

Copyright 2024, Society of Photo-Optical Instrumentation Engineers (SPIE). One print or electronic copy may be made for personal use only. Systematic reproduction and distribution, duplication of any material in this paper for a fee or for commercial purposes, or modification of the content of the paper are prohibited.

David Stelter, Ethan Brewer and Robert Sundberg "Atmospheric Correction using Diffusion Models and MODTRAN for Constrained Training," SPIE 13031, Algorithms, Technologies, and Applications for Multispectral and Hyperspectral Imaging XXX, (23 - 24 April 2024).

DOI: <https://doi.org/10.1117/12.3012882>

Paper available below:

# Atmospheric Correction using Diffusion Models and MODTRAN for Constrained Training

David Stelter, Ethan Brewer, and Robert Sundberg

Spectral Sciences, Inc., 30 Fourth Ave Suite 2, Burlington, MA 01803

## ABSTRACT

Diffusion models are a promising generative artificial intelligence (AI) technique for denoising and synthetic data generation. A de-noising diffusion model is applied to the atmospheric correction problem. In place of true noise, the model is trained based on predictions from the physics-based atmospheric radiative transfer tool, MODTRAN, to constrain the training environment. In this paper, we present results from a trained diffusion-based neural network model applied to hyperspectral image data and assess performance compared to conventional empirical atmospheric correction algorithms.

**Keywords:** Machine Learning, Diffusion Models, Atmospheric Correction

## 1. INTRODUCTION

Generative artificial intelligence has seen immense growth for techniques to create images, sounds, and other modalities. A key methodology is the denoising diffusion model which synthesizes content directly from sampled noise or compressed text input.<sup>1,2</sup> Denoising models have been deployed for a variety of modalities including image and video generation, text-to-speech, and 3-D model generation.<sup>3</sup> Other efforts have extended these techniques to predict accurate and physically meaningful states from a noise distribution, most notably for molecular structure prediction.<sup>4,5</sup> However, diffusion models can also learn noise that is not random and instead derived from a physical process. Recent work applied this strategy to atmospheric correction and showed that the effect of the atmosphere, as computed by a physics-based atmospheric modeling tool, can be viewed as “noise” and sufficiently corrected for by use of a diffusion model.<sup>6</sup>

Remote sensing and accurate retrieval of ground properties using overhead electro-optical/infrared (EO/IR) imagery requires compensation for the atmosphere. Subtracting these effects, including molecular absorption and scattering, is a crucial step for spectral and image processing and the focus of much past effort with physics-based, in-scene, and data-driven models.<sup>6-18</sup> Method development has focused in the visible / short-wave infrared (SWIR) which ranges from about 0.4 to 3  $\mu\text{m}$  and methodologies for mid-wave (MWIR, 3 - 5  $\mu\text{m}$ ) and long-wave (LWIR, 8 - 14  $\mu\text{m}$ ) compensation are more difficult. We aim to show that modern self-supervised machine learning techniques, based on denoising diffusion models, can be applied across this spectral range by identifying key input data and physics-based constraints.

Diffusion models leverage deep neural networks to learn a continuous, Markovian process to systematically repress “noise” from an input signal. For image generation as an example, a noisy input is typically sampled from a text input prompt, i.e., a large language model, and the diffusion process generates representative imagery given this initial state. During training, pristine images are input, noise is added in accordance to a sampled noise schedule, and the model is trained to separate the noise from the combined image. For processes where accurate forward models (i.e.- models that compute or predict addition of noise to an image) exist, the training process can be revamped where instead of adding random noise to a known pristine input, a forward model can be used to predict a physically meaningful “noise” that is added to an underlying signal. The diffusion model then learns the inverse problem, de-noising, constrained by the forward model.

In this paper, we couple a forward model for the atmospheric radiance, MODTRAN,<sup>19</sup> with the diffusion modeling training process. The forward model, MODTRAN or other atmospheric propagation code, computes

---

Further author information: (Send correspondence to D. S.)  
E-mail: dstelter@spectral.com

the path radiance of a line-of-sight through the atmosphere given inputs on the ground properties, atmospheric composition, and sensing geometry. Many of these inputs can be sampled on-the-fly to generate novel viewing scenarios and conditions. These results can be leveraged to train inverse models that capture the atmospheric correction process. Further, involving MODTRAN directly during training enables a self-supervised training loop that does not require labeled data, as the "labels" are the outputs from MODTRAN, and can be trivially expanded to arbitrarily large datasets. Thus, in place of sampling noise during training, the underlying truth values in the training data are used as inputs for a MODTRAN calculation. The noise is computed from the truth and forward model result, and used to train a denoising diffusion model. We train and demonstrate correction in two regions of the EO/IR spectrum, an expanded SWIR model based on prior work<sup>6</sup> and an initial demonstration MWIR correction model. For SWIR, the training conditions are expanded to include off-nadir viewing angles and varying atmospheric conditions. For MWIR, an initial demonstration for retrieval of the spectral emissivity is shown.

## 2. METHODS

### Training Procedure

The model is trained to predict the atmospheric "noise" based on the mean absolute error loss between the truth noise and predicted noise,

$$\mathcal{L} = \frac{\sum_i^b |N_i^{pred} - N_i^{truth}|}{b} \quad (1)$$

For each training batch with size  $b$ , altitude specific quantities are computed by MODTRAN given a database of surface properties. Then, the truth noise,  $N_{truth}$ , is computed using data output by MODTRAN as

$$N_{truth} = Q_{alt} - \frac{\sqrt{n_s} Q_{surf}}{\sqrt{1 - n_s}} \quad (2)$$

where  $Q$  represents the quantity of interest that will be retrieved, such as a reflectance or emissivity, and the subscript *alt* indicates an at-altitude (i.e., noisy) value compared to *surf* for the at-ground retrieved value (i.e., noiseless). The diffusive noise schedule,  $n_s$ , corresponds to the pressure fraction of atmosphere and is sampled for each batch of the training epoch. So at large  $n_s$ , there is a large pressure column in the line-of-sight corresponding to a high observer altitude and hence large noise. Thus, MODTRAN inputs  $Q_{surf}$  and the observer altitude derived from  $n_s$ , with any other ancillary inputs, and computes the at-altitude quantity  $Q_{alt}$  mapping to  $n_s$ .

While  $Q$  may appear as an optimal target for direct learning, instead the noise computed from Eq 2 is used for the training loss and to compute gradients to update the model weights. The truth noise computed from MODTRAN is normalized via local scaling,  $x' = x/(|x| + 1)$  which forces data to the [-1,1] range and then subsequently shifted to the range of [0,1]. This avoids issues with normalization in the input data (not crucial for SWIR, but important for other spectral regions) and also gives the network robustness to varying inputs of  $Q_{alt}$  with respect to  $n_s$ .

Once trained, the network predicts the altitude-specific noise,  $N^* = NN(Q_{alt}, n_s)$ , where the \* indicates a predicted value. The predicted surface quantity is computed as

$$Q_{surf}^* = Q_{alt} - \frac{\sqrt{1 - n_s} N^*}{\sqrt{n_s}} \quad (3)$$

Thus, the network predicts a unique  $Q_{surf}^*$  for a given choice of  $n_s$ . The optimal  $n_s$  that best characterizes the atmospheric noise at the measured condition is not known. As a final step,  $Q_{surf}^*$  is computed for many  $n_s$  (typically  $\sim 50$ ) and the median of  $Q_{surf}^*$  at each spectral bin is used as the final prediction.

## SWIR Implementation

For the SWIR, our primary retrieval target is the surface reflectance,  $R_{surf}$ , which we assume as Lambertian between 0 and 1. We define the altitude specific quantity normalized by the total solar irradiance and solar angle,  $R_{alt} = L(\lambda)\pi/I_{sol} \cos(\theta_{sol})$  which returns a well-behaved at-altitude quantity in the range of 0 to 1 in the absence of surface blackbody emission.

For training data we generate a large database of  $R_{surf}$  spectral albedos according to previous work.<sup>6</sup> The total dataset comprises 21,000 unique Lambertian reflectances which is randomly sampled down to 3,200 cases for training for quicker computational throughput. While the path length is determined during training, we introduce other ancillary parameters to the training database that will affect the MODTRAN calculation, namely the viewing angle which we allow to deviate  $\pm 5^\circ$  from nadir and the atmospheric composition which is selected randomly as one of the 6 model atmospheres in MODTRAN.<sup>19</sup> For completeness, the surface temperature in the range of [273, 320] is also added to the training database as a randomly sampled variable, but is not expected to have an effect for the SWIR correction performance.

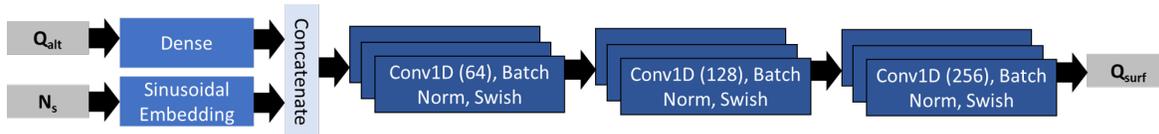


Figure 1. Neural network structure used for atmospheric correction.

We use a 1-D convolutional neural network (CNN) to capture dependence between the spectral channels, shown in Figure 1. Padding is used to ensure that convolutional filters do not reduce the number of desired spectral bins. Groupings of 3 consecutive convolution layers with batch normalization and swish activation<sup>20</sup> are used with successive increases in the total number of filters from 64 up to 256.  $n_s$  is input into a sinusoidal embedding layer with 25 embedding dimensions that maps it to a high frequency space,  $10^3$  Hz, enhancing sensitivity of the model to that input.<sup>21</sup> The embedded space is concatenated with an initial dense layer for the altitude specific data prior to any convolutions.

Training is performed in TensorFlow using AdamW<sup>22</sup> for gradient descent. The learning rate and weight decay are both set to  $10^{-3}$ . Spectral data are selected in the range of 1.5 microns to 2.5 microns and ignored in regions with large optically thick absorption features (i.e., around 1.8-2.0 for H<sub>2</sub>O). In total, 100 spectral bins are used with 0.008 micron resolution, but can be arbitrarily selected depending on the spectral region of interest.

## 3. RESULTS

The network is trained with the above hyperparameters for 500 epochs, though adequate results are obtained by epoch 40. Results are shown using the model weights from epoch 100 to avoid overfitting. Trained model performance is similar to our previous study,<sup>6</sup> however here we have provided more physically-relevant variations in the training data (atmospheric composition and view angle variations) in an attempt to train a more general atmospheric correction model. More spectral bins (100) have also been used for higher spectral resolution output of surface properties to illustrate the scalability of the approach and its potential to be tuned for specific spectral ranges.

Figure 2 shows the network inputs, noise prediction, and surface reflectance retrieval for several different Lambertian surfaces from the MODTRAN spectral albedo database. The network reproduces the spectral noise profile for each of the surfaces. Note that results for only a single  $n_s$  are shown, as computed according to Eq 2, but many spectral noise profiles are used for determination of the final spectral reflectance prediction. Results for highly reflective surfaces are particularly good with some drop in performance where the transmittance drops nearest large atmospheric molecular absorption features. At low reflectance, the predictions become noisy, but still qualitatively represent a non-reflective surface.

It is important to recognize that a single noise schedule does not best fit all predicted noises and, for the purpose of visualization, only a single  $n_s$  for each result/observation is presented in Figure 2. However, when determining the surface reflectance, instead of relying on a single  $n_s$  manifold, we compute network predictions

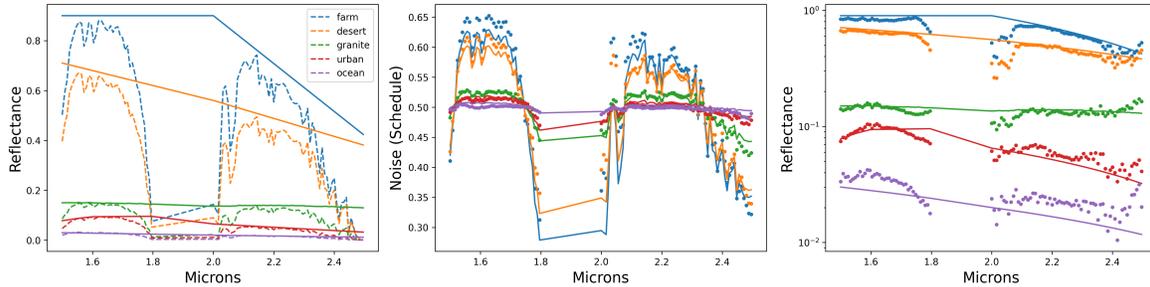


Figure 2. Neural network performance for several different surfaces. Left: Input data (dashes) compared to truth data (lines). Center: noise predictions from network (dots) compared to truth (lines). Right: Predicted surface retrieval values (dots) compared to truth (lines).

for a linear sweep in  $n_s$  space and select the median at each spectral point to avoid outliers. Better approaches for inference and selection of optimal  $n_s$  are likely possible and an aim for future work. We find that, in practice, the results are still smooth and able to capture spectral variation in the surface spectral reflectance albedo. Results for a  $n_s$  sweep for the “farm” case are shown in Figure 3. The noise happens to be best captured at  $n_s = 0.63$  (green) for the farm case and is different for all other cases and observation conditions.

It is somewhat surprising that the network does not reproduce the noise across all  $n_s$  values, though the results retain the general structure of the noise. Typical diffusion models reproduce noise accurately across the entire  $n_s$  manifold. We attribute this to one of several possibilities. First, that the truth noise computed by MODTRAN and used for training is not normalized to a unity variance. Second, the embedded space that combines the choice of  $n_s$  with the spectrally dependent input vector is simply too small to account for the full extent of  $n_s$  and noises sampled during training.

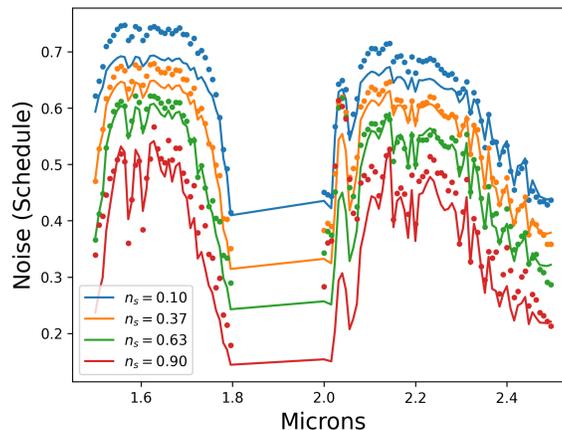


Figure 3. Variation in the noise schedule,  $n_s$ , for the “farm” case with truth noises (lines) and predictions (dots).

Network predictions for a variety of altitude observations are shown in Figure 4. The input data varies less in this case than when varying the surface albedo. Despite this, the noise profiles show large variance and a single network that does not require altitude as an input is able to recover the same surface reflectance curve ranging from 100 m near-nadir path lengths to paths containing > 99% of the atmosphere by pressure.

The above examples are all shown using MODTRAN’s “midaltitude summer” atmospheric profile. To assess the performance of the network on varying atmospheric conditions, we assess performance on atmospheric profiles inclusive of the training data paired with the above validation surface reflectances, not previously used while training. Figure 5 shows the effect of different atmospheric compositions on the corrected surface properties for the “granite” validation case. Note that the purple curve in Figure 5 is identical to the green curve in Figure 2. Overall, the corrected values for the different atmospheres agree well in spectral regions away from the water

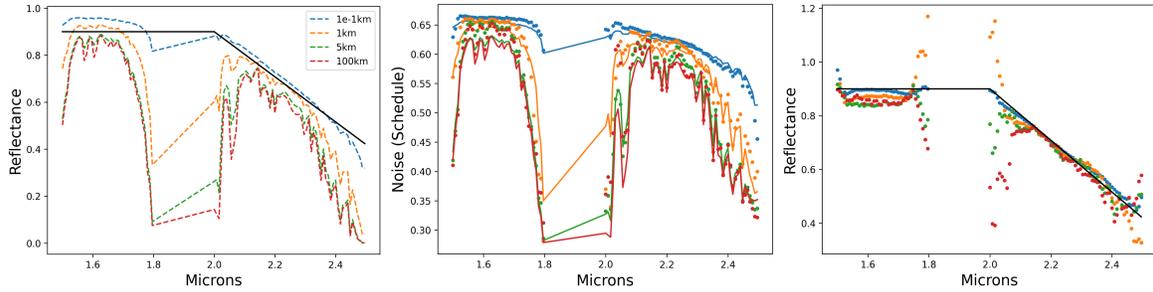


Figure 4. Network inputs (left) noise predictions (center) and surface reflectance retrieval (right) for the “farm” case at varying altitudes.

absorption bands. Interestingly, the “midlat summer” and “subarc winter” atmospheres correspond to the largest and smallest amounts of water, but in the regions nearest the H<sub>2</sub>O band absorption, “subarc winter” exhibits the worst performance. However, away from the water bands, the networks exhibits little to no difference when correcting to surface properties for different atmospheric conditions.

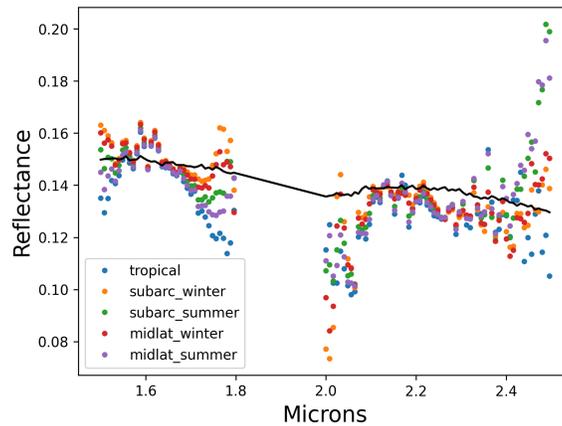


Figure 5. Truth (lines) and predicted (dots) surface reflectance for different atmospheres on the same surface pixel

Finally, we compare the trained network to the QUick Atmospheric Correction (QUAC) code, shown in Figure 6. A synthetic 1-D scene was constructed with 16 pixels each comprised of a unique Lambertian surface. An atmosphere was added based on the MODTRAN “midaltitude summer” composition. The QUAC results using the full 16-pixel scene as its database generally underpredict the true reflectance, which is likely attributed to a small database size and lower than ideal mean reflectance in the scene. Overall, the neural network estimates more accurate reflectance profiles than QUAC. However, there is more statistical scatter and in some cases, such as in black or specular surfaces, in which negative or larger than 1.0 values are sometimes predicted. There are no inherent constraints imposed on the network, but simple guardrails could be used to avoid nonphysical results. Regardless, the retrieval performance across a suite of surfaces shows good agreement with truth data.

### Extension to MWIR

Recently, we have considered the extension of the above approach for correction outside of the shortwave infrared which is dominated by solar photons. For the MWIR, the surface thermal emission becomes prominent and cannot be neglected. Thus the retrieval of the surface properties is not longer a single quantity and one aims to retrieve the spectral emissivity and the temperature at the surface. The spectral radiance,  $L(\lambda)$ , can be written with components of surface emission, downwelling, and upwelling terms as

$$L(\lambda) = B(T_{surf}, \lambda)\epsilon(\lambda)\tau(\lambda) + [1 - \epsilon(\lambda)]L^{\downarrow}(\lambda) + L^{\uparrow}(\lambda) \quad (4)$$

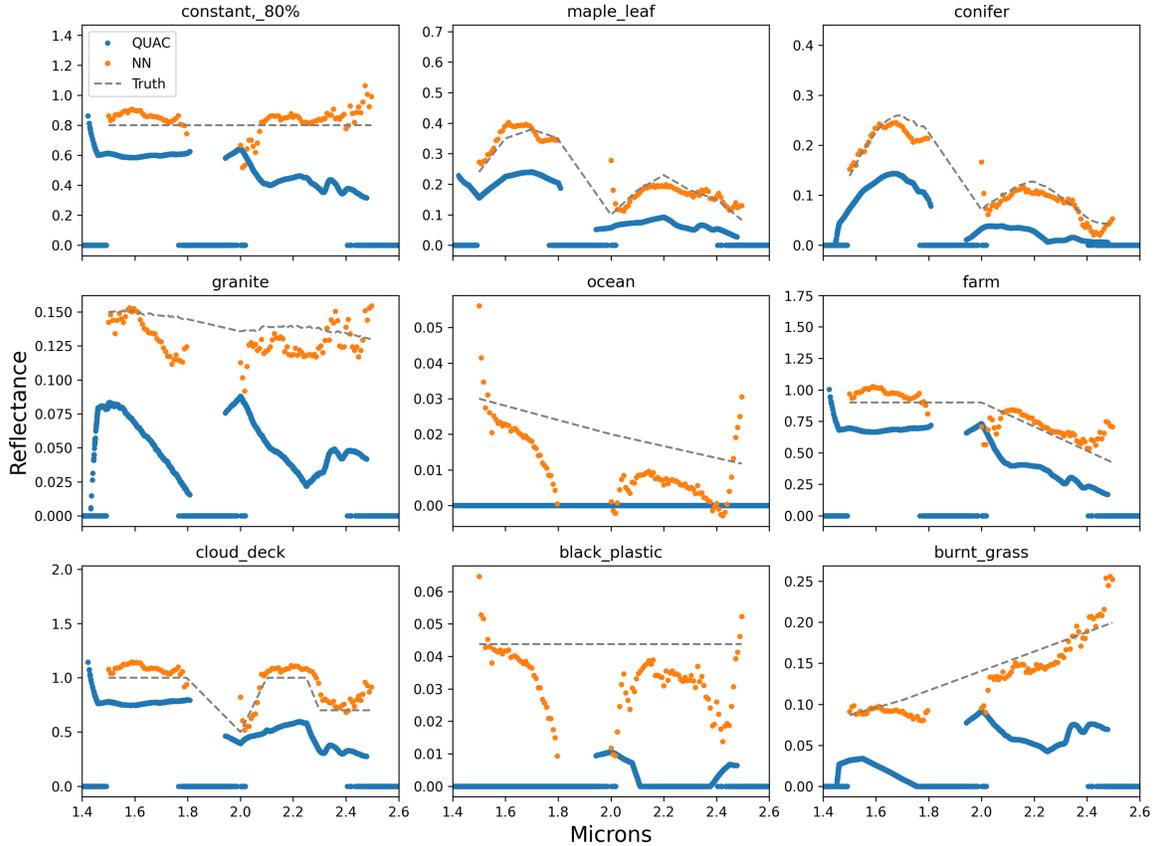


Figure 6. Truth (lines) and predicted (dots) surface reflectance for different atmospheres on the same surface pixel.

Here  $B(T, \lambda)$  is the Planck function,  $\epsilon(\lambda)$  is the spectral emissivity,  $\tau(\lambda)$  is the path transmittance, and  $L^\downarrow$  and  $L^\uparrow$  are the downwelling and upwelling to the sensor, respectively.

We seek the limit where the atmospheric “noise” is removed (i.e., an infinitesimal observation altitude), equivalent to  $\tau \rightarrow 1.0$ . In this limit, both  $L^\downarrow$  and  $L^\uparrow$  can be neglected and the true surface spectral emissivity is recovered as  $\epsilon(\lambda) = L(\lambda)/B(T_{surf}, \lambda)$ . This is the basis for the emissivity recovery diffusion network. Away from this limit, we view additional contributions from upwelling, downwelling, and surface emission as “noise”. The altitude specific quantity that is input along with  $n_s$  to the denoising diffusion model is then

$$Q_{alt} = \epsilon_{alt}(\lambda) = L(\lambda)/(B(T_{surf}, \lambda) + I_{sol}(\lambda)/\pi) \quad (5)$$

with normalization for solar photons from an overhead sun (important  $< \sim 4$  microns). Given this at-altitude quantity, the noise, and hence retrieved surface quantity, can be computed following the denoising diffusion model approach discussed above.

Of course, this input requires the surface temperature, which is not known *a priori*. As a first step, we assume the true surface temperature is known. This can be retrieved by conventional means, such as searching for blackbody pixels or spectral bins<sup>23</sup> or by an analogous neural network in future work.

### Emissivity Retrieval Network

As a proof-of-concept, we demonstrate a scenario to recover the spectral emissivity given knowledge of the surface temperature. A network was trained with the same architecture and hyperparameters as the SWIR example above. However, the training database is changed to also vary the surface temperature randomly for each pixel in the dataset between 273 K and 320 K with an overhead sun position. We omit other changes to

the database including the variable atmospheric composition and observation angles to reduce complexity in an initial proof-of-concept approach.

Figure 7 shows the results of retrieval using the trained network with the left side showing the effect of changing emissivity and the right showing the effect of surface temperature. Only small variations due to temperature differences are observed. On the left side, we evaluate four candidate pixels with varying emissivities and a constant temperature across each at 300 K. Performance is best at lower wavelengths where solar photons still make up a significant portion of the flux. At higher wavelengths, predictive performance is less robust. The near-blackbody pixel (orange) predicts an unphysical 1.5 spectral emissivity in the range of 4.5 to 5.0 microns. Similarly, the reflective pixel predicts near 0 or slightly negative values for much of the spectral range. These predictions are the result of noise predictions at the very edge of the normalization range.

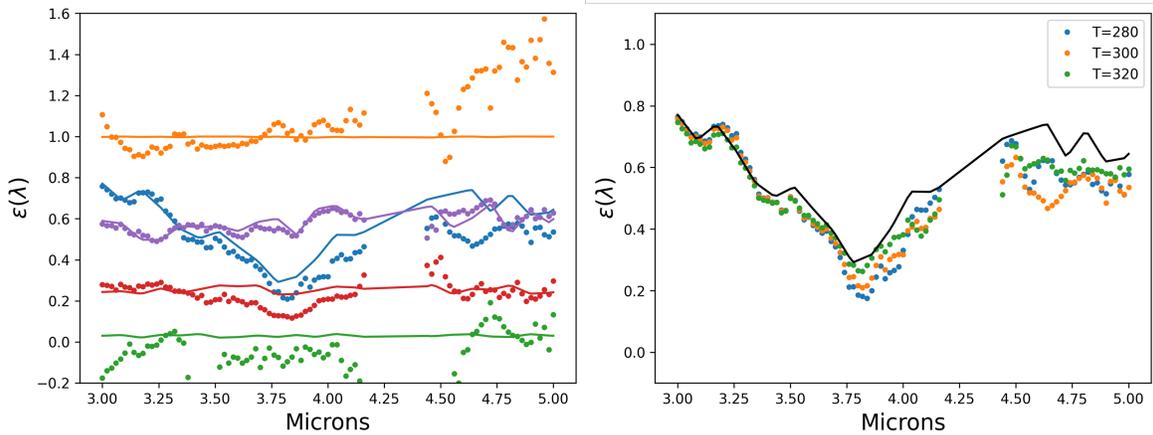


Figure 7. Truth (lines) and predicted (dots) emissivity for different pixels (colors) re-sampled from the training database (left) and variable temperature predictions for a single emissivity (right)

The 4.5 to 5.0 micron range remains challenging and predicted surface properties exhibit high variance. This is because there is little to no deviation between the input data in this region. We attribute any success of the network in this region to successive spectral convolutions that have learned to correlate the spectral emissivity at lower wavelengths to the 4.5-5.0 micron range.

If the downwelling term is known or an estimate is available, differentiation of the inputs in this spectral range can become more defined, as shown in Figure 8. In this case, the input data can be normalized as,

$$\epsilon_{alt}(\lambda) = \frac{L(\lambda) - L^{\downarrow}(\lambda)}{B(T_{surf}, \lambda) + \frac{I_{sol}(\lambda)}{\pi}} \quad (6)$$

However, this has the effect of eliminating input variation at the lower end of the spectral range (3.0 to 3.5 microns) where the surface emission is small and thus, once the downwelling is subtracted, the only significant contribution to the total radiance is the surface-independent upwelling. This tradeoff is unique to the MWIR, making it a notoriously difficult region for atmospheric compensation.

We retrained the model using the input as defined by Eq. 6. The resulting model exhibits better accuracy in the 4.5 to 5.0 micron range as expected with stronger variation in the input data with respect to the surface emissivity. However, results in the bluer regions of the spectrum are degraded and the model does a worse job at recovering the emissivity than when trained without the extra information of the downwelling. Future effort could focus on a redefining a model architecture with multiple inputs, such as inputting the downwelling term separately from the normalized total radiance.

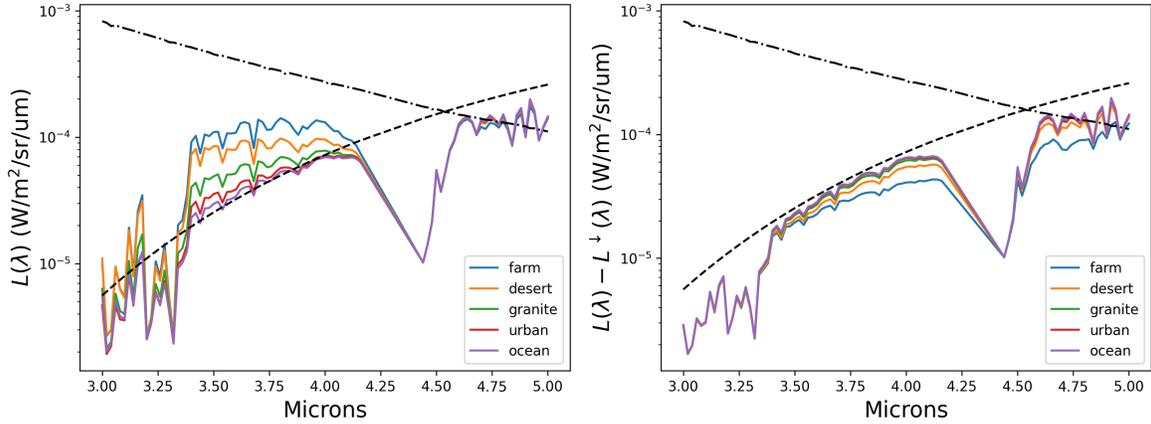


Figure 8. Pixel radiance (left) and radiance with downwelling subtracted (right), showing larger differences in the 4.5-5.0 micron region for validation cases (colors are surface type). Black lines indicate the Planck curve (black dashes) and the solar irradiance/ $\pi$  (black dash-dots).

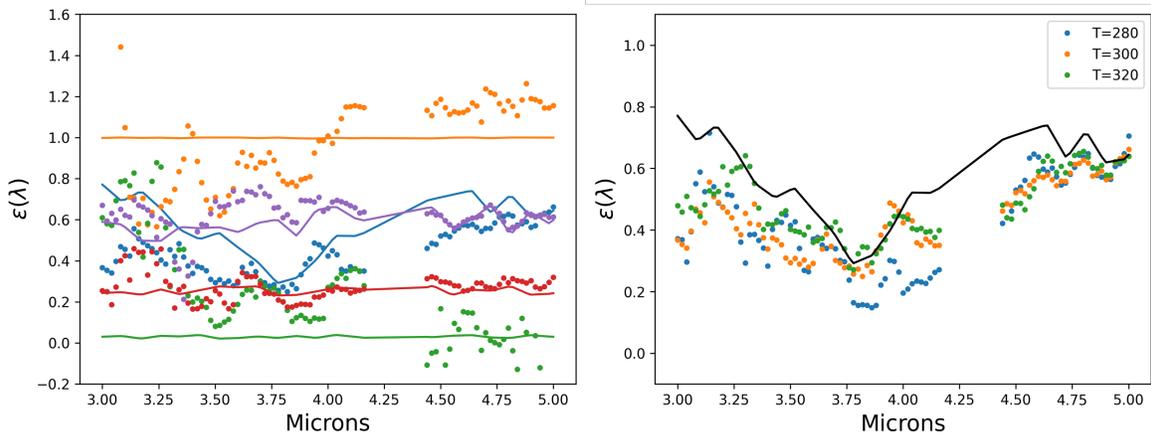


Figure 9. Truth (lines) and predicted (dots) emissivity for different pixels using the  $L^\downarrow$  subtracted input feature.

#### 4. CONCLUSIONS

The de-noising diffusion learning process was applied to atmospheric correction in the SWIR and MWIR. Synthetic surface data was generated for a large, self-supervised training dataset and radiometrically accurate at-sensor data was computed using MODTRAN. The trained model for the SWIR was able to accurately retrieve surface reflectance given the normalized at-sensor radiance for a broad range of conditions including different altitudes and atmospheric compositions. Results were compared to QUAC, an industry standard approach for in-scene Vis/SWIR atmospheric compensation and showed excellent performance. For the MWIR, results were spectrally dependent, and relied on input feature variation for certain parts of the spectrum. It was shown that accurate retrieval for all parts of the MWIR spectrum is possible, but likely requires multiple inputs.

Future work aims to address the long-wave infrared regions while also extending the MWIR approach for temperature retrieval. Other aims include verification of the trained SWIR model to alternative scenes, fine-tuning the approach for specific sensor bands, and establishing more robust computation across the full manifold of noise schedule space.

#### ACKNOWLEDGMENTS

The authors thank Timothy Perkins (SSI) for valuable discussion and insight throughout this effort and Spectral Sciences, Inc. for IRAD funding for this paper.

## REFERENCES

- [1] OpenAI, “Dall-e 2,” (2022).
- [2] Midjourney, “Midjourney user guide,” (2022).
- [3] Yang, L., Zhang, Z., Song, Y., Hong, S., Xu, R., Zhao, Y., Zhang, W., Cui, B., and Yang, M.-H., “Diffusion models: A comprehensive survey of methods and applications,” (2023).
- [4] Wu, K. E., Yang, K. K., van den Berg, R., Zou, J. Y., Lu, A. X., and Amini, A. P., “Protein structure generation via folding diffusion,” (2022).
- [5] Corso, G., Stärk, H., Jing, B., Barzilay, R., and Jaakkola, T., “Diffdock: Diffusion steps, twists, and turns for molecular docking,” (2023).
- [6] Stelter, D. and Sundberg, R., “Diffusion learning for atmospheric correction,” in [*Imaging Spectrometry XXVI: Applications, Sensors, and Processing*], **12688**, 1268801, SPIE (2023).
- [7] Ientilucci, E. J. and Adler-Golden, S., “Atmospheric compensation of hyperspectral data: An overview and review of in-scene and physics-based approaches,” *IEEE Geoscience and Remote Sensing Magazine* **7**(2), 31–50 (2019).
- [8] Berk, A., Adler-Golden, S., Ratkowski, A., Felde, G., Anderson, G., Hoke, M., Cooley, T., Chetwynd, J., Gardner, J., Matthew, M., et al., “Exploiting modtran radiation transport for atmospheric correction: The flaash algorithm,” in [*Proceedings of the Fifth International Conference on Information Fusion. FUSION 2002. (IEEE Cat. No. 02EX5997)*], **2**, 798–803, IEEE (2002).
- [9] Cooley, T., Anderson, G. P., Felde, G. W., Hoke, M. L., Ratkowski, A. J., Chetwynd, J. H., Gardner, J. A., Adler-Golden, S. M., Matthew, M. W., Berk, A., et al., “Flaash, a modtran4-based atmospheric correction algorithm, its application and validation,” in [*IEEE international geoscience and remote sensing symposium*], **3**, 1414–1418, IEEE (2002).
- [10] Miller, C. J., “Performance assessment of acorn atmospheric correction algorithm,” in [*Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery VIII*], **4725**, 438–449, SPIE (2002).
- [11] Thompson, D. R., Gao, B.-C., Green, R. O., Roberts, D. A., Dennison, P. E., and Lundeen, S. R., “Atmospheric correction for global mapping spectroscopy: Atrem advances for the hypsiri preparatory campaign,” *Remote Sensing of Environment* **167**, 64–77 (2015).
- [12] Bernstein, L. S., Adler-Golden, S. M., Jin, X., Gregor, B., and Sundberg, R. L., “Quick atmospheric correction (quac) code for vnir-swir spectral imagery: Algorithm details,” in [*2012 4th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*], 1–4 (2012).
- [13] Vermote, E., Roger, J.-C., Franch, B., and Skakun, S., “Lasrc (land surface reflectance code): Overview, application and validation using modis, viirs, landsat and sentinel 2 data’s,” in [*IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*], 8173–8176, IEEE (2018).
- [14] Basener, B. and Basener, A., “Gaussian process and deep learning atmospheric correction,” *Remote Sensing* **15**(3) (2023).
- [15] Koch, J., Forland, B., Doster, T., and Emerson, T., “A neural differential equation formulation for modeling atmospheric effects in hyperspectral images,” in [*Algorithms, Technologies, and Applications for Multispectral and Hyperspectral Imaging XXIX*], Velez-Reyes, M. and Messinger, D. W., eds., **12519**, 125190E, International Society for Optics and Photonics, SPIE (2023).
- [16] Shah, M., Raval, M. S., and Divakaran, S., “A deep learning perspective to atmospheric correction of satellite images,” in [*IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*], 346–349 (2022).
- [17] Rusia, P., Bhateja, Y., Misra, I., Moorthi, S. M., and Dhar, D., “An efficient machine learning approach for atmospheric correction,” *Journal of the Indian Society of Remote Sensing* **49**, 2539–2548 (2021).
- [18] Sun, J., Xu, F., Cervone, G., Gervais, M., Wauthier, C., and Salvador, M., “Automatic atmospheric correction for shortwave hyperspectral remote sensing data using a time-dependent deep neural network,” *ISPRS Journal of Photogrammetry and Remote Sensing* **174**, 117–131 (2021).
- [19] Berk, A., Conforti, P., Kennett, R., Perkins, T., Hawes, F., and Van Den Bosch, J., “Modtran® 6: A major upgrade of the modtran® radiative transfer code,” in [*2014 6th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*], 1–4, IEEE (2014).
- [20] Ramachandran, P., Zoph, B., and Le, Q. V., “Searching for activation functions,” (2017).

- [21] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I., "Attention is all you need," (2017).
- [22] Loshchilov, I. and Hutter, F., "Decoupled weight decay regularization," (2019).
- [23] Bernstein, L., Gelbord, J., Adler-Golden, S., Guler, N., Sundberg, R., and Conforti, P., "An improved in-scene atmospheric retrieval and correction algorithm for long-wavelength infrared hyperspectral imagery," in [*Imaging Spectrometry XXII: Applications, Sensors, and Processing*], Silny, J. F. and Ientilucci, E. J., eds., **10768**, 107680J, International Society for Optics and Photonics, SPIE (2018).