Variational Posterior Distribution Approximation in Bayesian Super Resolution Reconstruction of Multispectral Images

Rafael Molina [a,*,1], Miguel Vega [b,1], Javier Mateos [a,1],
Aggelos K. Katsaggelos [c]

[a] *Dept. Ciencias de la Computación e Inteligencia Artificial,*
Univ. de Granada, 18071 Granada, Spain.

[b] *Dept. de Lenguajes y Sistemas Informáticos,*
Univ. de Granada, 18071 Granada, Spain.

[c] *Dept. of Electrical Engineering and Computer Science,*
Northwestern University, Evanston, Illinois 60208-3118.

Abstract

In this paper we present a super resolution Bayesian methodology for pansharpening of multispectral images. By following the hierarchical Bayesian framework, and by applying variational methods to approximate probability distributions this methodology is able to: a) incorporate prior knowledge on the expected characteristics of the multispectral images, b) use the sensor characteristics to model the observation process of both panchromatic and multispectral images, c) include information on the unknown parameters in the model in the form of hyperprior distributions, and d) estimate the parameters of the hyperprior distributions on the unknown parameters together with the unknown parameters, and the high resolution multispectral image. Using real data, the pansharpened multispectral images are compared with the images obtained by other parsharpening methods and their quality is assessed both qualitatively and quantitatively.

Key words: Hierarchical Bayesian modeling, variational distribution approximation, super resolution image reconstruction, pansharpening, multispectral images.

* Corresponding author.

Email addresses: `rms@decsai.ugr.es` (Rafael Molina), `mvega@ugr.es` (Miguel Vega), `jmd@decsai.ugr.es` (Javier Mateos), `aggk@eecs.northwestern.edu` (Aggelos K. Katsaggelos).

1 Introduction

Multispectral images are of interest in commercial, civilian or military areas with a wide range of applications including GPS guidance maps, land type and usage measures and target detection, among others. Nowadays most remote sensing systems include sensors able to capture, simultaneously, several low resolution images of the same area on different wavelengths, forming a multispectral image, along with a high resolution panchromatic image. The main characteristics of such remote sensing systems are the number of bands of the multispectral image and the resolution of those bands and the panchromatic image. For instance, the Landsat 7 satellite (http://landsat.gsfc.nasa.gov/), equipped with the ETM+ sensor, allows for the capture of a multispectral image with six bands (three bands on the visible spectrum plus three bands on the infrared) with a resolution of 30 meters per pixel, a thermal band with a resolution of 60 meters per pixel and a panchromatic band (covering a large zone on the visible spectrum and the near infrared), with a resolution of 15 meters per pixel.

The main advantage of the multispectral image is to allow for a better land type and use recognition but, due to its lower resolution, information on the objects shape and texture may be lost. On the other hand, the panchromatic image allows for a better recognition of the objects in the image and their textures but provides no information about their spectral properties.

Throughout this paper the term *multispectral image reconstruction* will refer to the joint processing of the multispectral and panchromatic images in order to obtain a new multispectral image that, ideally, will exhibit the spectral characteristics of the observed multispectral image and the resolution and quality of the panchromatic image. The use of such an approach, also named pansharpening, will allow us to obtain, in the case of Landsat 7 ETM+, a multispectral image with a resolution of 15 meters per pixel.

A few approximations to this problem have been proposed in the literature. With the Intensity, Hue, Saturation (IHS) transformation method [1], the multispectral image is transformed from the RGB color space into the IHS domain. Then, the intensity component is replaced by the histogram matched panchromatic image and the hue and saturation components are resampled to the panchromatic resolution. The inverse IHS transformation is performed to return to the RGB domain. In [2] principal component analysis (PCA) is applied to the multispectral image bands, the first principal component is replaced by the panchromatic image, and the inverse PCA transform is computed to return to the image domain. Both IHS transformation and PCA based pansharpening, as well as other methods, are available in commercial remote sensing packages like ENVI (http://www.ittvis.com) and ERDAS imagine (http://gi.leica-
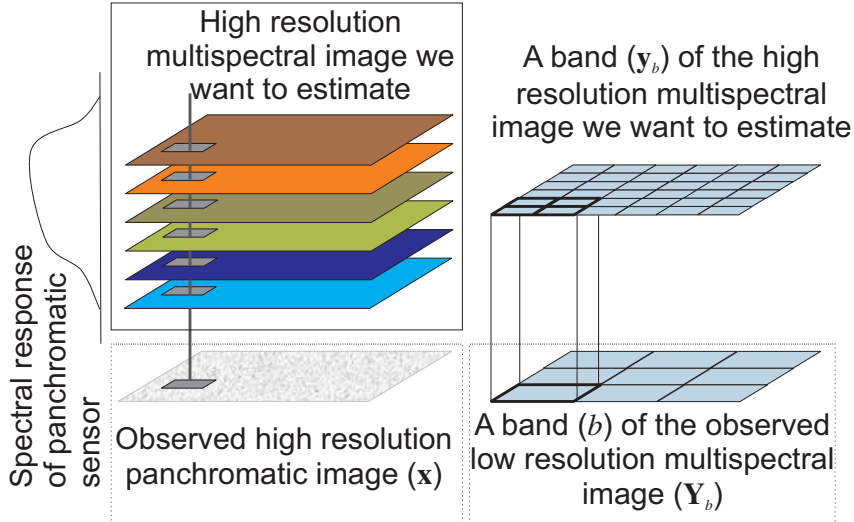
Fig. 1. Problem formulation, acquisition model, and used notation.

geosystems.com). Some wavelet based approaches have been also proposed. In [3], for instance, a redundant wavelet transform is applied to the multispectral and panchromatic images and some of the transformed bands of the multispectral image are either added or substituted by the transform bands of the panchromatic image. A comparison of such techniques can be found in [4]. Price [5] proposed a method relying on image based statistical relationships between the radiances in the low and high spatial resolution bands. Later, Park and Kang [6] modified the statistics estimation method to include spatial adaptivity. Recently a few super-resolution based methods have been proposed. Eismann and Hardie [7] proposed an MAP approach that makes use of a stochastic mixing model of the underlying spectral scene content to achieve resolution enhancement beyond the intensity component of the hyperspectral image. Akgun *et al.* [8] proposed a POCS based algorithm to reconstruct hyperspectral images where the hyperspectral observations from different wavelengths are represented as weighted linear combinations of a small number of basis image planes.

In this paper, in order to tackle the super resolution reconstruction of multispectral images we follow the hierarchical Bayesian framework to incorporate prior knowledge on the expected characteristics of the multispectral images, to model the observation process of both panchromatic and low resolution multispectral images, and also to include information on the unknown parameters in the model in the form of hyperprior distributions. Then, by applying variational methods to approximate probability distributions we estimate the parameters of the hyperprior distributions on the unknown parameters together with the unknown parameters, and the high resolution multispectral image.

The paper is organized as follows. In section 2 the Bayesian modeling and

inference for super resolution reconstruction of multispectral images is presented. The required probability distributions for the Bayesian modeling of the super resolution problem are formulated in section 3. The Bayesian analysis and posterior probability approximation to obtain the parameters and the super resolution reconstructed image is performed in section 4. Experimental results on a real Landsat 7 ETM+ image are described in section 5 and, finally, section 6 concludes the paper.

2 Bayesian Problem Formulation

Let us assume that $\mathbf{y}$, the multispectral image we would observe under ideal conditions with a high resolution sensor, has B bands $\mathbf{y}_b$, $b = 1, \ldots, B$, that is,

$$\mathbf{y} = [\mathbf{y}_1^t, \mathbf{y}_2^t, \ldots, \mathbf{y}_B^t]^t, \tag{1}$$

where each band is of size $p = m \times n$ pixels and t denotes the transpose of a vector or matrix. Each band of this image can be expressed as a column vector by lexicographically ordering the pixels in the band.

In real applications, this high resolution image is not available. Instead, we observe a low resolution multispectral image $\mathbf{Y}$ with B bands $\mathbf{Y}_b$, $b = 1, \ldots, B$, that is,

$$\mathbf{Y} = [\mathbf{Y}_1^t, \mathbf{Y}_2^t, \ldots, \mathbf{Y}_B^t]^t,$$

where each band is of size $P = M \times N$ pixels with $M < m$ and $N < n$. Each band of this image can also be expressed as a column vector by lexicographically ordering the pixels in the band. Figure 1 illustrates the acquisition model and the used notation.

The sensor also provides us with a panchromatic image $\mathbf{x}$ of size $p = m \times n$, obtained by spectrally averaging the unknown high resolution images $\mathbf{y}_b$. Figure 2 shows the spectral response covered by the observed low resolution and panchromatic Landsat 7 ETM+ bands (except the thermal band).

The Bayesian formulation of the high resolution multispectral image reconstruction problem requires the definition of the joint distribution $\mathrm{p}(\Omega, \mathbf{y}, \mathbf{Y}, \mathbf{x})$ of the panchromatic high resolution observation $\mathbf{x}$, the low resolution multispectral observation, $\mathbf{Y}$, the unknown high resolution multispectral image $\mathbf{y}$, and the hyperparameters Ω, describing their distributions. Then, the posterior distribution of the unknowns given the observed low resolution and panchromatic images $\mathrm{p}(\Omega, \mathbf{y}|\mathbf{Y}, \mathbf{x})$ has to be calculated and used to estimate the high resolution image $\mathbf{y}$.

To model the joint distribution, we utilize in this paper the hierarchical Bayesian paradigm (see, for example, [9]). This paradigm has been applied to
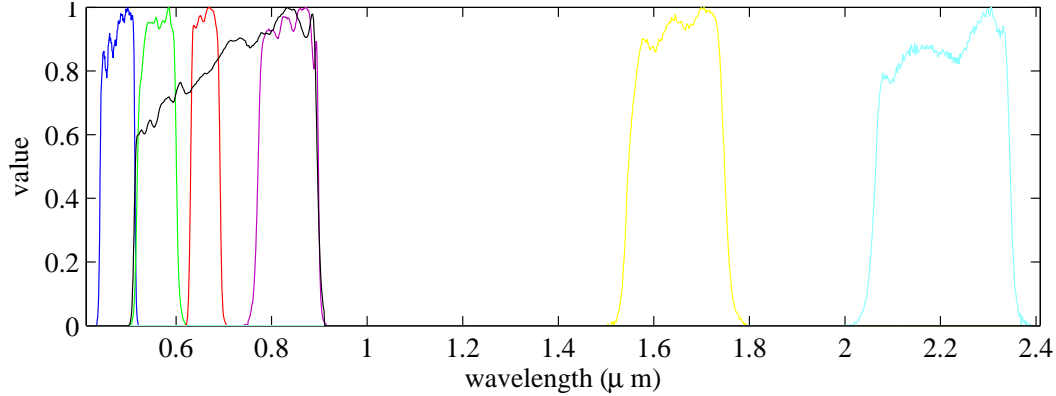
Fig. 2. Landsat 7 ETM+ band spectral response normalized to one.

various areas of research. For instance, Molina *et al.* [9] applied into image restoration, Mateos *et. al.* [10] to removing blocking artifacts in compressed images, and Galatsanos *et al.* [11] in deconvolution problems with partially known blurs.

In the hierarchical approach to our high resolution image reconstruction problem we have two stages. In the first stage, knowledge about the structural form of the low resolution and panchromatic image observation noise and the structural behavior of the high resolution multispectral image is used in forming $p(\mathbf{Y}, \mathbf{x}|\mathbf{y}, \Omega)$ and $p(\mathbf{y}|\Omega)$, respectively. These noise and image models depend on the unknown hyperparameters Ω. In the second stage a hyperprior on the hyperparameters is defined, thus allowing the incorporation of information about these hyperparameters into the process. We note here that each of the two above mentioned conditional distributions will depend only on a subset of Ω, but we use this more general notation until we precisely describe the hyperparameters that define Ω.

For Ω, $\mathbf{y}$, $\mathbf{Y}$, and $\mathbf{x}$ the following joint distribution is defined

$$p(\Omega, \mathbf{y}, \mathbf{Y}, \mathbf{x}) = p(\Omega)p(\mathbf{y}|\Omega)p(\mathbf{Y}, \mathbf{x}|\mathbf{y}, \Omega), \qquad (2)$$

and inference is based on $p(\Omega, \mathbf{y}|\mathbf{Y}, \mathbf{x})$.

The following questions have to be addressed when modeling and performing inference for high resolution multispectral image reconstruction using the hierarchical Bayesian paradigm. The first one relates to the definition of $p(\Omega)$. One can use the improper prior

$$p(\Omega) = const\,, \qquad (3)$$

which assigns *a priori* the same probability to all hyperparameters and makes the observation $\mathbf{Y}$ and $\mathbf{x}$ solely responsible for all the estimates (see [12]). However, we will see in the coming sections that alternative, more informative

hyperpriors can be used to better guide the estimation of the high resolution multispectral image.

The second question to be considered is how inference will be carried out. A commonly used approach consists of estimating the hyperparameters in Ω by using

$$\hat{\Omega} = \arg\max_{\Omega} p(\Omega|\mathbf{Y}, \mathbf{x}) = \arg\max_{\Omega} \int_{\mathbf{y}} p(\Omega, \mathbf{x}, \mathbf{Y}, \mathbf{y}) d\mathbf{y}, \qquad (4)$$

and then estimating the multispectral image by solving

$$\hat{\mathbf{y}} = \arg\max_{\mathbf{y}} p(\mathbf{y}|\hat{\Omega}, \mathbf{Y}, \mathbf{x}). \qquad (5)$$

This inference procedure aims at optimizing a given function and not at obtaining posterior distributions that can be simulated to obtain additional information on the quality of the estimates (see [13]). The solution of the above equations for $\hat{\Omega}$ and $\hat{\mathbf{y}}$ can be viewed as the approximation of posterior distributions by delta functions. Instead of having a distribution over all possible values of the parameters and high resolution multispectral images the above inference procedure chooses a specific set of values. This means that we have neglected many other interpretations of the data. If the posterior is sharply peaked, other values of the hyperparameters and image, will have a much lower posterior probability but, if the posterior is broad, choosing a unique value will neglect many other choices of them with similar posterior probabilities.

The third question to be answered when using the Bayesian paradigm on our problem is to decide how to calculate $p(\Omega, \mathbf{y}|\mathbf{Y}, \mathbf{x})$. The Laplace approximation of distributions has been used, for instance, in blind deconvolution problems when the blur is partially known [11, 14]. An alternative method is provided by variational distribution approximation. This approximation can be thought of as being between the Laplace approximation (see, for instance, [11, 14]) and sampling methods [15]. The basic underlying idea is to approximate $p(\Omega, \mathbf{y}|\mathbf{Y}, \mathbf{x})$ with a simpler distribution, usually one which assumes that $\mathbf{y}$ and Ω are independent given the data (see chapter II in [16] for an excellent introduction to variational methods and their relationships to other inference approaches).

The last few years have seen a growing interest in the application of variational methods [17,18] to inference problems. These methods attempt to approximate posterior distributions with the use of the Kullback-Leibler cross-entropy [19]. Application of variational methods to Bayesian inference problems include graphical models and neural networks [17], independent component analysis [18], mixtures of factor analyzers, linear dynamic systems, hidden Markov models [16], support vector machines [20], and blind deconvolution problems (see Miskin and MacKay [21], Likas and Galatsanos [22] and Molina *et. al.* [23]).

In the coming sections we address the modeling as well as the inference steps in our Bayesian formulation of the super resolution reconstruction of multispectral images.

3 Hyperpriors, priors, and observation models used in super resolution multispectral image reconstruction

In this section we describe the prior model for the multispectral image $p(\mathbf{y}|\Omega)$ and the observation model $p(\mathbf{Y}, \mathbf{x}|\mathbf{y}, \Omega)$ we propose for the first stage of the hierarchical Bayesian paradigm applied to our problem. Then, since these prior and observation models depend on unknown hyperparameters we proceed to explain the proposed hyperprior distribution $p(\Omega)$ on these hyperparameters.

3.1 First stage: prior model on the multispectral image

In this paper we assume no correlation between the different high resolution bands. Then, our prior knowledge about the smoothness of the object luminosity distribution within each band makes it possible to model the distribution of $\mathbf{y}$ by

$$p(\mathbf{y}|\Omega) = \prod_{b=1}^{B} p(\mathbf{y}_b|\alpha_b) \propto \prod_{b=1}^{B} \alpha_b^{(p-1)/2} \exp\left\{-\frac{1}{2}\alpha_b \parallel \mathbf{C}\mathbf{y}_b \parallel^2\right\}, \qquad (6)$$

where $\mathbf{C}$ denotes the Laplacian operator and $1/\alpha_b$ is the variance of the Gaussian distribution of $\mathbf{y}_b$, $b = 1, \ldots, B$.

3.2 First stage: observation model of the low resolution multispectral and panchromatic images

We assume that $\mathbf{Y}$ and $\mathbf{x}$, for a given $\mathbf{y}$ and a set of parameters Ω, are independent and consequently write

$$p(\mathbf{Y}, \mathbf{x}|\mathbf{y}, \Omega) = p(\mathbf{Y}|\mathbf{y}, \Omega)p(\mathbf{x}|\mathbf{y}, \Omega). \qquad (7)$$

Each band, $\mathbf{Y}_b$, is related to its corresponding high resolution image by

$$\mathbf{Y}_b = \mathbf{H}\mathbf{y}_b + \mathbf{n}_b, \quad \forall b = 1, \cdots, B, \qquad (8)$$

where $\mathbf{H}$ is a $P \times p$ matrix representing the blurring, the sensor integration function, and the spatial subsampling (we assume that this process is the same

over the whole set of spectral images), and $\mathbf{n}_b$ is the capture noise, assumed to be Gaussian with zero mean and variance $1/\beta_b$.

A simple but widely used model for the matrix $\mathbf{H}$ is to consider that each pixel (i, j) of the low resolution image is obtained according to (for $m = 2M$ and $n = 2N$)

$$Y_b(i, j) = \frac{1}{4} \sum_{(u,v) \in E_{i,j}} y_b(u, v) + n_b(i, j), \tag{9}$$

where $E_{i,j}$ consists of the indices of the four high resolution pixels $E_{i,j} = \{(2i, 2j), (2i + 1, 2j), (2i, 2j + 1), (2i + 1, 2j + 1)\}$.

We note here that $\mathbf{H}$ can be written as

$$\mathbf{H} = \mathbf{DB}, \tag{10}$$

where $\mathbf{B}$ is a $p \times p$ blurring matrix and $\mathbf{D}$ is a $P \times p$ decimation operator. Given the degradation model for multispectral image super-resolution described by Eq. (8) and assuming independence between the noise observed in the low resolution images, the distribution of the observed $\mathbf{Y}$ given $\mathbf{y}$ and a set of parameters Ω is

$$\mathrm{p}(\mathbf{Y}|\mathbf{y}, \Omega) = \prod_{b=1}^{B} \mathrm{p}(\mathbf{Y}_b|\mathbf{y}_b, \beta_b) \propto \prod_{b=1}^{B} {\beta_b}^{P/2} \exp \left\{ -\frac{1}{2} \beta_b \parallel \mathbf{Y}_b - \mathbf{H}\mathbf{y}_b \parallel^2 \right\}. \tag{11}$$

As already described, the panchromatic image $\mathbf{x}$ is obtained by spectral averaging of the unknown high resolution images $\mathbf{y}_b$. This relation is modeled as

$$\mathbf{x} = \sum_{b=1}^{B} \lambda_b \mathbf{y}_b + \mathbf{v}, \tag{12}$$

where $\lambda_b \geq 0$, $b = 1, 2, \cdots, B$, are known quantities that can be obtained, as we will see later, from the sensor spectral characteristics, and $\mathbf{v}$ is the capture noise that is assumed to be Gaussian with zero mean and variance γ^{-1}. Note that, usually, $\mathbf{x}$ does not depend on all the multispectral image bands but on a subset of them, i. e., some of the λ_b's are equal to zero. For example, for Landsat ETM+ images, the panchromatic image only covers the region from the end of band 1 to the end of band 4 and, so, the rest of the bands have no influence on $\mathbf{x}$.

Using the degradation model in Eq. (12), the distribution of the panchromatic image $\mathbf{x}$ given $\mathbf{y}$, and a set of parameters Ω is given by

$$\mathrm{p}(\mathbf{x}|\mathbf{y}, \gamma) \propto \gamma^{p/2} \exp \left\{ -\frac{1}{2} \gamma \parallel \mathbf{x} - \sum_{b=1}^{B} \lambda_b \mathbf{y}_b \parallel^2 \right\}. \tag{13}$$

3.3 Second stage: hyperprior on the hyperparameters

Let

$$\Omega = (\gamma, \beta_1, \ldots, \beta_B, \alpha_1, \ldots, \alpha_B). \tag{14}$$

A large part of the Bayesian literature is devoted to finding hyperprior distributions $p(\Omega)$ for which $p(\Omega, \mathbf{y}|\mathbf{x}, \mathbf{Y})$ can be calculated in a straightforward way or can be approximated. These are the so called conjugate priors [24]. Conjugate priors have, as we will see later, the intuitive feature of allowing one to begin with a certain functional form for the prior and end up with a posterior of the same functional form, but with the parameters updated by the sample information.

Taking the above considerations about conjugate priors into account, we will assume that each of the hyperparameters, $\omega \in \Omega$, has as a hyperprior the gamma distribution

$$p(\omega|a_\omega^o, c_\omega^o) = \Gamma(\omega|a_\omega^o, c_\omega^o), \tag{15}$$

defined by

$$\Gamma(\omega|a_\omega^o, c_\omega^o) = \frac{((a_\omega^o - 1)c_\omega^o)^{a_\omega^o}}{\Gamma(a_\omega^o)} \omega^{a_\omega^o - 1} \exp[-(a_\omega^o - 1)c_\omega^o \, \omega], \tag{16}$$

where $\omega > 0$ denotes a hyperparameter, and the two parameters $c_\omega^o > 0$ and $a_\omega^o > 1$ will be estimated in section 4.2. This gamma distribution has the following mean, mode, and variance

$$\mathbf{E}[\omega] = \frac{a_\omega^o}{(a_\omega^o - 1)c_\omega^o}, \quad \mathrm{mod}[\omega] = \frac{1}{c_\omega^o}, \quad \mathbf{var}[\omega] = \frac{a_\omega^o}{((a_\omega^o - 1)c_\omega^o)^2}. \tag{17}$$

Note that the mean and mode do not coincide. Using gamma distributions as hyperpriors for the hyperparameters allows us to incorporate in a straightforward manner prior knowledge about the expected value of the hyperparameters and, also, about the confidence on such expected value.

We will then use the following distribution as the hyperprior on the hyperparameters

$$p(\Omega) = p(\gamma, \beta_1, \ldots, \beta_B, \alpha_1, \ldots, \alpha_B) = p(\gamma)p(\beta_1), \ldots, p(\beta_B)p(\alpha_1), \ldots p(\alpha_B), \tag{18}$$

where the hyperprior for each hyperparameter $\omega \in \Omega$ is given by Eq. (15).

Finally, combining the first and second stage of the problem modeling we have the global distribution

$$p(\Omega, \mathbf{y}, \mathbf{Y}, \mathbf{x}) = p(\Omega)p(\mathbf{y}|\Omega)p(\mathbf{Y}|\mathbf{y}, \Omega)p(\mathbf{x}|\mathbf{y}, \Omega), \tag{19}$$

where $p(\Omega)$, $p(\mathbf{y}|\Omega)$, $p(\mathbf{Y}|\mathbf{y}, \Omega)$ and $p(\mathbf{x}|\mathbf{y}, \Omega)$ have been defined in Eqs. (18), (6), (11), and (13), respectively.

4   Bayesian inference and variational approximation of the posterior distribution for super resolution reconstruction of multispectral images

For our selection of hyperparameters in the previous section, the set of all unknowns is given by

$$(\Omega, \mathbf{y}) = (\gamma, \beta_1, \ldots, \beta_B, \alpha_1, \ldots, \alpha_B, \mathbf{y}). \tag{20}$$

As already known, the Bayesian paradigm dictates that inference on $(\Omega, \mathbf{y})$ should be based on

$$p(\Omega, \mathbf{y}|\mathbf{Y}, \mathbf{x}) = \frac{p(\Omega, \mathbf{y}, \mathbf{Y}, \mathbf{x})}{p(\mathbf{Y}, \mathbf{x})}, \tag{21}$$

where $p(\Omega, \mathbf{y}, \mathbf{Y}, \mathbf{x})$ is given by Eq. (19).

Once $p(\Omega, \mathbf{y}|\mathbf{Y}, \mathbf{x})$ has been calculated, $\mathbf{y}$ can be integrated out to obtain $p(\Omega|\mathbf{Y}, \mathbf{x})$. This distribution is then used to simulate or select the hyperparameters. If a point estimate,

$$\hat{\Omega} = (\hat{\gamma}, \hat{\beta}_1, \ldots, \hat{\beta}_B, \hat{\alpha}_1, \ldots, \hat{\alpha}_B), \tag{22}$$

is required, then the mode or the mean of this posterior distribution can be used. Finally, a point estimate of the original multispectral image $\hat{\mathbf{y}}$ can be obtained by maximizing $p(\mathbf{y}|\mathbf{Y}, \mathbf{x}, \hat{\Omega})$. Alternatively the mean value of this posterior distribution can be selected as the estimate of the multispectral image.

From the above discussion it is clear that in order to perform inference we need to either calculate or approximate the posterior distribution $p(\Omega, \mathbf{y}|\mathbf{Y}, \mathbf{x})$. Since $p(\Omega, \mathbf{y}|\mathbf{Y}, \mathbf{x})$ can not be found in closed form, we will apply variational methods to approximate this distribution by the distribution $q(\Omega, \mathbf{y})$.

The variational criterion used to find $q(\Omega, \mathbf{y})$ is the minimization of the Kullback–Leibler divergence, given by [19, 25]

$$C_{KL}(\mathrm{q}(\Omega, \mathbf{y}) \parallel \mathrm{p}(\Omega, \mathbf{y}|\mathbf{Y}, \mathbf{x})) = \int_{\Omega, \mathbf{y}} \mathrm{q}(\Omega, \mathbf{y}) \log \left(\frac{\mathrm{q}(\Omega, \mathbf{y})}{\mathrm{p}(\Omega, \mathbf{y}|\mathbf{Y}, \mathbf{x})} \right) d\Omega d\mathbf{y}$$

$$= \int_{\Omega, \mathbf{y}} \mathrm{q}(\Omega, \mathbf{y}) \log \left(\frac{\mathrm{q}(\Omega, \mathbf{y})}{\mathrm{p}(\Omega, \mathbf{y}, \mathbf{Y}, \mathbf{x})} \right) d\Omega d\mathbf{y}$$

$$+ \mathrm{const}, \tag{23}$$

which is always non negative and equal to zero only when $\mathrm{q}(\Omega, \mathbf{y}) = \mathrm{p}(\Omega, \mathbf{y}|\mathbf{Y}, \mathbf{x})$.

We choose to approximate the posterior distribution $\mathrm{p}(\Omega, \mathbf{y}|\mathbf{Y}, \mathbf{x})$ by the distribution

$$\mathrm{q}(\Omega, \mathbf{y}) = \mathrm{q}(\Omega)\mathrm{q}(\mathbf{y}), \tag{24}$$

where $\mathrm{q}(\mathbf{y})$ and $\mathrm{q}(\Omega)$ denote distributions on $\mathbf{y}$ and Ω, respectively. We now proceed to find the best of these distributions in the divergence sense.

Using Eq. (24) we have in Eq. (23)

$$C_{KL}(\mathrm{q}(\Omega, \mathbf{y}) \parallel \mathrm{p}(\Omega, \mathbf{y}|\mathbf{Y}, \mathbf{x})) = C_{KL}(\mathrm{q}(\Omega)\mathrm{q}(\mathbf{y}) \parallel \mathrm{p}(\Omega, \mathbf{y}|\mathbf{Y}, \mathbf{x}))$$

$$= \mathrm{const} + \int_{\Omega} \mathrm{q}(\Omega) \left(\int_{\mathbf{y}} \mathrm{q}(\mathbf{y}) \log \left(\frac{\mathrm{q}(\Omega)\mathrm{q}(\mathbf{y})}{\mathrm{p}(\Omega, \mathbf{y}, \mathbf{Y}, \mathbf{x})} \right) d\mathbf{y} \right) d\Omega \tag{25}$$

$$= \mathrm{const} + \int_{\mathbf{y}} \mathrm{q}(\mathbf{y}) \left(\int_{\Omega} \mathrm{q}(\Omega) \log \left(\frac{\mathrm{q}(\Omega)\mathrm{q}(\mathbf{y})}{\mathrm{p}(\Omega, \mathbf{y}, \mathbf{Y}, \mathbf{x})} \right) d\Omega \right) d\mathbf{y}. \tag{26}$$

Now, given $\hat{\mathrm{q}}(\Omega)$, an estimate of $\mathrm{q}(\Omega)$, we can obtain an estimate of $\mathrm{q}(\mathbf{y})$ by solving

$$\hat{\mathrm{q}}(\mathbf{y}) = \arg \min_{\mathrm{q}(\mathbf{y})} C_{KL} \left(\hat{\mathrm{q}}(\Omega)\mathrm{q}(\mathbf{y}) \right) \parallel \mathrm{p}(\Omega, \mathbf{y}|\mathbf{Y}, \mathbf{x})), \tag{27}$$

and given $\hat{\mathrm{q}}(\mathbf{y})$, an estimate of $\mathrm{q}(\mathbf{y})$, we can obtain an estimate of $\mathrm{q}(\Omega)$ by solving

$$\hat{\mathrm{q}}(\Omega) = \arg \min_{\mathrm{q}(\Omega)} C_{KL} \left(\mathrm{q}(\Omega)\hat{\mathrm{q}}(\mathbf{y}) \right) \parallel \mathrm{p}(\Omega, \mathbf{y}|\mathbf{Y}, \mathbf{x})). \tag{28}$$

The above equations lead to the following iterative procedure to find $\mathrm{q}(\Omega, \mathbf{y})$.

Algorithm 1
Given $\mathrm{q}^1(\Omega)$, *an initial estimate of the distribution* $\mathrm{q}(\Omega)$,

for $k = 1, 2, \ldots$ *until a stopping criterion is met:*

(1) Find

$$\mathrm{q}^k(\mathbf{y}) = \arg \min_{\mathrm{q}(\mathbf{y})} C_{KL}(\mathrm{q}^k(\Omega)\mathrm{q}(\mathbf{y}) \parallel \mathrm{p}(\Omega, \mathbf{y}|\mathbf{Y}, \mathbf{x})), \tag{29}$$

(2) Find

$$\mathrm{q}^{k+1}(\Omega) = \arg \min_{\mathrm{q}(\Omega)} C_{KL}(\mathrm{q}(\Omega), \mathrm{q}^k(\mathbf{y}) \parallel \mathrm{p}(\Omega, \mathbf{y}|\mathbf{Y}, \mathbf{x})). \tag{30}$$

The convergence of the parameters defining the distributions $q^k(\mathbf{y})$ and $q^{k+1}(\Omega)$ can be used as stopping criterion for the above iterations. In order to simplify such criterion, the following condition can also be used for terminating Algorithm 1 $\| \mathbf{E}[\mathbf{y}]_{q^k(\mathbf{y})} - \mathbf{E}[\mathbf{y}]_{q^{k-1}(\mathbf{y})} \|^2 / \| \mathbf{E}[\mathbf{y}]_{q^{k-1}(\mathbf{y})} \|^2 < \epsilon$, where ϵ is a prescribed bound. Note that this is a convergence criterion over the multispectral image but it normally also implies convergence on the posterior hyperparameter distribution, since its convergence is required for the convergence of the posterior distribution of the image.

Regarding the convergence of the algorithm we first note that, by construction, at every iteration (of the distributions of the multispectral image and hyperparameters) the value of the Kullback-Leibler divergence decreases. To gain further insight into the above algorithm, let us consider a degenerate distribution, $q(\Omega)$, that is,

$$q(\Omega) = \begin{cases} 1 \text{ if } \Omega = \underline{\Omega} \\ 0 \text{ otherwise} \end{cases}, \tag{31}$$

and use $q^*(\mathbf{y}) = p(\mathbf{y}|\mathbf{Y}, \mathbf{x}, \Omega)$.

If, at the $k - th$ iteration of Algorithm 1, $q^k(\Omega)$ is a degenerate distribution on $\underline{\Omega}^k$, then the step of Algorithm 1 for updating the multispectral image produces

$$q^{*k}(\mathbf{y}) = p(\mathbf{y}|\mathbf{Y}, \mathbf{x}, \underline{\Omega}^k). \tag{32}$$

and the step in Algorithm 1 for updating the degenerate distribution on the hyperparameters produces

$$\Omega^{k+1} = \arg \max_{\Omega} \mathbf{E}[\log(p(\Omega, \mathbf{y}, \mathbf{Y}, \mathbf{x})]_{q^{*k}(\mathbf{y})} \tag{33}$$

Interestingly, this is the EM formulation of the maximum *a posteriori* (MAP) estimation of the hyperparameters (see [12, 13]) for our super resolution problem. What Algorithm 1 does is to replace the search for just one hyperparameter by the search for the best distribution on the hyperparameters.

4.1 Calculating $q^k(\mathbf{y})$ and $q^{k+1}(\Omega)$

We now proceed to explicitly calculate the distributions $q^k(\mathbf{y})$ and $q^{k+1}(\Omega)$. Let us assume that at the k-th iteration step of Algorithm 1 we have

$$\mathbf{E}[\gamma]_{q^k(\Omega)} = \gamma^k, \tag{34}$$

$$\mathbf{E}[\alpha_b]_{q^k(\Omega)} = \alpha_b^k, \quad b = 1, \ldots, B, \tag{35}$$

$$\mathbf{E}[\beta_b]_{q^k(\Omega)} = \beta_b^k, \quad b = 1, \ldots, B. \tag{36}$$

Differentiating the right in the right hand side of Eq. (29) with respect to q($\mathbf{y}$) and setting it equal to zero we have that

$$q^k(\mathbf{y}) \propto \exp \mathbf{E}[\log p(\Omega, \mathbf{y}, \mathbf{Y}, \mathbf{x})]_{q^k(\Omega)}, \tag{37}$$

and so

$$q^k(\mathbf{y}) \propto \exp[-\frac{1}{2}(\sum_{b=1}^{B} \left(\alpha_b^k \parallel \mathbf{C}\mathbf{y}_b \parallel^2 + \beta_b^k \parallel \mathbf{Y}_b - \mathbf{H}\mathbf{y}_b \parallel^2\right) + \gamma^k \parallel \mathbf{x} - \sum_{b=1}^{B} \lambda_b \mathbf{y}_b \parallel^2)]. \tag{38}$$

Thus we have that $q^k(\mathbf{y})$ is a normal distribution with the following parameters

$$q^k(\mathbf{y}) = N\left(\mathbf{y} \mid \mathbf{E}^k[\mathbf{y}], \mathbf{cov}^k[\mathbf{y}]\right), \tag{39}$$

with

$$\left(\mathbf{cov}^k[\mathbf{y}]\right)^{-1} = \begin{pmatrix} \alpha_1^k \mathbf{C}^t \mathbf{C} & \mathbf{0}_p & \cdots & \mathbf{0}_p \\ \mathbf{0}_p & \alpha_2^k \mathbf{C}^t \mathbf{C} & \cdots & \mathbf{0}_p \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_p & \mathbf{0}_p & \cdots & \alpha_B^k \mathbf{C}^t \mathbf{C} \end{pmatrix} + \begin{pmatrix} \beta_1^k \mathbf{H}^t \mathbf{H} & \mathbf{0}_p & \cdots & \mathbf{0}_p \\ \mathbf{0}_p & \beta_2^k \mathbf{H}^t \mathbf{H} & \cdots & \mathbf{0}_p \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_p & \mathbf{0}_p & \cdots & \beta_B^k \mathbf{H}^t \mathbf{H} \end{pmatrix}$$
$$+ \quad \gamma^k \Lambda \otimes \mathbf{I}_p, \tag{40}$$

where $\otimes$ is the Kronecker product,

$$\Lambda = \begin{pmatrix} (\lambda_1)^2 & \lambda_1 \lambda_2 & \cdots & \lambda_1 \lambda_B \\ \lambda_2 \lambda_1 & (\lambda_2)^2 & \cdots & \lambda_2 \lambda_B \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_B \lambda_1 & \lambda_B \lambda_2 & \cdots & (\lambda_B)^2 \end{pmatrix}, \tag{41}$$

and

$$\mathbf{E}^k[\mathbf{y}] = \mathbf{cov}^k[\mathbf{y}] \phi^k \tag{42}$$

where $\mathbf{cov}^k[\mathbf{y}]$ has been defined in Eq. (40) and ϕ^k is the $(B \times p) \times 1$ vector

$$\phi^k = \begin{pmatrix} \beta_1^k \mathbf{H}^t & \mathbf{0}_p & \cdots & \mathbf{0}_p \\ \mathbf{0}_p & \beta_2^k \mathbf{H}^t & \cdots & \mathbf{0}_p \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_p & \mathbf{0}_p & \cdots & \beta_B^k \mathbf{H}^t \end{pmatrix} \mathbf{Y} + \gamma^k \begin{pmatrix} \lambda_1 \mathbf{I}_p & \mathbf{0}_p & \cdots & \mathbf{0}_p \\ \mathbf{0}_p & \lambda_2 \mathbf{I}_p & \cdots & \mathbf{0}_p \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_p & \mathbf{0}_p & \cdots & \lambda_B \mathbf{I}_p \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \mathbf{x} \\ \vdots \\ \mathbf{x} \end{pmatrix}. \tag{43}$$

For simplicity, the dependency of ϕ^k on both $\mathbf{x}$ and $\mathbf{Y}$ is not explicitly shown.

14

Once we know $q^k(\mathbf{y})$, the next step is to calculate $q^{k+1}(\Omega)$. Differentiating the right hand side of Eq. (30) with respect to $q(\Omega)$ and setting it to zero we have that $q^{k+1}(\Omega)$ satisfies

$$q^{k+1}(\Omega) \propto \exp \mathbf{E}_{q^k(\mathbf{y})}[\log p(\Omega, \mathbf{y}, \mathbf{Y}, \mathbf{x})], \qquad (44)$$

which produces

$$
\begin{aligned}
q^{k+1}(\Omega) \propto &\prod_{b=1}^{B} \alpha_b^{a^o_{\alpha_b}-1} \, e^{-(a^o_{\alpha_b}-1)c^o_{\alpha_b}\alpha_b} \, \alpha_b^{\frac{p-1}{2}} \, e^{-\frac{1}{2}\alpha_b \mathbf{E}[\|\mathbf{C}\mathbf{y}_b\|^2]_{q^k(\mathbf{y})}} \\
&\times \prod_{b=1}^{B} \beta_b^{a^o_{\beta_b}-1} \, e^{-(a^o_{\beta_b}-1)c^o_{\beta_b}\beta_b} \, \beta_b^{\frac{P}{2}} \, e^{-\frac{1}{2}\beta_b \mathbf{E}[\|\mathbf{Y}_b-\mathbf{H}\mathbf{y}_b\|^2]_{q^k(\mathbf{y})}} \\
&\times \gamma^{a^o_\gamma-1} \, e^{-(a^o_\gamma-1)c^o_\gamma\gamma} \, \gamma^{\frac{P}{2}} \, e^{-\frac{1}{2}\gamma \mathbf{E}[\|\mathbf{x}-\sum_{b=1}^{B} \lambda_b \mathbf{y}_b\|^2]_{q^k(\mathbf{y})}}, \qquad (45)
\end{aligned}
$$

where

$$\mathbf{E}\left[\|\, \mathbf{C}\mathbf{y}_b \,\|^2\right]_{q^k(\mathbf{y})} = \|\, \mathbf{C}\mathbf{E}^k[\mathbf{y}_b] \,\|^2 + \mathrm{tr}[\mathbf{C}^t\mathbf{C}\mathbf{cov}^k[\mathbf{y}_b]], \qquad (46)$$

$$\mathbf{E}\left[\|\, \mathbf{Y}_b - \mathbf{H}\mathbf{y}_b \,\|^2\right]_{q^k(\mathbf{y})} = \|\, \mathbf{Y}_b - \mathbf{H}\mathbf{E}^k[\mathbf{y}_b] \,\|^2 + \mathrm{tr}[\mathbf{H}^t\mathbf{H}\mathbf{cov}^k[\mathbf{y}_b]], \qquad (47)$$

$$
\begin{aligned}
\mathbf{E}[\|\, \mathbf{x} - \sum_{b=1}^{B} \lambda_b \mathbf{y}_b \,\|^2]_{q^k(\mathbf{y})} = & \|\, \mathbf{x} - \sum_{b=1}^{B} \lambda_b \mathbf{E}^k[\mathbf{y}_b] \,\|^2 \\
& + \sum_{i=1}^{B}\sum_{j=1}^{B} \lambda_i \lambda_j \mathrm{tr}[\mathbf{cov}^k[\mathbf{y}_i, \mathbf{y}_j]], \qquad (48)
\end{aligned}
$$

where $\mathbf{E}^k[\mathbf{y}]$ and $\mathbf{cov}^k[\mathbf{y}]$ have been calculated in Eqs. (42) and (40), respectively.

From Eq. (45) we have that

$$q^{k+1}(\Omega) = q^{k+1}(\gamma) \prod_{b=1}^{B} q^{k+1}(\alpha_b) q^{k+1}(\beta_b),$$

where

15

$$q^{k+1}(\gamma) = \Gamma\left(\gamma | a_\gamma^o + \frac{p}{2}, \left[(a_\gamma^o - 1)c_\gamma^o + \frac{1}{2}\mathbf{E}[\| \mathbf{x} - \sum_{b=1}^{B} \lambda_b \mathbf{y}_b \|^2]_{q^k(\mathbf{y})}\right] / (a_\gamma^o + \frac{p}{2} - 1)\right),$$

(49)

$$q^{k+1}(\alpha_b) = \Gamma\left(\alpha_b | a_{\alpha_b}^o + \frac{p-1}{2}, \left[(a_{\alpha_b}^o - 1)c_{\alpha_b}^o + \frac{1}{2}\mathbf{E}[\| \mathbf{C}\mathbf{y}_b \|^2]_{q^k(\mathbf{y})}\right] / (a_{\alpha_b}^o + \frac{p-1}{2} - 1)\right),$$
$$b = 1, \ldots, B,$$

(50)

$$q^{k+1}(\beta_b) = \Gamma\left(\beta_b | a_{\beta_b}^o + \frac{P}{2}, \left[(a_{\beta_b}^o - 1)c_{\beta_b}^o + \frac{1}{2}\mathbf{E}[\| \mathbf{Y}_b - \mathbf{H}\mathbf{y}_b \|^2]_{q^k(\mathbf{y})}\right] / (a_{\beta_b}^o + \frac{P}{2} - 1)\right),$$
$$b = 1, \ldots, B,$$

(51)

where the definition of the gamma distribution has been provided in Eq. (15).

These distributions have the following means

$$\mathbf{E}[\gamma]_{q^{k+1}(\Omega)} = \frac{a_\gamma^o + \frac{p}{2}}{(a_\gamma^o - 1)c_\gamma^o + \frac{1}{2}\mathbf{E}[\| \mathbf{x} - \sum_{b=1}^{B} \lambda_b \mathbf{y}_b \|^2]_{q^k(\mathbf{y})}},$$

(52)

$$\mathbf{E}[\alpha_b]_{q^{k+1}(\Omega)} = \frac{a_{\alpha_b}^o + \frac{p-1}{2}}{(a_{\alpha_b}^o - 1)c_{\alpha_b}^o + \frac{1}{2}\mathbf{E}[\| \mathbf{C}\mathbf{y}_b \|^2]_{q^k(\mathbf{y})}}, \quad b = 1, \ldots, B,$$

(53)

$$\mathbf{E}[\beta_b]_{q^{k+1}(\Omega)} = \frac{a_{\beta_b}^o + \frac{P}{2}}{(a_{\beta_b}^o - 1)c_{\beta_b}^o + \frac{1}{2}\mathbf{E}[\| \mathbf{Y}_b - \mathbf{H}\mathbf{y}_b \|^2]_{q^k(\mathbf{y})}}, \quad b = 1, \ldots, B, \quad (54)$$

which are then used to recalculate the distributions of $\mathbf{y}$ in Algorithm 1.

We can rewrite the above equations as follows

$$\frac{1}{\mathbf{E}[\gamma]_{q^{k+1}(\Omega)}} = \mu_\gamma \frac{(a_\gamma^o - 1)c_\gamma^o}{a_\gamma^o} + (1 - \mu_\gamma)\frac{\mathbf{E}[\| \mathbf{x} - \sum_{b=1}^{B} \lambda_b \mathbf{y}_b \|^2]_{q^k(\mathbf{y})}}{p}, \quad (55)$$

$$\frac{1}{\mathbf{E}[\alpha_b]_{q^{k+1}(\Omega)}} = \mu_{\alpha_b} \frac{(a_{\alpha_b}^o - 1)c_{\alpha_b}^o}{a_{\alpha_b}^o} + (1 - \mu_{\alpha_b})\frac{\mathbf{E}[\| \mathbf{C}\mathbf{y}_b \|^2]_{q^k(\mathbf{y})}}{p - 1}, \quad (56)$$
$$b = 1, \ldots, B,$$

$$\frac{1}{\mathbf{E}[\beta_b]_{q^{k+1}(\Omega)}} = \mu_{\beta_b} \frac{(a_{\beta_b}^o - 1)c_{\beta_b}^o}{a_{\beta_b}^o} + (1 - \mu_{\beta_b})\frac{\mathbf{E}[\| \mathbf{Y}_b - \mathbf{H}\mathbf{y}_b \|^2]_{q^k(\mathbf{y})}}{P}, \quad (57)$$
$$b = 1, \ldots, B,$$

where

$$\mu_\gamma = \frac{a_\gamma^o}{p/2 + a_\gamma^o}, \quad \mu_{\alpha_b} = \frac{a_{\alpha_b}^o}{(p-1)/2 + a_{\alpha_b}^o}, \quad \mu_{\beta_b} = \frac{a_{\beta_b}^o}{P/2 + a_{\beta_b}^o}, \quad b = 1 \ldots, B. \quad (58)$$

The above equations indicate that μ_γ and μ_{α_b}, and μ_{β_b}, $b = 1, \ldots, B$, can be understood as normalized confidence parameters taking values in the interval

16

$[0,1)$. That is, when they are zero no confidence is placed on the given hyperparameters, while when the corresponding normalized confidence parameter is asymptotically equal to one it fully enforces the prior knowledge of the mean (no estimation of the hyperparameters is performed). Furthermore, for each hyperparameter, the inverse of the mean of its posterior distribution approximation is a weighted sum of the inverse of the mean of its hyperprior distribution (see Eq. (17)) and its maximum likelihood estimate.

4.2 Variational super resolution reconstruction with hyperprior distribution parameter estimation

In the discussion so far in this section we have assumed that the values of the parameters a_ω^o and c_ω^o, $\omega \in \Omega$, are known. In this section we study how the application of the variational methodology to our super resolution reconstruction problem allows us to estimate those parameters as well.

Let us consider for each channel $b \in \{1, \ldots, B\}$ the following Bayesian modeling of our reconstruction problem

$$
\begin{aligned}
\mathrm{p}_b(\alpha_b, \beta_b, \gamma_b, \mathbf{y}_b, \mathbf{Y}_b, \mathbf{x}) &\propto \alpha_b^{(p-1)/2} \beta_b^{P/2} \gamma^{p/2} \\
&\times \exp\left\{ -\frac{1}{2}\alpha_b \parallel \mathbf{C}\mathbf{y}_b \parallel^2 -\frac{1}{2}\beta_b \parallel \mathbf{Y}_b - \mathbf{H}\mathbf{y}_b \parallel^2 -\frac{1}{2}\gamma_b \parallel \mathbf{x} - \lambda_b \mathbf{y}_b \parallel^2 \right\}.
\end{aligned} \quad (59)
$$

where we have assumed that $\mathrm{p}_b(\omega) \propto$ const, for $\omega \in \Omega_b = \{\alpha_b, \beta_b, \gamma_b\}$. Observe that in this formulation we are considering only the contribution of channel b to the panchromatic image, scaled by the weight λ_b (see Eq. (13)).

We can now use the variational methodology to approximate $\mathrm{p}_b(\alpha_b, \beta_b, \gamma_b, \mathbf{y}_b | \mathbf{Y}_b, \mathbf{x})$ by the distribution

$$
\mathrm{q}_b(\alpha_b, \beta_b, \gamma_b, \mathbf{y}_b) = \mathrm{q}_b(\alpha_b, \beta_b, \gamma_b)\mathrm{q}_b(\mathbf{y}_b) \quad (60)
$$

and apply Algorithm 1 to the problem of estimating $\mathbf{y}_b$ and Ω_b by replacing $\mathbf{y}$ by $\mathbf{y}_b$ and Ω by Ω_b.

Let us assume that at the k-th iteration step of the Algorithm 1 we have from the distribution of the hyperparameters

$$
\mathbf{E}[\gamma]_{q_b^k(\Omega_b)} = \gamma_b^{*k}, \quad (61)
$$

$$
\mathbf{E}[\alpha_b]_{q_b^k(\Omega_b)} = \alpha_b^{*k}, \quad (62)
$$

$$
\mathbf{E}[\beta_b]_{q_b^k(\Omega_b)} = \beta_b^{*k}. \quad (63)
$$

Then we obtain

$$q_b^k(\mathbf{y}_b) = N\left(\mathbf{y}_b \mid \mathbf{E}_b^k[\mathbf{y}_b], \mathbf{cov}_b^k[\mathbf{y}_b]\right), \tag{64}$$

where

$$\left(\mathbf{cov}_b^k[\mathbf{y}_b]\right)^{-1} = \alpha^{*k}_b \mathbf{C}^t\mathbf{C} + \beta^{*k}_b \mathbf{H}^t\mathbf{H} + \gamma^{*k}_b \mathbf{I}_p, \tag{65}$$

and

$$\mathbf{E}_b^k[\mathbf{y}] = \mathbf{cov}_b^k[\mathbf{y}_b](\beta^{*k}_b \mathbf{H}^t\mathbf{Y}_b + \gamma^{*k}\mathbf{x}). \tag{66}$$

Then

$$q_b^{k+1}(\Omega_b) = q_b^{k+1}(\gamma)q_b^{k+1}(\alpha_b)q_b^{k+1}(\beta_b),$$

where

$$q_b^{k+1}(\gamma) = \Gamma\left(\gamma \mid 1 + \frac{p}{2}, \left[\frac{1}{2}\mathbf{E}[\| \mathbf{x} - \lambda_b\mathbf{y}_b \|^2]_{q_b^k(\mathbf{y}_b)}\right]/\frac{p}{2}\right), \tag{67}$$

$$q_b^{k+1}(\alpha_b) = \Gamma\left(\alpha_b \mid 1 + \frac{p-1}{2}, \left[\frac{1}{2}\mathbf{E}[\| \mathbf{C}\mathbf{y}_b \|^2]_{q_b^k(\mathbf{y}_b)}\right]/\frac{p-1}{2}\right), \tag{68}$$

$$q_b^{k+1}(\beta_b) = \Gamma\left(\beta_b \mid 1 + \frac{P}{2}, \left[\frac{1}{2}\mathbf{E}[\| \mathbf{Y}_b - \mathbf{H}\mathbf{y}_b \|^2]_{q_b^k(\mathbf{y}_b)}\right]/\frac{P}{2}\right). \tag{69}$$

We can then use the parameters of the above posterior distributions as the parameters of the hyperpriors, that is,

$$a_\gamma^o = 1 + \frac{p}{2}, \qquad c_\gamma^o = \frac{1}{B}\sum_{b=1}^{B}\lim_{k\to\infty}\frac{\mathbf{E}[\| \mathbf{x} - \lambda_b\mathbf{y}_b \|^2]_{q_b^k(\mathbf{y}_b)}}{p}, \tag{70}$$

$$a_{\alpha_b}^o = 1 + \frac{p-1}{2}, \qquad c_{\alpha_b}^o = \lim_{k\to\infty}\frac{\mathbf{E}[\| \mathbf{C}\mathbf{y}_b \|^2]_{q_b^k(\mathbf{y}_b)}}{p-1}, \qquad b = 1,\ldots,B, \tag{71}$$

$$a_{\beta_b}^o = 1 + \frac{P}{2}, \qquad c_{\beta_b}^o = \lim_{k\to\infty}\frac{\mathbf{E}[\| \mathbf{Y}_b - \mathbf{H}\mathbf{y}_b \|^2]_{q_b^k(\mathbf{y}_b)}}{P}, \qquad b = 1,\ldots,B. \tag{72}$$

Finally, the confidence parameters in Eq. (58) take the values

$$\mu_\gamma = \frac{p+2}{2p+2}, \quad \mu_{\alpha_b} = \frac{p+1}{2p+1}, \quad \mu_{\beta_b} = \frac{P+2}{2P+2}, \ b = 1,\ldots,B. \tag{73}$$

5 Experimental Results

The proposed super resolution reconstruction algorithm has been tested on the set of Landsat ETM+ images [26] listed in Table 1. In the experiments we used regions of interest of size 128×128 pixels of the multispectral images and their corresponding regions of size 256×256 pixels of the panchromatic images. Figure 3 displays the region of interest for image D in Table 1. This

Image	Date	Path	Row
A	2000-07-30	200	031
B	2000-08-08	199	031
C	2002-08-05	015	034
D	2002-07-15	133	037

Table 1
Landsat 7 ETM+, L1G Orthorectified image sets.

figure depicts, on the left, a false RGB color image composed of bands 4, 3, and 2 of the Landsat ETM+ multispectral image and on the right its corresponding panchromatic image. Note that the multispectral image has been resized by zero-order hold to the size of the panchromatic image for displaying purposes.
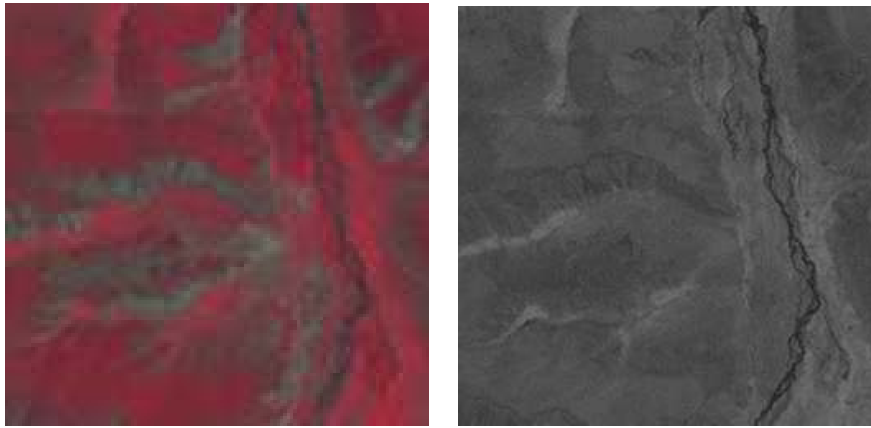
(a)   (b)

Fig. 3. Details of the (a) false color multispectral image D (see Table 1) and (b) its corresponding panchromatic image.

According to the ETM+ sensor spectral response, depicted in Fig. 2, the panchromatic image only covers the spectrum of a part of the first four bands of the multispectral image. Hence, we apply the proposed method with $B = 4$.

In order to apply Algorithm 1, we need to know the contribution of each band to the panchromatic image, that is, the values of λ_b, $b = 1, 2, \ldots, B$ in Eq. (12). These values can be obtained from the spectral response of the ETM+ sensor (see Fig. 2). Note that the panchromatic image covers a region of wavelengths from almost the end of band 1 to the end of band 4 and that the sensor sensibility is not constant over the whole range. Taking into account these considerations, we obtain values for λ_b, $b = 1, 2, 3, 4$ by summing up the spectral response of the panchromatic sensor weighted by the response of the sensor for each multispectral band. The obtained values are then normalized so that their sum equals one, thus producing the values of λ_b displayed on Table 2. The matrix $\mathbf{H}$ in Eq. (8) was modelled following Eq. (9).

The initial distribution on the parameters, $q^1(\Omega)$, in Algorithm 1 was chosen

Table 2

λ_1	λ_2	λ_3	λ_4
0.0078	0.2420	0.2239	0.5263

Estimated values for λ_b, $b = 1, \ldots, 4$.

(a)

(b)

(c)

(d)

Fig. 4. Reconstructions of image D using (a) bicubic interpolation (MI), (b) the Price method (MII), (c) the proposed method with $\mu_\omega = 0, \forall \omega \in \Omega$ ($MIII$), and (d) the proposed method with hyperprior distribution parameter estimation (MIV).

from Eqs. (49)–(51) using $a_\omega = 1$, $\omega \in \Omega$, and assuming that $q^0(\mathbf{y})$, is a degenerate distribution on the bicubic interpolation of the observed multispectral image. The value of $\epsilon = 10^{-6}$ was used as the prescribed bound to stop the iteration.

In the rest of the section we compare the proposed method with the result of applying bicubic interpolation to each low resolution band of the multispectral image (MI method) and the result of applying the method proposed by Price in [5] (MII method). We present results obtained by the proposed method in two cases. First, for the case when $a_\omega^o = 1$, $\omega \in \Omega$, that is, we make the observed data fully responsible for the estimation process ($MIII$ method). For the second case, following the variational approach for the hyperprior
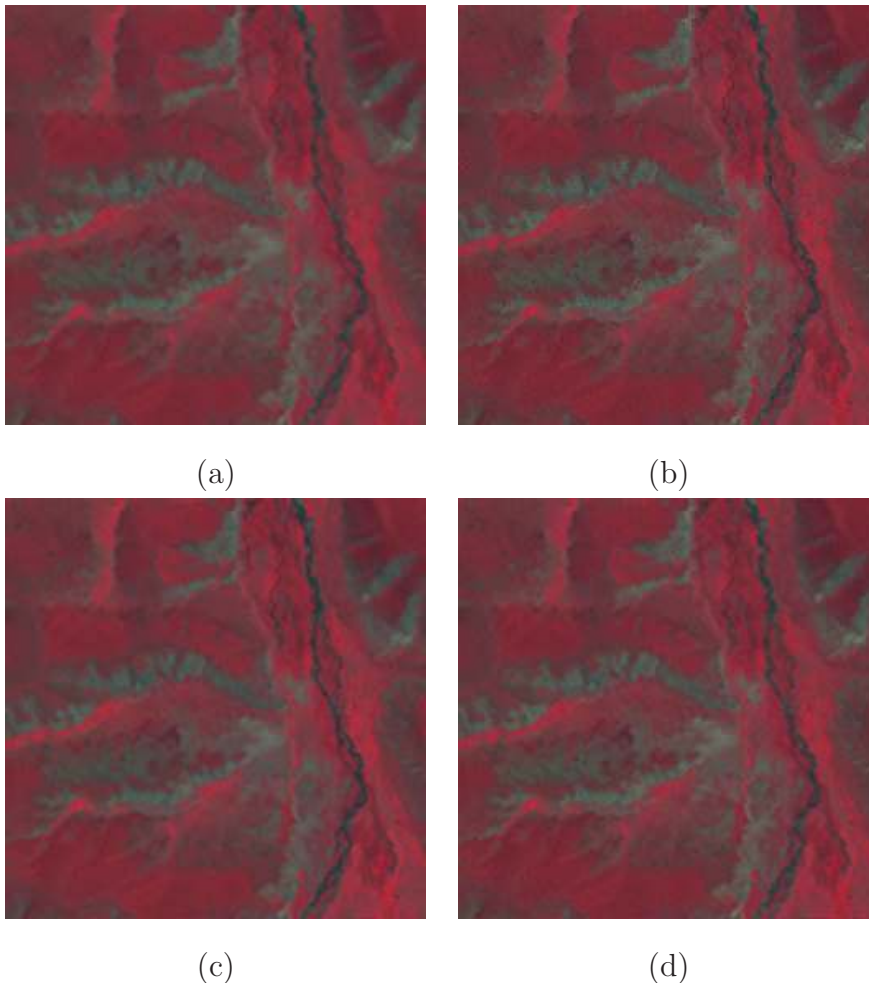
20

(a)   (b)

(c)   (d)

Fig. 5. Reconstructions of image B using (a) bicubic interpolation (MI), (b) the Price method (MII), (c) the proposed method with $\mu_\omega = 0, \forall \omega \in \Omega$ ($MIII$), and (d) the proposed method with hyperprior distribution parameter estimation (MIV).

distribution parameter estimation described in section 4.2, a_ω^o, c_ω^o, and the confidence parameters μ_ω, $\omega \in \Omega$, are selected using Eqs. (70)–(73) (MIV method). Note that the use of the MIV method implies that the variational approach is first applied to the model for each channel described in Eq. (59) to obtain estimates of a_ω^o and c_ω^o.

Table 3 shows the comparison of the four approaches in terms of

$$\text{PSNR}_b = 10 \log_{10} \left\{ 255^2 \times P/ \parallel \mathbf{Y}_b - \mathbf{D}\hat{\mathbf{y}}_b \parallel^2 \right\}.$$

For the $MIII$ and MIV methods their corresponding band estimates $\hat{\mathbf{y}}_b$, are the mean value of the limit distribution of their corresponding $q^k(\mathbf{y}_b)$. The obtained values show that the MIV method always obtains higher PSNR values than the Price method and the $MIII$ method. Figures 4 and 5 show the reconstructions corresponding to images D and B, respectively, and figure 6 details of the observed image B and its reconstructions. The observed multispectral image in Fig. 6a has been resized by zero-order hold to the size of the high
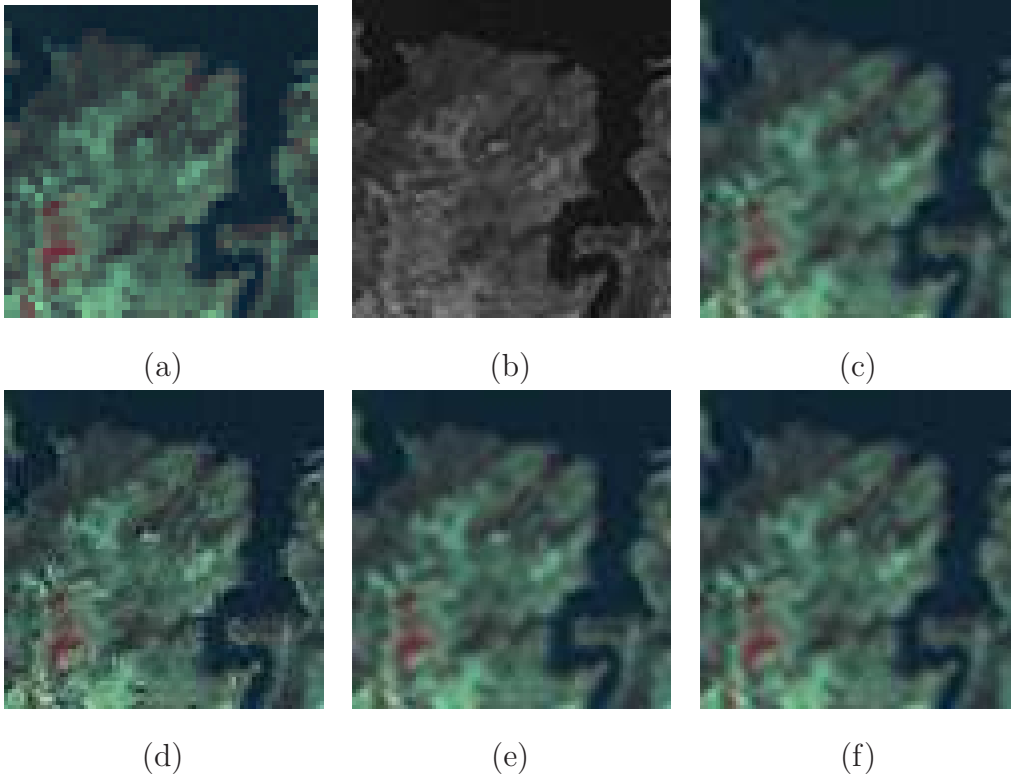
21

Fig. 6. Detail of the (a) observed multispectral image B, (b) its corresponding panchromatic image, and the reconstructions using (c) bicubic interpolation (MI), (d) the Price method (MII), (e) the proposed method with $\mu_\omega = 0, \forall \omega \in \Omega$ ($MIII$), and (f) the proposed method with hyperprior distribution parameter estimation (MIV).

resolution images. Visual inspection of the results shows that the proposed method MIV provides sharper images than the other methods and, also, they are not as noisy as the results obtained with the Price method (MII).

In a second experiment we simulate a multispectral image of size 64×64 pixels and its corresponding panchromatic image of size 128×128 pixels from the observed 128×128 pixels multispectral image and its corresponding 256×256 pixels panchromatic image using the observation process, that is, applying the sensor integration and downsampling the observed image according to Eq. (9). The objective is to obtain sets of low resolution images that, once reconstructed, could be numerically compared with their corresponding observed multispectral images. Evaluation of the results of pansharpened multispectral satellite images is a difficult task, since spatial improvement and spectral fidelity must be assessed separately, and a number of quality indices have been proposed.

Spatial improvement has been assessed bandwise in terms of the UIQI index, and Δ_{SNR} in dB. UIQI [27] is an overall image quality index, with a value in the range $[-1, 1]$. The closer the UIQI index to one the better the reconstruction. This quality index models any distortion as a combination of three different factors: loss of correlation, luminance distortion, and contrast distor-

22

Image	band	MI	MII	MIII	MIV
A	1	21.0	19.1	21.7	21.7
	2	19.8	18.1	20.4	20.6
	3	17.2	15.5	17.1	17.5
	4	20.1	17.8	19.7	20.7
B	1	18.1	17.0	18.9	18.9
	2	16.8	15.8	17.3	17.5
	3	14.8	13.7	15.1	15.4
	4	17.4	16.4	16.5	18.0
C	1	23.4	22.7	24.0	24.0
	2	22.6	21.9	22.4	22.9
	3	20.7	20.2	20.4	20.9
	4	19.7	18.7	13.8	19.6
D	1	21.7	20.4	22.4	22.4
	2	21.8	20.5	22.4	22.4
	3	19.5	18.1	19.6	20.0
	4	17.3	15.6	17.0	18.2

Table 3
PSNR (dB) obtained by bicubic interpolation (MI), the Price method (MII), the proposed method with $\mu_\omega = 0, \forall \omega \in \Omega$ ($MIII$), and the proposed method with hyperprior distribution parameter estimation (MIV).

tion. Numerical comparison of the different reconstruction methods, presented in Table 4, shows that the proposed methods perform better than bicubic interpolation (MI) and Price method (MII). Note, however, that in some cases, the proposed method $MIII$ performs worse than classical methods (see, for instance, the fourth band of image C in Table 4). This is due to the suboptimal estimation of the hyperparameters since no prior information about their values is included in the estimation process. However, when including information on the values of the parameters, as for the method MIV, our algorithms perform better than the other methods. Additional improvement could also be obtained by including more precise information about their possible values.

Spectral fidelity of the different fused images is quantified by the standard ERGAS index (from the French "Erreur Relative Globale Adimensionalle de Synthèse") [28], a dimensionless global criterion based on the ratio of Root Mean Square Error (RMSE) and bandwise mean. A low value of this index, specially a value smaller than 3.0, indicates a higher quality of the multispec-

Image	band	UIQI				Δ_{SNR}			
		MI	MII	MIII	MIV	MI	MII	MIII	MIV
A	1	0.70	0.68	0.76	0.75	26.2	24.5	26.8	26.8
	2	0.75	0.71	0.81	0.79	22.8	21.3	24.1	23.5
	3	0.76	0.72	0.77	0.76	18.3	16.8	18.8	18.3
	4	0.77	0.72	0.84	0.81	24.1	22.1	25.7	24.9
B	1	0.76	0.79	0.81	0.80	23.0	21.9	23.8	23.7
	2	0.78	0.81	0.85	0.78	19.2	18.5	20.7	19.2
	3	0.78	0.81	0.83	0.79	16.1	15.6	17.3	16.3
	4	0.79	0.81	0.89	0.80	18.5	17.8	21.3	18.7
C	1	0.57	0.58	0.63	0.62	31.3	29.8	31.9	31.9
	2	0.68	0.67	0.73	0.72	27.3	26.4	28.6	28.3
	3	0.68	0.66	0.67	0.72	21.5	20.5	22.1	22.5
	4	0.79	0.81	0.76	0.82	25.2	25.4	18.8	25.5
D	1	0.73	0.68	0.78	0.77	26.5	24.0	27.3	27.3
	2	0.79	0.72	0.82	0.83	26.8	23.8	27.5	27.6
	3	0.79	0.73	0.73	0.80	21.3	18.7	20.3	21.6
	4	0.83	0.77	0.89	0.87	23.0	21.4	24.6	24.1

Table 4

UIQI and Δ_{SNR} (dB) values obtained by bicubic interpolation (MI), the Price method (MII), the proposed method with $\mu_\omega = 0, \forall \omega \in \Omega$ ($MIII$), and the proposed method with hyperprior distribution parameter estimation (MIV).

Image	MI	MII	MIII	MIV
A	3.67	4.40	3.35	3.55
B	5.44	5.83	4.52	5.31
C	2.50	2.73	3.28	2.28
D	2.86	3.75	2.83	2.65

Table 5

ERGAS value obtained by bicubic interpolation (MI), the Price method (MII), the proposed method with $\mu_\omega = 0, \forall \omega \in \Omega$ ($MIII$), and the proposed method with hyperprior distribution parameter estimation (MIV).

tral image. Results, displayed in Table 5, show that bicubic interpolation (MI) provides very good spectral results since it does not take into account the panchromatic image or the information on other bands. Price method (MII) also

produces good spectral results although the obtained ERGAS values are always greater than the *MI* ones. The proposed *MIII* method performs better, giving a better spectral fidelity, except for the image C due, again, to a suboptimal estimation of the parameters. The proposed method *MIV*, which includes a hyperprior ditribution for parameter estimation, provides lower ERGAS values and, hence, a better spectral fidelity, for all the images when compared to classical methods.

We conclude this section by providing some information on the computing requirements of the algorithms. For our proposed methods, the most demanding computational task, both in terms of processing and memory requirements, is the calculation of the $\mathbf{cov}^k[\mathbf{y}]$ matrix, defined in Eq. (40). This calculation reduces, in the Fourier domain, to the inversion of p matrices of size $B \times B$ (in our experiments $p = 256 \times 256$ and $B = 4$). Each iteration of Algorithm 1, took 20 sec. to execute on an Xeon 3.2GHz processor, for observed multispectral images of size 128×128. The MIV method required less than 5 iterations to reach convergence for all considered cases using as stopping criterion $\epsilon = 10^{-6}$ (see section 4), while the MIII method required a higher number (between 8 and 20) of iterations. The Price method took only 0.2 sec. to reconstruct all four bands, while bicubic interpolation took 0.8 sec.

6   Conclusions

In this paper the reconstruction of multispectral images has been formulated from a superresolution point of view. A hierarchical Bayesian framework to incorporate prior knowledge on the expected characteristics of the multispectral images, to model the observation process of both panchromatic and low resolution multispectral images, and also to include information on the unknown parameters in the model in the form of hyperprior distributions has been presented. Then, by applying variational methods to approximate probability distributions, the parameters of the hyperprior distributions on the unknown parameters together with the unknown parameters, and the high resolution multispectral image have been estimated. Based on the presented experimental results, the proposed method outperforms bicubic interpolation and the method in [5].

References

[1]   W. J. Carper, T. M. Lillesand, R. W. Kiefer, The use of intensity-hue-saturation transformations for merging SPOT panchromatic and multispectral image data, Phot. Eng. & Rem. Sens. 56 (4) (1990) 459–467.

[2] P. S. Chavez, S. Sides, J. Anderson, Comparison of three different methods to merge multiresolution and multispectral data: Landsat TM and SPOT panchromatic, Phot. Eng. & Rem. Sens. 57 (3) (1991) 295–303.

[3] J. Nuñez, X. Otazu, O. Fors, A. Prades, V. Pala, R. Arbiol, Multiresolution-based image fusion with additive wavelet decomposition, IEEE Trans on Geosc. & Rem. Sens. 37 (3) (1999) 1204–1211.

[4] V. Vijayaraj, A quantitative analysis of pansharpened images, Master's thesis, Mississippi St. Univ. (2004).

[5] J. Price, Combining multispectral data of different spatial resolution, IEEE Trans. on Geosc. & Rem. Sens. 37 (3) (1999) 1199–1203.

[6] J. Park, M. Kang, Spatially adaptive multi-resolution multispectral image fusion, Int. Jour. of Rem. Sens. 25 (23) (2004) 5491–5508.

[7] M. Eismann, R. Hardie, Hyperspectral resolution enhancement using high-resolution multispectral imaginary with arbitray response functions, IEEE Trans. on Geosc. & Rem. Sens. 43 (3) (2005) 455–465.

[8] T. Akgun, Y. Altunbasak, R. Mersereau, Super-resolution reconstruction of hyperspectral images, IEEE Trans. on Img. Proc. 14 (11) (2005) 1860–1875.

[9] R. Molina, A. K. Katsaggelos, J. Mateos, Bayesian and regularization methods for hyperparameter estimation in image restoration, IEEE Trans. on Img. Proc. 8 (2) (1999) 231–246.

[10] J. Mateos, A. Katsaggelos, R. Molina, A Bayesian approach to estimate and transmit regularization parameters for reducing blocking artifacts, IEEE Trans. on Img. Proc. 9 (7) (2000) 1200–1215.

[11] N. P. Galatsanos, V. Z. Mesarovic, R. Molina, A. K. Katsaggelos, J. Mateos, Hyperparameter estimation in image restoration problems with partially-known blurs, Optical Eng. 41 (8) (2002) 1845–1854.

[12] R. Molina, M. Vega, J. Mateos, A. Katsaggelos, Hierarchical Bayesian super resolution reconstructiuon of multispectral images, in: 2006 European Signal Processing Conference (EUSIPCO 2006), 2006, p. Tue.4.4.

[13] R. Molina, M. Vega, J. Mateos, A. Katsaggelos, Parameter estimation in Bayesian reconstruction of multispectral images using super resolution techniques, in: 2006 International Conference on Image Processing (ICIP 2006), 2006, pp. 1749–1752.

[14] N. P. Galatsanos, V. Z. Mesarovic, R. Molina, A. K. Katsaggelos, Hierarchical Bayesian image restoration for partially-known blur, IEEE Trans. on Img. Proc. 9 (10) (2000) 1784–1797.

[15] C. Andrieu, N. de Freitras, A. Doucet, M. Jordan, An introduction to MCMC for machine learning, Mach. Lear. 50 (2003) 5–43.

[16] M. Beal, Variational algorithms for approximate Bayesian inference, Ph.D. thesis, The Gatsby Computational Neuroscience Unit, University College London (2003).

[17] M. I. Jordan, Z. Ghahramani, T. S. Jaakola, L. K. Saul, An introduction to variational methods for graphical models, in: Learning in Graphical Models, MIT Press, 1998, pp. 105–162.

[18] J. Miskin, Ensemble learning for independent component analysis, Ph.D. thesis, Astrophysics Group, University of Cambridge (2000).

[19] S. Kullback, Information Theory and Statistics, New York, Dover Publications, 1959.

[20] C. Bishop, M. Tipping, Variational relevance vector machine, in: Proceedings of the 16th Conference on Uncertainty in Articial Intelligence, Morgan Kaufmann Publishers, 2000, pp. 46–53.

[21] J. W. Miskin, D. J. C. MacKay, Ensemble learning for blind image separation and deconvolution, in: M. Girolami (Ed.), Advances in Independent Component Analysis, Springer-Verlag Scientific Publishers, 2000.

[22] A. C. Likas, N. P. Galatsanos, A variational approach for Bayesian blind image deconvolution, IEEE Trans. on Sig. Proc. 52 (8) (2004) 2222–2233.

[23] R. Molina, J. Mateos, A. Katsaggelos, Blind deconvolution using a variational approach to parameter, image, and blur estimation, IEEE Trans. on Img Proc. 15 (12) (2006) 3715–3727.

[24] J. O. Berger, Statistical Decision Theory and Bayesian Analysis, New York, Springer Verlag, 1985, Ch. 3 and 4.

[25] S. Kullback, R. A. Leibler, On information and sufficiency, Annals of Math. Stat. 22 (1951) 79–86.

[26] N. L. Program, Landsat ETM+ Scenes, in: Global Land Cover Facility, U.S. Geological Survey, Sioux Falls, South Dakota, http://glcf.umiacs.umd.edu.

[27] Z. Wang, A. C. Bovik, A universal image quality index, IEEE Sign. Proc. Lett. 9 (3) (2002) 81–84.

[28] L. Wald, T. Ranchin, M. Mangolini, Fusion of satellite images of different spatial resolutions: assessing the quality of resulting images, Phot. Eng. Rem. Sens. 63 (6) (1997) 691–699.