

Article

Hyperspectral Image Restoration under Complex Multi-Band Noises

Zongsheng Yue ¹, Deyu Meng ¹, Yongqing Sun ² and Qian Zhao ^{1,*}¹ Institute for Information and System Sciences and Ministry of Education Key Lab of Intelligent Networks and Network Security, Xi'an Jiaotong University, Xi'an 710049, China; zsy20110806207@stu.xjtu.edu.cn (Z.Y.); dymeng@xjtu.edu.cn (D.M.)² NTT Cyber Space Labs, Kanagawa 2360026, Japan; yongqing.sun@lab.ntt.co.jp

* Correspondence: timmy.zhaoqian@gmail.com; Tel.: +86-138-9187-3702

Received: 31 August 2018; Accepted: 6 October 2018; Published: 14 October 2018



Abstract: Hyperspectral images (HSIs) are always corrupted by complicated forms of noise during the acquisition process, such as Gaussian noise, impulse noise, stripes, deadlines and so on. Specifically, different bands of the practical HSIs generally contain different noises of evidently distinct type and extent. While current HSI restoration methods give less consideration to such band-noise-distinctness issues, this study elaborately constructs a new HSI restoration technique, aimed at more faithfully and comprehensively taking such noise characteristics into account. Particularly, through a two-level hierarchical Dirichlet process (HDP) to model the HSI noise structure, the noise of each band is depicted by a Dirichlet process Gaussian mixture model (DP-GMM), in which its complexity can be flexibly adapted in an automatic manner. Besides, the DP-GMM of each band comes from a higher level DP-GMM that relates the noise of different bands. The variational Bayes algorithm is also designed to solve this model, and closed-form updating equations for all involved parameters are deduced. The experiment indicates that, in terms of the mean peak signal-to-noise ratio (MPSNR), the proposed method is on average 1 dB higher compared with the existing state-of-the-art methods, as well as performing better in terms of the mean structural similarity index (MSSIM) and Erreur Relative Globale Adimensionnelle de Synthèse (ERGAS).

Keywords: Hyperspectral images; restoration; hierarchical Dirichlet process; variational Bayes

1. Introduction

Hyperspectral images (HSIs) are collected by high spectral resolution sensors, and consist of hundreds of bands ranging from ultraviolet to infrared wavelengths. Due to the richness of spatial and spectral information, this type of image has been widely used in many remote sensing applications [1,2]. However, HSIs inevitably suffer from various noise contamination in practice [3], which severely influences the image quality and limits the performance of the subsequent HSI processing tasks, such as unmixing [4], classification [5] and so on. Therefore, as a necessary pre-processing step of certain application tasks, HSI restoration is an important research topic and has attracted much attention in recent decades.

Regarding each band of HSIs as a gray image, we can simply use the traditional image denoising method to reduce noise in a band-by-band fashion, such as the sparse constraint methods [6–9], local polynomial approximation (LPA) methods [10–12] and so on. However, such methods ignore the intrinsic characteristics of HSIs, i.e., the spatial and spectral correlation among different bands, thus often resulting in unsatisfactory performance. To alleviate this issue, many methods have been proposed in order to better encode the spatial and spectral correlation knowledge of HSIs, e.g., [13–16]. Besides, nonlocal similarity [17], anisotropic diffusion [18] and conditional random fields [19], have also been utilized for the HSI restoration task recently.

Inspired by the spectral correlation of HSIs, various methods based on low-rank matrix factorization (LRMF) have been proposed in the past few years. For example, Zhang et al. [20] used a rank-constrained RPCA [21] while Wu et al. [22] and Xie et al. [23] employed weighted nuclear norm minimization (WNNM) [24,25] to enhance the restoration quality. In addition, considering the local similarity within a patch and the non-local similarity across the patches in a group simultaneously, a novel group low-rank representation (GLRR) model was proposed in [26]. Besides this kind of LRMF approach, a series of methods regarding HSIs as a 3-D tensor have also been proposed in recent years. Along this line, we can categorize these methods as tensor decomposition-based methods [27–31] and wavlet-based methods [32–35].

The current HSI restoration methods generally assume similar complexity of noise structures across all bands of an HSI dataset. In fact, most of them only assume simple independent and identical Gaussian/Laplacian noise on all HSIs, which implies that the noises on all bands share a similar noise type and extent. To make the approach better fit real noises with more complexity, some HSI restoration methods have been constructed to consider more complex noise beyond simple Gaussian/Laplacian [36,37]. Additionally, He et al. [38] considered the noise as a mixture of Gaussian and Laplacian distribution while Cao et al. [39] modelled the noise using a more general mixture of exponential power (MoEP) distributions. However, these methods all assume the noise in each HSI band as an identical distribution. Recently, Chen et al. [40] proposed modelling the noise distribution of each band as a mixture of Gaussian (MoG) distributions with different parameters sharing with the unique conjugate prior. This method, however, still assumes the same complexity (i.e., the number of components) of MoGs in all HSI bands.

For most practical HSIs, however, the noise across different bands is always evidently distinct, both in terms of distribution type and corruption extent. As shown in Figure 1, the noises in the three bands of the Urban (<http://lesun.weebly.com/hyperspectral-data-set.html>) HSI dataset (Figure 1a) and the Terrain (<https://www.agc.army.mil>) HSI dataset (Figure 1b) exhibit very different patterns. Specifically, band 76 of Urban and band 2 of Terrain are mainly contaminated by deadlines and the noises resemble Laplacian distributions; band 104 of Urban and band 139 of Terrain are more corrupted by Gaussian noise, while band 139 of Urban and band 151 of Terrain are simultaneously contaminated by several kinds of noise, so their noise distributions are multimodal. The current HSI restoration methods thus lack consideration on such band-noise-distinctness issue, which tends to weaken their robustness in handling real noisy HSI.

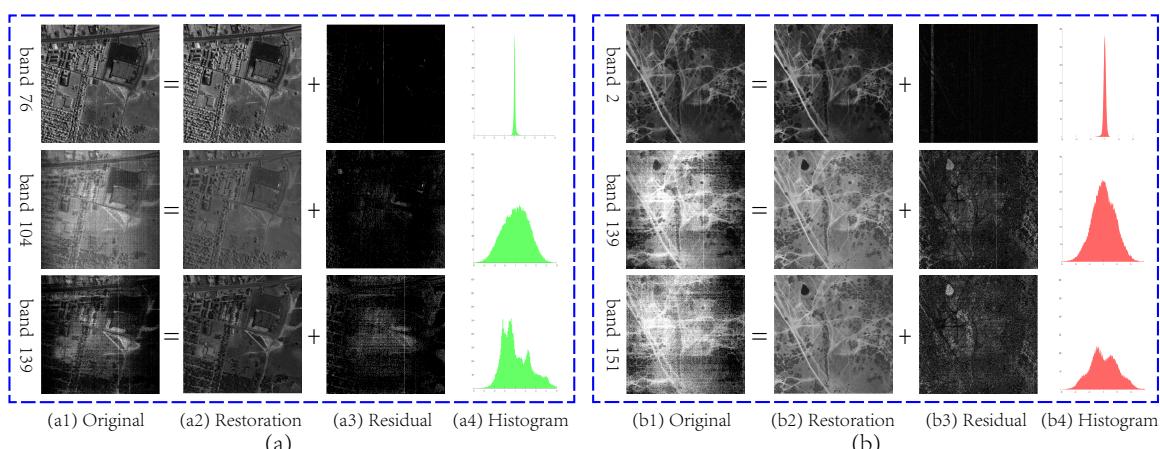


Figure 1. The distinctness of noises among different bands of hyperspectral images (HSIs). **(a)** Bands 76, 104 and 139 of the Urban HSI dataset; **(b)** Bands 2, 139 and 151 of the Terrain HSI dataset. (a1) and (b1): Original images, (a2) and (b2): restored images, (a3) and (b3): noises extracted by our proposed method, (a4) and (b4): histogram of noises in the same scale.

To address the aforementioned noise fitting issue, in this paper we propose a new HSI restoration method, which can more sufficiently encode such noise characteristics underlying HSIs. The main contribution of this work can be summarized as follows:

1. The noise in each HSI band is modelled by a Dirichlet process Gaussian mixture model (DP-GMM), in which the Gaussian components of MoG in each band are adaptively determined based on the specific noise complexity of this band. The distinctness of noise structures in different bands is thus able to be faithfully reflected by the model.
2. By using the hierarchical Dirichlet process (HDP) technique, we correlate the noise of different bands through a sharing strategy, in which the noise parameters of each band share a high-level noise configuration of the entire HSI.
3. A variational Bayes algorithm is readily designed to solve the model, and each of the involved parameters can be effectively updated in closed-form.

This paper is organized as follows. Section 2 provides some preliminaries on the utilized HDP technique. Section 3 presents the main model against the investigated problem, and the variational Bayes algorithm for solving it is given in Section 4. Experimental results on synthetic and real datasets are demonstrated in Section 5, and finally the paper is rounded up with a conclusion.

2. Preliminaries

In this section, we review some preliminary knowledge about the Dirichlet process and hierarchical Dirichlet process to set the stage for the presentation of our model later.

2.1. Dirichlet Process

The Dirichlet process (DP), developed firstly by Ferguson [41], is parameterized by a concentration parameter γ and a base distribution H . Consider a measurable set Θ and any of its finite partitions $\{A_1, A_2, \dots, A_r\}$; the Dirichlet process, denoted by $DP(\gamma, H)$, is a unique distribution over probability distributions on Θ such that

$$\text{if } G \sim DP(\gamma, H), \text{ then } (G(A_1), G(A_2), \dots, G(A_r)) \sim Dir(\gamma H(A_1), \gamma H(A_2), \dots, \gamma H(A_r)) \quad (1)$$

where $Dir(\cdot)$ denotes the Dirichlet distribution.

Besides, Ferguson [41] proved two important properties of the DP. The first is that any random sample G drawn from $DP(\gamma, H)$ is a discrete distribution over Θ . The second property considers repeated draws $\theta_1, \theta_2, \dots, \theta_n$ from this discrete distribution G , i.e.,

$$\theta_i \sim G, \quad i \in \{1, 2, \dots, n\},$$

where these θ_i s will exhibit a clustering property, i.e., they share repeated values with positive probability. In the DP-GMM, θ_i represents the parameters (i.e., mean and variance) of the corresponding Gaussian component to which the i -th data point belongs. Thus, the clustering property of θ_i s can adaptively determine the number of Gaussian components in MoG.

Sethuraman [42] provides an equivalent but more explicit representation of DP based on the stick-breaking construction. Considering two infinite collections of independent random variables, $\pi_i \sim Beta(1, \gamma)$ and $\eta_i^* \sim H$, for $i = 1, 2, \dots$, the stick-breaking representation of G is as follows:

$$\pi_i = \pi'_i \prod_{j=1}^{i-1} (1 - \pi'_j), \quad G = \sum_{i=1}^{\infty} \pi_i \delta_{\eta_i^*}$$

where $\delta_{\eta_i^*}$ represents the Dirac Delta function located at η_i^* . The stick-breaking construction depicts the DP from a different point of view, in which the clustering property of DP is realized by the mixing coefficient π_k . The parameter γ controls the dispersion level of π_k , and we simulate the stick-breaking

process under different γ in Figure 2. It is easy to see that this stick-breaking process is able to determine the number of Gaussian components of MoG through adjusting the value of π_k .

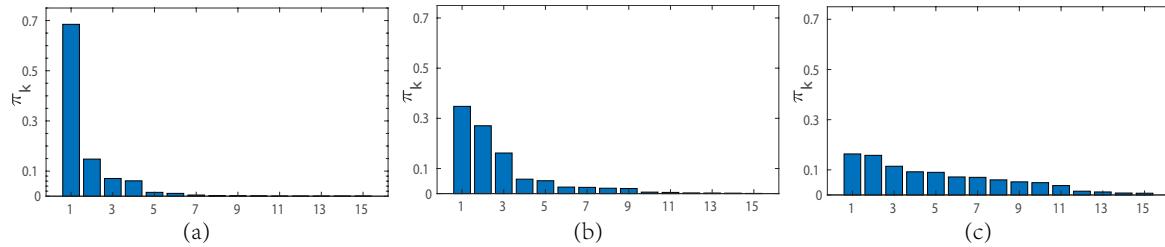


Figure 2. Stick breaking process under different concentration parameters γ . (a) $\gamma = 1$; (b) $\gamma = 5$; (c) $\gamma = 10$.

2.2. Hierarchical Dirichlet Process

The hierarchical Dirichlet process (HDP) [43] is originally proposed as a hierarchical Bayesian model for natural language processing, which provides a flexible framework for sharing mixture components among groups of related data. A two-level HDP is a collection of DPs, which is also drawn from a higher-level DP with base distribution H , i.e.,

$$G \sim DP(\gamma, H), G_j \sim DP(\alpha_j, G), \quad (2)$$

where j is an index for the data group. It should be noted that all distributions G_j possess different parameters while sharing the distribution G .

3. HSIs Restoration Model Based on DP-GMM

3.1. Notation Explanation

For convenience of formulating our model, we first introduce some necessary notations to be used in the following sections. We use light lowercase letters, bold lowercase letters and bold uppercase letters to denote scalars, vectors and matrices, respectively. Given a matrix X , we use $x_{\cdot j}$ to denote its j -th column vector, $x_{\cdot i}$ to denote its i -th row vector, and x_{ij} to denote the element in its i -th row and j -th column. For probability distributions, $\mathcal{N}(\mu, \Sigma)$ represents the multivariate Gaussian distribution with mean vector μ and covariance matrix Σ , $\mathcal{N}(\mu, \xi^{-1})$ represents the univariate Gaussian distribution with mean μ and precision ξ , $Beta(a, b)$ represents the Beta distribution with parameters a and b , $Gam(c, d)$ represents the Gamma distribution with parameters c and d , and $Multi(\pi)$ represents the multinomial distribution with parameter π .

3.2. Model Formulation

Now, we give the formulations of our proposed model. Let us present the observed HSI data contaminated by noises as $\mathbf{Y} \in \mathbb{R}^{d \times B}$, where $d = hw$, h and w are the height and width of the investigated HSI, respectively, and B is the number of bands. By considering a generative model, we can decompose the observed data into

$$\mathbf{Y} = \mathbf{X} + \mathbf{E}, \quad (3)$$

where \mathbf{E} denotes the noise term and \mathbf{X} is the expected restorated HSI.

3.2.1. The DP-GMM Model

In this study, we model the noise of each band as a DP-GMM, which is an advanced version of the non-parametric GMM model capable of adaptively rectifying the number of Gaussian components of MoG based on the data [44,45]. The noise complexity of each band can be adjusted by varying

the number of components of MoG. What is more, in order to simultaneously encode the correlation among different bands, we associate the corresponding DP-GMM of each band at a higher level and conduct a two-level HDP.

Specifically, we consider the noise located at the i -th row and j -th column of the noise term E . Following the idea of [40,46], we can model it as an element-wise Gaussian mixture distribution:

$$e_{ij} \sim \sum_{k=1}^K \pi_{jk} \mathcal{N}(0, \xi_k^{-1}), \quad (4)$$

which can be equivalently reformulated as a two-level generative process:

$$e_{ij} \sim \mathcal{N}(0, \xi_{z_{ij}}^{-1}), \quad z_{ij} \sim Multi(\boldsymbol{\pi}_j), \quad (5)$$

where $\boldsymbol{\pi}_j$ is a vector of length K , $1 \leq z_{ij} \leq K$. Then, we place a hierarchical Dirichlet distribution prior on $\boldsymbol{\pi}_j$ to associate the noise of different bands, i.e.,

$$\boldsymbol{\pi}_j \sim Dir(\alpha_j \boldsymbol{\beta}), \quad \boldsymbol{\beta} \sim Dir(\gamma/K, \gamma/K, \dots, \gamma/K). \quad (6)$$

Besides, ξ_k is also generated from its conjugate prior, i.e.,

$$\xi_k \sim Gam(a_0, b_0). \quad (7)$$

To allow the model to be flexible, we further assume $K \rightarrow +\infty$. Teh et al. [47] proved that the limit of Equations (5)–(7) is the following HDP noise model:

$$e_{ij} \sim \mathcal{N}(0, \psi_{ij}^{-1}), \quad \psi_{ij} \sim G_j, \quad G_j \sim DP(\alpha_j, G), \quad G \sim DP(\gamma, H), \quad (8)$$

where $H = Gam(a_0, b_0)$.

The noise model of Equation (8) is too metaphysical to understand, so we reformulate it equivalently but more explicitly with two coupled latent variables z_{ij} and c_{jt} in a stick-breaking constructive manner, i.e.,

$$\begin{aligned} c_{jt} &\sim Multi(\boldsymbol{\beta}), \quad \beta_k = \beta'_k \prod_{l=1}^{k-1} (1 - \beta'_l), \quad \beta'_k \sim Beta(1, \gamma), \quad \xi_k \sim Gam(a_0, b_0), \\ z_{ij} &\sim Multi(\boldsymbol{\pi}_j), \quad \pi_{jt} = \pi'_{jt} \prod_{s=1}^{t-1} (1 - \pi'_{js}), \quad \pi'_{jt} \sim Beta(1, \alpha_j), \quad e_{ij} \sim \mathcal{N}(0, \xi_{c_{z_{ij}}}^{-1}). \end{aligned} \quad (9)$$

It is worth noting that the ξ in Equation (9) and the ψ in Equation (8) have the following relation:

$$\xi_{c_{z_{ij}}} = \psi_{ij}.$$

We can explain the hierarchical structure of the noise model (8) using Figure 3a intuitively. In the first level, we attempt to fit the noise of the entire HSI data by a DP-GMM. In the second level, we also fit the noise of each band by a DP-GMM, but each of its components is shared from the first level. This sharing strategy associates the noise of different bands, and the DP-GMM, which allows an MoG to have a limited but unbounded number of Gaussian components, adjusts the noise complexity adaptivity for each band.

Note that we adopt the truncation strategy [48] for the DP-GMMs in the variational inference of Section 4, which means that the number of Gaussian components of DP-GMMs in the first and second levels are fixed as K and T (K and T is large enough, $T \leq K$) as shown in Figure 3a. Based on such a truncation strategy, z_{ij} ($1 \leq z_{ij} \leq T$) denotes the index of the Gaussian component from which e_{ij} is generated in the DP-GMM of band j ($1 \leq j \leq B$), and c_{jt} ($1 \leq c_{jt} \leq K, 1 \leq t \leq T$) denotes the index of

the Gaussian component of DP-GMM in the first level, with which the t -th Gaussian component of DP-GMM of band j in the second level is shared (c_{jt} is marked in Figure 3a).

Finally, the concentration parameters α and γ in Equation (8) mainly affect the number of Gaussian components of the second level MoG in each band and the first level MoG for the entire HSIs dataset, respectively. In order to enhance the adaptability of the model complexity, we can place the Gamma distribution on them [49], i.e.,

$$\gamma \sim \text{Gam}(c_0, d_0), \alpha_j \sim \text{Gam}(e_0, f_0). \quad (10)$$

In the above formulations, $a_0, b_0, c_0, d_0, e_0, f_0$ are hyperparameters of prior distributions, all of which can be easily set as small values to make the prior non-informative.

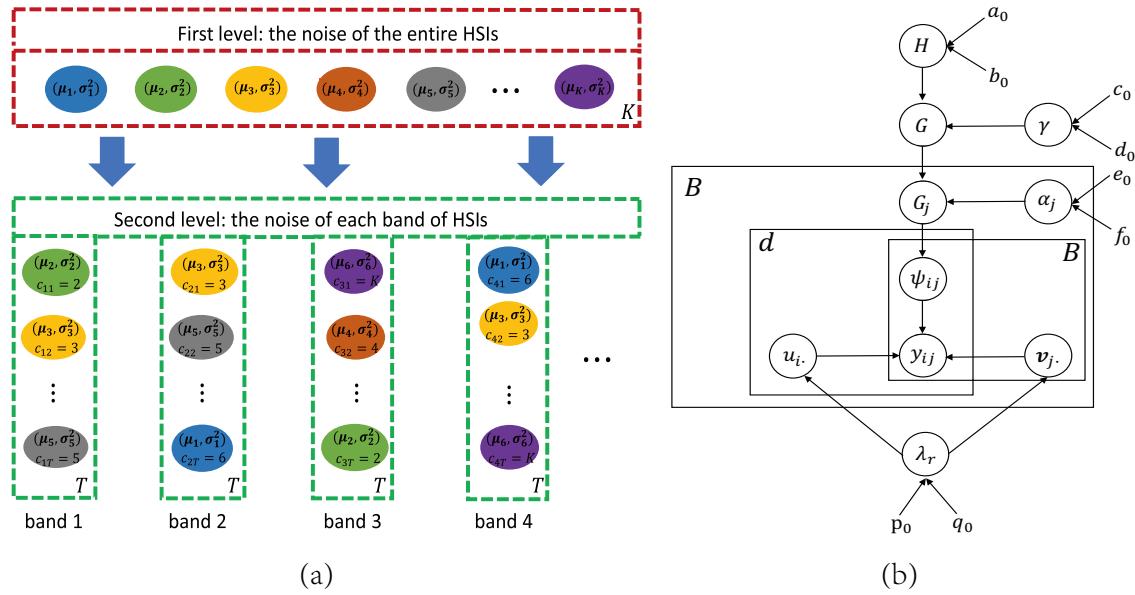


Figure 3. (a): Illustration of the hierarchical DP-GMM model for HSI restoration. The ellipses with different colors denote the Gaussian component with parameters (μ_k, σ_k^2) . The green dotted rectangles, representing the Gaussian compounds of each band, share Gaussian components from a top-level MoG in the red dotted rectangle. The number of Gaussian components of MoG at the two levels are dynamically adjusted during the HSI restoration until it converges. (b): The entire graphic model of DP-GMM.

3.2.2. LRMF Model

As the conventional HSI restoration model, we assume that the expected HSI X in Equation (3) has a low-rank property. Based on such an assumption, we formulate $X \in \mathbb{R}^{d \times B}$ as the product of two smaller matrices $U \in \mathbb{R}^{d \times l}$ and $V \in \mathbb{R}^{B \times l}$, i.e.,

$$X = UV^T = \sum_{r=1}^l u_{\cdot r} v_{\cdot r}^T, \quad (11)$$

where $l \leq B$ is fixed pre-manually. Furthermore, U and V can be generated from a multivariate zero-mean Gaussian distribution by the following method:

$$u_{\cdot r} \sim \mathcal{N}(\mathbf{0}, \lambda_r^{-1} I_d), \quad v_{\cdot r} \sim \mathcal{N}(\mathbf{0}, \lambda_r^{-1} I_B), \quad (12)$$

where I_d denotes the $d \times d$ identity matrix. Furthermore, the corresponding conjugate prior on each precision variable λ_r is:

$$\lambda_r \sim \text{Gam}(p_0, q_0). \quad (13)$$

3.2.3. The Entire Graphical Model

Combining Equations (3), (9)–(13), we can construct a full Bayesian model for HSI restoration. In brief, we call our model DP-GMM. The corresponding graphical model of our method is shown in Figure 3b. The goal turns to infer the posteriors of all involved variables:

$$p(\mathbf{U}, \mathbf{V}, \boldsymbol{\xi}, \mathbf{C}, \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\pi}, \boldsymbol{\lambda}, \boldsymbol{\alpha}, \boldsymbol{\gamma} | \mathbf{Y}). \quad (14)$$

4. Variational Inference

Directly computing the posterior of Equation (14) under an HDP mixture prior is intractable, and thus approximate inference methods need to be designed. Although the Markov chain Monte Carlo (MCMC) sampling method can provide very accurate approximation to the posterior [50], this method is limited for massive scale data due to its high computational cost and difficulty to detect convergence. As a more efficient and deterministic alternative to MCMC, we adopt a Variational Bayes (VB) method [51] to solve our model in this paper. Given the observed data \mathbf{Y} , maximizing the marginal log-likelihood $p(\mathbf{Y}) = \int p(\mathbf{Y}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}$ directly is computationally intractable. VB attempts to seek a lower bound of the log-likelihood and maximize it instead, i.e.,

$$\begin{aligned} \ln p(\mathbf{Y}) &= \int q(\boldsymbol{\theta}) \ln \frac{p(\mathbf{Y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{q(\boldsymbol{\theta})} d\boldsymbol{\theta} + KL(q(\boldsymbol{\theta})||p(\boldsymbol{\theta}|\mathbf{Y})) \\ &\geq \int q(\boldsymbol{\theta}) \ln \frac{p(\mathbf{Y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{q(\boldsymbol{\theta})} d\boldsymbol{\theta} \\ &\triangleq \mathcal{L}(q) \end{aligned} \quad (15)$$

where $KL(\cdot||\cdot)$ denotes the *Kullback-Leibler* (KL) divergence of two distributions. Assuming $q(\boldsymbol{\theta}) = \prod_i q_i(\theta_i)$, we can obtain the closed-form solution to $q_j(\theta_j)$ with other factors fixed through optimizing $\mathcal{L}(q)$ with respect to $q_j(\theta_j)$, i.e.,

$$q_j^*(\theta_j) = \frac{\exp \left\{ E_{\boldsymbol{\theta} \setminus \theta_j} [\ln p(\boldsymbol{\theta}, \mathbf{Y})] \right\}}{\int \exp \left\{ E_{\boldsymbol{\theta} \setminus \theta_j} [\ln p(\boldsymbol{\theta}, \mathbf{Y})] \right\} d\theta_j}, \quad (16)$$

where $E_x[\cdot]$ denotes the expectation over the variable x , and $\boldsymbol{\theta} \setminus \theta_j$ represents the set of $\boldsymbol{\theta}$ with θ_j removed. The entire inference can be accomplished by alternatively computing Equation (16) with respect to all the involved variables.

Based on the stick-breaking construction and the truncation strategy [48], we can approximate the posterior distribution of Equation (14) with the following factorized form:

$$q(\mathbf{U}, \mathbf{V}, \boldsymbol{\xi}, \mathbf{C}, \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\pi}, \boldsymbol{\lambda}, \boldsymbol{\alpha}, \boldsymbol{\gamma}) = \prod_{i=1}^d q(u_i) \prod_{j=1}^B q(v_j) \prod_{r=1}^l q(\lambda_r) \prod_{k=1}^K q(\xi_k) q(\beta'_k) \prod_{j,t}^{B,T} q(c_{jt}) q(\pi'_{jt}) \prod_{i,j}^{d,B} q(z_{ij}) \prod_{j=1}^B q(\alpha_j) q(\gamma) \quad (17)$$

Then we can analytically infer all the factorized distributions involved in Equation (17) as below. The computational details are provided in the Supplementary Material.

Update Noise Component: The parameters involved in the noise components are $\boldsymbol{\xi}$, \mathbf{C} , \mathbf{Z} , $\boldsymbol{\pi}$, $\boldsymbol{\beta}$. We use the stick-breaking procedure by setting a large enough value of K and T for truncated approximation, and then get the following updating equations on \mathbf{Z} and \mathbf{C} :

$$q(c_{jt}) = Multi(\boldsymbol{\varphi}_{jt}), \quad q(z_{ij}) = Multi(\boldsymbol{\rho}_{ij}), \quad (18)$$

where

$$\begin{aligned}\varphi_{jtk} &\propto \exp \left\{ \sum_{i=1}^d \rho_{ijt} \left(\frac{1}{2} E[\ln \xi_k] - \frac{1}{2} \ln 2\pi - \frac{1}{2} E[\xi_k] E[(y_{ij} - \mathbf{u}_i \cdot \mathbf{v}_{j \cdot}^T)^2] \right) + E[\ln \beta_k] \right\}, \\ \rho_{ijt} &\propto \exp \left\{ \sum_{k=1}^K \varphi_{itk} \left(\frac{1}{2} E[\ln \xi_k] - \frac{1}{2} \ln 2\pi - \frac{1}{2} E[\xi_k] E[(y_{ij} - \mathbf{u}_i \cdot \mathbf{v}_{j \cdot}^T)^2] \right) + E[\ln \pi_{jt}] \right\}, \\ E[\ln \beta_k] &= E[\ln \beta'_k] + \sum_{l=1}^{k-1} E[\ln(1 - \beta'_l)], \quad E[\ln \pi_{jt}] = E[\ln \pi'_{jt}] + \sum_{s=1}^{t-1} E[\ln(1 - \pi'_{js})].\end{aligned}$$

As for ξ , based on its conjugate priors, we have

$$q(\xi_k) = \text{Gam}(a_k, b_k), \quad (19)$$

where

$$a_k = \frac{1}{2} \sum_{i,j,t} \rho_{ijt} \varphi_{jtk} + a_0, \quad b_k = \frac{1}{2} \sum_{i,j,t} \rho_{ijt} \varphi_{jtk} E[(y_{ij} - \mathbf{u}_i \cdot \mathbf{v}_{j \cdot}^T)^2] + b_0.$$

Similarly, for π , β , we have

$$q(\pi'_{jt}) = \text{Beta}(r_{jt}^1, r_{jt}^2), \quad q(\beta'_k) = \text{Beta}(s_k^1, s_k^2), \quad (20)$$

where

$$r_{jt}^1 = \sum_i \rho_{ijt} + 1, \quad r_{jt}^2 = \sum_{i,s=t+1}^{d,T} \rho_{ijs} + \alpha_j, \quad s_k^1 = \sum_{j,t} \varphi_{jtk} + 1, \quad s_k^2 = \sum_{j,t,s=k+1}^{d,T,K} \varphi_{jts} + \gamma.$$

Update the Concentration Parameters: According to the conjugate priors of α and γ in Equation (10), we can obtain the following updating equations:

$$q(\alpha_j) = \text{Gam}(e_j, f_j), \quad q(\gamma) = \text{Gam}(c, d), \quad (21)$$

where

$$e_j = T + e_0, \quad f_j = f_0 - \sum_t E[\ln(1 - \pi'_{jt})], \quad c = K + c_0, \quad d = d_0 - \sum_k E[\ln(1 - \beta'_k)].$$

Update the LRMF Components: The parameters involved in the LRMF components are \mathbf{U} , \mathbf{V} , λ . For each row of \mathbf{U} , using the factorization in Equation (11), we can obtain

$$q(\mathbf{u}_{i \cdot}) = \mathcal{N}(\boldsymbol{\mu}_{i \cdot}^u, \boldsymbol{\Sigma}_{i \cdot}^u), \quad (22)$$

with mean $\boldsymbol{\mu}_{i \cdot}^u$ and covariance matrix $\boldsymbol{\Sigma}_{i \cdot}^u$, given by

$$\boldsymbol{\Sigma}_{i \cdot}^u = \left(\sum_{j,t,k} \rho_{ijt} \varphi_{jtk} E[\xi_k] E[\mathbf{v}_{j \cdot}^T \mathbf{v}_{j \cdot}] + \boldsymbol{\Lambda} \right)^{-1}, \quad \boldsymbol{\mu}_{i \cdot}^u = \sum_{j,t,k} \rho_{ijt} \varphi_{jtk} E[\xi_k] y_{ij} \mathbf{v}_{j \cdot} \boldsymbol{\Sigma}_{i \cdot}^u,$$

where $\boldsymbol{\Lambda} = \text{diag}(E[\lambda_1], E[\lambda_2], \dots, E[\lambda_l])$. Similarly, for each column of \mathbf{V} , we have

$$q(\mathbf{v}_{j \cdot}) = \mathcal{N}(\boldsymbol{\mu}_{j \cdot}^v, \boldsymbol{\Sigma}_{j \cdot}^v), \quad (23)$$

and the mean $\boldsymbol{\mu}_{j \cdot}^v$ and the covariance matrix $\boldsymbol{\Sigma}_{j \cdot}^v$ are calculated by

$$\boldsymbol{\Sigma}_{j \cdot}^v = \left(\sum_{i,t,k} \rho_{ijt} \varphi_{jtk} E[\xi_k] E[\mathbf{u}_{i \cdot}^T \mathbf{u}_{i \cdot}] + \boldsymbol{\Lambda} \right)^{-1}, \quad \boldsymbol{\mu}_{j \cdot}^v = \sum_{i,t,k} \rho_{ijt} \varphi_{jtk} E[\xi_k] y_{ij} \mathbf{u}_{i \cdot} \boldsymbol{\Sigma}_{j \cdot}^v.$$

For parameter λ , we have

$$q(\lambda_r) = \text{Gam}(p_r, q_r), \quad (24)$$

where

$$p_r = \frac{d}{2} + \frac{B}{2} + p_0, \quad q_r = \frac{1}{2} E[\mathbf{u}_{\cdot r}^T \mathbf{u}_{\cdot r}] + \frac{1}{2} E[\mathbf{v}_{\cdot r}^T \mathbf{v}_{\cdot r}] + q_0.$$

Settings of Hyperparameters: We set all the hyperparameters involved in our model in a non-informative manner to minimize their impacts on the inference of the posterior distribution [51]. Specifically, throughout our experiments, we easily set $a_0, b_0, c_0, d_0, e_0, f_0, p_0, q_0$ as 10^{-6} . Our proposed method behaves stably throughout all the experiments with this simple initialization choice.

5. Experimental Results

To demonstrate the effectiveness of the proposed DP-GMM model, we perform both synthetic and real HSI data experiments, and evaluate the experimental results quantitatively and visually. In order to comprehensively evaluate the performance of the proposed method, we compared the results with eight popular HSI restoration methods: the traditional image denoising method SVD [52], LRMF-based methods LRMR [20], LRTV [38], WNNM [24] and WSNM [23], tensor-based methods TDL [30] and BM4D [35], and mixture noise based-method NMoG [33]. These methods represent the state-of-the-art for the HSI restoration task. The parameters for these compared methods were manually adjusted according to their default strategies. In addition, to facilitate the numerical calculation and visualization, all the bands of the HSI datasets are normalized into [0,1].

5.1. Simulated Data Experiments

Two HSI datasets were employed in this experiment: Washington DC Mall (<http://lesun.weebly.com/hyperspectral-data-set.html>) with size $1208 \times 307 \times 191$ and RemoteImage (http://peterwonka.net/Publications/code/LRTC_Package_Ji.zip) with size $205 \times 246 \times 96$. Some bands of these two HSI datasets are heavily contaminated by noise and cannot be regarded as clean ground truth, and thus have to be deleted [33,36]. Due to the computer's memory limitation, we also cropped the subimage of the HSI and finally obtained two synthetic datasets with size $200 \times 200 \times 160$ and $200 \times 200 \times 89$, respectively. These two HSI datasets are considered as clean data in this simulated experiment.

To simulate noisy HSIs, we added six types of noises to the clean HSI datasets to test the performance of all the compared methods. The details of the added noise are listed as follows:

- Case 1: For different bands, the noise intensity was equal in this case. Furthermore, the same zero-mean Gaussian noise $\mathcal{N}(0, \sigma^2)$ with $\sigma = 0.05$ was added to all the bands
- Case 2: Entries in all bands were corrupted by zero-mean Gaussian noise but with different intensity. Furthermore, the standard deviation of Gaussian noise that was added to each band of HSIs was uniformly selected from 0.01 to 0.1.
- Case 3: All the bands were corrupted by Gaussian noise as Case 2. Besides, 40 bands of the DC Mall dataset (20 bands of the RemoteImage dataset) were randomly chosen to add deadlines, and the number of deadlines in each band is from 5 to 15 with width 1 or 2.
- Case 4: All the bands were contaminated by Gaussian noise as Case 2. Besides, 40 bands of the DC Mall dataset (20 bands of the RemoteImage) were randomly selected to add stripes, and the number of stripes is from 15 to 40 with width 1 or 2.
- Case 5: All the bands were corrupted by Gaussian noise as Case 2. In addition, different percentages of impulse noises which were uniformly selected from 0 to 0.15 were added to each band.
- Case 6: Each band was contaminated by Gaussian and impulse noise as Case 5. Besides, 20 bands of the DC Mall dataset (10 bands of the RemoteImage dataset) were randomly selected to add deadlines as Case 3, and another 20 bands of the DC Mall dataset (10 bands of the RemoteImage dataset) were randomly selected to add stripes as Case 4.

To fairly compare the overall performance of different methods, we set the rank of the expected HSIs as 7 in the DC Mall and RemoteImage dataset for all the compared methods. Furthermore, three measurements—the mean peak signal-to-noise ratio (MPSNR), mean structural similarity index (MSSIM) [53] and Erreur Relative Globale Adimensionnelle de Synthèse (ERGAS) [54]—are utilized to evaluate the restoration quality:

$$\begin{aligned} \text{MPSNR} &= \frac{1}{B} \sum_i^B \text{PSNR}_i, \\ \text{MSSIM} &= \frac{1}{B} \sum_i^B \text{SSIM}_i, \\ \text{ERGAS} &= \sqrt{\frac{1}{B} \sum_i^B \frac{\text{MSE}(\text{Ref}_i, \text{Res}_i)}{\text{Mean}^2(\text{Ref}_i)}}, \end{aligned}$$

where PSNR_i and SSIM_i denote the PSNR and SSIM values for the i -th band, and Ref_i and Res_i denote the i -th band of the reference image and the restorated image, respectively.

Tables 1 and 2 present the restoration results by all the compared methods in terms of the aforementioned three measurements. The best results for each quality measurement are highlighted in bold. It is clear that our proposed method provides the highest values in both MPSNR and MSSIM, and lowest ERGAS in most of the cases, which validates the superiority of the proposed method over the other methods. For Case 1 with the simple independent and identically distributed (i.i.d.) Gaussian noise, most of the methods have comparable performance, which confirms the effectiveness of these methods to some extent, especially for TDL. When the noise becomes more complex, e.g., the intensity of noise varies by bands, more stripes, more deadlines and more impulse noises are added, the advantages of our proposed method become more evident from Case 2 to Case 6. This advantage substantiates the robustness of our proposed method against complex noise, owing to its adaptability to distinguish noise complexities in different HSI bands in our model.

Figures 4 and 5 present some typical bands of the DC Mall and RemoteImage HSI dataset before and after restoration under Case 3, Case 4 and Case 6, respectively. By comparing the restoration results, it can be clearly seen that DP-GMM performs best, effectively removing all types of noise while preserving more detailed information. In Case 3, the restored HSIs of almost all of the compared methods contain some evident deadlines. Similarly, some stripes still exist in the restoration results of most of the compared methods in Case 4. Under the mixture noise of Case 6, LRTV, NMoG and DP-GMM all have relatively better results than the other methods. In conclusion, our proposed method can have stable and better performance in all cases, which verifies the effectiveness of the proposed method with better capability in fitting a much wider range of noises than current methods.

In order to further compare the performance of all the restoration methods, we also show the spectral signatures of some pixels before and after restoration; for example, Figure 6 shows the spectrum of pixels (55, 110) and (110, 184) for the DC Mall HSI dataset under Case 6, and Figure 7 shows similar results for the RemoteImage HSI dataset. It is easy to see that our proposed method provides the best spectral signature among all of the compared methods, which is closest to the original one. Besides, the NMoG method, which adopts non-i.i.d. MoG to model the noises of HSIs, also has comparable performance. These advantages of DP-GMM and NMoG indicate the necessity of giving more consideration to the complex noise structure of HSIs.

Table 1. Quantitative Evaluation of different methods on the simulated DC Mall dataset in different cases of noise.

Noise Case	Evaluation Index	Noise	SVD	BM4D	TDL	WNNM	WSNM	LRMR	LRTV	NMoG	DP-GMM
Case 1	MPSNR (dB)	26.02	38.96	36.66	40.23	36.98	36.98	38.79	38.45	39.05	39.60
	MSSIM	0.7627	0.9833	0.9728	0.9888	0.9806	0.9801	0.9848	0.9836	0.9865	0.9875
	ERGAS	187.97	41.37	53.24	35.34	52.74	52.74	41.98	43.27	42.19	38.22
Case 2	MPSNR (dB)	23.37	35.79	34.25	26.99	36.38	36.28	36.38	36.88	37.70	38.59
	MSSIM	0.6527	0.9600	0.9535	0.7904	0.9750	0.9732	0.9733	0.9739	0.9818	0.9842
	ERGAS	280.83	67.75	70.75	191.56	55.74	56.23	55.93	59.94	49.99	43.01
Case 3	MPSNR (dB)	22.51	33.90	31.67	25.12	35.99	35.34	34.45	35.96	36.99	37.99
	MSSIM	0.6317	0.9486	0.9275	0.7465	0.9730	0.9632	0.9619	0.9695	0.9767	0.9827
	ERGAS	308.89	102.41	128.78	240.34	59.22	81.04	84.31	76.22	65.47	47.72
Case 4	MPSNR (dB)	22.50	33.80	31.56	24.66	34.95	35.26	34.64	35.77	37.09	37.55
	MSSIM	0.6211	0.9280	0.9010	0.7035	0.9547	0.9690	0.9540	0.9642	0.9753	0.9790
	ERGAS	341.01	133.14	163.27	288.97	122.49	67.43	86.43	85.27	61.36	56.19
Case 5	MPSNR (dB)	16.47	27.17	25.12	21.80	34.33	34.45	35.07	36.28	36.15	37.61
	MSSIM	0.4122	0.8299	0.7339	0.5907	0.9263	0.9716	0.9644	0.9686	0.9688	0.9795
	ERGAS	601.66	206.72	225.50	319.36	183.85	72.97	64.44	78.02	126.97	48.80
Case 6	MPSNR (dB)	16.07	26.86	24.54	20.59	33.04	33.79	33.76	34.60	35.88	36.98
	MSSIM	0.3946	0.8202	0.7154	0.5412	0.8975	0.9616	0.9544	0.9599	0.9675	0.9782
	ERGAS	625.80	209.85	240.57	368.62	240.41	104.08	83.47	112.62	469.93	54.68

Table 2. Quantitative Evaluation of different methods on the simulated RemoteImage dateset in different cases of noise.

Noise Case	Evaluation Index	Noise	SVD	BM4D	TDL	WNNM	WSNM	LRMR	LRTV	NMoG	DP-GMM
Case 1	MPSNR (dB)	26.02	37.02	34.50	37.97	34.61	35.58	37.04	36.22	36.83	37.37
	MSSIM	0.6954	0.9655	0.9267	0.9705	0.9570	0.9575	0.9679	0.9588	0.9700	0.9700
	ERGAS	124.05	37.17	50.33	33.01	50.06	43.66	36.44	40.02	38.40	35.76
Case 2	MPSNR (dB)	23.25	33.85	32.12	27.65	34.18	33.02	34.67	34.72	35.56	35.79
	MSSIM	0.5554	0.9272	0.8785	0.7636	0.9399	0.8956	0.9464	0.9406	0.9610	0.9605
	ERGAS	186.22	54.26	64.95	112.77	50.80	82.17	47.62	47.44	44.30	42.69
Case 3	MPSNR (dB)	21.86	30.67	28.78	24.63	33.56	30.52	32.46	33.86	34.40	35.93
	MSSIM	0.5233	0.8764	0.8102	0.6693	0.9280	0.8246	0.9220	0.9314	0.9505	0.9600
	ERGAS	238.56	108.86	158.23	195.47	71.79	160.74	78.50	67.54	59.01	41.9869
Case 4	MPSNR (dB)	22.40	33.28	30.03	25.44	33.54	32.77	33.47	33.72	34.54	35.11
	MSSIM	0.5278	0.9056	0.8239	0.6750	0.9326	0.9152	0.9363	0.9310	0.9534	0.9575
	ERGAS	213.52	81.86	106.85	162.25	61.23	85.16	59.78	66.09	55.48	47.51
Case 5	MPSNR (dB)	17.15	27.98	27.65	23.99	29.82	30.75	33.77	33.52	33.61	35.63
	MSSIM	0.3155	0.7833	0.7090	0.5614	0.7858	0.8423	0.9351	0.9253	0.9490	0.9562
	ERGAS	394.87	124.32	108.07	169.46	229.53	182.01	52.78	75.27	60.58	43.26
Case 6	MPSNR (dB)	16.76	27.49	26.37	22.14	28.24	29.80	32.74	32.41	33.85	35.25
	MSSIM	0.3031	0.7781	0.6717	0.4797	0.7437	0.8153	0.9268	0.9151	0.9497	0.9550
	ERGAS	416.01	141.00	142.33	223.21	268.10	215.28	64.88	104.31	62.96	46.65

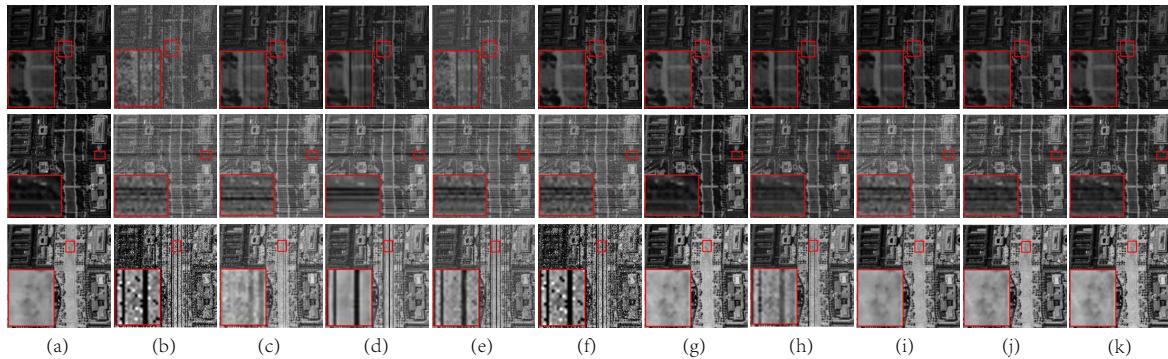


Figure 4. From top to bottom: restoration results of band 118, 20 and 66 of the DCmall HSI dataset under case 3, case 4 and case 6, respectively. (a) Original HSI; (b) Noisy HSI; (c) SVD; (d) BM4D; (e) TDL; (f) WNNM; (g) WSNM; (h) LRMR; (i) LRTV; (j) NMoG; (k) DP-GMM.

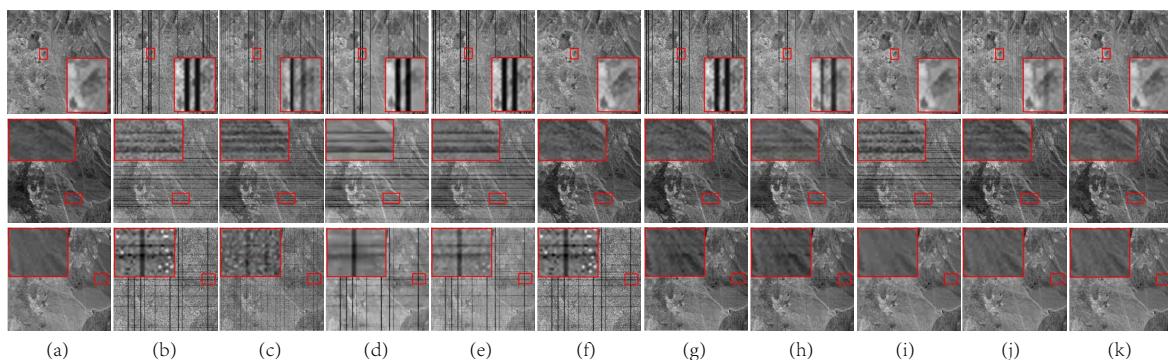


Figure 5. From top to bottom: restoration results of band 49, 11 and 74 of the RemoteImage HSI dataset under Case 3, Case 4 and Case 6, respectively. (a) Original HSI; (b) Noisy HSI; (c) SVD; (d) BM4D; (e) TDL; (f) WNNM; (g) WSNM; (h) LRMR; (i) LRTV; (j) NMoG; (k) DP-GMM.

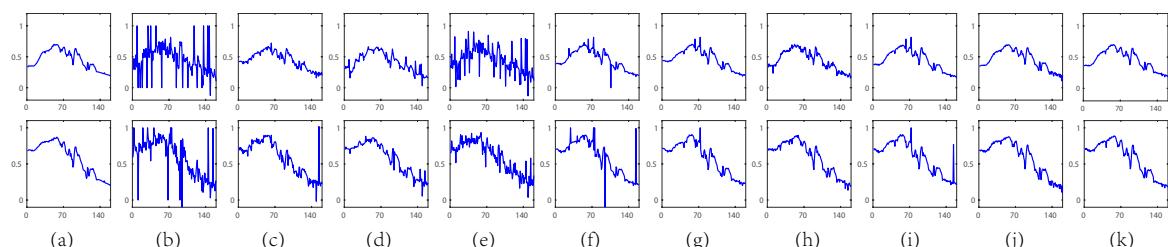


Figure 6. From top to bottom: Spectrum of pixels (55,110) and (110, 184) for the DC Mall HSI dataset. From left to right: (a) Original HSI; (b) Noisy HSI; (c) SVD; (d) BM4D; (e) TDL; (f) WNNM; (g) WSNM; (h) LRMR; (i) LRTV; (j) NMoG; (k) DP-GMM.

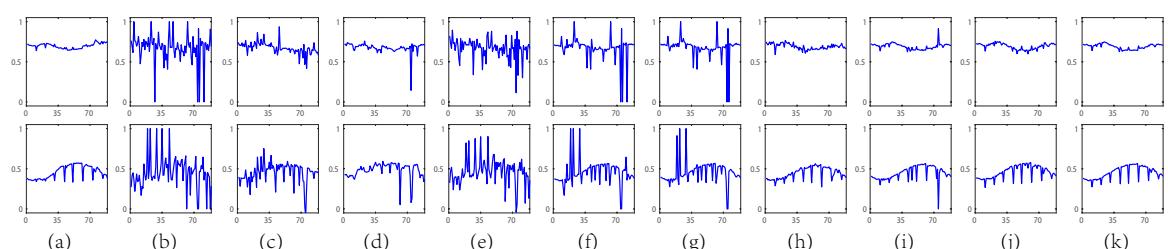


Figure 7. From top to bottom: Spectrum of pixels (78,61) and (137, 41) for the RemoteImage HSI dataset. From left to right: (a) Original HSI; (b) Noisy HSI; (c) SVD; (d) BM4D; (e) TDL; (f) WNNM; (g) WSNM; (h) LRMR; (i) LRTV; (j) NMoG; (k) DP-GMM.

5.2. Real Data Experiments

In this section, two real-world test HSI datasets are used in our experiments: the Hyperspectral Digital Imagery Collection Experiment (HYDICE) Urban (<http://lesun.weebly.com/hyperspectral-data-set.html>) dataset and the Army Geospatial Center (AGC) Terrain (<https://www.agc.army.mil>) dataset.

HYDICE Urban Dataset: The original image is $307 \times 307 \times 210$ in size, and we feed the entire dataset into the denoising algorithms directly. Some bands of Urban are seriously polluted by the atmosphere, water absorption and other unknown noises, such as bands 104–108, 139–151 and 206–210. Additionally, some other bands are contaminated by stripes, deadlines and Gaussian noise, such as bands 1–2, 76 and 198. In this first real data experiment, the rank is set as 7 for all the competing methods.

Figure 8 shows bands 76 (top), 104 (middle) and 139 (bottom) of the restored HSIs of the Urban dataset, and these three bands are typically polluted to varying degrees by different types of noise as analyzed in Figure 1a. Band 76 is relatively clean with several deadlines, and most of the competing methods can remove these deadlines except TDL and BM4D. As for bands 104 and 139, however, some stripes still exist in the restored images for almost all the competing methods. This is because these bands are heavily corrupted and can provide little useful information, and these methods lack noise adaptivity to deal with these complex cases. In contrast, our proposed DP-GMM method is able to completely suppress the stripes and effectively preserves the detailed information. What is more, the images restored by our proposed method are smoother than those restored by other compared methods. These visual results show that our proposed DP-GMM method can remove more complicated noise embedded in HSIs than other competing methods.

Figure 9 shows the vertical mean profiles of bands 76, 104 and 139, which is a quantitative comparison of the restorated results shown in Figure 8. The horizontal axis in Figure 8 represents the column number, and the vertical axis represents the mean digital number value of each column. Band 76 is almost noise-free with only one deadline. There is thus no visible change between the mean profiles before and after the restoration. When the noise becomes more complicated, the curves fluctuate rapidly in the original images for bands 104 and 139. After the restoration processing, the fluctuations are more or less suppressed. Obviously, the DP-GMM methods perform best and obtain the flattest curve without fluctuations. These results are in accordance with the visual results of Figure 8.

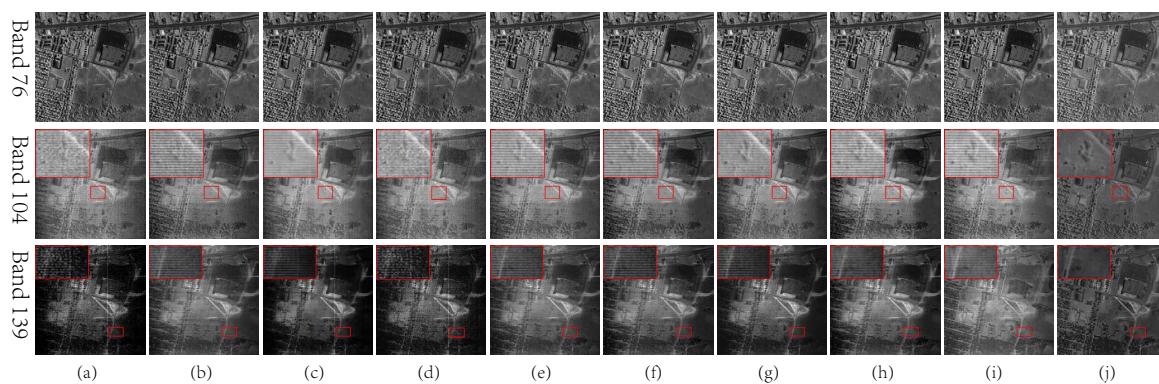


Figure 8. From top to bottom: restoration results of bands 76, 104 and 139 of the real Urban HSI dataset.

(a) Original HSI; (b) SVD; (c) BM4D; (d) TDL; (e) WNNM; (f) WSNM; (g) LRMR; (h) LRTV; (i) NMoG; (j) DP-GMM.

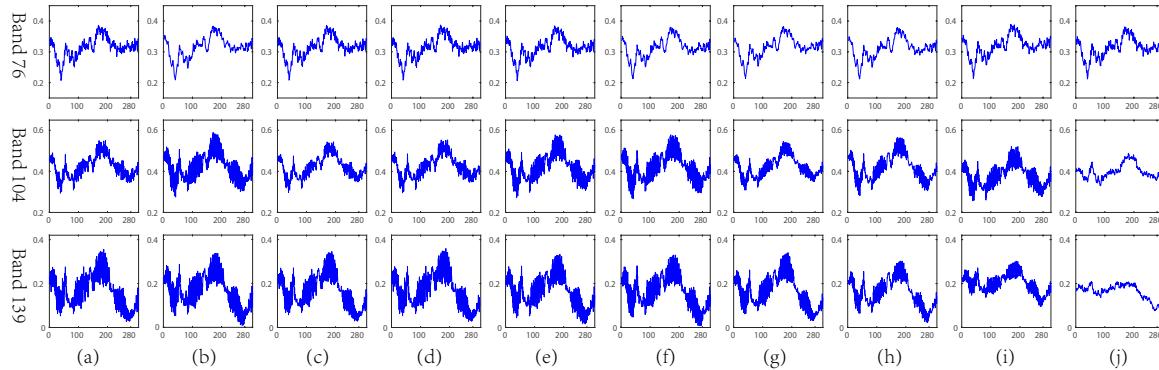


Figure 9. Vertical mean profiles of some typical bands of the Urban HSI dataset. From top to bottom: band 76 (top), band 104 (middle) and band 139 (down). From left to right: (a) Original HSI; (b) SVD; (c) BM4D; (d) TDL; (e) WNNM; (f) WSNM; (g) LRMR; (h) LRTV; (i) NMoG; (j) DP-GMM.

AGC Terrain Dataset: The second real data experiment employed the Terrain dataset as the test dataset, and it has 500×307 pixels and 210 bands. The full Terrain image is mainly corrupted by deadlines, atmosphere, water absorption and others, especially the bands 104–109, 138–152 and 204–210. We also set the rank as 7 in this experiment for all the methods.

We selected three typical bands 2, 139 and 151 to display the performance of all the compared methods in Figure 10. These three bands exhibit evident noise distinctness, as shown in Figure 1b, both in terms of distribution types and corruption extent. In band 2, there are some deadlines clustered together, as shown in the amplified area, and SVD, BM4D, WNNM, WSNM, LRMR, LRTV and our proposed DP-GMM can all remove these deadlines completely. As for bands 139 and 151, which contain many stripes and other mixed noise, it is easy to see that the DP-GMM removes more noise, especially stripes, compared with the other methods, owing to its noise adaptivity for different bands. These visual results show that our proposed DP-GMM method is more robust when dealing with the complicated real noise of HSI data.

Figure 11 shows the vertical mean profiles comparison before and after restoration for each competing method, corresponding to the visualization of Figure 10. The horizontal axis in this figure represents the column number, and the corresponding vertical axis represents the mean digital number value of each column. The curves obtained by different methods in band 2 appear to be slightly smoother to different degrees, as compared with that of the original images. However, as for bands 139 and 151, due to the interference of the mixed noise, especially stripes and deadlines, there are sharp fluctuations in the curve for the original images. After the restoration, the fluctuations are suppressed by all of the competing methods to some extent. Obviously, the curves obtained by our proposed method are smoother than others, which indicates the better restoration results of DP-GMM. These quantitative results are also in accordance with the visual results of Figure 10.

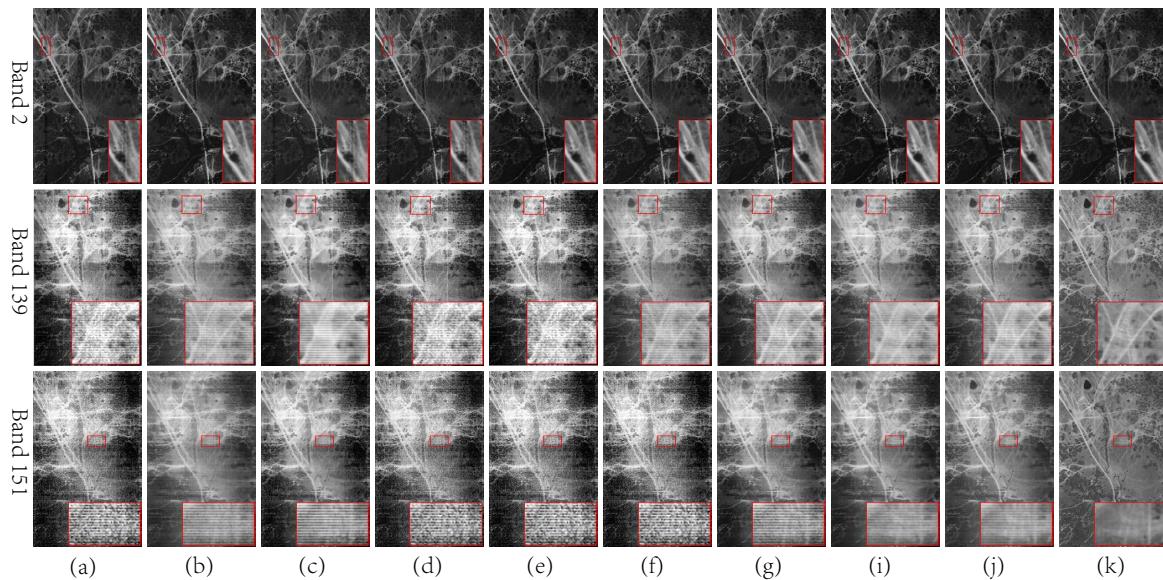


Figure 10. From top to bottom: restoration results of bands 2, 139 and 151 of the real Terrain HSI dataset. (a) Original HSI; (b) SVD; (c) BM4D; (d) TDL; (e) WNNM; (f) WSNM; (g) LRMR; (h) LRTV; (i) NMoG; (j) DP-GMM.

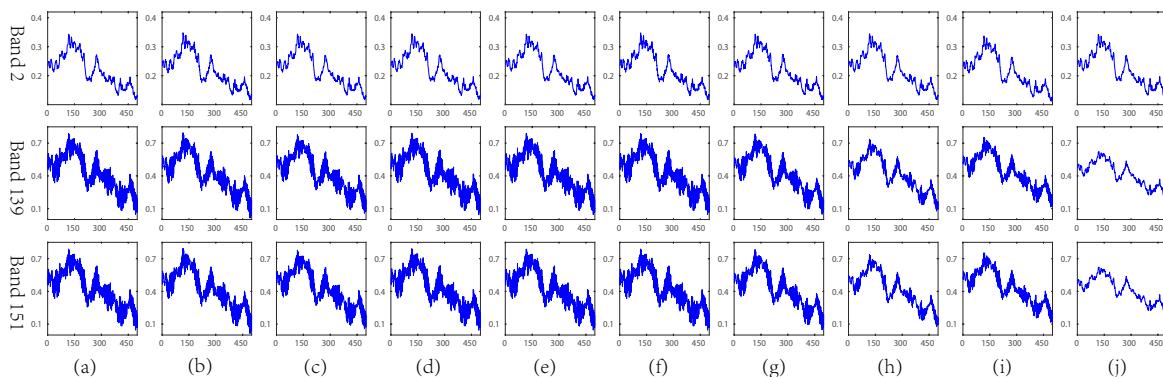


Figure 11. Vertical mean profiles of some typical bands in the Terrain HSI dataset. From top to bottom: band 2 (top), band 139 (middle) and band 151 (down). From left to right: (a) Original HSI; (b) SVD; (c) BM4D; (d) TDL; (e) WNNM; (f) WSNM; (g) LRMR; (h) LRTV; (i) NMoG; (j) DP-GMM.

6. Conclusions

We have proposed a new HSI restoration method by constructing the noise model as a DP-GMM under the Bayesian framework. Most of the current methods do not give much consideration to the variation of the noise complexity among the different bands of the natural HSI. In contrast, our method uses a hierarchical structure to fit the noise so that the noise of different HSI bands can be flexibly encoded by MoGs with different mixture component numbers. The noise distinctness among HSI bands can thus be faithfully reflected in the model, potentially making it more robust in real HSI scenarios. In order to validate the effectiveness of the proposed methods, we conducted a series of experiments on synthetic and real HSI datasets. For the synthetic HSI datasets, our proposed method performed better than all the competing methods in terms of MPSNR, MSSIM and ERGAS. Specifically, in terms of the MPSNR, our method is on average over 1 dB higher than the current state-of-the-art methods: LRTV and NMoG. For the real HSI dataset, our proposed method can remove almost all of the noises, and obtain better visual restoration results compared with other methods, especially for the heavily contaminated bands.

Supplementary Materials: The following are available online at www.mdpi.com/2072-4292/10/10/1631/s1.

Author Contributions: All authors designed the study and discussed the basic structure of the manuscript. Z.Y. carried out the experiments and finished the original version of this paper. D.M. administrated the project. Y.S. validated the results. Z.Y. and Q.Z. designed the experiments. D.M. and Q.Z. reviewed and edited the paper.

Funding: This research was supported by National Key R&D Program of China (2018YFB1004300), the China NSFC projects under contracts 61661166011, 61603292, 11690011, 61721002.

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

References

1. Goetz, A.F. Three decades of hyperspectral remote sensing of the Earth: A personal view. *Remote Sens. Environ.* **2009**, *113*, S5–S16. [[CrossRef](#)]
2. Willett, R.M.; Duarte, M.F.; Davenport, M.A.; Baraniuk, R.G. Sparsity and structure in hyperspectral imaging: Sensing, reconstruction, and target detection. *IEEE Signal Process. Mag.* **2014**, *31*, 116–126. [[CrossRef](#)]
3. Liu, X.; Bourennane, S.; Fossati, C. Nonwhite noise reduction in hyperspectral images. *IEEE Geosci. Remote Sens. Lett.* **2012**, *9*, 368–372. [[CrossRef](#)]
4. Bioucas-Dias, J.M.; Plaza, A.; Dobigeon, N.; Parente, M.; Du, Q.; Gader, P.; Chanussot, J. Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2012**, *5*, 354–379. [[CrossRef](#)]
5. Camps-Valls, G.; Bruzzone, L. Kernel-based methods for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 1351–1362. [[CrossRef](#)]
6. Elad, M.; Aharon, M. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Trans. Image Process.* **2006**, *15*, 3736–3745. [[CrossRef](#)] [[PubMed](#)]
7. Dabov, K.; Foi, A.; Katkovnik, V.; Egiazarian, K. Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Trans. Image Process.* **2007**, *16*, 2080–2095. [[CrossRef](#)] [[PubMed](#)]
8. Mairal, J.; Bach, F.; Ponce, J.; Sapiro, G.; Zisserman, A. Non-local sparse models for image restoration. In Proceedings of the 2009 IEEE 12th International Conference on Computer Vision, Kyoto, Japan, 29 September–2 October 2009; pp. 2272–2279.
9. Liu, L.; Chen, L.; Chen, C.P.; Tang, Y.Y.; Pun, C.M. Weighted joint sparse representation for removing mixed noise in image. *IEEE Trans. Cybern.* **2017**, *47*, 600–611. [[CrossRef](#)] [[PubMed](#)]
10. Lerga, J.; Sucic, V.; Vrankić, M. Separable image denoising based on the relative intersection of confidence intervals rule. *Informatica* **2011**, *22*, 383–394.
11. Lerga, J.; Sucic, V.; Sersic, D. Performance analysis of the LPA-RICI denoising method. In Proceedings of the 6th International Symposium on Image and Signal Processing and Analysis, Salzburg, Austria, 16–18 September 2009; pp. 28–33.
12. Mandic, I.; Peic, H.; Lerga, J.; Stajduhar, I. Denoising of X-ray images using the adaptive algorithm based on the LPA-RICI algorithm. *J. Imaging* **2018**, *4*, 34. [[CrossRef](#)]
13. Othman, H.; Qian, S.E. Noise reduction of hyperspectral imagery using hybrid spatial-spectral derivative-domain wavelet shrinkage. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 397–408. [[CrossRef](#)]
14. Chen, G.; Qian, S.E. Simultaneous dimensionality reduction and denoising of hyperspectral imagery using bivariate wavelet shrinking and principal component analysis. *Can. J. Remote Sens.* **2008**, *34*, 447–454. [[CrossRef](#)]
15. Chen, S.L.; Hu, X.Y.; Peng, S.L. Hyperspectral imagery denoising using a spatial-spectral domain mixing prior. *J. Comput. Sci. Technol.* **2012**, *27*, 851–861. [[CrossRef](#)]
16. Yuan, Q.; Zhang, L.; Shen, H. Hyperspectral image denoising employing a spectral-spatial adaptive total variation model. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 3660–3677. [[CrossRef](#)]
17. Qian, Y.; Ye, M. Hyperspectral imagery restoration using nonlocal spectral-spatial structured sparse representation with noise estimation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2013**, *6*, 499–515. [[CrossRef](#)]
18. Wang, Y.; Niu, R.; Yu, X. Anisotropic diffusion for hyperspectral imagery enhancement. *IEEE Sens. J.* **2010**, *10*, 469–477. [[CrossRef](#)]
19. Zhong, P.; Wang, R. Multiple-spectral-band CRFs for denoising junk bands of hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 2260–2275. [[CrossRef](#)]

20. He, W.; Zhang, H.; Zhang, L.; Shen, H. Hyperspectral image denoising via noise-adjusted iterative low-rank matrix approximation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 3050–3061. [[CrossRef](#)]
21. Candès, E.J.; Li, X.; Ma, Y.; Wright, J. Robust principal component analysis? *J. ACM (JACM)* **2011**, *58*, 11. [[CrossRef](#)]
22. Wu, Z.; Wang, Q.; Wu, Z.; Shen, Y. Total variation-regularized weighted nuclear norm minimization for hyperspectral image mixed denoising. *J. Electron. Imaging* **2016**, *25*, 013037. [[CrossRef](#)]
23. Xie, Y.; Qu, Y.; Tao, D.; Wu, W.; Yuan, Q.; Zhang, W. Hyperspectral Image Restoration via Iteratively Regularized Weighted Schatten p-Norm Minimization. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 4642–4659. [[CrossRef](#)]
24. Gu, S.; Zhang, L.; Zuo, W.; Feng, X. Weighted nuclear norm minimization with application to image denoising. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 2862–2869.
25. Peng, Y.; Suo, J.; Dai, Q.; Xu, W. Reweighted low-rank matrix recovery and its application in image restoration. *IEEE Trans. Cybern.* **2014**, *44*, 2418–2430. [[CrossRef](#)] [[PubMed](#)]
26. Wang, M.; Yu, J.; Xue, J.H.; Sun, W. Denoising of hyperspectral images using group low-rank representation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 4420–4427. [[CrossRef](#)]
27. Xie, Q.; Zhao, Q.; Meng, D.; Xu, Z.; Gu, S.; Zuo, W.; Zhang, L. Multispectral images denoising by intrinsic tensor sparsity regularization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1692–1700.
28. Karami, A.; Yazdi, M.; Asli, A.Z. Noise reduction of hyperspectral images using kernel non-negative tucker decomposition. *IEEE J. Sel. Top. Signal Process.* **2011**, *5*, 487–493. [[CrossRef](#)]
29. Liu, X.; Bourennane, S.; Fossati, C. Denoising of hyperspectral images using the PARAFAC model and statistical performance analysis. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 3717–3724. [[CrossRef](#)]
30. Peng, Y.; Meng, D.; Xu, Z.; Gao, C.; Yang, Y.; Zhang, B. Decomposable nonlocal tensor dictionary learning for multispectral image denoising. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 2949–2956.
31. Wang, Y.; Peng, J.; Zhao, Q.; Leung, Y.; Zhao, X.L.; Meng, D. Hyperspectral image restoration via total variation regularized low-rank tensor decomposition. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 1227–1243. [[CrossRef](#)]
32. Letexier, D.; Bourennane, S. Noise removal from hyperspectral images by multidimensional filtering. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 2061–2069. [[CrossRef](#)]
33. Chen, G.; Qian, S.E. Denoising of hyperspectral imagery using principal component analysis and wavelet shrinkage. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 973–980. [[CrossRef](#)]
34. Chen, G.; Zhu, W.P. Signal denoising using neighbouring dual-tree complex wavelet coefficients. *IET Signal Process.* **2012**, *6*, 143–147. [[CrossRef](#)]
35. Maggioni, M.; Katkovnik, V.; Egiazarian, K.; Foi, A. Nonlocal transform-domain filter for volumetric data denoising and reconstruction. *IEEE Trans. Image Process.* **2013**, *22*, 119–133. [[CrossRef](#)] [[PubMed](#)]
36. Zhang, H.; He, W.; Zhang, L.; Shen, H.; Yuan, Q. Hyperspectral image restoration using low-rank matrix recovery. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 4729–4743. [[CrossRef](#)]
37. Chen, X.; Han, Z.; Wang, Y.; Zhao, Q.; Meng, D.; Lin, L.; Tang, Y. A general model for robust tensor factorization with unknown noise. *arXiv* **2017**, arXiv:1705.06755.
38. He, W.; Zhang, H.; Zhang, L.; Shen, H. Total-variation-regularized low-rank matrix factorization for hyperspectral image restoration. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 178–188. [[CrossRef](#)]
39. Cao, X.; Zhao, Q.; Meng, D.; Chen, Y.; Xu, Z. Robust low-rank matrix factorization under general mixture noise distributions. *IEEE Trans. Image Process.* **2016**, *25*, 4677–4690. [[CrossRef](#)] [[PubMed](#)]
40. Chen, Y.; Cao, X.; Zhao, Q.; Meng, D.; Xu, Z. Denoising hyperspectral image with non-iid noise structure. *IEEE Trans. Cybern.* **2018**, *48*, 1054–1066. [[CrossRef](#)] [[PubMed](#)]
41. Ferguson, T.S. A Bayesian analysis of some nonparametric problems. *Ann. Stat.* **1973**, pp. 209–230. [[CrossRef](#)]
42. Sethuraman, J. A constructive definition of Dirichlet priors. *Stat. Sin.* **1994**, pp. 639–650.
43. Teh, Y.W.; Jordan, M.I.; Beal, M.J.; Blei, D.M. Sharing clusters among related groups: Hierarchical Dirichlet processes. In *Advances in Neural Information Processing Systems 17*; MIT Press: Cambridge, MA, USA; London, UK, 2005; pp. 1385–1392.

44. Rasmussen, C.E. The infinite Gaussian mixture model. In *Advances in Neural Information Processing Systems 12*; MIT Press: Cambridge, MA, USA; London, UK, 2000; pp. 554–560.
45. Chen, P.; Wang, N.; Zhang, N.L.; Yeung, D.Y. Bayesian adaptive matrix factorization with automatic model selection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1284–1292.
46. Meng, D.; De La Torre, F. Robust matrix factorization with unknown noise. In Proceedings of the IEEE International Conference on Computer Vision, Sydney, Australia, 1–8 December 2013; pp. 1337–1344.
47. Teh, Y.W.; Jordan, M.I.; Beal, M.J.; Blei, D.M. Hierarchical Dirichlet Processes. *Publ. Am. Stat. Assoc.* **2006**, *101*, 1566–1581. [[CrossRef](#)]
48. Blei, D.M.; Jordan, M.I. Variational inference for Dirichlet process mixtures. *Bayesian Anal.* **2006**, *1*, 121–143. [[CrossRef](#)]
49. MacEachern, S.N.; Müller, P. Estimating mixture of Dirichlet process models. *J. Comput. Graph. Stat.* **1998**, *7*, 223–238.
50. Neal, R.M. Markov chain sampling methods for Dirichlet process mixture models. *J. Comput. Graph. Stat.* **2000**, *9*, 249–265.
51. Bishop, C.M. *Pattern Recognition and Machine Learning*; Springer: Berlin/Heidelberg, Germany, 2006; pp. 140–155.
52. Haardt, M.; Strobach, P. Method for High-Resolution Spectral Analysis in Multi Channel Observations Using a Singular Value Decomposition (SVD) Matrix Technique. US Patent 5,560,367, 1996.
53. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [[CrossRef](#)] [[PubMed](#)]
54. Wald, L. *Data Fusion: Definitions and Architectures: Fusion of Images Of Different Spatial Resolutions*; Presses des MINES: Paris, France, 2002.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).