Spectral Unmixing and Design Requirements for a low-cost Crop Residue Cover Mapping Nanosatellite

Zoe Augspach[1,3], Andrei Akopian[1,3], Zara Graham[1,3], Fedor Pavlov[1,3], Ege Artan[2,3]*, Noam Tal-Siegel[1,3], Kemal Emirhan Uygun[2,3]

[1] Faculty of Arts and Science, University of Toronto, 100 St. George Street Toronto, ON M5S 3G3, Canada - (zoe.augspach, zara.graham, andrei.akopian, fedor.pavlov, noam.talsiegel)@mail.utoronto.ca
[2] University of Toronto Scarborough, Univ. of Toronto, 1265 Military Trail Toronto, ON M1C 1A4, Canada - (ege.artan, kemalemirhan.uygun)@mail.utoronto.ca
[3] Space Systems, University of Toronto Aerospace Team, 10 King’s College Road,D B740 Sandford Fleming Building, Toronto, ON M5S 3G8, Canada

Keywords: Crop residue cover mapping, spectral unmixing, design requirements, design validation, nanosatellite.

Abstract

FINCH is a student-led satellite mission whose novel sensor and cost effective form seek to provide crop residue mapping at a much lower cost and aid in smart-agriculture initiatives. To achieve this, crop residue must be accurately quantified using the limited reflectance range of $900\,\mathrm{nm}$ to $1700\,\mathrm{nm}$. Hence, novel unmixing methods must be developed. Two datasets were evaluated: a laboratory-acquired dataset and a simulated, atmospherically propagated dataset. Multiple unmixing methods were tested, including Linear Regression, a Bayesian Linear Dirichlet model, K-Nearest Neighbors, Random Forest, and deep learning approaches such as a Multi-Layered Perceptron. Strong performance was achieved on the laboratory dataset, with the Multi-Layered Perceptron achieving an R^2 for crop residue of 0.8436, total R^2 of 0.8935, and an RMSE of 0.0909 when plotting true to predicted abundances, demonstrating the feasibility of accurate unmixing in controlled conditions. However, performance decreased substantially on the atmospherically propagated dataset, likely due to nonlinearities and other stark differences between datasets that limit transfer learning. These findings indicate that while the lab results are highly promising, additional atmospheric measurements and model adaptations are necessary to achieve full confidence in FINCH’s predictions. Further testing and validation will be critical to establish robustness and guide the development of operational unmixing methods for determination of optical design and imaging requirements.

1. Introduction

The University of Toronto Aerospace Team Space Systems Division is currently developing FINCH - a Field Imaging Nanosatellite for Crop Residue Hyperspectral Mapping. The team is a student-led undergraduate design team, driven to create a financially accessible small satellite for crop residue cover mapping.

Crop residue cover mapping is a fixture in sustainable agriculture on a farm scale, as its impacts include carbon and nutrient cycles, and is an important indicator of tillage intensity, soil erosion, drought potential, ecosystem disturbances, and wildfire danger (Dennison et al., 2023).

However, to create a financially accessible mapping technique to accrue crop residue cover information on demand remains a challenge for farms. Common methods, such as unmanned aerial vehicles, are expensive, time-consuming, and require specialized labor. Other satellites such as EnMAP are not readily accessible for on-demand farm-level use. Using a specialized 3U satellite, FINCH aims to fill this gap by producing a proof of concept to demonstrate the feasibility of a more affordable alternative for farms to map crop residue cover by undertaking hyperspectral imaging in SWIR1 ($900\,\mathrm{nm}$ to $1700\,\mathrm{nm}$).

To keep FINCH financially accessible, cost-optimized reductions in the satellites size, power, and cooling result in a reduced spectral range and spatial resolution during imaging. Typical imaging ranges take advantage of the lignin-cellulose absorption band at $2100\,\mathrm{nm}$ to spectrally unmix crop residue from a mixed pixel, which is not covered in SWIR1 (Daughtry, 2001). Hence, we aim to develop an image analysis pipeline and data-driven spectral unmixing methods capable of quantifying crop residue abundance in the SWIR1 range.

* Corresponding author

2. Problem Statement

To develop a data analysis pipeline, unmixing methods must be tested and altered to achieve accurate crop residue estimations from satellite images. This will inform FINCH’s satellite design requirements such that the acquired images can be analyzed to achieve a target accuracy.

The pipeline’s output is the fractional abundance of crop residue, while its input is the restricted wavelength range between $900\,\mathrm{nm}$ and $1700\,\mathrm{nm}$ with a spectral resolution of $10\,\mathrm{nm}$. The reflectance will be affected by atmospheric effects and noise.

According to the linear mixing model, when atmospheric events are not considered, the reflectance can be described as the sum of the reflectance of the individual endmembers times their fractional abundances (Keshava and Mustard, 2002). Mathematically, a mixed pixel’s reflectance (ρ) can be written as:

$$\rho = \sum_{i=1}^{n} a_i \cdot \rho_i + \mathcal{N}(0, \mathbf{I}) \tag{1}$$

where n : number of endmembers in the scene
a_i : fractional abundance of the ith endmember
ρ_i : reflectance of the ith endmember
$\mathcal{N}(0, \mathbf{I})$: additive white Gaussian noise

To maintain physicality of the mixing, the following are imposed on abundances:

$$\sum_{i=1}^{n} a_i = 1, \quad a_i \geq 0 \quad \forall i \in \{i\}_{i=1}^{n} \tag{2}$$

Accuracy was defined as the coefficient of determination (R^2) for a true vs. predicted crop residue abundance plot. The target accuracy was set at $R^2 \geq 0.7$. This target accuracy was determined based on previous work done (Pacheco and McNairn, 2010). The problem can thus be reduced to a regression problem with an expected linear relationship between the predictor and target variables.

3. Approach and Designed Pipelines

The approach for taking to analyzing satellite images is shown in Figure 1. Images will be acquired under conditions that meet the identified design requirements, preprocessed by geolocation and noise removal, then atmospherically inverted and unmixed.

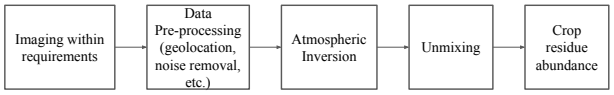

Figure 1. Image analysis pipeline for FINCH’s hyperspectral images

The imaging and design requirements, like signal-to-noise ratio (SNR), spatial resolution, and sun-sensor geometry, must be set prior to launching the satellite. To determine the exact conditions under which the unmixing accuracy achieves $R^2 \geq 0.7$, this pipeline must be stress tested under different parameter variables. The full pipeline developed for this purpose is shown in Figure 2. Pure spectra acquired in a laboratory setting and labeled with their corresponding crop residue, green vegetation and soil abundances (Dennison et al., 2019) were forward propagated using atmospheric simulation software under different assumed atmospheric and imaging conditions to acquire an atmospherically propagated dataset. The spectra were then atmospherically corrected and unmixed.

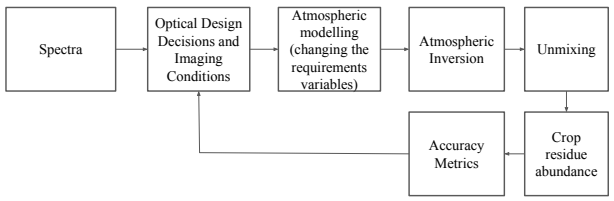

Figure 2. Pipeline to set design and mission operation requirements for FINCH

4. Data

4.1 Lab Dataset

The first dataset considered is lab-procured. It consists of 1723 spectra from 5 studies, labeled with their corresponding non-photosynthetic vegetation (NPV), green vegetation (GV), and soil abundances (Dennison et al., 2019).

The reflectances for a particular wavelength were found to be roughly normally distributed and centered around 0.25, as shown in the supplementary materials available in the Appendix. For all data points, the sum of the abundances is 1, which is consistent with the definition of fractional abundances and our model assumptions. However, in most data points just two endmember abundances add up to more than 0.90.

Figure 3 shows the abundance combinations represented in the dataset (colored according to their source). Unfortunately, large regions of the simplex remain unrepresented or severely underrepresented.

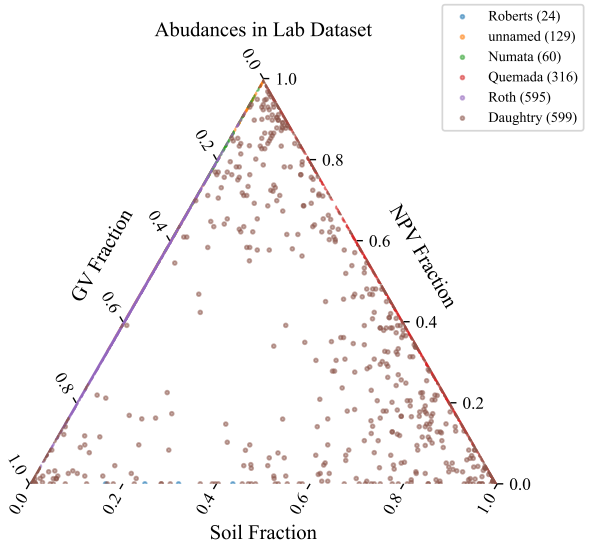

Figure 3. Distribution of abundances in the Lab Dataset, with points colored according to their source study

Finally, some reflectances were strongly correlated with each other, as seen in the white regions of Figure 4.

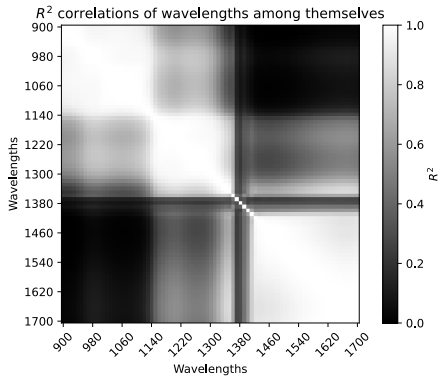

Figure 4. Correlations of Wavelengths among themselves in the lab dataset

4.2 Atmospherically Propagated Dataset

To create the atmospherically propagated dataset, NPV and soil endmembers were taken from the lab dataset. These endmembers were then mixed in different proportions. Next, the

mixed spectra were forward propagated through an atmospheric transmittance model while changing the viewing zenith angle (VZA). The model was developped in-house and is presented in a different publication from the same group. The signal-to-noise ratio was set to 200. The model was then employed to atmospherically correct the data, thereby retrieving bottom-of-atmosphere data suitable for spectral unmixing. In the end, 144 samples were created with reflectances between $900\,\mathrm{nm}$ and $1690\,\mathrm{nm}$ and the GV abundance set to 0, since we expect the imaged fields to have almost no GV.

This dataset presented two significant challenges. Firstly, the number of samples (144) is small compared to the number of wavelengths (80), making our models prone to overfitting. Additionally, wavelengths were highly correlated with each other, as seen in Figure 5.

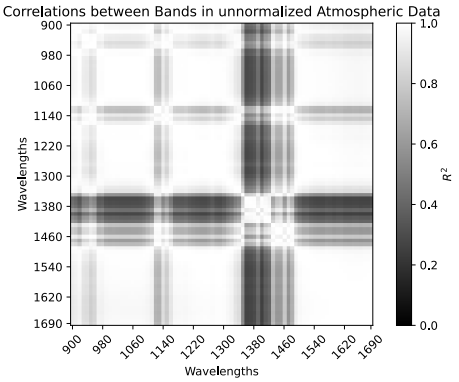

Figure 5. Correlations of wavelengths among themselves in the atmospherically propagated dataset

5. Data Analysis Methods

5.1 Normalization

Two different normalization methods were used. The first is L_1 sample normalization. This normalizes per sample across wavelengths and can be interpreted as normalizing for the amount of light reaching the sensor, which corrects for different lighting conditions. This normalization is motivated by the fact that the K-Nearest Neighbors model performed best when using the cosine definition of distance on unnormalized data.

The second is traditional z-score normalization of each wavelength across samples, motivated by the symmetric distribution of reflectances within single wavelengths in the dataset. Unless otherwise stated, this was the normalization method used. The mean and standard deviation were acquired from the training data only to prevent data leakage, but applied to both the training and testing datasets.

5.2 Unmixing Methods

Unless otherwise noted, an 80% to 20% training/testing split was used. For classical methods, Scikit-learn’s implementation (Pedregosa et al., 2011) was used.

5.2.1 Linear Regression: The first model considered was a multivariate linear regression with the sum-to-one constraint enforced. Symbolically:

$$\hat{\boldsymbol{\lambda}}_i = \mathbf{x_i}^\top W + \mathbf{b} \tag{3}$$

$$\hat{\boldsymbol{a}}_{ik} = \frac{\hat{\boldsymbol{\lambda}}_{ik}}{\sum_{k=1}^{3} \hat{\boldsymbol{\lambda}}_{ik}} \tag{4}$$

where $\hat{\boldsymbol{a}}_{ik} \in \mathbb{R}^3$: the predicted abundance of the kth endmember for sample i
$\hat{\boldsymbol{\lambda}}_i \in \mathbb{R}^3$: the initial linear prediction for the ith sample
$W \in \mathbb{R}^{80\times 3}$: is the weight matrix of best fit
$\mathbf{x}_i^\top \in \mathbb{R}^{80}$: is the input spectrum for sample i
$\mathbf{b} \in \mathbb{R}^3$: is the bias vector of best fit

5.2.2 Bayesian Linear Dirichlet Model: A Bayesian model was chosen in order to acquire credible intervals. A Dirichlet-based model was used to naturally integrate the non-negativity and sum-to-one constraints of the abundances. All training abundances equal to zero were replaced with $\epsilon = 10^{-6}$ and those equal to 1, with $1 - \epsilon$. The model was defined as follows:

$$\widehat{\mathbf{a}}_i \sim Dir(\boldsymbol{\alpha}_i) \tag{5}$$

$$\hat{\boldsymbol{\alpha}}_i = softplus(\mathbf{x}_i^\top W + b) \tag{6}$$

$$W_{lk} \sim \mathcal{N}(0, 2), \quad l = 1, \ldots, 80, \quad k = 1, 2, 3 \tag{7}$$

$$b_k \sim \mathcal{N}(0, 2), \quad k = 1, 2, 3 \tag{8}$$

where $\widehat{\mathbf{a}}_i \in \mathbb{R}^3$: the predicted abundance vector for sample i
$\hat{\boldsymbol{\alpha}}_i \in \mathbb{R}^3$: the transformed linear predictor for sample i
$\mathbf{x}_i \in \mathbb{R}^{80}$: the input spectrum for sample i
W_{lk} : is the weight connecting input feature l to output component k
b_k : the bias for output component k

The priors were chosen to be regularizing to prevent overfitting by shrinking weights towards smaller absolute values. We also expected each individual wavelength to have a small effect that could be either positive or negative.

Sampling was done using PyMC’s (Abril-Pla et al., 2023) implementation of NUTS (Hoffman and Gelman, 2011) with 4,000 samples total, as can be seen in the supplementary code.

Predictions were made twice: once by propagating aleatoric uncertainty and once by collapsing it to its mean, and hence propagating epistemic uncertainty only.

The linear component was acquired as follows:

$$\hat{\boldsymbol{\lambda}}_n^{(s)} = \mathbf{x}_n^\top W^{(s)} + \mathbf{b}^{(s)}, \quad \hat{\boldsymbol{\lambda}}_i^{(s)} \in R^3 \tag{9}$$

where $\hat{\boldsymbol{\lambda}}_i^{(s)}$: the linear predictor for the ith testing data point
$W^{(s)} \in \mathbb{R}^{80\times 3}$: the sth posterior draw for the weight matrix
$\mathbf{b}^{(s)} \in \mathbb{R}^3$: the sth posterior draw for the bias vector
$\mathbf{x}_i \in \mathbb{R}^{80}$: the ith testing data point

Then, the Dirichlet components were calculated as follows:

$$\hat{\boldsymbol{\alpha}}_i^{(s)} = \log(1 + e^{\hat{\boldsymbol{\lambda}}_i^{(s)}}) + \varepsilon, \quad \varepsilon = 10^{-6} \tag{10}$$

where $\hat{\boldsymbol{\alpha}}_i^{(s)}$: the predicted Dirichlet component for the ith data point and sth sample draw

The ε was used to ensure that the Dirichlet components were positive as required for Dirichlet parametrization.

Finally, the predictions for each testing data point per draw were acquired as follows when collapsing aleatoric uncertainty:

$$\widehat{\mathbf{a}}_i = \frac{\hat{\boldsymbol{\alpha}}_i^{(s)}}{\sum_{k=1}^{3} \hat{\alpha}_{ik}^{(s)}} \tag{11}$$

where $\hat{\alpha}_{ik}^{(s)}$: the predicted Dirichlet component for the ith data point and kth endmember

This takes the expected value for the abundance of each endmember, since the expected value of the kth Dirichlet component's abundance is the quotient of $\hat{\boldsymbol{\alpha}}_i$, and the sum of the three Dirichlet components for that point.

As for predictions that propagate aleatoric uncertainty as well, only the last step changes. Equation 11 is replaced by the following:

$$\widehat{\mathbf{a}}_i \sim Dir(\hat{\boldsymbol{\alpha}}_i^{(s)}) \tag{12}$$

5.2.3 K-Nearest Neighbors with Cosine Distance: K-Nearest Neighbor was chosen as it does not assume linearity, but is simpler and thus requires less data than deep learning models.

The algorithm uses the average of its k nearest points as its regression prediction. Different k values and definitions of distance were tested. The k between 3 and 7 produced best results, while the cosine definition of distance performed significantly better than euclidean.

5.2.4 Random Forest: Random Forest is a non-parametric, decision-tree based ensemble method. (Breiman, 2001) It was chosen because, like K-Nearest Neighbors, Random Forest does not make assumptions on the distribution of the data or relationship between predictor and target variables.

5.2.5 Deep Learning Models: Given the recent emergence of Deep Learning-based models as a powerful tool to analyze remote sensing images (Yang et al., 2025), Multi-Layered Perceptrons (MLP) and Convolutional Neural Networks (CNN) were a natural baseline choice to investigate the effectiveness of deep learning models. Neural Operators are capable of efficiently learning operator mappings on complicated problems such as fluid mechanics, plasma physics, and climate modeling. Given the significant use of Fourier Neural Operators (FNO) (Li et al., 2020) on surrogate modeling problems of various difficulties, they were chosen as another model for unmixing. Additionally, their capability in resolving global interactions throughout, compared to MLPs and CNNs, supported them as a worthwhile candidate.

MLPs were defined using leaky rectilinear unit (LeakyReLU) activations with 0.2 slope for the leak, applied after batch normalization which followed linear layers; dropout of 0.05 probability was applied after activation for all the models.

CNNs employed an hourglass shape, which resulted in the expansion of channels, and reduction of sequence lengths, which was reversed later in the model, to result in sequence length expansion and channel reduction. All the convolutional layers in the first half of the layers were followed by batch normalization, LeakyReLU of 0.2 leak slope, dropout of 0.05 probability, and max pool of kernel size 2 and stride 2. On the latter half of the CNNs, a convolutional transposition layer was applied to increase the sequence length and reduce the channel size, all of which were followed by batch normalization, LeakyReLU of 0.2 leak slope, channel and sequence preserving convolution, batch normalization, and another LeakyReLU of 0.2 leak slope.

The FNO models employed orthogonal normalization for the spectral convolutions, and geometric normalizations for all other layers. Additionally, all the activation functions were chosen as Gaussian Linear Error Unit (GELU), and a dropout of 0.05 probability was applied throughout all the layer parameters. On the output of the Q network, adaptive average pooling, linear layer, GELU, and a final linear layer was applied consecutively. Both MLP and CNN models employed ReLU followed by normalization for their last activations, and FNOs employed vertically shifted hyperbolic tangent activations followed by normalization.

All models were trained for 80 epochs using the lab dataset initially, with optimized learning rates employing cosine annealing learning rate updates after each training step. FNO models' learning rates were warmed up using linear updates for the first 10 epochs. All three models were optimized with the AdamW algorithm with 0.05 weight decay, where each model iterates training steps on the 1500 samples of the dataset each epoch. The validation dataset of 100 was kept constant throughout the dataset. The models' R^2 values were obtained from a testing dataset of 123 samples.

With models of varying parameter sizes trained on the lab dataset, models from each architecture were chosen as the best representative of the architecture class in terms of learning capacity and learned internal latent space mapping. The initial layers of the chosen models were then frozen, so that their final layers were fine-tuned on the atmospherically propagated dataset. The models were assumed to have sufficiently learned optimal feature extractions in their initial layers, providing a workable baseline for the fine-tuning of the latter layers. They were then fine-tuned over 80 epochs, with a train/validation/test split of 100, 24 and 20 samples respectively, determined by a randomly chosen seed (where the seed for seed generation was chosen using the ”secrets” module in Python), with 50 random splits in total.

A histogram of the testing accuracies of each model was plotted after fine-tuning on random data splits. This was critical because of the relatively low quantity of spectral data available in the atmospherically propagated dataset. The means of the determined distributions are reported in Table 1 for unmixing accuracies when fine-tuned on the atmospherically propagated dataset. All data in both pre-training and fine-tuning scenarios were normalized with the Spectral Normal Variate method.

6. Results

All results shown in Table 1 were acquired by splitting using random state 42.

6.1 Results on Lab Dataset

6.1.1 Linear Regression: The linear regression results can be seen in Table 1, where R^2 for the true vs. predicted abundances is equal to 0.8612, with an RMSE of 0.1153, and for NPV, R^2 is 0.8141. The coefficient weights can be seen in Figure 6.

Dataset	Method	RMSE	R^2	R^2 (NPV)
Lab Dataset	Linear Regression	0.1153	0.8612	0.8141
	Dirichlet Model	0.1248	0.8312	0.8024
	Dirichlet Model - aleatoric collapsed only	0.1248	0.8311	0.8023
	K-Nearest Neighbors	0.1355	0.8055	0.7432
	Random Forest	0.1252	0.8357	0.7868
	MLP_128_4L	0.0909	0.8935	0.8436
	CNN_32	0.1009	0.8683	0.8020
	FNO_32	0.1082	0.8513	0.7671
Atmospherically Propagated Dataset	Linear Regression	0.3553	-0.0730	-0.0730
	Dirichlet Model	0.2251	0.2432	0.3514
	K-Nearest Neighbors	0.2243	-0.0802	-0.0802
	Random Forest	0.2891	0.2899	-0.0652
	MLP_128_4L	0.2559	-6.2515	-8.4521
	CNN_32	0.3579	-3.7175	-2.7548
	FNO_32	0.2786	-4.4988	-6.3129

Table 1. Fit metrics for all models in both datasets (test size 0.2). (NPV) refers to the metric being measured based on the predicted abundance of non-photosynthetic vegetation only, instead of all three endmembers.

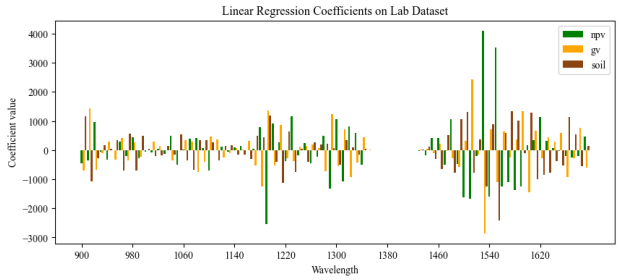

Figure 6. Coefficients weights for the linear regression model.

Such large weights could make the model prone to overfitting and more sensitive to noise.

Hence, numerous shrinkage and feature selection methods were implemented, although only the Dirichlet model is reported. The other techniques tried include Ridge regression and dropping columns that were highly correlated with each other. They all yielded similar results. Additionally, dimensionality reduction using Principal Component Analysis was applied. However, this yielded bad prediction accuracy, presumably because the linear relationship between the target and predictor variables was destroyed.

As can be seen in Figure 7, there was variability in terms of the R^2 value produced for different training/testing splits of the same size.

6.1.2 Dirichlet Model: The Dirichlet Model results are presented in Table 1, where R^2 for the true vs. predicted abundances is equal to 0.8312, with an RMSE of 0.1248, and for NPV, R^2 is 0.8024. The model with collapsed aleatoric uncertainty showed almost identical results, since the R^2 for the true vs. predicted abundances is equal to 0.8311, with an RMSE of 0.1248, and for NPV, R^2 is 0.8023. The median width of the credible interval was 0.3948 when considering both sources of uncertainty, and 0.0426 when collapsing aleatoric uncertainty for predictions. Although aleatoric uncertainty was extremely large, it did not seem to hurt prediction accuracy since the R^2 and RMSE were almost identical.

The true vs. predicted plot with collapsed aleatoric uncertainty can be seen in Figure 8, and for both sources of uncertainty

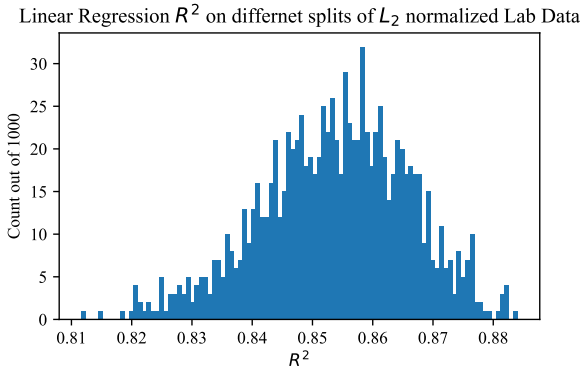

Figure 7. Frequency of R^2 values for different random states.

in the supplementary code. As can be seen from Figure 8, the degree of uncertainty still varied greatly, with some badly predicted plots having very small uncertainty when the aleatoric uncertainty was collapsed.

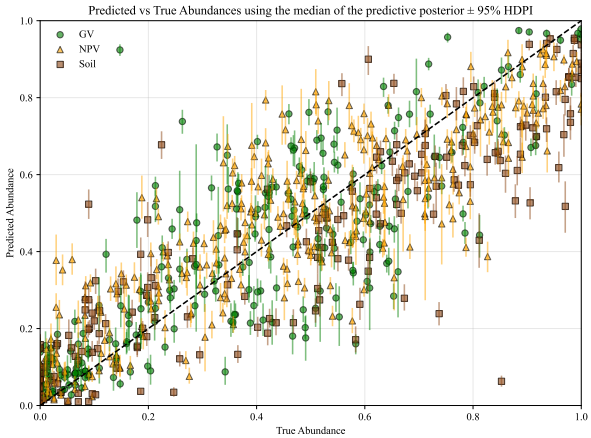

Figure 8. True vs. Predicted plot for the Dirichlet model on lab dataset, with collapsed aleatoric uncertainty, with 95% credible intervals in the error bars. The points represent the median of the prediction.

The coefficients were much smaller when compared to the fre-

quentist linear regression. The mean of the absolute values of the weights of the individual wavelengths never exceeded 6.

Although fewer random states could be trained due to the computational cost of training each one, there appeared to be lower variability in R^2 when the random state was changed and compared with the frequentist linear regression model.

6.1.3 Random Forest and K-Nearest Neighbors: The Random Forest Model results can be seen in Table 1, where R^2 for the true vs. predicted abundances is equal to 0.8357, with an RMSE of 0.1252, and for NPV, R^2 is 0.7868. Random Forest produced worse predictions than the linear regression model. This suggests that the linear model is appropriate, and that deviations from this relationship are subtle and complex.

Moreover, it has been shown that the best approach to determining the size of the training sample in the lab dataset, allowing to obtain the maximum value of R^2 while using Random Forest regression in the crop-residue analysis, is the k-fold division. Eventually, this method allowed us to achieve the R^2 value of 0.8953 with a k-fold size of 16. For the methodology, the k-fold dataset was formed in such a way that the best-performing splitting of size k was chosen as a representative. However, this method is incomplete for the current state of the pipeline and it would only work with more data to test on, since it would require having sepparate validation and testing datasets to mantain good statistical practice.

The points for which Random Forest produced poor predictions are the same points the Dirichlet model and standard linear model could not accurately predict. Moreover, the predictions for these points were fairly similar. This provides evidence that suggests some points do not follow the general trends and are thus harder to predict. However, these differences could not be traced back to a particular source dataset.

The prediction accuracy was extremely similar for K-Nearest Neighbors. Its low value compared to linear regression could have been caused by the under-representation of some abundance combinations in the dataset, as can be seen in Figure 3. It performed best when using the cosine definition of distance.

6.1.4 Multi-Layered Perceptron: The Multi-Layered Perceptrons (MLPs) have been observed to obey bias-variance tradeoff when considering the NPV R^2 accuracy during testing, with respect to the number of parameters as presented on Figure 9. The model with the highest testing accuracy has been identified as the MLP model with 4 hidden layers, each having 128 neurons, named as ”MLP_128_4L”, with ∼61k parameters. The flagship model, MLP_128_4L, exhibited an NPV R^2 accuracy of 0.8436, total R^2 accuracy of 0.8935, and an RMSE of 0.0909 during testing; this result positions MLP_128_4L as the highest performing model presented in this study for the lab dataset. Hence, MLP_128_4L was chosen to be fine tuned on the atmospherically simulated dataset. All MLP models reached convergence after 70 epochs of training.

6.1.5 Convolutional Neural Network: The results for the Convolution Neural Network (CNN) can be seen in Table 1, where R^2 for the true vs. predicted abundances is equal to 0.8683, with an RMSE of 0.1009, and for NPV, an R^2 of 0.8020. The Pareto front of the CNN model was systematically lower than the one defined by the MLP models. The highest performer of this architecture was employing 8-16-32-16-8 channels for the convolutional hidden layers, with ∼16k

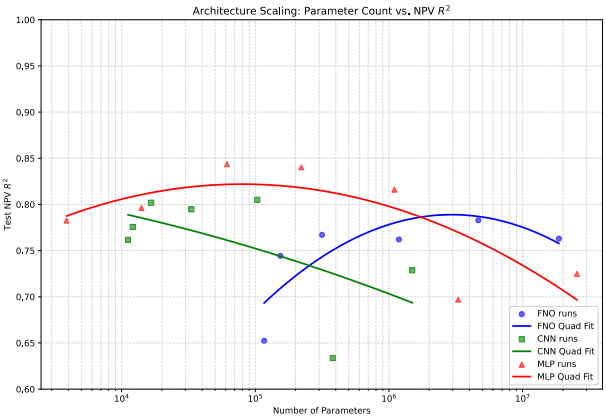

Figure 9. Pareto Fronts of Varying Architectures, for their NPV R^2 Testing Accuracy Given Parameter Numbers

parameters, henceforth referred to as ”CNN_32”. Hence, it was chosen to be used for fine-tuning on the atmospherically simulated dataset. An outlier model with ∼379k parameters was identified due to its unusually low NPV R^2 value during testing of 0.6338. All other CNN models obeyed the bias-variance trade-off, similar to the MLP models, as shown in Figure 9. All CNN models reached convergence after 65 epochs.

6.1.6 Fourier Neural Operator: The Fourier Neural Operators (FNO) results can be seen in Table 1, where R^2 for the true vs. predicted abundances plot is equal to 0.8513, with an RMSE of 0.1082, and for NPV, R^2 is 0.7671. The FNOs obeyed the bias-variance trade-off while consistently under-performing compared to the MLP models as presented in Figure 9. The highest performing FNO model, named ”FNO_32”, had ∼315k parameters. It was chosen for fine-tuning with atmospherically propagated data. Paradoxically, FNO_32 underperformed compared to CNN_32, despite having similar RMSE values during testing. All of the FNO models reached convergence after ∼75 epochs.

6.2 Results on Atmospherically Propagated Dataset

6.2.1 Linear Regression: The linear regression results can be seen in Table 1 where R^2 for the true vs. predicted abundances and NPV is equal to -0.0730, with an RMSE of 0.3553. Note that a negative R^2 implies the model’s predictions are worse than always predicting the mean.

Although training R^2 was 0.6, it had no predictive power on the testing dataset. Linear regression produced weights in the order of ± 400. Presumably the 80 predictor variables and only 115 training samples permit for severe overfitting. Textbook remedies like Ridge were unable to generate predictive power for the testing dataset. Ultimately, no linear regression based model produced meaningful results.

Normalization and dimensional reduction were also attempted (in the same manner as described for the Lab Dataset), but did not improve linear regression models in any way.

For comparison, unmixing was performed on a comparable set of 144 samples from the Lab Dataset. A testing R^2 of 0.4862 was obtained. Hence, lack of data was a serious limitation but not the only cause of poor results on atmospherically propagated data.

6.2.2 Dirichlet Model: The model was trained on both the lab dataset and 60% of the atmospherically propagated dataset. The rest of the 40% of the atmospherically propagated dataset was used for testing. Lab data was included in the training data because this is theoretically equivalent to training the model on this data and then using the results as the prior for training on the atmospherically propagated data.

The R^2 for is equal to 0.2432, with an RMSE of 0.2251, and for NPV, an R^2 of 0.3514. This shows the model was unable to make meaningful predictions. Similar results were obtained by training on the atmospherically propagated data only.

6.2.3 Random Forest: The results for the Random Forest model can be seen in Table 1, where R^2 is equal to 0.2899, with an RMSE of 0.2891, and for NPV, R^2 is -0.0652.

Even worse results were obtained for K-Nearest Neighbors, presumably because there were not enough data points to appropriately cover all areas of the simplex.

6.2.4 Multi-Layered Perceptron: Although MLP_128_4L was the highest performing unmixing architecture for the lab dataset, it failed to predict abundances when using the atmospherically simulated dataset. The model exhibited NPV R^2 value of -8.4521, total R^2 value of -6.2515, and an RMSE of 0.2559 during testing.

6.2.5 Convolutional Neural Network: Similar to MLP_128_4L, CNN_32 also failed to predict abundance using the atmospherically simulated dataset. The model presented an even higher RMSE of 0.3579 during testing, compared to MLP_128_4L’s 0.2559. The model exhibited NPV R^2 value of -2.7548, and a total R^2 value of -3.7175.

6.2.6 Fourier Neural Operator: FNO_32 does not accurately predict abundances when using the atmospherically simulated dateset. NPV R^2 is -6.3129, total R^2 as -4.4988, and RMSE is 0.2786.

7. Discussion

Excellent unmixing results were obtained for the lab dataset. Given the high R^2 values obtained by linear regression and the physical context presented in Equation 1, a linear model is appropriate for predicting abundances from spectral images. However, as can be seen by the even higher R^2 values obtained by deep learning models, there are some subtle deviations from this model. One possible explanation is that this is due to endmember variability, and that deep learning models were able to learn this endmember diversity.

Hence, more complex deep learning models could be built. Since MLP outperformed FNO and CNN, tabular and long-distance interactions seem to be better predictors than local or global interactions only. Thus, transformers and conformers should be considered in order to model long range and short range interactions.

Additional future directions for the Lab Dataset include trying different priors for the Dirichlet model and acquiring more data so that all areas of the abundance simplex are well-represented. The existing deep learning models could also have their hyperparameters tuned.

The Atmospherically Propagated Dataset, on the other hand, could not be unmixed. Various pieces of evidence indicate that the Lab and Atmospherically propagated dataset are extremely different. The first is the stark differences in the correlations between wavelengths, as can be seen by comparing Figure 4 and Figure 5. Moreover, transfer learning for deep learning models and training on the Lab Dataset as a prior did not yield good results. Finally, the fact that linear models performed so badly provide some evidence that the linear relationship between the reflectances and abundances was destroyed.

If the relationship is indeed not linear, then the statistical models must be completely rewritten. Unless the shape of this relationship can be identified, models that do not assume a family of functions for the distribution must be used.

Transfer learning did not work with this amount of data. A solution would be to acquire more data, but then this turns transfer learning obsolete. Hence, transfer learning is not an appropriate tool for the problem at hand.

Since linear models and transfer learning are discarded, the remaining options include either finding an algorithm that requires little data but does not make assumptions about the relationship between predictor and target variables or acquiring more atmospheric data. The former is extremely hard, as shown by the failure of K-Nearest Neighbors and Random Forest (R^2 of -0.0802 and 0.2899, respectively). This model would most likely have to be low capacity, which would hurt prediction accuracy given the complex relationships that this dataset seems to have. While the latter is hard to achieve due to limitations in the computing power, it would be extremely beneficial, especially if the abundance distribution covers all major regions of the simplex to avoid out-of-sample testing, as opposed to the lab dataset as seen in Figure 3.

8. Conclusion

A pipeline that allows for the determination of design and imaging requirements was developed, and multiple unmixing methods were tested. Excellent unmixing results were achieved using the lab dataset. The highest performing model was the Multi-Layered Perceptron (MLP), with an NPV R^2 accuracy of 0.8436, total R^2 accuracy of 0.8935, and an RMSE of 0.090. However, unmixing performed poorly on the atmospherically propagated dataset. Atmospherically propagated data seems to present stark differences to lab data, including non-linearity.

A possible explanation of deep learning’s performance is its ability to learn endmember variability, which classical methods were presumably unable to pick up. The atmospherically propagated data is non-linear. Hence, accurate prediction with this dataset is unlikely without more atmospheric data, because simple low-data models fail and low-capacity models can’t capture the complex relationships.

Further testing and validation are required to establish confidence in predictions from FINCH’s images. We hope to develop unmixing methods for infrared hyperspectral imagery, and show the satellite as a viable proof of concept for farm-level crop residue estimation.

Acknowledgments

We would like thank the University of Toronto Aerospace Team’s (UTAT) Space Systems Division for providing a collaborative research environment, and key operational support.

Furthermore, we would like to express our gratitude to our colleagues and teammates for their valuable proofreading.

References

Abril-Pla, O., Andreani, V., Carroll, C., Dong, L., Fonnesbeck, C. J., Kochurov, M., Kumar, R., Lao, J., Luhmann, C. C., Martin, O. A., Osthege, M., Vieira, R., Wiecki, T., Zinkov, R., 2023. PyMC: A Modern and Comprehensive Probabilistic Programming Framework in Python. *PeerJ Computer Science*, 9(e1516). https://doi.org/10.7717/peerj-cs.1516.

Breiman, L., 2001. Random Forests. *Machine Learning*, 45(1), 5-32. https://doi.org/10.1023/A:1010933404324.

Daughtry, C. S., 2001. Discriminating Crop Residues from Soil by Shortwave Infrared Reflectance. *Agronomy Journal*, 93(1), 125-131. https://doi.org/10.2134/agronj2001.931125x.

Dennison, P. E., Lamb, B. T., Campbell, M. J., Kokaly, R. F., Hively, W. D., Vermote, E., Dabney, P., Serbin, G., Quemada, M., Daughtry, C. S., Masek, J., Wu, Z., 2023. Modeling global indices for estimating non-photosynthetic vegetation cover. *Remote Sensing of Environment*, 295, 113715. https://doi.org/10.1016/j.rse.2023.113715.

Dennison, P. E., Qi, Y., Meerdink, S. K., Kokaly, R. F., Thompson, D. R., Daughtry, C. S. T., Quemada, M., Roberts, D. A., Gader, P. D., Wetherley, E. B., Numata, I., Roth, K. L., 2019. Comparison of Methods for Modeling Fractional Cover Using Simulated Satellite Hyperspectral Imager Spectra. *Remote Sensing*, 11(18). https://doi.org/10.3390/rs11182072.

Hoffman, M. D., Gelman, A., 2011. The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *arXiv preprint*. https://doi.org/10.48550/arXiv.1111.4246.

Keshava, N., Mustard, J., 2002. Spectral Unmixing. *Signal Processing Magazine, IEEE*, 19, 44 - 57. https://doi.org/10.1109/79.974727.

Li, Z., Kovachki, N. B., Azizzadenesheli, K., Liu, B., Bhattacharya, K., Stuart, A. M., Anandkumar, A., 2020. Fourier Neural Operator for Parametric Partial Differential Equations. *arXiv preprint*. https://doi.org/10.48550/arXiv.2010.08895.

Pacheco, A., McNairn, H., 2010. Evaluating multispectral remote sensing and spectral unmixing analysis for crop residue mapping. *Remote Sensing of Environment*, 114(10), 2219-2228. https://doi.org/10.1016/j.rse.2010.04.024.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830. https://doi.org/10.5555/1953048.2078195.

Yang, L., Lu, B., Schmidt, M., Natesan, S., McCaffrey, D., 2025. Applications of remote sensing for crop residue cover mapping. *Smart Agricultural Technology*, 11, 100880. https://doi.org/10.1016/j.atech.2025.100880.

Appendix

The data, code, configuration files, and images used in this study can be found in the following GitHub Repository Release: UTAT FINCH Mission Hyperspectral Unmixing Models for Lab and Atmospherically Simulated Data (v1.0.0). The frozen code used for this manuscript is persistently accessible via Zenodo in: `https://doi.org/10.5281/zenodo.19443830`. For the actively maintained codebase and future updates, please refer to the repository: `https://github.com/utat-ss/FINCH-Science_Unmixing`.