Metric	Interpolation Mean (SD)	Extrapolation Mean (SD)
RMSE ($\times 10^{-2}$)	0.04 (0.01)	0.15 (0.05)
R^2	0.91 (0.07)	-0.39 (0.70)
Coverage	0.95 (0.07)	0.89 (0.04)

Table 1: Generative model performance measured on the test set via RMSE and R^2 of the mean predicted spectrum and coverage of a 95% uncertainty interval, split into interpolation and extrapolation regimes.

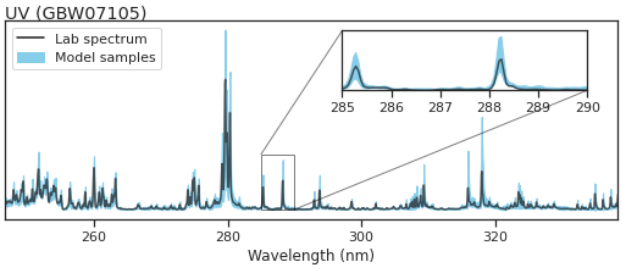

Figure 2: Generated spectra samples compared to a test set spectrum (target GBW07105) for the UV spectrometer, conditional on the true composition. Inset panel zooms in to show detail of an Si spectral line near 288.2 nm.

on the true composition. Table 1 gives the mean and standard deviation of the three performance metrics across all test set spectra, divided into interpolation and extrapolation regimes. In the interpolation regime, we note that the PCA decomposition alone incurs average RMSE near 0.0001, so our model introduces some additional error on average in reconstructing the spectra, but generally the R^2 is high. In the extrapolation regime, the RMSE and R^2 indicate worse performance (as expected, because no training data was seen in this area of input space). For assessing coverage, we used the 2.5% and 97.5% quantiles to obtain uncertainty intervals (nominal 95% coverage). In the interpolation regime, we achieve nominal coverage, with only a slightly lower coverage in the extrapolation regime. This demonstrates an important property of the model: while predictions may be inaccurate when extrapolating, the uncertainty intervals expand and can therefore still contain the true data values.

Fig 2 shows generated spectral samples for a given input composition (SiO_2 oxide weight percent 44.6%) with a test set spectrum corresponding to that composition shown in black; for simplicity, we show results only for the UV spectrometer. The generated spectral samples appear to capture the general shape of the true spectrum. Zooming in on a key Si spectral line near 288.2 nm, we see some variation across model samples, but the peak appears in all samples.

Inferring Generating Parameters

For a set of eighteen test set spectra representing distinct targets, we estimate the composition value corresponding to the maximum likelihood estimator (MLE) and generate asymmetric 95% confidence intervals for the MLEs using likelihood ratio intervals. Fig 3 shows the MLEs (blue horizontal

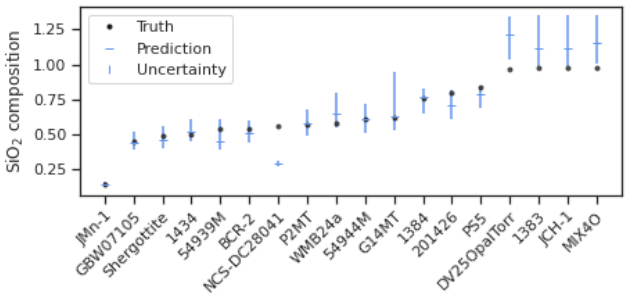

Figure 3: For the test set materials (named along the horizontal axis), we use the SNFGP likelihood to infer the composition with uncertainty. Many uncertainty intervals cover the true values. The rightmost four materials are the 'extrapolation regime' test set; as expected, the predictions are less accurate, but with wider uncertainty intervals.

bars) with the uncertainty intervals (blue vertical bars) compared to the true compositions (black dots) for the different test set materials. The intervals cover the true composition value for most of the materials, and in many cases the intervals are fairly tight. We note that the NCS-DC28041 material appears to be an outlier in principal components space (prior to learning the SNFGP), indicating that this sample may require further investigation. The four materials with the largest compositions correspond to the 'extrapolation regime'; we note that while while the predictions are more biased, the uncertainty intervals are larger.

Conclusions

In this work, we have presented SNFGP: a novel generative model that combines dimension reduction, normalizing flows, and Gaussian process regression. SNFGP conditions on input values to generate complex, structured, high-dimensional outputs. We demonstrate the model on LIBS spectra from the ChemCam instrument, where the model generates realistic spectra conditional on an input composition and provides a principled way to quantify uncertainty in predictions of the input composition given a new spectral observation. Importantly, we demonstrate that the SNFGP model has good properties when extrapolating from the training data, a property not shared by many machine learning models. In future work, we plan to compare our method to related methods such as the GPVAE in terms of performance and computational complexity and to expand the application to include ChemCam data from Mars (including modeling the Earth/Mars data discrepancy).

Acknowledgments

This project was supported by the Laboratory Directed Research and Development program of Los Alamos National Laboratory under project number LDRD-20210043DR.

Ducimus quas ratione nemo tenetur totam, vero iure molestiae doloremque consequuntur, suscipit quod odio, aspernatur officiis necessitatibus accusamus libero optio eligendi dolores similique, sit temporibus ut vel placeat delen-

iti.Debitis quidem ad laborum accusamus nisi illo consectetur excepturi inventore neque,