

Article

Probabilistic Mixture Model-Based Spectral Unmixing

Oliver Hoidn, Aashwin Ananda Mishra * and Apurva Mehta 

SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA; ohoidn@slac.stanford.edu (O.H.); mehta@slac.stanford.edu (A.M.)

* Correspondence: aashwin@slac.stanford.edu

Abstract: Spectral unmixing attempts to decompose a spectral ensemble into the constituent pure spectral signatures (called endmembers) along with the proportion of each endmember. This is essential for techniques like hyperspectral imaging (HSI) used in environment monitoring, geological exploration, etc. Several spectral unmixing approaches have been proposed, many of which are connected to hyperspectral imaging. However, most extant approaches assume highly diverse collections of mixtures and extremely low-loss spectroscopic measurements. Additionally, current non-Bayesian frameworks do not incorporate the uncertainty inherent in unmixing. We propose a probabilistic inference algorithm that explicitly incorporates noise and uncertainty, enabling us to unmix endmembers in collections of mixtures with limited diversity. We use a Bayesian mixture model to jointly extract endmember spectra and mixing parameters while explicitly modeling observation noise and the resulting inference uncertainties. We obtain approximate distributions over endmember coordinates for each set of observed spectra while remaining robust to inference biases from the lack of pure observations and the presence of non-isotropic Gaussian noise. As a direct impact of our methodology, access to reliable uncertainties on the unmixing solutions would enable robust solutions to noise, as well as informed decision-making for HSI applications and other unmixing problems.

Keywords: hyperspectral unmixing; probabilistic modeling; data-driven algorithms



Citation: Hoidn, O.; Mishra, A.A.; Mehta, A. Probabilistic Mixture Model-Based Spectral Unmixing. *Appl. Sci.* **2024**, *14*, 4836. <https://doi.org/10.3390/app14114836>

Academic Editor: Vittoria Guglielmi

Received: 6 March 2024

Revised: 13 May 2024

Accepted: 27 May 2024

Published: 3 June 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Mixing is ubiquitous in nature, ranging from the slow mixing in the Earth's mantle to manufacturing processes involving the blending of viscous polymers or even the dispersion of solids into rubber [1]. The process and mathematics of mixing are of great interest in physics and engineering applications, from mixing in process engineering to the chaotic mixing of horseshoe maps. The reverse of this process, referred to as unmixing, segregation, or separation, is also of great interest but challenging. It involves recovering information lost through the changes in entropy due to mixing. The process is possible computationally if the mixing is incomplete (i.e., there is variation in the mixing fractions at different sites in the material), and sufficient spectral signal is discernible from the individual mixture components. Unmixing is commonly attempted for cases where the equations governing mixing are linear [2], although non-linear unmixing methods, such as kernel-based approaches and neural networks, have also been developed. Computational unmixing is critical for understanding and controlling processes like electrolysis refining [3], petroleum refining [4], and chemical analysis via chromatography [5], besides others. Spectral unmixing, a specific application of computational unmixing, aims to identify and quantify material components in mixed signals by comparing sets of spectral measurements to a library of known signatures. Each observation in such a set of measurements contains a spectral footprint corresponding to the underlying material mixture. The subsequent task is inverting these data to identify each endmember's unique spectral signature and quantify the proportion of each endmember in each mixture. This process finds applications in diverse fields, from geology and agriculture to astronomy and material sciences. However,

unmixing represents an ill-posed problem, and the solutions may not be unique. In this light, it is essential to represent the uncertainties associated with the unmixing process.

To illustrate the property of mixture diversity, consider a color photograph represented as a set of multi-dimensional arrays of weights, with endmembers identifiable as distinct colors: red, blue, and green. A color photograph is an example of a heterogeneously mixed set of observations that can be mapped to points on a 3-simplex, and the maximally diverse color photograph is that which contains at least one each of a pure (or completely) red pixel, a pure blue pixel, and a pure green pixel (i.e., the vertices of the simplex).

The concept of an endmember extends beyond discrete colors in many scientific applications. The array of weights can be multi-dimensional, including non-spatial components like time, scattering vectors, chemical compositions, red-shifts, and more. Rather than viewing hyperspectral imaging as the sole technique, we adopt a broader perspective where the observations need not be a specific type of spatial map. Instead, they can be treated as an unordered set of spectra in which other descriptors (e.g., spatial parameters) are not necessarily needed for the endmember extraction, and thus, the problem we are addressing is the more general task of spectral unmixing.

The success of spectral unmixing relies on the selection of a spectral tool to analyze the data, the spectral distinctiveness of the endmembers, and the diversity of available mixtures. Given a particular choice of spectral measurement tool, the most salient considerations are noise and mixture diversity. Additional challenges arise from factors such as non-linear mixing and data distortions. Noise, or, more accurately, noise relative to the spectral separation between endmembers, determines the ease of distinguishing hypothetical pure spectral signatures. Thus, noisy observations can complicate extraction as can homogeneous mixtures, in which endmember extraction becomes nearly impossible even in the absence of noise. Conversely, if endmembers are immiscible—meaning the mixing fraction of an endmember in a given observation is either 0 or 1—extraction is trivial as long as the noise level is small compared to the spectral separation between endmembers. Here, we focus primarily on the challenging, but tractable, intermediate case of heterogeneously mixed linear observations.

2. Related Work

In the broad landscape of spectral unmixing, several techniques have emerged, including geometrical methods (e.g., N-FINDR), statistical approaches, non-negative matrix factorization (NMF), and sparse regression-based methods [6,7]. These methods can be broadly categorized into deterministic and probabilistic approaches. Deterministic methods, which are overwhelmingly more common among the state of the art, include approaches such as sparse regression and simplex-based techniques. These deterministic methods provide point estimates of endmembers and abundances. Some deterministic methods, like sparse unmixing with L1 regularization, can provide a limited form of uncertainty quantification through the selection of regularization parameters.

Geometric simplex methods, such as N-FINDR [6], rely on the assumption that the spectral signatures of the endmembers are distinct and that the observed mixed pixels lie within a simplex formed by the endmembers. This assumption is related to the underlying principle of manifold learning techniques, which assume that high-dimensional data lie on a lower-dimensional manifold. N-FINDR includes a dimensionality reduction step using orthogonal subspace projection (OSP) techniques such as SVD, PCA, or MNF before searching for the endmembers. However, while N-FINDR uses linear dimensionality reduction and aims to find the endmembers in the reduced space, manifold learning methods are designed to handle non-linear manifolds and aim to uncover the intrinsic low-dimensional structure of the data.

Non-probabilistic machine learning approaches have gained popularity in spectral unmixing due to their ability to learn complex representations and relationships from data. NMF, a widely used technique, decomposes the spectral data matrix into non-negative endmember and abundance matrices [8]. NMF-based methods have been extended to incorporate various constraints and regularizations to improve their performance and to handle different types of spectral variability [9]. Sparse unmixing methods, on the other hand, exploit the fact that each

pixel in hyperspectral images is often a mixture of only a few endmembers, incorporating sparsity constraints to estimate the abundances [10,11]. Manifold learning techniques [12], such as locally linear embedding (LLE) [13] and isometric feature mapping (ISOMAP) [14], have been applied to spectral unmixing by assuming that the high-dimensional spectral data lie on a low-dimensional manifold. These methods aim to uncover the intrinsic structure of the data and can handle non-linear spectral variability to some extent. Dictionary learning approaches, which learn a dictionary of endmember spectra directly from the hyperspectral data, have also shown potential in adapting to the specific characteristics of the input data.

More recently, deep learning models, such as convolutional neural networks (CNNs) and autoencoders, have been applied to spectral unmixing [15–17]. For example, Palsson et al. [15] propose an autoencoder-based method for blind hyperspectral unmixing. Their method uses a deep autoencoder to extract endmembers and estimate abundances in an unsupervised manner. The architecture is designed to enforce the linear mixing model, with the decoder weights representing the endmembers. While their approach is unsupervised, scalable, and compares favorably to benchmark methods in reconstruction accuracy, it does not model endmember uncertainties. Generally, while non-probabilistic machine learning approaches have shown promise in addressing specific challenges in spectral unmixing, they typically generate point estimates of the endmember spectral signatures and abundances.

Thus, this wide array of non-probabilistic approaches offers advantages in interpretability and robustness (with respect to spectral variability) over conventional methods but falls short in handling both epistemic uncertainties in the endmember spectra and fractional abundances, as well as the aleatoric uncertainties arising from noise in the measurements.

Our approach is probabilistic and demonstrates the ability of a hierarchical Bayesian mixture model to account for these uncertainties intrinsic to the unmixing problem fully. We retain the conceptual simplicity of a simplex method, but unlike N-FINDER, which relies on the assumption of fixed endmembers, our Bayesian mixture model explicitly accounts for the uncertainty in both individual observations and underlying endmember signatures, leading to more realistic and informative unmixing results.

The intent is to highlight the advantage of probabilistic modeling over point-estimate approaches, as the Bayesian approach explicitly acknowledges and incorporates the epistemic uncertainties inherent to the unmixing process. Additionally, Bayesian approaches can incorporate aleatoric uncertainties in the problem arising from (for example) noisy measurements. The final solutions engendered by such a Bayesian approach reflect the model’s confidence in its prediction. Having access to such predictive uncertainties can assist in informed decision-making under uncertainty.

3. Methods

We begin with a physical mixing model, where each endmember $k \in 1, \dots, K$ is characterized by a latent absolute spectral signature \vec{c}_k and a global relative concentration w_k , with $\sum_k w_k = 1$. We assume a linear mixing process, where the observed D -dimensional spectrum \vec{y}_i of mixture i is:

$$\vec{y}_i = \sum_k w_{ik} \vec{c}_k + \vec{e}_i$$

Here, w_{ik} is the relative concentration of endmember k in mixture i , satisfying $\sum_k w_{ik} = 1$, and $\vec{e}_i \sim \mathcal{N}(0, \Sigma)$ is Gaussian observation noise. The standard linear mixing model assumes fixed endmember signatures \vec{c}_k , limiting its flexibility to capture spectral variability across mixtures.

To address this limitation, we propose a hierarchical Bayesian model that treats the \vec{c}_k as random variables with prior distributions representing the variability of each endmember. The model priors are:

$$\begin{aligned} \vec{c}_k &\sim \mathcal{N}(\vec{\mu}_k, \Sigma_k), \quad \vec{\mu}_k = \vec{0}, \quad \Sigma_k = I \\ \vec{w}_i &\sim \text{Dir}(\alpha_1 \vec{w}), \quad \vec{w} \sim \text{Dir}(\alpha_0 \mathbf{1}) \\ \vec{e}_i &\sim \mathcal{N}(0, \Sigma), \quad \Sigma = \text{diag}(\vec{s}), \quad s_d \sim \mathcal{U}(\sigma/2, 3\sigma/2) \end{aligned} \tag{1}$$

The key components of this model, as shown in the plate diagram of Figure 1, are: Endmember spectra \vec{c}_k are drawn from multivariate Gaussian priors with zero mean and identity covariance. This allows for variation in the spectral signatures across mixtures. Mixture-specific abundances \vec{w}_i are drawn from a Dirichlet distribution with parameters $\alpha_1 \vec{w}$, where α_1 controls the diversity of mixtures, and \vec{w} are global relative concentrations. Global concentrations \vec{w} are drawn from a symmetric Dirichlet prior with concentration parameter α_0 . Observation noise \vec{z}_i is Gaussian with zero mean and diagonal covariance Σ parameterized by scale variables s_d with uniform priors.

The use of two Dirichlet distributions in our hierarchical model reflects the non-negativity and sum-to-one constraints on the endmember global weights and mixture fractions. The first Dirichlet distribution serves as a prior for the global endmember frequencies, while the second governs the variability of mixing fractions within each observed mixture.

In contrast to non-probabilistic methods—including the simplex-type approach N-FINDR, which relies on a strong assumption of noise-free data—this probabilistic framework explicitly includes observation noise, allows for incomplete sampling of the endmember simplex, and models the endmember parameters as distributions. By incorporating epistemic uncertainty and supporting the inference of full posterior distributions over endmember parameters, our approach offers a more flexible and principled avenue for unmixing tasks compared to traditional methods that rely on fixed endmember signatures.

The proposed approach can be seen as a Bayesian generalization of simplex-based unmixing methods like N-FINDR, which normally assume the presence of at least one pure pixel per endmember. Our model relaxes this assumption and allows endmembers to vary according to the specified priors, while still enabling the posterior estimates of \vec{c}_k to adapt to the data at hand.

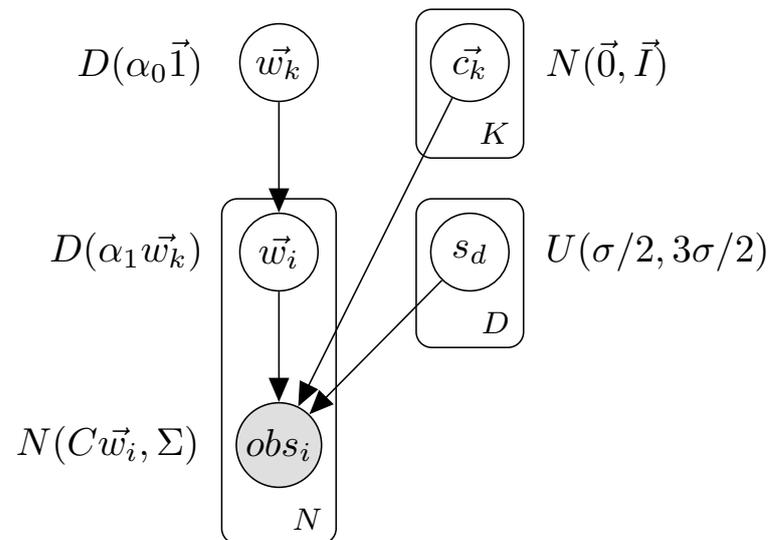


Figure 1. Plate diagram for the Bayesian endmember mixture model. The top-level variables \vec{w}_k , s_d and \vec{c}_k are given by Equation (1) and represent the prevalence of each mixture component, the extent of noise, and the endmember coordinates, respectively. ‘obs’, for ‘observation’, represents an observed mixture, of which N samples from the underlying distribution constitute one generated dataset.

Data Generation and Inference

Sampling under this model is performed once per mixture component k to determine endmember parameters (for a total of K , the number of endmembers); per d th spectral dimension to scale the variance of observation noise in each dimension (for a total of $D = K - 1$, the number of spectral dimensions); and per i th observed data point. The choice of $D = K - 1$ follows from the desire for a direct comparison between the Bayesian hierarchical model and N-FINDR [6]. N-FINDR, in turn, specifies this value of D because its convex set geometry formulation establishes a correspondence between the spectral space and the rank $K - 1$ space of allowed mixing coefficients.

This coupling of D to K simplifies the approach but would, in a practical setting, require prior dimensionality reduction in the data. Second, and more significantly, we have assumed that K is known. The suitability of this assumption is likely domain-dependent. To take a particular example, in the context of phase identification in materials science, the assumption may be reasonable to the extent that K is constrained by considerations such as the Gibbs phase rule.

In a bird's eye overview of our method, we generate simulated data of multiple points and then rely on two methods for inference of endmembers, Hamiltonian Monte Carlo (HMC) and variational inference (VI). Once a probabilistic model has been derived and established in the form of the model expression, one needs to infer the distributions of the parameters in the model expression. This inference is generally intractable for any realistic model due to the integrals involved in the Bayesian estimation [18]. To circumvent this intractability, alternative approaches are utilized. The gold standard in deriving parameter uncertainty is sampling-based Markov Chain Monte Carlo (MCMC) [19] methods, such as Metropolis Hastings. However, these have poor scaling and are computationally infeasible for high-dimensional problems. Recent developments of HMC and, specifically, the No U-Turn Sampler (NUTS) have transformed this status quo [20]. HMC avoids the random walk dynamics of classical MCMC algorithms by using the information in the gradient of the target distribution to inform sampling. NUTS regulates the frequency of sampling and together, HMC-NUTS enables sampling-based uncertainty estimation approaches to be extended to higher dimensions. However, in many cases with higher dimensions, HMC-NUTS may still not scale well and be time consuming. Thus, we also test inference using VI via the mean field approximation. Stochastic VI [21] reduces the sampling problem to an optimization exercise, wherein the divergence between the posterior distribution and a family of parameterized distributions is minimized. This enables the use of classical gradient-based optimizers to solve the problem rapidly and scale to even higher dimensions. However, in line with the no free lunch theorem, these advantages in scaling are accrued at the cost of being unbiased. Based on the assumptions made during VI to simplify the problem for computational and scaling expense, the final solution can be biased [19,22]. This may lead to over-confidence and under-prediction of the uncertainty. To this end, we compare and contrast the solutions that we generate via HMC-NUTS and mean-field VI.

4. Results

After generating 100 simulated datasets of $N = 500$ observations each in the aforementioned manner, we infer endmember coordinates in two ways: first, by using the model distribution function to sample from the posterior over model parameters with the no-U turn (NUTS) HMC sampler; second, with mean-field stochastic variational inference (SVI) using the ADAM optimizer and a learning rate of 0.005. To mitigate failures to converge (a common issue with mixture models), we run multiple rounds of inference with different random initializations. We select between 3 starts per dataset for HMC and 10 for VI; in each case, we retain only the inference round that produces the highest log likelihood for the dataset under the posterior samples (HMC) or approximation (VI). Per dataset, the sampling and inference are run for a total of 10 min on 12 CPU cores.

Figure 2 juxtaposes the endmember coordinates determined through N-FINDR and the hierarchical Bayesian model using HMC with the NUTS sampler. Specifically, it shows representative outcomes of these two methods under the four possible extreme conditions of hyperparameters α_1 and σ : low α_1 with low σ , low α_1 with high σ , high α_1 with low σ , and high α_1 with high σ . This comparison provides a qualitative contrast of the data distribution and the respective behaviors of the two endmember extraction approaches in different corners of the space of parameters α_1 and σ , which, we reiterate, capture the degree of mixing and severity of noise in the formulation. In this comparison, N-FINDR serves not as an example of a state-of-the-art model but rather as a representative instance of deterministic endmember extraction.

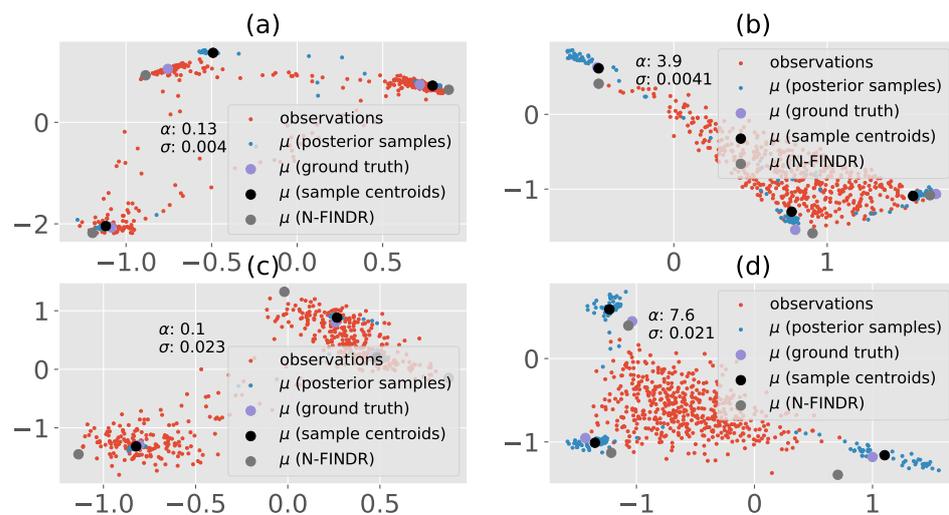


Figure 2. Comparison of endmember coordinates extracted using N-FINDR vs. through sampling of the model posterior using HMC and the no-U turn (NUTS) sampler in Pyro [23]. The panels (a,b) on the top row represent the results of endmember extraction on datasets generated under different conditions of the concentration hyperparameter α_1 : small and large, respectively, coupled with a small noise scale parameter (σ). The bottom row panels (c,d) depict inference results for large noise scale parameter (σ), again contrasting between small and large α_1 .

Further, Figure 3 compares the reconstruction accuracy for 100 instances of this simulation-inference procedure, with 100 distinct combinations of the hyperparameters α_1 and σ . For VI and HMC, we use Monte Carlo samples from the posterior to estimate the central tendency (i.e., estimated mean from MC samples) for the prediction. This is used as the approximation for the point prediction for the Bayesian approaches. With N finder, we use Monte Carlo samples over different seeds, and average these samples for the mean N-FINDR prediction. In each of the cases, due diligence is carried out to ensure that the number of MC samples is sufficient for a robust final prediction. Subsequently, we use the Frobenius norm of the difference between the point prediction and ground truth as our figure of merit for the reconstruction accuracy.

Figure 3 compares the reconstruction accuracy of HMC-NUTS, mean field VI, and the N-FINDR baseline under varying conditions of mixture diversity (α_1) and noise (σ). High α_1 and σ values pose the most difficulty for all approaches, while low values yield better performance. However, when either α_1 or σ is large, HMC-NUTS significantly outperforms N-FINDR. The superior performance of HMC-NUTS in these scenarios aligns with the expected benefits of the Bayesian mixture model approach. N-FINDR assumes the presence of pure endmembers and relies on finding the maximum-volume simplex enclosing the data, struggling when pure endmembers are absent (high α_1). In contrast, the Bayesian approach explicitly models mixing and uncertainty, handling low mixture diversity more effectively. Similarly, the deterministic nature of N-FINDR makes it sensitive to high noise levels (large σ), while the Bayesian model incorporates noise through the likelihood function, providing robustness. The results in Figure 3 confirm these expectations, highlighting the advantages of the probabilistic mixture model approach in challenging conditions where deterministic methods like N-FINDR may falter.

Finally, we investigate posterior approximation with VI as a computationally cheaper alternative to HMC. Figure 4 presents a three-way comparison of the reconstruction error for endmember coordinates under HMC, VI and N-FINDR. As is evident in the figures, HMC and N-FINDR generate the highest- and lowest-quality reconstructions, respectively. Further, mean field VI leads to reconstructions that lie between HMC and N-FINDR. However, VI provides an approximately 100-fold advantage in computation time over HMC, on average.

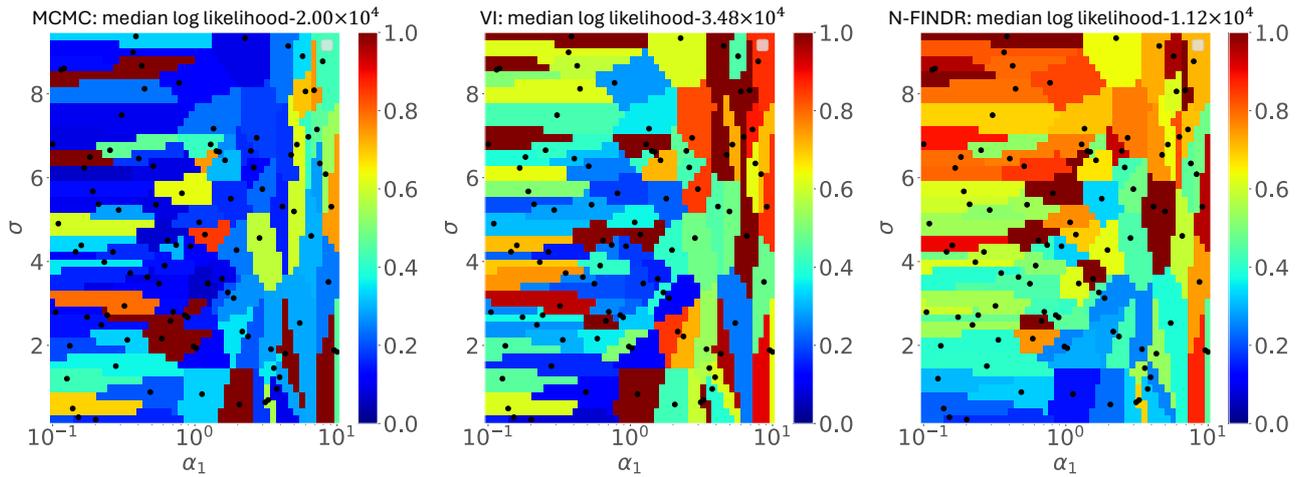


Figure 3. L2 reconstruction error for endmember coordinates via HMC posterior samples (**left**), mean field variational inference (**center**), compared to N-FINDR (**right**), with varying values of the hyperparameters α_1 and σ . The reconstruction error is a sum in quadrature of the L2 distance between the point prediction and the ground truth. The log likelihood of the predictions is also reported in the figure.

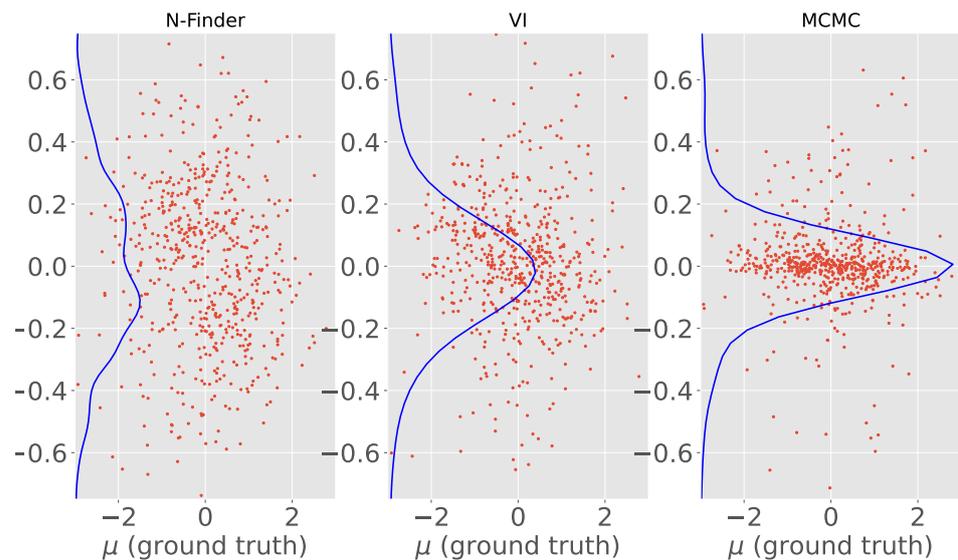


Figure 4. Scatter plots of reconstruction error (horizontal axis) and ground truth (vertical axis), concatenated over the individual scalar components of the endmember coordinates under HMC sampling (**right**), mean field variational inference (**center**), and N-FINDR (**left**). The blue curve in each panel is a smoothed density estimate along the horizontal axis—it represents the empirical probability distribution, over all datasets, for the scalar difference between the endmember coordinate estimates and ground truth.

5. Summary and Future Directions

In summary, we applied a probabilistic mixture model for the extraction of endmembers from spectral mixtures. Our approach formulates the retrieval of endmember spectra and per-observation endmember weights as a Bayesian inference problem using relatively uninformative priors for the phase-mixing and observation processes. Using simulated datasets, we sampled the model posterior with HMC and found that the resulting recovered endmember parameters are more robust to observation noise and absence, in the data, of pure endmember spectra compared to the popular N-FINDR algorithm. This Bayesian approach provides a principled framework for incorporating relevant prior infor-

mation without introducing undue assumptions and opens the door to other intrinsically probabilistic analyses, such as uncertainty quantification.

Both the baseline model (N-FINDR) and our Bayesian approach assume that the number of endmembers is known. Unlike deterministic approaches such as N-FINDR, the probabilistic formulation would allow one to relax this potentially onerous assumption by replacing the fixed number of mixture components and corresponding endmember weights by a set of samples from a stick-breaking distribution [24] as in a Dirichlet process mixture model. Assumptions about K would thus be condensed into a single hyperparameter, the concentration parameter of the stick-breaking process. Additionally, our approach can be generalized to incorporate more flexible, non-linear mappings using Bayesian Neural Networks (BNNs) to infer embeddings.

Author Contributions: Conceptualization, O.H., A.A.M. and A.M.; methodology, O.H. and A.A.M.; software, O.H. and A.A.M.; validation, O.H., A.A.M. and A.M.; formal analysis, O.H., A.A.M. and A.M.; investigation, O.H., A.A.M. and A.M.; resources, O.H., A.A.M. and A.M.; data curation, O.H.; writing—original draft preparation, O.H., A.A.M. and A.M.; writing—review and editing, O.H., A.A.M. and A.M.; visualization, O.H.; supervision, A.M.; project administration, A.M.; funding acquisition, A.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work was performed and partially supported by the US Department of Energy (DOE), Office of Science, Office of Basic Energy Sciences Data, Artificial Intelligence and Machine Learning at the DOE Scientific User Facilities program under the MLExchange Project (award No. 107514). Aashwin Ananda Mishra was partially supported by the SLAC ML Initiative.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Ottino, J.M. *The Kinematics of Mixing: Stretching, Chaos, and Transport*; Cambridge University Press: Cambridge, UK 1989; Volume 3.
- Heller, J.P. An unmixing demonstration. *Am. J. Phys.* **1960**, *28*, 348–353. [[CrossRef](#)]
- Zhang, L.; Gao, J.; Damoah, L.N.W.; Robertson, D.G. Removal of iron from aluminum: A review. *Miner. Process. Extr. Metall. Rev.* **2012**, *33*, 99–157. [[CrossRef](#)]
- Gary, J.H.; Handwerk, J.H.; Kaiser, M.J.; Geddes, D. *Petroleum Refining: Technology and Economics*; CRC Press: Boca Raton, FL, USA, 2007.
- Smith, I. *Chromatography*; Elsevier: Amsterdam, The Netherlands, 2013.
- Winter, M.E. N-FINDR: An algorithm for fast autonomous spectral end-member determination in hyperspectral data. In *Imaging Spectrometry V*; SPIE: Bellingham, WA, USA, 1999; Volume 3753, pp. 266–275.
- Miao, L.; Qi, H. Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 765–777. [[CrossRef](#)]
- Pauca, V.P.; Piper, J.; Plemmons, R.J. Nonnegative Matrix Factorization for Spectral Data Analysis. *Linear Algebra Its Appl.* **2006**, *416*, 29–47. [[CrossRef](#)]
- Jia, S.; Qian, Y. Constrained nonnegative matrix factorization for hyperspectral unmixing. *IEEE Trans. Geosci. Remote Sens.* **2008**, *47*, 161–173. [[CrossRef](#)]
- Iordache, M.D.; Bioucas-Dias, J.M.; Plaza, A. Sparse Unmixing of Hyperspectral Data. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 2013–2039. [[CrossRef](#)]
- Iordache, M.D.; Bioucas-Dias, J.M.; Plaza, A. Collaborative sparse regression for hyperspectral unmixing. *IEEE Trans. Geosci. Remote Sens.* **2013**, *52*, 341–354. [[CrossRef](#)]
- Izenman, A.J. Introduction to manifold learning. *Wiley Interdiscip. Rev. Comput. Stat.* **2012**, *4*, 439–446. [[CrossRef](#)]
- Roweis, S.T.; Saul, L.K. Nonlinear dimensionality reduction by locally linear embedding. *Science* **2000**, *290*, 2323–2326. [[CrossRef](#)] [[PubMed](#)]
- Balasubramanian, M.; Schwartz, E.L. The isomap algorithm and topological stability. *Science* **2002**, *295*, 7. [[CrossRef](#)] [[PubMed](#)]
- Palsson, F.; Sveinsson, J.R.; Ulfarsson, M.O.; Benediktsson, J.A. Hyperspectral Unmixing Using a Neural Network Autoencoder. *IEEE Access* **2018**, *6*, 25646–25656. [[CrossRef](#)]
- Zhang, X.; Sun, Y.; Zhang, J.; Wu, P.; Jiao, L. Hyperspectral unmixing via deep convolutional neural networks. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 1755–1759. [[CrossRef](#)]
- Qi, L.; Li, J.; Wang, Y.; Lei, M.; Gao, X. Deep spectral convolution network for hyperspectral image unmixing with spectral library. *Signal Process.* **2020**, *176*, 107672. [[CrossRef](#)]

18. Murphy, K.P. *Machine Learning: A Probabilistic Perspective*; MIT Press: Cambridge, MA, USA, 2012.
19. Murphy, K.P. *Probabilistic Machine Learning: An Introduction*; MIT Press: Cambridge, MA, USA, 2022.
20. Betancourt, M. A conceptual introduction to Hamiltonian Monte Carlo. *arXiv* **2017**, arXiv:1701.02434.
21. Hoffman, M.D.; Blei, D.M.; Wang, C.; Paisley, J. Stochastic variational inference. *J. Mach. Learn. Res.* **2013**, *14*, 1303–1347.
22. Mishra, A.A.; Edelen, A.; Hanuka, A.; Mayes, C. Uncertainty quantification for deep learning in particle accelerator applications. *Phys. Rev. Accel. Beams* **2021**, *24*, 114601. [[CrossRef](#)]
23. Bingham, E.; Chen, J.P.; Jankowiak, M.; Obermeyer, F.; Pradhan, N.; Karaletsos, T.; Singh, R.; Szerlip, P.; Horsfall, P.; Goodman, N.D. Pyro: Deep universal probabilistic programming. *J. Mach. Learn. Res.* **2019**, *20*, 973–978.
24. Teh, Y.W. Dirichlet Process. *Encycl. Mach. Learn.* **2010**, *1063*, 280–287.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.