

Probabilistic Super Resolution for Mineral Spectroscopy

Alberto Candela,¹ David R. Thompson,² David Wettergreen,¹
Kerry Cawse-Nicholson,² Sven Geier,² Michael L. Eastwood,² Robert O. Green²

¹The Robotics Institute, Carnegie Mellon University, Pittsburgh, PA 15213, USA

²Jet Propulsion Laboratory, California Institute of Technology, CA 91109, USA
{albertoc, dsw}@cmu.edu, {David.R.Thompson, Kerry-Anne.Cawse-Nicholson, Sven.Geier, Michael.L.Eastwood, Robert.O.Green}@jpl.nasa.gov

Abstract

Earth and planetary sciences often rely upon the detailed examination of spectroscopic data for rock and mineral identification. This typically requires the collection of high resolution spectroscopic measurements. However, they tend to be scarce, as compared to low resolution spectra. This work addresses the problem of inferring high-resolution mineral spectroscopic measurements from low resolution observations using probability models. We present the Deep Gaussian Conditional Model, a neural network that performs probabilistic super resolution via maximum likelihood estimation. It also provides insight into learned correlations between measurements and spectroscopic features, allowing for the tractability and interpretability that scientists often require for mineral identification. Experiments using remote spectroscopic data demonstrate that our method compares favorably to other analogous probabilistic methods. Finally, we show and discuss how our method provides human-interpretable results, making it a compelling analysis tool for scientists.

Introduction

Earth and planetary sciences rely upon the analysis of spectroscopic data. Instruments such as the Compact Reconnaissance Imaging Spectrometer for Mars (CRISM) (Murchie et al. 2007) and the Thermal Emission Imaging System (THEMIS) (Christensen et al. 2004) have been vital for understanding the geology of Mars, as well as for studying climate and habitability implications (Bishop 2018). Another example is the Moon Mineralogy Mapper (M3) instrument, which produced the first mineralogical map of the Moon and also found water ice in the polar regions (Li et al. 2018).

The reason is that spectrometers measure different wavelengths of the electromagnetic spectrum and thus capture more information than can be seen with the eye. The measured signals are called *spectra* and contain *spectral features* that are used for rock and mineral identification (Clark 1999). This is because each material reflects, emits, or absorbs electromagnetic radiation in a unique way.

The detection of spectral features often requires high resolution spectroscopic measurements (Clark 1999). An ex-

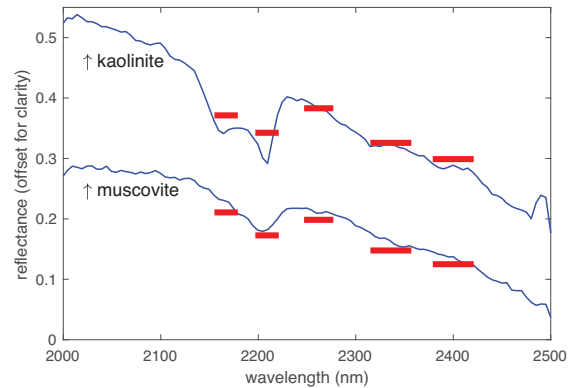


Figure 1: Shortwave infrared spectra from two different phyllosilicate minerals as measured by both low (red) and high (blue) resolution instruments. Low resolution measurements fail to capture important spectral features required for mineral identification.

ample of this phenomenon is described in Figure 1, showing how low resolution spectra may sometimes lose valuable information. Unfortunately, high resolution spectra are often scarce, whereas low resolution spectra are usually easier to obtain (Hubbard, Crowley, and Zimelman 2003; Zhang et al. 2016).

This work addresses the problem of inferring high-resolution spectra from low resolution measurements using probabilistic modeling. We present the Deep Conditional Gaussian Model (DCGM), a neural network architecture that emerges from the need for a spectroscopic analysis tool. It leverages ideas from different deep generative models, which have been highly successful for data generation and super resolution tasks (Goodfellow et al. 2014; van den Oord et al. 2016; Dahl, Norouzi, and Shlens 2017; Hu et al. 2018). But unlike most current methods, which use latent representations and are difficult to evaluate in probabilistic terms (Wu et al. 2017), our approach provides a means for log-likelihood estimation and for the inspection of learned statistical dependencies. This is particularly important in situations where a scientist needs to make informed

analytical decisions.

We validate our approach in a case study that consists of mineralogical investigations using remote spectroscopic data of Cuprite, Nevada; a well-studied test site for remote sensing algorithms (Swayze et al. 2014; Thompson et al. 2018). We first demonstrate that the DCGM compares favorably to other analogous probabilistic methods. We then show and discuss how the DCGM provides human-interpretable results that make it a compelling spectroscopic analysis tool.

Related Work

In the remote sensing community, many commonly used methods assume data follow Gaussian distributions, and also perform maximum likelihood estimation. For instance, one of the most popular algorithms is principal component analysis (PCA). Another one is the minimum noise fraction (MNF) transformation, which uses a noise-whitening operation before applying PCA in order to incorporate noise robustness (Green et al. 1988). Gaussian mixture models (GMM) are another classical probabilistic approach, and they have been used for tasks such as classification and unmixing of spectra (Zhou, Rangarajan, and Gader 2018).

Deep generative models also perform density estimation and have recently achieved impressive results in many different areas (Hu et al. 2018). In general, there are two learning approaches for generative models depending on whether they optimize a likelihood function or not.

There are deep generative models for which the likelihood is not explicitly defined (Wu et al. 2017). The main ones are generative adversarial networks (GANs) and variational autoencoders (VAEs). GANs consist of two neural networks contesting with each other in order to learn how to produce data, usually from a simple Gaussian random generator (Goodfellow et al. 2014). Despite their encouraging results in image generation, the density estimation process is highly opaque: it does not allow calculation of the likelihood of observed data, nor visualization of the learned statistical dependencies. VAEs (Kingma and Welling 2013), also known as Deep Latent Gaussian Models (DLGMs) (Rezende, Mohamed, and Wierstra 2014), have a more transparent probabilistic representation, which is learned by first encoding a data set into a latent space and then decoding it back into its original state. They assume the latent variables are normally-distributed. It is also simple to build conditional models connecting different types of inputs and outputs: VAEs have been used for spectral unmixing (Parente, Gemp, and Durugkar 2017; Borsoi, Imbiriba, and Bermudez 2019), and for the reconstruction of high resolution spectroscopic measurements from synthetic low resolution data (Candela, Thompson, and Wettergreen 2018). Although their probabilistic model is somewhat more tractable, it is not usually possible to compute the likelihood function analytically. That is why VAEs rely on variational approximations for training (hence their name). Additionally, the interpretation of the learned statistical dependencies is difficult because they lie in a hidden representation.

There are deep generative models that are explicitly driven by likelihood functions. An example are normalizing flows (Rezende and Mohamed 2015), which also work with

latent representations. They use change of variable transformations in order to generate complex models from simple probability distributions. In practice, they do not work well on high dimensional spaces, and the likelihood evaluation of unseen data is dependent on how invertible and tractable these series of transformations are. Autoregressive networks are a family of algorithms that have gained notable popularity for image generation (van den Oord et al. 2016) and super resolution (Dahl, Norouzi, and Shlens 2017) tasks. This is because they are scalable, relatively simple to train, and allow computation of log-likelihoods. They work by defining an arbitrary sequence of pixels and learning a series of conditional distributions. Consequently, they are unable to show correlations between any pair of non-sequential pixels. Conditional Gaussian distributions have been explored in the literature, but using spherical (isotropic) covariance matrices (Dahl, Norouzi, and Shlens 2017). Somewhat less constrained Gaussian models have been used for single-channel speech separation (Wang et al. 2017).

In summary, recent deep generative models have achieved remarkable results in data generation and super resolution, but have made little progress in producing conditional probabilistic models that are interpretable and easy to evaluate.

Method

Conditional Probability Model

Our super resolution method learns a conditional probability model that uses low resolution measurements in order to infer high resolution spectra. Let $x \in X \subset \mathbb{R}^m$ and $y \in Y \subset \mathbb{R}^n$ denote the low and high resolution spectra, respectively. We define the conditional probabilistic relationship between X and Y as $p_\theta(y|x)$. It is modeled using deep learning, where the network’s weights are defined as θ . This model assumes that the conditional probability can be represented with a multivariate Gaussian distribution, hence the name *Deep Conditional Gaussian Model* (DCGM). Formally, this means that:

$$p_\theta(y|x) \sim \mathcal{N}_n(\mu_y(x), \Sigma_y(x)), \quad (1)$$

where $\mu_y \in \mathbb{R}^n$ is the mean vector, and $\Sigma_y \in \mathbb{R}^{n \times n}$ is the covariance matrix. Note that both μ_y and Σ_y change as a function of x .

Maximum Likelihood Estimation

We rely on the well-known maximum likelihood estimation (MLE) method in order to estimate the parameters of the statistical model, that is, to tune the network’s weights θ . The training data set is comprised by pairs of inputs and ground-truth outputs $\mathcal{D} = \{(x_i, y_i^*)\}_{i=1}^N$. The network will try to maximize the following cumulative log-likelihood function:

$$\mathcal{L}(\theta|\mathcal{D}) = \sum_{(x, y^* \in \mathcal{D})} \log p_\theta(y^*|x). \quad (2)$$

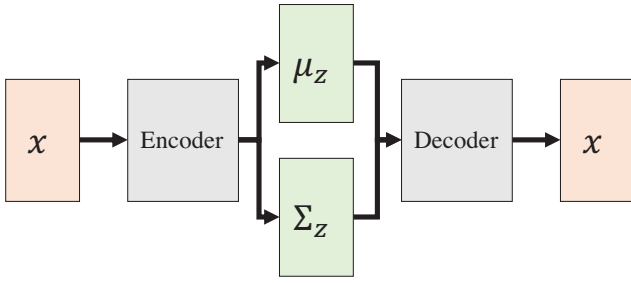


Figure 2: The architecture of the Variational Autoencoder, also known as the Deep Latent Gaussian Model.

The log-likelihood function for a multivariate Gaussian distribution is given by:

$$\log p_{\theta}(y|x) = -\frac{1}{2} [y - \mu_y(x)]^T \Sigma_y^{-1}(x) [y - \mu_y(x)] - \frac{1}{2} \log \{(2\pi)^n |\Sigma_y(x)|\}. \quad (3)$$

Intuitively, the model will try to approximate μ_y to y^* as closely as possible (i.e. regression), and simultaneously, Σ_y will reflect the model’s confidence in these predictions. After the learning process, high-confidence predictions should result in small values in Σ_y , and vice versa.

Architecture

The DCGM draws inspiration from the VAE (Kingma and Welling 2013) and the DLGM (Rezende, Mohamed, and Wierstra 2014), but differs in important ways. VAEs and DLGMs learn a probabilistic representation in a latent space Z by encoding it into a mean and a covariance (Figure 2). DCGMs, however, have a probabilistic representation that lies in the output layer (Figure 3). This enables our method to use an analytical likelihood function that allows for exact MLE, as opposed to VAEs and DLGMs, which require to learn via variational approximations.

VAEs typically work with diagonal covariances. DLGMs learn full covariance matrices by using a rank-1 approximation, which is just a slight improvement over a diagonal covariance (Rezende, Mohamed, and Wierstra 2014). We instead use a full-rank estimation that yields significantly better results, as will be shown later on. Additionally, Gaussian distributions in the output layer have been explored by other authors, but using either isotropic (Dahl, Norouzi, and Shlens 2017) or diagonal matrices (Wang et al. 2017).

Defining Σ_y as a full covariance matrix involves some practical challenges. First of all, Σ_y is a matrix of size $n \times n$, which could significantly increase the overall size of the output and the network. The network must also ensure Σ_y is always a symmetric positive definite matrix. The covariance matrix should also be numerically stable, since Equation 3 requires calculation of both its inverse and its determinant.

In order to face these challenges, we rely on the fact that the covariance can be represented with a Cholesky decomposition given by $\Sigma_y = LL^T$, where L is a lower triangular matrix. Numerical advantages related to this decomposition

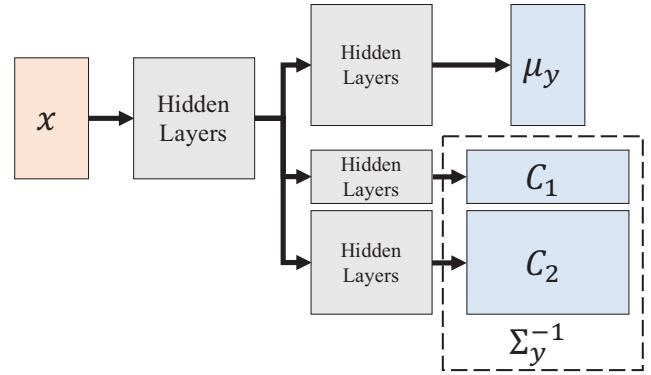


Figure 3: The architecture of the Deep Conditional Gaussian Model, based on an efficient Cholesky decomposition.

result in substantial computational simplifications. In order to avoid computing the inverse of the covariance matrix (first term of Equation 3), the model directly learns its Cholesky decomposition: $\Sigma_y^{-1} = CC^T$. Then, the determinant of the covariance matrix (second term of Equation 3) can be easily calculated with just the main diagonal of the decomposition: $|\Sigma_y| = (\prod_{i=1}^n C[i, i])^{-2}$.

For training and numerical stability purposes, the model decouples the Cholesky decomposition C into two parts: its main diagonal $C_1 \in \mathbb{R}_{>0}^n$, and the rest of the elements $C_2 \in \mathbb{R}^{n(n-1)/2}$. Figure 3 shows the corresponding architecture of the DCGM. The rationale is that C_1 must include positive numbers exclusively, and thus ensure both Σ_y and Σ_y^{-1} are truly positive definite. Furthermore, using a small threshold $\lambda > 0$ can add numerical stability.

The network can be pre-trained by fixing all of the values in C_2 to 0. Once training convergence is achieved, the model will learn a conditional probabilistic representation where all the channels in the output are considered to be independent. This translates into a μ_y that gives accurate *maximum a posteriori* predictions, and also a diagonal covariance matrix Σ_y that captures uncertainty in each channel individually. Then the network can be fine-tuned by “unfreezing” C_2 , allowing it to also learn correlations between channels.

Experiments

We present a case study that consists in the inference and reconstruction of high resolution mineral spectroscopy. The case study is focused on Cuprite, Nevada. It is a well-studied region of high mineralogical diversity that is amenable to remote sensing (Swayze et al. 2014), therefore making it an important test site for algorithms (Thompson et al. 2018).

Data Set

The data set is comprised by Cuprite data products from two imaging spectrometers: the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) (Fujisada et al. 1998), and the Airborne Visible Infrared Imaging Spectrometer Next Generation (AVIRIS-NG) (Hamlin et al. 2011). ASTER is a low resolution instrument, whereas

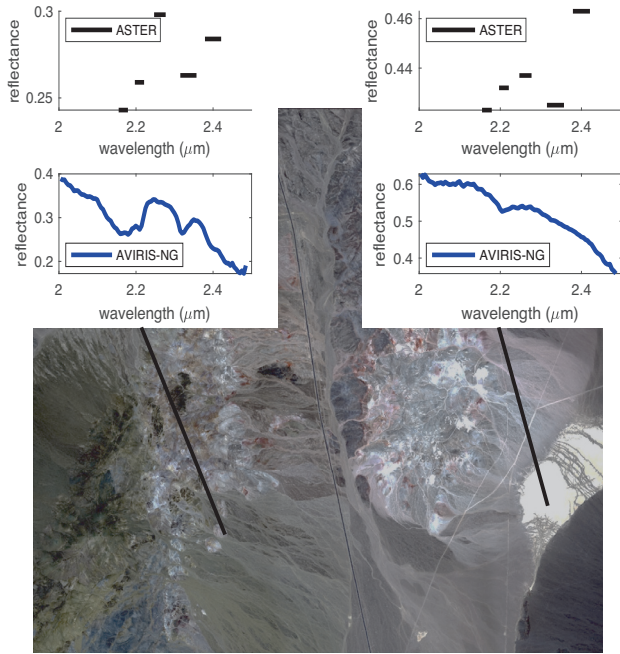


Figure 4: Example of remote spectroscopic measurements of Cuprite, Nevada, as seen by ASTER and AVIRIS-NG.

Table 1: Spectrometer measurement characteristics.

Instrument	Measurement	Spectral region	Channels
ASTER	Reflectance	2.0-2.5 μm	5
AVIRIS-NG	Reflectance	2.0-2.5 μm	85

AVIRIS-NG has a high spectral resolution. These data products were spatially aligned with ground control points that were selected manually. Figure 4 shows an example of a couple of representative spectroscopic measurements in the scene as seen by the two instruments. Table 1 summarizes the channels and wavelengths that were used for this study, which contain many of the diagnostic features needed for mineral identification at Cuprite (Swayze et al. 2014).

We perform a series of preprocessing operations on the data that seek to preserve spectral features (Clark 1999) and reduce the impact of misleading factors such as albedo and noise. Each spectrum is first scaled in the range [0,1] with min-max normalization. Afterwards, we apply the MNF transformation to add noise robustness (Green et al. 1988).

The data set consists of over 6×10^6 spectra recorded by each instrument. However, most of them are redundant and can lead to overfitting due to strong correlations between neighboring locations. Moreover, there are a few dozen mineral classes at Cuprite with imbalanced instances (Swayze et al. 2014). We solved these issues by first building a more balanced data set with the help of the mineral classification algorithm Tetracorder (Clark et al. 2003). We then eliminated spatially redundant measurements by randomly sampling 10,000 spectra from the scene. We divided them into three sets: training (5,000), validation (2,500), and test (2,500).

The test set was sampled from the south region of Cuprite, whereas the other two sets from the north region.

Experimental Setup

We make a quantitative comparison of the DCGM against the following four probabilistic baselines:

- **Gaussian (G)**: We concatenate the available and unavailable measurements, x and y respectively, and assume both follow one Gaussian distribution, i.e. $p(x, y) = \mathcal{N}_{m+n}(\mu, \Sigma)$. We then perform MLE by simply computing the joint sample mean and covariance. Finally, we predict the conditional mean $\mu_y(x) = \mu_{y|x}$ and covariance $\Sigma_y(x) = \Sigma_{y|x}$ using the standard formulas for conditional multivariate Gaussians (Eaton 1983).
- **Gaussian Mixture (GM)**: We learn a GMM for the joint distribution: $p(x, y) = \sum_{i=1}^K w^i \mathcal{N}_{m+n}(\mu^i, \Sigma^i)$. We then derive the conditional distribution, which also has the form of a GMM: $p(y|x) = \sum_{i=1}^K \bar{w}^i \mathcal{N}_n(\mu_{y|x}^i, \Sigma_{y|x}^i)$. The parameters of each Gaussian component $\mathcal{N}_n(\mu_{y|x}^i, \Sigma_{y|x}^i)$ are computed as in the previous baseline, and the weights are updated using Bayes' theorem (Gilardi, Bengio, and Kanevski 2002). We found that $K = 30$ components produced good results.
- **Unimodal Gaussian Mixture (UGM)**: This method is derived from the previous baseline. It estimates the overall mean $\mu_{y|x}$ and covariance $\Sigma_{y|x}$ of the conditional GMM after performing a prediction. These are given by:

$$\mu_{y|x} = \sum_{i=1}^K \bar{w}^i \mu_{y|x}^i, \quad (4)$$

$$\Sigma_{y|x} = \sum_{i=1}^K \bar{w}^i \left[\Sigma_{y|x}^i + \mu_{y|x}^i \mu_{y|x}^{i T} \right] - \mu_{y|x} \mu_{y|x}^T. \quad (5)$$

- **Variational Encoder (VE)**: It consists in the spectral reconstruction method in (Candela, Thompson, and Wettergreen 2018).

Furthermore, spherical, diagonal, and full covariance matrices were learned and tested for each of the previous algorithms. We evaluate the performance of these methods with the two following metrics:

- **Root mean squared error (RMSE)**: a measure of accuracy that uses *maximum a posteriori* predictions.
- **Negative log-likelihood (NLL)**: a measure of how probable the data are with respect to the learned model, defined as the negative of the log-likelihood (Equation 2).

The DCGM was implemented with the following network parameters (see Figure 3): one input layer, two sequential hidden layers, and then three parallel convolutional decoders for μ_y , C_1 , and C_2 . We used a dropout of 0.7 for all the hidden layers. The hidden layers and C_1 used a rectified linear unit (ReLU) activation function, whereas μ_y and C_1 used a sigmoid activation. We used a batch size of 4 and the Adam optimizer (Kingma and Ba 2014), with NLL as the loss function. The DCGM was implemented in Keras (Chollet, Allaire, and others 2017). As it is customary, we selected the

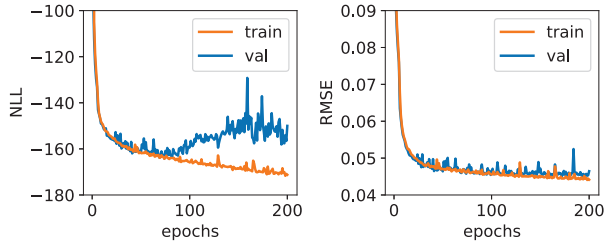


Figure 5: Training plots for the DCGM with a diagonal covariance.

Table 2: Average model performance (smaller is better). The asterisk indicates cases where paired t-tests ($p > 0.05$) are not statistically significant.

Spherical Covariance Matrix					
Metric	G	GM	UGM	VE	DCGM
RMSE	0.0938*	0.0685	0.0616	0.0510	0.0479
NLL	-79.6	-158.5	-129.2	-	-144.3
Diagonal Covariance Matrix					
Metric	G	GM	UGM	VE	DCGM
RMSE	0.0938*	0.0602	0.0565	0.0479	0.0458
NLL	-95.1	-174.7	-149.6	-	-162.2
Full Covariance Matrix					
Metric	G	GM	UGM	VE	DCGM
RMSE	0.0566	0.0538	0.0509	0.0490	0.0447
NLL	-272.8	-308.5	-282.8	-	-297.3

parameters that worked best with respect to the validation set in order to avoid overfitting. Figure 5 shows an example of the training process for a DCGM with a diagonal covariance matrix. Overall, we observed a good generalization. Given the reduced size of the overall data set, a 2.9 GHz Intel Quad-Core i7 laptop without a graphics processing unit (GPU) was sufficient to train the DCGM.

Results

The average performance of the different methods is shown in Table 2. We see that simpler covariance matrices lead to a deterioration in performance because they ignore valuable information. We observe that DCGM has the best performance in terms of RMSE because of its superior *maximum a posteriori* predictions. This is consistent across different types of covariance matrices. GM is the best in terms of NLL, probably because such low resolution inputs produce highly ambiguous relationships that are best modeled with a multimodal distribution (which needs to be tuned with the “right” number of components). However, DCGM still outperforms the other unimodal distributions. VE performs well in terms of RMSE, whereas its NLL is conventionally not defined. VE is consistent because of two reasons: it explicitly uses a RMSE loss function and its prior is a (spherical) standard Gaussian distribution. G, GM, and UGM perform poorly in terms of RMSE when using spherical and diagonal covariance matrices; but work well otherwise, especially GM and UGM because of the number of components they

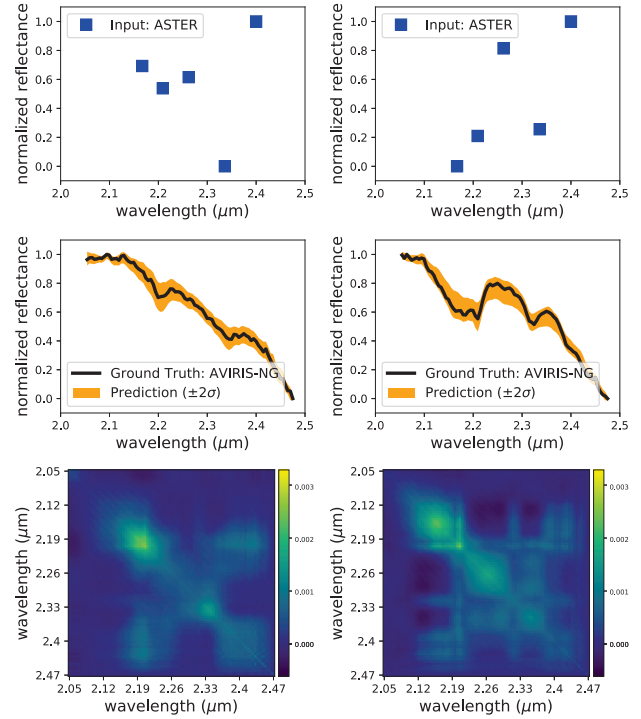


Figure 6: ASTER observations (top) being used to infer AVIRIS-NG measurements (middle) and their associated covariance matrices (bottom). This example shows two mineral mixtures at Cuprite: mica with calcite (left), and alunite with kaolinite (right).

use. As expected, GM is multimodal and thus outperforms UGM in terms of NLL, but has more unstable *maximum a posteriori* predictions that result in a higher RMSE.

Interpretability

This section discusses the potential of the DCGM as an analysis tool for geologists and spectroscopists via its human-interpretable results. Figure 6 shows how DCGM is using ASTER to infer AVIRIS-NG spectra of two common mineral mixtures at Cuprite: mica with calcite and kaolinite with alunite. There are clear differences between both instruments in terms of resolution, but also regarding offsets and noise. The predictions are highly accurate since the ground truth spectra are well within the error bars. But more importantly, the covariance matrices contain useful information about the mineral and spectral features. For instance, there are high covariances near the main diagonal, which is to be expected because adjacent channels are strongly correlated. In the first mineral mixture, mica has a characteristic absorption feature around $2.2 \mu\text{m}$, whereas calcite around $2.35 \mu\text{m}$. Many fractional abundances are possible, resulting in features and wavelengths with high variances. Since this is a common mineral mixture, the features are strongly correlated. In the alunite and kaolinite mixture, both constituent minerals have broader and more complex features with overlapping wavelengths. The identification of these minerals, as

well as the estimation of relative abundances, involves looking at subfeatures, hence the more complex correlations in the covariance matrix.

All of these notions are common for a scientist during the analysis and interpretation of spectroscopic data, and are certainly visible in the model. In summary, the DCGM not only has accurate predictions, but also provides human-comprehensible explanations regarding correlated features and feasible alternatives.

Conclusions and Future Work

This paper presents a deep generative model that performs super resolution for mineral spectroscopy. Unlike most recent methods, the DCGM truly performs log-likelihood estimation. And more importantly, it allows for the visualization and interpretation of learned statistical dependencies.

The quantitative results revealed that despite being a unimodal distribution, the DCGM generates accurate and often better predictions without the arduous need to estimate the “right” number of components in multimodal distributions such as GMMs. The qualitative results suggest that the DCGM shows promise as a scientific analysis tool.

In future work, we intend to use the predictive power of the DCGM for NASA’s Earth Surface Mineral Dust Source Investigation (EMIT), whose goal will be to study the role of atmospheric dust in Earth’s climate (Green 2019). Specifically, we plan to learn and visualize mineralogical and spectroscopic correlations between visible shortwave and thermal infrared measurements. Finally, we will explore the application of the DCGM to other remote mineralogical studies, as well as to other domains such as the oceanographic and agricultural sciences. When possible, we will also carry out user evaluation studies with expert scientists.

Acknowledgments

This research was supported by the National Science Foundation National Robotics Initiative Grant #IIS-1526667. We acknowledge support of the NASA Earth Science Division for the use of AVIRIS-NG data. A portion of this research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. Government sponsorship acknowledged.

References

Bishop, J. L. 2018. Chapter 3 - Remote Detection of Phyllosilicates on Mars and Implications for Climate and Habitability. In Cabrol, N. A., and Grin, E. A., eds., *From Habitability to Life on Mars*. Elsevier. 37 – 75.

Borsoi, R. A.; Imbiriba, T.; and Bermudez, J. C. M. 2019. Deep Generative Endmember Modeling: An Application to Unsupervised Spectral Unmixing. *arXiv preprint arXiv:1902.05528*.

Candela, A.; Thompson, D. R.; and Wettergreen, D. 2018. Automatic Experimental Design Using Deep Generative Models Of Orbital Data. In *International Symposium on Artificial Intelligence, Robotics and Automation in Space (iSAIRAS)*.

Chollet, F.; Allaire, J.; et al. 2017. R interface to keras. <https://github.com/rstudio/keras>.

Christensen, P. R.; Jakosky, B. M.; Kieffer, H. H.; Malin, M. C.; Mcswen, H. Y.; Nealon, K.; Mehall, G. L.; Silverman, S. H.; Ferry, S.; Caplinger, M.; and Ravine, M. 2004. The thermal emission imaging system (themis) for the mars 2001 odyssey mission. *Space Science Reviews* 110:85–130.

Clark, R. N.; Swayze, G. A.; Livo, K. E.; Kokaly, R. F.; Sutley, S. J.; Dalton, B.; McDougal, R. R.; and Gent, C. A. 2003. Imaging spectroscopy: Earth and planetary remote sensing with the USGS Tetracorder and expert systems. *Journal of Geophysical Research* 108(E12):5131.

Clark, R. N. 1999. Spectroscopy of rocks and minerals, and principles of spectroscopy. In Rencz, A. N., and Ryerson, R. A., eds., *Manual of Remote Sensing, Volume 3, Remote Sensing for the Earth Sciences*. New York: John Wiley and Sons, 3rd edition. chapter 1, 3—58.

Dahl, R.; Norouzi, M.; and Shlens, J. 2017. Pixel Recursive Super Resolution. In *Proceedings of the IEEE International Conference on Computer Vision*, 5449–5458.

Eaton, M. L. 1983. *Multivariate Statistics: a Vector Space Approach*. 605 THIRD AVE., NEW YORK, NY 10158, USA: John Wiley & Sons, 1st edition.

Fujisada, H.; Sakuma, F.; Ono, A.; and Kudoh, M. 1998. Design and preflight performance of aster instrument protoflight model. *IEEE Transactions on Geoscience and Remote Sensing* 36(4):1152–1160.

Gilardi, N.; Bengio, S.; and Kanevski, M. 2002. Conditional gaussian mixture models for environmental risk mapping. In *Proceedings of the 12th IEEE Workshop on Neural Networks for Signal Processing*, 777–786.

Goodfellow, I. J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; and Bengio, Y. 2014. Generative Adversarial Nets. In Ghahramani, Z.; Welling, M.; Cortes, C.; Lawrence, N. D.; and Weinberger, K. Q., eds., *Advances in Neural Information Processing Systems 27*, 2672—2680. Curran Associates, Inc.

Green, A.; Berman, M.; Switzer, P.; and Craig, M. 1988. A transformation for ordering multispectral data in terms of image quality with implications for noise removal. *IEEE Transactions on Geoscience and Remote Sensing* 26(1):65–74.

Green, R. 2019. The earth surface mineral dust source investigation (EMIT) using new imaging spectroscopy measurements from space. In *International Geoscience and Remote Sensing Symposium (IGARSS)*.

Hamlin, L.; Green, R. O.; Mouroulis, P.; Eastwood, M.; Wilson, D.; Dudik, M.; and Paine, C. 2011. Imaging spectrometer science measurements for terrestrial ecology: AVIRIS and new developments. *IEEE Aerospace Conference Proceedings* 1–7.

Hu, Z.; Yang, Z.; Salakhutdinov, R.; and Xing, E. P. 2018. On Unifying Deep Generative Models. In *International Conference on Learning Representations*.

Hubbard, B. E.; Crowley, J. K.; and Zimbelman, D. R. 2003. Comparative alteration mineral mapping using visible to

- shortwave infrared (0.4-2.4 μm) hyperion, ali, and aster imagery. *IEEE Transactions on Geoscience and Remote Sensing* 41(6):1401–1410.
- Kingma, D. P., and Ba, J. 2014. Adam: A Method for Stochastic Optimization. *arXiv preprint arXiv:1412.6980* 1–15.
- Kingma, D. P., and Welling, M. 2013. Auto-Encoding Variational Bayes. *arXiv preprint arXiv:1312.6114*.
- Li, S.; Lucey, P. G.; Milliken, R. E.; Hayne, P. O.; Fisher, E.; Williams, J.; Hurley, D. M.; and Elphic, R. C. 2018. Direct evidence of surface exposed water ice in the lunar polar regions. *Proceedings of the National Academy of Sciences* 115(36).
- Murchie, S.; Arvidson, R.; Bedini, P.; Beisser, K.; Bibring, J. P.; Bishop, J.; Boldt, J.; Cavender, P.; Choo, T.; Clancy, R. T.; Darlington, E. H.; Des Marais, D.; Espiritu, R.; Fort, D.; Green, R.; Guinness, E.; Hayes, J.; Hash, C.; Hefferman, K.; Hemmler, J.; Heyler, G.; Humm, D.; Hutcheson, J.; Izenberg, N.; Lee, R.; Lees, J.; Lohr, D.; Malaret, E.; Martin, T.; McGovern, J. A.; McGuire, P.; Morris, R.; Mustard, J.; Pelkey, S.; Rhodes, E.; Robinson, M.; Roush, T.; Schaefer, E.; Seagrave, G.; Seelos, F.; Silverglate, P.; Slavney, S.; Smith, M.; Shyong, W. J.; Strohbahn, K.; Taylor, H.; Thompson, P.; Tossman, B.; Wirzburger, M.; and Wolff, M. 2007. Compact Connaissance Imaging Spectrometer for Mars (CRISM) on Mars Reconnaissance Orbiter (MRO). *Journal of Geophysical Research E: Planets* 112(5):1–57.
- Parente, M.; Gemp, I.; and Durugkar, I. 2017. Unmixing in the presence of nuisances with deep generative models. In *International Geoscience and Remote Sensing Symposium (IGARSS)*, 5189–5192.
- Rezende, D. J., and Mohamed, S. 2015. Variational Inference with Normalizing Flows. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37, 1530—1538.
- Rezende, D. J.; Mohamed, S.; and Wierstra, D. 2014. Stochastic backpropagation and approximate inference in deep generative models. In *Proceedings of the 31st International Conference on International Conference on Machine Learning - Volume 32, ICML'14, II-1278-II-1286*. JMLR.org.
- Swayze, G. A.; Clark, R. N.; Goetz, A. F. H.; Livo, K. E.; Breit, G. N.; Kruse, F. A.; Sutley, S. J.; Snee, L. W.; Lowers, H. A.; Post, J. L.; Stoffregen, R. E.; and Ashley, R. P. 2014. Mapping Advanced Argillic Alteration at Cuprite, Nevada, Using Imaging Spectroscopy. *Economic Geology* 109(5):1179–1221.
- Thompson, D. R.; Candela, A.; Wettergreen, D. S.; Noe Dobra, E.; Swayze, G. A.; Clark, R. N.; and Greenberger, R. 2018. Spatial Spectroscopic Models for Remote Exploration. *Astrobiology* 18(7):934–954.
- van den Oord, A.; Kalchbrenner, N.; Vinyals, O.; Espenholt, L.; Graves, A.; and Kavukcuoglu, K. 2016. Conditional Image Generation with PixelCNN Decoders. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, 4797—4805.
- Wang, Y.; Du, J.; Dai, L. R.; and Lee, C. H. 2017. A maximum likelihood approach to deep neural network based nonlinear spectral mapping for single-channel speech separation. In *Proceedings of the Annual Conference of the International Speech Communication Association, INTER-SPEECH*, 1178–1182.
- Wu, Y.; Burda, Y.; Salakhutdinov, R.; and Grosse, R. 2017. On the Quantitative Analysis of Decoder-Based Generative Models. In *International Conference on Learning Representations*.
- Zhang, T.; Yi, G.; Li, H.; Wang, Z.; Tang, J.; Zhong, K.; Li, Y.; Wang, Q.; and Bie, X. 2016. Integrating data of ASTER and Landsat-8 OLI (AO) for hydrothermal alteration mineral mapping in duolong porphyry cu-au deposit, Tibetan Plateau, China. *Remote Sensing* 8(11).
- Zhou, Y.; Rangarajan, A.; and Gader, P. D. 2018. A Gaussian mixture model representation of endmember variability in hyperspectral unmixing. *IEEE Transactions on Image Processing* 27(5):2242–2256.