



HAL
open science

Hyperspectral super-resolution accounting for spectral variability: coupled tensor LL1-based recovery and blind unmixing of the unknown super-resolution image

Clémence Prévost, Ricardo A Borsoi, Konstantin Usevich, David Brie, José C. M. Bermudez, Cédric Richard

► To cite this version:

Clémence Prévost, Ricardo A Borsoi, Konstantin Usevich, David Brie, José C. M. Bermudez, et al.. Hyperspectral super-resolution accounting for spectral variability: coupled tensor LL1-based recovery and blind unmixing of the unknown super-resolution image. *SIAM Journal on Imaging Sciences*, 2022, 15 (1), pp.110-138. 10.1137/21M1409354 . hal-03158076v2

HAL Id: hal-03158076

<https://hal.science/hal-03158076v2>

Submitted on 2 Nov 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Hyperspectral super-resolution accounting for spectral variability: coupled tensor LL1-based recovery and blind unmixing of the unknown super-resolution image*

Clémence Prévost[†], Ricardo A. Borsoi^{‡,§}, Konstantin Usevich[†], David Brie[†], José C. M. Bermudez[‡], and Cédric Richard[§]

Abstract. In this paper, we propose to jointly solve the hyperspectral super-resolution problem and the unmixing problem of the underlying super-resolution image using a coupled LL1 block-tensor decomposition. We consider a spectral variability phenomenon occurring between the observed low-resolution images. Exact recovery conditions for the image and mixing factors are provided. We propose two algorithms: an unconstrained one and another one subject to non-negativity constraints, to solve the problems at hand. We showcase performance of the proposed approach on synthetic and real images.

Key words. Hyperspectral super-resolution, spectral variability, hyperspectral unmixing, image fusion, tensor decompositions.

AMS subject classifications. 68U10, 62H35, 94A08

1. Introduction.

1.1. Background. Hyperspectral devices are able to sample the electromagnetic spectrum into hundred of wavelengths, allowing for the acquisition of hyperspectral images (HSIs) that possess high spectral resolution. However, the tradeoff between spatial and spectral resolution forces the HSIs to have a small number of relatively large pixels [41]. On the other hand, multispectral sensors produce multispectral images (MSIs) with high spatial resolution (smaller pixels), at the cost of a restricted number of spectral bands. The composition of each pixel in HSIs and MSIs can be approximated by a sum of a small number of spectral signatures, or endmembers. This representation is known as the linear mixing model, and allows for identification of the materials and their abundances in a scene, a process termed unmixing. Many unmixing approaches have been proposed (see [3, 37, 36] and references therein).

The hyperspectral super-resolution (HSR) problem [50] was formulated to circumvent the physical limitations of each device. This problem aims at recovering a super-resolution image (SRI) that possesses both high spatial and high spectral resolutions from co-registered HSI and MSI of the same scene. The high spatial and spectral resolutions of the SRI can then be exploited in unmixing tasks.

[†]Centre de Recherche en Automatique de Nancy (CRAN), Université de Lorraine, CNRS, Vandoeuvre-lès-Nancy, France.

[‡]Department of Electrical Engineering, Federal University of Santa Catarina (DEE-UFSC), Florianópolis, SC, Brazil.

[§]Université Côte d'Azur, Nice, France, Lagrange Laboratory (CNRS, OCA). e-mail: firstname.lastname@unice.fr.

*Submitted to the editors October 25, 2021.

Funding: This work was partly supported by the ANR (Agence Nationale de Recherche) under grant LeaFleT (ANR-19-CE23-0021), by the National Council for Scientific and Technological Development (CNPq) under grants 304250/2017-1, 409044/2018-0, 141271/2017-5 and 204991/2018-8. The work of C. Richard was supported through the UCA JEDI Investments in the Future project with the reference number ANR-15-IDEX-0001.

We also thank the GdR CNRS ISIS for supporting the collaboration between R. A. Borsoi and C. Prévost, K. Usevich and D. Brie through a *bourse de mobilité*.

32 Since there exist few satellites that carry both hyperspectral and multispectral sensors
33 [15, 29], combining an HSI and an MSI acquired onboard of different missions has become
34 a task of prime interest [22, 17]. Since the HSI and the MSI are acquired at different time
35 instants, their acquisition conditions can differ by, e.g., illumination, atmospheric or seasonal
36 changes [5]. This can cause variations in the underlying endmembers and impact negatively
37 the HSR and unmixing algorithms. This phenomenon is known as spectral variability.

38 In this paper, we propose to formulate the HSR problem as a coupled tensor block-term
39 decomposition (BTD) of the HSI and MSI, accounting for spectral variability between the
40 endmembers. Inspired by the works of [54, 53, 13], we propose guarantees for noiseless exact
41 recovery of the SRI and its latent factors based on the linear mixing model. One advantage
42 of using the linear mixing model is that the factors of a properly chosen decomposition can
43 be seen as the high-resolution spectral signatures and abundance maps corresponding to the
44 materials in the underlying SRI, provided that they are entry-wise non-negative. Differently
45 from matrix-based models, our noiseless recovery conditions do not require additional con-
46 straints on the low-rank factors. We also propose a unified procedure that aims at estimating
47 the latent mixing factors and recovering the SRI in one single step. Our experiments illus-
48 trate the competitive performance of the proposed approach for HSR. The performance for
49 unmixing of the unknown SRI is compared to that of traditional unmixing algorithms applied
50 on an estimated SRI.

51 Many approaches have been proposed to solve the HSR problem. Most matrix approaches
52 [51, 49, 42, 48] are based on the linear mixing model and perform a coupled low-rank fac-
53 torization of the matricized HSI and MSI. Some matrix approaches provide exact recovery
54 conditions for the SRI in noiseless cases. However, these recovery guarantees usually require
55 additional priors on the model. See for example [33], which promotes sparsity of the factors. In
56 the absence of such hypotheses, only a bound on the recovery error can be obtained [34]. Some
57 matrix approaches are suitable for the HSR and unmixing problem as well, which consists of
58 recovering the underlying SRI by means of a physically-informed low-rank approximation. See
59 for instance [33] and [51]. However, identifiability of the mixing model could only be obtained
60 under additional constraints on the low-rank factors [14, 32]. As a result, to the best of our
61 understanding, recovery conditions for the joint HSR and unmixing problem in the literature
62 only consider specific classes of problems.

63 More recently, tensor-based approaches were proposed for the HSR problem. The works
64 of [27, 28] formulate the HSR problem as a coupled canonical polyadic (CP) decomposition,
65 while a coupled multilinear Tucker decomposition is used in [38]. However, the factors of these
66 decompositions lack physical interpretation, and thus the aforementioned methods cannot be
67 used for unmixing. Motivated by the usefulness of tensor models, approaches based on block-
68 tensor decomposition [53, 13] were proposed for solving the HSR problem. This decomposition
69 was also successfully used to perform unmixing [39] on the SRI directly.

70 Most of these approaches however share a common limitation: they assume that the
71 acquisition conditions of the HSI and MSI are the same, hence they ignore the variability
72 phenomenon. In [4], a super-resolution method accounting for seasonal spectral variability
73 was proposed. Using a low-rank matrix formulation, the spectral signatures underlying the
74 HSI and MSI were allowed to be different from each other, with variations introduced by a
75 set of multiplicative scaling factors [25]. This algorithm led to significant performance im-
76 provements when the HSI and MSI are subject to spatially uniform seasonal or acquisition
77 variations. However, the algorithm in [4] presented high computation times and did not offer
78 any theoretical guarantees. In [6], two tensor-based algorithms based on the Tucker decom-

79 position were proposed, accounting for spatially and spectrally-localized changes between the
 80 HSI and MSI. Noiseless unique recovery guarantees were proposed. Unfortunately, the Tucker
 81 decomposition being generally non-unique, the decomposition factors were not physically in-
 82 terpretable.

83 The paper is organized as follows. The remainder of Section 1 contains tensor algebra
 84 preliminaries. Section 2 introduces the low-rank tensor model, and the model for spectral vari-
 85 ability. Section 3 addresses recovery analysis for the joint HSR and unmixing task. Section 4
 86 describes the proposed algorithms and their computational complexity. Finally, Sections 5
 87 and 6 contain numerical experiments for HSR and coupled unmixing, respectively.

88 **1.2. Definitions and notations.** We follow the notations of [31, 10]. We use lower (a) or
 89 uppercase (A) plain font for scalars, boldface lowercase (\mathbf{a}) for vectors, boldface uppercase
 90 (\mathbf{A}) for matrices and calligraphic (\mathcal{A}) for tensors. The elements of vectors, matrices and
 91 tensors are denoted as a_i , $A_{i,j}$ and $\mathcal{A}_{i_1, \dots, i_N}$, respectively. The transpose of a matrix \mathbf{A} is
 92 denoted by \mathbf{A}^\top . We use \mathbf{I}_N for the $N \times N$ identity matrix and $\mathbf{0}_{L \times K}$ for the $L \times K$ matrix
 93 of zeros. Notation $\mathbf{1}_L$ denotes an all-ones vector of size $L \times 1$. For a matrix \mathbf{X} , the notation
 94 $\mathbf{X} \geq \mathbf{0}$ means that \mathbf{X} is entry-wise non-negative. Symbols \boxtimes and \odot denote the Kronecker
 95 and Khatri-Rao products, respectively. The Hadamard (element-wise) product is denoted by
 96 \square . We use $\text{vec}\{\cdot\}$ for the standard column-major vectorization of a matrix or a tensor. For
 97 two matrices \mathbf{A} and \mathbf{B} , the operator $\text{Diag}\{\mathbf{A}, \mathbf{B}\}$ produces a block-diagonal matrix whose
 98 diagonal blocks are \mathbf{A} and \mathbf{B} . Each dimension of a tensor is called a mode, and the number of
 99 dimensions is called order. A mode- p fiber of tensor \mathcal{X} is a vector obtained from \mathcal{X} by fixing
 100 all but the p -th dimension. A slab or slice of a tensor \mathcal{X} is a matrix whose columns are the
 101 vectors of \mathcal{X} obtained by fixing all but two of its modes. We restrict the scope of this paper
 102 to order-3 tensors.

103 **Definition 1.1. Outer product** – The outer product between three vectors $\mathbf{a} \in \mathbb{R}^I$, $\mathbf{b} \in \mathbb{R}^J$,
 104 $\mathbf{c} \in \mathbb{R}^K$ is an order-3 tensor of rank 1 defined as $\mathcal{X} = \mathbf{a} \otimes \mathbf{b} \otimes \mathbf{c} \in \mathbb{R}^{I \times J \times K}$. Each element of
 105 \mathcal{X} is accessed as $\mathcal{X}_{i,j,k} = a_i b_j c_k$.

106 **Definition 1.2. Tensor unfoldings** – The mode- p unfolding of a tensor \mathcal{X} , denoted by
 107 $\mathbf{X}^{(p)}$, is the matrix whose rows are the mode- p fibers of \mathcal{X} , ordered according to the vector-
 108 ization order. For a third-order tensor $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$, we have $\mathbf{X}^{(1)} \in \mathbb{R}^{JK \times I}$, $\mathbf{X}^{(2)} \in \mathbb{R}^{IK \times J}$
 109 and $\mathbf{X}^{(3)} \in \mathbb{R}^{IJ \times K}$.

110 **Definition 1.3. Mode product** – The mode- p product between a tensor \mathcal{X} and a matrix
 111 \mathbf{M} is denoted by $\mathcal{X} \bullet_p \mathbf{M}$ and is evaluated such that each mode- p fiber of \mathcal{X} is multiplied by
 112 \mathbf{M} . For instance, the elements of the mode-1 product between $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$ and $\mathbf{M} \in \mathbb{R}^{L \times I}$
 113 are determined as $(\mathcal{X} \bullet_1 \mathbf{M})_{\ell,j,k} = \sum_i \mathcal{X}_{i,j,k} \mathbf{M}_{\ell,i}$, $\ell \in \{1, \dots, L\}$. Moreover, it holds that
 114 $\mathcal{Y} = \mathcal{X} \bullet_k \mathbf{M} \Leftrightarrow \mathbf{Y}^{(k)} = \mathbf{X}^{(k)} \mathbf{M}^\top$.

115 **1.3. Block-term decomposition with ranks $(L, L, 1)$.** In this subsection, we introduce the
 116 block-term decomposition with ranks $(L, L, 1)$, that we will use to build our model. The main
 117 advantage of this decomposition is to link the low-rank factors to high-resolution abundance
 118 matrices and spectral signatures used in unmixing of the unknown SRI, under additional non-
 119 negativity priors on the low-rank factors. We also recall some sufficient uniqueness conditions
 120 for this decomposition, as well as useful properties.

121 **Definition 1.4. Block-term decomposition** – An order-3 tensor $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$ generally
 122 admits a block-term decomposition (BTB) with ranks $(L, L, 1)$ (LL1-BTD) as

$$123 \quad (1.1) \quad \mathcal{X} = \sum_{r=1}^R (\mathbf{A}_r \mathbf{B}_r^\top) \otimes \mathbf{c}_r,$$

125 where $\mathbf{A}_r \in \mathbb{R}^{I \times L}$, $\mathbf{B}_r \in \mathbb{R}^{J \times L}$, and $\mathbf{c}_r \in \mathbb{R}^K$, for $r \in \{1, \dots, R\}$. Moreover, we denote
 126 $\mathbf{A} = [\mathbf{A}_1, \dots, \mathbf{A}_R] \in \mathbb{R}^{I \times LR}$, $\mathbf{B} = [\mathbf{B}_1, \dots, \mathbf{B}_R] \in \mathbb{R}^{J \times LR}$ and $\mathbf{C} = [\mathbf{c}_1, \dots, \mathbf{c}_R] \in \mathbb{R}^{K \times R}$.

127 **Theorem 1.5.** [12, Theorem 4.7] Let $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ denote an LL1-BTD of a tensor \mathcal{X} for
 128 $r \in \{1, \dots, R\}$ as in (1.1). Assume that $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ are drawn from certain joint absolutely
 129 continuous distributions. If $IJ \geq L^2 R$ and

$$130 \quad \min\left(\left\lfloor \frac{I}{L} \right\rfloor, R\right) + \min\left(\left\lfloor \frac{J}{L} \right\rfloor, R\right) + \min(K, R) \geq 2R + 2,$$

132 then $\mathbf{A}_r \mathbf{B}_r^\top$ and \mathbf{c}_r are essentially unique almost surely for $r \in \{1, \dots, R\}$.

133 **Definition 1.6. Partition-wise Khatri-Rao product** – The partition-wise Khatri-Rao
 134 product between two partitioned matrices \mathbf{A} and \mathbf{C} defined as above can be expressed as

$$135 \quad \mathbf{C} \odot_p \mathbf{A} = [\mathbf{c}_1 \boxtimes \mathbf{A}_1, \dots, \mathbf{c}_R \boxtimes \mathbf{A}_R] \in \mathbb{R}^{IK \times LR}.$$

137 **Property 1.7. Tensor unfoldings and LL1** – Using the above notations, the unfoldings
 138 of a tensor \mathcal{X} admitting an LL1-BTD as above can be expressed as

$$139 \quad \mathbf{X}^{(1)} = (\mathbf{C} \odot_p \mathbf{B}) \mathbf{A}^\top,$$

$$140 \quad \mathbf{X}^{(2)} = (\mathbf{C} \odot_p \mathbf{A}) \mathbf{B}^\top,$$

$$141 \quad \mathbf{X}^{(3)} = [(\mathbf{A}_1 \odot \mathbf{B}_1) \mathbf{1}_L, \dots, (\mathbf{A}_R \odot \mathbf{B}_R) \mathbf{1}_L] \mathbf{C}^\top.$$

143 2. Proposed model.

144 **2.1. Degradation model and indeterminacies.** We consider an HSI data cube $\mathcal{Y}_H \in$
 145 $\mathbb{R}^{I_H \times J_H \times K}$ and an MSI data cube $\mathcal{Y}_M \in \mathbb{R}^{I \times J \times K_M}$. The scalars K and K_M denote the
 146 spectral dimensions, and (I, J) (resp. (I_H, J_H)) stand for the spatial dimensions. We suppose
 147 that the spatial resolution of the MSI is higher than that of the HSI (i.e., $I_H < I$ and $J_H < J$),
 148 while its spectral resolution is lower ($K_M < K$). Most previous works [27]–[13] considered the
 149 low resolution images as degraded versions of a single SRI $\mathcal{Z} \in \mathbb{R}^{I \times J \times K}$, that possesses high
 150 spatial and spectral resolutions. This model can be expressed as:

$$151 \quad (2.1) \quad \begin{cases} \mathcal{Y}_H &= \mathcal{Z} \bullet_1 \mathbf{P}_1 \bullet_2 \mathbf{P}_2 + \mathcal{E}_H, \\ \mathcal{Y}_M &= \mathcal{Z} \bullet_3 \mathbf{P}_3 + \mathcal{E}_M, \end{cases}$$

153 where the tensors \mathcal{E}_H and \mathcal{E}_M are additive noise terms. The matrix $\mathbf{P}_3 \in \mathbb{R}^{K_M \times K}$ contains
 154 the spectral response functions for each band of the MSI sensor. The spatial degradation
 155 matrices $\mathbf{P}_1 \in \mathbb{R}^{I_H \times I}$ and $\mathbf{P}_2 \in \mathbb{R}^{J_H \times J}$ perform Gaussian blurring and downsampling along
 156 each spatial dimension, i.e. we suppose that the spatial degradation operation is separable,
 157 as in the commonly used Wald's protocol [47].

158 However, this model implies that the acquisition conditions of \mathcal{Y}_H and \mathcal{Y}_M are the same,
 159 thus it ignores possible variations in atmospheric, seasonal or illumination conditions [44, 52]
 160 between the HSI and MSI. The variability phenomenon motivates the need for more flexible
 161 models. As a result, in this paper, we adopt a more general approach. As in [6], we consider
 162 two different SRIs $\mathcal{Z} \in \mathbb{R}^{I \times J \times K}$ and $\tilde{\mathcal{Z}} \in \mathbb{R}^{I \times J \times K}$, underlying the HSI and MSI, respectively.
 163 The SRIs \mathcal{Z} and $\tilde{\mathcal{Z}}$ contain possibly different spectral signatures and can be linked as

$$164 \quad (2.2) \quad \tilde{\mathcal{Z}} = \mathcal{Z} + \Psi,$$

166 where $\Psi \in \mathbb{R}^{I \times J \times K}$ is a tensor of spectral variability. This leads to the following model:

$$167 \quad (2.3) \quad \begin{cases} \mathcal{Y}_H &= \mathcal{Z} \bullet_1 \mathbf{P}_1 \bullet_2 \mathbf{P}_2 + \mathcal{E}_H, \\ \mathcal{Y}_M &= \tilde{\mathcal{Z}} \bullet_3 \mathbf{P}_3 + \mathcal{E}_M. \end{cases}$$

169 In this framework, the HSR problem consists recovering $\mathcal{Z} \in \mathbb{R}^{I \times J \times K}$ and $\Psi \in \mathbb{R}^{I \times J \times K}$
 170 under the assumption of the observation model (2.2)–(2.3). However, the presence of the
 171 variability tensor Ψ makes this problem ambiguous [6], as one cannot easily separate \mathcal{Z} and
 172 Ψ from $\tilde{\mathcal{Z}}$. We recall the following theorem [6]:

173 **Theorem 2.1.** [6, Theorem 1.a)] *Suppose that the HSI and MSI are generated according to*
 174 *(2.3) and that the observation noise is zero (i.e. $\mathcal{E}_H, \mathcal{E}_M = \mathbf{0}$). If either $\mathbf{P}_1, \mathbf{P}_2$ or \mathbf{P}_3 have*
 175 *non-trivial nullspace, then (\mathcal{Z}, Ψ) cannot be uniquely recovered from \mathcal{Y}_H and \mathcal{Y}_M .*

176 In [6], approaches based on model (2.3) were proposed, using a coupled Tucker approxima-
 177 tion. However, due to the non-uniqueness of the Tucker decomposition, the latent multilinear
 178 factors were not guaranteed to be unique, and no non-negativity constraints were enforced.
 179 Thus it was not possible to incorporate them into an interpretable mixing model. In what
 180 follows, we introduce a low-rank tensor model to circumvent the fundamental ambiguities of
 181 the proposed degradation model. Indeed, a wisely chosen low-rank decomposition might still
 182 allow for unique recovery of portions of the tensors. We propose to use the LL1-BTD, whose
 183 factors are suitable for physical interpretation. The LL1-BTD model was successfully used for
 184 unmixing [39] and HSR [53, 13]. However, these works ignored any variability phenomenon.

185 Differently from [6] (that considered spatially and spectrally localized changes), we con-
 186 sider that variability only impacts the spectral dimension of the SRI. This is reasonable, since
 187 spectral variability can occur even with short acquisition time differences. This assumption
 188 also allows for sometimes less restrictive noiseless recoverability guarantees than the ones from
 189 [6]*. As a result, the proposed model is more suitable for scenarios with low spatial variability.
 190 Nevertheless, we will show in Sections 5 and 6 that it is able to address large acquisition time
 191 differences as well.

192 **2.2. LL1-BTD mixing model for the underlying SRIs.** In the linear mixing model, each
 193 pixel of the SRI \mathcal{Z} (and therefore, of the HSI \mathcal{Y}_H) can be represented as a sum of a small
 194 number R of pure spectral signatures [30]. This property can be incorporated in a physically-
 195 informed low-rank approximation model, allowing to perform both HSR and unmixing in a
 196 unified procedure. Thus, as in [53], we can model the third-mode unfolding of \mathcal{Z} as:

$$197 \quad (2.4) \quad \mathbf{Z}^{(3)} = \mathbf{S}\mathbf{C}^T \in \mathbb{R}^{IJ \times K},$$

*because of the more general variability model.

199 where $\mathbf{C} = [\mathbf{c}_1, \dots, \mathbf{c}_R] \in \mathbb{R}^{K \times R}$ is a matrix containing the spectral signatures of the R
 200 endmembers underlying the SRI. The matrix $\mathbf{S} = [\text{vec}\{\mathbf{S}_1\}, \dots, \text{vec}\{\mathbf{S}_R\}] \in \mathbb{R}^{IJ \times R}$ contains
 201 the vectorized abundance maps of each material. In fact, (2.4) can be viewed as the linear
 202 mixing model for the SRI \mathbf{Z} , under the assumption that \mathbf{C} and \mathbf{S} are entry-wise non-negative[†].

203 We assume as in [53] that the abundance maps admit low rank L . The hypothesis of low-
 204 rank abundance matrices is reasonable, since the two spatial dimensions are often correlated
 205 along the rows and columns, respectively. Thus we have

$$206 \quad (2.5) \quad \mathbf{S}_r \approx \mathbf{A}_r \mathbf{B}_r^\top \in \mathbb{R}^{I \times J},$$

208 where $\mathbf{A}_r \in \mathbb{R}^{I \times L}$ and $\mathbf{B}_r \in \mathbb{R}^{J \times L}$ admit rank L . In [7], an upper bound on the reconstruction
 209 error of such matrices by (2.5) is provided in the general problem. In particular, this error
 210 can be as small as desired if L is large enough, which motivates the above assumption.

211 Reshaping (2.4) into tensor format yields the following:

$$212 \quad (2.6) \quad \mathbf{Z} = \sum_{r=1}^R (\mathbf{A}_r \mathbf{B}_r^\top) \otimes \mathbf{c}_r.$$

214 The above model can be seen as an LL1-BTD of the tensor \mathbf{Z} with factors $\mathbf{A} = [\mathbf{A}_1, \dots, \mathbf{A}_R]$,
 215 $\mathbf{B} = [\mathbf{B}_1, \dots, \mathbf{B}_R]$ and $\mathbf{C} = [\mathbf{c}_1, \dots, \mathbf{c}_R]$ related to the mixing factors.

216 State-of-the-art unmixing algorithms aim at recovering $\{\mathbf{S}_r = \mathbf{A}_r \mathbf{B}_r^\top\}_{r=1}^R$ and \mathbf{C} from the
 217 mixed pixels in \mathbf{Z} . Here, since \mathbf{Z} is unknown and only \mathbf{Y}_H is observed with high spectral reso-
 218 lution, these algorithms are only able to recover spatially-degraded versions of the abundance
 219 maps [13], namely

$$220 \quad (2.7) \quad \mathbf{P}_1 \mathbf{S}_r \mathbf{P}_2^\top \in \mathbb{R}^{I_H \times J_H} \text{ for } r \in \{1, \dots, R\}.$$

222 Differently from those works, fusion of an HSI with an MSI with high spatial resolution allows
 223 us to seek for abundance maps at a higher spatial resolution.

224 In Figure 1, our model and the joint unmixing-and-HSR strategy are summarized. Since
 225 the true SRIs \mathbf{Z} and $\tilde{\mathbf{Z}}$ are unknown, we utilize the fusion framework to decompose the HSI
 226 and MSI into interpretable mixing factors. While high-resolution spectra can be obtained
 227 from the HSI, high-resolution abundance maps can only be obtained from the MSI. The
 228 fusion framework allows to exploit fully the information contained in the observations. Once
 229 the mixing factors have been retrieved, the estimated SRI $\hat{\mathbf{Z}}$ can be approximated using the
 230 LL1-BTD[‡]. However, the spectral variability has to be modeled first.

231 **2.3. Modeling spectral variability.** In traditional unmixing applications, which only deal
 232 with a single SRI, spectral variability defines the fact that the spectrum of a material (e.g.,
 233 grass or soil) changes from pixel to pixel. This sort of spectral variability is widely considered
 234 in the literature; see e.g., [44] and references therein. However, two different images are

[†]In some traditional unmixing methods (see e.g. [36]), the sum-to-one constrained abundance matrices is also enforced. However, spatial illumination changes frequently introduce scaling variations in each pixel. Moreover, non-negativity constraints can be transformed equivalently to generalized sum-to-one constraint, as specified in [26]. As a result, we do not consider this additional constraint in this work.

[‡]It should be mentioned that this framework is different from a two-step procedure which would i) recover the SRI $\tilde{\mathbf{Z}}$ from data fusion, then ii) run a traditional unmixing algorithm with the estimated SRI.

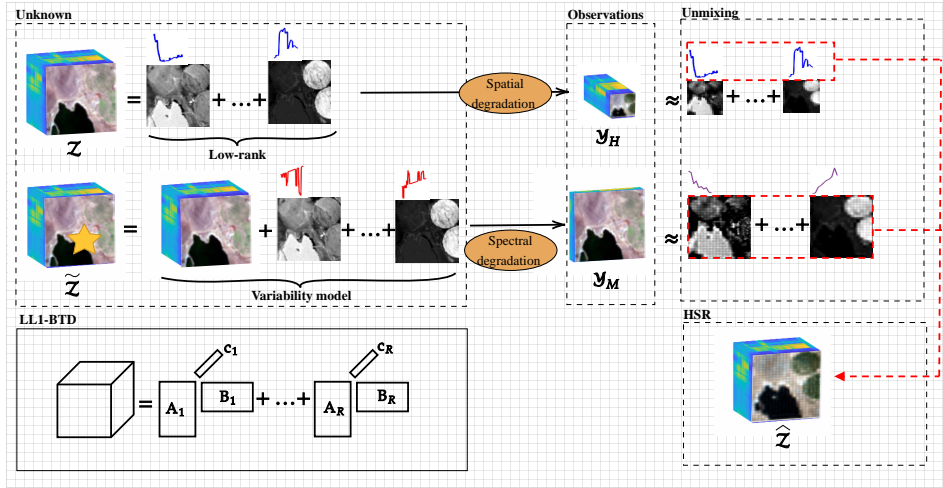


Figure 1. Summary of the model and resolution strategy.

235 considered in our framework. The variability between their corresponding spectra can be
 236 more significant, since their acquisition conditions may be very different. In this paper, we
 237 adopt a simple model for spectral variability, which characterizes which wavelengths are more
 238 impacted by different acquisition conditions for each material in the images. Although not
 239 describing pixel-by-pixel spectral variability, this model is appropriate to describe variations
 240 due to different acquisition conditions and will allow us to obtain recovery guarantees.

241 In [25], the generalized linear mixing model was proposed to model the spectra underlying
 242 the MSI as

$$243 \quad (2.8) \quad \tilde{C} = \psi_{\text{multi}} \square C,$$

245 where $\psi_{\text{multi}} \in \mathbb{R}^{K \times R}$ is a matrix of positive scaling factors. In this paper, motivated by model
 246 (2.1), we propose to use an equivalent additive model:

$$247 \quad (2.9) \quad \tilde{C} = \psi + C,$$

249 where $\psi \in \mathbb{R}^{K \times R}$ is different from ψ_{multi} . The choice of the additive variability model (2.9)
 250 allows us to keep the spectral variability explicit in ψ . Moreover, both models (2.8) and (2.9)
 251 are able to represent arbitrary endmember variations.

252 Since we allow only spectral variability to be present, the variability tensor Ψ also admits
 253 an LL1-BTD with the same factors A and B as for the SRI Z , but with spectral factor
 254 $\psi \in \mathbb{R}^{K \times R}$ representing the spectral variations. This allows to write \tilde{Z} as

$$255 \quad (2.10) \quad \tilde{Z} = \underbrace{\sum_{r=1}^R (A_r B_r^T) \otimes c_r}_Z + \underbrace{\sum_{r=1}^R (A_r B_r^T) \otimes \psi_r}_\Psi = \sum_{r=1}^R (A_r B_r^T) \otimes \tilde{c}_r,$$

256
 257 where $\tilde{c}_r = \psi_r + c_r$ is the r -th column of \tilde{C} .

258 From the above formulations, we can finally express (2.3) as a coupled LL1-BTD:

$$259 \quad (2.11) \quad \begin{cases} \mathcal{Y}_H &= \sum_{r=1}^R (\mathbf{P}_1 \mathbf{A}_r (\mathbf{P}_2 \mathbf{B}_r)^\top) \otimes \mathbf{c}_r + \boldsymbol{\mathcal{E}}_H, \\ \mathcal{Y}_M &= \sum_{r=1}^R (\mathbf{A}_r \mathbf{B}_r^\top) \otimes \mathbf{P}_3 \tilde{\mathbf{c}}_r + \boldsymbol{\mathcal{E}}_M. \end{cases}$$

261 Thus the joint unmixing-and-HSR problem consists in finding the LL1 factors $\{\mathbf{A}_r \mathbf{B}_r^\top\}_{r=1}^R$,
262 $\mathbf{C}, \tilde{\mathbf{C}}$ under the assumption of (2.11), subject to the constraints

$$263 \quad (2.12) \quad \{\mathbf{A}_r \mathbf{B}_r^\top\}_{r=1}^R \geq \mathbf{0}, \mathbf{C} \geq \mathbf{0}, \tilde{\mathbf{C}} \geq \mathbf{0}.$$

265 **3. Recoverability analysis.** The unmixing-and-HSR problem aims at recovering an SRI
266 \mathcal{Z} and a variability tensor Ψ underlying the HSI in MSI, admitting a coupled LL1-BTD as
267 in (2.11) under the constraints (2.12). The mixing factors underlying the images must also
268 be recovered uniquely. In this section, we provide a noiseless recovery analysis[§] for the SRI
269 \mathcal{Z} and variability tensor Ψ . We show that our results hold for both image recovery and
270 estimation of the mixing factors. Although the following results are inspired by those of
271 [54, 13], there exist two main differences in our work. First, differently from [13], our model
272 accounts for variability. Second, we explicitly address unique recovery of the low-rank LL1
273 factors as mixing factors, which was not addressed in [13].

274 **Theorem 3.1.** *Assume that the SRIs \mathcal{Z} and $\tilde{\mathcal{Z}}$ admit BTDs as in (2.6) and (2.10), respec-*
275 *tively, that the HSI and MSI follow the coupled model (2.11), and that $\boldsymbol{\mathcal{E}}_H, \boldsymbol{\mathcal{E}}_M = \mathbf{0}$. Suppose*
276 *that $\{\mathbf{A}_r, \mathbf{B}_r\}_{r=1}^R, \mathbf{C}, \tilde{\mathbf{C}}$ are drawn from any absolutely continuous joint distributions and that*
277 *$\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3$ are full row rank. Let $\{\mathbf{A}_r^*, \mathbf{B}_r^*\}_{r=1}^R, \mathbf{C}^*, \tilde{\mathbf{C}}^*$ denote any solution to the unmixing-*
278 *and-HSR problem. Then with probability one, the true SRI \mathcal{Z} and degraded SRI $\tilde{\mathcal{Z}} \bullet_3 \mathbf{P}_3$ are*
279 *uniquely recovered by*

$$280 \quad \mathcal{Z} = \sum_{r=1}^R (\mathbf{A}_r^* (\mathbf{B}_r^*)^\top) \otimes \mathbf{c}_r^*, \quad \tilde{\mathcal{Z}} \bullet_3 \mathbf{P}_3 = \sum_{r=1}^R (\mathbf{A}_r^* (\mathbf{B}_r^*)^\top) \bullet_3 \mathbf{P}_3 \tilde{\mathbf{c}}_r^*,$$

282 if $I_H J_H \geq LR$, $IJ \geq L^2 R$ and

$$283 \quad \min \left(\left\lfloor \frac{I}{L} \right\rfloor, R \right) + \min \left(\left\lfloor \frac{J}{L} \right\rfloor, R \right) + \min(K_M, R) \geq 2R + 2.$$

285 Moreover, the abundance maps and spectral signatures represented by the LL1 factors
286 $\{\mathbf{S}_r^* = \mathbf{A}_r^*, \mathbf{B}_r^*\}_{r=1}^R, \mathbf{C}^*, \mathbf{P}_3 \tilde{\mathbf{C}}^*$ are recovered uniquely up to permutation and scaling ambigu-
287 ties.

288 Let us first recall the following lemma:

289 **Lemma 3.2.** [27, Lemma 1] *Let us denote $\tilde{\mathbf{A}} = \mathbf{P}\mathbf{A} \in \mathbb{R}^{I' \times L}$, where $\mathbf{P} \in \mathbb{R}^{I' \times I}$ is full*
290 *row rank and $\mathbf{A} \in \mathbb{R}^{I \times L}$ is drawn from any absolutely continuous joint distribution. Then $\tilde{\mathbf{A}}$*
291 *follows an absolutely continuous joint distribution.*

292 We can now derive the proof for Theorem 3.1.

[§]See also [45, S.1] that proposes exact recovery conditions based on the LL1-BTD, from a tensor completion perspective.

293 *Proof.* Let $\{\mathbf{A}_r, \mathbf{B}_r\}_{r=1}^R, \mathbf{C}, \tilde{\mathbf{C}}$ denote the groundtruth factors of the SRI tensors and
 294 let $\{\mathbf{A}_r^*, \mathbf{B}_r^*\}_{r=1}^R, \mathbf{C}^*, \tilde{\mathbf{C}}^*$ denote a solution to problem the unmixing-and-HSR problem un-
 295 der constraints (2.12). Moreover, let $\{\mathbf{S}_r\}_{r=1}^R$ denote the groundtruth abundance maps and
 296 $\{\mathbf{S}_r^*\}_{r=1}^R = \{\mathbf{A}_r^*(\mathbf{B}_r^*)^\top\}_{r=1}^R$. Then for $\boldsymbol{\varepsilon}_H, \boldsymbol{\varepsilon}_M = \mathbf{0}$, it holds that

$$297 \quad (3.1) \quad \mathbf{Y}_H = \sum_{r=1}^R (\mathbf{P}_1 \mathbf{A}_r (\mathbf{P}_2 \mathbf{B}_r)^\top) \otimes \mathbf{c}_r = \sum_{r=1}^R (\mathbf{P}_1 \mathbf{A}_r^* (\mathbf{P}_2 \mathbf{B}_r^*)^\top) \otimes \mathbf{c}_r^*,$$

$$298 \quad (3.2) \quad \mathbf{Y}_M = \sum_{r=1}^R \mathbf{S}_r \otimes \mathbf{P}_3 \tilde{\mathbf{C}}_r = \sum_{r=1}^R \mathbf{S}_r^* \otimes \mathbf{P}_3 \tilde{\mathbf{C}}_r^*.$$

300 Since by assumption, $\{\mathbf{A}_r, \mathbf{B}_r\}_{r=1}^R, \mathbf{C}, \tilde{\mathbf{C}}$ are drawn from absolutely continuous joint distribu-
 301 tions and $\mathbf{P}_1, \mathbf{P}_2$ and \mathbf{P}_3 are full row rank, it follows from Lemma 3.2 that $\{\mathbf{P}_1 \mathbf{A}_r, \mathbf{P}_2 \mathbf{B}_r\}_{r=1}^R,$
 302 $\mathbf{P}_3 \tilde{\mathbf{C}}$ follow certain absolutely continuous joint distributions.

303 Therefore, by Theorem 1.5, the LL1-BTD of \mathbf{Y}_M is essentially unique almost surely if
 304 $IJ \geq L^2 R$ and

$$305 \quad \min\left(\left\lfloor \frac{I}{L} \right\rfloor, R\right) + \min\left(\left\lfloor \frac{J}{L} \right\rfloor, R\right) + \min(K_M, R) \geq 2R + 2.$$

307 This means that

$$308 \quad (3.3) \quad \mathbf{S}^* = \mathbf{S} \mathbf{\Pi} \mathbf{\Lambda}, \quad \mathbf{P}_3 \tilde{\mathbf{C}}^* = \mathbf{P}_3 \tilde{\mathbf{C}} \mathbf{\Pi} \mathbf{\Lambda}^{-1},$$

310 where $\mathbf{\Pi}$ is a permutation matrix and $\mathbf{\Lambda}$ is a non-singular diagonal scaling matrix.

311 Next, let us define $\tilde{\mathbf{S}} = (\mathbf{P}_2 \boxtimes \mathbf{P}_1) \mathbf{S}$. We can see that $\tilde{\mathbf{S}}^* = \tilde{\mathbf{S}} \mathbf{\Pi} \mathbf{\Lambda}$, where $\tilde{\mathbf{S}}^* =$
 312 $(\mathbf{P}_2 \boxtimes \mathbf{P}_1) \mathbf{S}^*$. From [11, Lemma 3.3] and the proof of [13, Theorem II], $\tilde{\mathbf{S}}$ has full column
 313 rank almost surely if $I_H J_H \geq LR$.

314 Let us continue by considering $\mathbf{Y}_H^{(3)}$. From (3.1), we have

$$315 \quad \mathbf{Y}_H^{(3)} = \tilde{\mathbf{S}} \mathbf{C}^\top = \tilde{\mathbf{S}}^* (\mathbf{C}^*)^\top = \tilde{\mathbf{S}} \mathbf{\Pi} \mathbf{\Lambda} (\mathbf{C}^*)^\top.$$

317 Since $\tilde{\mathbf{S}}$ has full column rank, we thus have

$$318 \quad (3.4) \quad \mathbf{C}^* = \mathbf{C} \mathbf{\Pi} \mathbf{\Lambda}^{-1}.$$

320 Following (3.3) and (3.4), the LL1 factors $\mathbf{S}, \mathbf{C}, \mathbf{P}_3 \tilde{\mathbf{C}}$ are recovered uniquely up to permutation
 321 and scaling ambiguities by $\mathbf{S}^*, \mathbf{C}^*$ and $\mathbf{P}_3 \tilde{\mathbf{C}}^*$, respectively.

322 Finally, we can express the third unfolding of the SRI \mathbf{Z} and degraded $\tilde{\mathbf{Z}} \bullet_3 \mathbf{P}_3$ as

$$323 \quad \mathbf{Z}^{(3)} = \mathbf{S}^* (\mathbf{C}^*)^\top, \quad (\tilde{\mathbf{Z}} \bullet_3 \mathbf{P}_3)^{(3)} = \mathbf{S}^* (\mathbf{P}_3 \tilde{\mathbf{C}}^*)^\top,$$

325 which are the third unfoldings of the tensors in (3.1)–(3.2). ■

326 *Remark 3.3.* In the proof for Theorem 3.1, we can see that the low-rank factors $\mathbf{C}, \mathbf{P}_3 \tilde{\mathbf{C}}$
 327 and \mathbf{S} can be uniquely identified up to permutation and scaling ambiguities. This means that
 328 they can be interpreted as mixing factors underlying \mathbf{Z} and $\tilde{\mathbf{Z}} \bullet_3 \mathbf{P}_3$. Hence Theorem 3.1
 329 proposes unique recovery conditions for both the fusion and unmixing parts of the problem.

330 Theorem 3.1 indicates that we can only recover $\mathbf{P}_3\tilde{\mathbf{C}}$ uniquely, up to permutation and
 331 scaling ambiguities. Following (2.9), the variability matrix $\boldsymbol{\psi}$ can only be recovered from the
 332 MSI up to the spectral degradation \mathbf{P}_3 as

$$333 \quad \mathbf{P}_3\boldsymbol{\psi} = \mathbf{P}_3(\tilde{\mathbf{C}} - \mathbf{C}).$$

335 Thus the proposed model only allows to recover uniquely a spectrally-degraded version of the
 336 variability tensor, that is, $\boldsymbol{\Psi} \bullet_3 \mathbf{P}_3$.

337 *Remark 3.4.* In practice, there exist cases where spatial degradation is unknown; previous
 338 tensor-based HSR methods [27, 38, 13] proposed unique noiseless recovery conditions for the
 339 SRI in this scenario, also referred to as “spatially-blind”. However, such methods cannot be
 340 envisioned with the proposed approach. Conversely, since knowledge of \mathbf{P}_3 is not required to
 341 establish the above theorem, it is possible to seek for “spectrally-blind” algorithms, that do
 342 not require the spectral degradation matrix \mathbf{P}_3 .

343 **4. Algorithms.** In this section, we propose two algorithms based on the LL1-BTD. The
 344 first one is unconstrained and solves the HSR problem only. The second one enforced non-
 345 negativity constraints on the factors of the mixing model and proposes a solution to the joint
 346 unmixing-and-HSR problem.

347 **4.1. Unconstrained optimization.** In the remaining of this paper, for simplicity, we de-
 348 note $\tilde{\mathbf{C}}_M = [\tilde{\mathbf{c}}_{M,1}, \dots, \tilde{\mathbf{c}}_{M,R}] = \mathbf{P}_3\tilde{\mathbf{C}}$ and $\boldsymbol{\psi}_M = \mathbf{P}_3\boldsymbol{\psi}$. Regarding super-resolution, we only
 349 aim at recovering the SRI \mathbf{Z} and variability tensor $\boldsymbol{\Psi}$. In this framework, the latent LL1
 350 factors do not need to be interpretable. Thus, we can consider unconstrained optimization.
 351 As in [27], one possible approach for solving the HSR problem is to consider the following
 352 optimization problem:

(4.1)

$$353 \quad \underset{\mathbf{A}, \mathbf{B}, \mathbf{C}, \tilde{\mathbf{C}}_M}{\text{minimize}} \quad \mathcal{J}(\mathbf{A}, \mathbf{B}, \mathbf{C}, \tilde{\mathbf{C}}_M) + \mu \|\tilde{\mathbf{C}}_M - (\mathbf{P}_3\mathbf{C} + \boldsymbol{\psi}_M)\|_F^2,$$

$$354 \quad \text{s. to } \|\mathbf{c}_r\|_2 = 1, \quad \|\tilde{\mathbf{c}}_{M,r}\|_2 = 1, \quad \text{where}$$

$$355 \quad \mathcal{J}(\mathbf{A}, \mathbf{B}, \mathbf{C}, \tilde{\mathbf{C}}_M) = \|\mathbf{Y}_H - \sum_{r=1}^R (\mathbf{P}_1\mathbf{A}_r(\mathbf{P}_2\mathbf{B}_r)^\top) \otimes \mathbf{c}_r\|_F^2 + \lambda \|\mathbf{Y}_M - \sum_{r=1}^R (\mathbf{A}_r\mathbf{B}_r^\top) \otimes \tilde{\mathbf{c}}_{M,r}\|_F^2,$$

357 and λ is a balance parameter that controls the weights on the HSI and MSI[¶]. The regular-
 358 ization parameter μ controls the weight on the structural constraint $\tilde{\mathbf{C}}_M = \mathbf{P}_3\mathbf{C} + \boldsymbol{\psi}_M$. The
 359 unit norm constraints on the columns of \mathbf{C} and $\tilde{\mathbf{C}}_M$ are enforced to avoid convergence issues,
 360 and are addressed during optimization with a projected gradient approach. Since (4.1) is a
 361 non-convex cost function, we adopt a block coordinate descent scheme: the latent factors are
 362 updated sequentially by solving unconstrained convex quadratic programs.

363 Below, we provide the framework of the unconstrained algorithm, named BTD-Var. The
 364 updates for \mathbf{A} , \mathbf{B} and \mathbf{C} can be seen as generalized Sylvester equations and solved by effi-
 365 cient solvers, for instance, Hessenberg-Schur or Bartels-Stewart algorithms; see [43] for a full
 366 overview. The update for $\tilde{\mathbf{C}}_M$ is solved using normal equations. Please refer to Appendix A
 367 for a full derivation.

[¶]As in previous works [27]–[13], we consider that $\lambda = 1$ in our experiments.

Algorithm 4.1 BT-D-Var

Input: $\mathcal{Y}_H, \mathcal{Y}_M, \mathbf{B}, \mathbf{C}, \tilde{\mathbf{C}}_M, \mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3; R, L, \text{iter}, \lambda, \mu$
Output: $\mathbf{Z} \in \mathbb{R}^{I \times J \times K}, \Psi \bullet_3 \mathbf{P}_3 \in \mathbb{R}^{I \times J \times K_M}$
Initialization: $\psi_M = \tilde{\mathbf{C}}_M - \mathbf{P}_3 \mathbf{C}$
for $m \in \{1, \dots, \text{iter}\}$ **do**
 for $r \in \{1, \dots, R\}$ **do**
 $\mathbf{c}_r = \mathbf{c}_r / \|\mathbf{c}_r\|,$
 $\tilde{\mathbf{c}}_{M,r} = \tilde{\mathbf{c}}_{M,r} / \|\tilde{\mathbf{c}}_{M,r}\|,$
 end for
 $\mathbf{A} \leftarrow \arg \min_{\mathbf{A}} \|\mathbf{Y}_H^{(1)} - (\mathbf{C} \odot_p \mathbf{P}_2 \mathbf{B}) \mathbf{A}^\top \mathbf{P}_1^\top\|_F^2 + \lambda \|\mathbf{Y}_M^{(1)} - (\tilde{\mathbf{C}}_M \odot_p \mathbf{B}) \mathbf{A}^\top\|_F^2,$
 $\mathbf{B} \leftarrow \arg \min_{\mathbf{B}} \|\mathbf{Y}_H^{(2)} - (\mathbf{C} \odot_p \mathbf{P}_1 \mathbf{A}) \mathbf{B}^\top \mathbf{P}_2^\top\|_F^2 + \lambda \|\mathbf{Y}_M^{(2)} - (\tilde{\mathbf{C}}_M \odot_p \mathbf{A}) \mathbf{B}^\top\|_F^2,$
 $\mathbf{S} \leftarrow [\dots, \text{vec}\{\mathbf{A}_r \mathbf{B}_r^\top\}, \dots],$
 $\mathbf{C} \leftarrow \arg \min_{\mathbf{C}} \|\mathbf{Y}_H^{(3)} - (\mathbf{P}_2 \boxtimes \mathbf{P}_1) \mathbf{S} \mathbf{C}^\top\|_F^2 + \mu \|\tilde{\mathbf{C}}_M - (\mathbf{P}_3 \mathbf{C} + \psi_M)\|_F^2,$
 $\tilde{\mathbf{C}}_M \leftarrow \arg \min_{\tilde{\mathbf{C}}_M} \lambda \|\mathbf{Y}_H^{(3)} - \mathbf{S} \tilde{\mathbf{C}}_M^\top\|_F^2 + \mu \|\tilde{\mathbf{C}}_M - (\mathbf{P}_3 \mathbf{C} + \psi_M)\|_F^2,$
 $\psi_M \leftarrow \tilde{\mathbf{C}}_M - \mathbf{P}_3 \mathbf{C}.$
end for
return $\mathbf{Z}^{(3)} = \mathbf{S} \mathbf{C}^\top, (\Psi \bullet_3 \mathbf{P}_3)^{(3)} = \mathbf{S} \psi_M^\top.$

368 The computational cost per iteration of BT-D-Var can be decomposed as follows:
369 • $\mathcal{O}(I^3 + J^3 + K^3 + L^3 R^3)$ for solving the Sylvester equations;
370 • $\mathcal{O}(IJK_M R + I_H J_H K R)$ for computing the right-hand side in the least squares sub-
371 problem.

372 **4.2. Constrained optimization.** Although BT-D-Var allows for reconstruction of \mathbf{Z} and
373 Ψ , it is not guaranteed that its result can be interpretable in a mixing model. Indeed, non-
374 negativity constraints must be imposed on factors \mathbf{C} and $\tilde{\mathbf{C}}_M$ to provide them with physical
375 meaning. Differently from [53], we also impose non-negativity on $\{\mathbf{S}_r\}_{r=1}^R$, rather than on
376 the individual factors \mathbf{A}_r and \mathbf{B}_r . This way, \mathbf{c}_r and $\tilde{\mathbf{c}}_{M,r}$ (resp. \mathbf{S}_r) can be seen as spectral
377 signatures (resp. abundance maps) of the underlying SRI \mathbf{Z} and MSI \mathcal{Y}_M .

378 The resulting constrained optimization problem is:

$$379 \quad (4.2) \quad \underset{\mathbf{A}, \mathbf{B}, \{\mathbf{S}_r\}_{r=1}^R, \mathbf{C}, \tilde{\mathbf{C}}_M}{\text{minimize}} \quad \mathcal{J} + \mu \|\tilde{\mathbf{C}}_M - (\mathbf{P}_3 \mathbf{C} + \psi_M)\|_F^2 + \gamma \|\mathbf{S}_r - \mathbf{A}_r \mathbf{B}_r^\top\|_F^2$$

$$380 \quad (4.3) \quad \text{s. to } \{\mathbf{S}_r = \mathbf{A}_r \mathbf{B}_r^\top\}_{r=1}^R \geq \mathbf{0}, \mathbf{C} \geq \mathbf{0}, \tilde{\mathbf{C}}_M \geq \mathbf{0}, \|\mathbf{c}_r\|_2 = 1, \|\tilde{\mathbf{c}}_{M,r}\|_2 = 1,$$

382 where γ is a regularization parameter that controls the weight on the low-rank constraint.

383 Differently from (4.1), in (4.2)–(4.3) the \mathbf{S}_r factors are no longer latent variables, and are
384 subject to non-negativity constraints. Such constraints can be handled by using alternating
385 direction method of multipliers [9, 23]. As in [23], a non-negativity constraint is relaxed by
386 considering the surrogate $\iota_+(\cdot)$ (see Appendix B). Algorithm 4.2 presents the optimization
387 framework for (4.2)–(4.3).

388 The computational cost per-iteration of CNN-BTD-Var is:

389 • $\mathcal{O}(I^3 + J^3 + K^3 + L^3 R^3)$ for solving \mathbf{A} , \mathbf{B} and \mathbf{C} ;

Algorithm 4.2 CNN-BTD-Var

Input: $\mathcal{Y}_H, \mathcal{Y}_M, \mathbf{A}, \mathbf{B}, \mathbf{C}, \tilde{\mathbf{C}}_M, \mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3; R, L, \text{iter}, \lambda, \mu, \gamma$
Output: $\mathbf{S} \in \mathbb{R}^{I \times J \times R}, \mathbf{C} \in \mathbb{R}^{K \times R}, \tilde{\mathbf{C}}_M \in \mathbb{R}^{K_M \times R}, \mathbf{Z} \in \mathbb{R}^{I \times J \times K}, \Psi \bullet_3 \mathbf{P}_3 \in \mathbb{R}^{I \times J \times K_M}$
Initialization: $\psi_M = \tilde{\mathbf{C}}_M - \mathbf{P}_3 \mathbf{C}, \{\mathbf{S}_r = \mathbf{A}_r \mathbf{B}_r^\top\}_{r=1}^R$
for $m \in \{1, \dots, \text{iter}\}$ **do**
 for $r \in \{1, \dots, R\}$ **do**
 $\mathbf{c}_r = \mathbf{c}_r / \|\mathbf{c}_r\|,$
 $\tilde{\mathbf{c}}_{M,r} = \tilde{\mathbf{c}}_{M,r} / \|\tilde{\mathbf{c}}_{M,r}\|,$
 end for
 $\mathbf{A} \leftarrow \arg \min_{\mathbf{A}} \|\mathbf{Y}_H^{(1)} - (\mathbf{C} \odot_p \mathbf{P}_2 \mathbf{B}) \mathbf{A}^\top \mathbf{P}_1^\top\|_F^2 + \lambda \|\mathbf{Y}_M^{(1)} - (\tilde{\mathbf{C}}_M \odot_p \mathbf{B}) \mathbf{A}^\top\|_F^2$
 $+ \gamma \|\mathbf{S}_r - \mathbf{A}_r \mathbf{B}_r^\top\|_F^2,$
 $\mathbf{B} \leftarrow \arg \min_{\mathbf{B}} \|\mathbf{Y}_H^{(2)} - (\mathbf{C} \odot_p \mathbf{P}_1 \mathbf{A}) \mathbf{B}^\top \mathbf{P}_2^\top\|_F^2 + \lambda \|\mathbf{Y}_M^{(2)} - (\tilde{\mathbf{C}}_M \odot_p \mathbf{A}) \mathbf{B}^\top\|_F^2$
 $+ \gamma \|\mathbf{S}_r - \mathbf{A}_r \mathbf{B}_r^\top\|_F^2,$
 $\mathbf{S}_r \leftarrow \arg \min_{\mathbf{S}_r} \gamma \|\mathbf{A}_r \mathbf{B}_r^\top - \mathbf{S}_r\|_F^2 + \iota_+(\mathbf{S}_r),$ for $r \in \{1, \dots, R\},$
 $\mathbf{C} \leftarrow \arg \min_{\mathbf{C}} \|\mathbf{Y}_H^{(3)} - (\mathbf{P}_2 \boxtimes \mathbf{P}_1) \mathbf{S} \mathbf{C}^\top\|_F^2 + \mu \|\tilde{\mathbf{C}}_M - (\mathbf{P}_3 \mathbf{C} + \psi_M)\|_F^2 + \iota_+(\mathbf{C}),$
 $\tilde{\mathbf{C}}_M \leftarrow \arg \min_{\tilde{\mathbf{C}}_M} \lambda \|\mathbf{Y}_M^{(3)} - \mathbf{S} \tilde{\mathbf{C}}_M^\top\|_F^2 + \mu \|\tilde{\mathbf{C}}_M - (\mathbf{P}_3 \mathbf{C} + \psi_M)\|_F^2 + \iota_+(\tilde{\mathbf{C}}_M),$
 $\psi_M \leftarrow \tilde{\mathbf{C}}_M - \mathbf{P}_3 \mathbf{C}.$
end for
return $\mathbf{Z}^{(3)} = \mathbf{S} \mathbf{C}^\top, (\Psi \bullet_3 \mathbf{P}_3)^{(3)} = \mathbf{S} \psi_M^\top.$

390 • $\mathcal{O}(IJK_M R + I_H J_H K R)$ for computing the right-hand sides in the least squares sub-
391 problems.

392 **4.3. Initialization.** Many options are available to initialize the LL1 factors. Here, as
393 suggested in [12, Theorem 4.1], we initialize the \mathbf{A} and \mathbf{B} factors by generalized eigenvalue
394 decomposition of the matrix pencil $((\mathcal{Y}_M)_{\dots,1}^\top, (\mathcal{Y}_M)_{\dots,2}^\top)$ (see [8, 16]), using the `ll1_gevd`
395 function of TensorLab [46]. The \mathbf{C} and $\tilde{\mathbf{C}}_M$ factors are recovered by solving least-squares
396 problems. We combine these steps in an algebraic algorithm called `BTDRec`^{||} (Algorithm 4.3):

Algorithm 4.3 BTDRec

Input: $\mathcal{Y}_H, \mathcal{Y}_M, \mathbf{P}_1, \mathbf{P}_2; R, L$
Output: $\mathbf{A} \in \mathbb{R}^{I \times RL}, \mathbf{B} \in \mathbb{R}^{J \times RL}, \mathbf{C} \in \mathbb{R}^{K \times R}, \tilde{\mathbf{C}}_M \in \mathbb{R}^{K_M \times R}$
 $\mathbf{A}, \mathbf{B} \stackrel{\text{LL1}}{\approx} \mathcal{Y}_M,$
 $\mathbf{S}_r = \mathbf{A}_r \mathbf{B}_r^\top$ for $r \in \{1, \dots, R\},$
 $\mathbf{C}^\top = ((\mathbf{P}_2 \boxtimes \mathbf{P}_1) \mathbf{S})^\dagger \mathbf{Y}_H^{(3)},$
 $\tilde{\mathbf{C}}_M^\top = \mathbf{S}^\dagger \mathbf{Y}_M^{(3)}.$

^{||}echoing the initialization algorithm in [27] (called `TenRec`)

397 **5. Experiments for image recovery.** All simulations were run on a MacBook Pro with
 398 2.3 GHz Intel Core i5 and 16GB RAM. For basic tensor operations we used TensorLab 3.0
 399 [46]. The code was implemented in MATLAB and is available online at [https://github.com/](https://github.com/cprevost4/LL1_HSR_HU)
 400 [cprevost4/LL1_HSR_HU](https://github.com/cprevost4/LL1_HSR_HU).

401 **5.1. Degradation model.** The real SRI and MSI were acquired with the same spatial
 402 resolutions by the AVIRIS and Sentinel-2A instruments at different time instants, resulting
 403 in variability between the images. The spectral bands of \mathcal{Z} and \mathcal{Y}_M were normalized such
 404 that the 0.999 intensity quantile corresponded to a value of 1. This ensured that the unit
 405 norm constraint for the columns of \mathbf{C} and $\tilde{\mathbf{C}}_M$ did not cause any convergence problems. The
 406 HSI was obtained by spatial degradation of \mathcal{Z} by \mathbf{P}_1 and \mathbf{P}_2 , *i.e.*, the SRI \mathcal{Z} and the MSI
 407 \mathcal{Y}_M represented images of the same scene acquired on board of different missions, and $\tilde{\mathcal{Z}}$ was
 408 unknown. Afterwards, the SRI \mathcal{Z} was denoised (as described in [40]) to yield the high-SNR
 409 reference image [50]. We also conducted experiments in a “no-variability” scenario, *i.e.* we
 410 considered that the HSI and MSI were obtained by spatial (resp. spectral) degradation of the
 411 same SRI \mathcal{Z} .

412 For spatial degradation, we followed the commonly used Wald’s protocol [47]. The matrices
 413 $\mathbf{P}_1, \mathbf{P}_2$ were computed with a separable Gaussian blurring kernel of size $q = 9$. Downsampling
 414 was performed along each spatial dimension with a ratio $d = 4$ between the SRI and HSI, as
 415 in previous works [27]–[13]. Refer to Appendix C for more details on the construction of $\mathbf{P}_1,$
 416 \mathbf{P}_2 . White Gaussian noise with 30dB SNR was added to the HSI and MSI.

417 For the spectral degradation matrix \mathbf{P}_3 , we used the spectral response functions of two
 418 multispectral instruments**. For images with spectral variability, the Sentinel-2 sensors span
 419 the electromagnetic spectrum from 412nm to 2022nm and produced a 10-band MSI corre-
 420 sponding to the wavelengths 433–453nm (atmospheric correction), 458–522nm (soil, vegeta-
 421 tion), 543–577nm (green peak), 650–680nm (maximum chlorophyll absorption), 698–712nm
 422 (red edge), 733–747nm (red edge), 773–793nm (leaf area index, edge of NIR), 785–900nm (leaf
 423 area index), 855–875nm (NIR plateau), 935–955nm (water vapour absorption). The LAND-
 424 SAT sensor spanned the spectrum from 400nm to 2500nm for the HSI and produced a 6-band
 425 MSI corresponding to wavelengths 450–520nm (black), 520–600nm (green), 630–690nm (red),
 426 760–900nm (near-IR), 1550–1750nm (shortwave-IR) and 2050–2350nm (shortwave-IR2). This
 427 spectral response was used for real images without spectral variability. The spectral degrada-
 428 tion matrix \mathbf{P}_3 was a selection-weighting matrix that selected the common spectral bands of
 429 the SRI $\tilde{\mathcal{Z}}$ and the MSI.

430 **5.2. Metrics.** We compared the groundtruth SRI \mathcal{Z} with the recovered SRI $\hat{\mathcal{Z}}$ obtained
 431 by the algorithms. The main performance metric used in comparisons was the *reconstruction*
 432 *Signal-to-Noise ratio* (R-SNR):

$$433 \quad (5.1) \quad \text{R-SNR} = 10 \log_{10} \left(\frac{\|\mathcal{Z}\|_F^2}{\|\hat{\mathcal{Z}} - \mathcal{Z}\|_F^2} \right).$$

**available for download at <https://earth.esa.int/web/sentinel/user-guides/sentinel-2-msi/document-library/-/assetpublisher/Wk0TKajiSaR/content/sentinel-2a-spectral-responses> and <https://landsat.gsfc.nasa.gov/landsat-8/>.

434 In addition to R-SNR, we considered different metrics described below:

$$435 \quad (5.2) \quad \text{CC} = \frac{1}{IJK} \left(\sum_{k=1}^K \rho \left(\mathcal{Z}_{::,k}, \widehat{\mathcal{Z}}_{::,k} \right) \right),$$

436 where $\rho(\cdot, \cdot)$ is the Pearson correlation coefficient between the estimated and original spectral
437 slices;

$$438 \quad (5.3) \quad \text{ERGAS} = \frac{100}{d} \sqrt{\frac{1}{IJK} \sum_{k=1}^K \frac{\|\widehat{\mathcal{Z}}_{::,k} - \mathcal{Z}_{::,k}\|_F^2}{\mu_k^2}},$$

439 where μ_k^2 is the mean value of $\widehat{\mathcal{Z}}_{::,k}$. ERGAS represents the relative dimensionless global error
440 between the SRI and the estimate, which is the root mean-square error averaged by the size
441 of the SRI. We also used Spectral Angle Distance (SAD):

$$442 \quad (5.4) \quad \text{SAD} = \frac{1}{R} \sum_{r=1}^R \arccos \left(\frac{\mathbf{c}_r^T \widehat{\mathbf{c}}_r}{\|\mathbf{c}_r\|_2 \|\widehat{\mathbf{c}}_r\|_2} \right),$$

443 which computes the spectral angle distance between original and estimated spectra, and can
444 be used to assess unmixing performance as well. Performance for recovery of the abundance
445 maps was assessed using the root mean-squared error between reference \mathbf{S} and estimate $\widehat{\mathbf{S}}$:

$$446 \quad (5.5) \quad \text{RMSE} = \frac{1}{R} \sum_{r=1}^R \sqrt{\frac{1}{IJ} \sum_{d=1}^{IJ} \left((\mathbf{S}_r)_d - (\widehat{\mathbf{S}}_r)_d \right)^2}.$$

447 Finally, we considered the computational time for each algorithm, given by the `tic` and `toc`
448 functions of MATLAB.

449 **5.3. Recovery of the SRI and variability tensor.** In this subsection, we assessed the per-
450 formances of Algorithm 4.1 (BTD-Var) and Algorithm 4.2 (CNN-BTD-Var) for reconstruction
451 of the SRI \mathcal{Z} and degraded variability tensor $\Psi \bullet_3 \mathbf{P}_3$. We ran our algorithms with 20 outer
452 iterations at most and 5 inner iterations for CNN-BTD-Var. For initialization, out of 20 trials
453 of BTDRec we picked the one that provided the best reconstruction of the HSI and MSI. For
454 CNN-BTD-Var, we used $\mu = \gamma = 1$. We chose the hyperparameters R and L jointly. While
455 L was as large as possible inside the identifiability region provided by Theorem 3.1, R was
456 selected according to the real number of endmembers, when possible. Other rank-selecting
457 algorithms are available for hyperspectral images, see e.g., [2, 19]. Due to page limitations,
458 please refer to the Supplementary materials for a thorough discussion on the choice of R and
459 L .

460 For the SRI \mathcal{Z} , we compared our results to matrix-based approaches, including HySure
461 [42], CNMF [51] and GLP-HS [1]. We also considered tensor methods, namely STEREO
462 [27] for CP decomposition, SCOTT [38] for Tucker and CNN-BTD [53], which is a coupled
463 LL1-based algorithm that does not account for spectral variability. Finally, we considered
464 matrix and tensor methods accounting for variability, namely FuVar [4] (a matrix-based algo-
465 rithm based on the generalized linear mixing model), CT-STAR and CB-STAR [6], which are

466 tensor approaches based on multilinear decomposition accounting for spectrally and spatially
 467 localized changes. Except for CNMF, the baseline algorithms are unable to perform the un-
 468 mixing task. For Hysure, CNMF, GLP-HS and FuVar, we chose the ranks and regularization
 469 parameters according to the original works [42, 51, 4].

470 For reconstruction of $\Psi \bullet_3 \mathbf{P}_3$, we compared the results of our algorithms with those of
 471 CT-STAR and CB-STAR.

472 **5.3.1. Lake Tahoe.** The first dataset was Lake Tahoe with $\mathcal{Z} \in \mathbb{R}^{100 \times 80 \times 173}$. The SRI
 473 \mathcal{Z} and MSI \mathcal{Y}_M were respectively acquired on 2014-10-04 and 2017-10-24 by the Sentinel-
 474 2A sensor, resulting in high variability in the crops and lake areas). We ran STEREO with
 475 $F = 30$ and 10 iterations, and SCOTT with $R = (40, 40, 7)$ as in [6]. We ran CT-STAR with
 476 ranks $(18, 15, 10), (3, 3, 1)$, and CB-STAR with ranks $(20, 20, 9), (20, 20, 4)$. For our algorithms,
 477 as well as for CNN-BTD, we chose $R = 3$, $L = 20$ and $\lambda = 1$. Tables 1 and 2 display the
 478 reconstruction metrics and computation time for \mathcal{Z} and $\Psi \bullet_3 \mathbf{P}_3$ and all considered algorithms.
 The two best results of each column are shown in bold.

Table 1
 Reconstruction metrics for \mathcal{Z} , Lake Tahoe dataset

Algorithm	R-SNR	CC	SAD	ERGAS	Time
BTD-Var	15.0746	0.9384	9.7688	5.2081	2.6390
CNN-BTD-Var	16.1371	0.9514	7.2173	4.5902	1.2251
STEREO	5.8368	0.75957	30.7346	15.2801	1.2148
SCOTT	1.918	0.50379	47.1781	23.3815	0.14701
CNN-BTD	6.0332	0.80003	27.7993	14.9491	1.2826
CNMF	12.1314	0.87494	9.2422	7.2804	1.7442
GLP-HS	11.7862	0.87408	11.6106	7.6011	4.507
HySure	9.2687	0.81256	12.8228	10.1511	7.2761
FuVar	14.54	0.92498	6.7013	5.528	761.3932
CT-STAR	11.7676	0.87843	13.3433	7.6236	0.20849
CB-STAR	19.2413	0.97539	6.4649	3.2231	8.3597

Table 2
 Reconstruction metrics for $\Psi \bullet_3 \mathbf{P}_3$, Lake Tahoe dataset

Algorithm	R-SNR	CC	SAD	ERGAS
BTD-Var	13.8652	0.8584	14.7252	11.9947
CNN-BTD-Var	14.7347	0.8654	9.7916	11.8278
CT-STAR	11.4131	0.84542	17.7857	12.8223
CB-STAR	16.6599	0.94161	10.4442	7.8569

479 We can see that algorithms accounting for variability provided the best reconstruction
 480 metrics: in particular, the high performance of CB-STAR resulted from the fact that the
 481 algorithm accounts for spectrally and spatially localized changes. BTD-Var and CNN-BTD-
 482 Var provided slightly higher metrics than FuVar, but with lower computation time. Among
 483 the matrix-based approaches, CNMF showed the best reconstruction performance. Finally,
 484 other tensor-based approaches, although fast, yielded worse reconstruction metrics, due to the
 485 fact that they did not consider the variability. CB-STAR also provided the best metrics for
 486 reconstruction of $\Psi \bullet_3 \mathbf{P}_3$. However, its computation time was large. The proposed algorithms
 487 showed competitive metrics and even slightly outperformed baseline methods in terms of CC,
 488

489 but with slightly higher computation time.

490 In addition, we plot in Figure 2 the 40th spectral band of the reference and estimated
 491 SRI. The proposed approaches recovered the SRI spectral band accurately.

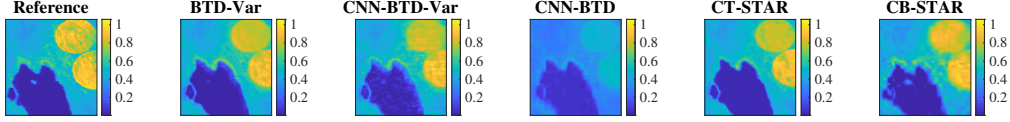


Figure 2. Single spectral band of the SRI, Lake Tahoe dataset

491

492 **5.3.2. Ivanpah Playa.** We also considered Ivanpah Playa with $\mathcal{Z} \in \mathbb{R}^{80 \times 128 \times 173}$ and large
 493 acquisition time difference: the SRI and MSI were acquired on 2015-10-26 and 2017-12-17
 494 respectively, by the Sentinel-2 sensor. We ran STEREO with $F = 10$ and 10 iterations and
 495 SCOTT with $R = (30, 30, 10)$. We ran CT-STAR with ranks $(10, 15, 8), (3, 3, 2)$, and CB-
 496 STAR with ranks $(40, 40, 4), (40, 40, 5)$. For our algorithms and CNN-BTD, we chose $R = 4$
 497 and $L = 18$.

Tables 3 and 4 show the reconstruction metrics and computation time.

Table 3
 Reconstruction metrics for \mathcal{Z} , Ivanpah Playa dataset

Algorithm	R-SNR	CC	SAD	ERGAS	Time
BTD-Var	19.5964	0.8684	2.4533	2.6223	3.2545
CNN-BTD-Var	24.7991	0.9533	1.5943	1.4725	2.2159
STEREO	6.0987	0.76283	29.0278	12.6747	0.93975
SCOTT	2.4445	0.34257	47.9598	19.372	0.2645
CNN-BTD	5.7515	0.33492	28.7006	13.1899	11.8775
CNMF	21.6059	0.90114	1.3019	2.1138	2.6656
GLP-HS	19.433	0.86261	3.3413	2.697	5.9218
HySure	18.4551	0.85218	3.3249	3.0653	10.4606
FuVar	22.0332	0.90354	1.5062	2.0189	526.1659
CT-STAR	21.1186	0.88849	1.9424	2.2386	0.15373
CB-STAR	25.7174	0.96003	1.3269	1.3228	8.2923

Table 4
 Reconstruction metrics for $\Psi \bullet_3 P_3$, Ivanpah Playa dataset

Algorithm	R-SNR	CC	SAD	ERGAS
BTD-Var	19.3624	0.7057	2.7901	39.5265
CNN-BTD-Var	23.6558	0.9207	1.3826	15.0054
CT-STAR	19.3597	0.73396	2.1977	33.853
CB-STAR	23.4888	0.90832	1.1567	16.9815

498

499 The best metrics were provided by CB-STAR, then CNN-BTD-Var. BTD-Var had a
 500 performance comparable to that of GLP-HS for reconstruction of \mathcal{Z} . Its performance was
 501 comparable to that of CT-STAR for $\Psi \bullet_3 P_3$. For this dataset as well, the proposed algorithms
 502 were faster than some other algorithms, including CNMF, CB-STAR, and FuVar. Other
 503 matrix-based approaches also provided satisfying reconstruction. However, STEREO, SCOTT
 504 and CNN-BTD provided the worst reconstruction metrics.

505 In Figure 3 we plot the 40th spectral band of the reference SRI and the estimated SRI for
 507 our algorithms, CNN-BTD, CNMF and CB-STAR for comparison. For this dataset, we can

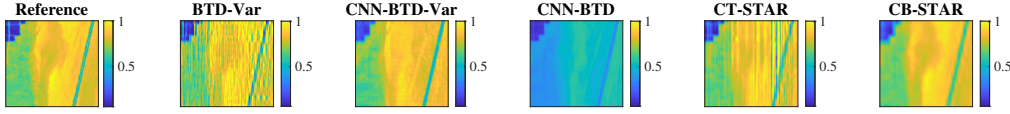


Figure 3. Single spectral band of the SRI, Ivanpah Playa dataset

506 see that CNN-BTD-Var yielded a better SRI spectral band reconstruction than BTD-Var.

508 **5.3.3. Lockwood.** The third dataset we considered was Lockwood with $\mathcal{Z} \in \mathbb{R}^{80 \times 100 \times 173}$.
 509 The SRI \mathcal{Z} and MSI were acquired on 2018-08-20 and on 2018-10-19. This was an example
 510 where only acquisition variations happened (which affected the image mostly uniformly), thus
 511 it illustrated the adequacy of the proposed variability model compared to more general ones
 512 proposed in the literature. We ran STEREO with $F = 100$ and 10 iterations and SCOTT
 513 with $\mathbf{R} = (60, 60, 5)$. We ran CT-STAR with ranks $(30, 30, 8), (3, 3, 2)$, and CB-STAR with
 514 ranks $(70, 70, 5), (40, 40, 3)$. For our algorithms, as well as for CNN-BTD, we chose $R = 9$ and
 515 $L = 16$. The reconstruction metrics are displayed in Tables 5 and 6.

Table 5
 Reconstruction metrics for \mathcal{Z} , Lockwood dataset

Algorithm	R-SNR	CC	SAM	ERGAS	Time (sec)
BTD-Var	20.1273	0.918432	2.92921	6.35566	5.46272
CNN-BTD-Var	19.4882	0.906525	3.0299	6.29101	4.11573
STEREO	6.552	0.80196	27.3623	25.1749	1.8835
SCOTT	2.2276	0.79276	28.5771	45.9608	0.2228
CNN-BTD	6.4909	0.81475	27.4245	25.436	2.3082
CNMF	18.7829	0.89063	2.9768	6.7014	4.353
GLP-HS	18.6734	0.88849	3.2079	6.9979	6.8463
HySure	14.125	0.8633	4.4044	11.6	6.9823
CT-STAR	18.4987	0.88287	4.571	8.2657	3.3013
CB-STAR	19.0751	0.89445	3.3707	7.2926	68.0282

Table 6
 Reconstruction metrics for $\Psi \bullet_3 P_3$, Lockwood dataset.

Algorithm	R-SNR	CC	SAM	ERGAS
BTD-Var	18.8768	0.810171	2.59862	11.9253
CNN-BTD-Var	18.3523	0.818424	2.76538	11.2095
CT-STAR	17.2744	0.73293	4.1677	15.8113
CB-STAR	17.5513	0.7402	3.2858	13.3116

516 For both \mathcal{Z} and $\Psi \bullet_3 P_3$, the best reconstruction metrics were generally provided by BTD-
 517 Var and CNN-BTD-Var. They were followed by CT-STAR and CB-STAR. The slightly better
 518 results obtained by our algorithms illustrate the fact that the variability model considered in
 519 [6] can represent spatially localized changes, but is not very appropriate or interpretable for

520 acquisition or illumination variations. In Figure 4 we plot the 40th spectral band of the
reference and estimated SRI.

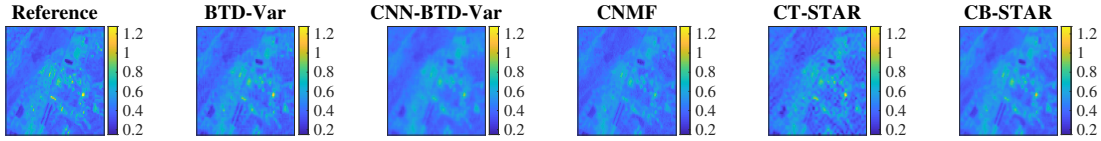


Figure 4. Spectral band no.40 of the SRI, Lockwood.

521

522 **5.4. Recovery without variability.** In this subsection, we assessed recovery performance
523 for the SRI \mathcal{Z} . We considered that the HSI and MSI are both degraded versions of the same
524 SRI \mathcal{Z} . In other words, we suppose that $\mathcal{Z} = \tilde{\mathcal{Z}}$, hence there is no spectral variability.
525 The dataset we considered was Indian Pines, where $\mathcal{Z} \in \mathbb{R}^{144 \times 144 \times 200}$ was degraded by a
526 LANDSAT sensor for the MSI and a downsampling ratio $d = 4$ for the HSI. We ran STEREO
527 with $F = 50$, SCOTT with $R = (40, 40, 6)$ and LL1-based algorithms with $R = 6$ and $L = 13$.
528 We tuned the other algorithms according to original works. The reconstruction metrics for
the SRI are presented in Table 7.

Table 7
Reconstruction metrics for \mathcal{Z} , Indian Pines dataset

Algorithm	R-SNR	CC	SAD	ERGAS	Time
BTD-Var	26.4299	0.8398	2.2384	1.1750	5.8435
CNN-BTD-Var	25.1166	0.8326	2.4127	1.2675	4.2968
STEREO	27.69	0.86669	1.9461	0.99959	1.8564
SCOTT	26.2451	0.86196	2.2694	1.1208	0.21087
CNN-BTD	25.2263	0.80949	2.5035	1.3497	24.5326
CNMF	27.2552	0.83978	1.9502	1.2056	8.2147
GLP-HS	26.2837	0.83813	2.2794	1.2918	14.2957
HySure	20.4281	0.66661	4.4916	2.5723	25.2202
CT-STAR	24.0398	0.84385	2.4839	1.3151	0.16528
CB-STAR	26.5216	0.86749	2.1265	1.0556	3.6761

529

530 The best reconstruction metrics were generally provided by STEREO. The proposed al-
531 gorithms had performance comparable to that of SCOTT, and computation time comparable
532 to that of CB-STAR. The slightly lower performance of constrained algorithms accounting
533 for variability can be explained by the use of more flexible models. In this specific scenario,
534 other methods based on a more restrictive model fitted the data more tightly. Nonetheless,
535 algorithms accounting for variability offered competitive performance in the “no-variability”
536 case. However, their computation time was usually higher than that of state-of-the-art tensor
537 approaches. In Figure 5 we plot the 40th spectral band of the reference and estimated SRI.

538

6. Performance for unmixing of an unknown SRI.

539

540 **6.1. Experiments setup.** In this section, we assessed the performance of CNN-BTD-Var
for unmixing of an unknown SRI on synthetic datasets, and real examples from Section 5.

541

542 We compared our results with those of CNMF [51] initialized by VCA [35]. We also
considered traditional unmixing algorithms: accelerated multiplicative algorithm (MU-Acc)

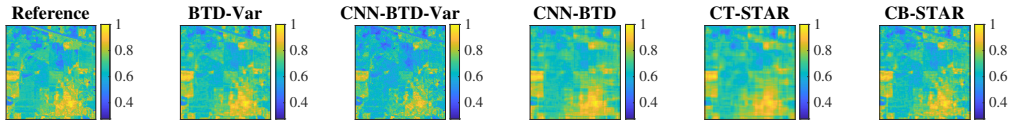


Figure 5. Single spectral band of the SRI, Indian Pines dataset

543 [21] and BMDR-ADMM [36] (enforcing minimum dispersion constraint on the spectra, and
 544 sum-to-one on the abundance maps). We ran these algorithms on the recovered SRI $\hat{\mathbf{Z}}$
 545 obtained from CB-STAR, which gave the best reconstruction metrics in the previous section.
 546 For these algorithms, we chose the parameters according to the original works. We set the
 547 number of materials to R . Since other fusion algorithms were not designed for unmixing, they
 548 were not directly included in this comparison.

549 For each dataset, we compared the abundance maps and spectral signatures obtained by
 550 the algorithms to references obtained by VCA or groundtruth materials. We also assessed the
 551 unmixing performance numerically by comparing the SAD, RMSE and computation time for
 552 the considered algorithms. The best results of each row are shown in bold in the tables.

553 **6.2. Unmixing with exact LL1 model.** We first assessed the unmixing performance with
 554 synthetic datasets. We tested our approach in the case where the SRI and variability tensor
 555 admit an exact LL1-BTD. Although these datasets did resemble real spectral images, they
 556 allowed us to assess unmixing performance in a case where the uniqueness conditions for the
 557 non-negative matrix factorization model (see [14, 32]) were not fulfilled.

558 **6.2.1. Generating synthetic datasets.** We considered $R = 3$ spectral signatures \mathbf{c}_r ($r \in$
 559 $\{1, \dots, R\}$) obtained from the Jasper Ridge reference data^{††}, corresponding to vegetation, soil
 560 and road materials. The SRI $\mathbf{Z} \in \mathbb{R}^{I \times J \times K}$ ($I = J = 90$, $K = 173$) was split into LR equal
 561 blocks in the spatial dimensions, with $L = 3$.

562 Each abundance map \mathbf{S}_r ($r \in \{1, \dots, R\}$) was a block matrix with $\frac{I}{L} \times \frac{J}{L}$ blocks. We
 563 generated a multiplicative variability matrix ψ_{multi} with random real entries drawn from the
 564 standard uniform distribution in the open interval $[0.9, 1.1]$. We then computed the altered
 565 spectra $\tilde{\mathbf{C}} = \psi_{\text{multi}} \square \mathbf{C}$. The variability matrix ψ that we aimed at recovering was obtained
 566 as $\psi = \tilde{\mathbf{C}} - \mathbf{C}$ so that it had zero mean.

567 Formally, we computed the high-resolution tensors as

$$568 \quad \mathbf{Z} = \sum_{r=1}^R \mathbf{S}_r \otimes \mathbf{c}_r, \quad \Psi = \sum_{r=1}^R \mathbf{S}_r \otimes \psi_r, \quad \tilde{\mathbf{Z}} = \mathbf{Z} + \Psi.$$

569

570 The HSI and MSI were obtained by degradation of the SRIs according to model (2.3). For
 571 $\mathbf{P}_1 = \mathbf{P}_2$, we had $q = 9$ and $d = 3$ so that $I_H = J_H = 30$. For \mathbf{P}_3 , we chose the spectral
 572 response of the Sentinel-2 MS sensor, which led to $K_M = 10$.

573 **6.2.2. Separable example.** In the first example, we generated a dataset for which the
 574 pure pixel assumption was valid. Thus in each $\frac{I}{L} \times \frac{J}{L}$ block, at most one material was active,
 575 as indicated by the numerals in the parcel map shown in Table 8. Each block in the parcel map
 576 was a patch composed of entries equal to one. The abundance maps resembled agricultural

^{††}Available for download at <http://lesun.weebly.com/hyperspectral-data-set.html>.

Table 8

Parcel map for the first synthetic dataset

1	2	3
3	1	2
2	3	1

577 fields. This was a case for which non-negative matrix factorization under minimal volume
 578 constraint was unique [20, 18, 24]. Only unconstrained non-negative matrix factorization was
 579 not unique.

580 We ran CNN-BTD-Var with $R = 3$ and $L = 3$; for other algorithms, we used $R = 3$. The
 581 spectral signatures \mathbf{c}_r and abundance maps \mathbf{S}_r are shown in Figures 6 and 7, respectively.
 The unmixing metrics and computation time are displayed in Table 9.

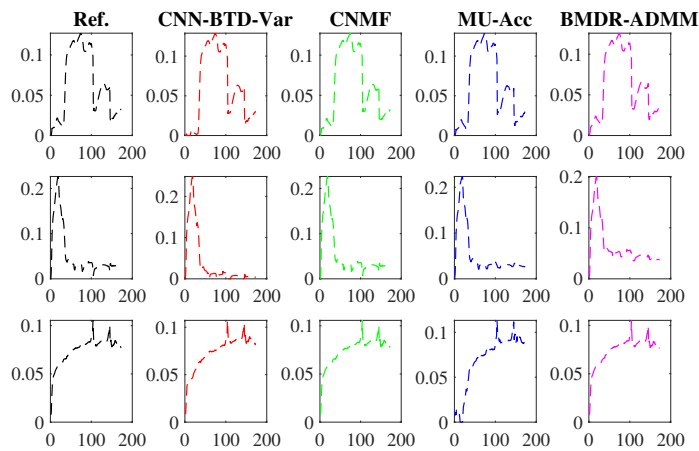


Figure 6. Reference and estimated spectra, synthetic dataset 1

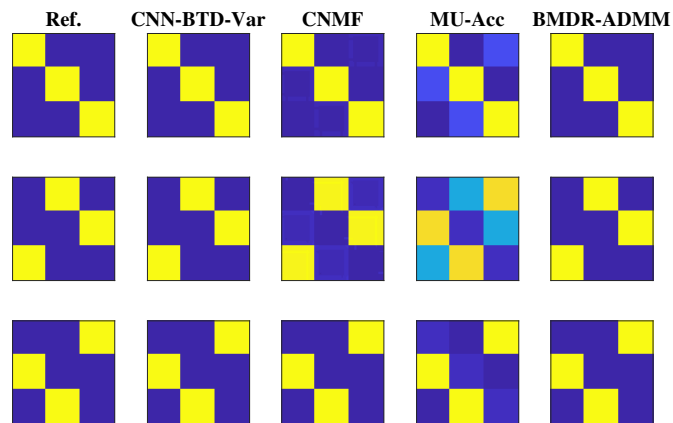


Figure 7. Reference and estimated abundance maps, synthetic dataset 1

582

Table 9
Unmixing, synthetic example 1

Algorithm	CNN-BTD-Var	CNMF	MU-Acc	BMDR-ADMM
SAD	0.012349	0.300049	0.132655	0.351785
RMSE	0.102441	0.274509	0.267861	0.201845
Time (sec)	0.958311	1.43277	1.73498	1.23186

583 We can see that all spectra and abundance maps were recovered accurately by CNN-
584 BTD-Var, with visual quality comparable to that of CNMF and BMDR-ADMM. The proposed
585 algorithm gave the best unmixing metrics and computation time. We also see some artifacts in
586 the abundance maps recovered by CNMF. Moreover, MU-Acc did not estimate all abundance
587 maps correctly for this example.

588 **6.2.3. Synthetic example with non-identifiable matrix factorization model.** In this sec-
589 ond example, we designed an example where the separability (or pure pixel) condition [14, 32]
590 was not fulfilled. This resulted in the traditional non-negative matrix factorization model
591 being non-identifiable. This was in fact a highly mixed situation for which we expected that
592 traditional unmixing algorithms fail at performing unmixing on this dataset. However, the
593 conditions in Theorem 3.1 were satisfied, which made the LL1 factors unique up to permuta-
594 tion and scaling ambiguities. The abundance maps \mathbf{S}_r were designed as follows:

$$595 \quad \mathbf{S}_1 = \frac{1}{12} \begin{bmatrix} 5 & 7 & 6 \\ 7 & 3 & 5 \\ 3 & 0 & 0 \end{bmatrix} \boxtimes \mathbf{H}, \quad \mathbf{S}_2 = \frac{1}{12} \begin{bmatrix} 7 & 5 & 3 \\ 0 & 6 & 0 \\ 3 & 5 & 7 \end{bmatrix} \boxtimes \mathbf{H}, \quad \mathbf{S}_3 = \frac{1}{12} \begin{bmatrix} 0 & 0 & 3 \\ 5 & 3 & 7 \\ 6 & 7 & 5 \end{bmatrix} \boxtimes \mathbf{H},$$

596 with \mathbf{H} a Gaussian of size 30×30 with standard deviation $\sigma = 5$.

597 We ran CNN-BTD-Var with $R = 3$ and $L = 3$; for other algorithms, we used $R = 3$. The
598 spectral signatures \mathbf{c}_r and abundance maps \mathbf{S}_r are shown in Figures 8 and 9, respectively.
599 The unmixing metrics are shown in Table 10.

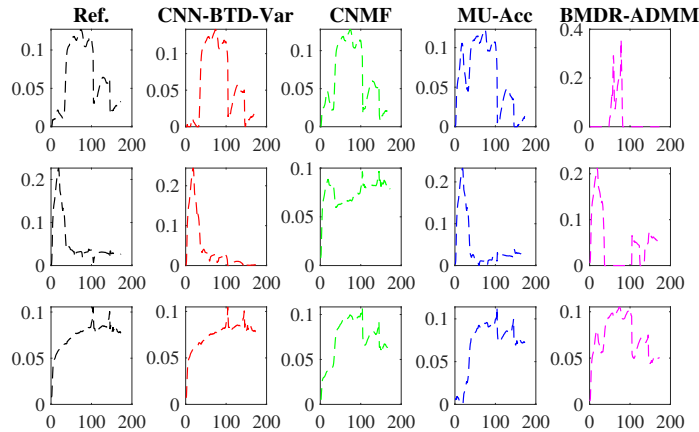


Figure 8. Reference and estimated spectra, synthetic dataset 2

600

601

602

The spectral signatures were best reconstructed by CNN-BTD-Var, although CNMF only reconstructed the first spectrum correctly. Moreover, only CNN-BTD-Var provided reasonable

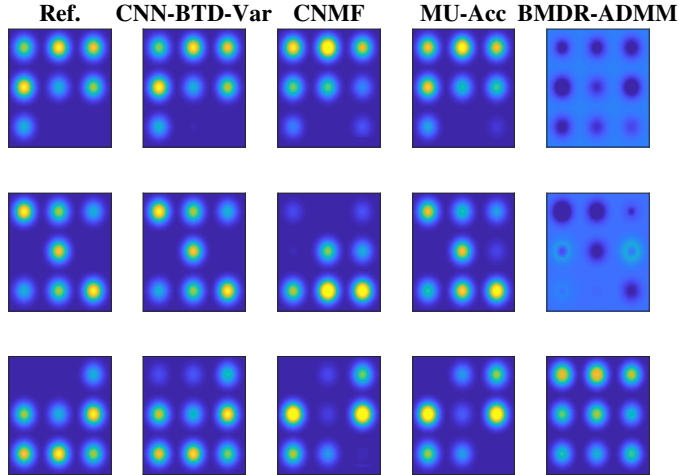


Figure 9. Reference and estimated abundance maps, synthetic dataset 2

Table 10
Unmixing, synthetic example 2

Algorithm	CNN-BTD-Var	CNMF	MU-Acc	BMDR-ADMM
SAD	0.160109	0.386413	0.355248	0.464474
RMSE	0.210952	0.333595	0.253813	0.339372
Time (sec)	0.972156	1.77138	1.57192	1.20186

603 estimates of the abundance maps, while other algorithms failed. Numerically, the proposed
604 algorithm yielded the best unmixing metrics and computation time.

605 **6.3. Unmixing for real datasets.** In this subsection, we assessed unmixing performance
606 of CNN-BTD-Var for real datasets Lake Tahoe and Ivanpah Playa. For the two considered
607 datasets, we followed the same degradation model as in Section 5. For these experiments,
608 the endmembers and abundance maps underlying \mathcal{Z} were unknown: as a result, we chose
609 as reference the spectra and abundance maps selected manually from the SRI $\mathcal{Z}^{\ddagger\ddagger}$. The
610 obtained abundance maps had very close correspondence with visual features in the image.
611 The columns of the abundance maps were rescaled with unit norm for comparison.

612 **6.3.1. Lake Tahoe.** We first considered the Lake Tahoe dataset. This dataset was mainly
613 composed of $R = 3$ materials: water (lake), soil and vegetation. As a result, we chose $R = 3$
614 and $L = 18$ as in the previous subsection. We compared our algorithm with CNMF, MU-Acc
615 and BMDR-ADMM with $R = 3$.

616 On Figures 10 and 11, we plot the estimated spectra and abundance maps.

617 The proposed approach estimated the spectra accurately. The abundance maps allowed for
618 identification of the areas corresponding to different materials, although with lower resolution
619 than other methods. Additionally, the abundance maps recovered by CNN-BTD-Var seemed
620 to be low-rank. The algorithms CNMF and MU-Acc did not recover the water abundance
621 map correctly, and CNMF did not recover the water spectrum. In Table 11, we show the

^{‡‡}In real applications, the SRI \mathcal{Z} is unknown. In this paper, we use it as a reference to evaluate the performance of our approach.

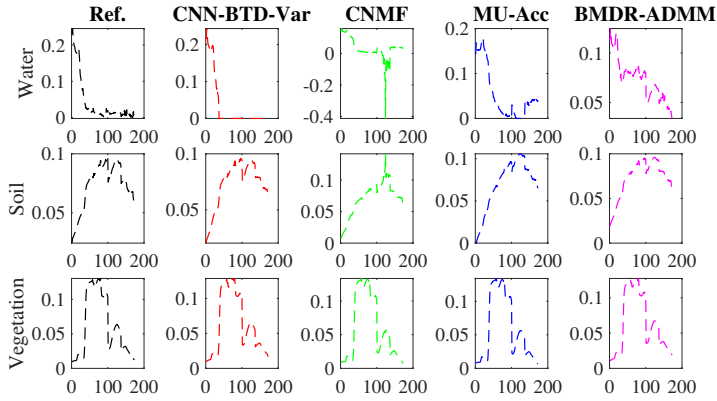


Figure 10. Reference and estimated spectra, Lake Tahoe dataset

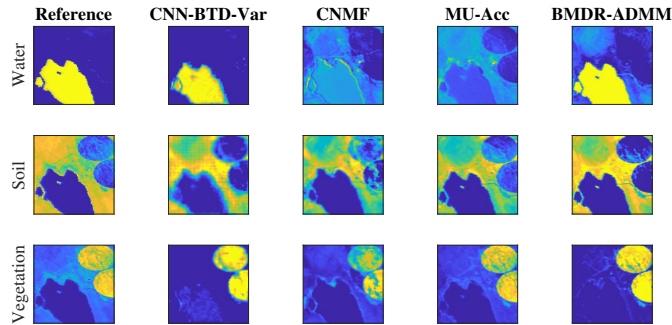


Figure 11. Reference and estimated abundance maps, Lake Tahoe dataset

unmixing and computation time for the considered algorithms. CNN-BTD-Var provided the

Table 11
Unmixing, Lake Tahoe dataset

Algorithm	CNN-BTD-Var	CNMF	MU-Acc	BMDR-ADMM
SAD	0.0794406	0.302813	0.1098101	0.255009
RMSE	0.466916	0.472743	0.637745	0.356724
Time (sec)	1.229906	1.98253	2.0130503	1.71438

622

best SAD and computation time, and the second best RMSE after BMDR-ADMM.

624

Additionally, on Figure 12, we plot the reference and estimated $\tilde{\mathbf{C}}_M$ and $\mathbf{P}_3\boldsymbol{\psi} = \tilde{\mathbf{C}}_M - \mathbf{P}_3\mathbf{C}$ obtained from CNN-BTD-Var.

626

In Figure 12, the water spectrum had high variability for the first MSI spectral band, which corresponds to the blue region. For the vegetation and soil spectra, the largest variability was found at spectral bands corresponding to the green and orange-red wavelengths. Moreover, CNN-BTD-Var recovered the reference $\tilde{\mathbf{C}}_M$ and $\mathbf{P}_3\boldsymbol{\psi}$ with a small discrepancy.

629

630

6.3.2. Ivanpah Playa.

Next, we considered the Ivanpah Playa dataset. This dataset was composed of $R = 4$ materials: solar panels, dark sand, yellow sand and road. We ran CNN-BTD-Var with $R = 4$, $L = 18$, and compared the results to other baseline algorithms with $R = 4$.

634

In Figures 13 and 14, we plot the reference and estimated spectra and abundance maps. In Table 12, we show the unmixing metrics and computation time.

635

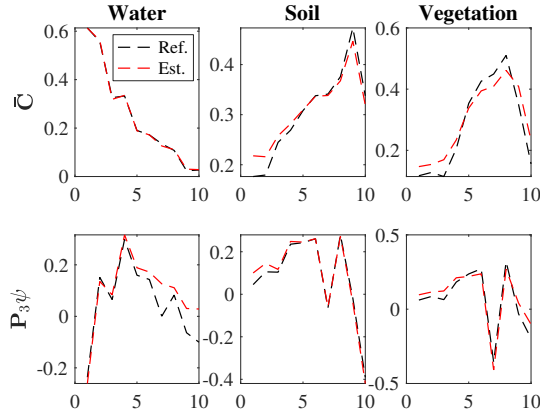


Figure 12. Reference and estimated \tilde{C}_M and $P_3\psi$, Lake Tahoe dataset

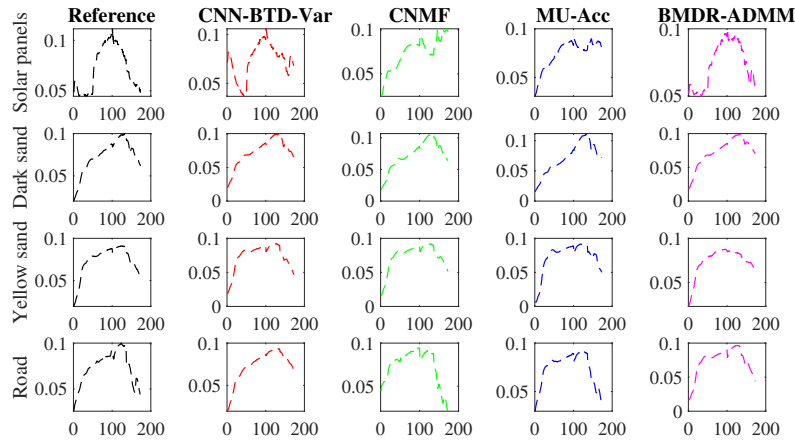


Figure 13. Reference and estimated spectra, Ivanpah Playa dataset

636 One difficulty for unmixing was the important similarity between the reference spectra
 637 (in particular, dark and yellow sand, road materials). This led to almost colinear columns
 638 in C , which resulted in CNN-BTD-Var giving the worse SAD. This issue was particularly
 639 visible in Figure 15 with estimated \tilde{C}_M and $P_3\psi$. High variability was found for the red and
 640 near-infrared spectral bands for all materials.

641 Despite this difficulty, CNN-BTD-Var yielded the best SAD and RMSE and recovered
 642 the solar panels and road abundance maps best, while the yellow sand map was slightly
 643 better for BMDR-ADMM. For this example, the low-rank assumption for abundance maps
 644 was reasonable: see that corresponding to solar panels. This assumption allowed for better
 645 visual reconstruction of this abundance map. Contrary to other algorithms, all spectra were
 646 correctly recovered.

647 **7. Conclusion.** In this paper, we proposed new algorithms for solving the HSR problem
 648 with variable images, using an LL1-BTD model. First, we showed that in the presence of
 649 variability, previous tensor models fail at recovering the SRI, since they do not account for
 650 spectral or spatial variability. Our approach allows to recover the SRI accurately for the
 651 considered datasets, as well as the degraded variability tensor.

652 An appropriate choice of ranks also allows our algorithms to estimate underlying spectra

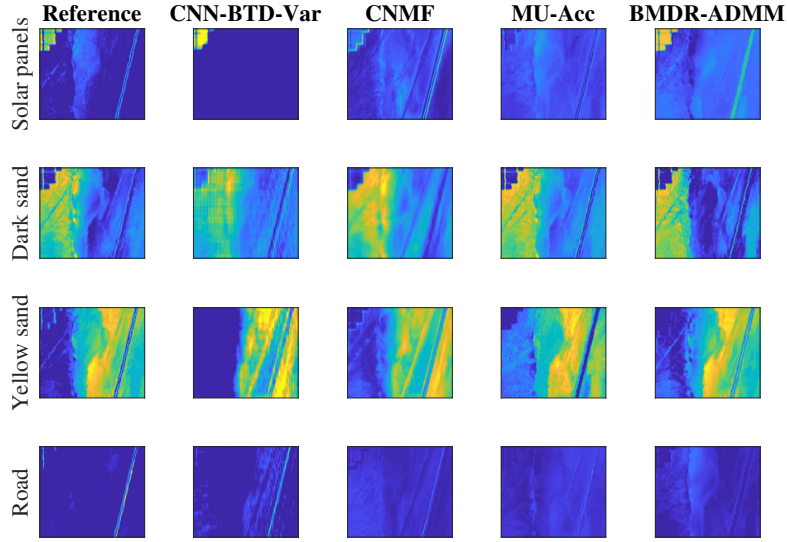


Figure 14. Reference and estimated abundance maps, Ivanpah Playa dataset

Table 12
Unmixing, Ivanpah Playa dataset

Algorithm	CNN-BTD-Var	CNMF	MU-Acc	BMDR-ADMM
SAD	0.094346	0.193547	0.134738	0.113456
RMSE	0.006693	0.007067	0.008188	0.007434
Time (sec)	1.30258	1.73402	1.56564	1.679304

653 and abundance maps of the unknown SRI, with performance comparable to those of tra-
 654 ditional unmixing algorithms applied on the SRI directly. Non-negativity priors allow the
 655 low-rank factors of our model to be interpretable, without having a high negative impact on
 656 the computation time.

657 Appendix A. Unconstrained factor updates in Algorithm 4.1.

658 In Algorithm 4.1, the least squares program for \mathbf{A} can be seen as a generalized Sylvester
 659 equation of the form $\mathbf{X}_1 \mathbf{A} \mathbf{X}_2 + \mathbf{X}_3 \mathbf{A} \mathbf{X}_4 = \mathbf{X}_5$, with

$$660 \quad \mathbf{X}_1 = \mathbf{P}_1^\top \mathbf{P}_1, \quad \mathbf{X}_2 = (\mathbf{C} \odot_p \mathbf{P}_2 \mathbf{B})^\top (\mathbf{C} \odot_p \mathbf{P}_2 \mathbf{B}), \quad \mathbf{X}_3 = \lambda \mathbf{I}_I,$$

$$661 \quad \mathbf{X}_4 = (\tilde{\mathbf{C}}_M \odot_p \mathbf{B})^\top (\tilde{\mathbf{C}}_M \odot_p \mathbf{B}), \quad \mathbf{X}_5 = \mathbf{P}_1^\top (\mathbf{Y}_H^{(1)})^\top (\mathbf{C} \odot_p \mathbf{P}_2 \mathbf{B}) + \lambda (\mathbf{Y}_M^{(1)})^\top (\tilde{\mathbf{C}}_M \odot_p \mathbf{B}),$$

663 and can be solved with efficient solvers. The updates for \mathbf{B} and \mathbf{C} can be solved similarly.

664 The pseudo-solution for $\tilde{\mathbf{C}}_M$ is expressed as $\text{vec}\{\tilde{\mathbf{C}}_M\} = (\mathbf{X}^\top \mathbf{X})^\dagger \mathbf{X}^\top \mathbf{z}$, with

$$665 \quad \mathbf{X} = (\sqrt{\lambda} \mathbf{S}^\top \mathbf{S} + \sqrt{\mu} \mathbf{I}_R) \boxtimes \mathbf{I}_{K_M},$$

$$666 \quad \mathbf{z} = \text{vec}\{\sqrt{\lambda} (\mathbf{Y}_M^{(3)})^\top \mathbf{S} + \sqrt{\mu} (\mathbf{P}_3 \mathbf{C} + \boldsymbol{\psi}_M)\}.$$

668

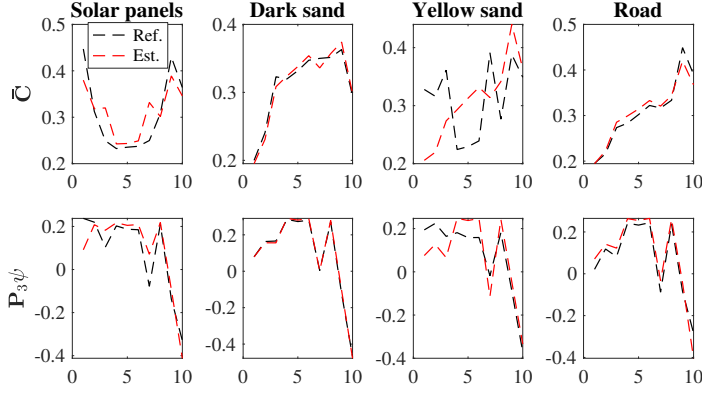


Figure 15. Reference and estimated \tilde{C}_M and $P_3\psi$, Ivanpah Playa dataset

669 Appendix B. Updates in Algorithm 4.2.

670 Similarly to Algorithm 4.1, the least squares programs for \mathbf{A} and \mathbf{B} in Algorithm 4.2 can
 671 be viewed as generalized Sylvester equations of the form $\mathbf{X}_1\mathbf{A}\mathbf{X}_2 + \mathbf{X}_3\mathbf{A}\mathbf{X}_4 = \mathbf{X}_5$. For
 672 instance, for \mathbf{A} , we have

$$\begin{aligned}
 673 \quad \mathbf{X}_1 &= \mathbf{P}_1^\top \mathbf{P}_1, \quad \mathbf{X}_2 = (\mathbf{C} \odot_p \mathbf{P}_2 \mathbf{B})^\top (\mathbf{C} \odot_p \mathbf{P}_2 \mathbf{B}), \quad \mathbf{X}_3 = \mathbf{I}_I, \\
 674 \quad \mathbf{X}_4 &= \lambda (\tilde{\mathbf{C}}_M \odot_p \mathbf{B})^\top (\tilde{\mathbf{C}}_M \odot_p \mathbf{B}) + \gamma \text{Diag}\{\mathbf{B}_1^\top \mathbf{B}_1, \dots, \mathbf{B}_R^\top \mathbf{B}_R\}, \\
 675 \quad \mathbf{X}_5 &= \mathbf{P}_1^\top (\mathbf{Y}_H^{(1)})^\top (\mathbf{C} \odot_p \mathbf{P}_2 \mathbf{B}) + \lambda (\mathbf{Y}_M^{(1)})^\top (\tilde{\mathbf{C}}_M \odot_p \mathbf{B}) + \gamma [\mathbf{S}_1 \mathbf{B}_1, \dots, \mathbf{S}_R \mathbf{B}_R],
 \end{aligned}$$

677 and likewise for \mathbf{B} .

678 At each iteration of the alternating directions method of multipliers scheme, we aim at
 679 solving the following equations:

$$\begin{aligned}
 680 \quad (\text{B.1}) \quad & (\gamma + \rho) \mathbf{S}_r = \gamma \mathbf{A}_r \mathbf{B}_r^\top + \rho (\mathbf{Z} + \mathbf{U}), \\
 681 \quad & \mu \mathbf{P}_3^\top \mathbf{P}_3 \mathbf{C} + \mathbf{C} (\mathbf{S}^\top (\mathbf{P}_2^\top \mathbf{P}_2 \boxtimes \mathbf{P}_1^\top \mathbf{P}_1) \mathbf{S} + \rho \mathbf{I}_R) \\
 682 \quad (\text{B.2}) \quad & = (\mathbf{Y}_H^{(3)})^\top (\mathbf{P}_2 \boxtimes \mathbf{P}_1) \mathbf{S} + \mu \mathbf{P}_3^\top (\psi_M - \tilde{\mathbf{C}}_M) + \rho (\mathbf{Z} + \mathbf{U}), \\
 683 \quad (\text{B.3}) \quad & \tilde{\mathbf{C}}_M (\mathbf{S}^\top \mathbf{S} + (\mu + \rho) \mathbf{I}_R) = (\mathbf{Y}_M^{(3)})^\top \mathbf{S} + \mu (\mathbf{P}_3 \mathbf{C} + \psi) + \rho (\mathbf{Z} + \mathbf{U}).
 \end{aligned}$$

685 For each equation, \mathbf{Z} is the projection of the considered variable onto the space of non-negative
 686 matrices, and \mathbf{U} denotes the dual variable for each subproblem [9]. The scalar ρ controls the
 687 convergence speed of the algorithm and is chosen according to [23].

688 Below, we present the framework for solving (B.2): the updates for \mathbf{S} and $\tilde{\mathbf{C}}_M$ can be
 689 handled in a similar fashion.

690 Here, the operator $[\cdot]_+$ zeroes out the negative values of the operand.

691 Appendix C. Spatial degradation matrices.

692 Here, we explain in details how the degradation matrices are constructed. For this ap-
 693 pendix, we consider that $\mathbf{P}_1 = \mathbf{P}_2$. As in previous works, \mathbf{P}_1 is constructed as $\mathbf{P}_1 = \mathbf{S}_1 \mathbf{T}_1$,
 694 where \mathbf{T}_1 is a blurring matrix and \mathbf{S}_1 is a downsampling matrix.

695 The blurring matrix is constructed from a Gaussian blurring kernel $\phi \in \mathbb{R}^{q \times 1}$ (in our case,

Algorithm B.1 Inner update for (B.2)**Input:** $\mathcal{Y}_H, \mathcal{Y}_M, \mathcal{S}, \mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3; \rho, \mu, R, \text{iter}$ **Output:** $\mathbf{Z} \in \mathbb{R}_+^{K \times R}$ **Initialization:** $\mathbf{Z} = \mathbf{U} = \mathbf{0}_{K \times M}$.**for** $m \in \{1, \dots, \text{iter}\}$ **do** $\mathbf{C} \leftarrow$ Solve (B.2) using fast solvers, $\mathbf{Z} \leftarrow [\mathbf{C} - \mathbf{U}]_+$, $\mathbf{U} \leftarrow \mathbf{U} + \mathbf{Z} - \mathbf{C}$.**end for****return** \mathbf{Z} .

696 $q = 9)$ with a standard deviation $\sigma = \frac{q\sqrt{2\log 2}}{4}$. For $m \in \{1, \dots, q\}$ and $m' = m - \lceil \frac{q}{2} \rceil$, we have

697
$$\phi(m) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-m'^2}{2\sigma^2}\right).$$

698 Thus, $\mathbf{T}_1 \in \mathbb{R}^{I \times I}$ can be expressed as

699
$$\mathbf{T}_1 = \begin{bmatrix} \phi(\lceil \frac{q}{2} \rceil) & \dots & \phi(q) & 0 & \dots & 0 \\ \vdots & \ddots & & & \ddots & \vdots \\ \phi(1) & & & & & 0 \\ 0 & \ddots & & & & \phi(q) \\ \vdots & \ddots & \ddots & & \ddots & \vdots \\ 0 & \dots & 0 & \phi(1) & \dots & \phi(\lceil \frac{q}{2} \rceil) \end{bmatrix}.$$

700 The downsampling matrix $\mathbf{S}_1 \in \mathbb{R}^{I_H \times I}$, with downsampling ratio d , is made of I_H inde-
 701 pendent rows such that for $i \in \{1, \dots, I_H\}$, $(\mathbf{S}_1)_{i, 2+(i-1)d} = 1$ and the other coefficients are
 702 zeros.

703

REFERENCES

- 704 [1] B. AIAZZI, L. ALPARONE, S. BARONTI, A. GARZELLI, AND M. SELVA, *MTF-tailored multiscale fusion of*
 705 *high-resolution ms and pan imagery*, Photogrammetric Eng. and Remote Sens., 72 (2006), pp. 591–
 706 596.
 707 [2] J. M. BIOCAS-DIAS AND J. P. NASCIMENTO, *Hyperspectral subspace identification*, IEEE Trans. Geosci.
 708 Remote Sens., 46 (2008), pp. 2435–2445.
 709 [3] J. M. BIOCAS-DIAS, A. PLAZA, N. DOBIGEON, M. PARENTE, Q. DU, P. GADER, AND J. CHANUSSOT,
 710 *Hyperspectral unmixing overview: geometrical, statistical, and sparse regression-based approaches*,
 711 IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., 5 (2012), pp. 354–379.
 712 [4] R. A. BORSOI, T. IMBIRIBA, AND J. M. BERMUDEZ, *Super-resolution for hyperspectral and multispectral*
 713 *image fusion accounting for seasonal spectral variability*, IEEE Trans. Image Process., 29 (2020),
 714 pp. 116–127.
 715 [5] R. A. BORSOI, T. IMBIRIBA, J. M. BERMUDEZ, C. RICHARD, J. CHANUSSOT, L. DRUMETZ, J.-Y.
 716 TOURNERET, AND C. ZARE, A. .AND JUTTEN, *Spectral variability in hyperspectral data unmixing: A*
 717 *comprehensive review*, arXiv preprint arXiv:2001.07307, (2020).
 718 [6] R. A. BORSOI, C. PRÉVOST, K. USEVICH, D. BRIE, J. M. BERMUDEZ, AND C. RICHARD, *Coupled tensor*
 719 *decomposition for hyperspectral and multispectral image fusion with inter-image variability*, IEEE J.
 720 Sel. Topics Signal Process., (2021).

- 721 [7] M. BOUSSE, O. DEBALS, AND L. DE LATHAUWER, *A tensor-based method for large-scale blind source*
722 *separation using segmentation*, IEEE Trans. Signal Process., 65 (2016), pp. 346–358.
- 723 [8] G. BOUTRY, M. ELAD, G. H. GOLUB, AND P. MILANFAR, *The generalized eigenvalue problem for non-*
724 *square pencils using a minimal perturbation approach*, SIAM J. Matrix Anal. Appl., 27 (2005), pp. 582–
725 601.
- 726 [9] S. BOYD, N. PARIKH, AND E. CHU, *Distributed optimization and statistical learning via the alternating*
727 *direction method of multipliers*, Now Publishers Inc, 2011.
- 728 [10] P. COMON, *Tensors: A brief introduction*, IEEE Signal Process. Mag., 31 (2014), pp. 44–53.
- 729 [11] L. DE LATHAUWER, *Decompositions of a higher-order tensor in block terms—part I: Lemmas for parti-*
730 *tioned matrices*, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 1022–1032.
- 731 [12] L. DE LATHAUWER, *Decompositions of a Higher-Order Tensor in Block Terms—Part II: Definitions and*
732 *Uniqueness*, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 1033–1066.
- 733 [13] M. DING, X. FU, T.-Z. HUANG, J. WANG, AND X.-L. ZHAO, *Hyperspectral super-resolution via inter-*
734 *pretable block-term tensor modeling*, arXiv e-prints, (2020), arXiv:2006.10248, p. arXiv:2006.10248,
735 <https://arxiv.org/abs/2006.10248>.
- 736 [14] D. DONOHO AND V. STODDEN, *When does non-negative matrix factorization give a correct decomposition*
737 *into parts?*, in Advances in neural information processing systems, 2004, pp. 1141–1148.
- 738 [15] A. ECKARDT, J. HORACK, F. LEHMANN, D. KRUTZ, J. DRESCHER, M. WHORTON, AND M. SOUTULLO,
739 *DESIS (DLR earth sensing imaging spectrometer for the ISS-muses platform)*, in 2015 IEEE IGARSS,
740 2015, pp. 1457–1459.
- 741 [16] M. ELAD, P. MILANFAR, AND G. H. GOLUB, *Shape from moments—an estimation theory perspective*,
742 IEEE Trans. Signal Process., 52 (2004), pp. 1814–1829.
- 743 [17] I. V. EMELYANOVA, T. R. McVICAR, T. G. VAN NIEL, L. T. LI, AND A. M. VAN DIJK, *Assessing the*
744 *accuracy of blending Landsat–MODIS surface reflectances in two landscapes with contrasting spatial*
745 *and temporal dynamics: A framework for algorithm selection*, Remote Sensing of Environment, 133
746 (2013), pp. 193–209.
- 747 [18] X. FU, K. HUANG, AND N. D. SIDIROPOULOS, *On identifiability of nonnegative matrix factorization*,
748 IEEE Signal Process. Letters, 25 (2018), pp. 328–332.
- 749 [19] X. FU, W.-K. MA, T.-H. CHAN, AND J. M. BIOUCAS-DIAS, *Self-dictionary sparse regression for hy-*
750 *perspectral unmixing: Greedy pursuit and pure pixel search are related*, IEEE J. Sel. Topics Signal
751 Process., 9 (2015), pp. 1128–1141.
- 752 [20] X. FU, W.-K. MA, K. HUANG, AND N. D. SIDIROPOULOS, *Blind separation of quasi-stationary sources:*
753 *Exploiting convex geometry in covariance domain*, IEEE Trans. Signal Process., 63 (2015), pp. 2306–
754 2320.
- 755 [21] N. GILLIS AND F. GLINEUR, *Accelerated multiplicative updates and hierarchical als algorithms for non-*
756 *negative matrix factorization*, Neural computation, 24 (2012), pp. 1085–1105.
- 757 [22] T. HILKER, M. A. WULDER, N. C. COOPS, J. LINKE, G. McDERMID, J. G. MASEK, F. GAO, AND
758 J. C. WHITE, *A new data fusion model for high spatial-and temporal-resolution mapping of forest*
759 *disturbance based on landsat and MODIS*, Remote Sensing of Environment, 113 (2009), pp. 1613–
760 1627.
- 761 [23] K. HUANG, N. D. SIDIROPOULOS, AND A. P. LIAVAS, *A flexible and efficient algorithmic framework for*
762 *constrained matrix and tensor factorization*, IEEE Trans. Signal Process., 64 (2016), pp. 5052–5065.
- 763 [24] K. HUANG, N. D. SIDIROPOULOS, AND A. SWAMI, *Non-negative matrix factorization revisited: Uniqueness*
764 *and algorithm for symmetric decomposition*, IEEE Trans. Signal Process., 62 (2013), pp. 211–224.
- 765 [25] T. IMBIRIBA, R. A. BORSOI, AND J. M. BERMUDEZ, *Generalized linear mixing model accounting for*
766 *endmember variability*, in 2018 IEEE ICASSP, 2018, pp. 1862–1866.
- 767 [26] M.-D. IORDACHE, J. M. BIOUCAS-DIAS, AND A. PLAZA, *Sparse unmixing of hyperspectral data*, IEEE
768 Trans. on Geosci. Remote Sens., 49 (2011), pp. 2014–2039.
- 769 [27] C. I. KANATSOULIS, X. FU, N. D. SIDIROPOULOS, AND W.-K. MA, *Hyperspectral Super-Resolution: A*
770 *Coupled Tensor Factorization Approach*, IEEE Trans. Signal Process., 66 (2018), pp. 6503–6517.
- 771 [28] C. I. KANATSOULIS, X. FU, N. D. SIDIROPOULOS, AND W.-K. MA, *Hyperspectral Super-Resolution:*
772 *Combining Low Rank Tensor and Matrix Structure*, in 2018 IEEE ICIP, Oct. 2018, pp. 3318–3322,
773 <https://doi.org/10.1109/ICIP.2018.8451733>.
- 774 [29] H. KAUFMANN, K. SEGL, S. CHABRILLAT, S. HOFER, T. STUFFLER, A. MUELLER, R. RICHTER,
775 G. SCHREIER, R. HAYDN, AND H. BACH, *EnMAP a hyperspectral sensor for environmental map-*
776 *ping and analysis*, in 2006 IEEE IGARSS, 2006, pp. 1617–1619.
- 777 [30] N. KESHAVA AND J. F. MUSTARD, *Spectral unmixing*, IEEE Signal Processing Magazine, 19 (2002),
778 pp. 44–57.

- 779 [31] T. G. KOLDA AND B. W. BADER, *Tensor Decompositions and Applications*, SIAM Review, 51 (2009),
780 pp. 455–500.
- 781 [32] H. LAURBERG, M. G. CHRISTENSEN, M. D. PLUMBLEY, L. K. HANSEN, AND S. H. JENSEN, *Theorems on*
782 *positive data: On the uniqueness of NMF*, Computational intelligence and neuroscience, 2008 (2008).
- 783 [33] Q. LI, W.-K. MA, AND Q. WU, *Hyperspectral super-resolution: Exact recovery in polynomial time*, in
784 2018 IEEE SSP, IEEE, 2018, pp. 378–382.
- 785 [34] H. LIU, R. WU, AND W.-K. MA, *Is there any recovery guarantee with coupled structured matrix factor-*
786 *ization for hyperspectral super-resolution?*, in 2019 IEEE CAMSAP, IEEE, 2019, pp. 480–484.
- 787 [35] J. NASCIMENTO AND J. DIAS, *Vertex component analysis: A fast algorithm to unmix hyperspectral data*,
788 IEEE Geosci. Remote Sens. Lett., 43 (2005), pp. 898–910.
- 789 [36] L. NUS, *Méthodes rapides de traitement d’images hyperspectrales. Application à la caractérisation en*
790 *temps réel du matériau bois*, PhD Thesis, University of Lorraine, France, 2019.
- 791 [37] M. PARENTE AND A. PLAZA, *Survey of geometric and statistical unmixing algorithms for hyperspectral*
792 *images*, in 2nd IEEE Workshop on Hyperspectral Image and Signal Process.: Evolution in Remote
793 Sens., 2010, pp. 1–4.
- 794 [38] C. PRÉVOST, K. USEVICH, P. COMON, AND D. BRIE, *Hyperspectral Super-Resolution with Coupled*
795 *Tucker Approximation: Identifiability and SVD-based algorithms*, IEEE Trans. Signal Process., 68
796 (2020), pp. 931–946.
- 797 [39] Y. QIAN, F. XIONG, S. ZENG, J. ZHOU, AND Y. TANG, *Matrix-vector nonnegative tensor factorization for*
798 *blind unmixing of hyperspectral imagery*, IEEE Trans. Geosci. Remote Sens., 55 (2016), pp. 1776–1792.
- 799 [40] R. E. ROGER AND J. F. ARNOLD, *Reliably estimating the noise in AVIRIS hyperspectral images*, Inter-
800 national Journal of Remote Sensing, 17 (1996), pp. 1951–1962.
- 801 [41] G. A. SHAW AND H. K. BURKE, *Spectral imaging for remote sensing*, Lincoln laboratory journal, 14
802 (2003), pp. 3–28.
- 803 [42] M. SIMOES, J. M. BIOUCAS-DIAS, L. B. ALMEIDA, AND J. CHANUSSOT, *A convex formulation for*
804 *hyperspectral image superresolution via subspace-based regularization*, IEEE Trans. Geosci. Remote
805 Sens., 53 (2015), pp. 3373–3388.
- 806 [43] V. SIMONCINI, *Computational methods for linear matrix equations*, SIAM Review, 58 (2016), pp. 377–441.
- 807 [44] B. SOMERS, G. P. ASNER, L. TITS, AND P. COPPIN, *Endmember variability in spectral mixture analysis:*
808 *A review*, Remote Sensing of Environment, 115 (2011), pp. 1603–1616.
- 809 [45] M. SØRENSEN AND L. DE LATHAUWER, *Fiber sampling approach to canonical polyadic decomposition and*
810 *application to tensor completion*, SIAM J. Matrix Anal. Appl., 40 (2019), pp. 888–917.
- 811 [46] N. VERVLIET, O. DEBALS, L. SORBER, M. V. BAREL, AND L. D. LATHAUWER, *Tensorlab 3.0*, Available
812 online, 2016.
- 813 [47] L. WALD, T. RANCHIN, AND M. MANGOLINI, *Fusion of satellite images of different spatial resolutions: As-*
814 *sessing the quality of resulting images*, Photogrammetric Eng. and Remote Sens., 63 (1997), pp. 691–
815 699.
- 816 [48] Q. WEI, J. M. BIOUCAS-DIAS, N. DOBIGEON, AND J.-Y. TOURNERET, *Multiband image fusion based on*
817 *spectral unmixing*, IEEE Trans. Geosci. Remote Sens., 54 (2016), pp. 7236–7249.
- 818 [49] Q. WEI, N. DOBIGEON, AND J.-Y. TOURNERET, *Fast fusion of multi-band images based on solving a*
819 *Sylvester equation*, IEEE Trans. Image Process., 24 (2015), pp. 4109–4121.
- 820 [50] N. YOKOYA, C. GROHNFELDT, AND J. CHANUSSOT, *Hyperspectral and multispectral data fusion: A*
821 *comparative review of the recent literature*, IEEE Trans. Geosci. Remote Sens., 5 (2017), pp. 29–56.
- 822 [51] N. YOKOYA, T. YAIRI, AND A. IWASAKI, *Coupled Nonnegative Matrix Factorization Unmixing for Hyper-*
823 *spectral and Multispectral Data Fusion*, IEEE Trans. Geosci. Remote Sens., 50 (2012), pp. 528–537,
824 <https://doi.org/10.1109/TGRS.2011.2161320>.
- 825 [52] A. ZARE AND K. C. HO, *Endmember variability in hyperspectral analysis: Addressing spectral variability*
826 *during spectral unmixing*, IEEE Signal Process. Mag., 31 (2013), pp. 95–104.
- 827 [53] G. ZHANG, X. FU, K. HUANG, AND J. WANG, *Hyperspectral super-resolution: A coupled nonnegative*
828 *block-term tensor decomposition approach*, in 2019 IEEE CAMSAP, 2019. Guadeloupe, West Indies.
- 829 [54] G. ZHANG, X. FU, J. WANG, X.-L. ZHAO, AND M. HONG, *Spectrum cartography via coupled block-term*
830 *tensor decomposition*, IEEE Trans. Signal Process., (2020).