



## Bayesian Nonnegative Matrix Factorization with Volume Prior for Unmixing of Hyperspectral Images

**Arngren, Morten; Schmidt, Mikkel Nørgaard; Larsen, Jan**

*Published in:*

2009 IEEE International Workshop on Machine Learning for Signal Processing (MLSP 2009)

*Link to article, DOI:*

[10.1109/MLSP.2009.5306262](https://doi.org/10.1109/MLSP.2009.5306262)

*Publication date:*

2009

*Document Version*

Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

*Citation (APA):*

Arngren, M., Schmidt, M. N., & Larsen, J. (2009). Bayesian Nonnegative Matrix Factorization with Volume Prior for Unmixing of Hyperspectral Images. In 2009 IEEE International Workshop on Machine Learning for Signal Processing (MLSP 2009) (pp. 1-6). IEEE. DOI: 10.1109/MLSP.2009.5306262

## DTU Library

Technical Information Center of Denmark

---

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# BAYESIAN NONNEGATIVE MATRIX FACTORIZATION WITH VOLUME PRIOR FOR UNMIXING OF HYPERSPECTRAL IMAGES

Morten Arngren<sup>†°</sup>, Mikkel N. Schmidt<sup>‡</sup> and Jan Larsen<sup>†</sup>

<sup>†</sup> Technical University of Denmark, DTU Informatics, Richard Petersens Plads, DK-2800 Lyngby

<sup>‡</sup> University of Cambridge, Department of Engineering, Trumpington Street, Cambridge CB2 1PZ, UK

<sup>°</sup> FOSS Analytical A/S, Slangerupgade 69, DK-3400 Hilleroed

## ABSTRACT

In hyperspectral image analysis the objective is to unmix a set of acquired pixels into pure spectral signatures (endmembers) and corresponding fractional abundances. The Non-negative Matrix Factorization (NMF) methods have received a lot of attention for this unmixing process. Many of these NMF based unmixing algorithms are based on sparsity regularization encouraging pure spectral endmembers, but this is not optimal for certain applications, such as foods, where abundances are not sparse. The pixels will theoretically lie on a simplex and hence the endmembers can be estimated as the vertices of the smallest enclosing simplex. In this context we present a Bayesian framework employing a volume constraint for the NMF algorithm, where the posterior distribution is numerically sampled from using a Gibbs sampling procedure. We evaluate the method on synthetic and real hyperspectral data of wheat kernels.

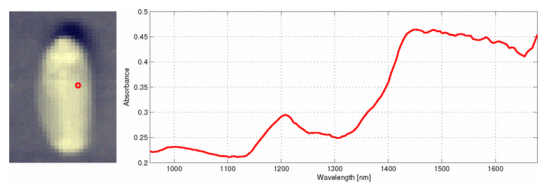
## 1. INTRODUCTION

Traditional image acquisition and analysis is based on three color bands, red, green, and blue, which is sufficient for human visualization. In the context of identifying or extracting material constituents of e.g. foods, three channels are rarely enough.

Hyperspectral image analysis offers the opportunity to capture detailed spectral information for improved image analysis. Each of the observed pixels can be considered a mix of latent pure constituents. In foods, these constituents are typically *protein*, *starch*, *H<sub>2</sub>O*, etc. The non-negative spectral signatures of these pure constituent are denoted *endmembers*. Each pixel also has a set of corresponding concentrations denoted *fractional abundances*, which are non-negative and must sum to one (additivity constraint). In a reflectance acquisition system the mixing can be considered linear according to Lambert-Beer's law. Non-linear mixing are usually neglected due to minor significance and increased modeling complexity [2].

Figure 1 illustrates a hyperspectral image of a wheat

kernel with a corresponding pre-processed spectrum from 950 – 1650nm.



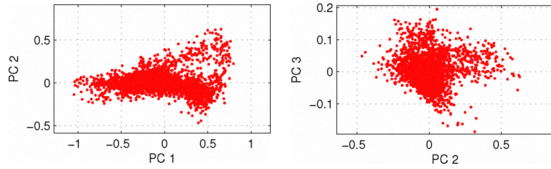
**Fig. 1.** Pseudo RGB image of a wheat kernel and the associated pre-processed spectrum for the selected pixel.

If we denote the  $N$  observed  $M$  dimensional spectra as the matrix  $\mathbf{X} \in \mathbb{R}^{M \times N}$ , the linear mixing can be expressed as a  $K$  rank linear factorization written as

$$\mathbf{X} = \mathbf{W}\mathbf{H} + \epsilon, \quad (1)$$

where the  $K$  non-negative endmembers are contained in  $\mathbf{W} \in \mathbb{R}_+^{M \times K}$ ,  $\mathbf{H} \in \mathbb{R}_+^{K \times N}$  holds the fractional abundances and  $\epsilon$  denote the residual noise. The objective of spectral unmixing is to decompose the pixels in order to extract the latent endmembers  $\mathbf{W}$  and corresponding fractional abundances  $\mathbf{H}$ .

The non-negativity and sum-to-unity assumption of  $\mathbf{H}$  implies a multidimensional simplex structure of the modeled data, where the vertices denote the endmembers. The observed data may not span the entire simplex due to the mixing of the constituents. For food applications pure endmembers appear rarely as the observed pixels are almost always a mix of constituents. This means the observed data will concentrate around the center of the simplex and very few samples can be expected at the vertices. The data acquisition is further subject to additive noise and thus the simplex structure will not be perfect. Figure 2 illustrates the simplex structure of wheat kernel image pixels. To unmix the hyperspectral image data, different approaches can be used. One approach is to analyse the data using convex geometrical methods. Plaza *et al.* (2004) [9] gives a comparative analysis of a set of endmember extraction methods



**Fig. 2.** Scatterplot of 1st, 2nd and 3rd principal component of the processed wheat kernel hyperspectral data.

based on geometrical analysis.

As the acquired data is non-negative by nature the Non-negative Matrix Factorization (NMF) has received much attention [8, 6, 7, 4]. The basic linear NMF model minimizes the squared error  $\|\mathbf{X} - \mathbf{WH}\|^2$  subject to non-negativity constraint on the elements  $\mathbf{W}$  and  $\mathbf{H}$ . This is however rarely sufficient to capture the endmembers in an unmixing problem and thus additional regularization is required. This can be expressed as part of the cost function as

$$C(\mathbf{W}, \mathbf{H}) = \frac{1}{2} \|\mathbf{X} - \mathbf{WH}\|^2 + \gamma J_w(\mathbf{W}) + \beta J_h(\mathbf{H}), \quad (2)$$

where  $J_w(\mathbf{W})$  and  $J_h(\mathbf{H})$  denote the regularization terms for the endmembers and fractional abundances respectively.

Sajda *et al.* [10] develop an NMF algorithm with additional constraints on the amplitude of the estimated spectral components, with improved endmember determination. An  $L_2$ -norm sparsity prior on both the endmembers  $\mathbf{W}$  and the fractional abundances  $\mathbf{H}$  is incorporated by Pauca *et al.* [8] also leading to improved estimation of the endmembers. A sparsity prior on the fractional abundances encourages pure spectra among the observed pixels, but this is rarely the case in foods and hence the sparsity prior is not useful.

Instead a prior can be based on the volume of the simplex. For a  $K$ -simplex in  $M$ -dimensional space for  $K = M$  and with vertices  $\{\mathbf{w}_0, \mathbf{w}_1, \dots, \mathbf{w}_M\}$  the volume is given by  $V_{simplex} = \frac{1}{K!} |\det(\tilde{\mathbf{W}})|$ , where  $\tilde{\mathbf{W}} = [\mathbf{w}_1 - \mathbf{w}_0 \ \mathbf{w}_2 - \mathbf{w}_0 \ \dots \ \mathbf{w}_M - \mathbf{w}_0]$  denote the vectors spanning the simplex. In our case  $K \ll M$  and thus the simplex volume can be expressed as

$$V_{simplex} = \frac{1}{K!} |\det(\tilde{\mathbf{W}}^T \tilde{\mathbf{W}})|^{1/2} \quad (3)$$

The simplex volume can be approximated by different means: Miao *et al.* [3] present an NMF algorithm that incorporates the following volume regularization (denoted *mvc*)

$$J_w(\mathbf{W}) \propto \det^2(\mathbf{C} + \mathbf{BU}_x^T(\mathbf{W} - \mu)), \quad (4)$$

where  $\mathbf{U}_x$  is a projection matrix defined as the  $K - 1$  principal components of the observed data  $\mathbf{X}$ ;  $\mu$  denote the mean of  $\mathbf{X}$ ; and  $\mathbf{B}$  and  $\mathbf{C}$  ensure a square matrix for the determinant. This volume regularization captures the volume of the centered simplex in the subspace defined by the  $K - 1$  eigenvectors. This approach is intuitive, but because of its

dependency on the observed data it does not have an interpretation as a prior in a Bayesian framework.

A different volume regularization is proposed in Schachtner *et al.* [11] based on the volume of the parallelepiped spanned by the endmembers. Based on this, the authors present a regularized NMF multiplicative update algorithm. The regularization is based on the determinant of the estimated endmembers expressed for the non-square case of  $\mathbf{W}$  as (denoted *pp*)

$$J_w(\mathbf{W}) = \det(\mathbf{W}^T \mathbf{W}). \quad (5)$$

This regularization can be seen as a direct approximation of (3), where the absolute vectors  $\mathbf{W}$  are used instead of the simplex spanning vectors  $\tilde{\mathbf{W}}$ . The omission of the square root leads to non-linear regularization, where large volumes will be penalized harder than small volumes. This regularization can be interpreted as a prior in a Bayesian framework, as it avoids the dependency to the observed data. It however suffers from including an offset mean  $\mu_w$  leading to potential movement toward origo when minimizing the volume.

As opposed to regularized least squares, a different approach is to build a Bayesian framework for hyperspectral unmixing. This allows to define appropriate likelihood and prior distributions equivalent to the NMF formulation and to extract confidence intervals on a MAP estimate. In Ochs *et al.* [5] a Bayesian Spectral Decomposition (BSD) algorithm is introduced based on an atomic prior. Moussaoui *et al.* [4] present a Bayesian approach to separate non-negative mixtures of NIR data based on a hybrid Gibbs-Metropolis-Hastings sampling procedure. Schmidt *et al.* [12] develop a Bayesian NMF based on a exponential sparsity prior and inverse gamma hyper priors.

In this paper we present a Bayesian NMF method with a volume prior for unmixing hyperspectral images. Our proposed method incorporates a determinant based prior inspired by the *pp* volume in (5). The additivity constraint on the fractional abundances is modeled as a uniform prior on the unit simplex. For inference in the model, a fast Gibbs sampling procedure is devised.

For comparison to our proposed Bayesian approach, the volume regularization terms *mvc* in (4) and *pp* in (5) is implemented in a projected gradient NMF framework. The additivity constraint is not included in the work of Schachtner *et al.* [11] and is implemented as a soft-constraint by Miao *et al.* [3] in their algorithm. This leads to a trade-off between describing the observed data and respecting the additivity constraint. In our projected gradient framework the additivity constraint is implemented with a variable substitution approach, which always fullfills the additivity constraint without sacrificing the description of the observed data. We evaluate the methods on artificial and real hyperspectral food image data.

In section 2 we present our Bayesian volume constrained NMF model as well as the Gibbs sampler. The synthetic and real data are described in section 3 with a short description of the initial pre-processing. Finally, in section 4 we evaluate how the proposed method can extract the true endmembers and corresponding fractional abundances, and compare to existing methods.

## 2. BAYESIAN NMF WITH VOLUME PRIOR

We here present a probabilistic Bayesian formulation of NMF with a prior that constrains the volume of the data simplex. We model the joint probability distribution of the endmembers,  $\mathbf{W}$ , and the fractional abundances,  $\mathbf{H}$ , as well as the noise, conditioned on the observed data,  $\mathbf{X}$ , and a set of model hyper-parameters,  $\Theta$ . To this end we must choose a suitable noise model as well as reasonable prior distributions over all parameters in the model.

We model the noise,  $\epsilon$ , as independent and identically distributed white Gaussian noise, which gives rise to the following likelihood function,

$$p(\mathbf{X}|\mathbf{W}, \mathbf{H}, \Theta) = \prod_{n=1}^N \prod_{m=1}^M \mathcal{N}(x_{mn}|\mathbf{W}_m \cdot \mathbf{H}_{:n}, \sigma^2), \quad (6)$$

where  $\mathcal{N}(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$  is the Gaussian probability density function.

In our notation, matrices are denoted by capital bold letters. A matrix with two subscripts denotes the submatrix with the corresponding rows and columns, where a colon denotes all indices, and  $\tilde{m}$  denotes all indices except  $m$ ; a single element is denoted by  $x_{mn}$ .

The likelihood has a single parameter, the noise variance,  $\sigma^2$ , for which we choose an conjugate prior, i.e., an inverse-Gamma distribution,

$$p(\sigma^2) = \mathcal{IG}(\sigma^2|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \left(\frac{1}{\sigma^2}\right)^{\alpha+1} \exp\left(-\frac{\beta}{\sigma^2}\right). \quad (7)$$

We choose a prior distribution for the endmembers, that encourages the simplex spanned by the estimated endmembers to be small. We base this prior on a determinant criterion, similar to [11], that measures the volume of the parallelepiped spanned by the endmembers. Furthermore, the prior includes the constraint that each element in the end-member matrix must be non-negative,

$$p(\mathbf{W}|\Theta) \propto \begin{cases} \exp(-\gamma \det(\mathbf{W}^\top \mathbf{W})) & w_{mk} \geq 0 \\ 0 & \text{otherwise.} \end{cases} \quad (8)$$

The prior has a single parameter,  $\gamma$ , that determines the strength of the volume penalty. This parameter could possibly be modeled hierarchically; however, presently we set the parameter by hand.

For the fractional abundances, the prior must enforce non-negativity as well as the constraint that the abundances for each pixel must sum to unity. We choose a uniform prior on the unit simplex,

$$p(\mathbf{H}|\Theta) \propto \begin{cases} 1 & h_{kn} \geq 0, \sum_{k=1}^K h_{kn} = 1 \\ 0 & \text{otherwise.} \end{cases} \quad (9)$$

### 2.1. Gibbs sampler

Inference in the proposed probabilistic NMF model can be conducted using a Gibbs sampling procedure, in which we sequentially draw samples from the posterior conditional distribution of each parameter of the model. Due to our choice of priors, we can sample from all conditional distributions directly using standard methods, which obviates slow sampling procedures such as rejection sampling.

The conditional distribution of the noise variance is an inverse-Gamma,

$$p(\sigma^2|\mathbf{X}, \theta \setminus \sigma^2) = \mathcal{IG}(\sigma^2|\bar{\alpha}, \bar{\beta}) \quad (10)$$

$$\begin{aligned} \bar{\alpha} &= \alpha + \frac{1}{2}NM - 1 \\ \bar{\beta} &= \beta + \frac{1}{2} \sum_{m=1}^M \sum_{n=1}^N (x_{mn} - \mathbf{W}_m \cdot \mathbf{H}_{:n})^2. \end{aligned} \quad (11)$$

The conditional distribution of the endmembers, which arises from the product of the Gaussian likelihood and the determinant-based prior, is a truncated Gaussian density,

$$p(w_{mk}|\mathbf{X}, \theta \setminus w_{mk}) \propto \begin{cases} \mathcal{N}(w_{mk}|\bar{\mu}_{mk}, \bar{\sigma}_{mk}^2) & w_{mk} \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (12)$$

$$\bar{\sigma}_{mk}^{-2} = (\mathbf{H}_{k:} \mathbf{H}_{k:}^\top) \sigma^{-2} + \gamma (D_{\tilde{k}\tilde{k}} - \mathbf{W}_{m\tilde{k}} \mathbf{A}_{\tilde{k}\tilde{k}} \mathbf{W}_{m\tilde{k}}^\top) \quad (13)$$

$$\begin{aligned} \bar{\mu}_{mk} &= \bar{\sigma}_{mk}^2 \left( (\mathbf{W}_{m\tilde{k}} \mathbf{A}_{\tilde{k}\tilde{k}} \mathbf{W}_{m\tilde{k}}^\top \mathbf{W}_{\tilde{m}k}) \gamma + \right. \\ &\quad \left. (\mathbf{X}_{m:} \mathbf{H}_{k:}^\top - \mathbf{W}_{m\tilde{k}} \mathbf{H}_{\tilde{k}:} \mathbf{H}_{k:}^\top) \sigma^{-2} \right) \end{aligned} \quad (14)$$

where  $D_{\tilde{k}\tilde{k}} = \det(\mathbf{W}_{\tilde{k}:}^\top \mathbf{W}_{\tilde{k}:})$  and  $\mathbf{A}_{\tilde{k}\tilde{k}} = \text{adj}(\mathbf{W}_{\tilde{k}:}^\top \mathbf{W}_{\tilde{k}:})$ . Samples from the truncated Gaussian distribution can be generated, e.g., using the method described by Geweke [1].

The conditional density of the fractional abundances is a constrained truncated Gaussian,

$$p(\mathbf{H}_{:n}|\mathbf{X}, \theta \setminus \mathbf{H}_{:n}) \propto \begin{cases} \mathcal{N}(\mathbf{H}_{:n}|\bar{\boldsymbol{\mu}}_n, \bar{\boldsymbol{\Sigma}}_n) & h_{kn} \geq 0, \sum_{k=1}^K h_{kn} = 1 \\ 0 & \text{otherwise} \end{cases} \quad (15)$$

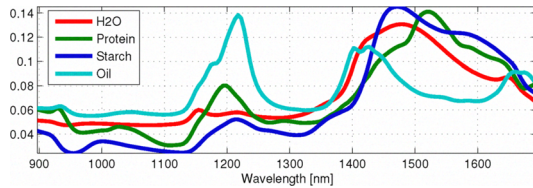
$$\bar{\mu}_n = (\mathbf{W}^\top \mathbf{W})^{-1} \mathbf{W}^\top \mathbf{X}_{:n}, \quad \bar{\Sigma}_n = \sigma^2 (\mathbf{W}^\top \mathbf{W})^{-1}. \quad (16)$$

In the remainder of our paper we denote the Bayesian framework presented *BayesNMF-Vol*.

### 3. DATA ACQUISITION AND PRE-PROCESSING

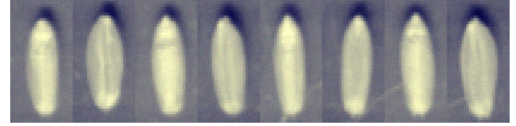
The hyperspectral image data was acquired using a line-scan NIR camera from 900-1700nm in 165 bands leading to a data cube of  $320 \times \text{lines} \times 165$ . Prior to the image data being subjected to our unmixing algorithm a series of pre-processing steps were carried out. Initially, the raw image data was compensated for white reference and dark current background spectra in order to remove light source characteristics and sensor offset. Because the camera sensor has very poor signal-to-noise ratio from 900 – 950nm and 1650 – 1700nm, these ranges were removed. In order to suppress noise, each line was scanned twice and averaged. Afterwards, each spectra was converted to absorbance in order to obey Lambert-Beer's law.

For reference we acquired spectra of pure food constituents, *protein*, *starch* and *oil*, using the camera system and pre-processed as described. Figure 3 illustrates the spectra for these pure compounds.



**Fig. 3.** Normalized spectra of pure basic food constituents. The water spectrum may suffer from poor SNR as  $\text{H}_2\text{O}$  has extremely high absorption rates from 1400 – 1700nm.

For our analysis, hyperspectral images of four wheat kernels were acquired on both front and back side leading to eight images. These images were pre-processed as described, and corrected for scatter effects. As sparse spectral peaks are not prominent in NIR data, low order scatter correction was applied as the residual from a first order polynomial fit. This approach preserves the simplex structure. The pure background pixels were then identified and removed from the data set, leaving a few around the kernel periphery. This segmentation was achieved by discriminating the first principal component applied on the reflectance data. Afterwards, the individual kernels were cropped. The final hyperspectral image data set was then represented as an unfolded matrix,  $\mathbf{X} \in \mathbb{R}^{M \times N}$ , where each  $M$  dimensional column vector  $\mathbf{x}$  is a pre-processed spectrum of a pixel. Figure 4 shows the pre-processed kernels prior to the background segmentation.



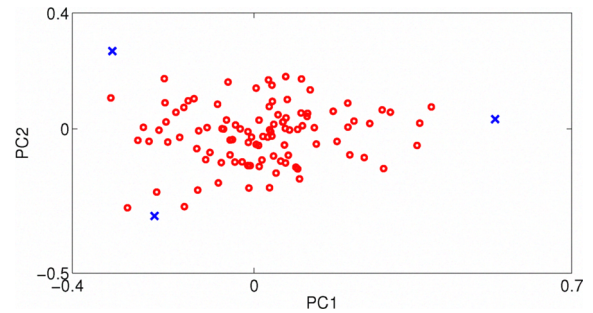
**Fig. 4.** Pseudo RGB image of wheat kernels 1-4 with pairwise front and backside.

### 4. SIMULATIONS

The two volume constrained NMF methods were evaluated in a set of simulations using synthetic data with different profiles and known labels. Afterwards our BayesNMF-Vol model was applied to real hyperspectral image data of wheat kernels.

#### 4.1. Synthetic data

A synthetic dataset was produced by forming a 3-simplex in 3 dimensions based on randomly generated endmembers  $\mathbf{W}$ . Based on these endmembers, 100 samples were generated with full mixing profile and a noise variance  $\sigma^2 = 10^{-3}$ . This leads to a data matrix  $\mathbf{X}$  of  $3 \times 100$  illustrated in figure 5.



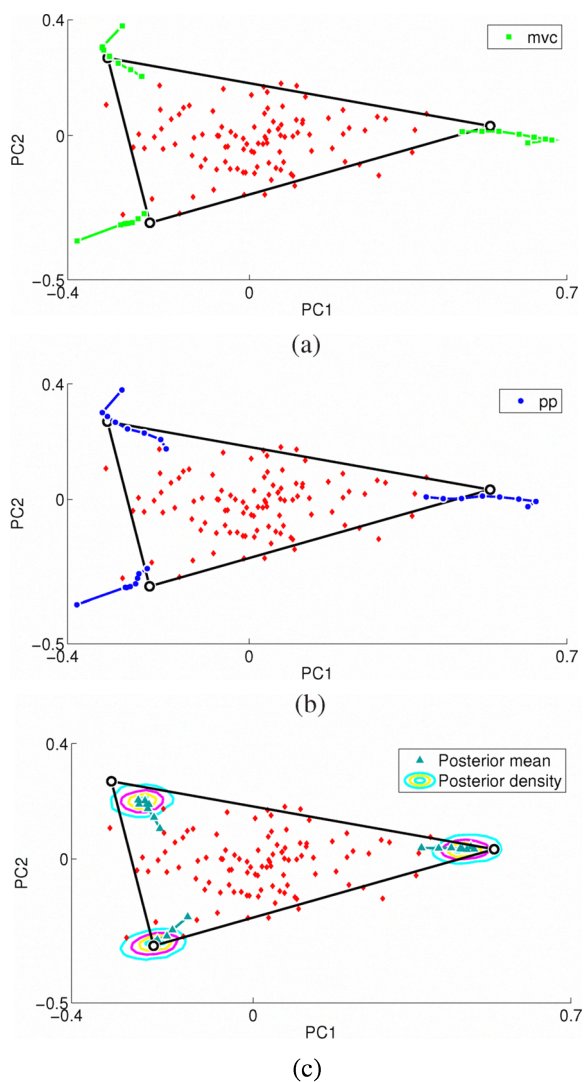
**Fig. 5.** Scatter plot of the synthetic data, 1st and 2nd PC. Circles denote datapoints and crosses denote endmembers.

Both the mvc, the pp regularized and the BayesNMF-Vol NMF algorithms were evaluated by applying the synthetic dataset to find their endmember MAP estimate  $\mathbf{W}_{map}$  for different values of the regularization parameter  $\gamma$ . The posterior mode estimate was computed for the BayesNMF-Vol method. All algorithms were initialized from randomly selected observed data points. A small range of regularization parameter values  $\gamma = \{0, .01, .02, .05, .1, .2, .5, 1\}$  were applied and the resulting scatter plots are shown in figure 6.

The plots depict the converged endmember estimates for all three approaches. The models successfully capture the data structure and increasing regularization encourages



smaller volumes. The advantage of our Bayesian model is that it allows us to estimate the full posterior producing confidence measures. The posterior mode estimates are further comparable with the performance of the two projected gradient NMF methods. In addition, the pp regularization indicate superior convergence properties, both in terms of significantly fewer iterations and slightly improved endmember estimate in average.



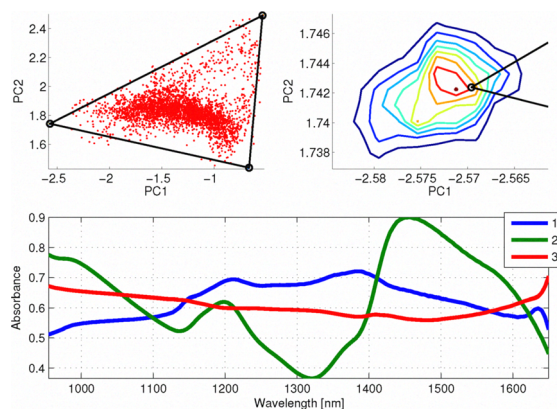
**Fig. 6.** Scatter plots of the observed data and the corresponding estimated endmembers for all 3 methods.

#### 4.2. Wheat kernel data

A wheat kernel consists of many different constituents, where the majority is  $H_2O$ , starch, protein and oil, i.e. we expect to extract 4 – 5 endmembers incl. background in our analysis. A set of reference concentrations for the compounds

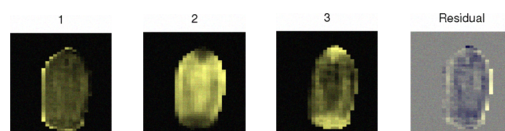
(typically protein) were unfortunately not available, so we are forced resort to subjective evaluation of the unmixing. The 4 kernels may also suffer from only small variations in the protein level between the pixel, i.e. strong correlation with other constituents. In such case the spectral profile of protein is almost impossible to extract. The corresponding PCA scatterplots of the wheat kernel pixels data indicate a simplex structure of 3-4 components (not shown).

The BayesNMF-Vol algorithm was initialized with pp regularized NMF endmember estimates in order to avoid long burn-in periods. A proper prior parameter was similarly found through empirical experiments. Figure 7 shows the estimated posterior mode endmembers and their spectral profiles for 10000 samples,  $K = 3$  endmembers and  $\gamma = 30$ . A single decomposed wheat kernel is illustrated in figure 8.



**Fig. 7.** Successful encapsulation of wheat kernel data. Note the burn-in samples on the right top illustration. The extracted spectral profiles, where the 2nd (green) is easily identified as starch comparing to figure 3.

Initially the background has been extracted by the model as the 3rd component. The starch spectral profile has also been identified very clearly along with a spatial distribution. Similarly the oil in the germ part can be identified primarily from the spatial distribution.



**Fig. 8.** Decomposition of backside of wheat kernel 3 (6th kernel in figure 4).

### 4.3. Discussion

All the volume based NMF methods described suffer from high sensitivity to the number of components,  $K$ . If  $K$  is too large, excessive components are typically split up in two.

The volume prior  $p(\mathbf{W})$  based on (5) has a fatal sensitivity to linear dependencies among the estimated endmembers leading to a collapsing volume. This can occur if the estimated amount of components  $K$  is higher than the true, where any excess endmember potentially fails to model the simplex. It can also occur for strong regularizations, i.e. large value of  $\gamma$ , leading to a collapse of the volume, as it is never allowed to expand and capture the data.

The experiments further reveal a high sensitivity to initializations as a bad start guess may require a temporary higher volume for rotation in order to settle correctly depending on data structure. If the volume prior  $p(\mathbf{W})$  is set to penalize high volumes, then we might get stuck in a local minimum.

In addition our BayesNMF algorithm indicate a latent regularization on the estimated endmembers. For a flat prior distribution  $p(\mathbf{W})$  with  $\gamma = 0$ , the BayesNMF-Vol still exhibit a small regularization encouraging small volumes. This means the observed data structure can not be encapsulated entirely. This effect is subject to further investigation.

The preliminary empirical analysis can easily be extended to more complex datasets in order to reveal a more detailed performance map in terms of large datasets with different mixing and noise profiles.

## 5. CONCLUSION

We have proposed a Bayesian NMF model employing a volume based prior more suitable for hyperspectral image analysis of foods. Using a synthetic data set our model indicate improved or similar unmixing performance compared to existing volume regulated NMF models.

Our BayesNMF-Vol model further extracts confidence intervals to the MAP estimate in terms of sampling variance.

Limitation to the performance of our model has also been identified and further research will focus on improving the framework.

## 6. REFERENCES

- [1] J. Geweke, *Efficient Simulation from the Multivariate Normal and Student-t Distributions Subject to Linear Constraints and the Evaluation of Constraint Probabilities*, 1991 in Computing Science and Statistics: Proceedings of the 23rd Symposium, p.571-578
- [2] Y. M. Masalmah, *Unsupervised Unmixing Of Hyperspectral Imagery Using The Constrained Positive Matrix Factorization*, Computing And Information Science And Engineering, University Of Puerto Rico, Mayaguez Campus, 2007
- [3] L. Miao and H. Qi, *Endmember Extraction From Highly Mixed Data Using Minimum Volume Constrained Nonnegative Matrix Factorization*, 2007 in Geoscience and Remote Sensing, IEEE Transactions on, Vol.45:3 p.765-777
- [4] S. Moussaoui, D. Brie, A. Mohammad-Djafari and C. Carteret, *Separation of non-negative mixture of non-negative sources using a bayesian approach and MCMC sampling*, 2006 in IEEE Transactions on Signal Processing, Vol.54:11 p.4133-4145
- [5] M. F. Ochs, R. S. Stoyanova, F. Arias-Mendoza and T. R. Brown, *A New Method for Spectral Decomposition Using a Bilinear Bayesian Approach*, 1999 in Journal of Magnetic Resonance, Vol.137:1 p.161-176
- [6] P. Paatero and U. Tapper, *Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values*, 1994 in Environmetrics, Vol.5:2 p.111-126
- [7] L. Parra, C. Spence, P. Sajda, A. Ziehe and K. Mller, *Unmixing Hyperspectral Data*, 1999 in Neural Information Processing Systems, Vol.12
- [8] P. V. Pauca, J. Piper and R. J. Plemmons, *Nonnegative Matrix Factorization for Spectral Data Analysis*, 2006 in Linear Algebra and Its Applications 416:1 p.29-47
- [9] A. Plaza, P. Martinez, R. Prez and J. Plaza, *A Quantitative and Comparative Analysis of Endmember Extraction Algorithms From Hyperspectral Data*, 2004 in IEEE Transactions on Geoscience and Remote Sensing, Vol.42:3
- [10] P. Sajda, S. Du and L. C. Parra, *Recovery of constituent spectra using non-negative matrix factorization*, 2003 in Proceedings of the SPIE - The International Society for Optical Engineering, Vol.5207:1 p.321-331
- [11] R. Schachtner, G. Poppel, A. M. Tom and E. W. Lang, *Minimum Determinant Constraint for Non-negative Matrix Factorization*, 2009 in Lecture Notes in Computer Science, Vol.5441 p.106-113
- [12] M. N. Schmidt, O. Winther and L. K. Hansen, *Bayesian non-negative matrix factorization*, 2009 in Lecture Notes in Computer Science (LNCS), Independent Component Analysis and Signal Separation, International Conference on, Vol.5441 p.540-547