Detail Enhancement of AIRS/AMSU Temperature and Moisture Profiles Using a 3D Deep Neural Network

Adam B. Milstein[a], Joseph A. Santanello[b], and William J. Blackwell[a]

[a] *MIT Lincoln Laboratory, Lexington, MA*

[b] *NASA Goddard Spaceflight Center, Greenbelt, MD*

Corresponding author: Adam B. Milstein, milstein@ll.mit.edu

File generated with AMS Word template 2.0

ABSTRACT

In recent decades, spaceborne microwave and hyperspectral infrared sounding instruments have significantly benefited weather forecasting and climate science. However, existing retrievals of lower troposphere temperature and humidity profiles have limitations in vertical resolution, and often cannot accurately represent key features such as the mixed layer thermodynamic structure and the inversion at the planetary boundary layer (PBL) top. Because of the existing limitations in PBL remote sensing from space, there is a compelling need to improve routine, global observations of the PBL and enable advances in scientific understanding and weather and climate prediction. To address this, we have developed a new 3D deep neural network (DNN) which enhances detail and reduces noise in Level 2 granules of temperature and humidity profiles from the Atmospheric Infrared Sounder (AIRS)/Advanced Microwave Sounding Unit (AMSU) sounder instruments aboard NASA's Aqua spacecraft. We show that the enhancement improves accuracy and detail including key features such as capping inversions at the top of the PBL over land, resulting in improved accuracy in estimations of PBL height.

1) Introduction

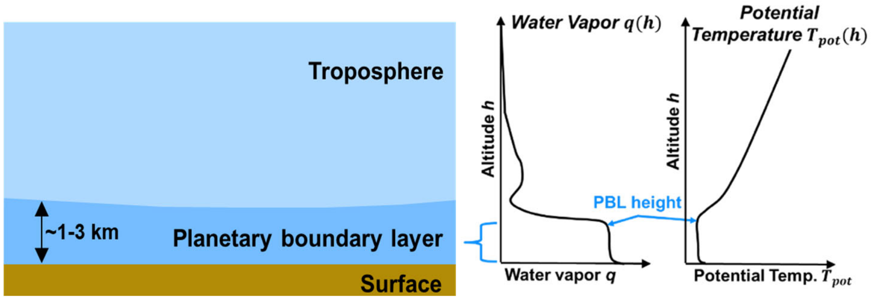

Figure 1: Planetary boundary layer (PBL) illustration, and illustrated water vapor and potential temperature profiles showing sharp gradient at the top of the PBL

Improved understanding of thermodynamics within the planetary boundary layer (PBL), including its structure and PBL height (PBLH) over land and water as a function of time of day, is of great importance to NASA, as recommended by the National Academy of Sciences in the 2017 Decadal Survey for Earth Science and Applications from Space ("ESAS 2017") (National Academies of Sciences & Medicine, 2018; "Planetary Boundary Layer Decadal

Survey Incubation Study Team,") and subsequently by the NASA PBL Incubation Study Team Report (STR; (Teixeira et al., 2021)). During the ESAS 2017 process, improved PBL monitoring from space was identified as a high priority across multiple interdisciplinary panels and science and application questions, leading to the current NASA PBL Decadal Survey Incubation (DSI) program that will invest in future spaceborne PBL mission development.

In the last two decades, spaceborne microwave and hyperspectral infrared sounding instruments on Aqua, Suomi NPP, and JPSS have significantly improved weather forecasting (Liu & Li, 2010; Pangaud, Fourrie, Guidard, Dahoui, & Rabier, 2009). These instruments use passive measurements of thermal radiation emitted by the atmosphere at different wavelengths to retrieve vertical profiles about the temperature T and water vapor q. The retrieval process makes use of the fact that different wavelengths are sensitive to thermal emissions from different altitudes. However, the sounding retrievals are fundamentally limited in vertical resolution, with a function called the "averaging kernel" describing how small point changes in the true state of the atmosphere become vertically spread out among the surrounding profile levels in the retrieval estimate (Rodgers, 2000). As a result, retrievals from sounders often cannot accurately represent key features such as the mixed layer thermodynamic structure and the inversion at the PBL top, the latter of which appears as a sharp gradient in q or potential temperature T_{pot} as illustrated in Figure 1. With the mixed layer itself being ~1-3 km thick, previously reported AIRS T and q profile resolution (and resultant PBLH) errors on the order of ~1-2 km (Martins et al., 2010) are not sufficient, and, alone, fall well short of the ESAS recommendation of ~100-300 m vertical resolution for new PBL observing systems. Because of the existing limitations in PBL remote sensing from space, there is an urgent need to improve routine, global observations of the PBL and enable advances in scientific understanding and weather and climate prediction. With the hyperspectral infrared sensor record continuing beyond the next decade with SNPP, JPSS, IASI, and planned GEO sounders, new methodologies that improve sounding capability will yield benefit for a long time to come.

Sounding retrieval algorithms reconstruct a vertical distribution of atmospheric temperature and water vapor from the observations, which consist of thermal IR and microwave radiation emitted by layers of the atmosphere and measured by a sounder instrument on orbit (Rodgers, 2000). The objective is to estimate the state of the atmosphere (represented by unknown parameter vector x, a vertical profile of $T(p)$ or $q(p)$), given spaceborne spectral radiance observations (represented by a radiance vector y). The IR observations used in y are typically a cloud-cleared spectrum derived from a 3 by 3 group of neighboring cloudy spectra beforehand (though single-FOV retrievals have recently advanced (Irion et al., 2018) (DeSouza-Machado et al., 2018) as well). This inverse problem is ill-posed, lacking a single unique solution, with vertical details beyond a certain (scene- and data-dependent) vertical resolution limit not directly observable from the spectral measurements, including in the lower troposphere and the PBL. This limitation in PBL information content from the instrument motivates interest in exploring the new, AI-informed technique described herein.

Two approaches to sounding retrievals are physical retrievals and statistical regression retrievals. The physical retrieval approach uses a forward model (the radiative transfer model) to calculate the expected measurements $f(x)$ given a specific atmospheric state x. The estimate $\hat{x}$ is iteratively adjusted to reduce the squared difference between the observations y and prediction $f(\hat{x})$. Due to the ill-posed nature of the inverse problem, regularization is typically required to stabilize the retrieval. Statistical regression approaches, including neural networks, learn an empirical relationship $\hat{x} = z(y)$ between an ensemble of measurements $y_{ensemble}$ and collocated truth datasets $x_{ensemble}$ (W.J. Blackwell & Chen, 2009). For a neural network, z is a nonlinear function composed of simple, interconnected computational elements, or nodes, defined by learned weight parameters and an activation function. A key advantage for neural networks over physical approaches is that they are fast and accurate, and, as universal function approximators (Hornik, 1991), they can empirically learn complex, often indirect and nonlinear dependencies embedded in the data that may be difficult to physically model(W.J. Blackwell & Chen, 2009). Neural networks have attracted increasing wide use from the sounding and remote sensing community in recent years (F. Aires, Rossow, Scott, & Chédin, 2002; W. J. Blackwell & Chen, 2005; Cai et al., 2020; Chase, Nesbitt, & McFarquhar, 2021; Maddy & Boukabara, 2021), including, recently, the use of convolutional neural networks in surface temperature retrievals (Filipe Aires, Boucher, & Pellet, 2021) and precipitation nowcasting (Samsi, Mattioli, & Veillette, 2019; Veillette, Samsi, & Mattioli, 2020).

In recent years, the Level 2 retrieval algorithm for Aqua's Advanced Microwave Sounding Unit (AMSU) (Susskind, Blaisdell, & Iredell, 2014; Susskind et al., 2020) has combined both of these approaches by using MIT-LL's Stochastic Cloud Clearing/Neural Network (SCC/NN) retrieval (W. J. Blackwell & Milstein, 2014; Milstein & Blackwell, 2016) as first guess for NASA's physical retrieval. The introduction of SCC/NN in versions 6 and 7 of the product as the first guess has led to improved accuracy and yield in down to the surface, including the PBL, versus previous versions with a different regression first guess(Susskind, Blaisdell, et al., 2014; Yue et al., 2020). SCC/NN builds upon years of work by Blackwell and colleagues (W. J. Blackwell, 2005; W. J. Blackwell & Chen, 2005; W.J. Blackwell & Chen, 2009; Cho & Staelin, 2006), and is a highly capable retrieval approach for T and q in its own right, achieving state of the art accuracy under a wide variety of cloud cover conditions. The most recent version of SCC/NN, which we call here "v7 NN" for convenience, improves upon the v6 NN by utilizing a more comprehensive training set, among other documented improvements (Susskind, Blaisdell, et al., 2014; Susskind et al., 2020), and includes architectural changes which have improved overall accuracy and precision in the PBL(Wong et al., 2018; Yue et al., 2020). In the current Level 2 retrievals, PBL phenomenology is, to a significant degree, introduced via the SCC/NN first guess, along with most of the fine vertical structure in the retrievals, with the physical retrieval adding coarse (~2 km) structure to the SCC/NN first guess (Susskind, Lee, & Iredell, 2014). This combination of neural networks and physical retrieval to improve operational science products resulted from over a decade of investment by NASA, long predating the recent intensified focus on AI from the larger science and technology community.

In this paper, we present a new AI technique for enhancing the resolution and accuracy of existing passive sensor retrievals, and show that it improves PBLH accuracy specifically. This technique uses a deep neural network that exploits temperature and moisture structure over a 3D volume to improve vertical resolution. In contrast to previous NN algorithms which performed regressions between the retrieved variable x and radiances y like those described above, the DNN described enhances a volumetric 3D dataset of the retrieved variable $\hat{x}$ itself, without using the radiances again. The DNN, trained using ensembles of realistic scenes from reanalysis fields, is a powerful way to incorporate prior knowledge, representing the joint statistics of the whole 3D scene being reconstructed with high fidelity. We describe our recent progress in developing this AI approach for the AIRS/AMSU sounders aboard Aqua. We show how the deep network results in improved thermodynamic profiles (T and q) in the lower troposphere, that have the physical (realistic) structure of the PBL mixed-layer and PBL-top inversion, and we demonstrate how these profiles can be used to derive PBLH over land more accurately than existing approaches. The AI approach that we present can be used in principle to enhance a wide variety of program-of-record retrievals, including regressions and physical retrievals. In the present work, we chose to enhance the v7 NN AIRS/AMSU retrievals. We found that using the v7 NN (first guess) retrievals rather than the v7 (final) retrievals was helpful in this initial investigation in a number of ways. For example, v7 NN provides the fine vertical structure and PBL information (of most interest to this study) used by the final v7 product, and, in contrast to recommended practice for the physical retrieval, is intended to be used down to the surface at all instrument footprints.

The DNN is trained using 3D T and q fields from the ERA5 reanalysis model (Hersbach et al., 2018) as an ensemble of realistic, detailed scenes of vertical thermodynamic profiles and PBL characteristics. One conceivable approach would be to assume that the ERA5 fields are the true state of the atmosphere, and pair them as training targets with the v7 NN retrievals or AIRS radiances as the corresponding DNN inputs. However, we do *not* choose this approach, as ERA5, while a good model, is not the true state of the atmosphere, and the temporal or spatial correspondence with the instrument data is not exact. Our goal is to restore the level of detail typical of ERA5 without relying on the accuracy or truth of the ERA5 model. Instead, we pair the ERA5 training targets with *simulated* v7-NN-like retrieval granules derived from the ERA5 fields in known fashion. First, we simulate how the retrieval process (the instrument and retrieval algorithm) degrades the "true" atmospheric state using a combination of vertical smoothing and noise. We then train the enhancement DNN to reverse, as best it can, the degradation process. Through this approach, we model what the retrievals would look like *if* the true atmospheric state were equal to ERA5, and train an enhancement function (our DNN) which restores the original atmosphere given a v7 NN-like retrieval granule as input. We then execute the trained enhancement function on real AIRS v7 NN retrieval granules to restore a better representation of the true atmosphere. In this paper we use ERA5 as a reference for preliminary validation to demonstrate performance improvements versus the v7 NN retrievals. Future work is planned to address validation versus in-situ truth data.

There is unique value in generating enhanced sounder retrievals to investigate the PBL rather than relying solely on modeled products such as ERA5. ERA5 has been shown to produce realistic estimates of PBL structure compared to observations (e.g. radiosondes) and other reanalysis models (Guo et al., 2021), typically underestimating PBLH by ~130 m. Nevertheless, reanalyses are limited in terms of assimilated PBL observations and lower tropospheric radiances, and contain inherent biases and compensating errors that require independent validation, beyond what is available from sparse, intermittent radiosondes. Hence, profile retrievals derived uniquely from satellite observations remain important, and can be used for global, gridded process studies and independent assessment of numerical weather and climate models. In addition, the retrieval enhancements presented here are compatible with near-real-time operation, while reanalyses such as ERA5 have significant latencies. We also note that multiple IR sounders with geostationary coverage are planned in the coming years (Li, Paul Menzel, Schmit, & Schmetz, 2022), with China’s Fengyun-4 series currently in orbit. Improved IR retrieval capabilities in combination with these new instruments will enable dramatic improvements in spatial and temporal coverage versus current polar-orbiting sounders. The technique presented here will be fully applicable in principle to these next-generation IR sounder missions.

2) Deep Neural Network Enhancement Approach

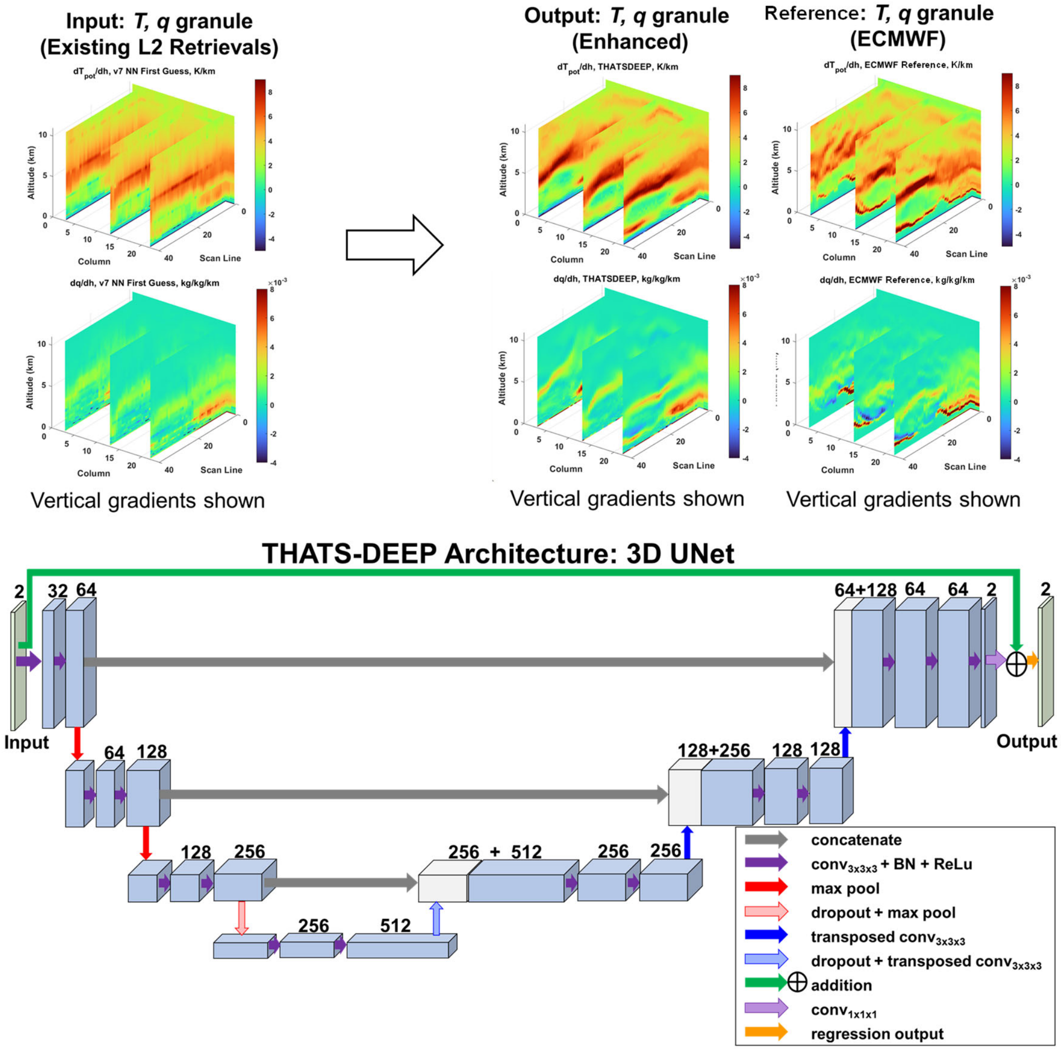

Figure 2: THATS-DEEP architecture and real example 3D granule images

Deep neural networks (DNN), including convolutional neural networks (Goodfellow, Bengio, & Courville, 2016) are highly expressive, and commonly used in AI to denoise, enhance, synthesize, or detect complicated phenomena in images or 3D datasets. They can be trained on a large ensemble of datasets and can provide a highly accurate representation of joint dependencies that conventional image enhancers or prior models aimed at regularizing solutions to inverse problems typically cannot. The AI agent we present here is a DNN aimed at removing typical sources of error (Rodgers, 2000) in sounding retrievals such as smoothing error (the vertical resolution limit of the observation and reconstruction technique in combination) and retrieval noise, resulting in enhanced vertical resolution with clearer PBL features, including sharper gradients, as well as more clearly-defined features in the mid- and upper atmosphere that can better constrain the sensor model. The agent is a 3D DNN operating on volumetric 3D granule datasets of size $N_{rows} \times N_{cols} \times N_{levels} \times 2$ (where the 2 channels are T and q, and “rows”/“columns” correspond to the 2D spot grid on which the profiles are geolocated). Specifically, the input is a 3D retrieval granule (in this case, the v7 NN, available to users in the Level 2 support product as the first guess), and the output is an enhanced retrieval granule of the same dimension. The DNN is therefore learning and utilizing complicated joint dependencies over the whole 3D volume, as well as joint dependencies between T and q to enhance the retrievals. We call this AI prior model *T*emperature and *H*umidity *AT*mospheric *S*ounding *DE*tail-*E*nhancing *P*rior, or “THATS-DEEP”.

The architecture for THATS-DEEP, along with a real example granule image, are illustrated in Figure 2. The DNN is a residual learning convolutional neural network with a 3D UNet (Ronneberger, Fischer, & Brox, 2015) architecture, an encoder-decoder network with skip connections, and a mixed gradient loss function (Lu & Chen, 2019). This loss function includes a sum of the mean squared error for the image and the mean squared error for the vertical gradients of the image, with the gradient term weighted by a factor of 2. (The mixed gradient was selected because it was found to offer a modest benefit in image detail enhancement, but overall gave visually and quantitatively similar results to the common mean squared error loss function.) As Figure 2 illustrates, the convolutional layers use rectified linear units (“ReLU”) and batch normalization, which are commonly used to maximize training performance in deep networks (Goodfellow et al., 2016). Dropout layers are used to regularize the network (Goodfellow et al., 2016). Kernels of size 3x3x3 were used.

The training set is a database of T and q fields from the European Centre for Medium-Range Weather Forecasts (ECMWF) ERA5 reanalysis model (Hersbach et al., 2018), collocated with the AIRS/AMSU retrieval spots from a total of 6,707 granules, with an additional 1,118 testing and 1,116 validation granules set aside. The training, testing, and validation granules are randomly interleaved, and are completely distinct from one another, and the results in this paper are from the testing set. Each granule is 2250 km by 1650 km in extent. We assumed that the large granule extent (encompassing a wide variety of surface and atmospheric conditions) makes even consecutive granules sufficiently distinct in coverage to preclude significant overfitting concerns, and thus, in the current work, we did not try to guarantee larger temporal separation between the testing granules from the closest training granules. The training set size was determined by initial trial and error and inspection of the results to be sufficient to generalize well, though further study of the optimal training set size is likely needed. The training outputs, or targets, are the ECMWF fields..

The inputs at training are highly realistic simulated retrievals formed by degrading the ECMWF fields with synthetic smoothing error and noise representative of the real v7 NN retrievals we aim to enhance, for reasons highlighted below in a). The inputs at execution are the real v7 NN granules. The ERA5 fields were vertically interpolated from the provided 37 pressure levels to 91 sigma levels used by both SCC/NN and the ECWMF IFS (ECMWF, 2014) forecast fields used in SCC/NN's training set. (While all processing is done on the sigma levels, profiles in this paper are plotted versus altitude in kilometers, approximating height above surface for visualization purposes.) The 3D AIRS granules are initially of size 45x30x91 (where 45x30 is the AIRS L2 retrieval and SCC/NN granule format), and are cropped to size 40x24x56, making each dimension a multiple of 8 for convenience and simplicity with the encoder depth of the selected 3D UNet architecture. (Future iterations of this work will aim to refine the architecture with resizing layers to use the full set of instrument granule footprints, though we sidestep this additional effort for now.) The v7 NN granules and corresponding ERA5 data were selected from throughout 2010 and 2013 and interpolated to AIRS spot locations, and are distributed globally, though care was taken to maintain an even balance between ocean and land retrievals and between nonpolar versus polar regions.

THATS-DEEP is currently implemented using Matlab's Deep Learning toolbox (Beal, Hagan, & B., 2022), and trained using acceleration from the GPU nodes in the Lincoln Laboratory Supercomputing Center (LLSC) (Byun et al., 2012; Reuther et al., 2018) using the well-known stochastic gradient descent with momentum (SGDM) optimizer in randomly shuffled minibatches, with 65 epochs required for convergence before noticeable increases in validation error that would indicate overfitting. The training options used are listed in Table 1. The hyperparameters specifying the model were chosen by experimenting with different values and using the resulting training and validation loss values to select the best-performing model. Specifically, the "Momentum", "InitialLearnRate", and "L2Regularization" parameter were chosen by looping over a range of values and reviewing the loss values. The SGDM and Adam (Kingma & Ba, 2014) optimizer were evaluated, and gave overall similar results. The 3D UNet encoder is generated by the Matlab-supplied "unet3dlayers" function (with layers and connections modified for image enhancement rather than the defaults for segmentation). The UNnet encoder depth of "3" selected by searching over another loop. The use of gradient clipping was inspired by Kim et al.(Kim, Lee, & Lee, 2016), from work that used a different architecture for image enhancement.

Option Name	**Value**
Momentum	0.9
InitialLearnRate	0.1
LearnRateSchedule	Piecewise
LearnRateDropPeriod	60
GradientThresholdMethod	Absolute-value
GradientThreshold	0.005
L2Regularization	0.0001

Table 1 Matlab Deep Learning training options used in this work

a) Synthetic Training Input Model

As described above, we train the DNN using simulated retrievals with realistic errors rather than real v7 NN retrievals, while we execute with the real v7 NN retrievals. Our rationale for training with simulated rather than actual retrievals derived from instrument data is that the reanalysis fields and the real retrievals inevitably contain undesired temporal or spatial collocation discrepancies between the reanalysis and instrument as well as errors in the reanalysis itself, while the simulated retrievals have a known, exact correspondence to the inputs used to generate them. We have found that this known correspondence is essential for this technique, and our prior attempts to train a deep network with only real retrievals paired with ECMWF produced only minor enhancement. However, as a result, a high degree of realism is required for the training inputs, with errors similar to those that would be encountered in the real retrievals. We considered (and remain open to) a variety of methodologies for generating the simulated retrieval errors, including generative adversarial networks (GANs) for style transfer (Zhu, Park, Isola, & Efros, 2017) and first-principles approaches. For example, one could:

- Compute and apply retrieval averaging kernels on each footprint
- Compute realistic, synthetic cloudy radiance spectra for inputs into the AIRS/AMSU v7 NN algorithm.

However, in our current implementation, we chose a different approach, which balanced ease of implementation with direct control of both the smoothing and noise error models. We generate synthetic proxy data profiles from the ECMWF fields, using a purpose-built neural network to degrade them such that errors are very similar to those in retrievals from the AIRS/AMSU v7 NN. We call this training approach “Sim-NN”, and illustrate it in Figure 3. The inputs to this network are the ECMWF profiles (both T and q), transformed into their principal components for dimensionality reduction. These are augmented with ancillary inputs similar to those that would be required for running a cloudy radiative transfer model, including cloud liquid water path (from the AIRS/AMSU L2 product), the v7 NN cloudiness flags (“BTCorr”) indicating degree of cloud spectrum correction applied the stochastic cloud clearing algorithm, land fraction, surface pressure, cos(latitude), cos(scan angle), and cos(solar zenith angle), all suitably normalized and centered. The sim-NN network is a shallow, fully connected network with two hidden layers (17 and 13 nodes), and uses sigmoid activation functions. The network is trained in two passes. The first pass uses more dimensionality reduction on the inputs (25 PCs for T and 35 PCs for q), and the well-known Levenberg-Marquardt algorithm. The second pass relaxes the dimensionality reduction (90 and 90 PCs for T and q, respectively, with input layer nodes added and initialized to zero on the second pass), and uses the GPU-friendly scaled conjugate gradient algorithm, to refine the results of the first pass, enabling slight improvement in fidelity without the Levenberg-Marquardt algorithm’s memory requirements. For the training targets, the outputs were the corresponding v7 NN profiles, divided into the same vertical layer segments used in the generation of that product to match its characteristics. The training set was about one fourth the size used for THATS-DEEP, with limited improvement found for additional profiles. Training time on LLSC was under 1 hour.

An example of this approach is illustrated in Figure 3, for an example log(q) granule, including the input ECMWF granule, the output “Sim-NN” granule, and the corresponding v7 NN granule. The Sim-NN granule visually appears very similar to the v7 NN granule in vertical smoothness, which we quantify in future sections. However, the Sim-NN profiles paired with exact ECMWF “truth” profiles, with no temporal or spatial discrepancy between them, as desired. Sim-NN smooths each profile with a spatially varying vertical smoothing error based on cloud and other ancillary inputs, as would be expected in the real v7 NN datasets. However, retrieval noise also needs to be added to the Sim-NN retrievals. As with the smoothing error, we considered first principles approaches for estimating and adding retrieval noise to the proxy retrievals. However, in the current implementation, we choose a simple heuristic approach, which we found worked remarkably well. This approach is illustrated in Figure 4. In this approach, a “noise” pattern is derived from a real v7 NN granule, and added to the corresponding Sim-NN proxy granule. This noise is determined by denoising the v7 NN granule with a 3D median filter and differencing the filtered data from the original. The median filter was chosen due to its simplicity and its well-known edge-preserving denoising capabilities. The noise-like 3D pattern we obtain was found to be representative of 3D real errors in the v7 NN granules, including scene-dependent errors that would be complex to model synthetically otherwise.

We considered that the “noise” determined using the median filter may include some 3D T and q features other than true retrieval noise, and thus might cause THATS-DEEP to mistakenly reject true image features (such as some sharp gradients) as if they were noise. However, we have found empirically that this was not a significant or driving issue, as such sharp image features of interest do not exactly align between the Sim-NN and corresponding v7 NN granules. Hence, treating such features as additive noise does not typically cause confusion for THATS-DEEP in practice.

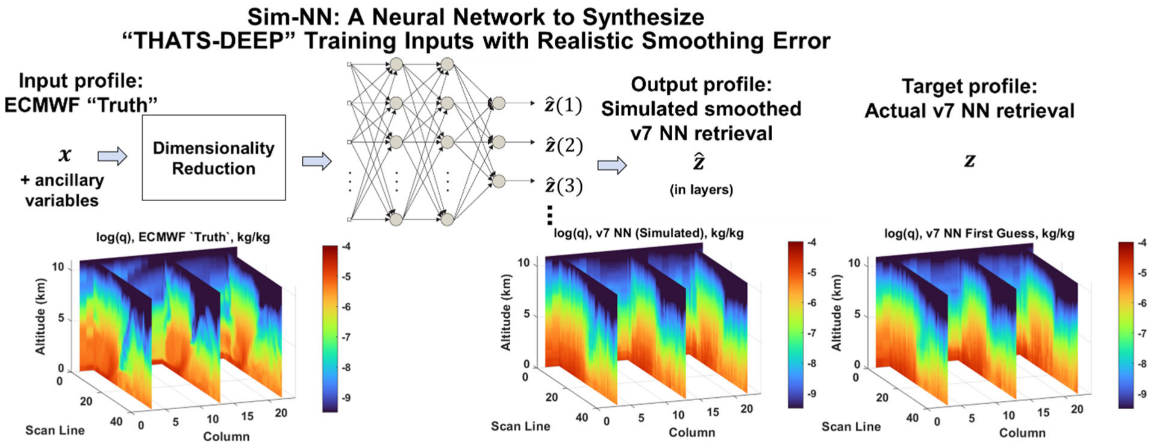

Figure 3: Sim-NN approach for generating realistic proxy v7 NN data to use as training inputs for THATS-DEEP

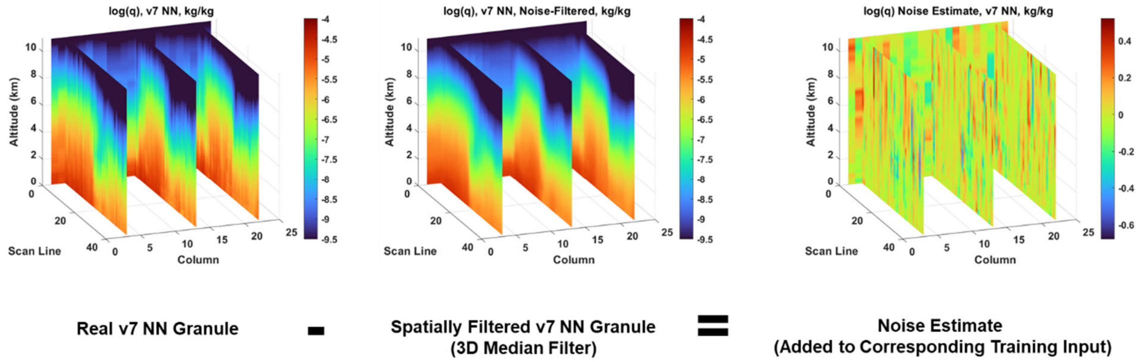

Figure 4: Heuristic approach used to add noise derived from real v7 NN retrieval granules to synthetic inputs used to train THATS-DEEP

3) Results

a) Example granules

First, we demonstrate the impact of THATS-DEEP with three example test granules from AIRS/AMSU, all taken from 2010 (which, along with 2013, was included in our training, validation, and independent testing datasets). These example scenes are illustrated in Figure 5 as VIS/NIR or LWIR preview images from the AIRS, and include (a) a July 16, 2010 UTC 11:53 daytime scene over Africa; (b) a December 16, 2012 UTC 18:47 nighttime scene over China and Southeast Asia; and (c) an August 21, 2010 UTC 06:53 scene over northern Russia. The arrows superimposed on the images depict the direction and approximate location of slice plots shown in the subsequent figures.

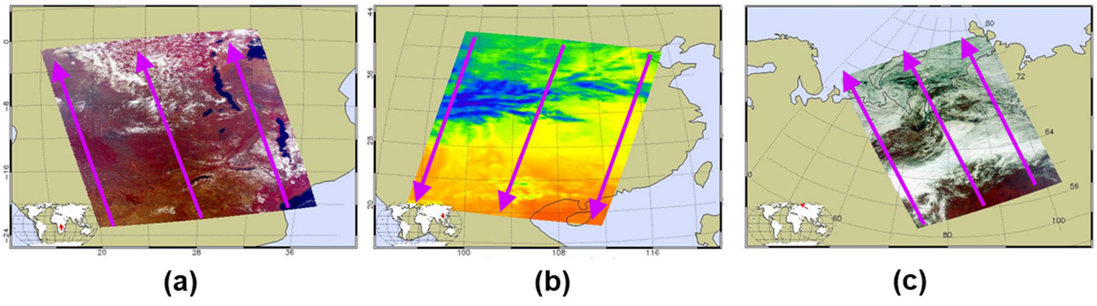

Figure 5: Three example scenes, shown as VIS/NIR or LWIR images from AIRS, with arrows depicting the direction and approximate location of slice plots in the subsequent figures. (a) July 16, 2010 UTC 11:53 daytime scene over Africa; (b) December 16, 2012 UTC 18:47 nighttime scene over China and Southeast Asia; (c) August 21, 2010 UTC 06:53 daytime scene over northern Russia.

Figure 6 shows example results, as 3D slices through the granule in Figure 5 (a). The granule is shown as slices of potential temperature T_{pot}, the log of specific humidity ($\log q$), the specific humidity q and vertical gradients $\frac{dT_{pot}}{dz}$ and $\frac{dq}{dz}$, all chosen to highlight key details. Slices through the original AIRS/AMSU v7 NN are shown, along with the enhanced THATS-DEEP result, and the corresponding ECMWF fields. The scene includes features of a convective boundary layer under mostly clear conditions, with the height of the mixed layer slowly varying across the granule. The original v7 NN retrievals, while a significant improvement over previous AIRS L2 retrievals (Susskind, Blaisdell, et al., 2014; Susskind, Lee, et al., 2014) contain smoothing error and retrieval noise which are intrinsic to the hyperspectral IR sounding process (including both the measurement and retrieval algorithm)(Rodgers, 2000). These errors vary from profile to profile, as well as vertical gradient artifacts at regular intervals due to the layer-wise retrieval method. As a result, key PBL features as well as other vertical structure are not always clearly discernible. However, while ECMWF is not exact truth (as it is for the synthetic training profiles), the enhanced outputs from THATS-DEEP look far more similar to the ECMWF granules than the input v7 NN retrievals do, including sharper, more well-defined moisture and potential temperature gradients at the top of the PBL and more accurately defined vertical structure in temperature and moisture, a key science objective of this effort. An example profile of q and T_{pot} versus altitude from this granule is plotted in Figure 7, demonstrating how THATS-DEEP enhances the accuracy of the inversion location at the top of the mixed layer versus the v7 NN profile.

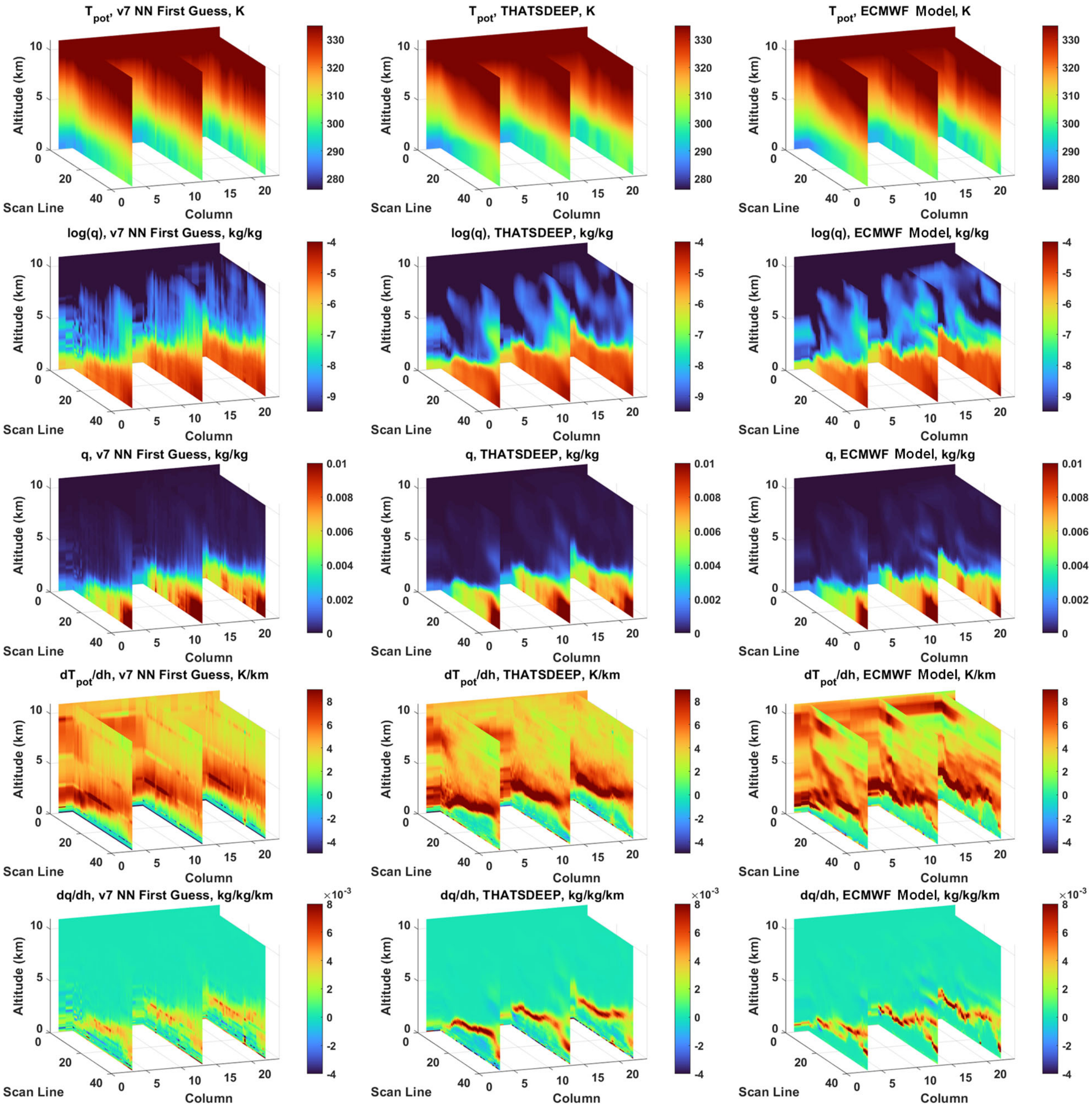

Figure 6: Example THATS-DEEP enhancement results for real AIRS/AMSU v7 NN granule slices of the scene in Figure 5 (a)

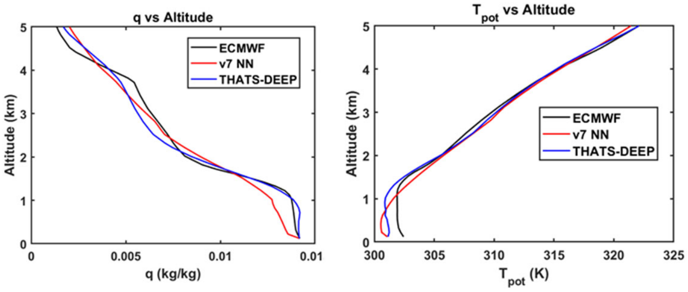

Figure 7: Example plot of *q* and T_{pot} versus altitude from example granule of scene in Figure 5 (a) and Figure 6, for v7 NN, THATS-DEEP, and ECMWF

Figure 8 shows slices through the granule in Figure 5 (b), a night-time overcast scene over China and southeast Asia (including Vietnam, Laos, and Myanmar). The scene is complex in terms of variable structure, including nighttime/stable PBL close to the surface and one more layer around 2-3 km, as demonstrated in the humidity gradient. The original v7 NN retrievals, while again significantly improved relative to previous L2 retrievals, also contain the aforementioned smoothing error and retrieval noise. The enhanced outputs from THATS-DEEP enhance key vertical structure and detail, showing that elevated layers of moisture and moisture gradients can be restored. This is particularly important for processes such as elevated convection layers that are important for severe weather and aviation forecasting, respectively. These types of scenes are particularly challenging from space, and demonstrate the potential of this approach.

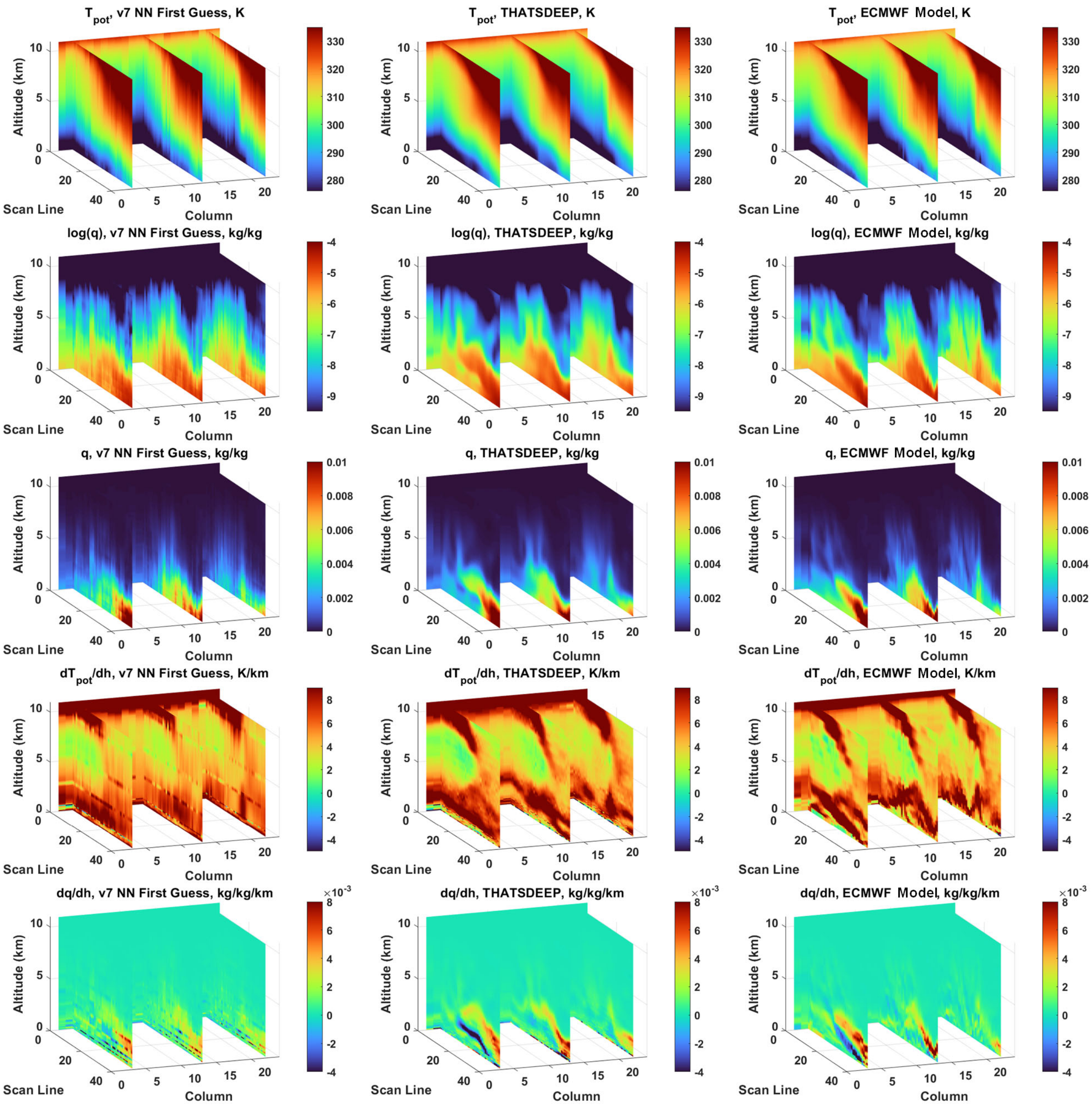

Figure 8: Example THATS-DEEP enhancement results for real AIRS/AMSU v7 NN granule slices of the scene in Figure 5 (b)

Figure 9 shows slices through the granule in Figure 5 (c), a daytime scene over northern Russia which contained rainy weather and low clouds. This scene also contains complex features and shows variable PBL structure and height. The restored scene in THATS-DEEP, once again, removes noise and layer artifacts from the original retrievals and sharpens key details and gradient features, including the varying height of the near-surface moisture layer.

b) Simulated vs Real Granule Example

As outlined in 2)a), THATS-DEEP is trained using simulated input data and meant to be executed using real input data. Therefore, it is instructive to look at an example of the simulated versus real test data for the same granule. Figure 10 shows results using the simulated Sim-NN granule corresponding to the real one in Figure 9. As in Figure 3, the Sim-NN inputs are comparable in appearance with the real v7 NN inputs, with similar layer artifacts, smoothing, and noise. As is typically the case, the simulated granule restored using THATS-DEEP shows similar removal of errors and enhancement of detail, as well as similar phenomenology overall to the real granule. These similarities between results for simulated and real datasets provide reassurance that the simulated datasets used to train the algorithm are representative, and that the algorithm performance is comparable on both as well. One notable, and expected, difference is that the simulated granules have features with close alignment with the corresponding features in the ECMWF granule, while this correspondence of key features is less closely aligned in the real dataset.

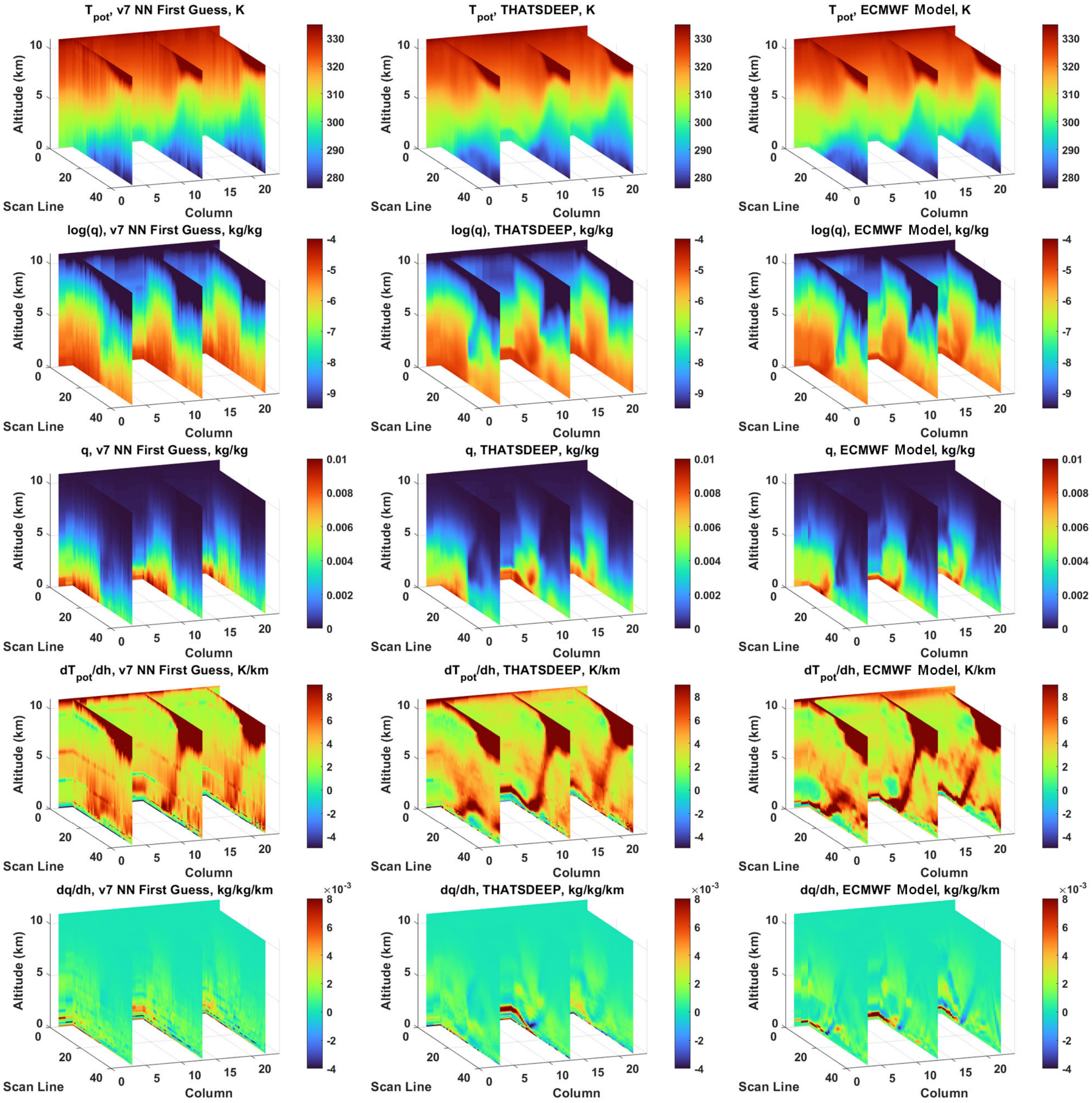

Figure 9: Example THATS-DEEP enhancement results for real AIRS/AMSU v7 NN granule slices of the scene in Figure 5 (c)

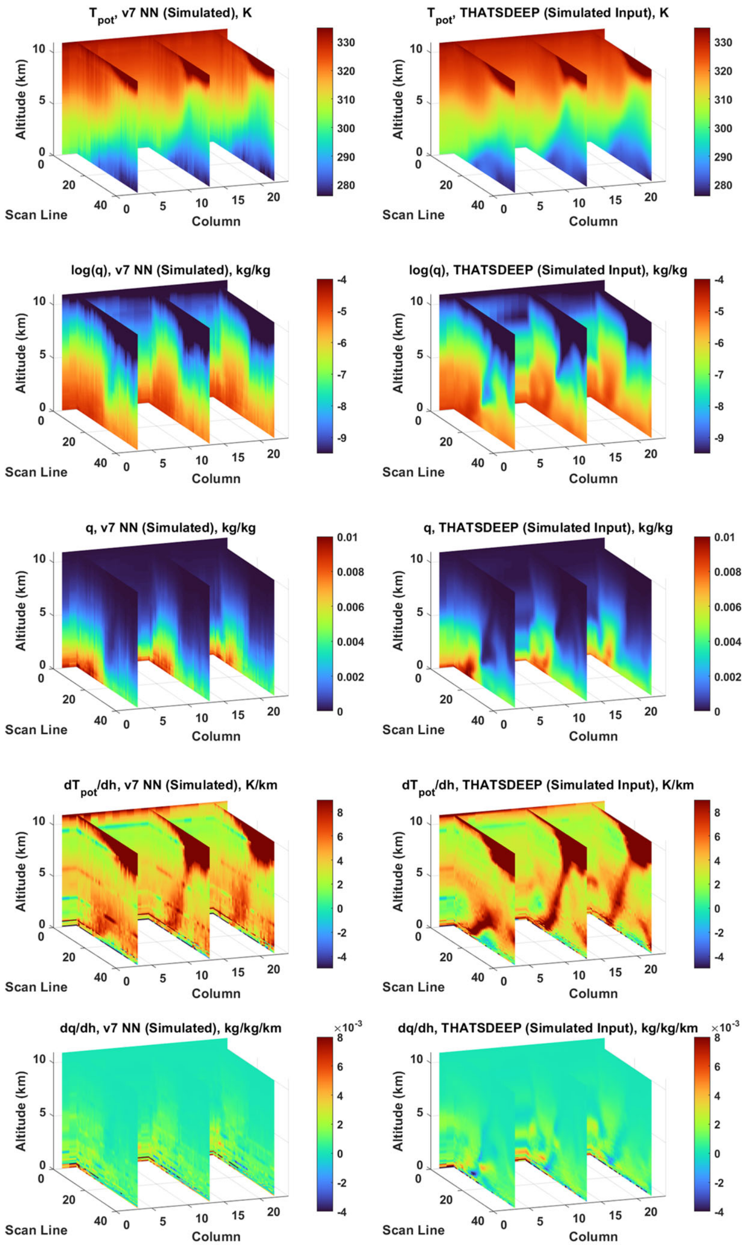

Figure 10: Example THATS-DEEP enhancement results for simulated granule slices corresponding to the real scene in Figure 5 (c) and Figure 9.

c) Daytime Land PBL Retrieval Study

A key science objective of this effort is to assess PBL retrieval capability over land, which is inherently more difficult from space due to land surface heterogeneity and emissivity uncertainties. To demonstrate the utility of THATS-DEEP, we assessed its impact using an ensemble of 149,622 profiles (daytime, |lat|<65, 2010 and 2013, land) from the test granule set described in 2). We show example PBLH assessments over land for both simulated and real datasets, with results described as follows.

d) Principal Component Enhancement

To assess vertical detail enhancement due to THATS-DEEP, we performed principal component (PC) analysis on the ensemble of T and $\log q$ test profiles including the original v7 NN profiles, the enhanced THATS-DEEP profiles, and the ECMWF profiles. The analysis was also performed on the corresponding simulated datasets. The PC transform vectors were computed using the ECMWF training set profiles. Figure 11 shows the PC analysis results for T. Figure 11 (a) shows the first 20 PCs of T profiles for ECMWF, the v7 NN, THATS-DEEP, Sim-NN, and THATS-DEEP (with Sim-NN inputs). Figure 11 (b) offers a closer look by showing the same PCs but normalized by the PCs for ECWMF, highlighting how each of the retrieval techniques falls off from unity. In particular, the blue curves showing THATS-DEEP PCs fall off from unity at a significantly slower rate than the red curves showing PCs for the v7 NN and Sim-NN inputs. This indicates that for both real and simulated detail, THATS-DEEP restores detail in the retrieved profiles, bringing them significantly closer to the level of detail in the ECMWF fields (unity in Figure 11 (b)). The first 20 principal vectors corresponding to these PCs are shown in Figure 11 (c), visualized as functions of altitude. The components boosted by THATS-DEEP include higher resolution features in their basis vectors, particularly in the PBL. Beyond the first 20 PCs, which contain the vast majority of the signal in the retrieved T profiles, the blue curves representing THATS-DEEP intersect and even dip below the red curves representing the corresponding inputs, indicating that THATS-DEEP attenuates those PCs to suppress noise rather than try to enhance them. The overall agreement between the dashed curves (representing simulation) and the solid curves (representing real retrievals) is reassuring, indicating that the simulated inputs and their level of smoothing are good models for the real retrievals, and that similar enhancement is provided to both by THATS-DEEP. (The Sim-NN curve is greater than the corresponding v7 NN curve at higher PCs, suggesting that the simulated inputs are noisier than the real ones.) Figure 12 shows a similar PC analysis for $\log q$, from which we draw similar conclusions about the utility of

THATS-DEEP in enhancing vertical detail and the good agreement between simulation and reality.

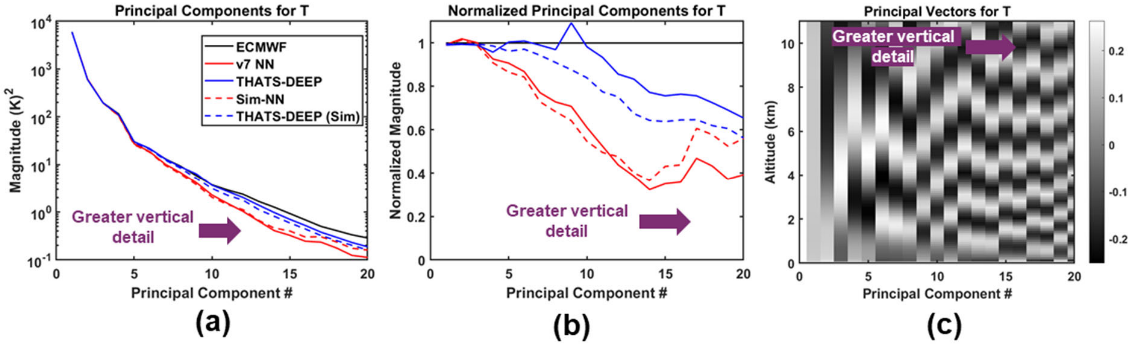

Figure 11: Principal component analysis of *T* profiles. (a) First 20 PCs of *T* profiles for ECMWF, v7 NN, THATS-DEEP, Sim-NN, and THATS-DEEP (with Sim-NN inputs). (b) The same PCs as (a) but normalized by the PCs for ECWMF, highlighting THATS-DEEP enhancement. (c) The first 20 PC vectors vs altitude, showing that enhanced components include greater vertical detail features

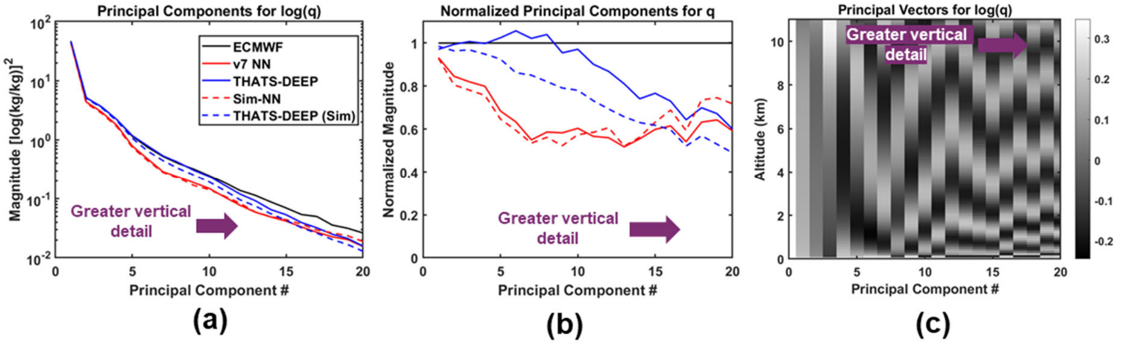

Figure 12: Principal component analysis of log *q* profiles. (a) First 20 PCs of log *q* profiles for ECMWF, v7 NN, THATS-DEEP, Sim-NN, and THATS-DEEP (with Sim-NN inputs). (b) The same PCs as (a) but normalized by the PCs for ECWMF, highlighting THATS-DEEP enhancement. (c) The first 20 PC vectors vs altitude, showing that enhanced components include greater vertical detail features

e) RMS Errors

We evaluated the retrieval error performance by computing RMS errors over the test profile ensemble. While improved vertical detail is a goal of this work, coarser layer-averaged errors are commonly used in the sounding community as a less noisy benchmark of overall error in profile data. Hence, we show layer-averaged and unaveraged RMS errors below. We examined performance for both simulated and real retrievals. Simulations provide a best-case bound of feasible performance because the ECMWF atmospheric state is the exact truth used to generate them, and THATS-DEEP was trained using Sim-NN inputs. On the other hand, real retrievals provide an upper bound on actual RMS error, due to the mismatch between ECMWF and the true atmospheric state at the time of the instrument measurements.

First, we examined the simulated retrievals, for which the ECMWF profiles are known, exact ground truth. Figure 13 shows RMS error for Sim-NN and corresponding THATS-DEEP T profiles and $\log(q)$ profiles versus ECMWF, both averaged to 2 km layers and unaveraged. The results indicate dramatic improvement due to THATS-DEEP relative to its Sim-NN inputs, with RMS errors typically reduced by 40-50%. While these retrievals are simulated, they are representative of real retrievals in terms of the level of smoothing error and noise relative to the ECMWF fields, as the PC analysis of d) establishes. Hence, Figure 13 provides a best-case bound on RMS error improvement that can be obtained for realistic input retrievals using THATS-DEEP.

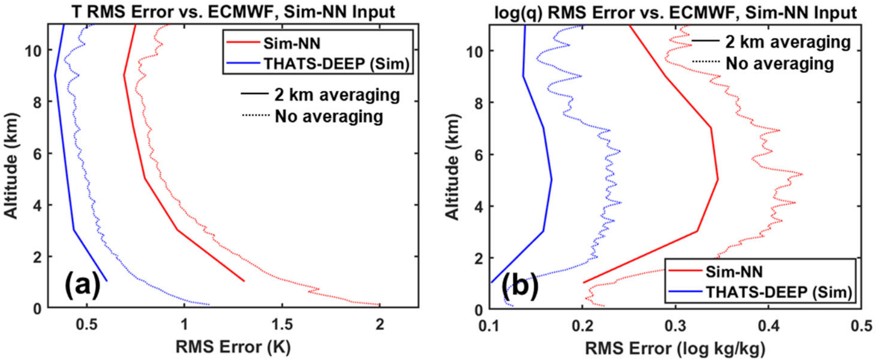

Figure 13: (a) RMS error as a function of altitude for T and log(q) for simulated retrievals, including Sim-NN and THATS-DEEP vs corresponding ECWMF profiles

We also determined errors for retrievals from real AIRS/AMSU, for which the ECMWF profiles do not correspond exactly to the true atmospheric state at the time of the measurement. Figure 14 shows RMS error for v7 NN and corresponding THATS-DEEP T profiles and $\log(q)$ profiles versus ECMWF, both layer-averaged and unaveraged. The layer-averaged results indicate improvements on the order of 5-20% due to THATS-DEEP relative to its v7 NN inputs. This improvement is statistically significant, as determined by a one-tailed F-test for variance reduction applied on all the profiles with a p-value of 0.05. However, the reductions, as expected, are less dramatic than the changes seen in the simulated datasets. We expect that this difference is due to two factors: 1) the aforementioned differences between the ECMWF fields and the true atmospheric state at the time of the AIRS measurements and 2) remaining discrepancies between the Sim-NN model used to train THATS-DEEP and the actual retrieval error statistics which impact performance on real data. The qualitative agreement in phenomenology between simulated and real granules noted in 2)a) suggests that the second of these factors is relatively small, but future validation efforts will work to quantify more precisely, in part using in-situ truth data.

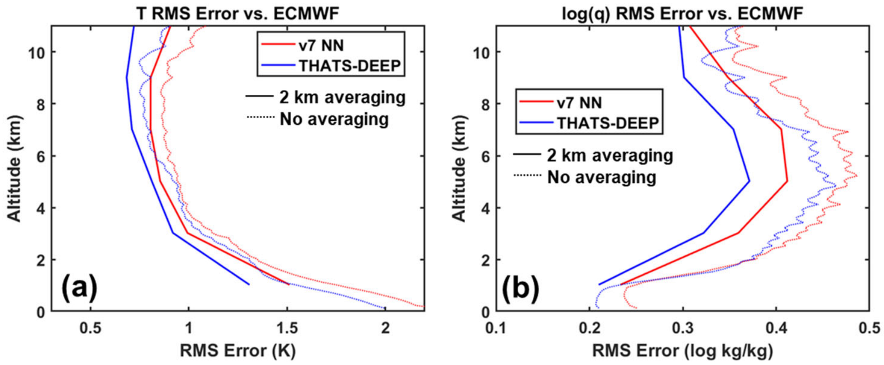

Figure 14: (a) RMS error as a function of altitude for T and log(q) for real retrievals, including v7 NN and THATS-DEEP vs corresponding ECWMF profiles

f) PBLH Results

We also evaluated PBLH from the retrievals in comparison to ECMWF for the same land, daytime test profiles. While many approaches can be used to evaluate PBLH and the best choice is typically regime-dependent, we used a vertical gradient approach similar to Seidel et al.(Seidel, Ao, & Li, 2010).While relative humidity gradient is one quantity which has previously been used to evaluate PBLH (Ding, Iredell, Theobald, Wei, & Meyer, 2021), we evaluated gradients in both q and T_{pot} individually, as both are useful indicators of the PBL top as illustrated in Figure 1, and both are more directly related to the variables retrieved by the algorithms. We also check PBLH derived from q against the PBLH derived from T_{pot} to confirm whether they are mutually consistent as physically expected. Specifically, the PBLH was determined using two different measures: The minimum value of $\frac{dq}{dz}$, and the maximum value of $\frac{dT_{pot}}{dz}$. To reduce errors, the search for the minimum or maximum gradient was limited to a certain interval within the vertical profile, as illustrated in Figure 15. For q, only segments of the profile for which q was within the interval $[0.8 \times \max(q), \max(q)]$ were included, reducing inadvertent selection of sharp gradient features too far away from the PBL. The PBLH was also bounded by the interval [0.29 km, 5 km]. Similarly, for T_{pot}, only segments of the profile for which T_{pot} was within the interval $[\min(T_{pot}), 1.02 \times \min(T_{pot})]$ were included. PBLH was bounded by [0.61 km, 5 km] and we took the additional step of rejecting any profile for which the maximum gradient was out of that range. This additional step was included because a distinct above-surface gradient was not identified for all T_{pot} profiles, including ECMWF profiles, even when it was for the corresponding q profile. All search intervals were selected after empirical tradeoffs to obtain best results.

As in e), we evaluate PBLH performance versus ECMWF for both simulated and real retrievals, with simulated retrievals offering a best case performance bound and real retrievals offering an upper bound on PBLH errors due to the mismatch between ECWMF and the true atmospheric state.

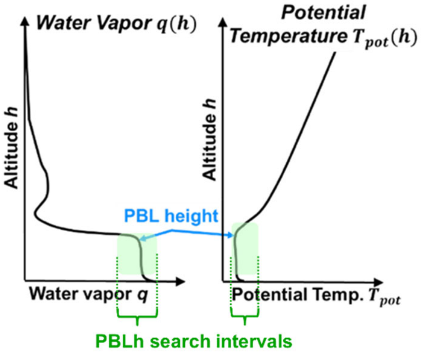

Figure 15: Illustration of the bounded search interval used to locate the top of the PBL for both q and T_{pot}

Before evaluating PBLH using gradients for v7 NN retrievals, we noted that these retrievals were generated using vertical layers, each from its own neural network, and thus gradient discontinuity artifacts were sometimes apparent at the boundaries between the layers. Figure 16(a) and (b) show an example of this with a slice through a v7 NN granule of q and its vertical gradient, respectively. We found that these artifacts impacted gradient-based PBLH assessments, causing them to cluster near layer boundaries. Hence, before calculating PBLH, we used a simple heuristic method to reduce these artifacts on a per-profile basis. Specifically, we assumed that the desired gradient at each layer boundary should be equal to a local three-point average (on the retrieval grid) of the actual gradient, and we added a small ramp function to each retrieval layer such that the gradients at the layer boundaries were equal to this desired value. Example results, including the modified granule of q and its vertical gradient, are shown in Figure 16(c) and (d), respectively. The q profiles are less blocky in appearance, and the layer artifacts are reduced in gradient. The same process was performed on T profiles. While this simple artifact reduction approach is not guaranteed to be optimal in any particular sense (and indeed appears to smooth the retrievals to a small degree), it significantly reduced the aforementioned clustering of PBLH estimates from the v7 NN retrievals in the results that follow.

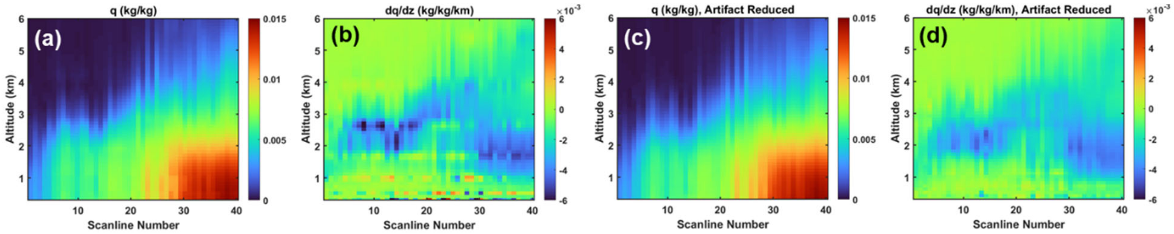

Figure 16: Example of reduction of vertical layer blocking artifacts from v7 NN and Sim-NN retrievals before attempting gradient-based PBL top location. (a) original *q*; (b) gradient of *q*, showing blocking artifact; (c) *q* after artifact reduction and (d) gradient of *q*, showing reduction in artifact

Figure 17 shows the assessment of PBLH, computed from *q*, shown as 2D histogram versus the PBL computed using the gradient of the corresponding ECMWF *q* profile. Figure 17(a) shows the results for v7 NN, while Figure 17 (b) shows the results for THATS-DEEP with real v7 NN inputs. The PBLH histogram peaks for THATS-DEEP maintain a slope of nearly 1 with respect to the ECMWF-derived PBLH on the horizontal axis over the whole plotted range, demonstrating its overall accuracy. The agreement between PBLH from THATS-DEEP and ECMWF is typically within ~0.5 km (as seen in the half-width of the histogram peak). In contrast, the PBLH for v7 NN is less accurate, tending to remain concentrated between 1 and 2 km, even when the ECMWF-derived PBLH is outside that range. Figure 17 (c) shows the results for THATS-DEEP with Sim-NN inputs, for which the ECMWF exactly represents the true atmospheric state. As in Figure 17 (b), the PBLH histogram peaks in in Figure 17 (c) maintain a slope of nearly 1 with respect to ECMWF, demonstrating overall accuracy. However, the histogram peaks for the simulated cases are noticeably narrower than for the real ones, with agreement typically between ~0.25 km and ~0.5 km. As in e), THATS-DEEP's performance with Sim-NN inputs represents a best-case bound due to the absence of model mismatch with the training set inputs, while performance with real inputs represents a worst-case bound due to ECMWF differing from the true atmospheric state in reality. However, for both real and simulated profiles, THATS-DEEP ultimately improved assessed PBLH accuracy relative to their inputs, which are typical of current program-of-record retrieval approaches in quality. We note that the results shown here do not include screening for cloudiness or "PBest" quality flags, and that we found such screening changes the distribution of PBLH values in the accepted profiles but does not change the relative performance. Figure 18 shows the assessment of PBLH, computed from T_{pot} for (a) v7 NN, (b) THATS-DEEP with real inputs, and (c) THATS-DEEP with simulated inputs. The results are consistent with those of Figure 17 which used q. As mentioned above, one notable difference in how we assessed PBLH from T_{pot} is that the T_{pot} profile's PBLH estimate was not accepted if it was below the search interval [0.61 km, 5 km], while a PBLH estimate was accepted for all q profiles. The percentage of profiles with accepted PBLH from T_{pot} is shown for all three cases.

We also compared PBLH values computed from q versus those computed from T_{pot} for consistency. Figure 19 shows 2D histograms of the PBLH estimates derived from T_{pot} versus the PBLH estimates derived from the corresponding q for (a) v7 NN, (b) THATS-DEEP with real inputs, and (c) ECMWF. For the ECMWF profiles, the agreement between T_{pot} and q is excellent, with a slope close to 1, and agreement precision typically within ~0.25 km. PBLH from T_{pot} is systematically about 0.25 km lower than T_{pot} from q. For THATS-DEEP, the slope is also close to 1, with similar systematic offset to the ECMWF profiles and tight agreement showing the overall consistency between T_{pot} and q in the enhanced profiles. The consistency is also significantly improved versus the v7 NN in (a) used as input.

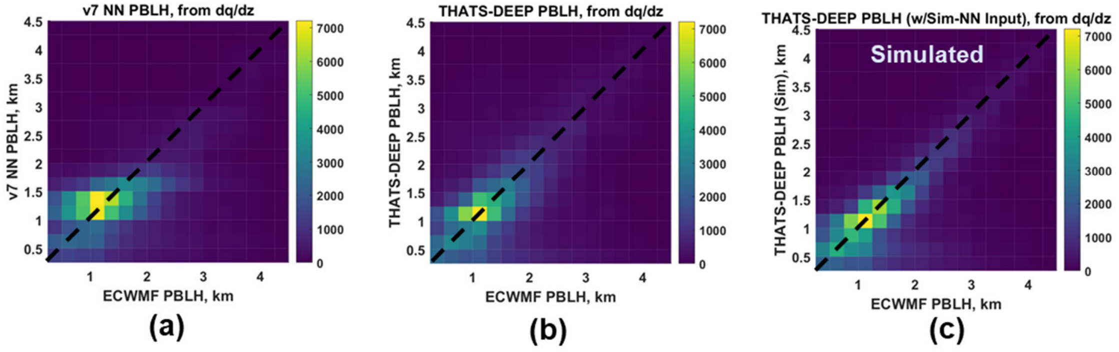

Figure 17: 2D histogram for PBLH for (a) v7 NN vs corresponding ECMWF profiles, (b) THATS-DEEP (with real v7 NN inputs), and (c) THATS-DEEP (with Sim-NN inputs) vs corresponding ECMWF profiles. PBLH was computed from the gradient of q

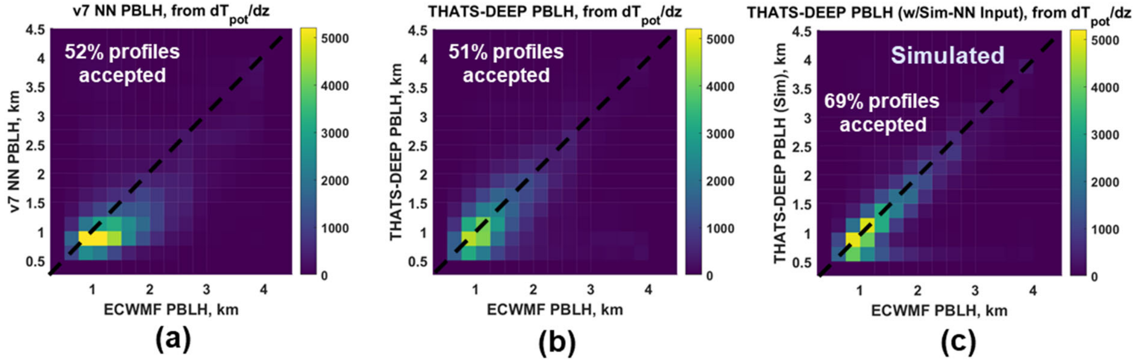

Figure 18: 2D histogram for PBLH for (a) v7 NN vs corresponding ECMWF profiles, (b) THATS-DEEP (with real v7 NN inputs), and (c) THATS-DEEP (with Sim-NN inputs) vs corresponding ECMWF profiles. PBLH was computed from the gradient of T_{pot}

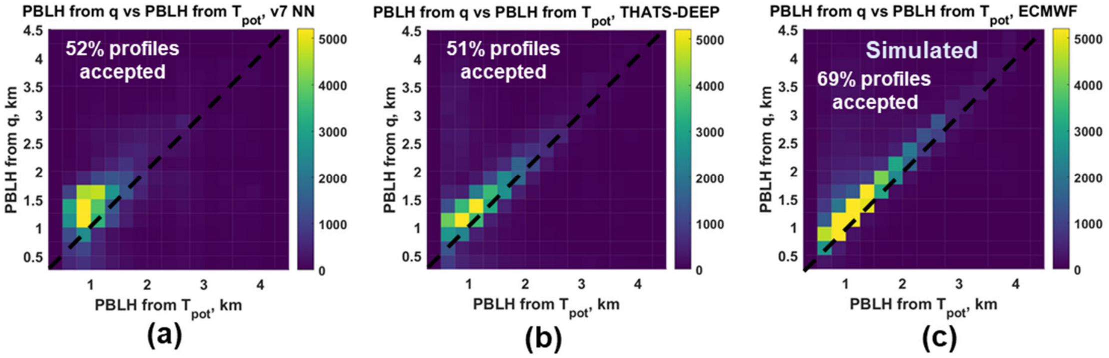

Figure 19: 2D histogram for PBLH derived from q versus PBLH derived from T_{pot}, for (a) v7 NN, (b) THATS-DEEP (with real v7 NN inputs), and (c) ECWMF profiles

To quantify the improvement in PBLH derived from q, we computed both the median and mean absolute error in PBLH with respect to ECMWF as a function of altitude for both v7NN and THATS-DEEP. The median error is plotted in Figure 20 (a), and shows that the overall bias in PBLH is reduced by THATS-DEEP at almost all altitudes by roughly a factor of 2. The mean absolute error (MAE), a measure of the spread, in PBLH is plotted in Figure 20 (b), showing reduction in MAE from THATS-DEEP for altitudes above 1.7 km, though the MAE remains similar or slightly increased versus v7NN below that altitude.

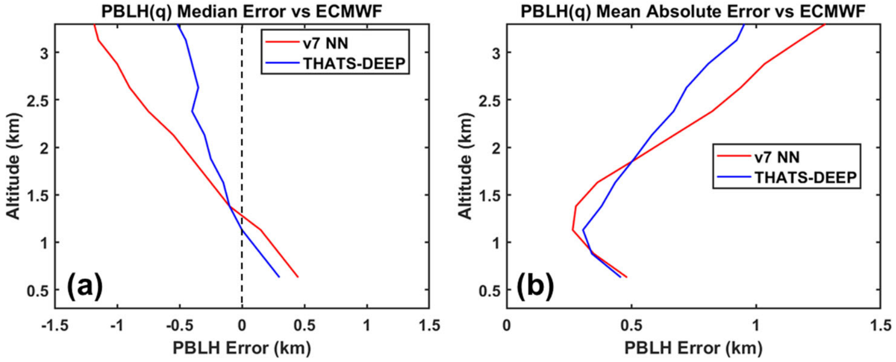

Figure 20: Quantitative PBLH results (from the gradient of *q*), including (a) median PBLH error versus ECMWF and (b) mean absolute error of PBLH versus ECWMF, for both v7NN and THATS-DEEP, as a function of altitude

4) Conclusion

We have presented a new DNN approach for enhancing detail and reducing noise in 3D granules containing current state of the art retrievals of temperature and humidity, and we have shown that this approach improves their scientific utility, including representation of key PBL features over land such as inversions representing PBLH, with Figure 20 a) showing PBLH median errors reduced by a factor of 2 at most PBLH altitudes. This technique, while demonstrated here on the v7 NN retrievals, can in principle be applied to enhance the scientific return from a wide variety of sensors and retrieval algorithms (for example, those presented by Smith and Barnett (Smith & Barnet, 2020) and by Irion (Irion et al., 2018)) provided it is trained accordingly. While we apply this technique as a standalone enhancement step to existing Level 2 retrieval granules, we envision that this could be used in recently proposed frameworks (for example, (Buzzard, Chan, Sreehari, & Bouman, 2017)) which balance AI and physics, extending and drawing upon classic optimal estimation approaches familiar to the remote sensing community. Such techniques are expected to benefit from recent progress in uncertainty estimation for remote sensing retrievals (for example, (Braverman, Hobbs, Teixeira, & Gunson, 2021)), including a neural network error prediction technique (Tao, Blackwell, & Staelin, 2013) which we are testing for v7 NN retrievals. We also expect that further improvements in accuracy and detail may be possible, including DNN architecture improvements and training improvements. For example, any improvements to the fidelity of

Sim-NN as a realistic model for real retrievals will improve THATS-DEEP accuracy. Future work will address more validation versus in-situ data sources in addition to other remote sensing modalities (active and passive) and reanalysis.

Once trained, THATS-DEEP can in principle execute in near-real time on operational retrievals, enhancing their utility to the forecast and Earth science community. A key consideration of the current work is the dependence on an existing operational product. For example, any revision to the v7 NN product would potentially require retraining of both Sim-NN and THATS-DEEP. In practice, this is mitigated by the fact that significant AIRS/AMSU version upgrades are relatively infrequent and have typically involved reprocessing over the life of the mission to avoid data discontinuities, resulting in a long, stable data record from which to train the subsequent steps in this work. However, in principle, the need to retrain Sim-NN and THATS-DEEP if the source data is updated may be inconvenient. In addition, THATS-DEEP, as currently implemented, likely requires retraining for retrievals derived from different sensors, even if they are similar hyperspectral sounder missions. Future work will consider ways to address this, such as transfer learning and more general-purpose, less sensor-specific style transfer techniques which might replace Sim-NN.

Acknowledgments.

DISTRIBUTION STATEMENT A. Approved for public release. Distribution is unlimited. This material is based upon work supported by the National Aeronautics and Space Administration under Air Force Contract No. FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Aeronautics and Space Administration.

We gratefully acknowledge support from NASA's Earth Science research program as part of "The Science of Terra, Aqua, and Suomi NPP" element. We are also grateful for support from Dr. Tsengdar Lee, and helpful discussions with colleagues including Dr. Antonia Gambacorta and the members of NASA's Sounder Science Team. We also thank the anonymous reviewers for very helpful comments which improved the submitted manuscript.

Data Availability Statement.

The AIRS data used in this study are all openly available for download from the NASA Goddard Earth Sciences Data Information and Services Center (GESDISC) and can be accessed on NASA's Earth Data Gateway (AIRS, 2019). Specifically, the v7 NN fields are available as part of the AIRS/AMSU v7 Level 2 Support Product, named “AIRX2SUP” by the AIRS project, which can be found at DOI: 10.5067/TZ6I8E3ODIQB.

The ERA5 pressure-level data (Hersbach et al., 2018) used in this study are all openly available for download from ECMWF’s Copernicus Climate Change Service (C3S) Climate Data Store (CDS), at DOI: 10.24381/cds.bd0915c6.

As part of this work, we created a large dataset file in Matlab’s “MAT” format containing collocated Sim-NN, v7 NN, and ERA5 data, for training, validation, and testing of THATS-DEEP. This dataset is too large to be retained or publicly archived with available resources, but will be provided on request to anyone who requests it from the corresponding author, along with the trained THATS-DEEP model.

REFERENCES

Aires, F., Boucher, E., & Pellet, V. (2021). Convolutional neural networks for satellite remote sensing at coarse resolution. Application for the SST retrieval using IASI. *Remote Sensing of Environment, 263*, 112553. doi:https://doi.org/10.1016/j.rse.2021.112553

Aires, F., Rossow, W. B., Scott, N. A., & Chédin, A. (2002). Remote sensing from the infrared atmospheric sounding interferometer instrument 2. Simultaneous retrieval of temperature, water vapor, and ozone atmospheric profiles. *Journal of Geophysical Research: Atmospheres, 107*(D22), ACH 7-1-ACH 7-12. doi:https://doi.org/10.1029/2001JD001591

AIRS. (2019). Aqua/AIRS L2 Support Retrieval (AIRS+AMSU) V7.0. Retrieved from https://disc.gsfc.nasa.gov/datasets/AIRX2SUP_7.0/summary doi:10.5067/TZ6I8E3ODIQB

Beal, M. H., Hagan, M. T., & B., D. H. (2022). *Deep Learning Toolbox™ User's Guide*. Natick, MA: The MathWorks, Inc.

Blackwell, W. J. (2005). A neural-network technique for the retrieval of atmospheric temperature and moisture profiles from high spectral resolution sounding data. *Geoscience and Remote Sensing, IEEE Transactions on, 43*(11), 2535-2546. doi:10.1109/tgrs.2005.855071

Blackwell, W. J., & Chen, F. W. (2005). Neural network applications in high-resolution atmospheric remote sensing. *Lincoln Laboratory Journal, 15*(2), 299-322.

Blackwell, W. J., & Chen, F. W. (2009). *Neural Networks in Atmospheric Remote Sensing*: Artech House, Incorporated.

Blackwell, W. J., & Milstein, A. B. (2014). A Neural Network Retrieval Technique for High-Resolution Profiling of Cloudy Atmospheres. *7*(4), 1260-1270. doi:10.1109/jstars.2014.2304701

Braverman, A., Hobbs, J., Teixeira, J., & Gunson, M. (2021). Post hoc Uncertainty Quantification for Remote Sensing Observing Systems. *SIAM/ASA Journal on Uncertainty Quantification, 9*(3), 1064-1093. doi:10.1137/19m1304283

Buzzard, G. T., Chan, S., Sreehari, S., & Bouman, C. A. (2017). Plug-and-Play Unplugged: Optimization-Free Reconstruction Using Consensus Equilibrium. *SIAM J. Imaging Sci., 11*(3), 2001-2020.

Byun, C., Arcand, W., Bestor, D., Bergeron, B., Hubbell, M., Kepner, J., . . . Yee, C. (2012, 10-12 Sept. 2012). *Driving big data with big compute.* Paper presented at the 2012 IEEE Conference on High Performance Extreme Computing.

Cai, X., Bao, Y., Petropoulos, G. P., Lu, F., Lu, Q., Zhu, L., & Wu, Y. (2020). Temperature and Humidity Profile Retrieval from FY4-GIIRS Hyperspectral Data Using Artificial Neural Networks. *Remote Sensing, 12*(11), 1872.

Chase, R. J., Nesbitt, S. W., & McFarquhar, G. M. (2021). A Dual-Frequency Radar Retrieval of Two Parameters of the Snowfall Particle Size Distribution Using a Neural Network. *Journal of Applied Meteorology and Climatology, 60*(3), 341-359. doi:10.1175/jamc-d-20-0177.1

Cho, C., & Staelin, D. H. (2006). Cloud clearing of Atmospheric Infrared Sounder hyperspectral infrared radiances using stochastic methods. *111*(D9), 10 pp. doi:10.1029/2005jd006013

DeSouza-Machado, S., Strow, L. L., Tangborn, A., Huang, X., Chen, X., Liu, X., . . . Yang, Q. (2018). Single-footprint retrievals for AIRS using a fast TwoSlab cloud-representation model and the SARTA all-sky infrared radiative transfer algorithm. *Atmos. Meas. Tech., 11*(1), 529-550. doi:10.5194/amt-11-529-2018

Ding, F., Iredell, L., Theobald, M., Wei, J., & Meyer, D. (2021). PBL Height From AIRS, GPS RO, and MERRA-2 Products in NASA GES DISC and Their 10-Year Seasonal Mean Intercomparison. *Earth and Space Science, 8*(9), e2021EA001859. doi:https://doi.org/10.1029/2021EA001859

ECMWF. (2014). IFS Documentation CY40R1 - Part I: Observation Processing. In *IFS Documentation CY40R1*: ECMWF.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. In. Retrieved from https://www.deeplearningbook.org/

Guo, J., Zhang, J., Yang, K., Liao, H., Zhang, S., Huang, K., . . . Xu, X. (2021). Investigation of near-global daytime boundary layer height using high-resolution radiosondes: first results and comparison with ERA5, MERRA-2, JRA-55, and NCEP-2 reanalyses. *Atmos. Chem. Phys., 21*(22), 17079-17097. doi:10.5194/acp-21-17079-2021

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., . . . Thépaut, J.-N. (2018). ERA5 hourly data on pressure levels from 1979 to present. Retrieved from https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-pressure-levels?tab=overview doi:10.24381/cds.bd0915c6

Hornik, K. (1991). Approximation capabilities of multilayer feedforward networks. *Neural Networks, 4*(2), 251-257. doi:http://dx.doi.org/10.1016/0893-6080(91)90009-T

Irion, F. W., Kahn, B. H., Schreier, M. M., Fetzer, E. J., Fishbein, E., Fu, D., . . . Yue, Q. (2018). Single-footprint retrievals of temperature, water vapor and cloud properties from AIRS. *Atmos. Meas. Tech., 11*(2), 971-995. doi:10.5194/amt-11-971-2018

Kim, J., Lee, J. K., & Lee, K. M. (2016, 27-30 June 2016). *Accurate Image Super-Resolution Using Very Deep Convolutional Networks.* Paper presented at the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

Kingma, D., & Ba, J. (2014). Adam: A Method for Stochastic Optimization. Retrieved from https://arxiv.org/abs/1412.6980

Li, J., Paul Menzel, W., Schmit, T. J., & Schmetz, J. (2022). Applications of geostationary hyperspectral infrared sounder observations – progress, challenges, and future perspectives. *Bulletin of the American Meteorological Society*. doi:10.1175/bams-d-21-0328.1

Liu, H., & Li, J. (2010). An Improvement in Forecasting Rapid Intensification of Typhoon Sinlaku (2008) Using Clear-Sky Full Spatial Resolution Advanced IR Soundings. *Journal of Applied Meteorology and Climatology, 49*(4), 821-827. doi:10.1175/2009jamc2374.1

Lu, Z., & Chen, Y. (2019). Single Image Super Resolution based on a Modified U-net with Mixed Gradient Loss. Retrieved from https://arxiv.org/abs/1911.09428

Maddy, E. S., & Boukabara, S. A. (2021). MIIDAPS-AI: An Explainable Machine-Learning Algorithm for Infrared and Microwave Remote Sensing and Data Assimilation Preprocessing - Application to LEO and GEO Sensors. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14*, 8566-8576. doi:10.1109/JSTARS.2021.3104389

Martins, J. P. A., Teixeira, J., Soares, P. M. M., Miranda, P. M. A., Kahn, B. H., Dang, V. T., . . . Fishbein, E. (2010). Infrared sounding of the trade-wind boundary layer: AIRS and the RICO experiment. *Geophysical Research Letters, 37*(24). doi:https://doi.org/10.1029/2010GL045902

Milstein, A. B., & Blackwell, W. J. (2016). Neural network temperature and moisture retrieval algorithm validation for AIRS/AMSU and CrIS/ATMS. *Journal of Geophysical Research: Atmospheres*.

National Academies of Sciences, E., & Medicine. (2018). *Thriving on Our Changing Planet: A Decadal Strategy for Earth Observation from Space*. Washington, DC: The National Academies Press.

Pangaud, T., Fourrie, N., Guidard, V., Dahoui, M., & Rabier, F. (2009). Assimilation of AIRS Radiances Affected by Mid- to Low-Level Clouds. *Monthly Weather Review, 137*(12), 4276-4292. doi:10.1175/2009mwr3020.1

Planetary Boundary Layer Decadal Survey Incubation Study Team. Retrieved from https://science.nasa.gov/earth-science/decadal-pbl

Reuther, A., Kepner, J., Byun, C., Samsi, S., Arcand, W., Bestor, D., . . . Michaleas, P. (2018, 25-27 Sept. 2018). *Interactive Supercomputing on 40,000 Cores for Machine Learning and Data Analysis.* Paper presented at the 2018 IEEE High Performance extreme Computing Conference (HPEC).

Rodgers, C. D. (2000). *Inverse Methods for Atmospheric Sounding: Theory and Practice*: World Scientific Publishing Company, Incorporated.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. Retrieved from https://arxiv.org/abs/1505.04597

Samsi, S., Mattioli, C. J., & Veillette, M. S. (2019, 24-26 Sept. 2019). *Distributed Deep Learning for Precipitation Nowcasting.* Paper presented at the 2019 IEEE High Performance Extreme Computing Conference (HPEC).

Seidel, D. J., Ao, C. O., & Li, K. (2010). Estimating climatological planetary boundary layer heights from radiosonde observations: Comparison of methods and uncertainty analysis. *Journal of Geophysical Research: Atmospheres, 115*(D16). doi:https://doi.org/10.1029/2009JD013680

Smith, N., & Barnet, C. D. (2020). CLIMCAPS observing capability for temperature, moisture, and trace gases from AIRS/AMSU and CrIS/ATMS. *Atmos. Meas. Tech., 13*(8), 4437-4459. doi:10.5194/amt-13-4437-2020

Susskind, J., Blaisdell, J., & Iredell, L. (2014). Improved methodology for surface and atmospheric soundings, error estimates, and quality control procedures: the atmospheric infrared sounder science team version-6 retrieval algorithm. *Journal of Applied Remote Sensing, 8*(1), 084994.

Susskind, J., Blaisdell, J., Iredell, L., Lee, J., Milstein, A., Barnet, C., . . . Ruzmaikin, A. (2020). *AIRS-Team Retrieval For Core Products and Geophysical Parameters: Versions 6 and 7 Level 2*. Retrieved from https://docserver.gesdisc.eosdis.nasa.gov/public/project/AIRS/L2_ATBD.pdf

Susskind, J., Lee, J., & Iredell, L. (2014, September). *Version 6 Surface Skin/Surface Air Temperature Differences and their Interannual Variability.* Paper presented at the NASA Sounder Science Team Meeting, Greenbelt, MD.

Tao, Z., Blackwell, W. J., & Staelin, D. H. (2013). Error Variance Estimation for Individual Geophysical Parameter Retrievals. *51*(3), 1718-1727. doi:10.1109/tgrs.2012.2207728

Teixeira, J., Piepmeier, J. R., Nehrir, A. R., Ao, C. O., Chen, S. S., Clayson, C. A., . . . Zeng, X. (2021). *Toward a Global Planetary Boundary Layer Observing System: The NASA PBL Incubation Study Team Report*. Retrieved from https://science.nasa.gov/files/science-red/s3fs-public/atoms/files/NASA_PBL_Incubation_Final_Report_2.pdf

Veillette, M., Samsi, S., & Mattioli, C. (2020). *SEVIR : A Storm Event Imagery Dataset for Deep Learning Applications in Radar and Satellite Meteorology.* Paper presented at the Advances in Neural Information Processing Systems 33 (NeurIPS 2020), Vancouver, Canada.

Wong, S., Yue, Q., Milstein, A., Blaisdell, J., Manning, E., Fetzer, E., . . . Teixeira, J. (2018, April). *Validation of pre-V7 AIRS and neural network temperature in the PBL against the IGRA radiosondes.* Paper presented at the Sounder Science Team Meeting, Pasadena, CA.

Yue, Q., Lambrigtsen, B., Blaisdell, J. M., Farahmand, A., Fetzer, E. J., Fishbein, E., . . . Wong, S. (2020). *AIRS Version 7 Level 2 Performance Test and Validation Report*. Retrieved from https://docserver.gesdisc.eosdis.nasa.gov/public/project/AIRS/V7_L2_Performance_Test_and_Validation_report.pdf

Zhu, J.-Y., Park, T., Isola, P., & Efros, A. (2017). *Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks*. Paper presented at the IEEE International Conference on Computer Vision (ICCV).