Journal of Computational Physics **181,** 398–406 (2002)
doi:10.1006/jcph.2002.7109

Estimating Anomalies from Indirect Observations

Jari P. Kaipio* and Erkki Somersalo†

*Department of Applied Physics, University of Kuopio, P.O. Box 1627, FIN-70211 Kuopio, Finland; and
†Institute of Mathematics, Helsinki University of Technology, P.O. Box 1100, FIN-02015 Hut, Finland
E-mail: kaipio@venda.uku.fi; erkki.somersalo@hut.fi

Received June 5, 2001; revised May 15, 2002

Indirect observations often lead to ill-posed inverse problems, where the use of prior information plays a crucial role. When correctly used, informative priors should be based on our belief of what the typical situation looks like, while the aim of the measurement procedure is often to find deviations from the typical. This work proposes a method of implementing priors that on one hand reflects the belief concerning the typical situation and on the other retains sensitivity to anomalies of the prescribed type. Numerical results are based on Bayesian sampling methods.

Key Words: Bayesian inference; Markov chain Monte Carlo sampling; inverse problems; anomalies; image reconstruction.

1. INTRODUCTION

In several fields of physics and engineering, direct observations of a quantity of primary interest are difficult or even impossible to obtain. In such cases, one must perform indirect measurements and by proper mathematical modelling must extract the desired information of the quantity by computational methods. This approach leads often to an inverse problem that is ill-posed. In ill-posed problems small measurement or modelling errors can cause large variations in the estimates, and the problem may suffer from a nonuniqueness; i.e., several different values of the quantity of interest may produce the same measurement output even in the ideal noiseless case.

To get around these problems, one must augment the measurement with *prior information* of the quantity of interest. Prior information can be based on previous observations of the quantity under similar conditions using the same or different measurement modality, or it can be based on our understanding of the underlying physics.

From a practical point of view, an informative prior may be problematic. Indeed, the prior should express what one believes to be the distribution of the parameter of interest in *typical* cases. However, in many applications it is not the typical case one is interested in but rather an

0021-9991/02 $35.00

anomaly, a deviation from the typical. For instance, in medical imaging applications, prior information could be based on anatomical information, while the whole procedure aims at finding anomalies such as tumors or fractures normally absent from anatomically typical structures. These anomalies represent outliers for an anatomically correct prior in the sense that their probability of occurrence is low. A similar situation appears in nondestructive material evaluation in which any deviation from the designed structure is of interest. Hence, estimation based on these priors easily fails to find the anomalies.

In this paper, we propose an approach in which it is possible to implement a prior distribution based on normal or typical cases, yet which allows the presence of anomalies of very different nature to occur. The framework in this paper is Bayesian statistics with sampling-based estimation methods.

2. BAYESIAN MODEL

In this section, we review briefly the Bayesian interpretation of indirect observations. In this approach, all the variables included in the observation model are considered random variables. This randomness reflects one's degree of belief of their values, and this belief is expressed in terms of probability distributions. Rather than trying to find one single estimate of the parameter of interest, one seeks to solve the probability distribution of the corrensponding random variable conditioned on the observation.

In the sequel, the following notations are used. Capital letters refer to random variables, and lowercase letters to their realizations. We write $X \in \mathbb{R}^n$ when X is a real-valued n-variate random variable.

Let

$$(F, N) \in \mathbb{R}^n \times \mathbb{R}^k, \quad G \in \mathbb{R}^m,$$

be random vectors. Here, the vector (F, N) represents all those quantities that cannot be directly observed, while G represents a vector of observable quantities. In our notation, $F \in \mathbb{R}^n$ represents those variables that one is interested in, while $N \in \mathbb{R}^k$ contains unknown but directly uninteresting variables such as the measurement noise or model parameters of which one has incomplete knowledge. The physical model that ties these quantities together is assumed to be of the form

$$G = \mathcal{A}(F, N), \tag{1}$$

where $\mathcal{A} : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^m$ is a deterministic function; i.e., the vectors F and N determine the observable G uniquely. In this model, the probability distribution of the observation G conditioned on $F = f$ and $N = n$ is formally given by

$$\pi(g \mid f, n) = \delta(g - \mathcal{A}(f, n)),$$

where δ is the Dirac delta in $\mathbb{R}^m$.

Let $\pi(f, n)$ denote the probability density of the unknown vector (F, N). Then the joint probability density of (F, N) and G can be written as

$$\pi(f, n, g) = \pi(g \mid f, n)\pi(f, n) = \delta(g - \mathcal{A}(f, n))\pi(f, n).$$

To obtain the joint probability density of the variables F and G, we integrate out the variable n to get

$$\pi(f, g) = \int_{\mathbb{R}^k} \pi(f, n, g)\, dn = \int_{\mathbb{R}^k} \delta(g - \mathcal{A}(f, n))\pi(f, n)\, dn. \tag{2}$$

Assume that the variables F and N are independent, i.e., that we have

$$\pi(f, n) = \pi_{\mathrm{pr}}(f)\pi_{\mathrm{noise}}(n),$$

where we have identified the variable N as noise, and where the prior information about the variable F is coded in the prior distribution π_{pr}. Furthermore, assume for simplicity that the noise is additive. Then the model equation (1) is written as

$$G = A(F) + N.$$

Under these assumptions, Eq. (2) gives

$$\pi(f, g) = \int_{\mathbb{R}^k} \delta(g - A(f) - n)\pi_{\mathrm{pr}}(f)\pi_{\mathrm{noise}}(n)\, dn = \pi_{\mathrm{pr}}(f)\pi_{\mathrm{noise}}(g - A(f)). \tag{3}$$

The density of F conditioned on the observation $G = g$, which is denoted $\pi(f \mid g)$ and is also referred to as the posterior density, is then obtained by the Bayes formula

$$\pi(f \mid g) = \frac{\pi(f, g)}{\int \pi(f, g)\, df} = \frac{\pi_{\mathrm{pr}}(f)\pi_{\mathrm{noise}}(g - A(f))}{\int \pi(f, g)\, df}. \tag{4}$$

The estimation of F given the observation $G = g$ is based on this formula.

The most commonly used estimators are the maximum a posteriori (MAP) estimate,

$$f_{\mathrm{MAP}} = \arg\max_f \pi(f \mid g),$$

and the conditional expectation,

$$f_{|g} = \int_{\mathbb{R}^n} f\pi(f \mid g)\, df. \tag{5}$$

More generally, the conditional expectation for any derived quantity $R(f)$ is

$$R(f)_{|g} = \int_{\mathbb{R}^n} R(f)\pi(f \mid g)\, df.$$

It is often found that the conditional mean is in most cases a more robust point estimate than the MAP [2]. However, the computation of the conditional mean is a futile task in large-dimensional problems. Conventional numerical integration methods such as quadrature, including also conventional Monte Carlo integration, are not feasible methods in the evaluation of the integral in (5) or in other expectations and interval estimates. An increasingly popular method for exploring the posterior distributions is to use more advanced sampling methods such as the Markov chain Monte Carlo (MCMC) methods. MCMC methods allow the drawing of samples from the target distribution. The estimates are then computed from this ensemble. General references on MCMC methods include [3–5, 7] and a classical text in statistical inverse problems is [6]. Our analysis in this article is also based on MCMC. More details will be given when the examples are discussed below.

3. ANOMALY PRIORS

Assume that under normal conditions, the prior distribution of F is given by $\pi_{\rm pr}(f) = \pi_1(f)$ and this prior is such that it more or less rules out an interesting class of anomalies. Then clearly the use of the Bayes formula is not suitable for estimating these anomalies. On the other hand, assume that one knows in terms of statistics what kind of anomalies are possible. In this case, we write simply

$$F = U + V, \quad U \sim \pi_1(u), \quad V \sim \pi_2(v),$$

where U represents the normal or nonanomalous part of F, with the probability distribution being π_1, while V represents the anomalous part with the probability distribution π_2. If U and V are independent, their joint probability distribution is

$$\pi(u, v) = \pi_1(u)\pi_2(v).$$

It is possible to express the prior density of F in terms of π_1 and π_2. Indeed, let us denote

$$H = U - V.$$

Then,

$$U = \frac{1}{2}(F + H), \quad V = \frac{1}{2}(F - H),$$

and we have

$$\pi_{\rm pr}(f) = \int_{\mathbb{R}^n} \pi_1\left(\frac{1}{2}(f + h)\right)\pi_2\left(\frac{1}{2}(f - h)\right)dh. \tag{6}$$

In practice, the computation of the prior for F using formula (6) may be a formidable integration task. The catch in the sampling-based approach is that the integral (6) *need never be integrated*. Indeed, we write the Bayes formula directly for u and v, yielding (with the normalization constant in the denominator ignored)

$$\pi(u, v \mid g) \propto \pi_1(u)\pi_2(v)\pi_{\rm noise}(g - A(u + v)), \tag{7}$$

and compute sampling-based estimates for $(u, v) \in \mathbb{R}^n \times \mathbb{R}^n$ directly from this formula.

4. NUMERICAL EXAMPLE: IMAGE DEBLURRING

To demonstrate the ideas of the previous section, consider a deconvolution problem with noisy observations. We have a pixelized $p \times p$ image $f \in \mathbb{R}^{p\times p}$ that has been convolved with a point spread function $a \in \mathbb{R}^{p\times p}$ and corrupted by noise n to give a blurred image,

$$g(i, j) = \sum_{k,q} a(i - k, j - q) f(k, q) + n(i, j), \quad 1 \le i, j \le p.$$

For the sake of simplicity, we assume in this example that the additive noise is zero mean, Gaussian, and uncorrelated with constant variance,

$$\pi_{\rm noise}(n) \sim \exp\left(-\frac{1}{2\sigma^2}\|n\|_2^2\right).$$

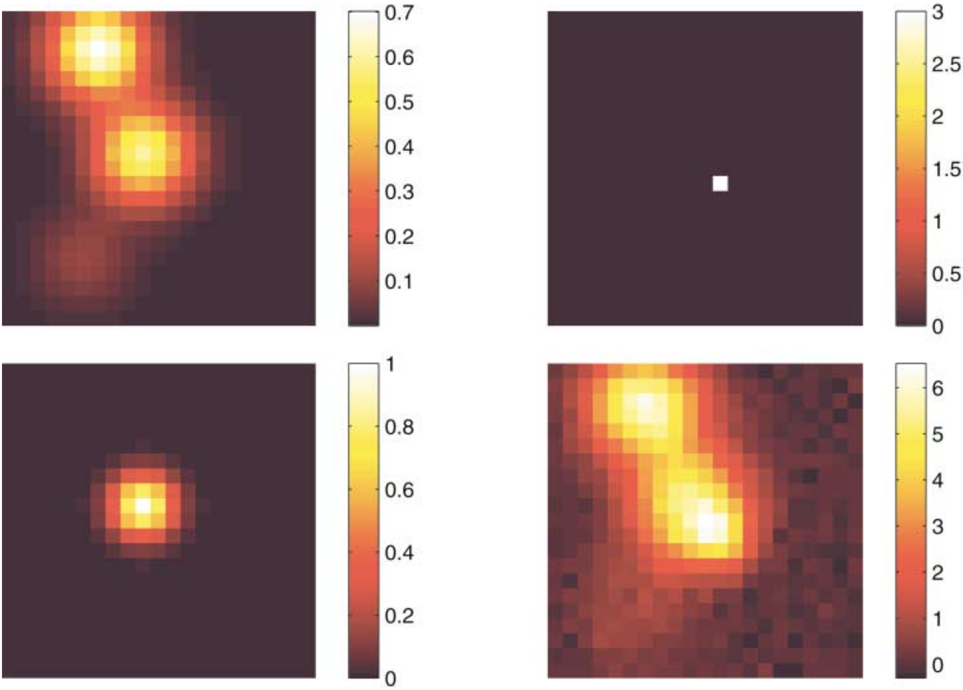

FIG. 1. Smooth part of the true image (top left), and anomalous component (top right), the blurring kernel (bottom left), and the blurred image with 2% Gaussian noise added (bottom right).

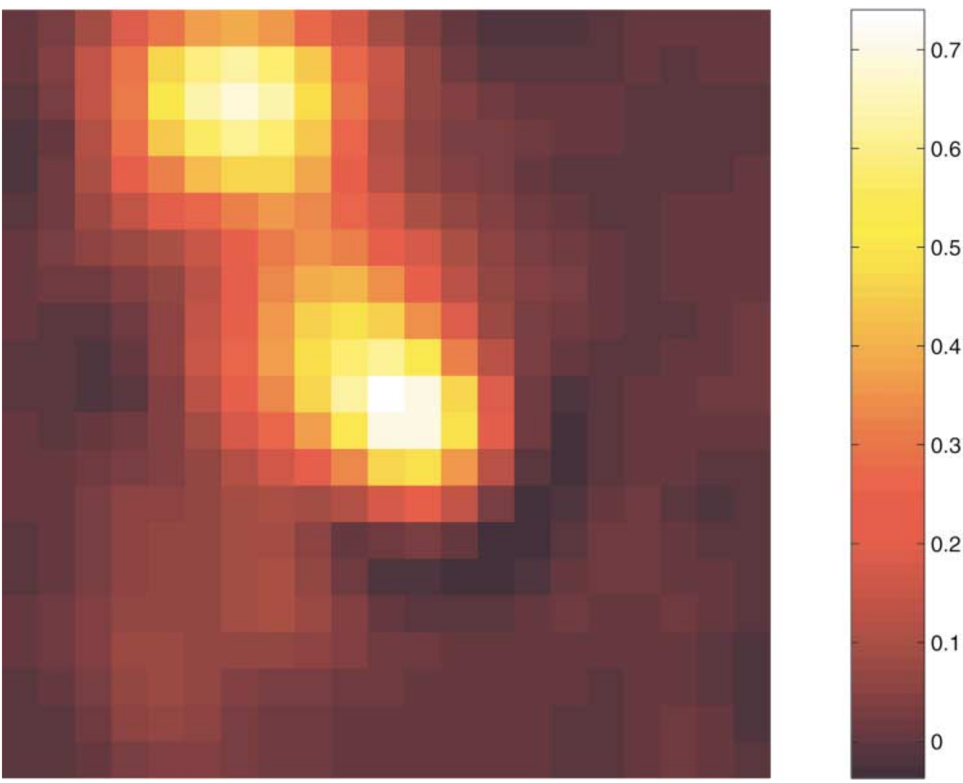

FIG. 2. Maximum a posteriori estimate of the image when the smoothness prior alone is used.

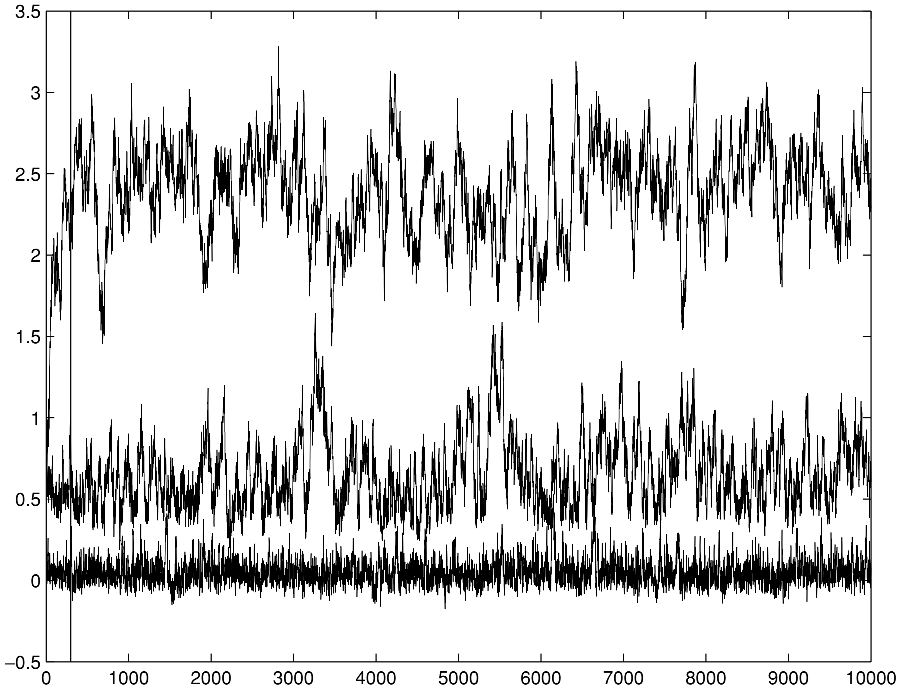

FIG. 3. Sampling histories of three single pixels. The topmost curve is the anomalous pixel. The vertical line denotes the burn-in period.

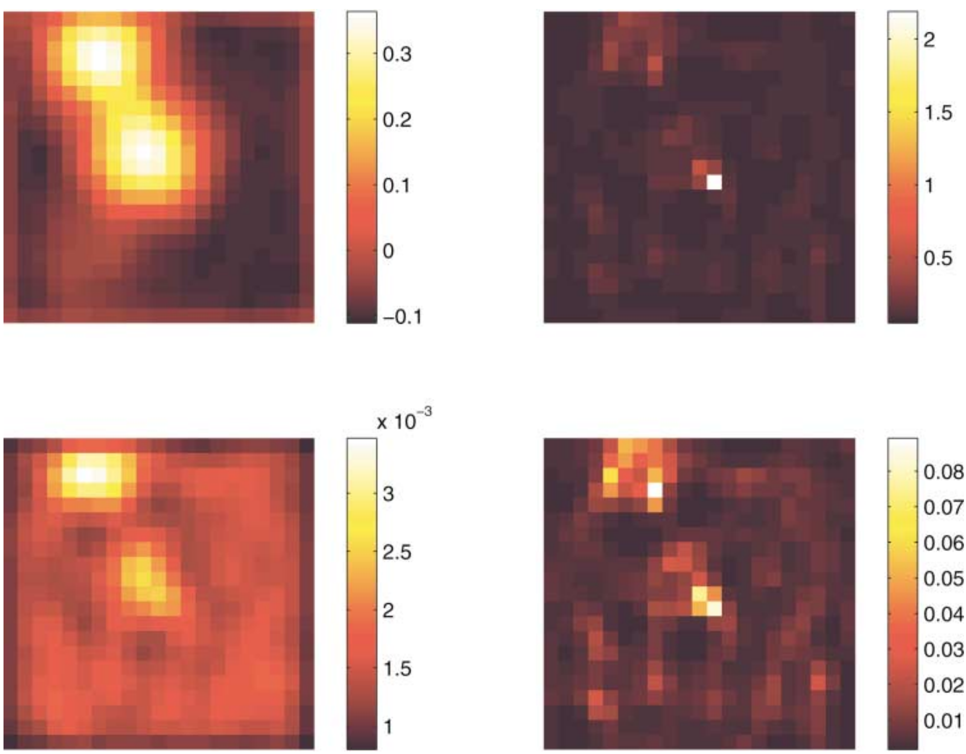

FIG. 4. Sampling-based estimates of the conditional mean of the nonanomalous part, the anomalous part (to row), and their pixel autocovariances (bottom row). The visible boundary effect is due to image truncation.

The noise level σ in our test runs in 2% of maximum amplitude of the noiseless image. Suppose that when no anomalies are present in the image, it is likely to be a smooth image. This belief is expressed here is terms of a smoothness prior,

$$\pi_1(f) \sim \exp\left(-\frac{\alpha}{2}\|Lf\|_2^2\right),$$

where $\alpha > 0$ and L is chosen as a five-point approximation of the Laplacian. Assume now that we suspect that the image may be in fact anomalous and could possibly contain nonsmooth features. More precisely, we have a reason to look for impulse-type anomalies in the image, that is, single pixels that differ considerably from their neighboring values. Clearly, the smoothness prior is in strong conflict with the presence of this type of anomaly in the image. One the other hand, we know that a prior that allows images to contain single outstanding pixels is, e.g., one that is based on the ℓ^1-norm of the image,

$$\pi_2(f) \sim \exp(-\beta\|f\|_1), \quad \|f\|_1 = \sum |f(i, j)|.$$

For more information, see [1] for theoretical discussion and [8] and [4] for applications. Hence, as in the previous section we write $F = U + V$ and formula (7) gives the posterior for (U, V) as

$$\pi(u, v \mid g) \sim \exp\left(-\frac{\alpha}{2}\|Lu\|_2^2 - \beta\|v\|_1 - \frac{1}{2\sigma^2}\|A(u + v) - g\|_2^2\right).$$

Here, A stands for the convolution matrix with kernel $a(i - k, j - q)$. What is more, we impose a positivity constraint on the anomalous part, $v(i, j) \geq 0$. This constraint is not restrictive, since the prior π_1 is insensitive to subtraction of a constant.

To demonstrate the importance of modelling the anomalous part of the image separately from the smooth background, we consider an image of size 21×21 consisting of three Gaussian humps as the nonanomalous smooth part plus an impulse anomaly of the size of one single pixel. The smooth and the anomalous parts of the image are depicted separately in the top row of Fig. 1, with the true image being the sum of these two. To produce the data, this image is first convolved with a blurring kernel $a(i, j)$ and Gaussian noise is added to it. The blurring kernel a is plotted in the left bottom image of Fig. 1, and the blurred noisy data are shown bottom right of Fig. 1.

In this work, we use the same pixel grid for the true image and the inverse solution, although strictly speaking, this choice constitutes what is sometimes called the *inverse crime* in the literature. However, this simplified situation suffices to demonstrate the main ideas here. Assuming first that the smoothness prior π_1 represents the prior distribution of the image, we compute the maximum a posteriori estimator of the image. In this case, the posterior distribution is Gaussian, and the maximum posterior estimate coincides with the Tikhonov regularized solution of the inverse problem,

$$f_s = (A^T A + \alpha\sigma^2 L^T L)^{-1} A^T g. \tag{8}$$

The value of the parameter $\alpha > 0$ is adjusted here to give satisfactory estimates for smooth images. This estimator is plotted in Fig. 2. As expected, the anomalous part of the true image is badly smeared out and thus indistinguishable from the smooth background.

Next we introduce the anomaly prior and estimate both the smooth and the anomalous part from the data. To calculate the estimates by sampling, we use a version of a block-form Gibbs sampler. In algorithmic form, the sampler can be implemented as follows:

Initialize (u^0, v^0), $k = 0$.
do
    $(u, v) = (u^k, v^k)$
    draw u^{k+1} from $\pi(u \mid v, g)$
    for $i, j = 1 : p$
        draw $v^{k+1}(i, j) \geq 0$ from $\pi(v(i, j) \mid u, v(-i, -j), g)$
    end
    $k \leftarrow k + 1$
end

Above, the notation $v(-i, -j)$ denotes the matrix v with the element $v(i, j)$ deleted. The conditional mean estimate is then obtained as the pixelwise mean of the sample ensemble. Observe that the drawing of u is done in one batch operation, as the conditional distribution is Gaussian, while the updating of the anomalous part is done componentwise and requires more numerical work. For more details, see, for example, [4].

In our test, we run the above updating loop 10,000 times, starting from initial values $u^0 = v^0 = 0$. For a criterion of the burn-in (or warm-up) period, we observe the samples of a set of chosen pixel values of the image $f = u + v$. Figure 3 shows the sample values of three pixels. The lowest curve is a background pixel, the middle curve corresponds to a pixel on one of the smooth humps, and the uppermost one is the value of the anomalous pixel. This figure indicates that the burn-in period can be assumed to be relatively short, a few hundred iterations. In our calculations, we used the burn-in sequence of 300 samples, indicated by a vertical line in Fig. 3. This plot also indicates that while the background pixel samples are well uncorrelated, the pixel values of higher amplitude show relatively strong correlation. This is especially true for the anomalous pixel samples. Thus, to obtain reliable estimates for single pixel values with the above updating method, one could expect that rather long runs need to be performed.

To study the performance of the algorithm in *locating* the anomaly, consider the sampling-based conditional means of the regular and anomalous parts. In Fig. 4, we have depicted the conditional means of u and v and the diagonals of their sampling-based posterior covariance matrices, i.e.,

$$u_{|g}(i, j) \approx N^{-1} \sum_k (u^k(i, j)) = \tilde{u}_{|g}(i, j),$$

$$\mathrm{cov}(u_{|g}(i, j)) \approx N^{-1} \sum_k (u^k(i, j) - \tilde{u}_{|g}(i, j))^2,$$

and similarly for v. Above, N is the sample size after the removal of the burn-in sequence, here $N = 9700$. In particular, even if the pixel values of the anomalous pixel are correlated and therefore the sample average converges, possibly slowly, towards the conditional mean in a pixelwise sense, the algorithm in able to locate the anomaly quite accurately. One can argue that in this example, the pixel-value-based statistics is a poor indicator for the convergence while the one based on locating the anomaly is a better one. In fact, once the location of the anomaly is found, one could design a fast Metropolis–Hastings-type algorithm to obtain an estimate for the actual pixel value of the anomaly. The discussion of this issue is not included in this article.

5. CONCLUDING REMARKS

The basic idea presented in this article is to split the parameters of interest in a nonunique way to a typical and an anomalous part and use separate statistics for these parts. This approach can be applied to a wide class of inverse problems, both linear and nonlinear. The implementation of the Monte Carlo sampling is based on Gibbs sampling, which is definitely not an effective algorithm in general. The design of effective sampling algorithms is an application-specific issue and therefore it is not discussed in this work.

REFERENCES

1. D. Donoho, I. M. Johnstone, J. C. Hoch, and A. S. Stern, Maximum entropy and the nearly black object, *J. R. Stat. Soc. Ser. B* **54**, 41 (1992).
2. C. Fox and G. Nicholls, Exact MAP states and expectations from perfect sampling: Greig, Porteous and Seheult revisited, in *Bayesian Inference and Maximum Entropy Methods in Science and Engineering* (CNRS Gif-sur-Yvette, 2000) *AIP Conf. Proc.* **568**, 252–263.
3. W. Gilks, S. Richardson, and D. Spiegelhalter, *Markov Chain Monte Carlo in Practice* (Chapman and Hall, London, 1996).
4. J. P. Kaipio, V. Kolehmainen, E. Somersalo, and M. Vauhkonen, Statistical inversion and Monte Carlo sampling methofs in electrical impedance tomography, *Inverse Probl.* **16**, 1487 (2000).
5. A. F. M. Smith and G. O. Roberts, Bayesian computation via Gibbs sampler and related Markov Chain Monte Carlo methods, *J. R. Stat. Soc. Ser. B* **55**, 3 (1993).
6. A. Tarantola, *Inverse Problem Theory* (Elsevier, Amsterdam, 1987).
7. L. Tierney, Markov chains for exploring posterior distributions, *Ann. Stat.* **22**, 1701 (1994).
8. K. Uutela, M. Hämäläinen, and E. Somersalo, Visualization of magnetoencephalographic data using minimum current estimates, *NeuroImage* **10**, 173 (1999).