# NONLINEAR SPECTRAL UNMIXING OF MODEL EFFERVESCENT PARACETAMOL FORMULATIONS USING Bézier SURFACES

*Zakaria Bnoulkacem[1] Bikram Koirala[1], Paul Scheunders[1]*

*[1] Imec-Visionlab, University of Antwerp (CDE) Universiteitsplein 1, B-2610 Antwerp*

## ABSTRACT

Estimating the composition of intimately mixed pharmaceutical powders from spectral reflectance is complicated by multiple scattering, particle-size and packing effects, and the nonlinear relation between composition and the measured spectrum. We propose a supervised nonlinear unmixing approach to model effervescent paracetamol formulations. The main idea is to represent the spectral data manifold using a high-dimensional Bézier surface whose uniformly distributed control points correspond to measured mixtures of citric acid, sodium bicarbonate, and paracetamol. To demonstrate the potential of the proposed approach, a total of 66 mixtures uniformly covering a ternary composition simplex at 10% intervals were prepared. The spectral reflectances of these mixtures were acquired using an ASD spectrometer over the wavelength range of 350-2500 nm. To estimate the mixture composition, the proposed method minimizes the reconstruction error between the input spectral reflectance and the reflectance sampled from the Bézier surface, subject to non-negativity and sum-to-one constraints.

When utilizing the entire dataset of 66 mixtures to generate the Bézier surface, the method achieves an abundance root mean squared error (RMSE) of 5.62%, suggesting that the method can accurately estimate the mixture compositions, while the smooth and continuous nature of the Bézier surface may help mitigate overfitting. In comparison, the fully constrained linear unmixing (FCLSU) method achieves an RMSE of 14.2% on the same dataset, indicating the presence of nonlinear mixing effects. When only 21 mixtures are used to generate the Bézier surface, the method achieves an RMSE of 7.67% on the remaining 45 test samples. These results demonstrate that the measured powder spectra lie on a curved data manifold and suggest that nonlinear unmixing can improve mixture-composition estimation.

***Index Terms—*** spectral unmixing, nonlinear unmixing, Bézier surface, near-infrared spectroscopy, paracetamol, acetaminophen, effervescent formulation, composition estimation.

## 1. INTRODUCTION

Spectral unmixing estimates the fractional abundances of pure materials from a measured spectrum. Using the linear mixing model (LMM), a spectrum is represented as a convex combination of endmember spectra. The corresponding abundances are constrained to sum-to-one and be non-negative [1–3]. LMM assumes that every incoming ray of light interacts only once with a specific pure material in the pixel before reaching the sensor. This model is effective when the surface contains large flat areas with clearly separated regions containing different pure materials [4]. However, this model is ineffective for intimate powder mixtures, where photons interact with several particles before reaching the detector. Multiple scattering and variations in particle size, density, and the sample's surface topography can distort the spectral manifold away from the linear simplex [5, 6].

Near-infrared (NIR) spectroscopy is particularly attractive for pharmaceutical analysis because it allows for rapid, non-destructive spectral acquisition without the need for slow and lengthy sample preparation. It has been effectively used for raw-material identification, blend monitoring, content uniformity, and final product control [7, 8]. Paracetamol has, in particular, been quantified in intact tablets by reflectance and transmittance NIR spectroscopy [9, 10]. These studies generally rely on multivariate regression. Although these methods can achieve accurate predictions, they do not explicitly model the underlying physical complexity associated with intimate mixing and strong light scattering; instead, the regression absorbs that difficulty into an empirical calibration.

This work addresses the same analytical objective from an unmixing perspective. The intended application is the estimation of the components of an effervescent paracetamol formulation. Such formulations typically combine the active pharmaceutical ingredient with an organic acid and a carbonate or bicarbonate [11]. In the simplified formulation studied here, the three components are paracetamol, citric acid, and sodium bicarbonate. They are all ground to fine, visually white powders. Consequently, composition cannot be reliably inferred from visible colour alone, as the measured spectral response is governed by both chemical absorption and light-scattering effects associated with the physical structure of the powder particles.

We used a data-driven nonlinear model instead of prescribing a particular radiative-transfer or polynomial mixing law. Following the Bézier surface unmixing framework of Koirala *et al.* [6], we used measured mixtures with known mass fractions as control points of a smooth spectral data manifold. An unknown spectrum is unmixed by locating the point on that manifold that gives the smallest reconstruction error. The approach preserves the physically meaningful abundance simplex while allowing a nonlinear mapping from composition to reflectance.

The contribution of this work is threefold. First, a ternary spectral calibration set is constructed specifically from components relevant to a model effervescent paracetamol formulation. Second, the formulation composition is linked directly to a degree-10 Bézier surface, and fractional abundances are estimated by constrained nonlinear optimization. Third, the effect of model order and control-point density is examined using held-out mixtures.

## 2. DATASET AND METHODOLOGY

### 2.1. Prepared mixtures and spectral acquisition

Citric acid, sodium bicarbonate, and paracetamol were reduced to fine, visually white powders. They were weighed and mixed to generate a uniform ternary composition grid. Each sample had a total powder mass of 3.0 g. The component masses were

$$\boldsymbol{m}_{ijk} = 0.3[i, j, k]^{\mathsf{T}}\ \mathrm{g}, \qquad i+j+k = 10, \qquad i, j, k \in \mathbb{N}_0. \tag{1}$$

After division by the total mass, the ground-truth fractional abundances are

$$\boldsymbol{a}_{ijk} = [i/10, j/10, k/10]^{\mathsf{T}}. \tag{2}$$

The grid therefore contains

$$M = \binom{10+3-1}{3-1} = 66 \tag{3}$$

samples: three pure-component samples, 27 binary mixtures, and 36 ternary mixtures.

The samples were measured in the dry state with an ASD spectrometer. Each spectrum contains 2151 reflectance values spanning 350–2500 nm at 1 nm sampling.

### 2.2. Unmixing formulation

Let $\boldsymbol{y} \in \mathbb{R}^L$ denote a measured spectrum with $L$ spectral bands, and let

$$\Delta^2 = \left\{ \boldsymbol{a} \in \mathbb{R}^3 : a_r \geq 0,\ \sum_{r=1}^{3} a_r = 1 \right\} \tag{4}$$

be the ternary abundance simplex. A general spectral mixing model is written as

$$\boldsymbol{y} = F(\boldsymbol{a}) + \boldsymbol{\eta}, \tag{5}$$

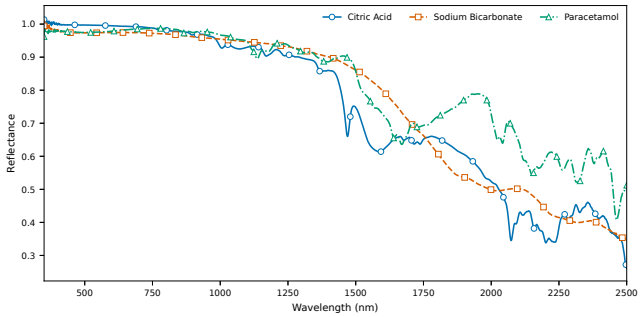



**Fig. 1**. Measured ASD reflectance spectra of the pure components used as the ternary endpoints: citric acid, sodium bicarbonate, and paracetamol.

where $F$ maps composition to reflectance and $\boldsymbol{\eta}$ contains measurement noise and model discrepancy. The linear model sets $F(\boldsymbol{a}) = \mathbf{E}\boldsymbol{a}$, where the columns of $\mathbf{E}$ are the three pure spectra. The FCLSU estimates

$$\widehat{\boldsymbol{a}}_{\mathrm{lin}} = \arg\min_{\boldsymbol{a}\in\Delta^2} \|\boldsymbol{y} - \mathbf{E}\boldsymbol{a}\|_2^2 . \tag{6}$$

This serves as the linear baseline.

### 2.3. Bézier representation of the spectral manifold

For a Bézier surface of degree $n$, let $\boldsymbol{C}_{ijk} \in \mathbb{R}^L$ be the spectral control point associated with the integer triple $(i, j, k)$, where $i + j + k = n$. The reconstructed spectrum at composition $\boldsymbol{a} = [a_1, a_2, a_3]^{\mathsf{T}}$ is

$$\widehat{\boldsymbol{y}}_n(\boldsymbol{a}) = \sum_{\substack{i,j,k\geq 0 \\ i+j+k=n}} \frac{n!}{i!j!k!} a_1^i a_2^j a_3^k\, \boldsymbol{C}_{ijk}. \tag{7}$$

The Bernstein basis in Eq. (7) is non-negative and forms a partition of unity on $\Delta^2$. The surface therefore provides a smooth map from the physically feasible composition simplex to the spectral space [6]. When $n = 1$, Eq. (7) reduces to the linear simplex. Orders $n \geq 2$ can represent curved manifolds and higher-order interactions without selecting a defined model.

For the complete calibration, $n = 10$ and the measured spectrum at each composition in Eq. (2) is assigned to its corresponding control point $\boldsymbol{C}_{ijk}$. It is important to distinguish a Bézier control point from an interpolating knot, except at the vertices, the surface does not generally pass through an individual control point. The complete set of spectra collectively determines the surface geometry.

### 2.4. Fractional-abundance estimation

The complete procedure is summarised in Fig. 2. For a target spectrum $\boldsymbol{y}$, the fractional abundances are estimated by solving

$$\widehat{\boldsymbol{a}} = \arg\min_{\boldsymbol{a}\in\Delta^2} \frac{1}{L} \|\boldsymbol{y} - \widehat{\boldsymbol{y}}_n(\boldsymbol{a})\|_2^2 . \tag{8}$$

| **Prepared composition lattice** $\longrightarrow$ | **ASD reflectance spectra** $\longrightarrow$ | **Bézier spectral surface** $\longrightarrow$ | **simplex-constrained inversion** |
|---|---|---|---|
| $\{\boldsymbol{a}_{ijk} : i+j+k=10\}$ | $\{\boldsymbol{C}_{ijk} \in \mathbb{R}^L\}$ | $\widehat{\boldsymbol{y}}_n(\boldsymbol{a})$ | $\widehat{\boldsymbol{a}} = \arg\min_{\boldsymbol{a}\in\Delta^2} \|\boldsymbol{y} - \widehat{\boldsymbol{y}}_n(\boldsymbol{a})\|_2^2$ |

**Fig. 2**. Overview of the supervised nonlinear unmixing procedure. Known mixtures provide spectral control points; the fitted composition of a target spectrum is its constrained projection onto the Bézier surface.

A discrete search over a uniform simplex grid supplies the initial composition. The solution is then refined by sequential least-squares quadratic programming with bounds $0 \le a_r \le 1$ and the equality constraint $\sum_r a_r = 1$. If the continuous optimizer fails, the best grid point is retained.

The abundance error is summarized by

$$\mathrm{RMSE}_a = \sqrt{\frac{1}{3N}\sum_{r=1}^{3}\sum_{s=1}^{N}\left(\widehat{a}_{rs} - a_{rs}\right)^2} \times 100, \qquad (9)$$

where $N$ is the number of evaluated spectra. Component-wise RMSE values are computed by restricting the sum to one component. Spectral reconstruction error is reported in reflectance units as

$$\mathrm{RMSE}_{y,s} = \sqrt{\frac{1}{L}\left\|\boldsymbol{y}_s - \widehat{\boldsymbol{y}}_s\right\|_2^2} \times 100. \qquad (10)$$

### 2.5. Experimental design

The analysis was threefold. First, the complete set of 66 spectra was used to define a degree-10 surface, and the same spectra were projected back onto that surface. This experiment measures calibration consistency.

Second, FCLSU was applied to all 66 spectra using only the three measured pure component spectra as endmembers. This gives a basis for comparison for the benchmark.

Third, interpolation was evaluated by extracting lower-degree control lattices from the degree-10 grid and evaluating only the remaining mixtures. Degrees 1, 2, and 5 were used because all of their required compositions are present exactly on the 10% lattice. They require 3, 6, and 21 control spectra and leave 63, 60, and 45 spectra, respectively, for evaluation. Other degrees were not interpolated from approximate neighbours because doing so would mix control-point selection error with model-order error.

The experiment used 400 bands from 2001 to 2400 nm. This interval was selected using an exploratory search that favoured smooth, non-crossing trajectories of constant paracetamol fraction in a two-dimensional principal-component projection.

## 3. RESULTS AND DISCUSSION

### 3.1. Composition estimation on the complete lattice

Table 1 compares the degree-10 Bézier estimate with FCLSU. The global abundance RMSE decreased from 14.21 % for the linear model to 5.62 % for the nonlinear surface, a relative reduction of 60.45%. The improvement was observed for each component. The smallest Bézier error was obtained for citric acid (3.78%), while paracetamol was the most difficult component (6.72 %). Component-wise agreement is visualized in Fig. 3.

**Table 1**. Abundance RMSE on the complete 66-sample lattice. The degree-10 Bézier values are calibration-consistency results.

| Component | FCLSU | Degree-10 Bézier |
|---|---|---|
| Citric acid | 11.37% | 3.78% |
| Sodium bicarbonate | 14.29% | 5.96% |
| Paracetamol | 16.50% | 6.72% |
| All components | 14.21% | 5.62% |

The proposed method obtained a global spectral reconstruction RMSE of 0.5% compared to 1.1% from FCLSU. The smaller abundance error of the Bézier approach, together with its lower spectral residual, indicates that the linear simplex does not follow the measured composition manifold. The comparison must nevertheless be interpreted in light of supervision. FCLSU uses three pure spectra, whereas the degree-10 surface uses the entire calibration lattice. The result demonstrates the value of measured nonlinear control points; it does not establish that a 66-control model is universally preferable to every linear or chemometric alternative.

### 3.2. Held-out interpolation and control-point density

The lower-degree experiment provides a stricter test because control spectra are not evaluated as targets. Figure 4 shows how the degree-1, degree-2, and degree-5 control sets are nested within the measured 10% composition lattice; all remaining locations are held out at the corresponding degree. Table 2 shows that the abundance error decreased systematically as the surface order and the number of control points increased. A degree-1 surface, equivalent to the linear simplex, gave an abundance RMSE of 14.21% on 63 held-out mixtures. The degree-2 surface reduced this error to 11.41%. The degree-5 surface, defined by 21 controls and evaluated on 45 held-out mixtures, obtained an RMSE of 7.67%. This is a 46.0% reduction relative to the degree-1 result.

The mean fitted spectral RMSE values were 1.106%, 0.920%, and 0.513% for degrees 1, 2, and 5, respectively. They are smaller than the reconstruction errors evaluated

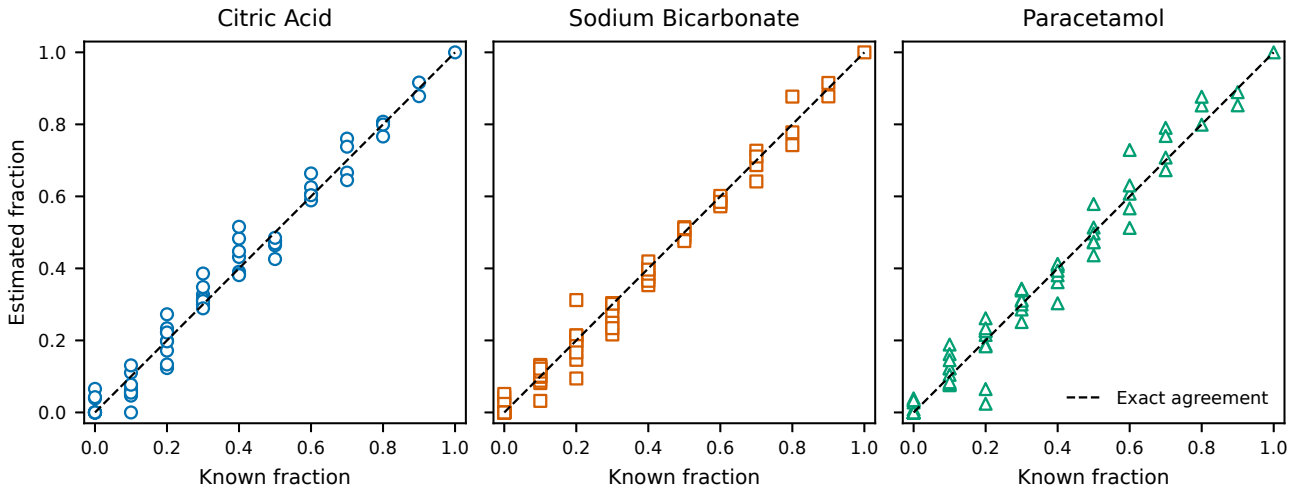



**Fig. 3**. Known versus estimated component fractions for projection onto the complete degree-10 Bézier surface. The diagonal denotes exact agreement.

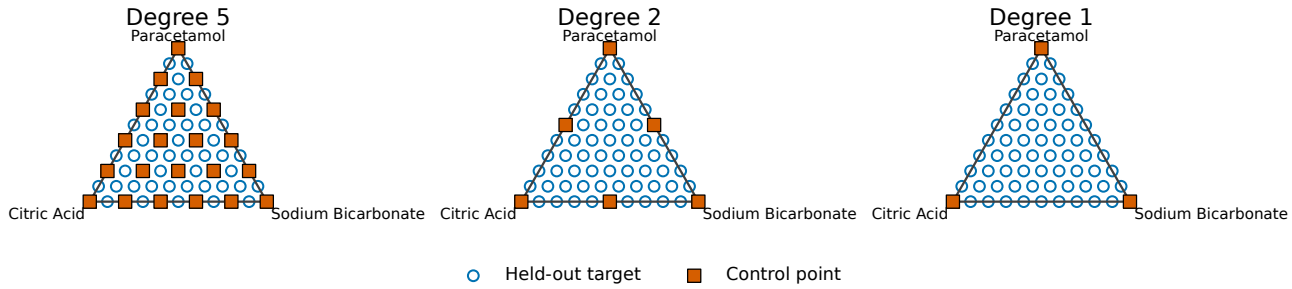



**Fig. 4**. Control-point and held-out-sample layouts for the lower-degree interpolation experiment. Red squares identify spectra used as Bézier control points, and blue circles identify targets excluded from the corresponding surface construction.

at the true compositions because Eq. (8) is free to move the abundance estimate to compensate for local model discrepancy. This distinction is important. A low spectral residual does not, by itself, guarantee an accurate composition when two regions of the manifold are spectrally similar or when the forward model is biased.

The error trend supports two conclusions. First, the mapping from the ternary mass fractions to reflectance is measurably nonlinear. Second, control-point density matters. A higher-order surface can follow more complex curvature, but the number of required controls grows as $\binom{n+p-1}{p-1}$. For three components, the degree-5 model requires 21 measurements; a degree-10 model requires 66. The selection of surface order is therefore a calibration-design decision, not just a numerical hyperparameter. This is consistent with previous Bézier unmixing studies of intimate mineral powders [6, 12].

## 4. CONCLUSION

In this work, we proposed a supervised nonlinear unmixing approach to model effervescent paracetamol formulations. The basic idea is to model the spectral data manifold using a Bézier surface. The proposed methodology was validated on a dataset generated under controlled laboratory conditions. In particular, 66 samples were produced by homogeneously mixing citric acid, sodium bicarbonate, and paracetamol. The modeling error was approximately 5.6% when the spectral reflectances of all 66 samples were used. For these mixtures, the FCLSU produced an error of approximately 14.2%, indicating a nonlinear relationship between the measured spectra and the fractional abundances. When only 21 mixtures were used to model the data manifold, the abundance error for the remaining 45 test samples was approximately 7.7%. This represents an approximately 46% reduction relative to the FCLSU. These results demonstrate the potential of the proposed approach for the pharmaceutical analysis of effervescent paracetamol formulations in a rapid and non-destructive manner.

**Table 2**. Held-out Bézier interpolation results for the selected 2001-2400 nm interval. Spectral RMSE at the true abundance evaluates forward model fidelity; fitted spectral RMSE evaluates the best spectral projection.

| Degree | Controls | Held out | Global abundance RMSE (in %) | Citric acid RMSE (in %) | Sodium bicarbonate RMSE (in %) | Paracetamol RMSE (in %) | Mean spectral RMSE at true abundance (in %) | Mean spectral RMSE at fitted abundance (in %) |
|---|---|---|---|---|---|---|---|---|
| 1 | 3 | 63 | 14.21 | 11.37 | 14.29 | 16.50 | 3.602 | 1.105 |
| 2 | 6 | 60 | 11.41 | 9.28 | 11.58 | 13.05 | 2.911 | 0.920 |
| 5 | 21 | 45 | 7.67 | 5.59 | 8.70 | 8.34 | 2.121 | 0.513 |


# Acknowledgment

The research is supported by the Research Foundation–Flanders (project G031921N). Zakaria Bnoulkacem is a doctoral fellow of the Research Foundation–Flanders, Belgium (FWO grant 1SH1M24N). Bikram Koirala is a postdoctoral fellow of the Research Foundation–Flanders, Belgium (FWO grant 1250824N-7028).


## 5. REFERENCES


[1] Nirmal Keshava and John F. Mustard, "Spectral unmixing," *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 44–57, Jan. 2002, https://doi.org/10.1109/79.974727.

[2] Daniel C. Heinz and Chein-I Chang, "Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 3, pp. 529–545, Mar. 2001, https://doi.org/10.1109/36.911111.

[3] José M. Bioucas-Dias, Antonio Plaza, Nicolas Dobigeon, Mario Parente, Qian Du, Paul Gader, and Jocelyn Chanussot, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 2, pp. 354–379, Apr. 2012, https://doi.org/10.1109/JSTARS.2012.2194696.

[4] Bikram Koirala, Rob Heylen, and Paul Scheunders, "A neural network method for nonlinear hyperspectral unmixing," in *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, 2018, pp. 4233–4236.

[5] Nicolas Dobigeon, Jean-Yves Tourneret, Cédric Richard, José Carlos M. Bermudez, Stephen McLaughlin, and Alfred O. Hero, "Nonlinear unmixing of hyperspectral images: Models and algorithms," *IEEE Signal Processing Magazine*, vol. 31, no. 1, pp. 82–94, Jan. 2014, https://doi.org/10.1109/MSP.2013.2279274.

[6] Bikram Koirala, Behnood Rasti, Zakaria Bnoulkacem, and Paul Scheunders, "Nonlinear spectral unmixing using Bézier surfaces," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–16, 2024, https://doi.org/10.1109/TGRS.2024.3422495.

[7] Yves Roggo, Pascal Chalus, Lene Maurer, Carmen Lema-Martinez, Agnes Edmond, and Nelly Jent, "A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies," *Journal of Pharmaceutical and Biomedical Analysis*, vol. 44, no. 3, pp. 683–700, July 2007, https://doi.org/10.1016/j.jpba.2007.03.023.

[8] Marcel Blanco, Raquel Cueva-Mestanza, and Amparo Peguero, "Controlling individual steps in the production process of paracetamol tablets by use of NIR spectroscopy," *Journal of Pharmaceutical and Biomedical Analysis*, vol. 51, no. 4, pp. 797–804, Mar. 2010, https://doi.org/10.1016/j.jpba.2009.09.038.

[9] Andrew D. Trafford, Roger D. Jee, Anthony C. Moffat, and Paul Graham, "A rapid quantitative assay of intact paracetamol tablets by reflectance near-infrared spectroscopy," *Analyst*, vol. 124, no. 2, pp. 163–167, 1999, https://doi.org/10.1039/A806629I.

[10] A. Eustaquio, M. Blanco, R. D. Jee, and A. C. Moffat, "Determination of paracetamol in intact tablets by use of near infrared transmittance spectroscopy," *Analytica Chimica Acta*, vol. 383, no. 3, pp. 283–290, Mar. 1999, https://doi.org/10.1016/S0003-2670(98)00815-0.

[11] Claude Dubray, Philippe Maincent, and Jean-Yves Milon, "From the pharmaceutical to the clinical: The case for effervescent paracetamol in pain management. a narrative review," *Current Medical Research and Opinion*, vol. 37, no. 6, pp. 1039–1048, 2021, https://doi.org/10.1080/03007995.2021.1902297.

[12] Bikram Koirala, Behnood Rasti, Zakaria Bnoulkacem, Andrea de Lima Ribeiro, Yuleika Madriz, Erik Herrmann, Arthur Gestels, Thomas De Kerf, Sandra Lorenz, Margret C. Fuchs, Koen Janssens, Gunther Steenackers, Richard Gloaguen, and Paul Scheunders, "A multisensor hyperspectral benchmark dataset for unmixing of intimate mixtures," *IEEE Sensors Journal*, vol. 24, no. 4, pp. 4694–4710, Feb. 2024, https://doi.org/10.1109/JSEN.2023.3343552.