Bayesian Scale Space Analysis of Temporal Changes in Satellite Images

Leena Pasanen¹ and Lasse Holmström
Department of Mathematical Sciences
University of Oulu
Finland

abstract

We consider the detection of land cover changes using pairs of Landsat ETM+ satellite images. The images consist of eight spectral bands and to simplify the multidimensional change detection task, the image pair is first transformed to a one dimensional image. When the transformation is non-linear, the true change in the images may be masked by complex noise. For example, when changes in the normalized difference vegetation index (NDVI) is considered, the variance of noise may not be constant over the image and methods based on image thresholding can be ineffective. To facilitate detection of change in such situations, we propose an approach that uses Bayesian statistical modeling and simulation based inference. In order to detect both large and small scale changes, our method uses a scale space approach that employs multi-level smoothing. We demonstrate the technique using artificial test images and two pairs of real Landsat ETM+ satellite images.

keywords: Remote sensing; Change detection; NDVI; Bayesian methods; Scale space; iBSiZer

1 Introduction

Over the years, various approaches have been proposed for the detection of land cover change using remote sensing images; see for example the review articles [23, 17, 4]. For Landsat ETM+ satellite images, particular challenges are posed by the multidimensionality of the data due to the eight image spectral bands that represent different wavelengths of light measured by the sensor. Three basic approaches to detect changes in pairs of such images have been suggested. First, one can simply classify pixels in each image separately and then identify those pixels whose class has changed. The downside of this approach is the risk of producing a large number of erroneous change indications [23]. The second approach is to classify corresponding pixels in the two images simultaneously. This is time consuming and the results are often unsatisfactory, in particular when high-quality training sample data are not available [17].

We will adopt the third commonly used approach where change detection is applied to a transformed image. Possible transformations include band-wise differences, difference of vegetation indexes, temporal difference of components obtained by Principal Component Analysis (PCA) or Tasseled Cap Transformation, image regression, and the magnitude of the difference vector [17, 23]. In band-wise differencing and PCA, one can analyze one band of the transformed image at a time and with Tasseled Cap one can choose one of the components, brightness, greenness or wetness for the analysis. The image thus transformed is one-dimensional and the detection of changed pixels is simpler. However, as some information is inevitably lost, the transformation must be chosen carefully to match the specific type of change one is interested in.

Many approaches have been proposed for change detection in transformed images, such as thresholding,
exemplified by density slicing [23] and the algorithm of Kittler and Illingworth (e.g. [19]). Other alternatives include the Bayesian rule of minimum error, as well as spatial-context based classification where one assumes that a pixel is likely to be surrounded by pixels belonging to the same class [1]. Another example of spatial-contextual classification was proposed in [6] where fuzzy clustering is applied to change detection from magnitude of difference images.

In [18], the changes are detected from MODIS satellite image pairs using the difference of normalized difference vegetation index (NDVI) images. The images are first filtered by removing poor quality data values, anomalously high and low values and delineating the water areas from the images. The changed pixels are then detected by thresholding. NDVI images have been utilized also for example in [24] where the unchanged pixels in the two NDVI images are assumed to have a bivariate normal distribution and the changed pixels are detected using hypothesis testing.

The difficulty in detecting change from nonlinearly transformed images, such as the difference of NDVI images or the magnitude of difference images, is that the noise in the transformed image may be heteroscedastic. For example, NDVI in the case of Landsat ETM+ images is computed as the difference of bands 4 and 3 divided by their sum. Therefore, the noise standard deviation in the difference of NDVI images is usually not constant, being much higher in the areas where the intensity values are small and lower in the areas with high intensity values. However, all the aforementioned change detection methods assume that the noise variance in fact is constant.

Another important goal to consider is the detection of change in different spatial scales. One might for instance be interested both in high-intensity local changes and low-intensity, large scale changes. Then simple thresholding based methods again will not work and more sophisticated approaches are called for. For example, wavelets have been used for optical remote sensing images [20] and, for SAR-images, the Bayesian method of multiscale change profiles (MCP) has been proposed in [13]. We propose to use for change detection iBSiZer, the Bayesian SiZer for images, originally designed for the detection of differences in digital images [11] and now here extended for the analysis of multidimensional satellite images. The extension was briefly outlined in [21]. This method has its origin in the SiZer scale space technique for one-dimensional nonparametric probability density estimation and curve fitting [2, 3] which since its inception has been extended into many directions (cf. [9, 10]). In [7] SiZer was used in the context of digital images. A scale space approach for detecting features from a single remote sensing image was also described in [8]. Using iBSiZer, we are able to carry out spatial-contextual statistical inference even in the presence of non-linear image transformations and at the same time detect changes in different spatial scales. Although our focus will be on NDVI differences, the proposed method is equally applicable in connection with other image transformations.

The proposed method consists of three steps. First, the posterior distribution of the transformed image is obtained. The precise mathematical expression for the posterior distribution can be very complex, in particular with the image transformations needed for vegetation indices and the magnitude of the difference vector. Therefore, instead of defining the posterior analytically, we simulate a large sample from the joint posterior of the multi-channel satellite images and then apply the desired transformation to the sample images. Prior knowledge about the properties of the images can be utilized in defining the image priors. One option is a Gaussian smoothing prior, that assumes that the neighboring pixels are relatively similar. However, as images sometimes contain edges in known locations, we also propose a Gaussian prior that allows the image to have edges at known locations. In addition, since satellite images may sometimes contain a spatial trend, we have also experimented with a prior that models a polynomial trend in the image. As the results depend on the values of the parameters in the model, we have applied the full Bayesian approach, and treated the parameters in the models as random variables. In addition, we have tried the empirical Bayesian approach of estimating the parameters using the maximum likelihood method.
In iBSiZer, change detection in different spatial scales is implemented by making inferences about features in the smooths of the underlying true image. Therefore, the second step of our method is to obtain the posterior distributions of the smooths of a transformed image. Finally, in the third step these posteriors are used to detect the credibly changed pixels. The credibility analysis is performed jointly over all pixels in the image. The computations required by simulation based inference are greatly facilitated by the use of Fourier transformation based algorithmic speed-ups.

The rest of the paper is organized as follows. Section 2 explains our basic idea for change detection in satellite images. In 2.1 the first step of the proposed method, the statistical model of the image transformation is described. In 2.2 the other two steps, smoothing and change detection, are discussed. In Section 3, we report experimental results for three test image pairs. The first pair consists of two artificial images that mimic features of satellite images. The second image pair consist of subimages of a real Landsat ETM+ satellite image pair from eastern Finland. The third test image pair contains an analysis of another real satellite image pair obtained from northern Norway. In Section 4, we compare the results of our method with thresholding using the artificial test images. The main points of the paper are summarized in Section 5. Most of the mathematical details of the posterior distribution of the NDVI difference are given in Appendices A, B and C. Appendix D explains the maximum likelihood estimation of the model parameters.

2 Bayesian multiscale analysis of temporal change

2.1 Modeling a transformation

Consider a pair of two $M \times N$ Landsat ETM+ satellite images taken at two different instants of time. We write each $M \times N$ image band columnwise as an $n \times 1$ vector with $n = MN$ and then combine all 8 bands of the two images into a single $16n \times 1$ vector $y = [y_{11}^T, \ldots, y_{18}^T, y_{21}^T, \ldots, y_{28}^T]^T$, where $y_{ij}$ is the band $j$ at time $i$. Each band of the observed satellite image is assumed to contain additive Gaussian noise. Hence, for time $i = 1, 2$ and band $j = 1, \ldots, 8$,

$$y_{ij} = x_{ij} + \epsilon_{ij},$$

where $\epsilon_{ij} \sim N(0, \sigma_{ij}^2 I)$ and $x_{ij}$ is the noiseless image corresponding to the observed image $y_{ij}$. We also assume that the noise vectors $\epsilon_{ij}$ for different times and bands are independent. Hence, $\epsilon = [\epsilon_{11}^T, \ldots, \epsilon_{18}^T, \epsilon_{21}^T, \ldots, \epsilon_{28}^T]^T$ is Gaussian with mean zero and the $16n \times 16n$ covariance matrix

$$\Sigma = \text{diag}[\sigma_{11}^2 I, \ldots, \sigma_{18}^2 I, \sigma_{21}^2 I, \ldots, \sigma_{28}^2 I],$$

where $I = [1, \ldots, 1]$. Although only this particular model is considered here, our method can be easily modified to accommodate other noise types, too (cf. [11]).

To detect change, a transformation is first applied to the $16n \times 1$ vector $y$ that reduces its dimension to $n$. One possibility is the difference between two NDVI images. An NDVI image of a Landsat ETM+ multispectral image $v = [v_1^T, \ldots, v_8^T]^T$ is computed as

$$\text{NDVI}[v_1^T, v_1^T] = \frac{v_4 - v_3}{v_4 + v_3},$$

where division is carried out pixelwise. The range of pixels in an NDVI image is $[-1, 1]$ and the closer the value of a pixel is to 1, the more vegetation there is in the corresponding area on the ground. The difference of the NDVI images of a pair of satellite images represented by $v = [v_{11}^T, \ldots, v_{18}^T, v_{21}^T, \ldots, v_{28}^T]^T$ is denoted by

$$N_v \equiv \frac{v_{24} - v_{23}}{v_{24} + v_{23}} - \frac{v_{14} - v_{13}}{v_{14} + v_{13}}.$$
Accordingly, the true and the noisy NDVI difference image is denoted by $N_x$ and $N_y$, respectively. Other possible transformations were described in the Introduction and, although only NDVI differences are considered here, the proposed method works similarly with them, too.

Change detection in the transformed $n \times 1$ image is based on Bayesian inference. Our prior beliefs about true image bands $x = [x_1^T, \ldots, x_{18}^T, x_{21}^T, \ldots, x_{28}^T]^T$ are modeled by a prior probability density $p(x)$ and the likelihood of the data $y$ is given by the conditional density $p(y|x)$. By the Bayes' formula, the posterior density of $x$ is then

$$p(x|y) = \frac{p(y|x)p(x|y)}{p(y)} \propto p(x)p(y|x).$$  \hspace{1cm} (4)

Note that only the bands actually involved in the transformation need to be included in the posterior. For example, in the case of NDVI, we need to obtain the posterior only for $[x_{13}^T, x_{14}^T, x_{23}^T, x_{24}^T]^T$.

The random variable of interest $N_x$, the transformation of the noiseless satellite images, can be computed from $x$. However, for example in the case of NDVI, $p(N_y|x)$ is a difference of quotients of Gaussian random variables, whereas the likelihood function $p(y|x)$ is simply a multivariate Gaussian. Our strategy, therefore, is to generate a sample from $p(x|y)$ and apply the transformation (3) to produce a sample from the posterior $p(N_x|y)$.

A common prior assumption about the properties of $x$ is that the noiseless satellite images are relatively smooth. This motivates our use of Gaussian smoothing priors, that penalize from large differences between neighboring pixels. The degree of prior smoothness is controlled by a smoothing parameter $\lambda_{ij} > 0$, with larger $\lambda_{ij}$ corresponding to a smoother $x_{ij}$. a priori. We will also allow prior temporal dependence in the images corresponding to the same band for the two instants of time considered. The strength of dependence for band $j$ is controlled by a parameter $\theta_j > 0$. Note that, if desired, dependence between different bands could be modeled similarly. The details of the models used are described in Appendices A and B.

With a Gaussian smoothing prior and a Gaussian likelihood, the posterior distribution is also Gaussian and therefore simulation from the posterior $p(x|y)$ is straightforward. If all $\lambda_{ij}$’s and all $\sigma_{ij}^2$’s are equal, $\theta_3 = \theta_4$, and toroidal image boundary conditions are used, then one can use Fourier methods to accelerate the computations considerably. This is discussed in more detail in Appendix A (cf. [11]). However, the satellite images may also feature sharp edges that arise for example from lake boundaries. Hence, in addition to the smoothing priors, we have also employed a Gaussian prior capable of preserving sharp edges at known locations; see Appendix B for details. Although the posterior in this case is still Gaussian, Fourier methods cannot be utilized to speed up computations.

The particular boundary conditions needed for the use of Fourier methods imply an assumption that the intensity values of $x$ do not exhibit a spatial gradient over the image area. The satellite images nevertheless often do contain such a distortion where one corner or edge is darker or lighter than the rest of the image. Since such a feature cannot be modeled adequately by a simple Gaussian smoothing prior, we propose to add a polynomial trend to the model. It is then assumed that the observed satellite image is a sum of three parts: a Gaussian random variable containing the small scale features, a polynomial trend, and the noise,

$$y = x + Z\beta + \epsilon,$$

where $x$ is the Gaussian random variable (6), $\epsilon$ contains the noise and $Z\beta$ contains the large scale features modeled as a polynomial trend with $Z$ as the design matrix and $\beta$ as a random parameter vector. In our examples a second degree polynomial trend appeared to be sufficient. Details about this model are described in Appendix C.

The likelihood function $p(y|x)$ depends on the covariance matrix (2) and the prior $p(x)$ depends on the smoothing parameters $\lambda_{ij}$ and $\theta_j$. If we have no prior knowledge about these parameters, we can try different values and choose ones that seem to produce posterior means that are in line with our idea of the smoothness.
of the underlying images. We however preferred full Bayesian inference that considers the parameters as random variables with log-normal prior distributions. When this is not feasible, an alternative approach is to estimate the values of the parameters for example by maximum likelihood. We have implemented this for a model where the images are assumed independent, that is \( \theta_3 = \theta_4 = 0 \). The details are presented in Appendix D.

### 2.2 Multiscale analysis of change

When land cover changes are analyzed using remote sensing images, one is often interested both in sharp local changes as well as changes over larger geographic areas. Scale-dependent changes can be revealed by multiresolution analysis techniques, such as wavelets. Here we propose to use a scale space approach where image features in different spatial scales are explored by smoothing the transformed image \( N_x \) using several different smoothing levels.

Thus, let \( S_\lambda \) be a smoothing operator corresponding to a smoothing level \( \lambda \geq 0 \), and denote by \( N_\lambda x \) the smooth \( S_\lambda N_x \) of \( N_x \). Bayesian inference about the features of the smooth \( N_\lambda x \) is based on its posterior density \( p(N_\lambda^2|y) \). This posterior density can be obtained by applying the transformation \( S_\lambda \) to the density \( p(N_u|y) \). In practice, we first simulate a sample from \( p(x|y) \), then transform the sample vectors \( x \) (e.g. (3)), and finally apply the smoother \( S_\lambda \) to the transformed sample vectors \( N_x \). Inference about \( N_\lambda x \) is based on this transformed and smoothed sample.

In scale space analysis we employ the smoother

\[
S_\lambda = (I + \lambda R_2)^{-1},
\]

that minimizes the penalized loss defined by \( R_2 \),

\[
S_\lambda w = \arg\min_{u} \{ ||w - u||^2 + \lambda u^T R_2 u \}.
\]

The matrix \( R_2 \) is a positive semidefinite matrix with the property that \( w^T R_2 w \) measures the 'roughness' of \( w \) by the variability of its discrete second partial derivatives. The matrix \( R_2 \) is defined in (8), where so-called Neumann boundary conditions are used. This smoother was used also in the original iBSiZer method [11].

Consider now the transformed and smoothed image \( N_\lambda x \) and denote by \( N_{\lambda,s}^x \) the intensity value at a pixel location \( s = 1, \ldots, n \). We detect changes by identifying those pixels \( s \) for which \( N_{\lambda,s}^x \) differs from zero with marginal posterior probability \( p(N_{\lambda,s}^x|y) \) that exceeds a given threshold \( \alpha \). Typically, \( \alpha = 0.95 \). At these pixels the difference, either positive or negative, is flagged as credible. In practice, the marginal posterior probability is computed as the proportion of positive or negative pixel values in the transformed sample. However, instead of such pixelwise (PW) inference we in fact prefer to perform simultaneous inference over all pixels of the image. The result of inference is then a global pattern of credible change at the level \( \lambda \) of smoothing, or spatial scale, and at the chosen level of posterior probability \( \alpha \). Here we use the simultaneous inference method of 'simultaneous credible intervals' (CI) that was first proposed for one dimensional data in [5] and then extended for digital images in [12, Appendix A.2]. The idea is to first select \( \Delta > 0 \) so that

\[
P \left( \max_{s \in \{1, \ldots, n\}} \left| \frac{N_{\lambda,s}^x - E(N_{\lambda,s}^x|y)}{Std(N_{\lambda,s}^x|y)} \right| \leq \Delta \middle| y \right) = \alpha
\]

and then define the set of jointly credible positive and negative pixel sets as

\[
\{ s \mid E(N_{\lambda,s}^x|y) - \Delta \text{Std}(N_{\lambda,s}^x|y) > 0 \}, \text{ and } \{ s \mid E(N_{\lambda,s}^x|y) + \Delta \text{Std}(N_{\lambda,s}^x|y) < 0 \}.
\]
Three image pairs were used to test the proposed change detection method, one consisting of artificial images and the other two of subimages of real Landsat ETM+ images. In the artificial example, we focus on the effect of noise heteroscedasticity as well as temporal dependence within a spectral band. The second real data example also considers the effect of a polynomial trend and an image prior that allows sharp edges along lake boundaries.

When changes are detected using differences of satellite images, it is crucial that the original images are properly aligned and that the image acquisition conditions are properly controlled. While a good vegetation index aims to mitigate the effect of external conditions, they still usually need to be controlled by applying some correction method to the images [4, 16]. A wide variety of methods exists for this purpose with some methods aiming at absolute correction while others just try to statistically set the conditions of the two images similar in order to make change detection as independent of the external conditions as possible. In our examples, we have used relative normalization based on simple linear regression adjusted, if necessary, to preserve known features, where the subject image is regressed against the reference image using least-squares fitting [26, 14].

In posterior sampling, some negative pixel intensities were observed. For some transformations, such as a difference of bands, this is not a problem. It is, however, unacceptable in case of an NDVI difference, at least if negative values are relatively common. To guarantee non-negativity, one could assume noise with a truncated normal distribution or use a multiplicative noise model with a log-normal distribution. If a truncated normal distribution is employed, posterior sampling should be based on iterative methods such as the Gibbs algorithm which would slow computations considerably. On the other hand, a multiplicative noise model is not commonly used with Landsat ETM+ images. A third alternative that allows direct posterior sampling is to assume that the observed image $y$ contains heteroscedastic Gaussian additive noise. The iBSiZer method for change detection allows such a noise model (cf. [11]) but then the posterior covariance matrix is no longer block circulant and fast Fourier methods cannot be used for computations. However, none of this is of concern here since negative simulated pixel values were exceedingly rare in our experiments. We therefore simply rounded intensity values smaller than $10^{-3}$ to the constant value $10^{-3}$.

### 3.1 Artificial test images

Our artificial test image pair is constructed from a small $176 \times 165$ subimage of a real Landsat ETM+ satellite image taken over eastern central Finland on August 2, 1999. The pixel size is $25m \times 25m$. As the observed satellite image can be assumed to contain noise, it was first smoothed to obtain the supposed noiseless band images $x_{13}$ and $x_{14}$. To obtain the images $x_{23}$ and $x_{24}$, we made manually some changes to $x_{13}$ and $x_{14}$. Finally, to obtain the noisy, 'observed images' $y_{13}, y_{14}, y_{23},$ and $y_{24}$, Gaussian iid noise with standard deviation 4 was added to each band. Only pixel values between $[0.5, 255.5]$ were allowed, and hence the added noise actually came from a truncated normal distribution and our model (1) is only an approximation. The band images, their time differences, as well as the corresponding NDVI images are shown in Figure 1. The NDVI difference of the true images, $N_x$, and of the noisy images, $N_y$, are presented in the last row of Figure 1.

Clearly, the noise in the NDVI difference image $N_y$ is heteroscedastic. The noise has large variance in the lake areas where the intensity in the underlying band images is low. Correspondingly, noise variance is smaller in areas where the underlying image intensity is higher.

While the difference of the NDVI images appears to be rather rough, the original underlying images themselves seem smooth. We therefore used the prior (6) with the second difference matrix (8) with toroidal
Figure 1: The artificial test image pair. Original Landsat ETM+ satellite image: Copyright 1999 ESA. 1st row: components of the true image $x$ for the two supposed instants of time for the bands 3 and 4. 2nd row: corresponding observed images. 3rd row: differences between times in bands 3 and 4 and the true NDVI images for the two times. 4th row: observed versions of the images on the third row. 5th row: true and observed NDVI difference images, respectively.
boundary conditions as $R_{ij}$ to specify prior dependence of pixels in an image $x_{ij}$ (cf. Appendices A and B). The temporal difference between the bands is nearly zero and we defined the temporal dependence matrices $\Theta_j$ also as second difference matrices with toroidal boundary conditions (cf. Appendices A and B).

As a large value of the parameter $\theta_j$ suppresses the temporal difference between bands and we actually try to make inferences about this very quantity, $\theta_j$ should be chosen rather small. We took $\theta_3 = \theta_4 = 0.05$. The parameters $\lambda_{ij}$ control prior smoothness of the underlying images $x_{ij}$ and we took $\lambda_{ij} = 0.1$ for all $i$ and $j$. Inferences about changes were based on a posterior sample of size 4000 and the results are shown in the three leftmost columns of Figure 2. Black and white are used to flag negative and positive credible change, respectively, and gray indicates no credible change. The level of credibility here and in the rest of the examples is $\alpha = 0.95$ (cf. Section 2.2). The pixelwise (PW) iBSiZer maps clearly show many false alarms, whereas the simultaneous maps based on the CI method detect the actual true changes without false alarms. None of the three considered smoothing levels alone reveals all true changes which justifies the scale space approach where a whole range of scales is analyzed simultaneously.

To evaluate the influence of temporal dependence in the image prior, we then assumed independence by setting $\theta_3 = \theta_4 = 0$. Inspection of the CI maps in the last column of Figure 2 reveals that now the large positive area at the center of the image is not fully detected at the smallest spatial scale. This is due to the lack of extra smoothing provided by temporal dependence. However, the small negative change in the middle of the image that was not detected with the previous model is now detected. Hence, it seems that when $\theta_3, \theta_4 > 0$, there is more power to detect large changed areas but small areas may be missed.

The marginal posterior standard deviations for these two models are shown in the first row of Figure 2. The posterior standard deviation is highest in the lake area where the underlying image intensities are lowest. The posterior standard deviation is also slightly higher for the model that assumes temporal independence.

The results seem to be rather robust with respect to the $\lambda_{ij}$ and $\theta_j$ parameters. If smaller values would be used for $\lambda_{ij}$, fewer credible features would be detected with small values of the scale space smoothing parameter $\lambda$ while increasing $\lambda$ would make all truly changed areas credible. If larger values of $\lambda_{ij}$ would be used, also the spatially larger but lower intensity areas are detected with small smoothing levels $\lambda$. However, the problem with too large $\lambda_{ij}$’s is that some of the spatially smallest features may then be smoothed out even with very small scale space smoothing.

However, the noise variances $\sigma^2_{ij}$ need to be handled with care as the results depend strongly on them. For example, with $\sigma^2_{ij} = 8^2$ the smallest scale true positive change in the middle of the images is only barely detected, whereas with $\sigma^2_{ij} = 2^2$ there are numerous false detections. In this example the true values of $\sigma^2_{ij}$ are known but when they are not, they along with the smoothing parameters can be set random or estimated e.g. using ML-method (c.f. Appendix D). Yet another approach is to estimate the $\sigma^2_{ij}$’s from the data as smoothing residuals, as we do in Section 3.3.

### 3.2 A Landsat test image pair

Our first real test image pair consists of $260 \times 280$ subimages of two Landsat ETM+ satellite images. The pixel values are expressed as digital numbers as opposed to calibrated reflectance values and each pixel represents a $25m \times 25m$ area on the ground. The first image of the pair is a piece of the same satellite image from which our artificial test images were constructed. The second image of the pair was obtained from the same location on May 29, 2002.

The original subimages are shown in Figure 3. The imaging conditions were different at the two times considered, especially in band 4. In order not to interpret atmospheric changes as real changes, the intensity values of time 2 band images were regressed on the corresponding values of time 1 images. The results,
Figure 2: Scale space analysis of the artificial test image. First row: The true and the noisy NDVI difference images and the posterior standard deviations of the two posterior models. Rows 2-4: 1st column: the posterior means of the smooths of the model assuming temporal dependence. 2nd column: pointwise iBSiZer maps, third column: simultaneous iBSiZer maps drawn using the CI method. 4th-5th column: posterior means of smooths and CI maps of the model assuming temporal independence. Black/white flags credible negative/positive change and gray means no credible change.
however, were unsatisfactory. By examining the scatter plot of the pixel values, we observed that band 3 actually did not need to be transformed at all and we adjusted manually the regression line of band 4 to better fit the data. The corrected image $\tilde{y}_{24}$ obtained is presented in Figure 3. The NDVI images and their difference image are displayed in Figure 4. Note that, just as for artificial test image, the noise variance in the NDVI images and in their difference seems heteroscedastic, being higher in the lake region. Note also the stripy artefacts in the lake area. These stripes originate from band 3. However, due to the low intensity in the lake area, the contrast of the images is not high enough to make these features visible in the original images.

Since the images seem rather smooth, we used a model with the smoothing prior (6) with the matrices $R_{ij}$ and $\Theta_j$ defined by (8) and using Neumann boundary conditions. Full Bayesian inference was applied with vague log-normal hyperpriors for all the model parameters: $\lambda_{ij} \sim \log-N(\log(0.1), 10^3)$, $\theta_j \sim \log-N(\log(0.05), 10^3)$, $\sigma_{ij}^2 \sim \log-N(\log(4^2), 10^3)$.

The posterior sample was drawn with hybrid Gibbs sampling. The satellite image $x$ given the other parameters was sampled from a multinormal distribution and the parameters were sampled from their full conditional distributions using Metropolis-Hastings steps. After a burn-in period of 1000 iterations, we kept every 4th sample until 3000 samples were obtained. Convergence was controlled by examining the trace plots of the parameters. The marginal posteriors were symmetric, and therefore summarized in Table 1 by their means and standard deviations.

The posterior mean and pixelwise standard deviation of the true NDVI difference image $N_x$ are shown in the first row of Figure 5. The posterior mean seems slightly smoother than the observed $N_y$. The marginal
posterior standard deviations of the pixels in the lake area are again higher than elsewhere. Three iBSiZer maps are presented in Figure 5. The sharpest changes are detected at the smallest scale and larger areas that have changed on average are detected at the higher smoothing levels.

To demonstrate the feasibility of handling larger data sets, we also analyzed a 800 × 800 subimage of this Landsat pair. The observed NDVI difference is displayed in the left-most image of Figure 6. The location of the smaller subimage analyzed above is marked with a rectangle. We used the same prior for x as before but with toroidal boundary conditions and fixed parameter values obtained as marginal posterior means from the previous analysis. A sample of 3000 images was drawn from the posterior distribution (cf. Appendix A). The key constraint on the computation is the size of the available computer RAM. Each sampled image was therefore saved to the hard drive. The posterior mean and the corresponding iBSiZer maps are shown in Figure 6.

3.3 Another Landsat test image example

Our second real data example consists of a pair of 300 × 218 Landsat ETM+ images taken over an area in the subarctic Norway (latitude 70°N) in 1986 and 2000. The images are shown in the first and second column of Figure 7. Before the images were made available to us they had been preprocessed by averaging the pixels over disjoint 4 × 4 neighborhoods of the original images. Each pixel then represents a 100m × 100m area on ground. The averaging reduced noise variance while keeping the assumptions about its Gaussianity and pixelwise independence still plausible. We eliminated the effect of changes in atmospheric conditions as
Figure 5: Scale space analysis of change in NDVI for the first ETM+ satellite image pair. 1st row: the observed difference of NDVI images, posterior mean of the true NDVI difference and the pixelwise posterior standard deviation. 2nd row: corresponding CI maps with smoothing levels $\lambda = 0, 1$ and 100.

Figure 6: Scale space analysis of the larger NDVI difference image. From left to right: The observed NDVI difference image and three iBSiZer maps. The area corresponding to the smaller Landsat images is marked with a rectangle.
in Section 3.2. The corrected images $\tilde{y}_{23}$ and $\tilde{y}_{24}$ are presented in the third column of Figure 7. The NDVI images and their difference are displayed in Figure 8.

Posterior modeling was tried using three different image prior models. The first uses the smoothing prior (6) with the matrices $R_{ij}$ and $\Theta_j$ defined by (8) and with equal values for the parameters $\lambda_{ij}$, $\sigma^2_{ij}$, and $\theta_j$. It also involves a second degree polynomial gradient. Full Bayesian approach was tried, but the results were unsatisfactory. We suspect that this is due to the coarser nature of the data caused by image preprocessing. However, when the image bands were assumed temporally independent, we were able to utilize an empirical Bayesian approach based on maximum likelihood estimates of the parameters. In principle, full Bayesian inference would also have been applicable, but because of computational constraints the empirical Bayesian approach was taken, instead. Hence, the second model assumes $\theta_3 = \theta_4 = 0$ while other parameters are independently estimated with maximum likelihood. The third model uses an edge preserving prior.

In the first model, just like in Section 3.1, the matrices $R_{ij}$ and $\Theta_j$ employed toroidal boundary conditions.
Figure 8: NDVI images of times 1 and 2 and their difference. In the gray-scale bar, the upper intensity values refer to the NDVI images, while the lower values pertain to the difference image.

Table 2: Maximum likelihood estimates for the model parameters

<table>
<thead>
<tr>
<th></th>
<th>$\lambda_{i3}$</th>
<th>$\lambda_{i4}$</th>
<th>$\sigma^2_{i3}$</th>
<th>$\sigma^2_{i4}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$i = 1$</td>
<td>$7.8 \cdot 10^{-3}$</td>
<td>$3.3 \cdot 10^{-3}$</td>
<td>6.1</td>
<td>7.0</td>
</tr>
<tr>
<td>$i = 2$</td>
<td>$1.4 \cdot 10^{-2}$</td>
<td>$4.1 \cdot 10^{-3}$</td>
<td>6.2</td>
<td>9.9</td>
</tr>
</tbody>
</table>

(cf. Appendices A and B). There is now a strong intensity gradient especially in band 3 and it turned out that the use of the toroidal boundary conditions required that the image prior model needed to include also a polynomial trend (cf. Appendix C). We took $\lambda_{ij} = 0.1$ and $\theta_{j} = 0.05$ for all $i$ and $j$ as these values appeared to produce reasonably good image reconstructions. For this model, the noise standard deviation was assumed to be equal in all four images. The value used was $\sigma_{ij} = 2.3$ and it was determined as follows. For each $i$ and $j$, a smooth $\hat{x}_{ij}$ was computed from $y_{ij}$ and the noise standard deviation $\hat{\sigma}_{ij}$ was estimated from the residual $y_{ij} - \hat{x}_{ij}$ over an area where $\hat{x}_{ij}$ seemed relatively smooth. The value 2.3 was then obtained as the average of the estimates $\hat{\sigma}_{ij}$. The posterior sample size was 3000. In order to demonstrate the smoothness of the posterior of $x$, the posterior mean of $x_{24}$ is displayed in Figure 9.

In the second model, we assumed that the images $x_{ij}$ are independent for all $i$ and $j$ and estimated the values of $\lambda_{ij}$ and $\sigma^2_{ij}$ using maximum likelihood (cf. Appendix D). The estimated parameter values are given in Table 2. Neumann instead of toroidal boundary conditions were applied (cf. Appendix B). As the top and the bottom of the image and its left and the right sides are no longer assumed to be neighbors, the spatial intensity gradient is not a problem and a spatial trend was therefore not included in the model. As can be seen from Figure 9, the posterior mean of $x_{24}$ is now much coarser.

Preprocessing by pixel averaging had made the images coarser and hence smoothness assumptions about $x$ less plausible. The intensity also changes rather abruptly at lake boundaries. As the third model, an edge preserving prior that smooths less at the lake boundaries was therefore tried in (6) with $R_{ij}$ and $\Theta_{j}$ defined by (9). The lakes were manually identified as follows. The images $y_{24}$ and $y_{14}$ were first truncated
using thresholds of 20 and 19, respectively. Then pixels were assigned to two segments by flagging as ‘lake’ all pixels whose intensity was below the threshold for both instants of time (see Appendix B). The prior parameter values used were $\gamma = 0.05$, and $\theta_j = 0.05$ and $\lambda_{ij} = 0.1$, $i = 1, 2$ and $j = 3, 4$. The noise standard deviations were $\sigma_{13} = 1.6$, $\sigma_{14} = 2.7$, $\sigma_{23} = 1.9$, $\sigma_{24} = 2.8$ and they were obtained in the same way as for the first model. Again, Neumann boundary conditions were used. As desired, the posterior mean of $x_{24}$ with this model is smooth but the lake boundaries remain sharp (Figure 9).

In Figure 10 the three rightmost panels in the first row show the pixelwise posterior standard deviations for the three models. The results of iBSiZer inference are illustrated in rows 2-4. As was the case with the artificial example, low intensity areas in the observed images $y_{ij}$ exhibit large posterior standard deviation. That is why the lakes that show as small brighter areas especially in std images of the first row do not show as credible in the iBSiZer maps. The clearest differences in the CI maps appear in the smallest scale.

4 Comparison with thresholding

In order to emphasize the importance of correctly handling the heterogeneity of noise in the NDVI difference images, we have analyzed the artificial test image of Section 3.1 also with simple thresholding often used with NDVI images [24]. In the first row of Figure 11 the reference change image, two iBSiZer maps and a thresholded image are presented. The smoothing levels of the iBSiZer maps are $\lambda = 1$ and $\lambda = 100$, respectively, and they were chosen so that the first map detects the sharpest features and the second map detects also the changed area in the lower right corner. The threshold used was 0.1647, chosen visually to fit the data as well as possible.

In the second row of Figure 11, the correctness of the iBSiZer maps are analyzed. The correctly detected changed areas are white and unchanged areas with no false alarms are gray. Areas with a false alarm are red and missed changes are blue. The few pixels where positive change is detected as negative or vice versa are colored magenta.

In the iBSiZer map with smoothing level 1, the changed areas are mostly correctly detected. However, the large area in the lower right corner is detected only partially. With the smoothing level 100, the area in the
Figure 10: iBSiZer analysis of NDVI difference images. First row: the observed difference of NDVI images and posterior marginal standard deviations of the second difference model, second difference model with maximum likelihood estimates and the edge preserving model, respectively. Rows 2-4: Posterior means and CI maps for the three models. The smoothing levels used are $\lambda = 0, 1,$ and 100. For more information see the text.
true changes  iBSiZer maps  thresholded

Figure 11: Comparison of iBSiZer maps and thresholding. 1st row: True change image, iBSiZer maps with smoothing levels $\lambda = 1$ and $\lambda = 100$, respectively, and a change map obtained by thresholding. 2nd row: maps indicating correctly and incorrectly detected pixels for the two iBSiZer maps and thresholding, respectively. White and gray signifies here correctly detected changed and unchanged pixels. Pixels that have undetected change are colored blue and falsely detected unchanged pixels are red. The few pixels where positive change is detected as negative or vice versa are colored magenta.
Table 3: Numerical comparison of the three maps in Figure 11. In the column names T=true, F=false, C=change, NC=no change. Thus, TC(TNC, respectively) shows counts of truly changed (truly unchanged) pixels correctly identified. FC (FCN, respectively) shows counts of changed (unchanged) pixels incorrectly identified. The $+\leftrightarrow -$ column gives the number of pixels with true positive change but flagged as negative change or vice versa, and the total F=FC+FNC.

<table>
<thead>
<tr>
<th>map</th>
<th>TC</th>
<th>TNC</th>
<th>FC</th>
<th>FNC</th>
<th>$+\leftrightarrow -$</th>
<th>total F</th>
</tr>
</thead>
<tbody>
<tr>
<td>CI, $\lambda=1$</td>
<td>2983</td>
<td>23864</td>
<td>192</td>
<td>2001</td>
<td>0</td>
<td>2193</td>
</tr>
<tr>
<td>CI, $\lambda=100$</td>
<td>4569</td>
<td>22277</td>
<td>1779</td>
<td>396</td>
<td>19</td>
<td>2175</td>
</tr>
<tr>
<td>Thresholding</td>
<td>596</td>
<td>20979</td>
<td>3077</td>
<td>4386</td>
<td>2</td>
<td>7463</td>
</tr>
</tbody>
</table>

lower right corner is also detected but the extent of the other changed areas are then overestimated. Note that there are no isolated false alarm blobs in either of the iBSiZer maps. In the thresholded map most of the changed areas are not detected and the lake area in the middle of the image is filled with false alarms. With a higher threshold, fewer false alarms would appear but also fewer true changes would then be detected. Correspondingly, with a lower threshold more true changes would be detected but the number of false alarms would also be higher. Therefore, changing the threshold would not improve the results considerably. These results are also summarized numerically in Table 3.

We also tried the hypothesis-test-based method proposed in [24]. The unchanged pixels are then assumed to have a bivariate Gaussian distribution. We used the method based on conditional distributions, where the changed pixels can be identified as the pixels with a small p-value of time 2, under the null hypothesis of no change given the value at time 1. The results (not shown) were not better than with simple thresholding.

5 Summary

The iBSiZer method, originally designed for the analysis of differences between digital images, was extended for the detection of land cover changes in pairs of Landsat ETM+ multispectral images. The method detects both small and large scale changes, provides a statistical interpretation of the pattern of changes and it is able to detect changes also in nonlinear transformations of the original data, such as in NDVI images. The performance of the method was illustrated using three pairs of test images, the first one artificial and two real ones, one taken over eastern Finland and the second taken over subarctic Norway. For the artificial images, the tests highlighted the importance of a good noise model and the effect of temporal dependence in the form of extra smoothing was noted.

For the first real image pair, we used the fully Bayesian approach treating the model parameters as random variables. We also demonstrated the feasibility of our method for larger images by analyzing a bigger part of the original satellite image. Still larger images can be analyzed in overlapping blocks.

The second pair of real satellite images contained a spatial trend which was therefore also considered as a possible component of the image prior model. Preprocessing by pixel averaging had made the images coarser especially at the lake boundaries. This motivated testing also an edge preserving prior. We observed that the biggest differences between the models appear in the smallest scale.

All computations were carried out on a Dell OptiPlex 9010 PC with Intel Core i7-3770 CPU and 32Gb of RAM using Matlab 2012B under the Windows 7 64-bit operating system. The computational burden can be heavy for the kind of simulation based modeling approach and the type of image data considered here. This was alleviated by the possibility to sample from Gaussian posteriors and the speed-ups obtained from Fourier methods. For example, in the case of the $176 \times 165$ artificial images, a posterior sample of size 4000

18
can be generated in about 30 seconds and all three reported maps were drawn in 15 seconds. However, for the model using an edge preserving prior, generation of a posterior sample of size 3000 took 3.5 minutes for the 300 × 218 Landsat image pair and for the model with a prior intensity gradient, 2.5 minutes was required. The three CI maps were drawn in 25 seconds. The hybrid Gibbs sampling for the first real satellite image took 21 hours. For the 800 × 800 large test image, computations of the iBSiZer maps together with posterior sampling took 8 hours. The Matlab codes used are available at http://cc.oulu.fi/~lpasanen/iBSiZer/ChangeDetection/.

For the artificial test images, we also tested the performance of simple thresholding but observed that, predictably, it was unable to cope with noise heteroscedasticity in the NDVI difference image. False detections were made in the lake areas where the noise was most pronounced. When the threshold was set so high that false alarms were avoided, most true changes were missed. This underlines the need for an adequate model for the NDVI difference image distribution. Overall, the results reported in this paper suggest that the proposed method has potential as a tool for change detection in remote sensing, especially for the vegetation index and other images that involve a nonlinear transformation of the original satellite images.

Acknowledgements

We are grateful to Erkki Tomppo and Kai Mäkisara from the Finnish Forest Research Institute for the Landsat images used in the first two examples and to Miska Luoto from the Department of Geosciences and Geography of the University of Helsinki for the image pair used in the third example.

References


A The Gaussian posterior distribution

Assuming the independence of noise within each band image and between them, the likelihood \( y \) given \( x \) and the noise variances is

\[
p(y|x, \{\sigma_{ij}^2\}) \propto \prod_{i=1,2, j=3,4} \sigma_{ij}^{-n} \exp \left( -\frac{1}{2\sigma_{ij}^2} |y_{ij} - x_{ij}|^2 \right).
\]

The prior \( p(x) \) models dependence both within each \( x_{ij} \) and between different \( x_{ij} \)'s. Within each \( x_{ij} \), near-by pixels can be assumed to depend on each other more than pixels further apart. Dependence is also allowed between pixels of the images \( x_{ij} \) and \( x_{2j} \) that correspond to the same band observed at different times. These assumptions are captured by the Gaussian prior

\[
p(x|\lambda, \theta) \propto \exp \left( -\frac{1}{2} \sum_{i=1,2, j=3,4} \lambda_{ij} x_{ij}^T R_{ij} x_{ij} + \sum_{j=3,4} \theta_j (x_{1j} - x_{2j})^T \Theta_j (x_{1j} - x_{2j}) \right),
\]

where \( \lambda = \{\lambda_{13}, \lambda_{14}, \lambda_{23}, \lambda_{24}\}, \theta = \{\theta_3, \theta_4\}, \lambda_{ij}, \theta_j > 0 \) for all \( i \) and \( j \), and the structure of the matrices \( R_{ij} \) and \( \Theta_j \) that model spatial dependence within \( x_{ij} \) and temporal dependence between \( x_{1j} \) and \( x_{2j} \), respectively, is discussed in Appendix B. When smoothing is preferred, the matrix \( R_{ij} \) is constructed to penalize for large differences between neighboring pixels and if there is a desire to preserve edges in the image, a different kind of matrix is needed. Setting \( \Theta_j \) as the identity matrix means that little temporal change is expected within band \( j \). If, on the other hand, \( \Theta_j \) is defined to cause smoothing, temporal differences are allowed in large smooth areas but sharp differences are penalized.

The prior (6) can be written in a more compact form as

\[
p(x|\lambda, \theta) \propto \exp \left( -\frac{1}{2} x^T Q(\lambda, \theta) x \right),
\]

where

\[
Q(\lambda, \theta) = \begin{pmatrix}
Q_{13} + Q_3 & 0 & -Q_3 & 0 \\
0 & Q_{14} + Q_4 & 0 & -Q_4 \\
-Q_3 & 0 & Q_{23} + Q_3 & 0 \\
0 & -Q_4 & 0 & Q_{24} + Q_4
\end{pmatrix},
\]

and \( Q_{ij} = \lambda_{ij} R_{ij}, Q_j = \theta_j \Theta_j \). To simplify the notation, we will denote \( Q := Q(\lambda, \theta) \).

With the prior (6) and the Gaussian likelihood (5) the posterior density of \( x \) is

\[
p(x|y) \propto \exp \left[ -\frac{1}{2} (x^T Q x + (y - x)^T \Sigma^{-1} (y - x)) \right],
\]

where

\[
\Sigma = \text{diag}[\sigma_{i1}^2, \sigma_{i4}^2, \sigma_{23}^2, \sigma_{24}^2].
\]

By completing the square with respect to \( x \) we have

\[
x|y \sim N(\mu, \Sigma_{\text{post}}),
\]

where

\[
\mu = (I + \Sigma Q)^{-1} y, \quad \Sigma_{\text{post}} = (\Sigma^{-1} + Q)^{-1}.
\]

Computing is faster if the random variables \( x_{ij} \) are assumed to be isotropic, have equal variances \( \sigma_{ij}^2 \), and the matrices \( Q_{ij} \) and \( Q_j \) are equal for all \( i \) and \( j \). This is because then the structure of both the noise covariance matrix \( \Sigma \) and the matrix \( Q \) is hierarchically block circulant and therefore fast Fourier methods can be used to obtain their eigenvalue decompositions [22, Chapter 2.6].
B  Image priors

Here we concentrate on modeling the dependency between pixels in one satellite image band and therefore denote \( x_{ij}, R_{ij}, \) and \( \lambda_{ij} \) simply by \( x, R, \) and \( \lambda_0, \) respectively. The prior of \( x \) is then

\[
p(x|\lambda_0) \propto \exp\left( -\frac{\lambda_0}{2} x^T R x \right).
\]

(7)

We write \( s \sim t \) for two pixel locations \( s \) and \( t \) if they are neighbors. An interior pixel location \( s = (l, k) \) has four neighbors, \((l - 1, k), (l + 1, k), (l, k - 1)\) and \((l, k + 1)\). For \( R \), we consider \( R_2 \) defined by

\[
x^T R_2 x = \| C x \|^2 = \sum_t \left( \sum_{s \sim t} (x_s - 4x_t)^2 \right)^2,
\]

(8)

where \( C \) can be interpreted as the discrete Laplace operator. This prior penalizes for image roughness as measured by the second differences of neighboring pixel intensities and the level of penalty is controlled by \( \lambda_0 \). In order to have four neighbors for the boundary pixels too, we modified \( R_2 \) in two different ways. In the first modification, the boundary values themselves are extended beyond the actual image to produce the required neighbors. The second modification is to assume that \( R_2 \) is a circulant matrix. This means that the left and the right edges of the image are neighbors and a pixel \((M, k)\) on the bottom edge is a neighbor of the pixel \((1, k + 1)\) on the top edge. With these assumptions the image topologically resembles a torus. We refer to these two modifications as Neumann and toroidal boundary conditions, respectively. The \( n \times n \) matrix \( R_2 \) is symmetric and positive semidefinite and with these boundary conditions its rank is \( n - 1 \). The null space of \( R_2 \) consists of constant images and therefore these priors are sensitive only to the intrinsic variation of the image pixels relative to their mean, not the absolute level of the pixel values. Although the function (7) is not a proper density, it can be used to express our belief about the smoothness of the image \( x \). Such a ‘smoothing prior’ is an example of an intrinsic Gaussian Markov random field [22, Chapter 3].

When the image is known to contain edges, instead of a smoothing prior, an edge preserving prior may be needed, (e.g. [25]). This will usually lead to a non-Gaussian posterior and slow Metropolis-Hastings simulation. In [15] a Gaussian prior has been proposed which is able to preserve sharp intensity variation between image segments if the edges of the segments are known. Based on their ideas we propose to define matrix \( R \) as follows. Suppose that the image has been divided into finitely many segments and denote by \( I_s \) the segment that the pixel \( s \) belongs to. The matrix \( R \) is then defined by

\[
x^T R x = \sum_{s \sim t} \gamma_{st} (x_s - x_t)^2,
\]

(9)

where

\[
\begin{align*}
\gamma_{st} &= \gamma, & \text{if} & \ I_s \neq I_t, \\
\gamma_{st} &= 1, & \text{if} & \ I_s = I_t,
\end{align*}
\]

and \( 0 < \gamma \leq 1 \). The corresponding prior (7) does smooth edges between segments but, if \( \gamma < 1 \), not as strongly as the interiors of the segments. The parameter \( \gamma \) controls the level of dependence between neighboring pixels that lie in different segments. In the example considered in 3.3, pixels can belong only to two different segments, lake or non-lake. In other applications more segment types could be defined and different values of \( \gamma \) could be used between them.

C  A model with an intensity gradient

The prior model (6) assumes that each \( x_{ij} \) is an intrinsic Gaussian random field that evolves around the mean value. Further, with toroidal image boundary conditions needed for fast Fourier methods, the left and
the right edges as well as the upper and lower edges are assumed to be similar. Such a model may not be suitable in the presence of large scale features such as intensity gradients. Then we propose an image model

\[ y = x + Z\beta + \epsilon, \]

where \( x \) contains the small scale features modeled as a Gaussian random variable (6), \( Z\beta \) contains the large scale features modeled as a polynomial trend, and \( \epsilon \) contains the noise. In our experiments reported in Section 3.3 we assumed that the trend \( Z\beta \) at a pixel \((k, l)\) in the image \( y_{ij} \) is of the polynomial form

\[ f_{ij}(k, l) = a_{ij}k + b_{ij}l + c_{ij}kl, \quad k = 1, \ldots, M, \ l = 1, \ldots, N, \tag{10} \]

where \( a_{ij}, b_{ij} \) and \( c_{ij} \) are constants. Hence, the design matrix \( Z \) is a block diagonal matrix

\[ Z = \text{diag}(z, z, z), \]

where

\[ z = [1_N \otimes m, n \otimes 1_M, (1_N \otimes m) \otimes (n \otimes 1_M)], \]

\[ 1_N = [1, \ldots, 1]^T, \ 1_M = [1, \ldots, 1]^T, \ m = [1, 2, \ldots, M]^T, \ n = [1, 2, \ldots, N]^T, \ \otimes \text{ denotes elementwise multiplication. Further, } \beta = [\beta_1^T, \beta_2^T, \beta_3^T, \beta_4^T]^T \text{ is a random vector with } \beta_{ij} = [a_{ij}, b_{ij}, c_{ij}]^T, \text{ and } a_{ij}, b_{ij}, c_{ij} \text{ are as in (10)}. \]

For \( \beta \) we will use an uninformative prior \( \beta \sim [1, \ldots, 1]^T \). Then the joint posterior of \( x \) and \( \beta \) is

\[ p(x, \beta | y) \propto \exp \left[ -\frac{1}{2} (x^T Q x + (y - x - Z\beta)^T \Sigma^{-1} (y - x - Z\beta)) \right]. \]

A posterior sample from \( p(x + Z\beta | y) \) can now be drawn without having to resort to slow iterative procedures by first drawing \( \beta \) from its marginal posterior distribution \( p(\beta | y) \) and then \( x \) from the conditional posterior \( p(x | \beta, y) \). The distributions are

\[ \beta | y \sim N \left( (Z^T Q \beta Z)^{-1} Z^T Q \beta y, (Z^T Q \beta Z)^{-1} \right), \]

\[ x | \beta, y \sim N \left( (I + \Sigma Q)^{-1} (y - Z\beta), (\Sigma^{-1} + Q)^{-1} \right), \]

where \( Q\beta = -(\Sigma^{-1})^T (\Sigma^{-1} + Q)^{-1} \Sigma^{-1} + \Sigma^{-1} \).

### D Maximum likelihood estimation for independent images

Here we consider maximum likelihood (ML) estimation of the parameters for the model proposed in Appendix A assuming that the images are independent. Thus, \( \theta_j = 0 \) for \( j = 3, 4 \). As we estimate the parameters \( \lambda_{ij} \) and \( \sigma_{ij}^2 \) for each image \( ij \) separately, we will drop the subscripts as we did in Appendix B.

The maximum likelihood estimates of \( \lambda \) and \( \sigma^2 \) are the zeros of the gradient of the likelihood function. For the mathematical details, we refer to [11] where the ML estimation of \( \lambda \) is considered in the case of a general \( \Sigma \). As the estimates must be positive, we actually maximized here the likelihood function relative to the logarithms of the parameters by numerically solving the equations

\[ \frac{\partial l}{\partial (\log \lambda)} = \left( \frac{(N - 1)}{2\lambda} - \frac{1}{2} \text{tr}(SR) - \frac{1}{2\sigma^2} \text{tr}(Sy)R(Sy) \right) \lambda = 0, \]

\[ \frac{\partial l}{\partial (\log \sigma^2)} = \left( -\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4} \left( \text{tr}(S) + y^T y \right) - \frac{1}{\sigma^6} y^T Sy + \frac{1}{2\sigma^8} y^T S^2 y \right) \sigma^2 = 0, \]

where \( S = (\sigma^{-2} I + \lambda R)^{-1} \). If a second differences prior with Neumann or torus boundary conditions is used, Fourier methods can be used to accelerate computations considerably [11].