

Article

Fast and Effective Techniques for LWIR Radiative Transfer Modeling: A Dimension Reduction Approach

Nicholas Westing^{1,†,*} , Brett Borghetti^{1,‡}  and Kevin C. Gross^{2,‡}

¹ Department of Electrical and Computer Engineering, Air Force Institute of Technology, WPAFB OH 45433

² Department of Engineering Physics, Air Force Institute of Technology, WPAFB OH 45433

* Correspondence: nicholas.westing@afit.edu

† Current address: Air Force Institute of Technology, WPAFB OH 45433

‡ These authors contributed equally to this work.

Abstract: The increasing spatial and spectral resolution of hyperspectral imagers yields detailed spectroscopy measurements from both space-based and airborne platforms. Machine learning algorithms have achieved state-of-the-art material classification performance on benchmark hyperspectral data sets; however, these techniques often do not consider varying atmospheric conditions experienced in a real-world detection scenario. To reduce the impact of atmospheric effects in the at-sensor signal, atmospheric compensation must be performed. Radiative Transfer (RT) modeling can generate high-fidelity atmospheric estimates at detailed spectral resolutions, but is often too time-consuming for real-time detection scenarios. This research utilizes machine learning methods to perform dimension reduction on the transmittance, upwelling radiance, and downwelling radiance (TUD) data to create high accuracy atmospheric estimates with lower computational cost than RT modeling. The utility of this approach is investigated using the instrument line shape for the Mako long-wave infrared hyperspectral sensor. This study employs physics-based metrics and loss functions to identify promising dimension reduction techniques. As a result, TUD vectors can be produced in real-time allowing for atmospheric compensation across diverse remote sensing scenarios.

Keywords: Hyperspectral Imagery; Machine Learning; Atmospheric Compensation; Autoencoders; Radiative Transfer Modeling

1. Introduction

Next generation hyperspectral imagers continue to improve in both spatial and spectral resolution with increasingly lower noise-equivalent spectral radiance (NESR) values, presenting unique challenges and opportunities in efficiently characterizing pixel materials [1]. A pixel in a hyperspectral image can be represented as a vector across all spectral channels, producing a three dimensional data cube for an entire image, width by height by spectral channel [2]. Hyperspectral imagers have been deployed in both airborne and space-based platforms with uses ranging from precision agriculture to search and rescue operations [3]. The spectral bands making up a hyperspectral cube can span from the visible to the longwave infrared (LWIR), sampled across hundreds of narrow spectral channels [4]. The visible to short-wave infrared is dominated by scattering while the LWIR is dominated by material emission [5]. The atmospheric state — which includes the altitude-dependent temperature and pressure; how water, carbon dioxide, ozone, and other trace gases are distributed vertically; the kind and size distributions of various aerosols — has a significant impact on the at-sensor radiance. Understanding and accounting for these atmospheric effects is critical for quantitative exploitation of hyperspectral imagery, especially in the domain of material identification.

Considerable research aimed at improving material identification and classification in hyperspectral data has been conducted in recent years. Machine learning and deep learning approaches

have proved advantageous for both high accuracy results and low latency predictions [6,7]. While machine learning has long been a useful tool for hyperspectral image exploitation, the number of papers applying machine learning tools, in particular deep artificial neural networks, has surged in recent years. A significant limitation with many supervised classification approaches is a lack of labeled data to fully represent a data space that is complex, nonlinear, and high-dimensional. Since many bands in hyperspectral data are often highly correlated, dimension reductions techniques such as [Principal Component Analysis \(PCA\)](#) have been used to transform hyperspectral data to a small number of uncorrelated features [8]. [Support Vector Machine \(SVM\)](#) classifiers have proven useful for hyperspectral data because of their ability to project the sparse, high-dimensional data using a kernel function for high accuracy classification [9].

Scattering and absorption by atmospheric constituents between the sensor and target affect the at-sensor radiance in ways that vary greatly with wavelength or frequency. Since this creates highly nonlinear relationships in the data, deep learning models have also been studied for extracting meaningful features. One of the first deep learning approaches applied to hyperspectral classification was the work performed by Chen *et al.* where [Stacked Autoencoders \(SAEs\)](#) were trained to create a useful feature representation for pixel classification [10]. This work was followed by studies utilizing Deep Belief Networks and Restricted Boltzman Machines to further improve classification accuracy on benchmark hyperspectral data sets [11]. Based on successes in computer vision tasks [12], convolutional neural networks have also been applied to hyperspectral data as another method for creating useful feature representations. These approaches include one-dimensional convolutions performed on the spectral axis of the data cube, combinations of two-dimensional convolutions on the spatial axis and three-dimensional convolutions performed on the data cube [11,13,14]. Additionally, recurrent neural network approaches have been used, where the spectral data is considered sequentially to create useful feature representations [15,16].

In the majority of the classification research discussed thus far, the approach is typically trained and tested using benchmark hyperspectral data sets such as Kennedy Space Center, Indian Pines and Pavia University [7]. Models trained on these individual hyperspectral data cubes would likely show poor generalization performance as atmospheric conditions deviate from the conditions observed in the training data. For many real-world applications, atmospherically-robust methods must be developed. Towards this goal, this study focuses on efficient and accurate estimation of atmospheric parameters using neural network approaches for the [LWIR](#) domain. Specifically, we demonstrate dimension reduction techniques that are useful for emulating high-fidelity [RT](#) models supporting atmospheric compensation techniques.

The contributions and findings of this research include:

- Employing machine learning techniques which: (1) are computationally faster than traditional [RT](#) calculation methods; (2) reduce the dimension of both the [TUD](#) and atmospheric measurement vectors; (3) produce the desirable latent-space-similarity property such that small deviations in the low-dimension latent space result in small deviations in the high-dimension [TUD](#)
- Developing a data augmentation method using [PCA](#) and [Gaussian Mixture Models \(GMMs\)](#) on real atmospheric measurements that lead to improved model training and generalizability
- Improving machine learning model training by introducing a physics-based loss function which encourages better fit models than traditional loss functions based on mean squared error
- Demonstrating an effective [Autoencoder \(AE\)](#) pre-training strategy that leverages the local-similarity properties of the latent space to reproduce [TUDs](#) from atmospheric conditions

Together, these contributions form the basis of a novel method for efficient and effective [RT](#) modeling, using a small number of parameters, which is applicable for atmospheric compensation across diverse remote sensing scenarios.

1.1. Background

Atmospheric compensation techniques estimate the atmospheric effects imposed on the at-sensor signal, leading to atmospherically-corrected data for material classification and identification. In the **LWIR**, the simplest **RT** model for describing the at-sensor radiance, $L(\lambda)$, from a diffuse (lambertian) thermal emitter and scatterer, can be expressed as:

$$L(\lambda) = \tau(\lambda) \left[\epsilon(\lambda) B(\lambda, T) + [1 - \epsilon(\lambda)] L_d(\lambda) \right] + L_a(\lambda) \quad (1)$$

λ : wavelength

T : material temperature

$\tau(\lambda)$: atmospheric transmission

$\epsilon(\lambda)$: material emissivity

$B(\lambda, T)$: Planckian distribution

$L_d(\lambda)$: downwelling atmospheric radiance

$L_a(\lambda)$: atmospheric path (upwelling) radiance

Both τ and L_a are specific to the line of sight between the sensor and the surface, whereas L_d represents a cosine-weighted average of the downwelling radiance for the hemisphere above the surface. Planck's blackbody distribution function, $B(\lambda, T)$, is given by

$$B(\lambda, T) = \frac{2hc^2}{\lambda^5} \frac{1}{e^{hc/\lambda kT} - 1} \quad (2)$$

where k is Boltzmann's constant, c is the speed of light and h is Planck's constant. Atmospheric compensation recovers the surface leaving radiance $L_s(\lambda)$ by estimating $\tau(\lambda)$ and $L_a(\lambda)$ in Equation 1 as shown in Equation 3.

$$L_s(\lambda) = \epsilon(\lambda) B(\lambda, T) + [1 - \epsilon(\lambda)] L_d(\lambda) \quad (3)$$

One of the most popular **LWIR** atmospheric compensation techniques is the **In-Scene Atmospheric Compensation (ISAC)** method which first identifies blackbody pixels within the scene to estimate $\tau(\lambda)$ and $L_a(\lambda)$ [17]. By using only pixel spectra from blackbodies, the surface leaving radiance is equivalent to Planck's blackbody distribution and the simplified **LWIR** at-sensor radiance can be expressed as

$$L(\lambda)_{\epsilon(\lambda) \rightarrow 1} = \tau(\lambda) B(\lambda, T) + L_a(\lambda). \quad (4)$$

Under the assumption that distinct blackbody pixels can be identified and their temperatures known, a linear fit can be performed across all spectral channels to identify $\tau(\lambda)$ and $L_a(\lambda)$. In practice, temperatures estimates are made in the most transmissive part of the at-sensor spectral radiance but they are often systematically biased since $\tau(\lambda')$ and $L_a(\lambda')$ are unknown and assumed to be 0 and 1, respectively, in that particular spectral channel λ' . A common method to remove the biases introduced into τ and L_a by inaccurate surface temperatures relies on spectral analysis near the isolated water absorption feature near 11.73 μm . This method is very similar to the **Autonomous Atmospheric Compensation (AAC)** method, which estimates a transmittance ratio and an upwelling radiance parameter derived from the off- and on-resonance spectral values at the same isolated water band [18]. By assessing this water feature, both transmittance and path radiance contributions can be independently estimated, allowing the biased ISAC estimates of τ and L_a to be fixed, or under further assumptions about the atmosphere, allowing full estimates of τ and L_a to be made. To ensure algorithmic efficiency for both **ISAC** and **AAC**, pre-computed look-up tables of $\tau(\lambda)$ and $L_a(\lambda)$ are forward modeled with a **RT** model over a wide range of possible atmospheric water and temperature

profiles. In the **LWIR** portion of the spectrum, **Temperature-Emissivity Separation (TES)** follows atmospheric compensation to estimate material emissivity and surface temperature from $L_s(\lambda)$ [19].

Model-based atmospheric compensation methods use **RT** models such as **MODTRAN** or **Line-by-Line Radiative Transfer Model (LBLRTM)** to simulate scattering and absorption by atmospheric constituents to support surface leaving radiance retrievals [5,20–22]. **RT** calculations produce high-fidelity **TUD** estimates constrained by physics, but come at a computational cost often precluding their use for real-time atmospheric compensation. **RT** models divide the atmosphere into thin, homogeneous layers, conducting high spectral resolution **RT** calculations (e.g., **LBLRTM**), or approximations thereof (e.g., **MODTRAN**). Due to the large number of discrete absorption lines of the many trace gases in the atmosphere, millions of calculations are required to model a sensor's entire spectral range with high fidelity [23]. While **RT** models require estimates of gas and aerosol concentrations as a function of altitude, **RT**-based atmospheric correction methods often apply a variable scale factor to standard vertical profiles of gases such as H_2O , CO_2 , and O_3 . To avoid the high computational cost of line-by-line **RT** calculations, approximate **RT** models are used to increase computational speed while trading off accuracy [24]. One of the most widely used approaches to improve **RT** computation time is the correlated-k method, which divides opaque spectrum into a subset of b bands and then applies a weighting k to these bands, dependent on the opacity distribution of the b bands [25]. Similar to the weighting scheme employed in the correlated-k method, **PCA** has also been implemented to reduce **RT** computation time [23]. In [23], **PCA** was performed on a database of **TUD** vectors identifying a lower dimensional space of only a few hundred components compared to the thousands of dimensions in the original data space. By sampling the lower dimensional representation correctly, **RT** calculations could be reduced significantly. These approximate **RT** models can result in systematic errors, leading to residual atmospheric features which complicate material identification. Challenging hyperspectral imaging use-cases, such as search and rescue missions, require accurate, robust, and real-time material identification, justifying continued effort in developing **RT** models that are both highly accurate and computationally efficient.

In this study, we conduct dimension reduction on the **TUD** vectors ($\tau(\lambda)$, $L_a(\lambda)$, $L_d(\lambda)$) in Equation 1 which span a wide range of global atmospheric variability. Specifically, this research uses the **Thermodynamic Initial Guess Retrieval (TIGR)** database comprising myriad and diverse atmospheric conditions in the form of temperature, water vapor, and ozone profiles on a fixed pressure grid. The 2311 atmospheric profiles provided in the **TIGR** database are based on 80,000 radiosonde measurements collected worldwide [26,27]. The **TIGR** atmospheric profiles are first filtered for cloud-free conditions and then forward modeled using the **LBLRTM** (version 12.8) to create realistic **TUD** vectors which also span the majority of expected variability within **TUD** vector space.

Conducting dimension reduction on the **TIGR**-derived **TUD** vectors creates a low-dimensional representation that can be sampled to create new **TUD** vectors without the need of costly **RT** calculations. Research performed in [28], specifically considered a low-rank subspace of $\tau(\lambda)$ and $L_a(\lambda)$ for atmospheric compensation in the **LWIR** spectrum. They performed a singular value decomposition on representative $\tau(\lambda)$ and $L_a(\lambda)$ vectors generated by **MODTRAN** for a given seasonal model and flight altitude. The approach taken in [28] is similar to the work performed in this study, however, we also consider the downwelling radiance, $L_d(\lambda)$, and investigate the utility of neural network-based dimension reduction using **TUD** vectors derived from real atmospheric measurements such as the **TIGR** database.

Creating a low-dimensional **TUD** representation is also important for data augmentation applications, distinctly different from atmospheric compensation. To identify a material of interest, many augmented representations of that material through diverse atmospheric conditions can be created using a low-dimensional **TUD** representation. By providing many of the commonly investigated classification techniques augmented at-sensor data representative of diverse and realistic **TUD** vectors, atmospherically-robust classification can be improved. This was the approach employed in [29], where a small neural network was trained to detect specific materials in the **LWIR** across

varying atmospheric conditions. The neural network-based approaches investigated here offer a highly efficient method to generate realistic **TUD** vectors to support data augmentation.

Additionally, the utility of the low-dimensional representation is explored by mapping temperature, water vapor and ozone profiles to the low-dimensional space, thus creating a highly-efficient and accurate **RT** model. This **RT** model can be used for data augmentation as discussed above or to support model-based compensation techniques where hundreds of possible transmittance and upwelling vectors can be computed in real-time, avoiding the use of precomputed look-up tables. By performing dimension reduction prior to fitting this mapping, similar atmospheric conditions cluster together in the low-dimensional space. Additionally, based on this clustering, small deviations in the low-dimensional space correspond to small changes in generated **TUD** vectors, further improving the mapping from atmospheric measurements to the low-dimensional space.

In the next section dimension reduction techniques are reviewed and the **TIGR** dataset is explained in further detail. Metrics based on Equation 1 are also derived to ensure dimension reduction techniques are correctly evaluated. Sampling of the low-dimensional **TUD** representation is also outlined to demonstrate the utility of these techniques and highlight the importance of latent-space-similarity where deviations in the latent space correspond to similar deviations in high-dimension **TUD** space.

Following the methodology section, results are presented comparing dimension reduction performance derived from the **TIGR** data and an augmented version of the **TIGR** data. After confirming improved performance with the augmented data, a novel physics-based loss function is compared to **mean squared error (MSE)** to further improve dimension reduction reconstruction error. Finally, a **RT** model is formulated from the dimension reduction algorithms, showing the importance of the dimension reduction pre-training step toward reduced **TUD** prediction error.

2. Methodology

This section first reviews the atmospheric measurement data and corresponding forward modeled **TUD** vectors used for dimension reduction. Metrics for comparing the performance of each technique are also reviewed, with a focus on incorporating properties from the simplified **RT** model in Equation 1. A unique data augmentation scheme is also discussed to increase the number of **TUD** samples for model fitting.

2.1. Data

The **TIGR** database consists of 2311 atmospheres selected from over 80,000 worldwide radiosonde reports. These atmospheric conditions encompass the expected variability of temperature, water content and ozone at 43 discrete pressure levels ranging from the earth's surface (1013 hPa) to > 30 km (<1hPa) [26,27]. Cubic interpolation was used to upsample these profiles to 66 pressure levels, with finer sampling in the lower, most dense part of the atmosphere. Additionally, the profiles are grouped by air mass category such as polar, tropical and mid-latitude. The entire **TIGR** data matrix shape is 2311 atmospheric profiles by 198 measurements, where the 66 pressure level measurements for temperature, water content and ozone are concatenated.

By using atmospheric profiles that span nearly all expected atmospheric variability, **RT** can be conducted to generate **TUD** vectors encapsulating nearly every possible atmospheric scenario. The **LBLRTM** was used to create high resolution **TUD** vectors, however, the spectral resolution must be downsampled for a particular sensor to ensure the dimension reduction techniques are applicable to real-world sensor resolutions.

The **LWIR** Mako hyperspectral sensor is a high-performance, airborne sensor imaging across 7.8 – 13.4 μm into 128 spectral channels with a noise-equivalent temperature difference of 0.02K at 10 μm and 300 K [1,30]. The high-resolution **LBLRTM** generated **TUD** vectors (11,513 spectral channels) are downsampled according to the Mako instrument line shape creating representative **TUD** vectors for this sensor. Additionally, the **TUD** vectors are generated to represent a sensor altitude of 3.3 km.

The result of this process is shown in Figure 1, where after downsampling, the TUD data matrix shape is 2311 samples by 384 spectral measurements ($\tau(\lambda)$, $L_a(\lambda)$, $L_d(\lambda)$ concatenated). The goal of the dimension reduction algorithms discussed next is to project the length 384 TUD vectors to a lower dimensional space such that reconstruction error is minimized.

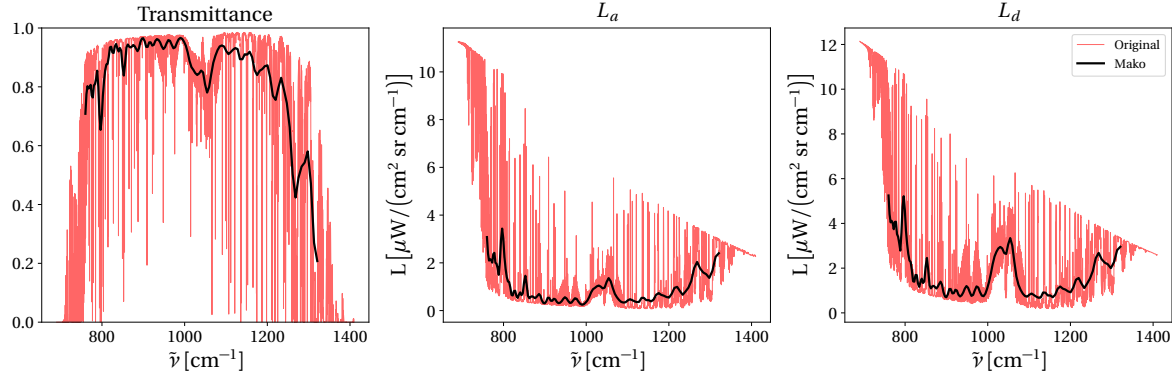


Figure 1. The high-resolution LBLRTM transmittance, upwelling and downwelling vectors are shown with their downsampled counterparts for the Mako LWIR sensor. The downsampled vectors are the data used in the remainder of this study.

2.2. TUD Dimension Reduction Techniques

PCA removes correlation from data by projecting it onto a new coordinate system which maximizes data variance. The reduced dimensions are the eigenvectors and are orthogonal to one another. The eigendecomposition is performed on the data covariance matrix C_x such that

$$C_x = ADA^T \quad (5)$$

where the matrix A is an orthonormal matrix consisting of N eigenvectors and D is a diagonal matrix consisting of the corresponding eigenvalues [31]. The eigenvalues are sorted in descending order and the eigenvectors corresponding to the largest eigenvalues are kept. This selection is based on the cumulative sum of data variance explained with the eigenvectors. The smallest eigenvalue components are considered noise and have little impact on reconstructing the original signal. The subset of selected eigenvectors are used to linearly transform the data to the low-dimensional representation y by:

$$y_i = A^T x_i.$$

An AE is a neural network designed for performing nonlinear compression by projecting data to a low-dimensional latent space, followed by nonlinear reconstruction from the latent space. An AE is composed of two networks to perform this operation: an encoder network and a decoder network. The encoder compresses the input data, x , into a lower dimensional latent space, z , and the decoder reconstructs the data based on the latent space mapping into y [32]. Equations for these two transformations are

$$\begin{aligned} z &= f(W_z x + b_z) \\ y &= f(W_y z + b_y) \end{aligned} \quad (6)$$

where $x \in R^d$ is the input data, $z \in R^l$ is the latent space representation with $l \ll d$. The reconstructed data is $y \in R^d$ and b_y and b_z are the biases of the hidden and output layer layers respectively. W_z and W_y are the weight matrices from the input to hidden layer and hidden layer to output layer respectively. An AE diagram is shown in Figure 6 specifically for TUD compression and decompression using the encoder, decoder and latent space nomenclature. This figure is only notional, and does not represent the number of nodes actually used in this study.

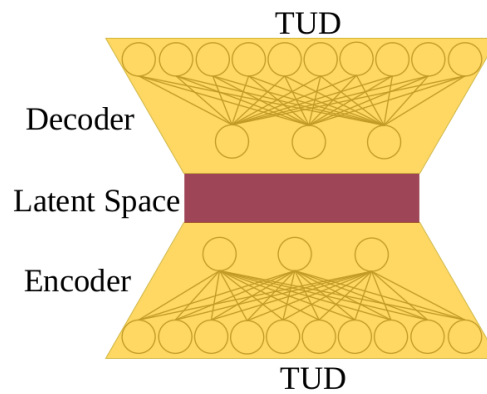


Figure 2. An example **AE** model where the **TUD** vectors are compressed through one or more encoder layers to a low-dimensional latent space. The decoder transforms the low-dimensional latent space back to the original **TUD** vector.

The **AE** predicted **TUD** vectors are compared to the **LBLRTM** generated **TUD** vectors through a loss function to determine model performance and update the weight matrices. The loss function used for measuring reconstruction error between **TUD** vectors is an important design variable influencing how the **AE** structures the latent space and ultimately what the network understands about **TUD** reconstruction. A commonly used loss function is **MSE**, calculated according to:

$$\text{MSE} = \frac{1}{K} \sum_{i=1}^{K+1} (\mathbf{x}_i - \mathbf{y}_i)^2 \quad (7)$$

where K equals the number of dimensions in the **TUD** vector, and \mathbf{x}_i and \mathbf{y}_i are the predicted and truth **TUD** vectors respectively. **MSE** will be used in this study, but an additional loss function will be derived later based on the underlying **LWIR RT** model. In many applications, a series of hidden layers are used to create a latent space representation of the data. This architecture is commonly referred to as a **SAE**, where the functions shown in Equation 6 are nested to include additional layers. The activation function, f , can be linear or non-linear. A comprehensive search of activation functions found the **Rectified Linear Unit (ReLU)** [33] yielded the best results over functions such as hyperbolic tangent, sigmoid, exponential linear units, and scaled exponential linear units. The **ReLU** function used in this study is

$$\text{ReLU}(x) = \begin{cases} x, & \text{if } x > 0 \\ \alpha x, & \text{if } x < 0 \end{cases}$$

where α controls how much information is passed through the network for negative inputs. This small slope increases information flow during backpropagation allowing more weights to be influenced by training samples [34].

The number of latent components is an important design parameter controlling model complexity and reconstruction performance. Using two latent components allows for visualizations of the latent space by overlaying measurement parameters such as surface temperature and total atmospheric water content. For the **PCA** model, the first two components capture 99.50% of the data variance. Plotting just these two components shows a smoothly varying relationship between the components and these physical parameters as shown in Figure 3.

Similar plots are shown in Figure 4 when considering a 2 component **SAE** model. Interestingly, the **SAE** disperses points throughout the latent space which is beneficial for sampling the low dimensional representation. Small changes in latent space components should result in small changes in the generated **TUD** vectors. This latent-space-similarity is an important property when fitting a sampling method to correctly identify a small number of components to generate a **TUD** vector. Small sampling

errors should not result in large TUD deviations. Since it's difficult to visualize higher dimension latent spaces, this property can be observed by fitting a small neural network to correctly predict the latent components for a known TUD.

The PCA model has little change in components 1 and 2 for cold, dry atmospheric conditions as shown by the tight clustering of these points. There appears to be a stronger dependence on component 2 for cold, dry atmospheres, while hot, humid conditions are dependent on both components. The SAE components are both influenced by surface temperature, while component 1 appears more dependent total water content. Based on the preliminary results shown in Figures 3 and 4, sampling the two-dimensional AE latent space will result in lower reconstruction error because of the tight clustering of cold dry atmospheric conditions shown in the PCA latent space. Using only two components results in large reconstruction error, therefore, we will consider additional latent components to cluster similar atmospheric conditions together in a low-dimensional space while minimizing reconstruction error.

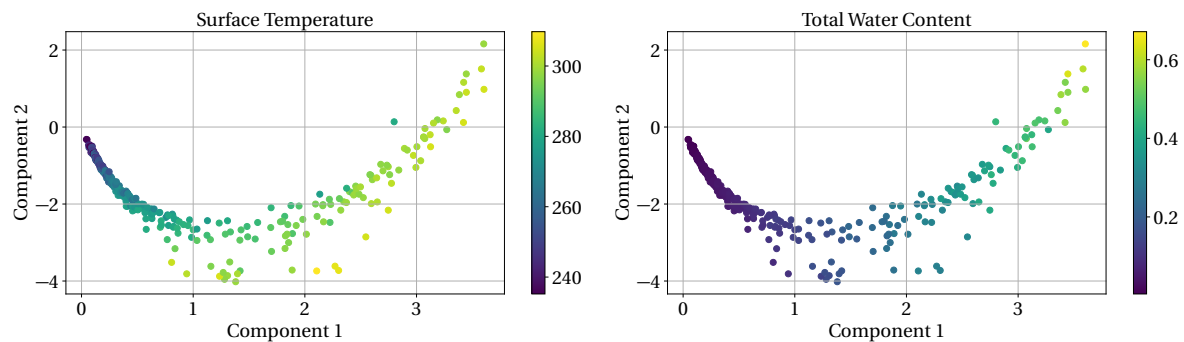


Figure 3. The first two principal components using the 6 component PCA model with the augmented TIGR data. Hot, humid atmospheric conditions vary with component 1 and 2 while cold, dry atmospheres are more dependent on component 2

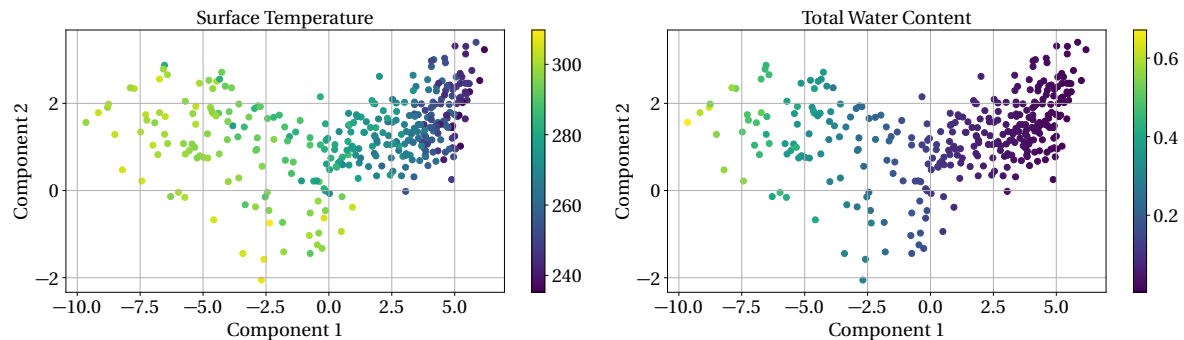


Figure 4. Autoencoder latent space when trained using 2 components. The points are scattered throughout the latent space with an overall clustering of similar atmospheric conditions. Both components appear dependent on surface temperature. Component 1 also appears more dependent on total water content versus component 2.

2.3. Metrics

For all methods considered, the reconstruction error must be placed in context of the at-sensor radiance to provide meaningful reconstruction performance. At-sensor radiance errors are dependent on the material emissivity as shown in Equation 4 where downwelling radiance does not play a role in the total error. However, if the surface material is reflective ($\epsilon(\lambda) = 0$), the simplified LWIR RT equation becomes

$$L(\lambda)_{\epsilon(\lambda) \rightarrow 0} = \tau(\lambda)L_d(\lambda) + L_a(\lambda) \quad (8)$$

where errors in $\tau(\lambda)$ and $L_d(\lambda)$ are now exaggerated. Using a standard metric, such as MSE, does not capture this dependence on material emissivity and provides misleading model performance for reflective versus emissive materials. A more appropriate metric for this domain considers the material emissivity in the at-sensor radiance error calculation.

For a test emissivity, $\epsilon_t(\lambda)$, the estimated at-sensor radiance, $\hat{L}(\lambda)$ is calculated based on the reconstructed **TUD** vector. Additionally, the original **TUD** vector is used in conjunction with $\epsilon_t(\lambda)$ to calculate the true at-sensor radiance $L(\lambda)$. The **root mean square error (RMSE)**, E_t , is calculated across all spectral channels such that

$$E_t = \sqrt{\frac{1}{K} \sum_{i=1}^K (L(\lambda_i) - \hat{L}(\lambda_i))^2} \quad (9)$$

where K represents the number of spectral channels and E_t is now in radiance units representing the emissivity dependent **RMSE**. For the **LWIR** domain, errors are typically expressed in terms of temperature where conversion of radiance to brightness temperature is defined as [3]:

$$T_{BB}(\lambda) = \frac{hc}{\lambda k \ln \left(\frac{2hc^2}{\lambda^5 I(\lambda)} + 1 \right)} \quad (10)$$

By transforming at-sensor radiance to brightness temperature, the at-sensor error between $\hat{L}(\lambda)$ and $L(\lambda)$ can now be expressed in Kelvin. In general, reconstruction performance improves as $\epsilon_t(\lambda)$ approaches 1.0 based on Equation 4. The actual emissivity values used are assumed grey bodies (spectrally flat) and linearly sampled between 0 and 1. Calculating the emissivity dependent **RMSE** provides additional information over standard **MSE** between predicted and truth **TUD** values. Model selection is performed based on performance across the entire emissivity domain resulting in lower error for reflective materials.

Constructing a model with low error across a range of emissivity values requires modifications to model training. As shown by the emissivity dependent **RMSE** metric, standard **MSE** will not provide sufficient information to properly update model weights. Instead, the loss function for training the **SAE** must include emissivity dependent information. The loss function must still be differentiable and result in stable training performance. To achieve this, we use the **TUD MSE** calculation for stabilized training, but also include an at-sensor radiance **MSE** dependent on material emissivity:

$$\mathcal{L}(\mathbf{x}, \mathbf{y}) = \frac{1}{K} \sum_{i=1}^{K+1} (\mathbf{x}_i - \mathbf{y}_i)^2 + \frac{\gamma}{MN} \sum_{j=1}^{M+1} \sum_{i=1}^{N+1} (L_{\mathbf{x}}(\lambda_i, \epsilon_j) - \hat{L}_{\mathbf{y}}(\lambda_i, \epsilon_j))^2 \quad (11)$$

where \mathbf{x} is the truth **TUD** vector and \mathbf{y} is the reconstructed **TUD** vector. The terms $L_{\mathbf{x}}(\lambda_i, \epsilon_j)$ and $\hat{L}_{\mathbf{y}}(\lambda_i, \epsilon_j)$ represent the at-sensor radiance using the truth and predicted **TUD** vectors respectively. The at-sensor radiance loss is calculated using a linear sampling of M emissivity values between 0 and 1, noted by ϵ_j in the loss calculation. A regularization term, γ , is included in Equation 11 to trade-off at-sensor radiance error and the **TUD** reconstruction error. In this study we only consider $\gamma = 1$, however, future work will consider this additional hyperparameter in the network optimization.

By including a loss component for the at-sensor radiance, network weights are updated to minimize at-sensor radiance error rather than strictly **TUD** reconstruction error. This is an important additional constraint since material emissivity impacts the difficulty of the reconstruction problem as shown by Equation 8. The **MSE** component in Equation 11 is necessary to stabilize training since errors in one component of the **TUD** vector can cause a reduction in overall loss depending on material emissivity. In practice, training networks without the **MSE** component caused large deviations in loss values as the network weights tried to simultaneously optimize for a range of emissivity values.

2.4. Radiative Transfer Modeling

We consider the utility of the low-dimensional TUD representation by applying it to the problem of RT modeling. A generative TUD model is created by considering only the latent space and decoder of the developed SAE model. Sampling the latent space, generates new TUD samples with similar characteristics for similar latent space samples (latent-space-similarity). To create a RT model, the TIGR atmospheric measurements are mapped to the generative model's low-dimensional input. In this study, a neural network is used to create this mapping. An overview of this process is shown in Figure 5, where the encoder and decoder can be replaced with their corresponding PCA transformations. After training the dimension reduction algorithm to compress TUD data through a low-dimensional latent space (first step in Figure 5), a sampling network is trained to correctly predict the latent space components without updating the decoder network as shown by the second step in Figure 5.

Once the sampling network weights have converged, the entire RT model is trained end-to-end, allowing small weight updates in both the sampling network and the decoder. Only a small number of training iterations are needed for the final training step as network weights are nearly optimized for the TUD regression task. This combination of weight updates is similar to pre-training performed in other domains using AEs to create useful feature maps for classification [10]. As shown later, we compare the results of this process to a fully-connected neural network without the two pre-training steps shown in Figure 5.

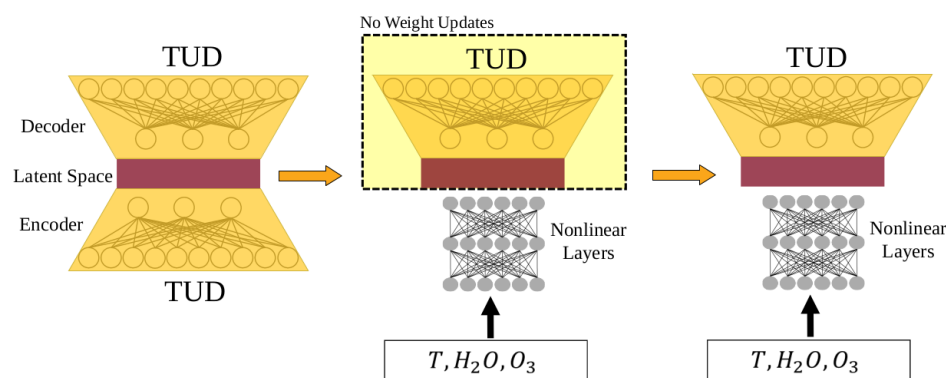


Figure 5. The RT model is created by first creating a low-dimensional representation of the TUD vectors with acceptable at-sensor radiance reconstruction errors. The latent space and decoder parameters are locked and a sampling model is fit to correctly identify the low-dimensional components to map atmospheric measurements to their corresponding TUD vectors. This diagram is specific for the SAE approach, but the encoder and decoder can be replaced with equivalent PCA transformations.

Creating a RT model also highlights differences in the latent space construction among differing dimension reduction techniques. Ideally, similar atmospheric conditions will form clusters in the low-dimensional space. Sampling anywhere within these clusters should result in similar TUD vectors reducing the impact of sampling errors. Additionally, small changes in generative model components should lead to small deviations in the generated TUD vectors. We found that pre-training an AE to reconstruct TUD vectors was useful for enforcing a similarity between generative model components and their corresponding TUD vectors. Both of these properties allow a sampling method to quickly learn a relationship between atmospheric measurements and the corresponding generative model components. Difficulties in sampling the latent space to generate TUD vectors may be the result of one or both of these properties not being satisfied.

The loss function for updating the nonlinear sampling layers is dependent on the dimension reduction method used to form the latent space. For SAE dimension reduction, the loss function is simply the MSE calculated between predicted components and truth components. In this case, truth components are derived from the encoder model. For PCA, components are ordered according to the variance they capture, therefore, it is important for the sampling method to correctly predict

components capturing higher variance. The loss function used in this case is a weighted MSE described as:

$$\text{MSE}_{\text{PCA}} = \frac{1}{K} \sum_{i=1}^{K+1} \mathbf{w}_i (\mathbf{x}_i - \mathbf{y}_i)^2 \quad (12)$$

where \mathbf{w}_i corresponds to the percentage of variance captured by component i during the PCA fitting process. This scaling ensures a weighted reconstruction reflecting component importance. Next, TIGR data augmentation is discussed as more atmospheric measurement samples are needed to fit the large number of dimension reduction model parameters.

2.5. Atmospheric Measurement Augmentation

The 2311 TIGR atmospheric measurements span expected atmospheric variability, providing a set of basis measurements to fit dimension reduction models. To accurately fit the thousands of weights within a neural network, additional TUD vectors are needed to interpolate between the TIGR samples. In this section, a data augmentation approach is introduced, resulting in over 11,000 new TUD vectors derived from the TIGR database.

This study will only consider cloud free conditions requiring a relative humidity calculation to be performed on each TIGR measurement. Using a threshold of 96% relative humidity, we downselect the TIGR data to 1755 samples. Each remaining temperature, water content and ozone measurement is concatenated forming vectors of length 198. A weighted PCA approach is employed by air mass type (Polar, Tropical, etc.) on the concatenated measurement vectors such that reconstruction error is minimized at low-altitudes. Low-altitude atmospheric dynamics have a larger impact on the resulting TUD vector, requiring more accurate reconstruction at these altitudes to generate realistic TUD values. The data considered here used 15 components to capture nearly all variance using the weighted scheme.

Next, a 10 mixture GMM is fit to the 15 dimensional latent space created by the weighted-PCA approach. Sampling the multivariate normal distribution results in new latent space samples that are inverse transformed using the weighted-PCA model. This creates new temperature, water content and ozone measurement vectors for a particular air mass category. The relative humidity of the generated measurements is again calculated, removing new measurements exceeding 96%. Measurement vectors exceeding 10% of filtered TIGR bounds are removed and any measurements with pressure level gradients larger than the TIGR data are also removed. By filtering the generated results, the generated measurements closely match the statistics of the original data as shown in Figure 6. These measurements are forward-modeled with LBLRTM increasing the number of samples in the TUD training data. This data augmentation step is important as the number of parameters in most AE models is significantly higher than the number of samples in the original TIGR database. Model validation and testing will only consider held out sets of original TIGR samples to ensure model performance isn't based on possible errors in the augmentation process.

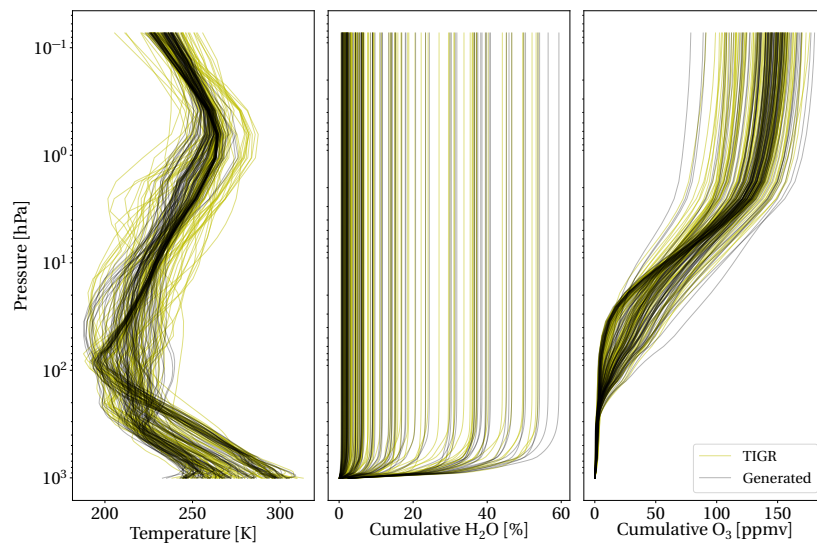


Figure 6. Generated atmospheric measurements based on sampled PCA components from the GMM.

Using the metrics, models and augmented data outlined in this section, algorithm performance is compared in the next section. The best performing methods are considered for the RT modeling problem where we show the overall effectiveness of using SAE pre-training to improve RT performance.

3. Results

In this section, we first consider the impact of including the augmented atmospheric measurements in fitting the dimension reduction algorithms. After validating the augmented data improves model performance, we next compare the loss function described in Equation 11 against MSE. Finally, the latent space created by each dimension reduction technique is sampled following the process outlined in Figure 5 to compare RT model performance.

3.1. Atmospheric Measurement Augmentation

Using the data augmentation approach outlined in Section 2.5, over 11,000 new atmospheric measurements were created from the original 1755 filtered TIGR measurements. These measurements were forward-modeled through LBLRTM to create high-resolution TUD vectors. The augmented TUD vectors were downsampled to the Mako instrument line shape (ILS) resulting in 128 spectral channels for each component of the TUD vector. To test the validity of this augmentation strategy, we consider dimension reduction performance with and without the use of the augmented TUD samples.

All results are reported on test TUD vectors derived from the original TIGR database to verify the models generalize to real measurement data. The validation and test TIGR data points were selected based on total optical depth. Optical depth, $OD(\lambda)$, is related to transmittance by $\tau(\lambda) = e^{-OD(\lambda)}$. The validation and test sets contain the entire range of optical depth encountered in the TIGR data, ensuring these sets are not biased toward a particular atmospheric condition. To extract average performance information, 5 fold cross validation was used for the PCA model where the training and validation sets were still configured to contain the entire range of total optical depths in the data. For the SAE model, random weight initialization was used to derive performance statistics.

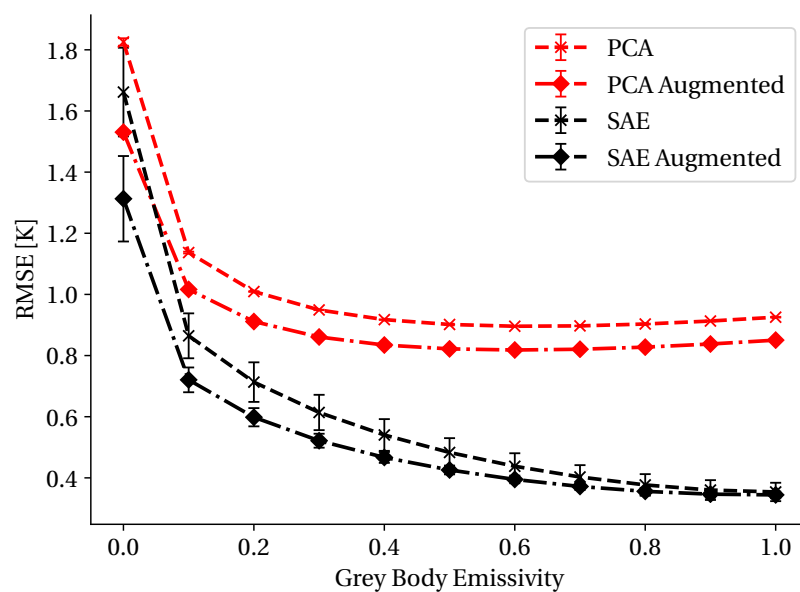


Figure 7. Dimension reduction techniques show improved results when using the augmented **TIGR** data. All models reduce the input data to 5 components in this plot, however, the number of components is an additional hyperparameter that will be considered later.

As shown in Figure 7, reconstruction performance improves when using the augmented data to train both the **SAE** and **PCA** algorithms. Since the **SAE** has many parameters to fit, the additional information encoded in the augmented data allows this technique to extract these underlying relationships with lower error. This additional information also improves **PCA** reconstruction performance by enforcing the axes of maximum variance within the data.

The **SAE** encoder and decoder have mirrored configurations with the encoder consisting of 40 nodes followed by 15 nodes connecting to the latent space. The overall network node structure is 384-40-15-N-15-40-384 where N represents the number of latent components and 384 corresponds to the **TUD** vector dimension. This configuration was found by conducting a hyperparameter sweep across the number of layers, nodes per layer, learning rate, batch size, and activation functions. In Figure 7, all models use 5 components to evaluate the utility of the augmented data. A learning rate of 0.001 was found to achieve acceptable results when training for 500 epochs. Additionally, the **ReLU** activation function was used for all nodes, except the output layer which consisted of linear activation functions. **MSE** loss was used for all **SAE** models in Figure 7. Next, the augmented data is used to evaluate the utility of the physics-based loss function described in Equation 11.

3.2. At-Sensor Loss Constraint

The same **SAE** architecture used to create Figure 7 is used again in this analysis where we evaluate the utility of the loss function in Equation 11. To compare the **MSE** against our new loss function, two identical networks were trained. Specifically, each network was initialized with the same weights and samples were presented in identical order, where the only difference between the networks is in the loss calculation. Both networks utilize 5 components in the latent space for dimension reduction to demonstrate differing loss characteristics for a particular network configuration. As shown in Figure 8, the physics-based loss function provides lower reconstruction error for reflective materials ($\epsilon(\lambda) < 0.5$). This is the designed behavior of the loss function since the at-sensor radiance error for reflective materials increases based on the relationship shown in Equation 8. As the material emissivity trends toward one, the at-sensor radiance can be described by Equation 4, where errors are no longer multiplicative. In this regime, **MSE** and the physics-based loss function converge. The error bars in Figure 8 are based on random weight initialization of the networks for repeated training trials.

MSE shows significantly less variance across repeated training but is unable to reach the lower RMSE values for reflective materials observed with the physics-based loss function. MSE outperforms the physics-based loss function for emissive materials since the reconstruction problem no longer benefits from this additional information and inhibits the model the training process.

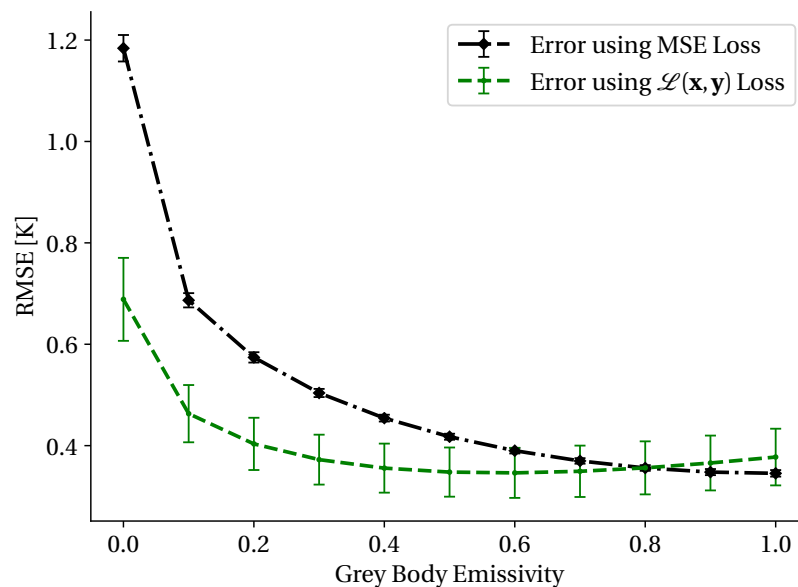


Figure 8. Comparison of SAE performance when using strictly MSE loss or the loss function described in Equation 11. Updating the model using information from the at-sensor radiance error improves reconstruction performance for reflective materials. The error bars represent the performance standard deviation when training multiple networks with identical architectures and random weight initialization.

3.3. Dimension Reduction Performance

Using the same SAE structure discussed in Section 3.1, the augmented training data and the physics-based loss function, the number of components were adjusted to compare dimension reduction performance. Rather than creating plots similar to Figure 7 for each component configuration, the area under the RMSE curve was calculated (smaller is better), creating the plot shown in Figure 9. When the area under the curve is similar for multiple methods, we cannot determine which method is better without further analysis. This is because the individual curves demonstrate different performance characteristics for emissive and reflective materials. For example, we cannot say which method is best when using 8 components without also considering the material emissivity.

From Figure 9, it is clear lower reconstruction error can be achieved with the SAE when using a low number of components. While PCA can achieve overall lower error with greater than 8 components, this isn't ideal for sampling the low-dimensional space as additional components complicate the sampling process. For the SAE model, 4 components is adequate for reconstructing the data, when considering the 176 test samples used to create these results.

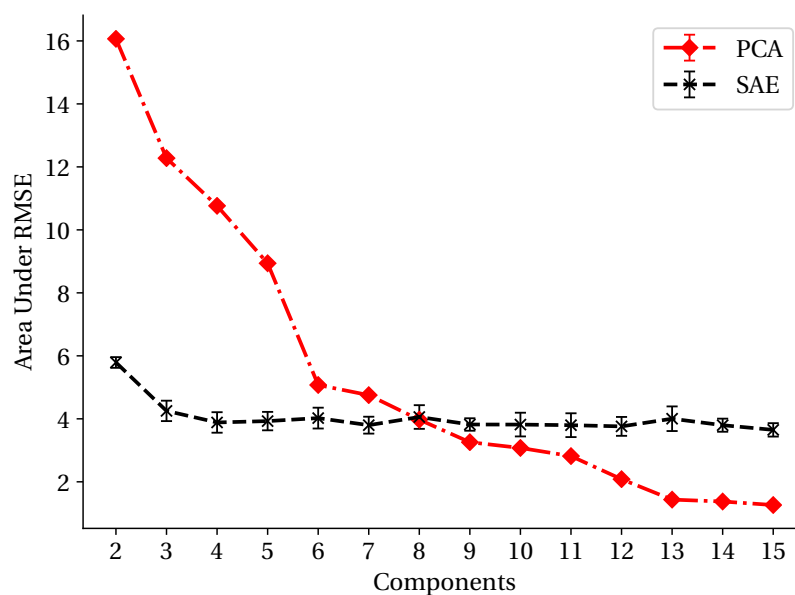


Figure 9. Varying the number of latent components and calculating the area under the RMSE curve shown in Figure 8 shows how many components are necessary to reconstruct the TIGR data. Results are plotted for the validation set consisting of 158 samples using the augmented data for training and the loss function outlined in Equation 11 for SAE training. The PCA error bars correspond to performance standard deviation when using 5 fold cross-validation. The SAE errors bars show the performance standard deviation when random weight initialization is used.

3.4. Radiative Transfer Modeling

The low-dimensional representations created by PCA and SAE can be used for efficient radiative transfer modeling by mapping TIGR atmospheric measurements to the encoder-predicted latent components. This mapping is more difficult if diverse atmospheric conditions are closely grouped in the latent space, or similarly, if small deviations in the latent space create large TUD vector differences. The same metrics used for developing the dimension reduction methods are also used to compare RT models as we are primarily concerned with at-sensor radiance reconstruction error across a range of material emissivities.

The 66 pressure level measurements for air temperature, water content and ozone interpolated from the TIGR database are concatenated together forming a 198 dimensional input vector for latent space prediction. A two layer, fully-connected neural network (58-29) is used to map the atmospheric measurement vector to the latent space. This network utilizes ReLU activation functions, a learning rate of 0.001, and a batch size of 16. The atmospheric measurements were z-score standardized by feature and the latent space components were normalized between 0 and 1. This network configuration was used for the 6 component PCA model and the 4 component SAE model. The 6 component PCA model was selected for this analysis because of the large reduction in RMSE error from 5 to 6 components. In both cases, the network was trained for 500 epochs with validation and training loss stabilizing between 200 and 300 epochs. This model configuration was found by performing a hyperparameter sweep for the PCA model and identifying the configuration with minimum validation set at-sensor reconstruction error for a range of emissivity values. The sampling network weights were updated based on latent space component prediction errors with SAE utilizing standard MSE loss. For sampling the PCA latent space, weighted MSE loss was used as outlined in Equation 12.

The resulting RMSE for each RT model is shown in Figure 10 where the RT model derived from the SAE decoder has the lowest error across all emissivity values. The Artificial Neural Network (ANN) model results shown in Figure 10 represent a baseline approach where an end-to-end neural network was trained with the same network configuration as the SAE RT model (198-58-29-4-15-40-384)

using **MSE** for the model loss function. The performance difference between the **SAE RT** model and the **ANN** model highlight the advantages of first using an **AE** to initially fit network weights before training the **RT** model. Initially training the **SAE** weights clusters similar atmospheric conditions together, limiting the impact of **RT** model sampling errors, improving overall **RT** model performance. Additionally, the at-sensor radiance loss function used in the **SAE** approach significantly improves model performance for reflective materials. Without the at-sensor radiance loss function or the weight initialization imposed by training an **AE** model, the **ANN** model is unable to reconstruct **TUD** vectors with the accuracy of the **SAE RT** model.

The **SAE RT** model was trained by first fitting encoder and decoder networks to minimize at-sensor radiance reconstruction error, following by training a small sampling network to correctly predict latent components. This training methodology did not allow updates to the decoder network after training of the sampling network. The **SAE RT Tuned** model result shown in Figure 10, has the same configuration as the **SAE RT** model, but the decoder weights were also updated after the sampling network training converged. This final training step utilized the same physics-based loss function used for the initial **SAE** training, improving reconstruction performance for reflective emissivity values. Since the previous training steps had already created weight matrices resulting in high performance, only small changes were needed to further reduce reconstruction error.

Additionally, the 15 component **PCA** model results are shown. In this case, even higher **RMSE** error is observed since correctly sampling the 15 components is a more complex task. While the 15 component **PCA** model has the lowest reconstruction error during the **TUD** reconstruction training phase, the added complexity in latent space fitting limits the utility of this model. In all cases the highest errors are observed for reflective materials since errors in transmittance and downwelling radiance are multiplicative in this region.

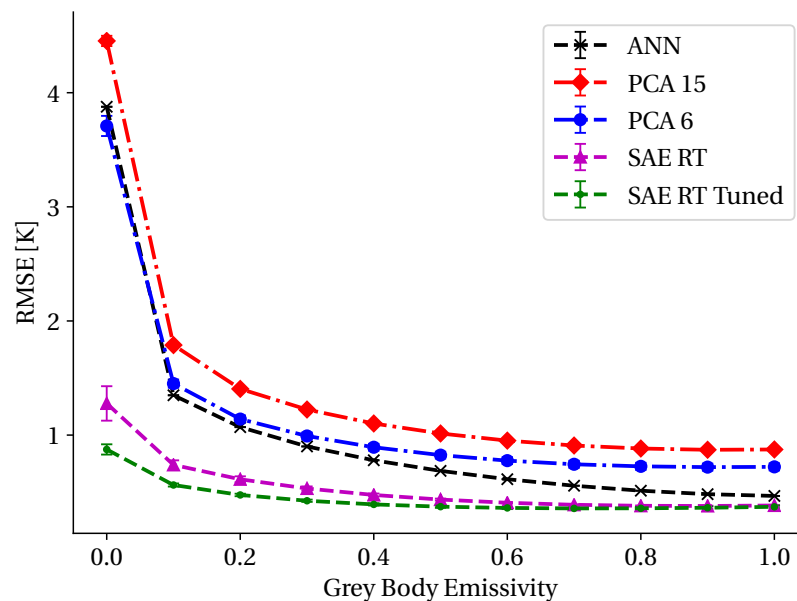


Figure 10. The performance of the **RT** models is shown as a function of emissivity where it is clear the **SAE** derived **RT** models create a latent space that is easier to sample with a small neural network. In all cases performance improves as materials become more emissive since downwelling radiance plays a less significant role in these cases. The 15 component **PCA** model is also shown, where sampling the 15 components correctly becomes a complex problem resulting in lower overall performance.

Considering only the **SAE RT Tuned** model, the at-sensor radiance **RMSE** as a function of wavenumber and emissivity is shown in Figure 11. These results are the average **RMSE** for the 176 test TIGR samples at each emissivity level. For emissive materials, **RT** model errors are below 0.5 K

for most bands. The ozone absorption bands between 1050 and 1100 cm^{-1} lead to larger errors because of limited transmittance in this domain of the spectrum resulting in large deviations in the at-sensor radiance. The challenge of correctly identifying a **TUD** vector for reflective materials is seen by the high **RMSE** for $\epsilon = 0.01$ in Figure 11. While these errors appear large on the scale shown in Figure 11, these errors are significantly larger using other models based on the results for $\epsilon = 0$ in Figure 10.

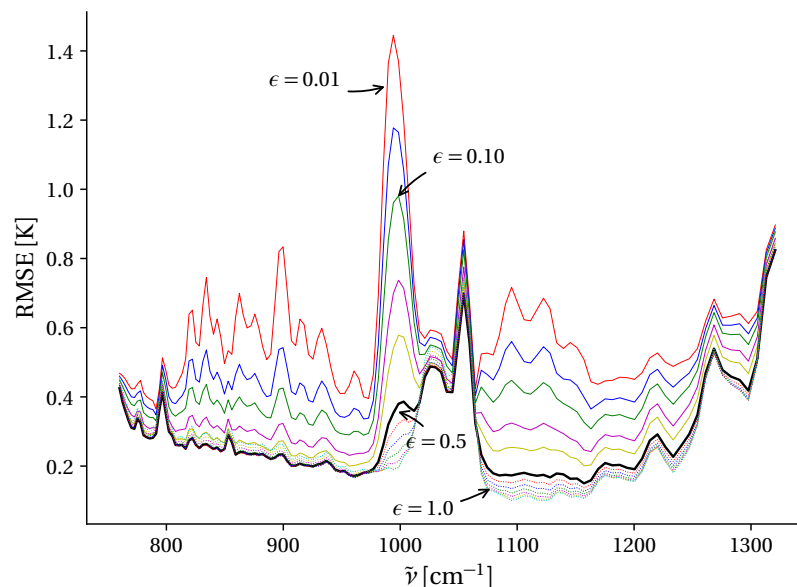


Figure 11. The at-sensor radiance **RMSE RT** model errors for 176 test TIGR samples as a function of surface emissivity expressed in spectral brightness temperature. Model errors decrease with increasing emissivity, consistent with the findings in Figure 10.

Based on the results shown in Figure 11, the **SAE-based RT** model can estimate **TUD** vectors with errors below 0.5 K for most spectral bands and a range of emissivity values. These generated **TUD** vectors are useful for estimating surface leaving radiance described in Equation 3 if estimates of atmospheric conditions can be provided. Also, the **RT** model can be used to quickly estimate **TUD** vectors for data augmentation purposes when multiple estimates of the atmospheric conditions are assumed. This provides a classification algorithm efficient access to a locally varying range of likely spectral signatures for a given material. The data was constructed such that this method could be used for the Mako **LWIR** hyperspectral sensor, however, resampling of the high-resolution **LBLRTM** data can be performed for other sensors.

3.5. Atmospheric Measurement Estimation

Finally, we consider estimation of the most likely atmospheric measurements for a given **TUD** vector using the formulated **RT** model. Instead of propagating inputs forward through the **RT** model, this section considers estimation of the model's inputs for a given output. Since the **RT** model is composed of a sampling network (atmospheric measurements to latent components) and a decoder network (latent components to **TUD** vectors) the estimation problem can be partitioned into two steps.

First, the latent space components are identified that correspond to the **TUD** vector. Since the latent space only consists of 4 components, finding these 4 components using an optimizer takes little time and results in predicted **TUD** vectors closely matching the given **TUD** vector. This optimization is performed with respect to the decoder network of the **RT** model. As an example of this process, we select a **TUD** vector with a 50th percentile reconstruction error from the test data set. Figure 12 shows the **LBLRTM** generated **TUD** components and the predicted **TUD** from optimizing the 4 latent components.

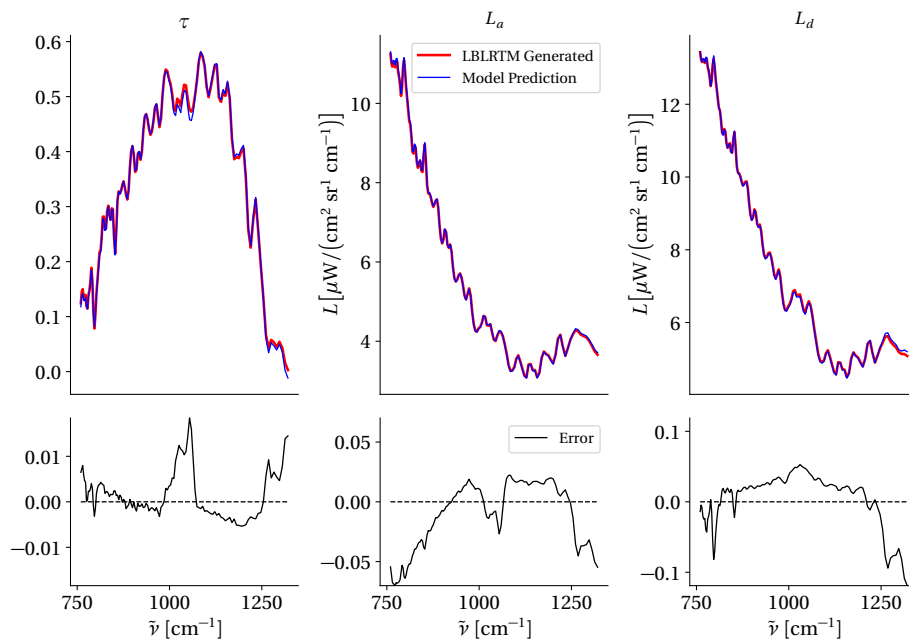


Figure 12. The top 3 panels are the predicted TUD components plotted against the LBLRTM generated TUD components. The predicted TUD components were generated by optimizing the 4 latent components. The bottom 3 panels are the TUD component residual curves, showing low error across most spectral channels.

After identifying the latent components, inputs to the sampling network (atmospheric measurements to latent components) must be optimized to identify measurements that will produce the estimated latent components. Unfortunately, the input measurement vector contains 198 values (temperature, water content and ozone at 66 pressure levels) and optimizing this large number of values for 4 components is a time-consuming and difficult task.

To make this problem more tractable, a PCA transform was fit to the training data atmospheric measurements. Using 10 components to represent the atmospheric measurements captures 90% of data variance and simplifies the optimization problem as only 10 values must be optimized to predict the 4 latent components. For the same TUD vector used in Figure 12, the result of the atmospheric measurement estimation process is shown in Figure 13. Interestingly, the largest errors occur at high altitudes, where deviations in these measurements have less impact due to lower air density.

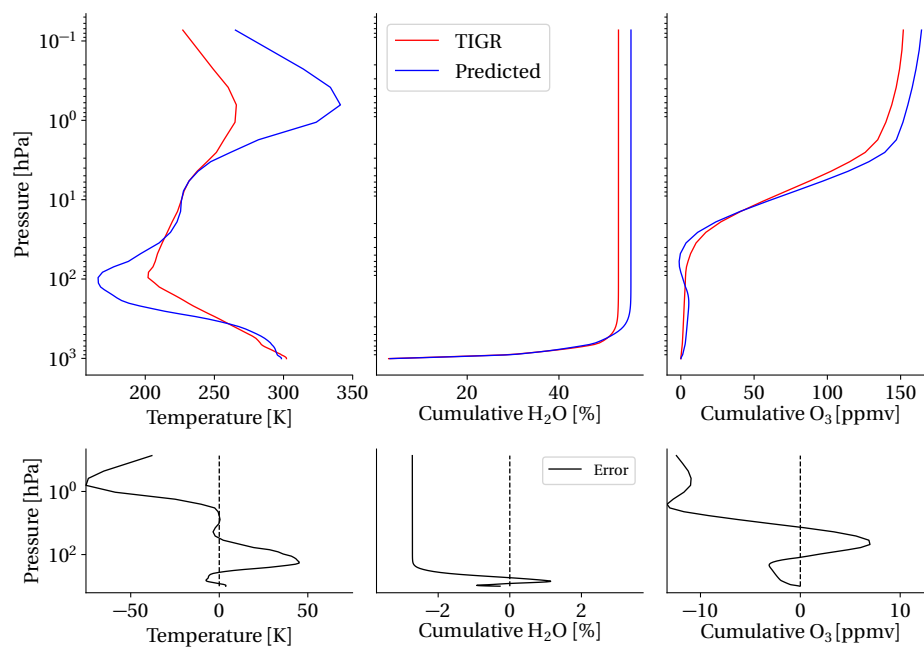


Figure 13. Predicted atmospheric measurements compared to the TIGR atmospheric measurements. The predicted atmospheric measurements will produce a close match to a given TUD vector but do show some deviations from the original TIGR measurement, specifically at high altitudes. Fortunately, high altitude errors has less impact on the at-sensor radiance error because of lower air density.

The results shown in Figures 12 and 13 are for a single TUD vector. This process was repeated for all TUD vectors in the test data set to determine overall performance metrics. Errors between predicted temperature and water content measurements and the corresponding TIGR atmospheric measurements are weighted by the density at the discrete pressure levels as errors at high altitudes will have a lower impact on the TUD vector prediction. For all 176 test set TUD vectors, we observe an average error of 2.61 K for predicted temperature profiles and 0.45 cumulative H₂O % for predicted water content profiles. For ozone measurements, an atmospheric density weighting was not applied. The observed average error for ozone measurement estimation was 0.79 ppmv.

Overall, the goal of the RT model is predict TUD vectors, based on known or estimated atmospheric measurements. By showing the model's ability to estimate atmospheric measurements from a given TUD vector (inverse problem), we have demonstrated the utility of low-dimension representations of the TUD vectors. Additionally, since the latent space clusters similar atmospheric conditions together (latent-space-similarity), an ensemble of likely atmospheric measurements can be generated for a given TUD by applying small deviations to latent components.

4. Conclusions

This study has leveraged SAEs with a novel physics-based loss function to reduce TUD vector dimensionality such that fast and effective LWIR RT models could be constructed by sampling the low-dimensional TUD representation. By using an AE pre-training step, the low-dimensional TUD representation clustered similar atmospheric conditions together reducing sampling errors. Additionally, the pre-training step verified that small deviations in the low dimensional TUD representation corresponded to small deviations in the high dimensional TUD vector. These approaches were shown to reconstruct at-sensor radiance with errors below 0.5 K for most emissivity values.

The dimension reduction results utilized real atmospheric measurements from the TIGR database and augmented data derived from this same database. Using augmented atmospheric measurements improved both PCA and SAE performance for a range of material emissivities. PCA was shown to reconstruct data with lower error than the SAE when using beyond 8 components. The SAE

performance did not improve significantly when using more than 4 components, demonstrating adequate capacity for the augmented TIGR data.

Sampling the low-dimensional representations created by these methods highlighted significant differences in TUD reconstruction. This study found the SAE latent space easier to sample, resulting in lower RT model errors. Additionally, training the entire RT model after pre-training the sampling network and decoder networks improved RT performance. The RT model was further explored to identify the most likely atmospheric measurements for a given TUD vector. This analysis revealed that RT model inputs could easily be optimized resulting in predicted atmospheric measurements with some agreement to the TIGR measurements. While this was not the goal of this work, we will explore the utility and limitations of SAEs in the inverse problem of estimating atmospheric conditions from spectral measurements.

Optimizing the latent components for a particular TUD vector is a straightforward process, however, no known physical quantities are directly correlated with these components. No constraints were placed on the formulation of this latent space other than the overall network loss function. This unconstrained representation creates a disentanglement problem limiting the utility of the SAE as a generative model when limited atmospheric information is available. Future work in this area will consider additional constraints on the latent space to improve upon this disentanglement problem, offering a means for creating TUD vectors with properties representative of specific atmospheric conditions. Specifically, if complete atmospheric measurements are not available, this analysis will determine what information is required to predict the most likely TUD vector using a small number of components. Methods such as Variational AEs, multi-modal AEs and Generative Adversarial Networks coupled with physical constraints will be investigated toward this goal.

Author Contributions: All coauthors contributed to the research methodology and manuscript. Nicholas Westing led the research investigation, wrote the first draft, and developed all the dimension reduction, analysis, and visualization code. Kevin Gross and Brett Borghetti provided mentorship and ideas related to the blending of physics principles and machine learning in the dimension reduction approach.

Funding: This research was supported by Dr. Jacob Martin at Air Force Research Laboratory through the effort "Developing Physics-Based Machine Learning Algorithms to Exploit Hyperspectral Imagery."

Acknowledgments: The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the United States Air Force or the U.S. Government.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hall, J.L.; Boucher, R.H.; Buckland, K.N.; Gutierrez, D.J.; Keim, E.R.; Tratt, D.M.; Warren, D.W. Mako airborne thermal infrared imaging spectrometer: performance update. *Imaging Spectrometry XXI*. International Society for Optics and Photonics, 2016, Vol. 9976, p. 997604.
2. Goetz, A.F.; Vane, G.; Solomon, J.E.; Rock, B.N. Imaging spectrometry for earth remote sensing. *Science* **1985**, *228*, 1147–1153.
3. Eismann, M.T. *Hyperspectral Remote Sensing*; Bellingham, WA USA: SPIE, 2012.
4. Manolakis, D.G.; Lockwood, R.B.; Cooley, T.W. *Hyperspectral Imaging Remote Sensing: Physics, Sensors, and Algorithms*; Cambridge University Press, 2016.
5. Gao, B.C.; Montes, M.J.; Davis, C.O.; Goetz, A.F. Atmospheric correction algorithms for hyperspectral remote sensing data of land and ocean. *Remote Sensing of Environment* **2009**, *113*, S17–S24.
6. Ghamisi, P.; Plaza, J.; Chen, Y.; Li, J.; Plaza, A.J. Advanced spectral classifiers for hyperspectral images: A review. *IEEE Geoscience and Remote Sensing Magazine* **2017**, *5*, 8–32.
7. Benediktsson, J.A.; Ghamisi, P. *Spectral-Spatial Classification of Hyperspectral Remote Sensing Images*; Artech House, 2015.
8. Rodarmel, C.; Shan, J. Principal component analysis for hyperspectral image classification. *Surveying and Land Information Science* **2002**, *62*, 115–122.

9. Melgani, F.; Bruzzone, L. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on Geoscience and Remote Sensing* **2004**, *42*, 1778–1790.
10. Chen, Y.; Lin, Z.; Zhao, X.; Wang, G.; Gu, Y. Deep learning-based classification of hyperspectral data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2014**, *7*, 2094–2107.
11. Chen, Y.; Zhao, X.; Jia, X. Spectral-spatial classification of hyperspectral data based on deep belief network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2015**, *8*, 2381–2392.
12. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 2012, pp. 1097–1105.
13. Hu, W.; Huang, Y.; Wei, L.; Zhang, F.; Li, H. Deep convolutional neural networks for hyperspectral image classification. *Journal of Sensors* **2015**.
14. Liang, H.; Li, Q. Hyperspectral imagery classification using sparse representations of convolutional neural network features. *Remote Sensing* **2016**, *8*, 99.
15. Mou, L.; Ghamisi, P.; Zhu, X.X. Deep recurrent neural networks for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing* **2017**, *55*, 3639–3655.
16. Wu, H.; Prasad, S. Convolutional recurrent neural networks for hyperspectral data classification. *Remote Sensing* **2017**, *9*, 298.
17. Young, S.J.; Johnson, B.R.; Hackwell, J.A. An in-scene method for atmospheric compensation of thermal hyperspectral data. *Journal of Geophysical Research: Atmospheres* **2002**, *107*.
18. Gu, D.; Gillespie, A.R.; Kahle, A.B.; Palluconi, F.D. Autonomous atmospheric compensation (AAC) of high resolution hyperspectral thermal infrared remote-sensing imagery. *IEEE Transactions on Geoscience and Remote Sensing* **2000**, *38*, 2557–2570.
19. Borel, C.C. ARTEMIS—an algorithm to retrieve temperature and emissivity from hyper-spectral thermal image data. 28th Annual GOMACTech Conference, Hyperspectral Imaging Session. Citeseer, 2003, Vol. 31, pp. 1–4.
20. Clough, S.A.; Iacono, M.J.; Moncet, J.L. Line-by-line calculations of atmospheric fluxes and cooling rates: Application to water vapor. *Journal of Geophysical Research: Atmospheres* **1992**, *97*, 15761–15785.
21. Clough, S.; Shephard, M.; Mlawer, E.; Delamere, J.; Iacono, M.; Cady-Pereira, K.; Boukabara, S.; Brown, P. Atmospheric radiative transfer modeling: a summary of the AER codes. *Journal of Quantitative Spectroscopy and Radiative Transfer* **2005**, *91*, 233–244.
22. Berk, A.; Conforti, P.; Kennett, R.; Perkins, T.; Hawes, F.; van den Bosch, J. MODTRAN6: a major upgrade of the MODTRAN radiative transfer code. *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XX*. International Society for Optics and Photonics, 2014, Vol. 9088, p. 90880H.
23. Liu, X.; Smith, W.L.; Zhou, D.K.; Larar, A. Principal component-based radiative transfer model for hyperspectral sensors: theoretical concept. *Applied Optics* **2006**, *45*, 201–209.
24. Kopparla, P.; Natraj, V.; Spurr, R.; Shia, R.L.; Crisp, D.; Yung, Y.L. A fast and accurate PCA based radiative transfer model: Extension to the broadband shortwave region. *Journal of Quantitative Spectroscopy and Radiative Transfer* **2016**, *173*, 65–71.
25. Goody, R.; West, R.; Chen, L.; Crisp, D. The correlated-k method for radiation calculations in nonhomogeneous atmospheres. *Journal of Quantitative Spectroscopy and Radiative Transfer* **1989**, *42*, 539–550.
26. Chedin, A.; Scott, N.; Wahiche, C.; Moulinier, P. The improved initialization inversion method: A high resolution physical method for temperature retrievals from satellites of the TIROS-N series. *Journal of climate and applied meteorology* **1985**, *24*, 128–143.
27. Chevallier, F.; Chérut, F.; Scott, N.; Chédin, A. A neural network approach for a fast and accurate computation of a longwave radiative budget. *Journal of Applied Meteorology* **1998**, *37*, 1385–1397.
28. Acito, N.; Diani, M.; Corsini, G. Coupled Subspace-Based Atmospheric Compensation of LWIR Hyperspectral Data. *IEEE Transactions on Geoscience and Remote Sensing* **2019**.
29. Martin, J.A. Target detection using artificial neural networks on LWIR hyperspectral imagery. *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XXIV*. International Society for Optics and Photonics, 2018, Vol. 10644, p. 1064402.
30. Warren, D.; Boucher, R.; Gutierrez, D.; Keim, E.; Sivjee, M. MAKO: a high-performance, airborne imaging spectrometer for the long-wave infrared. *Imaging Spectrometry XV*. International Society for Optics and Photonics, 2010, Vol. 7812, p. 78120N.

31. Friedman, J.; Hastie, T.; Tibshirani, R. *The elements of statistical learning*; Vol. 1, Springer series in statistics New York, NY, USA; 2001.
32. Hinton, G.E.; Salakhutdinov, R.R. Reducing the dimensionality of data with neural networks. *Science* **2006**, *313*, 504–507.
33. Nair, V.; Hinton, G.E. Rectified linear units improve restricted boltzmann machines. Proceedings of the 27th International Conference on Machine Learning (ICML-10), 2010, pp. 807–814.
34. Maas, A.L.; Hannun, A.Y.; Ng, A.Y. Rectifier nonlinearities improve neural network acoustic models. in ICML Workshop on Deep Learning for Audio, Speech and Language Processing, 2013.