SNFGP Model Specification

The SNFGP model is specified through the conditional density $p_{Y|X}(\mathbf{y}|\mathbf{x})$ which is written in terms of GP marginal likelihoods via the structured normalizing flow as follows. We presume that the K-dimensional normalizing flow latent space can be modeled as a set of K independent univariate-output GP models $p_{Z_k|X}(z_k|\mathbf{x})$:

$$p_{Z|X}(\mathbf{z}|\mathbf{x}) = p_{Z|X}(z_1, ..., z_K|\mathbf{x}) = \prod_{k=1}^{K} p_{Z_k|X}(z_k|\mathbf{x}). \quad (7)$$

Recalling that $\mathbf{w} = g_\theta(\mathbf{y}) = \mathbf{W}^T\mathbf{y}$ is the dimension reduction transform and $\mathbf{z} = f_\phi(\mathbf{w})$ the normalizing flow, we obtain the density of $\mathbf{y}$ conditional on inputs $\mathbf{x}$ by application of the change of variables formula Eq 5:

$$p_{Y|X}(\mathbf{y}|\mathbf{x}) = p_{Z|X}(\mathbf{z}|\mathbf{x}) \left| \frac{\partial f_\phi(\mathbf{w})}{\partial \mathbf{w}} \right| \frac{1}{\sqrt{\prod_{k=1}^{K} w_{kk}}} \quad (8)$$

where $\mathbf{w} = g_\theta(\mathbf{y})$. Eq 8 can then be maximized with respect to parameters γ_k for each individual Gaussian process, in addition to the NF parameters ϕ.

Implementation Details

In this work, we first estimate g_θ via PCA with $K = 15$ components (capturing over 96% of the variance in the LIBS training set). We then learn the NF parameters ϕ along with the GP parameters γ by maximizing the transformed GP marginal likelihood in Eq 8. For the normalizing flow f_ϕ, we use a RealNVP architecture with six coupling layers. Note that $p_{Z|X}$ involves all of the data examples for evaluation; with a computational complexity of $O(N^3)$, the GP can become prohibitive for large data sets and will dominate the computational cost (since RealNVP architectures are designed for efficient forward, backward, and Jacobian evaluation). However, recent work suggests that mini-batch training for Gaussian process models is computationally efficient and accurate (**?**), so we use batch sizes of 512 data points for each update (via Adam optimizer, learning rate 0.0005). We implement the model in Pytorch (**?**) with some custom layer functions from `pytorch_flows` (https://github.com/ikostrikov/pytorch-flows). We randomly select distinct materials for the training and test sets, but also select a set of 'extrapolation regime' test materials by including all materials with SiO_2 composition greater than 0.9 in the test set and using only materials with composition less than 0.8 in the training set. This results in 2,109 unique spectra corresponding to 422 unique materials in the training set, 150 spectra corresponding to 30 unique materials in the validation set, and 18 unique materials (disjoint from the training/validation sets) with a total of 90 spectra in the test set.

Model Evaluation

To demonstrate goodness of fit of the model, we investigate the Gaussian process predictive accuracy in $\mathbf{z}$ space on both the training and test data to determine whether the model residuals indicate lack of fit. To assess generative model performance, we sample from the fitted Gaussian process conditional on the true test set composition values, then propagate the samples through the inverse normalizing flow and pseudo-inverse dimension reduction functions to obtain samples in the LIBS spectral space. Across wavelengths for each spectrum, we compute the root mean squared error (RMSE), the R^2, and the coverage. RMSE and R^2 measure how well the mean predicted spectra matches the true spectrum, while the coverage metric constructs nominal $(1-\alpha)$ uncertainty intervals using the $[\alpha/2,\ 1-\alpha/2]$ quantiles of the predicted spectral samples and estimates the proportion of wavelengths for which the true spectrum fell within the uncertainty intervals. Finally, we ask at whether, given a new test spectrum, its latent representation $\mathbf{z}^*$ can be used to infer the corresponding composition $\mathbf{x}^*$ using grid search to find the maximum of the likelihood function (Eq 8) with respect to unknown $\mathbf{x}^*$ based on the GP predictive distribution and likelihood intervals (**?**) to describe uncertainty.

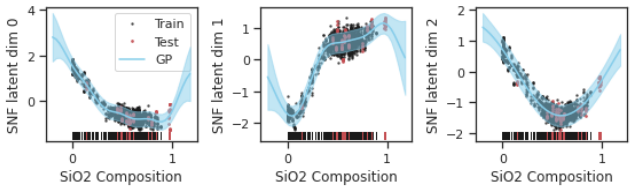

Figure 1: Gaussian process predictive distribution (mean line with shading for 95% uncertainty interval) as a function of SiO_2 composition for three of the 15 normalizing flow latent dimensions. The predictive mean describes the patterns in the data accurately and the uncertainty intervals appear to capture the majority of the data.

Results

Goodness of Fit

First, we investigate how well the GP model describes the data in the latent space. Fig 1 shows, for three of the 15 total latent dimensions, scatter plots of the SiO_2 composition against the latent representation value. Black points represent training data while red points represent test data; rug plots along the horizontal axis indicate the distributions of test and training data, including the 'extrapolation regime' test set with $SiO_2 > 0.9$. The GP predictive distribution is shown as a function of composition, with the mean function prediction as a blue line and 95% uncertainty intervals as shaded blue areas. It appears that the GP models adequately capture variation in the latent space conditional on the composition, and the uncertainty intervals capture the data distribution well. Some of the inherent variability in the data (reflected by the spread around the mean prediction) comes from uncontrollable sources (shot-to-shot variations for the same target) while other variability may come from the unmodeled influence of other elemental compositions.

Generative Model Performance

Across the test set, we evaluate how well the generative model captures the characteristics of the spectra conditional

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.

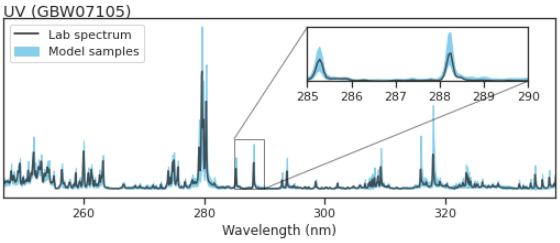

Figure 2: Generated spectra samples compared to a test set spectrum (target GBW07105) for the UV spectrometer, conditional on the true composition. Inset panel zooms in to show detail of an Si spectral line near 288.2 nm.

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 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.

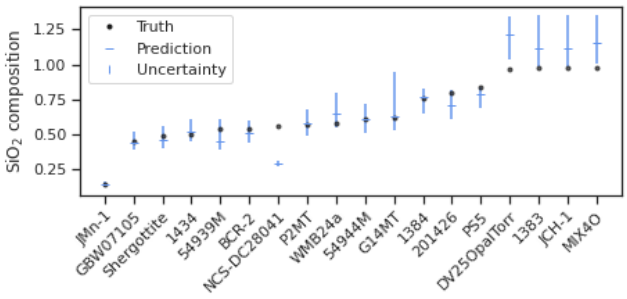

Figure 3: For the test set materials (named along the horizontal axis), we use the SNFGP likelihood to infer the composition with uncertainty. Many uncertainty intervals cover the true values. The rightmost four materials are the 'extrapolation regime' test set; as expected, the predictions are less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

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