A statistical approach to the problem of restoring damaged and contaminated images

Article

Accepted Version

A statistical approach to the problem of restoring damaged and contaminated images

R. G. Everitt[a] and R. H. Glendinning[b] *
QinetiQ ltd,
Great Malvern,
Worcestershire, WR14 3PS, UK

February 22, 2010

ABSTRACT

We address the problem of automatically identifying and restoring damaged and contaminated images. We suggest a novel approach based on a semi-parametric model. This has two components, a parametric component describing known physical characteristics and a more flexible non-parametric component. The latter avoids the need for a detailed model for the sensor, which is often costly to produce and lacking in robustness. We assess our approach using an analysis of electroencephalographic images contaminated by eye-blink artefacts and highly damaged photographs contaminated by non-uniform lighting. These experiments show that our approach provides an effective solution to problems of this type.

Keywords: Bayesian statistics; Damaged images; EEG artefacts; illumination variations; photographs; Semi-parametric model

1 Introduction

Low cost sensors are making the routine collection of images in non-laboratory conditions a practical reality. Typical automatic decision support applications relying on image data include document processing [1], archiving photographic material [2], object recognition from infra-red images [3] and medical diagnosis [4] using electroencephalography (EEG), medical imaging techniques based on sparsely sampled frequencies [5] and projections [6].

Our focus is on the automatic processing of damaged and contaminated images which are ubiquitous features of image capture in non-laboratory conditions.

*Corresponding author. Tel.: +44 (0) 7 990 667344. [a]rgeveritt@qinetiq.com (R. G. Everitt), [b]rhglendinning@yahoo.co.uk (R. H. Glendinning).

These have a wide variety of characteristics and generating mechanisms, which often preclude detailed modelling due to cost considerations and drift in sensor characteristics. Typical examples of images of this type include

- *Damage*. This category includes drop outs and outliers. These are generated by a wide variety of mechanisms, with *drop outs* generated by display monitor faults [7] and abrasion on photographic material [2] [7], with calibration errors and impulsive noise generating outliers (anomalous pixel values) in a wide range of sensor types.

- *Contamination*. A typical example of contamination with relatively well defined structure is an eye blink artefact in EEG images. These are generated by the contact of an eyelid with the positively charged cornea which generates an electrical potential field on the scalp which decreases rapidly with the distance from the eye. Their presence can result in significant reductions in the performance of subsequent procedures [8]. Other mechanisms include vibration induced blur [9] and certain non-uniform lighting effects [10]. Other sources of contamination have effects which are less well defined, but often exhibiting local smoothness. These include atmospheric scattering and absorption of radiation over several frequencies using hyperspectral sensors [11].

  Some contamination mechanisms depend on a vector of explanatory variables (known as covariates) often including *time*. These include the orientation of the sun in infra-red imaging and the degree of scattering and absorption in hyper- spectral images [11] which vary with atmospheric conditions.

We propose a means of recovering the underlying image from damaged and contaminated data using a spatial semi-parametric model. This describes the value of an image $Y(x,z)$ at the location $s=(x,z)$ by

$$Y(x,z) = f(\beta,x,z) + h(x,z) + Z(x,z), \tag{1}$$

where $Z(x,z)$ describes independent and identically distributed white noise with zero mean and variance σ^2. This model decomposes the expected value of $Y(x,z)$ into a non-linear component $f(\beta,x,z)$ with known (up to a small number of parameters) non-linear function and locally smooth component $h(x,z)$.

A vector of covariates can be included into the parametric component to include the effect of spatially varying covariates $\alpha(x,z)$ to give $f(\beta,x,z,\alpha(x,z))$. While they are not used in this study, they are easily incorporated into our approach and provide an important means of incorporating additional information. A typical example is the detection of land mines from infra-red images. Here the parametric component $f(\beta,x,z,\alpha(x,z))$ describes the temperature generated by a land mine at (x,z), where β is a vector with the components (location of the

centre of the mine, depth, mine orientation) and $\alpha(x, z)$ describing the orientation of the land surface determined from map information, say. The locally smooth component $h(x, z)$ describes temperature measurements over the ground.

In contrast to much recent work on recovering sparse signals [12], we adopt a fully Bayesian approach which is used to specify our prior knowledge about the nature of the parameters describing the signal and noise characteristics. The former including a constraint based on continuity which essentially ensures that high frequency Fourier coefficients are small. The Bayesian approach allows uncertainty in the parameter values and form of the model to be propagated in subsequent inferential procedures.

We demonstrate the value of models of this type in two practically important problems without covariate effects to simplify the presentation. The first is described in section 6 concerned with EEG potentials over the scalp. These are often contaminated by eye-blink artefacts which are modelled using the non-linear parametric component of a semi-parametric model, with the instantaneous EEG potentials describe by the locally smooth component (without a model for propagation through the skull). The second example is described in section 7 and is concerned with damaged photographs with high levels of drop outs and non-uniform illumination effects. The underlying image described by the locally smooth components with the non-uniform illumination effect described by the non-linear component.

We see from these examples that semi-parametric models exploit the advantages of locally smooth models (often referred to as non-parametric) and parsimonious non-linear models. As a wide range of characteristics can be modelled by the locally smooth component, this removes the need for a detailed physical model for the sensor output, which can be costly to produce and lacking in robustness to drift. The locally smooth component provides a means of modelling the output from drifting sensors, as its mathematical form is not specified in advance. Locally smooth components are also of value in modelling contaminated sensor output. A typical example is the frequency response of a pixel in hyperspectral imaging which is often contaminated by a locally smooth component which varies with atmospheric conditions [11]. For other examples of the use of semi-parametric models for scalar signals, see [13] or [14].

2 The Bayesian paradigm

The Bayesian paradigm provides a means of incorporating prior knowledge about likely values of the parameters describing the data generating mechanism into the analysis. Prior knowledge is captured by a joint probability distribution associated with the parameters of the model describing the data generating mechanism. Prior knowledge is modified by the data using Bayes theorem to give a (posterior) distribution which may be used to generate (point) estimates of the model parameters.

This paradigm is applied to semi-parametric models by parameterizing the locally smooth component $h(x, z)$ in an appropriate manner. This is a non-trivial issue as locally smooth functions are essentially infinite-dimensional (non-parametric). A range of techniques [15] have been developed for this purpose, with significant effort applied to *geo- additive* models. These have several non-parametric terms (possibly depending on spatial position [16]) and are observed at irregularly sampled points $s_i = ((x_i, z_i), i = 1, \ldots, n)$ in the two-dimensional domain S.

Two issues are addressed in this study. These are meaningful ways of describing *a priori* knowledge about the smoothness of $h(x, z)$ and the inclusion of non-linear components in a computationally feasible manner. We propose an approach based on the coefficients of the Fourier representation of $h(x, z)$ and posterior distributions generated using a Markov-chain Monte-Carlo (MCMC) sampler. This is used to generate point estimates of the parametric and non-parametric components of our model (1) using sample estimates of the expected value of the posterior distribution associated with their parameters.

The Bayesian paradigm has significant advantages over classical methods (such as surface fitting by least squares), as it can explicitly incorporate prior knowledge about the parameters describing the data generating mechanism. This is exploited in the example in section 6, where EEG images are subject to contaminated by eye-blink artefacts. Our second example is described in section 7 and describes a means of reconstructing damaged photographs contaminated by the presence of spatially varying illumination. Here the Bayesian paradigm provides a means of incorporating prior knowledge about the characteristics of the underlying image. The Bayesian paradigm also allows uncertainty about parameter values and the structure of the model to be propagated in subsequent inferential operations.

3 Priors for the locally smooth component

We describe *a priori* assumptions about the smoothness by regarding $h(x, z)$ as a realization from a Gaussian random field on S with zero mean. The smoothness of $h(x, z)$ is controlled by the spatial covariance kernel defining this process. A wide range of procedures are used to describe suitable covariance kernels. These can be defined using ideas from *Kriging* (see [17] for a non-Bayesian perspective) or indirectly using a suitably parameterized model. The latter include partial differential equations (PDE), Markov models which describe the distribution of $h(x, z)$ in terms of their spatial neighbours [16] and the use of basis function expansions.

Our focus is on the use of the latter approach which provides a simple means of dealing with non-uniformly sampled functions (images damaged by drop-outs). In addition, some studies [18] suggest that this approach out-performs a range of competitors, although the emphasis of these experiments is on non-Bayesian

techniques. Formally

$$h(x,z) = \sum \theta_j \phi_j(x,z), \qquad (x,z) \in S, \quad j = (j_1, j_2) \tag{2}$$

where $(\phi_j(x,z))$ is a collection of basis functions. Various approaches differ in their choice of basis functions and the way in which the smoothness of $h(x,z)$ is described. Significant work is based on spatially localized basis functions. These include $B-$splines [19] [20] and radial basis functions [21]. Smoothness constraints are typically imposed by limiting the number of basis functions, where [21] [22] and [23] focus on free knot splines and [24] [20] on their fixed knot analogues, although sparse basis techniques of this type are computationally costly [24].

Other approaches use a probabilistic model for the spatially distributed coefficients (θ_j). These priors are generated by the conditional distribution of $\theta_j|(\theta_k, k \in \delta_j)$ in terms of its neighbours $(\theta_k, k \in \delta_j)$, which specify the spatial characteristics [25] of the underlying Gaussian process $h(x,z)$. These are two-dimensional analogues of the usual auto-regressive or random walk priors on the real-line and generate Bayesian analogues of $p-$Splines in [19] [26]. The first order smoothness prior for observations on a regular lattice $X \in S$ is

$$\theta_j|(\theta_k, k \in \delta_j) \sim N(\sum_{k \in \delta_j} \theta_k, \tau^2), \qquad k = (k_1, k_2) \tag{3}$$

where $N(., \tau^2)$ is a normal distribution with conditional variance τ^2. This prior can be written in the form

$$exp(-\frac{1}{\tau^2}\theta' W \theta), \quad \theta = (\theta_j)^T \tag{4}$$

which emphasizes the relationship with classical splines [27]. Priors of this form can be modified [28] [29] to deal with spatially varying smoothness by replacing τ^2 with a spatially varying alternative (denoted by τ_j^2) with appropriate hyper-parameters. This generates a long tailed prior for $\theta_j|(\theta_k, k \in \delta_j)$.

We incorporate prior knowledge about the smoothness of $h(x,z)$ using the characteristics of its Fourier expansion $g(x,z)$. This basis function expansion provides a simple means of implicitly describing the covariance kernel of the Gaussian process generating $h(x,z)$ through the rate of decay of its coefficients. As we assume that $h(x,z)$ has zero mean

$$g(x,z) = \sum_{k=1}^{\infty}\sum_{j=1}^{\infty} \theta_{kj}\phi_k(x)\phi_j(z) + \sum_{k=1}^{\infty} \theta_{k0}\phi_k(x) + \sum_{j=1}^{\infty} \theta_{0j}\phi_j(z) \tag{5}$$

over $(x,z) \in S = [a,b] \times [c,d]$, where

$$\phi_k(x) = \left(\frac{2}{b-a}\right)^{1/2} \cos\left\{\pi k \left(\frac{x-a}{b-a}\right)\right\}, \qquad \text{for } k \geq 1 \tag{6}$$

and

$$\theta_{kj} = \int_c^d \int_a^b g(x,z)\phi_k(x)\phi_j(z)dxdz, \quad \theta_{k0} = \int_c^d \int_a^b g(x,z)\phi_k(x)dxdz \tag{7}$$

for $k \geq 1$ in the usual way. Similar expressions describe $\phi_j(z)$ and θ_{0j}. It is easy to see that the rate of decay of the coefficients in (16) is closely related to the smoothness of $h(x,z)$ by noting that

$$||\frac{\delta^p \delta^q g}{\delta x^p \delta z^q}||^2 < \infty \Rightarrow \theta_{kj} = o((kj)^{-(p+q)}), \quad k,j \to \infty \tag{8}$$

where $||.||$ is the usual norm on S. We propose priors of the form

$$\theta_{kj} \sim N(0, \tau^2 \exp(-\gamma c_{kj})) \tag{9}$$

where $N(.,.)$ is a Gaussian density. This prior imposes constraints on the magnitude of the Fourier coefficients, rather than their joint distribution which are though to pose problems due to their irregular nature [12]. We impose prior constraints on the magnitude of $h(x,z)$ using τ^2 and its smoothness by γ. We use a novel two-dimensional generalization of [30] and put

$$c_{kj} := c_j + c_k \tag{10}$$

where ($c_j = j$ and $c_k = k$) for a *geometric smoother* or ($c_j = \log(j)$ and $c_k = \log(k)$) for an *algebraic smoother*. We use an inverse Gaussian (IG) prior

$$\tau^2 \sim \text{IG}(u_0/2, v_0/2) \tag{11}$$

and an exponential prior

$$\gamma \sim \text{E}(1/w_0) \tag{12}$$

for γ with mean $1/w_0$ for a geometric smoother or

$$\gamma \sim 1 + \text{E}(1/w_0) \tag{13}$$

for the algebraic smoother. It is easy to see that $g(x,z)$ converges almost surely for $\gamma > \eta$, where $\eta = 0$ for the geometric smoother or $\eta = 1$ for the algebraic case. Other patterns can be imposed on c_{kj} to incorporate assumptions about directional effects in the frequency domain. All of our the experiments are based on the algebraic smoother with problem specific choices of the hyper-parameters u_0, v_0, w_0 which encapsulate our prior knowledge about the properties of $h(x,z)$..

4 Semi-parametric regression with a locally smooth component

We propose a Bayesian approach to the problem of fitting a semi-parametric model to spatial data. We describe the key steps in our approach in this section for an algebraic smoother for clarity, with obvious changes for other smoothers.

- First we specify the priors for the vector of parameters β describing the parametric component of (1). We assume that

$$\beta \sim \mathrm{N}(b_0, B_0) \tag{14}$$

is a multivariate normal distribution with mean vector b_0 and covariance matrix B_0, although other distributions can be used where appropriate. The values of b_0 and B_0 are selected in a problem specific manner, with b_0 describing the most likely value for β with B_0 describing our uncertainty. Non-informative prior can be used to describe our prior knowledge about β, see [30]. Priors for the parameters describing spatial covariates $\alpha(x, z)$ are added at this stage.

- Specify the prior for the variance σ^2 of the additive noise. We use the usual inverse Gaussian (IG) prior

$$\sigma^2 \sim \mathrm{IG}(r_0/2, s_0/2) \tag{15}$$

with problem specific parameters r_0 and s_0. These parameters describe the magnitude and uncertainty in our prior knowledge.

- Choose the number of basis functions used to describe the Fourier expansion $g(x, z)$ and equivalently $h(x, z)$. Here

$$g(x,z) = \sum_{k=1}^{N}\sum_{j=1}^{M}\theta_{kj}\phi_k(x)\phi_j(z) + \sum_{k=1}^{N}\theta_{k0}\phi_k(x) + \sum_{j=1}^{M}\theta_{0j}\phi_j(z), \tag{16}$$

giving $K = (M+1)(N+1) - 1$ Fourier coefficients. The values of M and N can be selected in a number of ways. One approach is to ensure that the semi-parametric model fits the data exactly with smaller numbers of basis functions used for smooth functions. This can reduce the computational costs associated with our approach.

- Impose our prior knowledge about the smoothness of $h(x, z)$ by specifying the likely decay in the Fourier coefficients θ by choosing the values of u_0, v_0 and w_0. Suitable values for these parameters are suggested in our simulation experiments and used in our eye blink and damaged image examples, with one dimensional functions considered in [30]. We assume that the prior distributions of β and θ are independent to encourage separation of the locally smooth and parametric component of our model.

We carry out posterior inference using the full posterior distribution $P(\beta, \sigma^2, \theta, \tau, \gamma \mid Y)$ of the parameters given the data Y, where $Y = (y_i, i = 1, \ldots, n)^T$, $y_i = Y(x_i, z_i)$, $(x_i, z_i) \in X \subseteq S$ and $\theta = (\theta_{jk}, j = 1, \ldots, N; k = 1, \ldots, m)$. This is generated by

$$P(\beta, \sigma^2, \theta, \tau, \gamma \mid Y) \propto P(Y, \beta, \sigma^2, \theta, \tau^2, \gamma) \tag{17}$$

with

$$P(Y, \beta, \sigma^2, \theta, \tau^2, \gamma) = P(Y \mid \beta, \sigma^2, \theta)P(\beta)P(\sigma^2)P(\theta \mid \tau^2, \gamma)P(\tau^2)P(\gamma) \quad (18)$$

which follows from our assumptions. The likelihood is given by

$$P(Y \mid \beta, \sigma^2, \theta) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp(-R^T R/2\sigma^2) \quad (19)$$

where $R = (r_i, i = 1, \ldots, n)^T$ and $r_i = y_i - f(\beta, x_i, z_i) - h(x_i, z_i)$. Next

- Select the number of iterations (samples) used to *burn in* the MCMC sampler. The aim is to ensure that it has *forgotten* its initial values.
- We draw realizations from the full posterior (17) using the MCMC sampling scheme described in the appendix until a convergence diagnostic is satisfied, see [31].
- The posterior sample means associated with β and θ are used to give point estimates of our semi-parametric model.

The extension of this algorithm to deal with uncertainties in the mathematical form of the parametric component $f(\beta, x, z)$ is described in our eye blink example. Our sampling scheme is immediately applicable to images in higher-dimensional images, an advantage of our approach.

5 Performance assessment using simulated data

We assess the performance of our approach using a range of semi-parametric models with components exhibiting different characteristics. These are recovered from observations at regular or irregularly distributed design points $X = ((x_i, z_i), i = 1, \ldots, n)$ using different values of n to emulate the effect of image damage. The effect of un-controlled noise levels is assessed by repeating our experiments for different values of σ^2. The sensitivity of our approach to changes in the priors is addressed in the same way.

We use the familiar R^2 statistic

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{f}(\beta, x_i, z_i) - \hat{h}(x_i, z_i))^2}{\sum_{i=1}^{n}(y_i - \bar{y}^2)}, \quad \bar{y} = \frac{i}{n}\sum_{i=1}^{n} y_i, \quad (20)$$

as a performance measure, where $\hat{f}(\beta, x_i, z_i)$ and $\hat{h}(x_i, z_i)$ are the posterior estimates of parametric and non- parametric components of our semi-parametric model. This statistic is averaged over five reconstructions to assess the convergence of our MCMC sampler. High values of the resulting statistic suggest good performance.

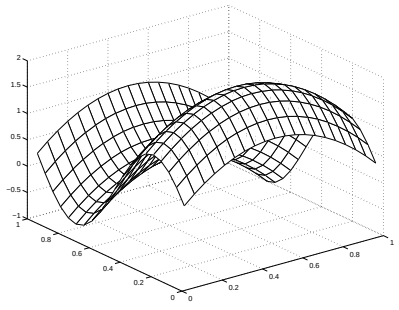

Figure 1: Function I: $\sin(2\pi x) - 4z(z-1)$.

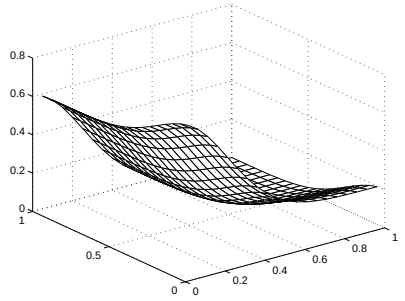

Figure 2: Function II: $0.4\exp(-8x^2) + 0.6\exp(-8z^2)$

5.1 Recovering a locally smooth component

The first set of experiments are based on a number of semi-parametric models with constant parametric component to assess the performance of our approach in recovering locally smooth functions over the square $S = [0,1] \times [0,1]$. We use two functions

- Function I: $h(x,z) = \sin(2\pi x) - 4z(z-1)$
- Function II: $h(x,z) = 0.4\exp(-8x^2) + 0.6\exp(-8z^2)$

which mirror the characteristics of the functions used to assess the performance of complexity based methods [32]. These functions are clearly aperiodic on S, see Figures 1 and 2.

We take $M = N = 6$ components in the Fourier expansion (16) to describe $h(x,z)$, as this closely approximates this function in noise free scenarios for a

range of values for τ and γ. We keep v_0 constant (at 5), u_0 in the range 0.01 to 20000 and w_0 in the range 0.0001 to 40. We use a fixed Gaussian prior for β with $b_0 = 0$ and $B_0 = 3$ and an inverse Gaussian prior for σ^2 with r_0 and s_0 equal to 1.

We present selected results in Table I for functions I and II, with observations drawn from a regular lattice over the square S. We see that good performance is achieved for moderate signal to noise ratios ($\sigma^2 = 0.01$) for all values of n. The best performance is associated with function I in Figure 1. The performance of our approach deceases as n increases and dominates our smoothness priors.

Table I
The effect of the number of observations (in brackets) and the noise variance on reconstruction performance. The average R^2 values are presented for functions I and II. We use the priors $u_0 = 5$, $v_0 = 5$, $r_0 = 4$, $s_0 = 1$, $w_0 = 1$.

	function I				function II			
σ^2	(25)	(100)	(400)	(1600)	(25)	(100)	(400)	(1600)
0.01	0.98	0.95	0.95	0.95	0.87	0.84	0.84	0.82
0.10	0.93	0.89	0.83	0.82	0.83	0.38	0.37	0.31
0.25	0.86	0.73	0.70	0.69	0.80	0.26	0.16	0.16
0.50	0.87	0.56	0.54	0.52	0.66	0.22	0.13	0.10
1.00	0.53	0.47	0.46	0.38	0.87	0.14	0.12	0.05

The performance of our approach deteriorates as the noise variance σ^2 increases. This generally results in less smooth reconstructions. An example is given in Figure 3 for function I in a high noise scenario. This effect can be partly mitigated by imposing stronger smoothness constraints, although our approach is highly robust to changes in the priors for moderate noise levels ($\sigma^2 = 0.01$). Here u_0 must be increased to 2000 before significant effects can be observed in the reconstructed functions.

5.2 Non-linear parametric components

The second set of experiments is based on combinations of non-linear parametric and non-parametric components. We take $h(x, z)$ to be functions I or II, with the following parametric Components with $\beta = (\beta_1, \beta_2, \beta_3)$

- A: $f(\beta, x, z) = \cos(\beta_1 x) + \beta_2 z + \beta_3$ with $\beta = (10, 1, 2)^T$
- B: $f(\beta, x, z) = \tan(\beta_1 x) + \beta_2 z + \beta_3$ with $\beta = (1.5, 1, 1)^T$
- C: $f(\beta, x, z) = \beta_1 x + \beta_2 z + \beta_3$ with $\beta = (1, 1, 1)^T$
- D: $f(\beta, x, z) = \beta_2$ if $x \leq \beta_1$ and $5 + \beta_2$ if $x > \beta_1$, with $\beta = (0.5, 1)^T$

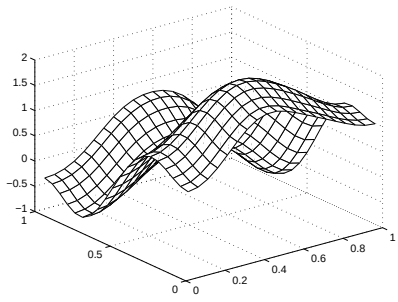

Figure 3: Function I recovered in a high noise scenario

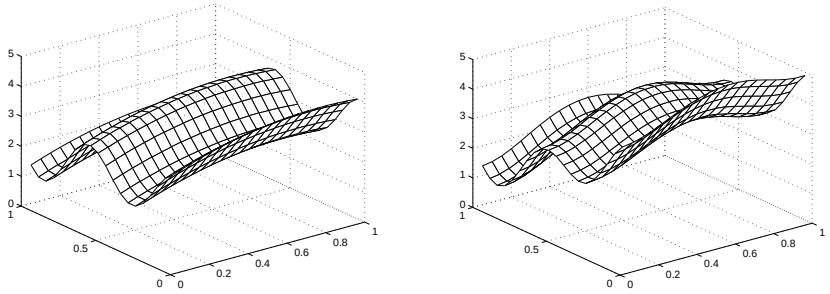

Figure 4: (Left) The function (A + II). (Right) The recovered version

on the unit square S. A typical semi-parametric model generated by functions $A + I$ is given in the left hand image in Figure 4. Representative results are presented for semi-parametric models observed at 400 points on a regular lattice over S and $\sigma^2 = 0.5$, as the corresponding noise substantially disrupts the visual appearance of the underlying function. Five realizations of each semi-parametric model are used for assessment.

We describe the non-parametric component by a tensor product Fourier basis with 6×6 components and the priors used in our first set of experiments. The prior on the parametric component is a Gaussian distribution centred on the true parameter values and diagonal covariance matrix.

We see that the reconstruction of the semi-parametric model (A+I) is visually close to the underlying function (compare the left and right hand images in Figure 4). The corresponding parametric component II and its reconstruction are presented in Figure 5. The latter is generated by the posterior mean

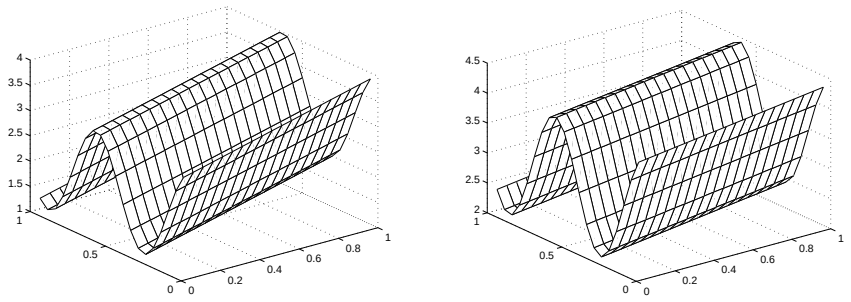

Figure 5: (Left) The parametric component and (Right) The recovered version

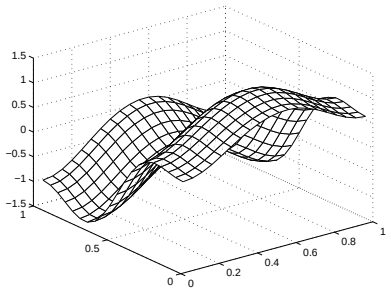

Figure 6: The non-parametric component (II) recovered by our approach

$(10.26, 2.99, 0.44)^T$ associated with β, although this includes the mean associated with the non-parametric component.

The corresponding estimate of the non-parametric component is presented in Figure 6 and is relatively close to the underlying function in Figure 1. Our approach achieves R^2 values of approximately 0.7 with the smoothness priors generated by $u_0 = 4$ and $v_0 = 1$, giving an inverse gamma distributed prior with a mean of $2/7$ for τ^2.

Next, we consider the effect of including a non-linear parametric component in our model. We note that there is relatively little interaction between the posterior estimates of the parametric and non-parametric components for B_0 with large diagonal elements (100). These results are typical, with high values of τ^2 allowing some interaction between the estimates of the parametric and non-parametric components.

The convergence of the sampling scheme described in the appendix appears to be influenced by the presence of a non-linear parametric component. The number of samples needed to ensure convergence is problem dependent, with several thousand typically used in our experiments, although there are still a small number of instances of the underlying chain becoming trapped in a local minimum. The latter appear to be associated with the use of a non-linear parametric component and are a common problem associated with sampling methodologies in non-linear regression problems.

The results presented in this section are typical of those conducted in this study, which suggest that our approach can recover the components of our semi-parametric model in adverse scenarios generated by non-laboratory conditions.

6 Detecting and identifying eye blink artefacts

The electrical potential on the scalp of an individual is measured by an array of electrodes in a process known electroencephalography (EEG). This provides a non-invasive means of sensing brain functions in real time, as potentials are generated by the coherent activation of large groups of neurons in the brain. The resulting signals are known as evoked potentials when they are generated by a stimulus. EEG data is used to study cognition, identify individuals [33], detect various stages in sleep [34] and provide a non-invasive means of controlling machines [35].

Eye blink artefacts are generated by the contact of an eyelid with the cornea which is positively charged. This adds a new component to the potential field on the scalp. An example of a scalp potential with an eye blink artefact is given in Figure 8. Here the scalp potentials are projected onto the extended standard electrode positions in Figure 7 and exclude the reference electrodes A1, A2 and NZ.

It is easy to see that the magnitude of the new component decreases rapidly as it moves away from the eye [36]. Its initial magnitude is typically ten times larger than the background scalp potential and last around 200 to 400 ms. The presence of eye blink artefacts can seriously degrade in the performance of procedures used to analyse evoked potentials [8].

We propose a mixture of Gaussian components to describe an eye blink artefact. This is defined using a spherical skull model [37], although our approach can be used with more realistic skull models. Here

$$Y(\phi_i, \varphi_i) = f(\beta, \phi_i, \varphi_i) + h(\phi_i, \varphi_i) + Z_i \tag{21}$$

for the i^{th} electrode at (ϕ_i, φ_i) in spherical co- ordinates on the sphere S, were $\phi_i \in [0, 2\pi]$ and $\varphi_i \in [0, \pi/2]$. The locally smooth component $h(\phi, \varphi)$ is described by a Fourier expansion (16) with $M = N = 7$. Earlier experiments suggest that these give sufficient flexibility to capture significant characteristics of scalp potentials.

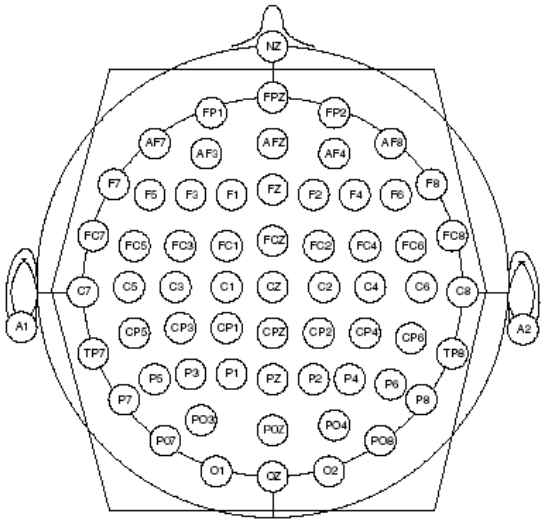

Figure 7: Extended standard electrode positions

Extensive clinical investigations have generated strong *a priori* knowledge about the position, form and extent of eye blink artefacts. We describe this by a two component expansion

$$f(\beta, \phi, \varphi) = \beta_1^{(w)} \Psi_1(\phi, \varphi) + \beta_2^{(w)} \Psi_2(\phi, \varphi) \tag{22}$$

with

$$\Psi_j(z) = (2\pi)^{-1/2} |\Sigma_j(\beta)|^{-1/2} exp\left(\frac{1}{2}(z - \mu_j(\beta))^T \Sigma_j^{-1}(\beta)(z \ - \ \mu_j(\beta))\right) \tag{23}$$

where $\beta = (\beta^{(w)}, \beta^{(\mu)}, \beta^{(\Sigma)})$ and $z = (\phi, \varphi) \in S$. Our priors for β are suggested by examining historical experimental data and describe likely locations of eye blink artefacts. We assume that $\mu_j(\beta) = (\beta_{j1}^{(\mu)}, \beta_{j2}^{(\mu)})$ and

$$\beta_{j1}^{(\mu)} \sim N(1.75, 0.1), \quad \beta_{j2}^{(\mu)} \sim N(1.75, 0.1) \tag{24}$$

with diagonal covariance matrix $\Sigma_j(\beta) = diag(\beta_{j1}^{(\Sigma)}, \beta_{j2}^{(\Sigma)})$,

$$\beta_{j1}^{(\Sigma)} \sim IG(3, 2), \quad \beta_{j1}^{(\Sigma)} \sim IG(3, 2) \tag{25}$$

using prior weights $\beta_1^{(w)} \sim N(1.75, 0.1)$, $\beta_2^{(w)} \sim N(1.75, 0.1)$. The use of prior information ensures that atypical positions of eye blink artefacts in high noise scenarios are less likely.

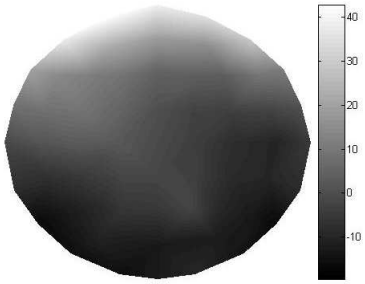

Figure 8: The scalp potential field on the standard electrode positions (excluding the reference electrodes A1, A2 and NZ) with an eye blink artefact

The Bayesian approach provides a principled means of determining whether an eye blink artefact is present or absent and provides an estimate of the underlying electrical potential. These tasks are currently achieved by a variety of techniques ranging from the use of a threshold [33], regression estimates using reference electrodes near the eye [38] and blind signal separation [36] [8].

We introduce two models M_1 and M_0 with equal prior probabilities. The no-blink model given by

$$M_0 : Y(\phi_i, \varphi_i) = \beta^{mean} + h(\phi_i, \varphi_i) + Z_i \tag{26}$$

with the prior $\beta^{mean} \sim N(0, 1000)$ and the eye-blink artefact model

$$M_1 : Y(\phi_i, \varphi_i) = f(\beta, \phi_i, \varphi_i) + h(\phi_i, \varphi_i) + Z_i \tag{27}$$

with the priors ($u_0 = 5$, $v_0 = 5$, $r_0 = 4$, $s_0 = 1$, $w_0 = 1$) for the locally smooth component with $M = N = 7$ basis functions. These values are selected using the experience gleaned from our experiments on recovering locally smooth function described earlier. Additional models describing other artefacts (eye movement) and known effects of head shape can be incorporated into this framework.

Model M_m is selected using $\hat{m} = argmax_m \; p(m \mid Y)$

$$p(m \mid Y) \propto p(m) \int_{\Lambda_m} p(\Lambda_m \mid Y, m) d\Lambda_m \tag{28}$$

with

$$p(\Lambda_m \mid Y, m) = p(Y \mid \Lambda_m, m) p(\Lambda_m \mid m) \tag{29}$$

where Λ_m denotes the set of parameters of the model. The quantity $p(Y \mid m)$ is known as the marginal likelihood.

A range of procedures [39] can be used to estimate this quantity (often known as the evidence). We use *importance sampling*. This is based on the generation of

L realizations from an importance proposal $q(\Lambda_m \mid m)$. Let Λ^j be the j^{th} sample from this distribution and

$$\hat{p}(Y \mid m) = \frac{1}{\sum_{j=1}^{L} c_j} \sum_{j=1}^{L} c_j p(Y \mid \Lambda^j, m) \tag{30}$$

where $c_j = p(\Lambda^j \mid m)/q(\Lambda^j \mid m)$. The optimal choice of $q(\Lambda_m \mid m)$ should be as close as possible to $p(\Lambda_m \mid Y, m)$. We use a multi-variate Gaussian importance proposal generated by the posterior mean and covariance of samples from $p(\Lambda_m \mid Y, m)$ generated by the MCMC scheme in the Appendix. This approach gives good results with $c_j = 1$ in our experiments, although long tailed importance proposals are often used [39] in this context.

Our experiments are based on several scalp potentials from different individuals with and without eye blink artefacts. The data was collected in a study [40] of the genetic pre-disposition to alcoholism. Here a group of alcoholics and non-alcoholics view line drawings from the Snodgrass and Vanderwart picture set. These are a collection of visual stimuli used to study cognition (such as name and image agreement, familiarity and visual complexity). We generate 10^5 samples for M_0 and M_1 using the MCMC algorithm in the appendix (after a burn-in of 0.5×10^5 iterations).

We present representative results in Figure 9 which corresponds to the raw scalp potentials in the right hand image in Figure 8. Here

$$\log\ \hat{p}(Y \mid M_0) = -98.4, \qquad \log\ \hat{p}(Y \mid M_1) = -92.8 \tag{31}$$

which suggests that M_1 should be selected with equal prior probabilities of each model. We have applied this approach to several examples of scalp potentials with and without eye blink artefacts. Our approach consistently calculates higher probabilities for the correct model.

7 Analysis of damaged and contaminated images

Successful decision support systems must be able to deal with damaged or corrupted images. These include photographic material damaged by scratches and abrasion [2], vibration induced blur [9], non-uniform lighting effects [10] and drop-outs generated by display faults [7].

We suggest that effects of this type can be mitigated by the use of a semi-parametric model. Let $Y(x_i, x_j)$ be the intensity associated with the $(i,j)^{th}$ pixel in an $m \times n$-dimensional array. We assume that intensities $Y(x_i, x_j)$ are generated by

$$Y(x_i, x_j) = f(\beta, x_i, x_j) + h(x_i, x_j) + Z_{ij} \tag{32}$$

where $h(x_i, x_j)$ describes the underlying image over the square S. The parametric component describes global image contamination, perhaps generated by spatially varying illumination.

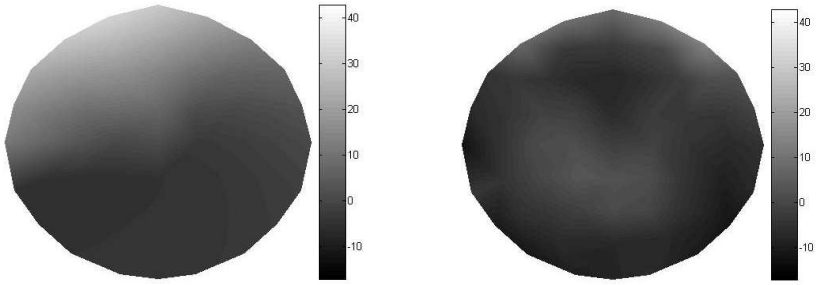

Figure 9: (Right) An estimate of the eye blink artefact from the potential field in Figure 8. (Left) The corresponding non-parametric component

Figure 10: A 120×120 section of the standard Lena image

7.1 Experimental results

This experiment is based on a section of the standard *Lena* image in Figure 10 with grey values in the range 0 to 256. This is corrupted by additive Gaussian noise with zero mean and variance $\sigma^2 = 5$ and global non-linear lighting effect

$$f(\beta, x_i, x_j) = \beta_2 exp(-\beta_3 x_i) + \beta_1 \tag{33}$$

with $\beta_1 = -50$, $\beta_2 = 130$ and $\beta_3 = 0.04$. For a discussion of image degradation models, see [41]. This is added to the original Lena image to give left-hand image in Figure 11.

We also examine the effects of differing degrees of image damage. These are generated by a binomial model with probability p of a pixel being missing. The corresponding pixels are coloured black in subsequent figures (these extend over the image, but are difficult to see in the dark background). We conduct experiments covering a range of values of $p = (0.50, 0.70, 0.80, 0.90$ and $0.95)$.

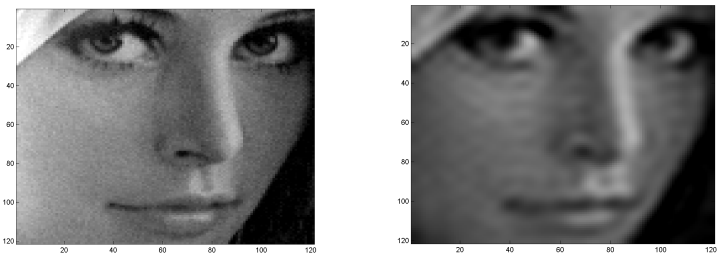

Figure 11: (Left) Lena with additive Gaussian noise with $\sigma^2 = 5$ and non-linear lighting effect. (Right) The restored version

We restore the underlying image using a semi-parametric model with a spatially varying illumination effect with known parametric form (33) and a locally smooth component describing the underlying image. The prior for the parameters describing the parametric component are independently distributed Gaussian random variables centred at the true values and variance of $5, 20$ and 0.1 for β_1, β_2 and β_3 respectively.

The prior for the non-parametric component is based on a Fourier expansion with $M = N = 5$ components. This reduces computational costs and has little effect on performance when the underlying image is smooth. We note that appropriate values of M and N can be generated automatically using a combination of variable selection methods [32] and our approach. We make the assumption that the underlying image does not show any directional effects in the frequency domain and use the priors ($u_0 = 5$, $v_0 = 1000$, $r_0 = 5$, $s_0 = 25$, $w_0 = 10$) to reflect the variance of the noise and the range of pixel values. The relatively large value of w_0 reflects the fact that the underlying image is less smooth than the images considered earlier.

We present results for an image with no missing values ($p = 0$) in Figure 11. We see that the estimated image is free from spatially varying illumination artefacts and noise, but exhibit high frequency artefacts. This appears to be due to the use of a spatially homogeneous model [42]. Our approach can be readily extended to deal with locally smooth components with spatially varying smoothness using a mixture of models approach [43].

The key features of our results are illustrated in Figures 12, 13 and 14. Recognizable reconstructions are generated in all cases, with substantial distortions only appearing at extreme levels of damage (less than 95 percent missing values). The global lighting effect is removed in all cases. This is consistent with the results obtained using non-Bayesian technologies [44] [45], although they cannot deal with non-linear spatially varying illumination, incorporate strong prior knowledge or propagate model uncertainty in subsequent processing. In addition,

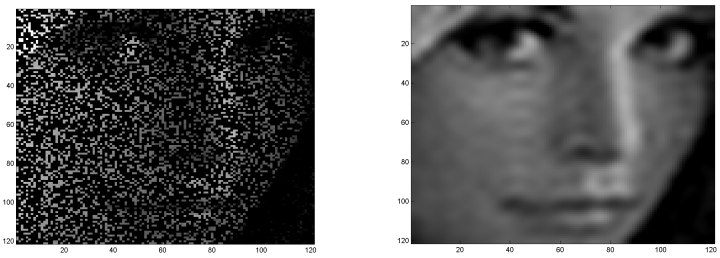

Figure 12: (Left) A 120×120 section of the standard Lena image with additive Gaussian noise with $\sigma^2 = 5$, non-linear exponential lighting effect and 70 percent missing pixel values. (Right) The restored version

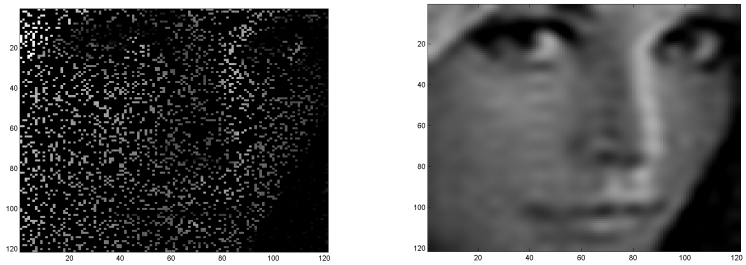

Figure 13: (Left) A 120×120 section of the standard Lena image with additive Gaussian noise with $\sigma^2 = 5$, non-linear lighting effect and 80 percent missing pixel values. (Right) The restored version

we can incorporate a library of different spatially varying illumination and select the most appropriate using the approach used for eye blink artefacts in EEG.

Semi-parametric models can be used to provide enhanced performance in scenarios generating contiguous areas of image damage when there is strong prior knowledge about the structure of the underlying image. This is described by the parametric component with local variations modelled by the non-parametric component [13].

8 Conclusions

Successful decision support systems must be able to deal with damaged and corrupted images. These are ubiquitous feature of data collected in non- laboratory environments and have a significant effect on performance. Our approach is

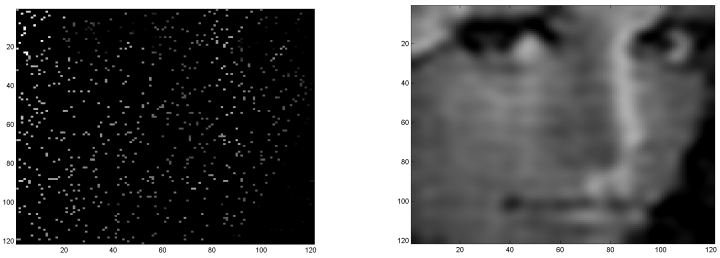

Figure 14: (Left) A 120×120 section of the standard Lena image with additive Gaussian noise with $\sigma^2 = 5$, non-linear lighting effect and 95 percent missing pixel values. (Right) The restored version

based on a novel generalization of [30] to images and fits locally smooth components using a Bayesian approach. This aspect of our approach is related to spline smoothing [46], although it is able to deal with irregular sampled images and provides a means of incorporating prior knowledge about parameter values and propagating uncertainty.

We demonstrate the value of this approach using real and simulated data. The former dealing with eye blink artefacts in EEG and the latter a reconstruction of damaged photographs contaminated by spatially varying lighting effects, each providing an effective and novel solution to these problems. Our experiments suggest that our approach is an effective means of dealing with damaged and contaminated images without the need for a detailed model of the sensor output. The need for the latter is mitigated by the use of a locally smooth component which can adapt to variations in the shape of an image.

The computational cost associated with our approach is dominated by the number of samples generated by our MCMC scheme, as we would expect. The number of samples needed to ensure convergence is problem dependent, with around 10^5 samples used in our image and EEG experiments. The typical cpu time used to carry out each iteration of our MCMC sampler using the Lena image with ninety five percent missing pixels was around 0.01 seconds on a standard laptop computer. A range of criteria used to monitor convergence are described in [31]. There are a small number of instances of the underlying chain becoming trapped in a local minimum. This is a common problem encountered by sampling methodologies in non-linear problems and can be mitigated by generating several independent chains.

While the performance of our approach can be improved by using problem specific information to limit the number of Fourier basis, recent advances in the development of efficient parallel algorithms for computational Bayesian techniques may be of value, see [47].

We have seen that good reconstructions can be achieved when our continuity conditions are satisfied. This is consistent with recent work on the almost sure recovery of signals and images from random samples, provided that they have a sparse representation relative to some basis system. In contrast to this work, we adopt a fully Bayesian approach which allows for a complete specification of the underlying signal and noise characteristics. In contrast to techniques based on penalizing complexity, [48] this approach provides posterior measures of uncertainty which can be incorporated within Bayesian based data fusion hierarchies. A limited numerical comparison of the performance of the novel smoothing technique used in our approach images and a comparable sparse basis method has been carried out. The latter uses the Bayesian approach described in [24] to give a fair comparison. This technique removes redundant basis functions from an initial set which included Fourier and B-splines in our experiments. The results suggest that both approaches give similar performance on smooth functions, as we would expect.

While smoothness is a key characteristic of our approach, we note that discontinuities can be incorporated using a hybrid approach. This is based on an additional term describing discontinuities in some appropriate basis system, [49] for a simple example. A specific discontinues function is selected using the model selection approach described in our EEG example.

Our approach can also be used for high dimensional images encountered in many medical imaging applications and can be readily extended to deal with non-parametric components with spatially varying smoothness using a mixture of models approach [43] [50].

9 Appendix: Monte-Carlo Markov Chain (MCMC) sampler

Here we describe the MCMC sampler used in our experiments. At the first ($i = 0$) iteration, we choose an initial point $(\beta_0, \sigma_0^2, \theta_0, \tau_0^2, \gamma_0)$. The i^{th} iteration of the MCMC sampler is

1. Draw a realization from $P(\beta_i \mid \sigma_{i-1}^2, \theta_{i-1}, \tau_{i-1}^2, \gamma_{i-1}, Y)$. When $f9\beta, x, z)$ is linear

$$\beta_i \sim \mathrm{N}(\mu_\beta, \Sigma_\beta) \tag{34}$$

where

$$\mu_\beta = \Sigma_\beta \left(D^T (Y - \Phi\Theta)/\sigma_{i-1}^2 + B_0^{-1} b_0 \right) \tag{35}$$

and

$$\Sigma_\beta = ((D^T D)/\sigma_{i-1}^2 + B_0^{-1}) \tag{36}$$

with D a $n \times 1$ matrix with 1 in every position, Φ is the $n \times K$ matrix with $(i,j)^{th}$ entry $\phi_j(x_i, z_i)$ and Φ is the vector of the most recent values of θ. This can be sampled directly.

When f is non-linear, the full conditional density $P(\beta \mid \sigma^2, \theta, \tau^2, \gamma, Y)$ for β is proportional to

$$\exp\left(-(Y - \Phi\Theta - f(\beta, x, z))^T (Y - \Phi\Theta - f(\beta, x, z))/2\sigma^2\right) P(\beta) \quad (37)$$

using the most recent values of the parameter values. This cannot be sampled from exactly, in general, and we use a Metropolis-Hastings step [39] with an appropriate proposition. Our experiments suggest that our approach is relatively insensitive to the choice of proposition, with a range of techniques used for non-linear regression. The development of more complex propositions which exploit $f(\beta, x, z)$ is a significant area for future research.

2. Draw a realization from $P(\sigma_i^2 \mid \beta_{i-1}, \theta_{i-1}, \tau_{i-1}^2, \gamma_{i-1}, Y)$ using

$$\sigma_i^2 \sim \text{IG}((r_0 + n)/2, 2/(R^T R + s_0)) \quad (38)$$

where $R = (r_i, i = 1, \ldots, n)^T$ with $r_i = y_i - f(\beta, x_i, z_i) - h(x_i, z_i)$ is calculated using the most recent values of the parameters.

3. Draw a realization from $P(\theta_i \mid \beta_i, \sigma_i^2, \tau_{i-1}^2, \gamma_{i-1}, Y)$ using

$$\theta_i \sim \text{N}(\mu_\theta, \Sigma_\theta) \quad (39)$$

where

$$\mu_\theta = \Sigma_\theta \Phi^T (Y - f(\beta_i, x, z))/\sigma_i^2 \quad (40)$$

and

$$\Sigma_\theta = ((\Phi^T \Phi)/\sigma_{i-1}^2 + \Psi^{-1}/\tau_{i-1}^2)^{-1} \quad (41)$$

with a $K \times K$ diagonal matrix Ψ with elements $\exp(-\gamma c_j)$ at entry (j,j).

4. Draw a realization from $P(\tau_i^2 \mid \beta_i, \sigma_i^2, \theta_i, \gamma_{i-1}, Y)$:

$$\tau_i^2 \sim \text{IG}((u_0 + K)/2, 2/(\theta_i^T \Psi^{-1} \theta_i + v_0)). \quad (42)$$

5. Draw a realization from $P(\gamma_i \mid \beta_i, \sigma_i^2, \theta_i, \tau_i^2, Y)$ using the *slice* sampler described in [30] and using the prior generated by w_0.

These steps are repeated at each iteration in a random order to ensue that the corresponding Markov chain is reversible.

Acknowledgements

The authors wish to thank Dr. R. Paolo of Bristol University and Mr. W. D. Addison of QinetiQ ltd for their comments on model selection procedures. This work was partially funded by the British Ministry of Defence Corporate Research Programme.

References

[1] T. Hastie, A. Buja, R. Tibshirani, Penalised discriminant analysis, Ann. Statist. 23 (1995) 73-102.

[2] A. C. Kokaram, On missing data treatment for degraded video and film archives: A survey and a new Bayesian approach, IEEE Trans. Image. Proc. 13 (2004) 395-413.

[3] D. C. I. Walsh, A. E. Raftery, Detecting mines in mine fields with linear characteristics, Technometrics 44 (2002) 34-44.

[4] C. W. Anderson, E. A. Stolz, S. Shamsunder, Multivariate auto-regressive models for classification of spontaneous Electroencephalographic signals during mental tasks, IEEE Trans. Biomedical Engineering 45 (1998) 277-286.

[5] M. Lustig, D. Donoho and J. M. Pauly, Sparse MRI: The application of compressed sensing for rapid MR imaging, Magnetic Resonance in Medicine, 58(6) (2007) 1182- 1195.

[6] J. Haupt, R. Nowak, Signal reconstruction from noisy random projections, IEEE Trans. Inform. Theory, 52(9) (2006) 4036-4048.

[7] R. Everson, L. Sirovich, Karhunen-Loéve procedure for gappy data, J. Opt. Soc. America A 12 (1995) 1657-1664.

[8] T-P. Jung, S. Makeig, M. Westerfield, J. Townsend, E. Courchesne, T. J. Sejnowski, Removal of eye activity artefacts from visual event-related potentials in normal and clinical subjects, Clinical Neurophysiology 111 (2000) 1745-1758.

[9] P. Xu, Q. Hao, C. N. Huang, Y. T. Wang, Degradation of modulation transfer function in push-broom camera caused by mechanical vibration, Optronics and Laser technology 35 (2003) 547-552.

[10] Y-B. Yang, H. Yan, An adaptive logical method for binarization of degraded document images, Pattern Rec. 33 (2000) 787-807.

[11] W. Philpot, The derivative ratio algorithm: avoiding atmospheric effects in remote sensing, IEEE Trans. Geoscience and Remote Sensing 29 (1991) 350-357.

[12] E. J. Candes, J. Romberg, T. Tao, Stable signal recovery from incomplete and inaccurate measurements, Communications on Pure and Applied Mathematics, 59(8) (2006) 120-1223.

[13] R. H. Glendinning, A. J. Goode, Semi-parametric classification of noisy curves, Pattern Rec. 36 (2003) 35-44.

[14] W. Guo, Y-D. Wang, M. B. Brown, A signal extraction approach to modelling hormone time series with pulses and a changing baseline, J. Amer. Statist. Assoc. 94 (1999) 746-756.

[15] G. Koop, D. J. Poirier, J. Tobias, Semiparametric Bayesian inference in multiple equation models, J. Appl. Econometrics 20 (2005) 723-747.

[16] L. Fahrmeir, T. Kneib, S. Lang, Penalized structured additive regression for space-time data: a Bayesian perspective, Statistica Sinica 14 (2004) 731-761.

[17] E. E. Kammann, M. Wand, Geoadditive models, J. Royal Statist. Soc. C 52 (2003) 1-18.

[18] H. G. Müller, U. Stadtmüller, F. Tabnak, Spatial smoothing of geographically aggregated data, with application to the construction of incidence maps, J. Amer. Statist. Assoc. 92 (1997) 61-71.

[19] S. Lang, A. Brezger, Bayesian P-splines, J. Comp. Graph. Statist. 13 (2004) 183-212.

[20] M. Smith, P. Yau, T. Shively, Estimating long-term trends in Tropospheric ozone levels, Internat. Statist. Rev. 70 (2002) 99-124.

[21] C. C. Holmes, B. K. Mallick, Generalized nonlinear modelling with multivariate free-knot regression splines, J. Amer. Statist. Assoc. 98 (2003) 352-368.

[22] C. Biller, L. Fahrmeir, Adaptive Bayesian regression splines in semiparametric generalized linear models, J. Comp. Graph. Statist. 9 (2000) 122-141.

[23] I. DiMatteo, C. R. Genovese, R. E. Kass, Bayesian curve-fitting with free-knot splines, Biometrika 88 (2001) 1055-1071.

[24] R. Kohn, M. Smith, D. Chan, Nonparametric regression using linear combinations of basis functions, Statist. Comput. 11 (2001) 313-322.

[25] J. Besag, J. York, A. Mollie, Bayesian image restoration with applications in spatial statistics (with discussion), Ann. Inst. Statist. Math. 43 (1991) 1-59.

[26] L. Fahrmeir, S. Lang, Bayesian semiparametric regression analysis of multi-categorical time-space data, Ann. Inst. Statist. Math. 53 (2001) 11-30.

[27] A. Van der Linde, Splines from a Bayesian point of view, Test 4 (1995) 63-81.

[28] S. Lang, E. M. Fronk, L. Fahrmeir, Function estimation with locally adaptive dynamic models, Comput. Statist. 17 (2002) 479-500.

[29] R. Aykroyd, Bayesian estimation for homogeneous and inhomogeneous Gaussian random fields, IEEE Trans. Patt. Anal. Mach. Intell. 20 (1998) 535-539.

[30] P. J. Lenk, Bayesian inference for semiparametric regression using a Fourier representation, J. Roy. Statist. Soc. B 61 (1999) 863-880.

[31] K. L. Mengersen, C. P. Robert, and C. Guihenneuc-Jouyaux, MCMC convergence diagnostics: A review. In: Baysian Statistics, Vol. 6, edited by J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith. Oxford: Oxford University Press, 1999, 415440.

[32] M. Smith, R. Kohn, A Bayesian approach to non-parametric bivariate regression, J. Amer. Statist. Assoc. 92 (1997) 1522-1535.

[33] R. Palaniappan, Method of identifying individuals using VEP signals and neural network, IEE Proc.-Sci. Meas. Technol. 151 (2004) 16-20.

[34] P. Sykacek, S. J. Roberts, Bayesian time series classification, in: T. G. Dietterich, S. Becker, Z. Ghahramani (Eds.), Advances in Neural Information Processing Systems-14, MIT, 2002, pp. 937-944.

[35] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, T. M. Vaughan, Brain-computer interfaces for communications and control, Clinical Neurophysiology 11 (2002) 767-791.

[36] C. A. Joyce, I. F. Gorodnitsky, M. Kutas, Automatic removal of eye movement and blink artefacts from EEG data using blind component separation, Psychophysiology 41 (2004) 1-13.

[37] F. Perrin, J. Pernier, O. Bertrand, J. F. Echallier, Spherical splines for scalp potential and current density mapping, EEG and Clinical Neurophysiology 72 (1998) 184-187.

[38] J. L. Kenemans, P. Molenaar, M. N. Verbaten, J. L. Slangen, Removal of ocular artefacts from EEG: a comparison of time and frequency domain methods with simulated and real data, Psycholophysiology 28 (1991) 114-121.

[39] J. J. K. O'Ruanaidh, W. J. Fitzgerald, Numerical Bayesian Methods Applied to Signal Processing, Springer-Verlag, Berlin, 1996.

[40] X. L. Zhang, H. Begleiter, B. Porjesz, W. Wang, A. Litke, Event related potentials during object recognition tasks, Brain Research Bulletin 38 (1995) 531-538.

[41] T. Kanungo, R.M. Haralick, and I. Phillips, Nonlinear local and global document degradation models, Int. J. Imaging Systems and Technology 5 (1994) 220-230.

[42] M. Karczewicz, M. Gabbouj, Robust B-spline Image modelling with applications to image analysis, IEEE. Trans. Image. Proc. 7 (1998) 912-917.

[43] S. A. Wood, W. Jiang, M. Tanner, Bayesian mixture of splines for spatial adaptive non-parametric regression, Biometrika 89 (2002) 513-528.

[44] M. Petrou, R. Piroddi, S. Chandra, Irregularly sampled scenes, Proc. of SPIE - The International Society for Optical Engineering 5573 (2005) 319-333.

[45] S. Mazzola, A k-nearest neighbor based method for the restoration of damaged images, Patt. Rec. 23 (1990) 179-184.

[46] M. Berman, Automated smoothing of image and other regularly spaced data, IEEE. Trans. Patt. Anal. Mach. Intell. 16 (1994) 460-468.

[47] J. Yan, M. K. Cowles, S. Wang and M. P. Armstrong, Parallelizing MCMC for Bayesian spatiotemporal geo-statistical models, Statistics and Computing 17(4) (2007) 323-335.

[48] Y. Tsaig, D. L. Donoho, Extensions of compressed sensing, Sig. Proc. 86(3) (2006) 549-571.

[49] J.-J. Shiau, G. Wahba, D. R. Johnson, Partial spline models for the inclusion of Tropopause and frontal boundary information in otherwise smooth two and three dimensional objective analysis, J. Atmospheric and Oceanic Technology 3 (1986) 714-725.

[50] P. Yau, R. Kohn, S. Wood, Bayesian variable selection and model averaging in high-dimensional multinomial nonparametric regression, J. Comput. Graph. Statist. 12 (2003) 23-54.