.

To avoid abrupt discontinuities over stratification boundaries, we use linear interpolation to gradually phase in transitions over the neighboring stratifications. For season transitions, we stratify using 5 month overlapping seasons and linearly interpolate retrievals over season in the overlap interval. We also interpolate over latitude region with a 10° transition interval and interpolate between land and ocean retrievals by using the land fraction as the interpolation weight. The global plots of SCC/NN results shown by Blackwell and Milstein [*Blackwell and Milstein*, 2014] demonstrate the lack of abrupt regional discontinuities.

In the polar regions, sea ice detection is needed to determine whether the "frozen ocean" surface-type stratification should be used. Following the approach of *Ferraro et al.* [2005], we use AMSU measurements to compute sea ice concentration and assume ice if this product exceeds a threshold (empirically chosen at 120%).

2.4. AIRS/AMSU Channels and Bad Channel Filling

For AIRS, we selected 573 of the total 2378 channels to use in the SCC and NN algorithms, based on a combination of the 314 channels previously used by *Cho and Staelin* [2006] and additional channels historically used in AIRS retrievals [*Susskind et al.*, 2003]. For AMSU, we use all channels as NN inputs, and channels 5, 6, and 8–10 as SCC inputs, prior to 2007. However, for 2007 onward, some AMSU channels have degraded [*Fetzer and Manning*, 2012], showing increased noise. Hence, our post-2007 retrieval approach uses all channels apart from 4, 5, and 7 as NN inputs and use

Table 1. Number of Collocated RAOB Data Sets

Year	Day	Night	Total
2003	64,568	104,745	168,196
2010	32,388	59,744	91,836

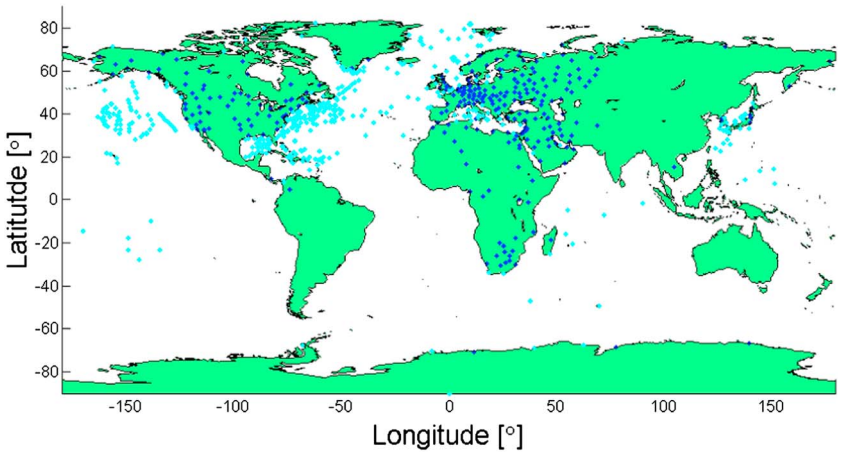

Figure 7. AIRS/RAOB collocation sites from 2003.

channels 2, 3, 6, and 8–10 as SCC inputs. We train a different set of coefficients corresponding to each time interval where different inputs are used. We note that for the AIRS v6 SCC/NN implementation used as the first guess to generate the L2 retrieval product, the post-2007 approach is used for all years, for consistency in the data record. An AIRS-only version of SCC/NN was also implemented to allow the AIRS v6 algorithm to degrade gracefully in the event of AMSU data being unavailable. This implementation uses the same algorithmic approach described above, but with AMSU data omitted as inputs. Sea ice is detected via a threshold test on forecast surface temperature in place of AMSU data, with the threshold, 271.3 K, chosen based on prior AIRS team experience [*Hearty and Olsen*, 2007].

Occasionally, an AIRS channel will be flagged by AIRS Level1b quality control flags at the granule or scanline level as having problems, such as "popping," or estimated noise that significantly exceeds the expected

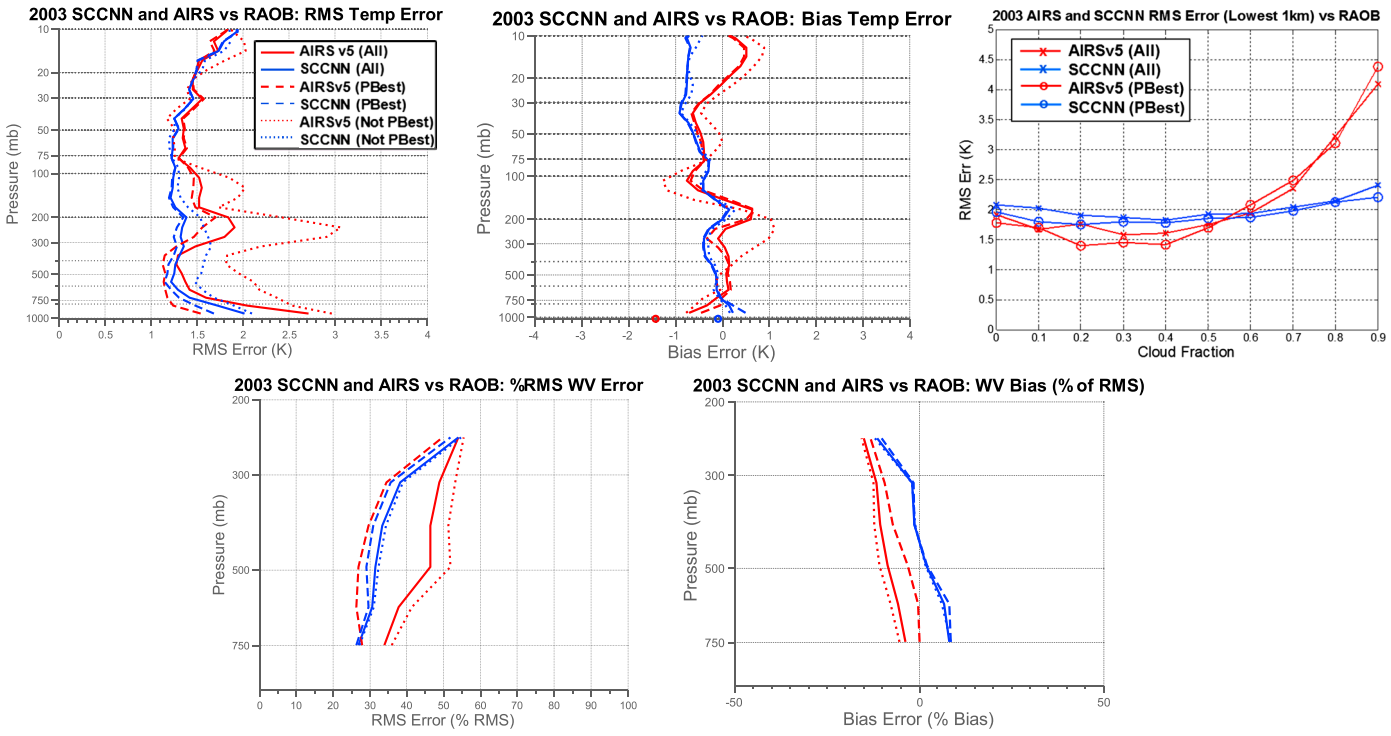

Figure 8. AIRS/AMSU SCC/NN validation result versus radiosondes, with global coverage spanning all of 2003: Temperature RMS error and water vapor mixing ratio %RMS error and bias (as percent of water vapor RMS).

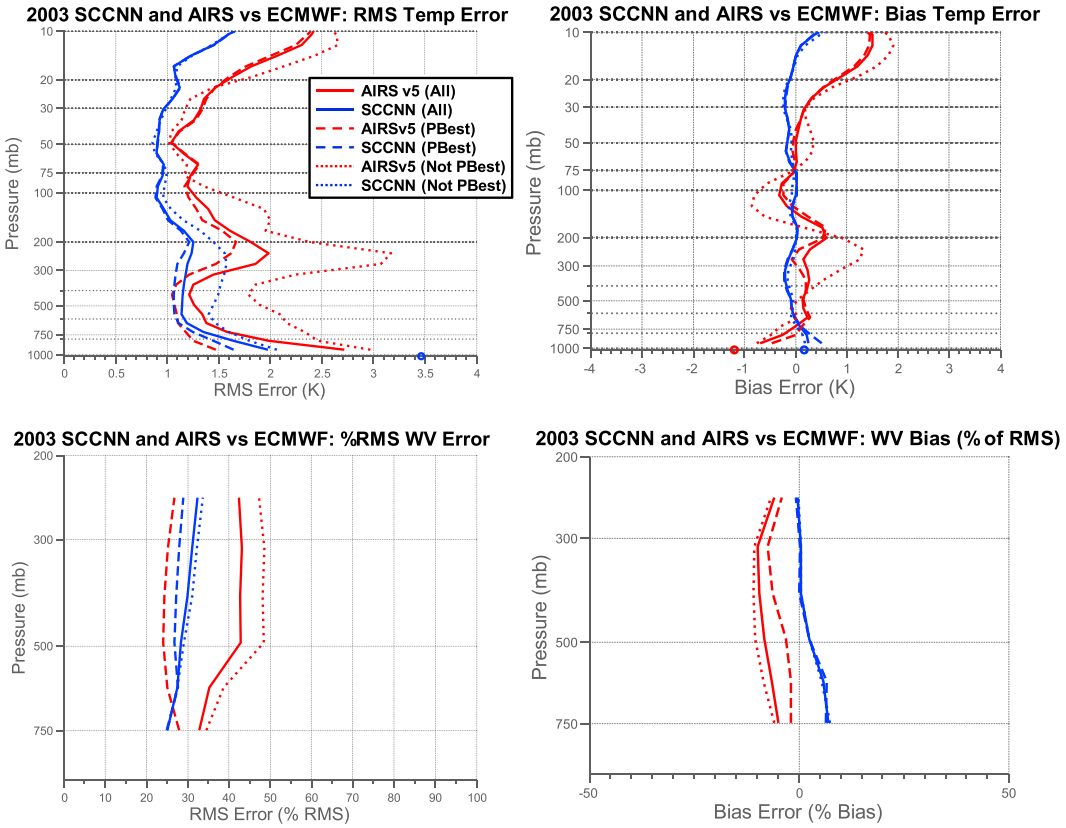

Figure 9. AIRS/AMSU SCC/NN validation result versus coincident ECMWF, with global coverage spanning all of 2003: Temperature RMS error and water vapor mixing ratio %RMS error and bias (as percent of water vapor RMS).

levels. When a channel measurement or set of channel measurements is flagged, we fill the "bad" channels with a predicted value obtained using linear regression on the remaining "good" AIRS channels, using previously computed covariance matrices for the radiances calculated from the training set inputs. Specifically, let G be the good portion of the radiance spectrum with covariance C_{GG}, B be the bad portion of the radiance spectrum, with the cross-covariance C_{BG}. Then the estimated bad radiances $\hat{B}$ are given by

$$\hat{B} = \mathbf{C}_{BG}(\mathbf{C}_{GG} + \mathbf{C}_{NN})^{-1} G \tag{3}$$

where C_{NN} is the instrument noise covariance, assumed to be diagonal with values derived from the instrument noise equivalent differential temperature. While (3) includes a matrix inverse that must be computed every time a different set of channels are flagged, the computation can be reduced by precomputing the expected covariance and covariances of the whole radiance spectrum from the training set and exploiting known block matrix inversion identities [*Bernstein*, 2009].

2.5. Cris/ATMS Version

We have recently adapted SCC/NN to the CrIS/ATMS sensors on the Suomi National Polar-orbiting Partnership (Suomi NPP) spacecraft. This version of SCC/NN employs the same basic methodology used for AIRS/AMSU. The training inputs consist of CrIS Sensor Data Record (SDR) and ATMS resampled SDR products drawn from every fourth day between 1 November 2012 and 31 October 2013. The training targets are collocated ECMWF profiles. Because of ECMWF improvements since early 2006, the ECMWF fields have 91 levels. The NNs have the same architecture as the AIRS/AMSU version, apart from having more outputs to accommodate the additional ECMWF levels. The CrIS radiances are apodized using a Hamming window. We currently use 340 out of 1305 CrIS instrument channels, the closest available channels to the 573 channels used in the AIRS version. All 22 ATMS channels are used in the NN, and channels 6, 7, and 9–11 are used in SCC.

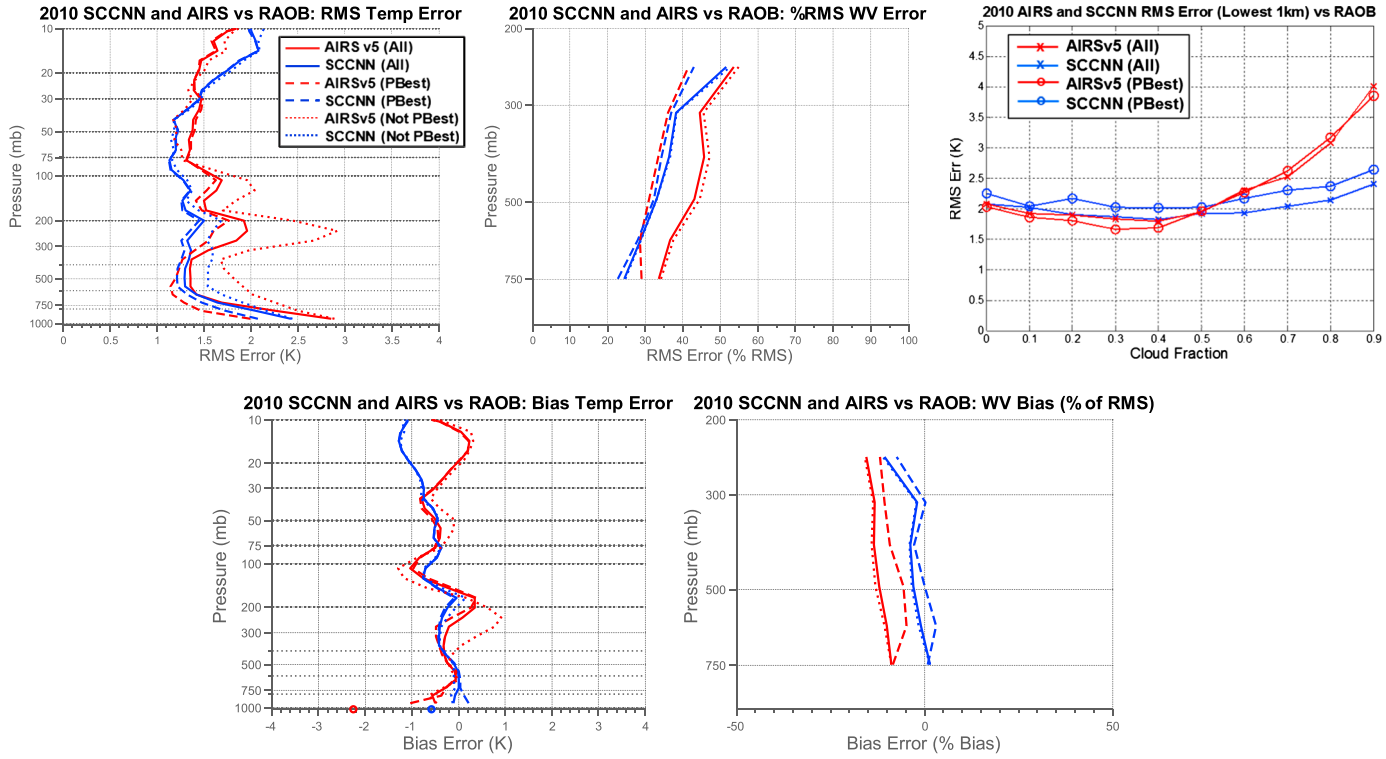

Figure 10. AIRS/AMSU SCC/NN validation result versus radiosondes, with global coverage spanning all of 2003: Temperature RMS error and water vapor mixing ratio %RMS error and bias (as percent of water vapor RMS).

3. Validation Results for AIRS/AMSU

In order to assess the performance of SCC/NN throughout the life of the Aqua mission, extensive and ongoing validation is needed. Here we present results using a variety of independent data sources, including ECMWF reanalysis fields, synoptic radiosonde measurements, and initial results for dedicated radiosondes.

3.1. AIRS/AMSU Golden Day Results

Figure 4 shows example SCC/NN validation performance obtained to date versus ECMWF reanalysis fields, for eight "golden" days (9/6/02, 25 January 2003, 8 June 2003, 21 August 2003, 3 September 2003, 12 October 2003, 5 December 2003, and 29 September 2004), along with AIRS Version 5 results for comparison. (We show comparisons to Version 5 rather than Version 6 because Version 5 retrievals are completely independent of the NN. Comparisons of Version 6 results to the NN first guess are the subject of separate work.) The data set includes global coverage with a total of 2,592,000 profiles. Figure 4 shows plots for temperature profile root-mean-square (RMS) error and bias with 1 km averaged layers, and water vapor %RMS error with 1 km averaged layers, and temperature RMS error versus cloud fraction for the lowest 1 km layer. The %RMS error used for water vapor is equivalent to computing

$$\%\ \text{RMS error} = \frac{\sqrt{\sum_{i=1}^{N}\left(q_{i,\text{retrieval}} - q_{i,\text{truth}}\right)^2}}{\sqrt{\sum_{i=1}^{N}\left(q_{i,\text{truth}}\right)^2}} \times 100\% \tag{4}$$

where q_i is the water vapor mixing ratio (in kg/kg) of retrieval i at the layer being assessed. Water vapor bias results are also shown, also normalized as a percentage of water vapor RMS. For every retrieval, the AIRS algorithm includes a quality indicator called "PBest," which is the maximum pressure for which the profile is determined to have high quality. In Figure 4, we show, at each level, a curve with results for all profiles, a curve for profiles which pass the PBest indicator, as well as a separate curve, "Not PBest," showing results for the profiles which did not pass the PBest indicator. The plots demonstrate significant improvement of

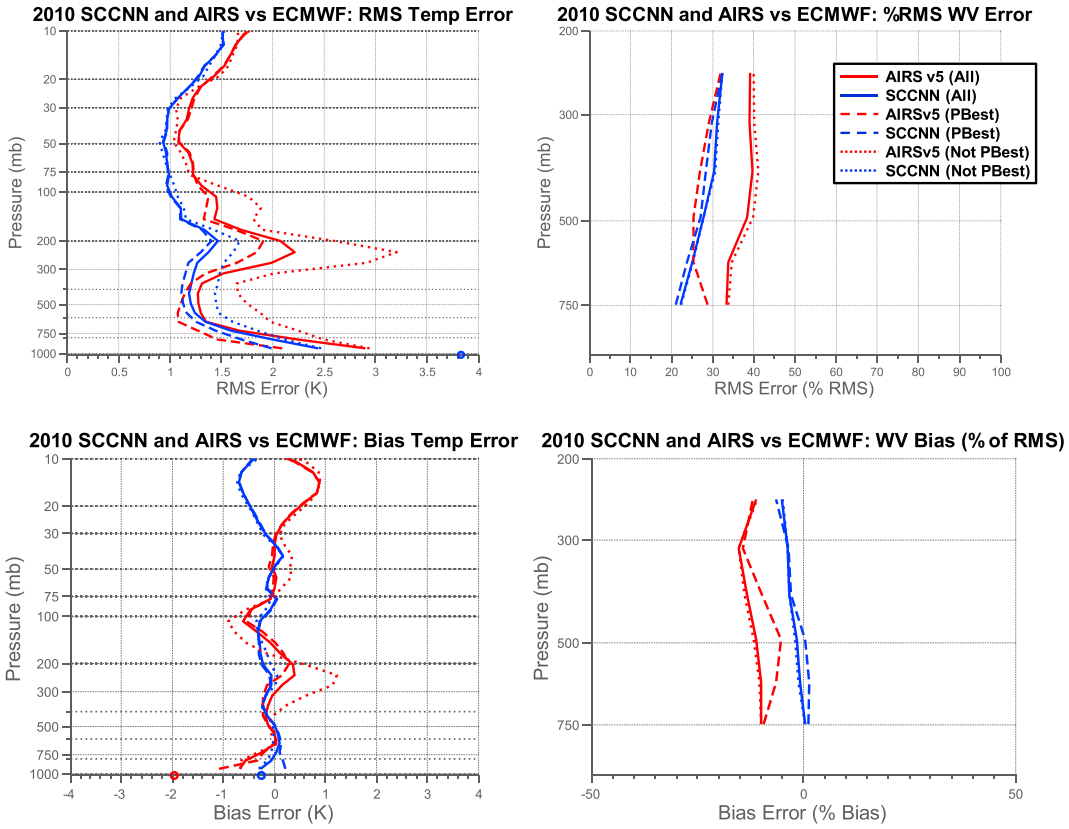

Figure 11. AIRS/AMSU SCC/NN validation result versus coincident ECMWF, with global coverage spanning all of 2010: Temperature RMS error and water vapor mixing ratio %RMS error and bias (as percent of water vapor RMS).

SCC/NN over the AIRS v5 retrievals under cloudy conditions, including profiles not designated PBest, the most stringent screening for quality control, with the AIRS v5 PBest retrievals occasionally slightly better than the SCC/NN PBest retrievals. The systematic bias in the SCC/NN retrievals is generally lower than with the AIRS v5 retrievals, apart from lower altitudes for PBest-accepted water vapor retrievals. The increase in RMS temperature error for both SCC/NN and AIRS v5 for the 0 cloud fraction bin versus the 0.1 cloud fraction bin is notable. The cloud fraction is from the AIRS v5 level 2 product. We hypothesize that the cloud fraction product may be assigning 0 to some uniformly cloud covered scenes, though further study is needed to confirm this.

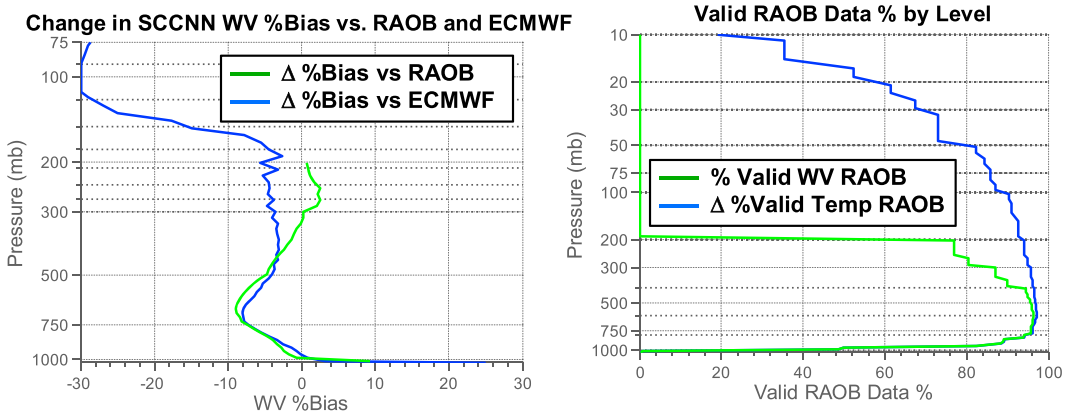

Figure 12. Change in water vapor bias (as a percent of RMS) between 2003 and 2010, for SCC/NN versus RAOB and ECMWF, and percentage of valid RAOB data for temperature and water vapor as a function of pressure level.

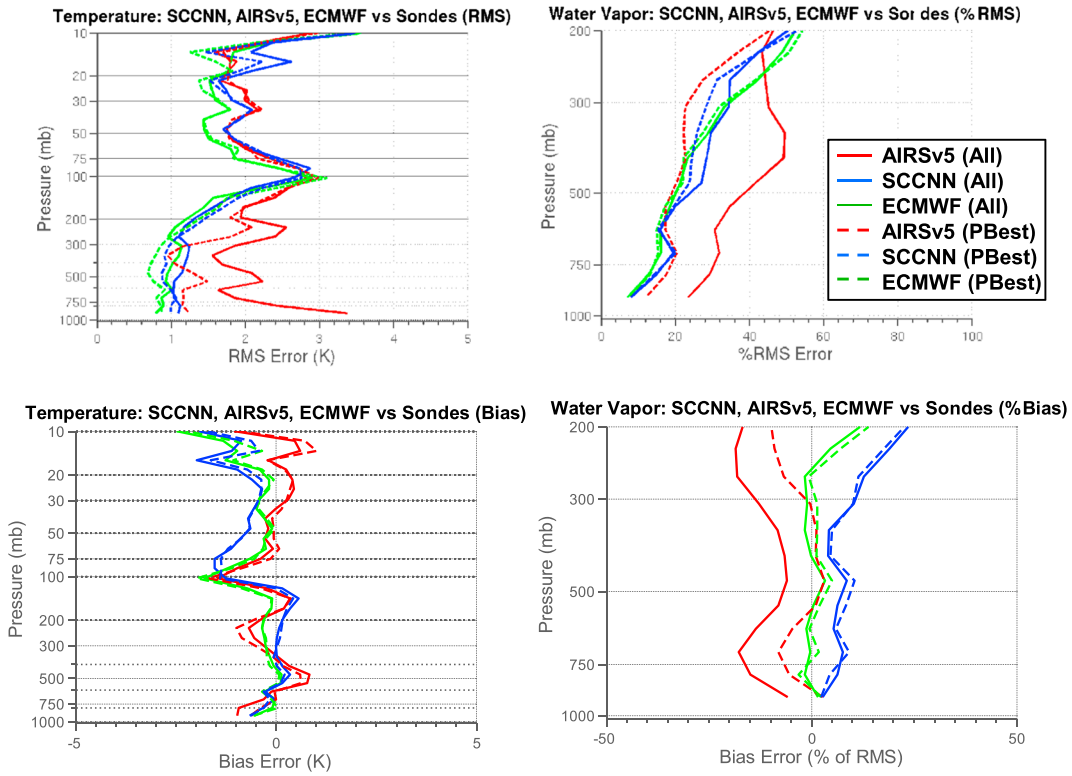

Figure 13. Dedicated radiosonde collocation sites.

3.2. Degraded AMSU and AIRS-Only Results

As described in section 2.4, AMSU channels 4, 5, and 7 are not used after 2007 due to noise, and different SCC and NN coefficients are used after that year, with the post-2007 approach being employed for all years in the v6 first guess. To assess the effect of using fewer AMSU channels, the "golden day" assessment was repeated using the post-2007 coefficients. Figure 5 shows the results, as compared with the results of using the pre-2007 coefficients. The results show generally small changes for water vapor, but an increase in temperature RMS error versus ECMWF near the lowest 1 km layer ranging from approximately 0.2 K for clear profiles to 0.3 K for cloudy profiles.

Figure 5 also shows the results using only AIRS, without AMSU data, compared with the results of using AMSU with the pre-2007 coefficients. For water vapor, a slight increase of approximately 5% RMS error is apparent near 800 mbar. For temperature, the temperature RMS error near the lowest 1 km layer shows an increase ranging from approximately 0.5 K for clear profiles to 0.8 K for cloudy profiles. Figure 6 shows global results for 11 days in 2010 the first of every month apart from December, covering a year-round time interval with more recent data than the "golden days" set. The RMS results are similar to the results of Figure 5 using the post-2007 coefficients but show slight error increases (approximately 0.1 K RMS near the surface). However, for both AIRS v5 and SCC/NN, the water vapor bias relative to ECMWF above 200 mbar increased in magnitude in the dry direction, relative to that seen in the golden days increases to about 10% of bias. Some changes between the golden days and the 2010 sets may be a result of additional instrument degradation over time, or changes in the atmosphere since the 2005-centered training set. However, for water vapor bias, with AIRS v5 and SCC/NN both showing the same bias trend relative to ECMWF above 200 mbar, a change in the ECMWF model over time might also explain such results.

3.3. Radiosondes

For additional validation, we assessed SCC/NN performance versus radiosondes, as *Divakarla et al.* [2006] have done for previous AIRS algorithm versions. Extensive synoptic radiosonde observation (RAOB) reports for 2003 and 2010 were obtained via NOAA [*Reale et al.*, 2012]. We extracted interpolated temperature and

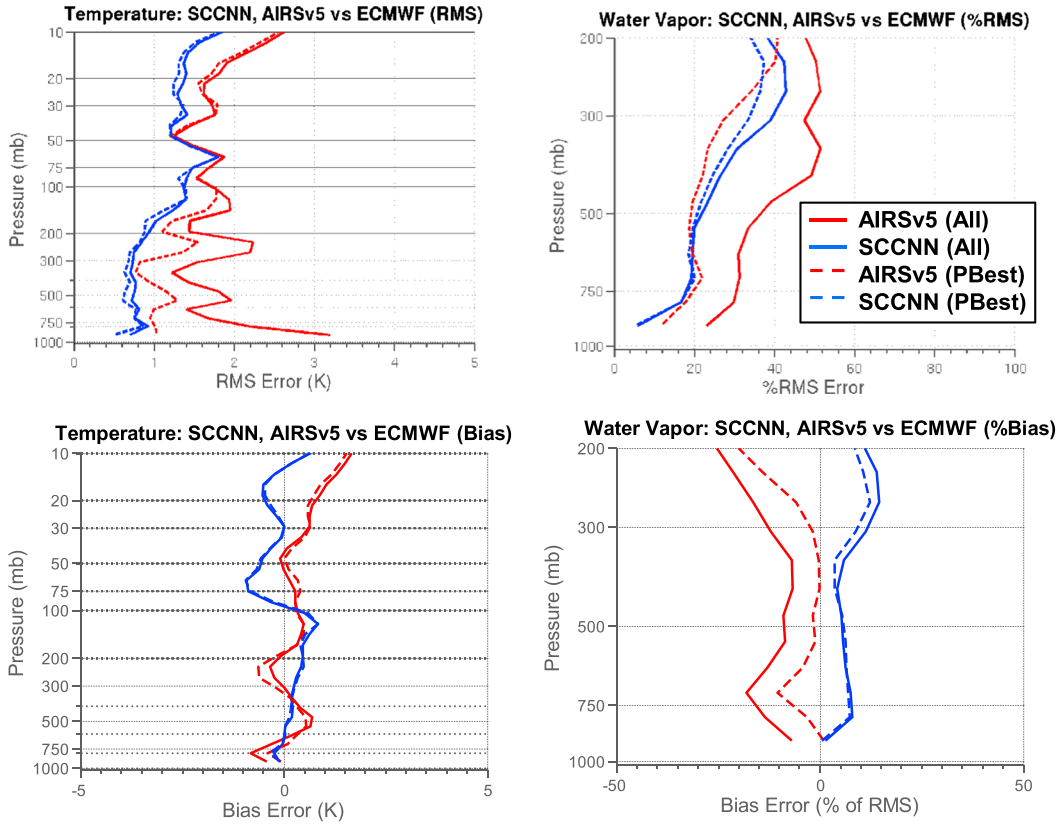

Figure 14. Comparison of SCC/NN, AIRS v5, and ECMWF profiles versus 169 dedicated sonde measurements from 2003 to 2004.

water vapor profiles from the reports and collocated them to year-round SCC/NN retrievals and ECMWF fields. The RAOBs were collocated to SCC/NN within ±3 h time and 100 km. Table 1 summarizes the number of collocations, totaling 260,032 for both years. (We also obtained and processed data sets for 2004, but with the results being very similar to 2003, we omit them for brevity.) The collocated radiosonde sites were mostly land based, while 4% of them were sea based. Figure 7 shows maps of the collocation positions for 2003, which are representative of both years. The sea-based measurements are widely distributed due to the fact

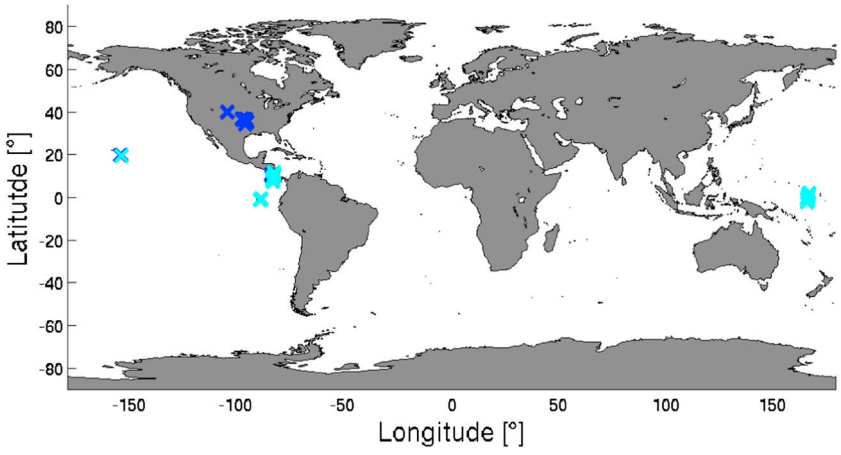

Figure 15. Comparison of SCC/NN, AIRS v5 versus ECMWF profiles, from collocations with versus 169 dedicated sonde measurements.

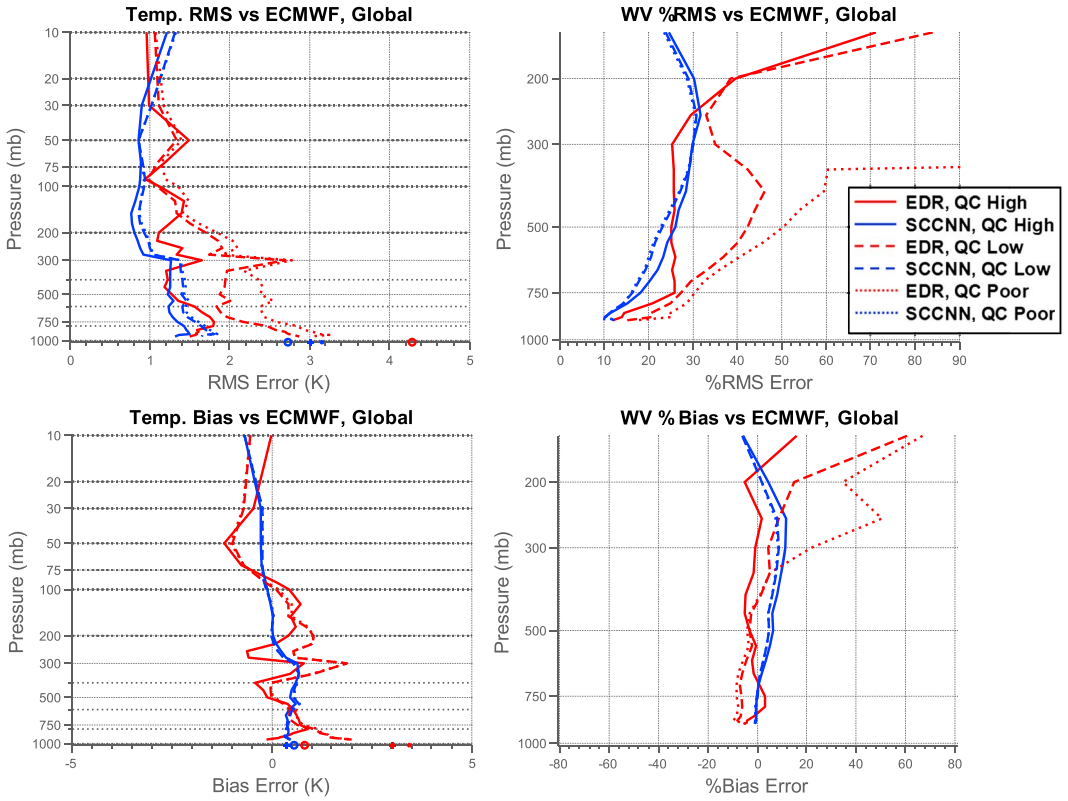

Figure 16. SCC/NN for CrIS/ATMS: Initial temperature RMS results versus ECMWF.

that many were made via ship rather than a stationary site. The geographic distribution of the collated measurements depends on how the AIRS/AMSU overpasses coincide with the 12 h synoptic RAOB measurement schedule.

Figure 8 shows SCC/NN validation performance versus coincident radiosondes for 2003, while Figure 9 shows results versus collocated ECMWF data sets. Figures 10 and 11 show similar results for 2010. Temperature profile RMS errors are shown with 1 km averaged layers and water vapor % RMS error with approximately 1.6 km averaged layers, chosen to be comparable to layer averaging used by *Divakarla et al.* [2006]. The SCC/NN retrievals show improved RMS error relative to AIRS v5 in both the ECMWF and RAOB comparisons. The SCC/NN RMS errors are generally slightly greater in the RAOB comparisons than in the ECMWF comparisons but show consistent behavior. While direct comparison with different data sets is not available, the RMS errors also appear comparable to or lesser than earlier published RAOB validation studies of accepted AIRS retrievals by *Divakarla et al.* [2006]. Figures 8 and 10 show the SCC/NN and AIRS v5 temperature RMS error versus RAOB as a function of cloud fraction for the lowest 1 km layer, for both 2003 and 2010. As in section 3.1, the SCC/NN errors degrade more slowly with heavy cloud cover than the AIRS v5 retrievals. The 2010 SCC/NN temperature RMS errors are approximately 0.2 K greater than the 2003 results, consistent with the 2010 results in section 3.2. Approximately 0.1 K of that difference appears attributable to the use of the post-2007 coefficients with reduced AMSU channels. Slight systematic differences between the layer-averaged temperature retrieval errors versus ECMWF are apparent between 2003 and 2010, with the 2010 retrievals appearing to be warmer above 50 mbar. However, the bias versus RAOBs, while greater than the bias versus ECMWF, appears to change by smaller amount. This suggests that the change in bias versus ECMWF may be at least partly attributable to changes in the ECMWF model over time. Clear trends in the water vapor bias over time are hard to discern in the collocated, layer-averaged results, which are also cut off above 200 mbar due to the vertical extent of RAOB data. Figure 12 shows the change in the level-by-level water vapor bias between 2003 and 2010, for SCC/NN versus RAOB, and for SCC/NN versus ECMWF. Below 500 mbar, a similar systematic change of up to 8% of water vapor RMS is shown versus both ECMWF and RAOB. However, between 200 mbar and 500 mbar, Figure 12 shows more change in the bias

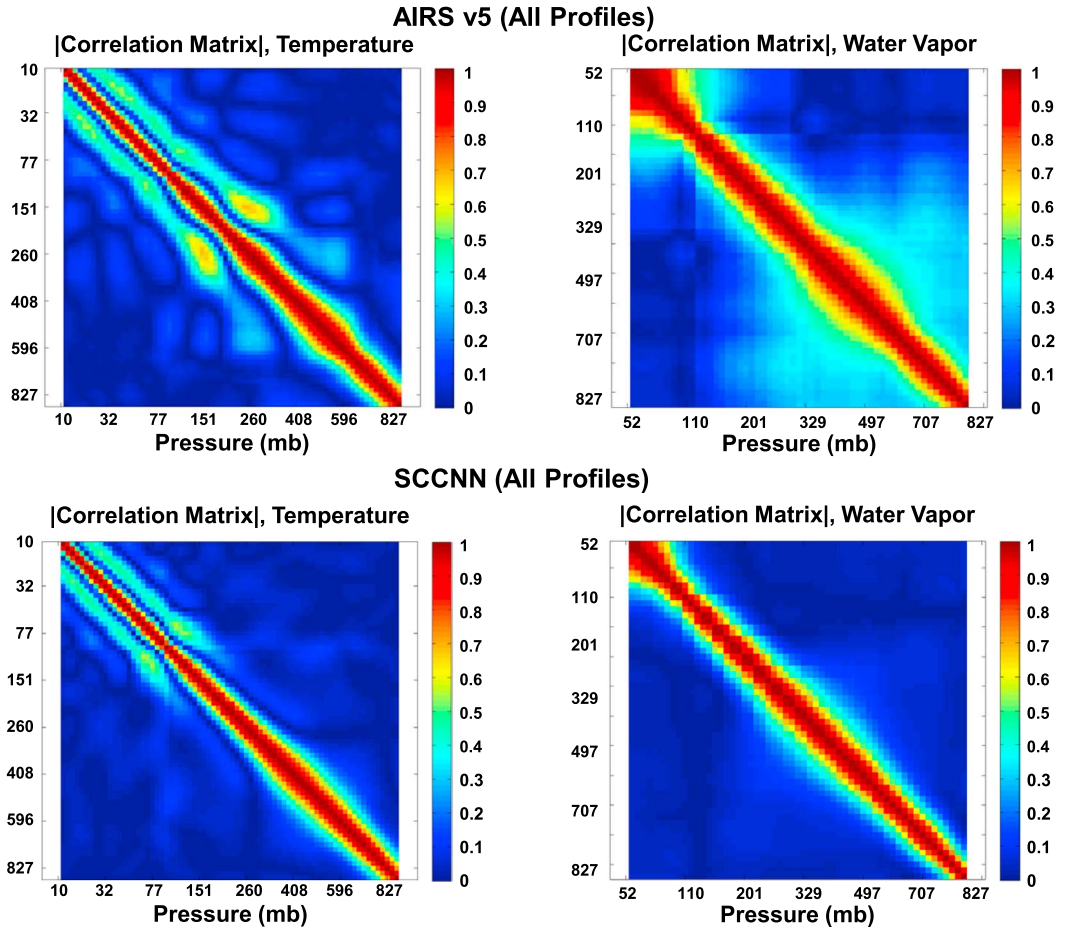

Figure 17. AIRS SCC/NN retrieval error correlations: Golden days.

(up to 5% of RMS) versus ECMWF than versus RAOB, suggesting that the ECMWF model might have changed slightly relative to the RAOB, in addition to any changes that might be due to instrument degradation. Above 200 mbar, only ECMWF data are available, making a check of the model changes versus the RAOB changes infeasible. The ECMWF data show a larger change (about 30% of RMS) similar to that seen in the global 2010 results in Figure 6. While all available ECMWF data were used at every level, rather than just those which were collocated with RAOB data, the ECMWF comparison plot below 200 mbar changes little when only RAOB-collocated ECMWF data are included. Figure 12 also shows the percentage of valid RAOB data for temperature and water vapor as a function of pressure level.

Temperature RMS errors above 200 hPa for SCC/NN and AIRS v5 retrievals are larger in validation with conventional RAOBS than with ECMWF. It is possible that ECMWF is more accurate than RAOBS in the upper troposphere and lower stratosphere where conventional RAOBS suffer from radiation-induced warm bias [*Sun et al.*, 2013]. Because of the overall dry bias in conventional RAOB data [*Sun et al.*, 2010], the water vapor RMS error in comparison with RAOBs are greater than in with ECMWF data. (Such radiosonde radiation temperature bias and humidity bias are corrected during their assimilation in ECMWF.)

Additional independent validation was performed by studying SCC/NN performance versus dedicated radiosonde, supplied by Jet Propulsion Laboratory, which are used only for validation studies and not assimilated into the ECMWF reanalysis [*Fetzer et al.*, 2003]. Figure 13 shows the locations of the small number (169) of dedicated radiosonde data sets from 2003 to 2004 used in this study. The same collocation criteria were applied for the dedicated sondes as with the synoptic ones, though the geographic diversity and size of the data set were much more limited. Figure 14 shows results, including RMS and bias errors versus the sondes for temperature and water vapor mixing ratio for SCC/NN, AIRS v5, and the closest ECMWF fields

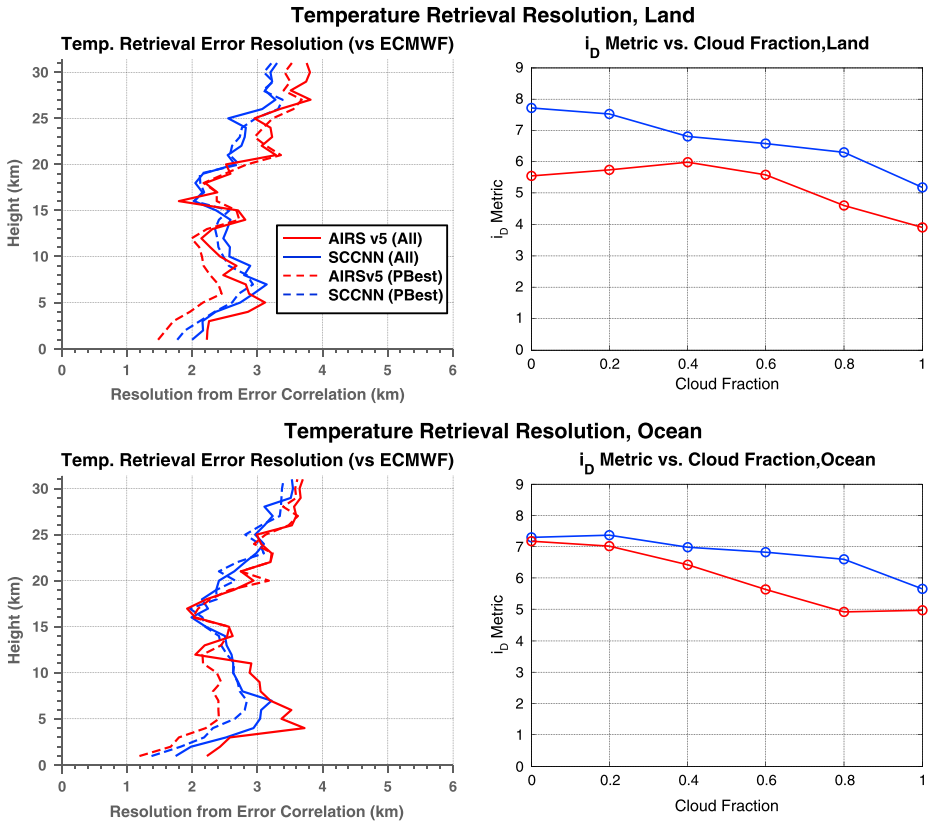

Figure 18. AIRS/AMSU temperature resolution from retrieval error correlation.

available. Figure 15 shows results for SCC/NN and AIRS v5 versus the collocated ECMWF data. The results indicate comparable RMS error of SCC/NN and ECMWF versus the dedicated radiosondes, and the expected performance improvement versus AIRS v5 under cloudy, non-PBest conditions. Systematic errors versus both the sondes and ECMWF are slightly greater for SCC/NN versus AIRS v5 for temperature above 100 mbar, and improved overall for SCC/NN versus AIRS v5 for water vapor as well, though AIRS v5 has better performance for PBest-accepted retrievals.

4. Preliminary Results for CrIS/ATMS

Figure 16 shows initial validation results for the CrIS/ATMS version of SCC/NN versus ECMWF truth data. The retrievals were calculated globally for 3 days: 19 February 2014, 7 May 2014, and 19 August 2014, which are all after the training epoch. Figure 16 shows RMS and errors for temperature and for water vapor, with SCC/NN and ECMWF averaged into the layers used by the NOAA CrIS/ATMS Environmental Data Record (EDR) [*Liu et al.*, 2012]. The EDR product obtained from NOAA's Comprehensive Large Array-data Stewardship System archive [*Rank and Reynolds*, 2005] is shown for comparison. This product is a physical retrieval, with the microwave retrieval as the first guess. The EDR product convergence flag was used to sort all retrievals into three categories, based on whether the EDR algorithm converged ("QC High"), the EDR algorithm only converged for ATMS ("QC Low"), and the EDR algorithm did not converge ("QC Poor"). The preliminary results for SCC/NN demonstrate good RMS performance consistent with the AIRS/AMSU version, and consistent performance across all 3 days and for all the EDR QC categories.

5. Retrieval Vertical Correlation for AIRS/AMSU

The extent to which retrieval error is correlated across neighboring vertical levels can be used as an indicator of retrieval resolution [*Serio et al.*, 2008]. One method of assessing resolution is to compute the retrieval error correlation matrix and compute the full width at half maximum (FWHM) of the correlation peak in the vicinity

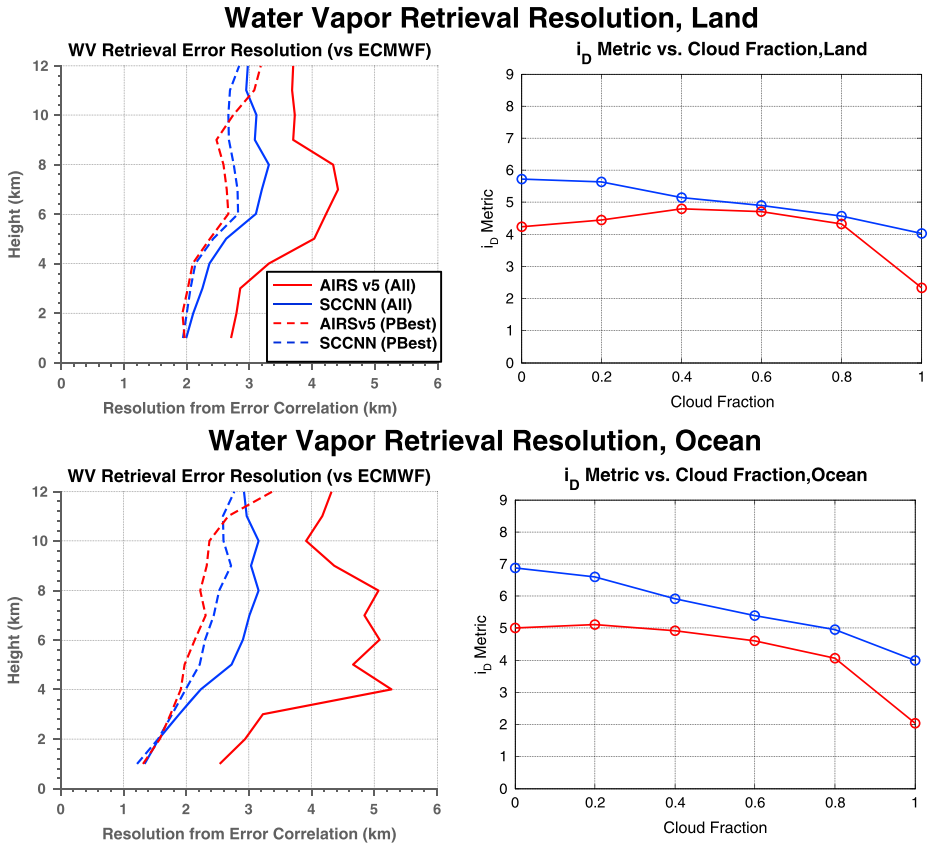

Figure 19. Water vapor resolution from retrieval error correlation.

of the diagonal. The resolution can be quantified using the degree of diagonalization i_D, a metric described by *Serio et al.* [2008]. This metric represents the relative contribution to correlation matrix norm from the diagonal and can be interpreted as the number of independent layers in retrieval.

Figure 17 shows the absolute value of the correlation matrix for AIRS/AMSU SCC/NN retrieval error versus ECMWF for temperature and water vapor mixing ratio on the golden day data set, with $|\text{latitude}| < 60$. From these matrices we compute vertical resolution by interpolating to a height grid up to 32 km and computing FWHM. We also compute i_D binned by cloud fraction. Figures 18 and 19 show the results for temperature and water vapor, respectively, along with results for AIRS v5 for comparison, plotted by land and ocean. (The noisy appearance of the plots of resolution versus height is a result of artifacts in thresholding on 0.5 to compute the FWHM where the correlation function may be slowly varying.) For temperature, SCC/NN and AIRS v5 generally have comparable resolution varying between 1 km and 3.8 km by height. Over land, SCC/NN shows slightly better resolution than AIRS v5 at high altitudes, leading to greater value of i_D for clear and cloudy profiles. Over ocean, SCC/NN and AIRS v5 show closer i_D, with the SCC/NN curve decreasing more slowly with cloudiness. For water vapor, SCC/NN shows a more pronounced resolution improvement over AIRS v5, with slower decrease in i_D at high cloudiness values.

6. Conclusion

We have described the current versions of SCC/NN for AIRS/AMSU and CrIS/ATMS. We have also shown comprehensive validation for SCC/NN with multiple data sets, demonstrating improved retrieval performance over previous physical retrievals that did not incorporate SCC/NN as a first guess, robustness to increased cloud fraction, and consistency over time. In future work, we will expand validation using dedicated sources, further develop and validate the CrIS/ATMS version of SCC/NN, and train and validate statistical QC approaches such as mixture density networks [*Tao et al.*, 2013]. NASA is currently evaluating several first

guess options for a first guess for CrIS/ATMS retrieval products that are expected to extend the AIRS/AMSU record, including SCC/NN, as well as other approaches similar to those used by the NOAA sounder science teams [*Susskind et al.*, 2015]. This work is expected to lead to a better understanding of the relative benefits or drawbacks of these respective approaches. While we plan a separate investigation into the benefits of the SCC/NN first guess on the AIRS v6 retrievals, recent work [*Susskind et al.*, 2014] indicates that the first guess has improved retrieval yields and performance in AIRS v6, while generally improving even the Not PBest results relative to the PBest ones.

Acknowledgments
The authors gratefully acknowledge contributions from Tony Reale, Frank Tilley, Frederick W. Irion, and Christy F. Cull. For questions about the sharing of data used in this paper, please contact the corresponding author at milstein@ll.mit.edu. This work was supported in part by the National Oceanic and Atmospheric Administration and the National Aeronautics and Space Administration under Air Force contract FA8721-05-C-0002 and in part by NASA contract NNH08AH88I. Opinions, interpretations, conclusions, and recommendations are those of the author and are not necessarily endorsed by the U.S. Government.

References

Bernstein, D. S. (2009), *Matrix Mathematics: Theory, Facts, and Formulas*, 2nd ed., Princeton Univ. Press, Princeton, N. J.

Blackwell, W. (2012), Neural network Jacobian analysis for high-resolution profiling of the atmosphere, *EURASIP J. Adv. Signal Process.*, *2012*(1), 1–11.

Blackwell, W. J. (2005), A neural-network technique for the retrieval of atmospheric temperature and moisture profiles from high spectral resolution sounding data, *IEEE Trans. Geosci. Remote Sens.*, *43*(11), 2535–2546.

Blackwell, W. J., and F. W. Chen (2009), *Neural Networks in Atmospheric Remote Sensing*, Artech House Inc., Boston, Mass.

Blackwell, W. J., and A. B. Milstein (2014), A neural network retrieval technique for high-resolution profiling of cloudy atmospheres, *IEEE J. Selec. Top. Appl. Earth Obs. Remote Sens.*, *7*(4), 1260–1270.

Cho, C., and D. H. Staelin (2006), Cloud clearing of Atmospheric Infrared Sounder hyperspectral infrared radiances using stochastic methods, *J. Geophys. Res.*, *111*, D09S18, doi:10.1029/2005JD006013.

Divakarla, M. G., C. D. Barnet, M. D. Goldberg, L. M. McMillin, E. Maddy, W. Wolf, L. Zhou, and X. Liu (2006), Validation of Atmospheric Infrared Sounder temperature and water vapor retrievals with matched radiosonde measurements and forecasts, *J. Geophys. Res.*, *111*, D09S15, doi:10.1029/2005JD006116.

Ferraro, R. R., W. Fuzhong, N. C. Grody, Z. Limin, M. Huan, C. Kongoli, P. Pellegrino, Q. Shuang, and C. Dean (2005), NOAA operational hydrological products derived from the advanced microwave sounding unit, *IEEE Trans. Geosci. Remote Sens.*, *43*(5), 1036–1049.

Fetzer, E., and E. Manning (2012), Failure of AMSU-A channel 4 and degradation of AMSU-A channel 5 NeDT affecting AIRS retrieval performance, edited.

Fetzer, E., D. Tobin, and B. Irion (2003), Comparison of AIRS retrievals and dedicated validation observations, paper presented at Optical Remote Sensing, Opt. Soc. of Am., Quebéc City, 3 Feb.

Hearty, T. J., and E. T. Olsen (2007), *AIRS/AMSU/HSB Version 5 Level 2 Performance and Test Report*, Jet Propul. Lab., Pasadena, Calif.

Lee, J. B., A. S. Woodyatt, and M. Berman (1990), Enhancement of high spectral resolution remote-sensing data by a noise-adjusted principal components transform, *IEEE Trans. Geosci. Remote Sens.*, *28*(3), 295–304.

Liu, X., et al. (2012), Retrieving atmospheric temperature and moisture profiles from SUOMI NPP CrIS/ATMS sensors using CrIMSS EDR algorithm, paper presented at Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International, 22-27 July.

Rank, R., and R. G. Reynolds (2005), Comprehensive Large Array-data Stewardship System (CLASS): A fully-distributed system, paper presented at AGU Fall Meeting.

Reale, T., B. Sun, F. H. Tilley, and M. Pettey (2012), The NOAA Products Validation System (NPROVS), *J. Atmos. Oceanic Technol.*, *29*(5), 629–645.

Rodgers, C. D. (2000), *Inverse Methods for Atmospheric Sounding: Theory and Practice*, World Sci. Publ. Co. Inc., Boston.

Serio, C., F. Esposito, G. Masiello, G. Pavese, M. R. Calvello, G. Grieco, V. Cuomo, H. L. Buijs, and C. B. Roy (2008), Interferometer for ground-based observations of emitted spectral radiance from the troposphere: Evaluation and retrieval performance, *Appl. Opt.*, *47*(21), 3909–3919.

Strow, L. L., S. E. Hannon, S. De Souza-Machado, H. E. Motteler, and D. Tobin (2003), An overview of the AIRS radiative transfer model, *IEEE Trans. Geosci. Remote Sens.*, *41*(2), 303–313.

Sun, B., A. Reale, D. J. Seidel, and D. C. Hunt (2010), Comparing radiosonde and COSMIC atmospheric profile data to quantify differences among radiosonde types and the effects of imperfect collocation on comparison statistics, *J. Geophys. Res.*, *115*, D23104, doi:10.1029/2010JD014457.

Sun, B., A. Reale, S. Schroeder, D. J. Seidel, and B. Ballish (2013), Toward improved corrections for radiation-induced biases in radiosonde temperature observations, *J. Geophys. Res. Atmos.*, *118*, 4231–4243.

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

Susskind, J., J. M. Blaisdell, and L. Iredell (2014), Improved methodology for surface and atmospheric soundings, error estimates, and quality control procedures: The Atmospheric Infrared Sounder science team version-6 retrieval algorithm, *J. Appl. Remote Sens.*, *8*, doi:10.1117/1.JRS.8.084994.

Susskind, J., J. M. Blaisdell, L. Iredell, and F. Keita (2011), Improved temperature sounding and quality control methodology using AIRS/AMSU data: The AIRS science team version 5 retrieval algorithm, *IEEE Trans. Geosci. Remote Sens.*, *49*(3), 883–907.

Susskind, J., L. Kouvaris, J. M. Blaisdell, and L. Iredell (2015), CrIS/ATMS retrievals using the latest AIRS/AMSU retrieval methodology, AIRS Science Team Meeting.

Tao, Z., W. J. Blackwell, and D. H. Staelin (2013), Error variance estimation for individual geophysical parameter retrievals, *IEEE Trans. Geosci. Remote Sens.*, *51*(3), 1718–1727.