PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Global and Local Priors, and the Location of Lesions Using Gamma-Camera Imagery

R. G. Aykroyd and P. J. Green

Phil. Trans. R. Soc. Lond. A 1991 **337**, 323-342
doi: 10.1098/rsta.1991.0128

Global and local priors, and the location of lesions using gamma-camera imagery

By R. G. Aykroyd† and P. J. Green

Department of Mathematics, University of Bristol, Bristol BS8 1TW, U.K.

After a brief review of the paradigm of bayesian image restoration, we pose the question: If high-level prior information is available and usable, what is lost by modelling at the pixel level instead? Our discussion is based on a real application where this question is relevant: the use of gamma-camera imagery in the location of lesions. Procedures using 'global' prior information in the form of a structural model for the image are compared with those using more conventional 'local' priors, modelling only interactions among neighbouring pixels. We address in detail physical modelling and algorithms, and present results with both real and simulated data.

1. Introduction

Most modern work on statistical approaches to the extraction of information from digital images is based on the paradigm of bayesian image analysis pioneered by Besag (1983, 1986) and Geman & Geman (1984). This aims to embrace a wide variety of image analysis tasks, arising in applications right across the applied sciences from geography to medicine, within a framework based on probabilistic modelling and statistical inference. The essence of this framework is as follows.

1. The observed image or *record*, a finite array of numbers representing the pixellated signals or intensities, is regarded as a realization of a random vector.

2. The true 'state of nature', about which the record provides partial information, is a realization of a random array (or function) called the true image or *truth*.

3. Information about the truth available *prior* to observation is represented by a probability distribution.

4. The process generating the record, typically incorporating various forms of degradation such as blur, noise, geometrical distortion and discretization, is represented by the record's conditional distribution given the truth.

5. It is required to make inference about the truth, or some function thereof: this will be based on the conditional distribution of the truth given the record.

We will use the symbols x and y to denote the truth and the record respectively. All probability distributions will be expressed as densities with respect to appropriate measures, and will be denoted generically by p: thus $p(x)$, $p(y|x)$ and $p(x|y)$ represent the three distributions mentioned above.

Within this broad framework, there is considerable flexibility: we briefly elaborate on each of the items above. Even what constitutes the raw data y may be ambiguous when the image is recorded by modern instrumentation incorporating data

† Present address: Department of Mathematics, University of Bradford, Bradford BD7 1DP, U.K.

Phil. Trans. R. Soc. Lond. A (1991) **337**, 323–342
Printed in Great Britain

compression and pre-processing. The truth x may represent some physical reality such as the concentration of a chemical in human tissue, or the variety of a crop growing at a particular point in a remotely sensed scene, or it may simply represent an idealized version of the observed image y: the image that would be obtained by a perfect sensor.

The modelling of $p(x)$ may be based on a physical stochastic mechanism generating variation in x over replications, as might be appropriate in an industrial inspection application for example, or it may be a conventional representation of the form of variation that is expected. In some applications, the representation of information about x in probability terms is completely artificial, and the methodology used is really that of penalized likelihood. (In maximum penalized likelihood estimation, instead of maximizing the log-likelihood $\ln p(y\,|\,x)$ alone, one maximizes a penalized version, $\ln p(y\,|\,x)-J(x)$, say. This is formally equivalent to estimating by the posterior mode, when using a prior $p(x)$ proportional to $\exp(-J(x))$. For a review of penalized likelihood in regression problems (see Green 1987).) Whatever the basis for $p(x)$, it is important that this component be included. The dimension of x is usually comparable with or greater than that of y, so that some form of regularization is necessary in order to render inferences about x from y well-defined and correct. In this paper, we shall adopt the bayesian terminology without further comment.

The degradation model, $p(y\,|\,x)$, is often mechanistic in nature, aiming to capture the image production through an understanding of the physical processes involved, but it may be simply empirical.

Finally, although in this model-based scenario the posterior distribution is completely specified in principle, the large dimensionality of the arrays involved means that computational considerations may influence the specific inferential statements that may be made in practice. For example, one functional of the posterior distribution may be computed much more cheaply than another that would otherwise have been preferred, or approximations may be necessary.

The past six years or so have seen a number of successful applications of the framework described above, and also spin-offs of the same ideas into different areas, such as general multi-parameter bayesian inference (Gelfand & Smith 1990; Smith 1991), pedigrees in genetics (Sheehan & Thomas 1991), and geographical epidemiology (Besag *et al.* 1991). On the other hand, many traditional aspects of image processing technology have been barely touched by this statistical approach, and continue to make use of standard, mostly linear, techniques including spatial filtering, frequency domain methods, etc. Where traditional methods provide satisfactory results, the statistical approach may still have something to add, in that it offers the prospect of more complete inferential statements about the truth, beyond simply provision of a 'cleaned-up image': these include estimates of error, confidence bounds, tests of hypotheses about anomalies, and go on to higher level vision tasks such as probabilistic diagnosis and object recognition. Within the more mundane realm of image restoration and reconstruction, probabilistic methods seem to have had most success in applications where there is substantial stochastic noise, especially where this is not gaussian, where there is awkward geometrical degradation in addition to noise (as in the case of tomographic reconstruction, for instance), and where estimation of parameters in the models is also an issue.

One general field in which these ingredients are often present is that of medical imaging, particularly those techniques where image formation is a result of the counting of individual photons, such as emission tomography and other 'nuclear

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

medicine' methods. This paper will focus on a particular problem arising in this area, that of locating lesions in organs from gamma camera imagery: this application will be introduced in the next section.

Considerable research effort in bayesian image analysis has been to do with the construction of prior distributions $p(x)$. From one perspective, this is just the general problem of modelling spatial stochastic processes in a tractable and flexible way, with parsimonious parametrization. Since we are interested in spatially discrete processes on regular lattices, the seminal work of Besag (1974) on lattice processes has been extremely important; here processes are specified as conditional autoregressions, that is, defined by the local conditional distribution of the value of x at each pixel i given the values at all other pixels, $p(x_i \mid x_{S\setminus i})$ say. If, for each i, this distribution is a function only of x_j for j in a set of pixels spatially close to i (that we denote by ∂i and term neighbours of i), then the resulting distribution is a Markov random field with respect to this neighbour structure. The local conditional probabilities must satisfy consistency conditions that are captured in the Hammersley–Clifford Theorem, which plays a central role by identifying positive Markov random fields precisely as Gibbs distributions.

Much of the modern statistical literature on image analysis has used such processes as prior distributions for the true image, not from principle, but for the thoroughly practical reason that most of the restoration and reconstruction algorithms based on posterior image distributions $p(x \mid y)$ can only be implemented efficiently when the prior is Markov with respect to a fairly sparse graph. For example, approaches using stochastic relaxation (e.g. the Gibbs sampler (Geman & Geman 1984)) and coordinate-wise maximization (e.g. iterated conditional modes (Besag 1986)) use the local conditional *posterior* distributions

$$p(x_i \mid x_{S\setminus i}, y),$$

while methods based on modifying the EM algorithm (see, for example, Green 1990) require the partial derivatives $\partial/\partial x_i \ln p(x)$. It is only with Markov fields $p(x)$ that these expressions can be computed very rapidly.

Although such prior distributions are important for practical reasons, it is clear that they can only to a very limited extent represent real prior information about the truth. Even with additional structure such as 'edge sites' between the pixels to help to organize boundaries in spatial classification problems (Geman & Geman 1984) or hierarchical structure such as that given by allowing the interaction parameter to vary stochastically in a spatially coherent way (see, for example, Clifford 1986), these remain *low-level* or local priors, capturing only relations involving small sets of pixels, such as continuity, gradients, boundary curvature, and so forth. In this sense they are a stochastic analogue of the penalty functionals that are used in non-parametric regression to fit regression relations assuming only smoothness.

Now, however, suppose there really is *high-level* or global prior information about the truth x available. Three pressing questions arise.

(*a*) Can this information be represented in probabilistic form?

(*b*) Can practical algorithms be constructed to extract information about the resulting posterior distribution $p(x \mid y)$?

(*c*) What is lost by not doing so, and using a low-level prior instead?

The answer to (*a*) is very context-specific, of course, and greatly influences the answer to (*b*). Much recent work in statistical image analysis seems to built on the implicit belief that the answer to (*c*) is 'not much'. This is expressed in various ways:

through arguments that properties of posterior distributions or restoration algorithms are sensitive only to *local* properties of the prior, through informal appeal to Savage's principle of precise measurement, and through special-case arguments such as that of Green (1986) in a situation where certain global structure is almost irrelevant to pixel-wise restoration algorithms.

Our principal aim in this paper is to examine a situation where the answers to both (*a*) and (*b*) are 'yes', and where, through some extensive experimentation, we can provide an answer to (*c*). The context is the use of gamma-camera imagery for the detection and location of tumours. The models we use are slightly idealized, but we believe that the comparisons we provide between use of local and global priors in bayesian image analysis have more general relevance. In the next section, we describe the process of image capture using the gamma camera. Section 3 includes a discussion of prior modelling of tumours in this context, and both local and global priors are defined. In §4 we describe our methodology, and in §5 we report on some of our experiments in the use of the resulting techniques, initially on artificial examples, and then on some real gamma-camera data.

For completeness, we close this introduction by mentioning some other alternatives to pixel-based priors that have been discussed. Recent work that comes quite close in spirit to the present paper is that of Dupuis *et al.* (1991) concerning optical astronomy data obtained from the defective Hubble space telescope. Here the authors are concerned with the classification of astronomical objects, such as binary stars, of simple basic form, so that realistic parametric models can be constructed. Avoiding pixellation altogether in modelling the truth can be accomplished by use of appropriate sets of basis functions. For example, if trigonometrical polynomials were used, this would amount to discretization in the frequency domain instead of the original space. Jones & Silverman (1989) use Zernike and Chebyshev polynomials in an orthogonal series approach to the PET reconstruction problem. When the true image is of known, highly structured form, but uncertain in terms of fine detail, the approach of Chow *et al.* (1988) and Amit *et al.* (1991) is attractive: here the stochastic aspect of the prior is provided by representing the truth as the result of a random deformation of a fixed ideal template. Ripley & Sutherland (1990) use related ideas in modelling spiral structure in galaxies.

The underlying effect of all of the approaches just discussed is a drastic reduction in the number of degrees of freedom of the random elements in the true image model from the very large number inevitable if one works at the pixel level. This sounds attractive, but what is the cost in lost generality and robustness? Which approach, global or local, gives better performance when computational effort is also taken into account?

2. Gamma-camera imagery

Gamma-camera imaging is a modern medical diagnostic technique aimed at studying function rather than form. The patient is injected with, or inhales, an appropriate drug or some other substance that is known to become concentrated in the organ of interest, in a manner related to the phenomenon under study, for example, metabolic activity. Before injection, the substance is labelled with a suitable radio-isotope: photon emission therefore occurs in the organ, at a rate varying spatially according to the local concentration of the substance. Indirect measurements of this concentration can thus be made from counts of emitted particles collected in appropriate detectors.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

The physical and operational details of the gamma camera are described in the monograph of Larsson (1980). In essence it consists of a parallel-bore collimator fronting a fluorescing crystal. Incoming photons successfully traversing the narrow tubes of the lead collimator strike the crystal, causing a fluorescence, whose location is measured by an array of photo-multiplier tubes behind the crystal. The circuitry and software of the system record these locations, applying spatial discretization on to a square grid. The whole device is mounted on a ring in a vertical plane so that it can be swung round the patient, who lies on a horizontal couch, and thus used to collect tomographic information: this is the methodology of single photon emission computed tomography (SPECT). A bayesian approach to reconstruction from SPECT data has been investigated by Geman & McClure (1987) and Green (1990). The gamma camera is also routinely used to create sequences of images, to allow the study of dynamic processes such as the flow of material through the kidneys, or the function of the beating heart.

In this paper, however, we are concerned with the most fundamental use of the device, that of making a single, direct, two-dimensional image representing a projection onto the plane of the camera of the isotope concentration within the patient. The recorded image y thus consists of a finite array, usually square, of counts of detected photons, which we will index as $\{y_t, t \in T\}$. The truth x in this case is the spatial distribution of the isotope: let $\{x(\boldsymbol{u}), \boldsymbol{u} \in \mathbb{R}^3\}$ denote the isotope concentration at spatial coordinates $\boldsymbol{u}$ with respect to appropriately chosen axes. We need to model $p(y \mid x)$, the joint probability function of $\{y_t, t \in T\}$ given x. In one sense, this is entirely straightforward: to a close approximation, photons are emitted in an inhomogeneous space-time Poisson process, with rate proportional to $x(\boldsymbol{u})$, and each photon taking an independent, completely random direction of flight. Through nuclear interactions with other particles, both within the body and elsewhere, that lead to scattering and absorption, most of the photons are lost to the system. Some small fraction arrive at the collimator with paths that are sufficiently close to perpendicular to its face that they pass through and are detected. We neglect 'dead-time' effects in detection, so that we assume all photons striking the crystal are recorded. Each photon is recorded at most once, so that we are observing independent, superimposed, thinned Poisson processes of arrival at each detector, and thus the model

$$y_t \sim \text{Poisson}\left(\int a_t(\boldsymbol{u})\, x(\boldsymbol{u})\, \mathrm{d}\boldsymbol{u}\right),$$

independently for all $t \in T$, is derived. Here $a_t(\boldsymbol{u})$ is the mean rate of arrival at detector t from emission at unit rate from a point source of unit concentration at $\boldsymbol{u}$. These functions express the physical factors of absorption and scattering, and the geometrical relation between $\boldsymbol{u}$ and t, including the traversal of the collimator bores. This statistical model for nuclear medicine data is similar to that derived by Shepp & Vardi (1982).

In any reconstruction method based on this model, values for these functions $\{a_t(\boldsymbol{u})\}$ will be needed, and these may be obtained either by associated experiments with known sources, or by appropriate physical modelling. An example of the latter approach is given in some detail in Green (1990) for the case of SPECT.

We are concerned here, however, with the straightforward use of the gamma camera where a single direct view is obtained, so that the integral $\int a_t(\boldsymbol{u})\, x(\boldsymbol{u})\, \mathrm{d}\boldsymbol{u}$ represents a pixellated, blurred, attenuated projection. As a result of the narrow angle of view due to the collimator, there is very little information in these

projections about the third dimension of x, orthogonal to this projection. To be more precise, suppose that the u_3 axis represents that orthogonal direction, oriented into the patient's body, that we can approximate the skin surface nearest the camera by the (u_1, u_2) plane, and that the coefficient of linear attenuation within the body is a constant μ. Then, if detector t has coordinates (t_1, t_2, t_3) with respect to the same axes, we will have approximately

$$a_t(\boldsymbol{u}) = \mathrm{e}^{-\mu u_3} h(u_1 - t_1, u_2 - t_2),$$

for an appropriate point-spread function, h, that will be assumed known. Thus

$$\int a_t(\boldsymbol{u})\, x(\boldsymbol{u})\, \mathrm{d}\boldsymbol{u} = \int h(u_1 - t_1, u_2 - t_2)\, x^*(u_1, u_2)\, \mathrm{d}u_1\, \mathrm{d}u_2,$$

where

$$x^*(u_1, u_2) = \int_0^\infty x(u_1, u_2, u_3)\, \mathrm{e}^{-\mu u_3}\, \mathrm{d}u_3.$$

This expresses the blur, attenuation and projection. Finally, we will discretize x spatially onto the same grid as that on which the data are recorded, and obtain a model in the form

$$y_t \sim \text{Poisson}\left(\sum_s h_{ts}\, x_s\right), \tag{1}$$

where x_s is the discretized version of the attenuated projection $x^*(u_1, u_2)$, and h_{ts} is the discretized point spread function. Other approaches to the discretization would be possible, of course: Jennison & Jubb (1991) discuss restoration on multiple pixel scales, including some much finer than that corresponding to the data, or indeed we could work in continuous space by using a set of smooth basis functions.

3. Priors for ellipsoidal lesions

In this section we discuss both global and local prior distributions that we will use to model a lesion in the organ being imaged. We suppose that it is known that exactly one lesion is present, and the problem is to locate it. There are fairly obvious extensions, which we have not pursued, to problems of detection, and problems involving several lesions: one can simply place a prior distribution over the number of lesions, including zero.

The simplest credible model for a lesion is to regard it as ellipsoidal in shape with a constant isotope density, surrounded by healthy tissue, again with a constant but higher density. In ideal imagery the lesion would thus be seen as an elliptical 'cold spot'. If the organ is reasonably regular in shape, then, in the absence of a lesion, the attenuated projected density x_s would be a smooth function, which we will regard as locally planar. The ellipsoidal lesion will introduce a depression into the surface, which, if its diameter is small compared with the attenuation length constant μ^{-1}, will be hemi-ellipsoidal in profile. Thus, before blur, pixellation and noise, the ideal image surface will be, locally at least, a ramp with a 'hole' of known shape.

Such an intensity surface can be characterized very economically, with three parameters to specify the ramp, two for the hole position, one for its depth and three for its size and shape. Our global prior will be the model that takes the truth to have this ideal form, with a uniform prior distribution over each of the ramp and hole location parameters. In this study, we assume the hole depth, size and shape to be

known. The approach should extend readily to deal with the general case, where informative prior distributions for the additional parameters would usually be appropriate, but that is beyond the scope of this paper.

Two pixel-based priors will also be considered, for comparison. One is an 'implicit discontinuity' model similar to that introduced by Geman & McClure (1987) for use in SPECT:

$$p(x) \propto \exp\left(-\beta \sum_{s\sim r} w_{sr}\, \phi\left(\frac{x_s - x_r}{\delta}\right)\right), \tag{2}$$

where β and δ are parameters, the sum is over pairs of neighbours in an 'eight nearest neighbour' system on the square lattice, and the weights w_{sr} are 1 for orthogonal and $1/\sqrt{2}$ for diagonal neighbours.

If ϕ were a positive definite quadratic function, this would represent a gaussian random field: our experience, agreeing with that of others, is that discontinuities in the image surface are excessively smoothed out in restorations based on a gaussian prior. What is needed is a function that increases less rapidly than $\phi(u) = u^2$ for large (absolute) values of its argument: we use $\phi(u) = \mathrm{logcosh}\,(u)$, as in Green (1990), which has this property, yet for which $p(x)$ remains log-concave.

The second pixel-based prior we consider is a simple modification to this, that gives equal maximum probability to all planar image surfaces, not just the constant ones. As suggested, for example, by Geman (1991), we replace the difference in (2) by a laplacian, so that the prior has the form

$$p(x) \propto \exp\left(-\beta \sum_{s} \phi\left(\frac{x_s - \bar{x}_s)}{\delta}\right)\right), \tag{3}$$

where $\bar{x}_s$ is the mean of the four nearest neighbours. For pixels at the edge of the image, with fewer than four orthogonal neighbours within the image, we replace $\bar{x}_s$ in expression (3) by the value at s of the least squares plane fitted to those of the eight nearest neighbours to s on the lattice that are within the image, and scale δ so as to standardize the variance of all resulting arguments of ϕ. Effectively, (3) penalizes discontinuities in gradient whereas (2) penalizes discontinuities in level.

Both (2) and (3) are chosen with the ideal ramp-plus-ellipsoidal-hole model in mind, and might be expected not to smooth out the abrupt changes in gradient at the perimeter of the hole; however, they only reflect 'local knowledge' and in no sense represent the complete profile of the image surface. On the other hand, their very lack of specificity to the lesion model suggests robustness to departures from this ideal form, which would clearly be a desirable property in handling real images. Further, there is more accumulated expertise in using such low-level models, and the prospects of general purpose software being developed to deal with them.

4. Methodology and algorithms

We are here aiming at Bayes inference using a fully specified model, but may have to accept compromises, forced by computational considerations, over what aspects of the posterior distribution are extracted.

The methods for image restoration and lesion location described here are based on approximate posterior distributions calculated using various forms of the Metropolis algorithm. This is a dynamic Monte Carlo technique, in which an ergodic image-valued Markov chain is constructed that has the posterior distribution of interest as

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

its limiting distribution. For discussion see Hammersley & Handscomb (1964), Hastings (1970), Geman (1991) or Green & Han (1991).

We first consider this in detail for the case of the global prior. The methodology has been implemented at present only for the case where the hole in the true image surface x caused by the lesion has a known profile (depth, size and shape). There remain only five degrees of freedom to describe the surface (intercept, two gradients and two hole location coordinates). These reduce to three in the case of a one-dimensional image, a situation studied in greater detail in our simulation experiments.

The Metropolis algorithm is used in the following manner for this problem (Hastings (1970) is a convenient general reference). Each of the five (or three) parameters is considered in turn. A proposed new value for the parameter is drawn from a Normal distribution centred at the current value. It is valid to use any positive value τ^2 for the variance of this proposal distribution; however, the rate of convergence to equilibrium is greatly influenced by τ^2. This issue is discussed in detail by Green & Han (1991). In the present context, separate experiments suggested that the best performance was obtained with τ^2 somewhat larger than the variance of a gaussian approximation to the Gibbs sampler for this problem. Let the true discretized image surface corresponding to the current parameter values be x, and that obtained with the proposed change be x'. The proposal is accepted, and the parameter value updated accordingly with probability $\min\{1, p(x'\,|\,y)/p(x\,|\,y)\}$. Otherwise, it is rejected, and no change is made. The expression for the acceptance probability reduces to $\min\{1, p(y\,|\,x')/p(y\,|\,x)\}$ since the prior is uniform, and further simplifies considerably for numerical calculation when the point-spread function h has limited width. Each of the parameters is considered, in the same way, and the whole cycle repeated until stability is apparent.

Convergence is assessed by monitoring the values of several one-dimensional functionals of the evolving process, including the total posterior energy, the relative frequency of acceptance, and parameters of a least-squares plane fitted to the image surface. The stochastic processes thus defined are plotted against number of sweeps, and a judgement made as to how many iterations must be discarded to be confident that the initial transient is effectively finished. A full discussion of this difficult issue is not appropriate here: the reader is referred to the excellent lecture notes on the subject by Sokal (1989). It is clear that there can be no absolute guarantee that convergence has been achieved, given a finite realization. The process is continued for several hundred more sweeps, accumulating the sample means of the evolving parameter values. Our estimate of x is the surface corresponding to these mean values.

Turning now to the pixel-based priors, a similar procedure is followed, except that now it is individual pixel values that are considered in turn. Again Normal distributions centred at current values, with variances chosen in a similar way, are used to generate proposals. The acceptance probability has a more complicated form in the presence of a non-trivial prior: it reduces to

$$\min\left\{1, \frac{p(y\,|\,x')}{p(y\,|\,x)}\frac{p(x'_i\,|\,x_{\partial i})}{p(x_i\,|\,x_{\partial i})}\right\} \tag{4}$$

when pixel i is being considered. The last factor involves computing only a few values of the function ϕ, whether the pixel difference prior (2) or the Laplacian based prior

(3) is used. After discarding the initial transient, we estimate the posterior expectations of pixel values x_s by accumulating sample means.

Noise and blur are reduced or removed by this process, but of course the result cannot be expected closely to approximate the idealized ramp-plus-ellipsoidal-hole model. This will always be true when using a pixel-based prior, even in the case of artificial data generated from the model, in the presence of noise of the magnitude we are dealing with (the Poisson means are in the range (20, 100)). In our experience, the position of the hole is always well-defined visually in a display of the estimated posterior mean true image, which provides a clean and acceptable function estimate.

For some purposes in practice, however, a parametric point estimate of lesion location would be needed. This will also allow us here to make a fair comparison in numerical terms with the approach using the global prior. Thus some post-processing of the posterior mean is required to extract numerical estimates of the parameters of the surface. The algorithm we use is simplistic, and it is certainly quite inadequate as an estimator of lesion location if applied directly to the raw data. We subtract our estimated posterior mean truth from an over-smoothed version, that reflects the general trend of the surface, but smooths out any holes. Such a surface can be obtained by any convenient method: for example, we used a cubic spline smooth of the posterior mean in our one-dimensional experiments, and another bayesian estimate, with much greater β, in two dimensions. The difference image consists of a pattern of positive and negative lobes: the centroid of the highest positive lobe provides a cheaply computed estimate of hole location, that proves experimentally to be fairly robust to smooth departures from the ramp-plus-hole model. With the lesion located, the remainder of the surface is estimated by a least squares plane fitted to the posterior mean with the hole profile subtracted. This proposed procedure is arbitrary, but the estimates it produces are stable and robust to perturbations in the method. It is extremely cheap computationally compared with any formal method of fitting the ramp-plus-hole model, which would require searching or iterating as the model is severely nonlinear.

As in any application of a stochastic algorithm to solve a deterministic problem, it is appropriate to attempt to quantify the errors introduced into the values of our estimates by the algorithm. We should generally be satisfied that sufficient precision was being achieved if such 'Monte Carlo variances' did not exceed, say, 1% of the sampling variances of the estimators. Estimation of Monte Carlo variance is discussed by Sokal (1989) and Green & Han (1991). The truncated periodogram estimator advocated there is not appropriate in the present context because the functionals of the process of interest, the parameter estimates extracted by post-processing, are nonlinear. An estimator obtained from 'blocking' the Monte Carlo realization, related to a proposal by Hastings (1970), is more easily used. Suppose that the Metropolis sampler is run for $N = bk$ sweeps after the initial transient has been discarded, so that it may be assumed the process is an equilibrium. Let $\bar{x}_{k,i}$ be the pixel-wise mean of the realized value of x, over sweeps $(i-1)\,k+1$ to ik inclusive, for $i = 1, 2, 3, \ldots, b$. This is the estimated posterior mean from the ith of b blocks of k consecutive sweeps: $\bar{x}_{N,1}$ is the mean over the whole run. Let the post-processing used to extract an estimate of a particular parameter be represented by the function f. Then if $s^2_{b,k}$ is the sample variance of the post-processed block means:

$$s^2_{b,k} = \frac{1}{b-1}\sum_{i=1}^{b}\left[f(\bar{x}_{k,i}) - \frac{1}{b}\sum_{j=1}^{b} f(\bar{x}_{k,j})\right]^2,$$

we will have, under weak conditions on the Markov chain being simulated (see Green & Han 1991),

$$E(s^2_{b,k}) \sim \mathrm{var}\{f(\bar{x}_{k,i})\}, \quad \mathrm{var}\{f(\bar{x}_{k,i})\} \sim c/k \quad \text{as } k \to \infty,$$

so it follows that $b^{-1} s^2_{b,k}$ is an approximately unbiased estimator of $\mathrm{var}\{f(\bar{x}_{N,1})\}$ as $k \to \infty$. For stability of this estimator, we will also require b to be large, so for fixed $N = bk$ there will be a trade-off between number of blocks and block length. A practical suggestion would be to compute the estimate of Monte Carlo variance for several block lengths, and to use the smallest blocks for which any trend has levelled out.

One important aspect of this methodology has not been addressed at all in this study. That is the choice of values for the hyperparameters β and δ appearing in (2) and (3). From a formal Bayes perspective, these parameters will either be known, or themselves subject to prior modelling. An empirical Bayes viewpoint would allow them to be estimated from the data, perhaps using Besag's pseudo-likelihood method as implemented for similar problems, not involving blur, by Qian & Titterington (1991*a*), or the Gibbs/EM approach used in Geman & McClure (1987). It should be recognized that this estimation problem is not straightforward, when taking account of the Poisson variation, the substantial blur, the fact that not only canonical parameters have to be estimated and the fact that the prior distribution will only be an approximate representation of reality. These difficulties are additional to those identified in the very useful discussion by Qian & Titterington (1991*b*). Further, good estimation is not the same as successful selection for the ultimate goal of lesion location. In this paper, values of β and δ are chosen, quite informally, to give good performance in practice. Clearly, further work would be needed on this issue before routine application of this methodology.

5. Experiments

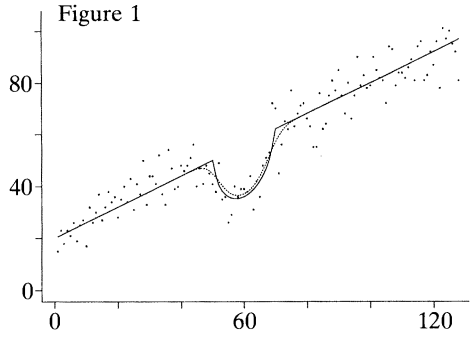
We have conducted a rather extensive series of experiments using the models and methods described above. In this section, we report on some of these in two subsections, one concerning simulation experiments with one-dimensional images, and the other using real gamma-camera data.

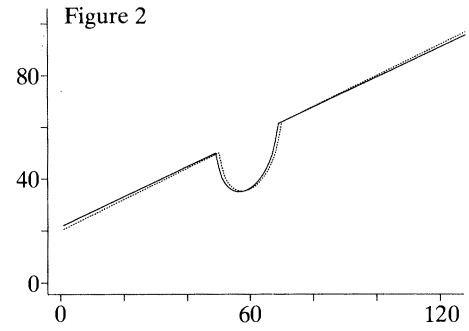
(*a*) *One-dimensional simulations*

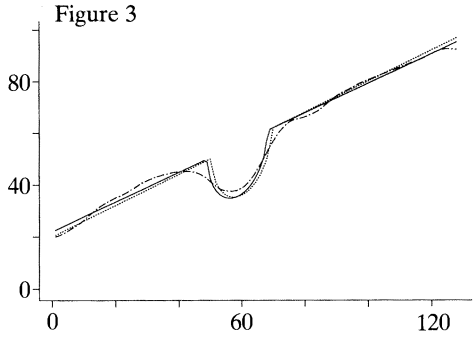
Artificial data from the model (1) were generated in order to examine the performance of our methods under repeated sampling. The truth was constructed using the ramp-plus-hole model in one dimension, in which the ramp had the form

$$x_s = \alpha_0 + \alpha_1 s$$

using values $\alpha_0 = 20$ and $\alpha_1 = 0.6$. The hole had a fixed profile, that of a semi-ellipse, with vertical semi-axis 20 and horizontal semi-axis 9.5, centred at location $\gamma = 60$. This surface was discretized into 128 pixels, not by sampling, but by integrating over 128 contiguous intervals $[0.5, 1.5]$, $[1.5, 2.5], \ldots, [127.5, 128.5]$. This seemed to be a better approximation of the effect of discretization in the real gamma camera. The point-spread function h_{ts} appearing in (1) was taken to be gaussian in shape, integrating to 1, with spread specified by its standard deviation σ; for numerical efficiency, the gaussian curve was truncated at $\pm 3\sigma$. (Point-spread functions are commonly parameterized in image-processing by the full width at half maximum







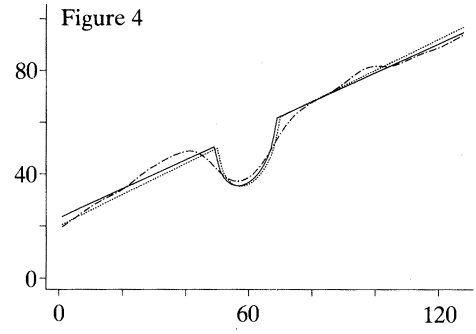


Figure 1. Truth, blurred truth and example data-set ($\sigma = 3$). ——, truth; – – –, blurred truth; ●, example data-set.

Figure 2. Reconstruction using global prior ($\sigma = 3$). ——, posterior mean; – – –, truth.

Figure 3. Reconstruction using local difference prior ($\sigma = 3$). ——, post processed; – – –, truth; —·—, posterior mean.

Figure 4. Reconstructions using local laplacian prior ($\sigma = 3$). ——, post processed; – – –, truth; —·—, posterior mean.

(FWHM). For a gaussian function, FWHM is 2.35σ.) In our experiments σ was varied, taking values 0, 1, 3 and 5, and its value assumed known in the analysis. A modest complication is introduced by edge effects, which are quite pronounced when the blur is severe. Values of x_s needed from pixels s beyond the boundary of the observed image were replaced by values extrapolated by linear regression through values close to but within the boundary. (For efficiency, this was actually implemented by appropriately modifying h_{ts}.) The effect is that the blur leaves a linear ramp unaltered.

For each level of blur, 100 independent replicate data-sets were generated from the model (1), using an exact Poisson random number generator. The range of variation of the artificial data in this study was chosen to correspond broadly to that of the real gamma-camera imagery we have encountered.

In figure 1, we display the true ramp-plus-hole curve, the result after application of the gaussian point spread function with $\sigma = 3$, and one of the 100 corresponding data-sets. It is evident that both blur and noise make the problem of locating the hole considerably more difficult.

Each of the three methods proposed in the paper was applied to each data-set. In the case of the 'local' prior methods, values for β and δ were chosen by trial and error,

Table 1. *Mean parameter estimates from one-dimensional simulation experiments*

global prior (number of sweeps = 2000 (after discarding 500), $\tau^2 = \{4.0, 9.0 \times 10^{-4}, 4.0\}$)

σ	$\bar{\alpha}_0$	$\bar{\alpha}_1$	$\bar{\gamma}$	mse (l)	mse (α_0)	mse (α_1)	mse (γ)
0	19.85	0.6008	60.05	1.472	1.466	3.01×10^{-4}	0.422
1	19.84	0.6044	60.11	1.728	1.675	3.68×10^{-4}	0.500
3	19.90	0.6016	60.01	1.843	1.140	3.06×10^{-4}	0.775
5	20.07	0.6015	59.92	2.362	1.117	2.88×10^{-4}	1.395

local difference prior (number of sweeps = 5000 (after discarding 500), $\tau^2 = 25$)

σ	β	$\bar{\alpha}_0$	$\bar{\alpha}_1$	$\bar{\gamma}$	mse (l)	mse (α_0)	mse (α_1)	mse (γ)
0	3.0	20.40	0.5890	60.18	1.878	1.891	4.828×10^{-4}	0.560
1	2.5	20.27	0.5946	60.12	1.849	1.532	4.401×10^{-4}	0.716
3	2.0	20.17	0.5959	60.11	1.929	1.293	3.306×10^{-4}	0.941
5	1.5	20.29	0.5968	60.10	2.624	1.471	3.681×10^{-4}	1.631

local laplacian prior (number of sweeps = 5000 (after discarding 2000), $\tau^2 = 0.25$)

σ	β	$\bar{\alpha}_0$	$\bar{\alpha}_1$	$\bar{\gamma}$	mse (l)	mse (α_0)	mse (α_1)	mse (γ)
0	40	19.74	0.6031	60.01	1.844	1.771	4.104×10^{-4}	0.664
1	35	19.93	0.6030	60.01	2.172	1.784	5.190×10^{-4}	0.799
3	25	20.02	0.6011	60.00	2.620	1.602	4.655×10^{-4}	1.396
5	25	20.04	0.6009	59.92	2.886	1.521	3.927×10^{-4}	1.952

to give good average performance across all data-sets at a given level of blur. The values used for β are indicated in table 1: δ was fixed throughout at 2.0 for the pixel difference prior, and 0.5 for the laplacian prior. These values were chosen as giving best performance in terms of mean squared error. The table also indicates the values of τ^2 and the number of sweeps that were used. Figures 2, 3 and 4 display the results of our methods in graphical form, for the single data-set illustrated in figure 1. These figures show the estimated true images, obtained from the estimated posterior means as described in §4, and also the post-processed versions in the cases of the two 'local' methods.

In table 1, the performance across all levels of blur and all data-sets is summarized by a tabulation of means and mean-squared errors across replications. For a particular level of blur, and a particular method, let $\hat{\alpha}_0^{(r)}$, $\hat{\alpha}_1^{(r)}$ and $\hat{\gamma}^{(r)}$ be the parameter estimates for the rth data-set obtained from the posterior means, after post-processing if necessary. The columns headed $\bar{\alpha}_0$, $\bar{\alpha}_1$ and $\bar{\gamma}$ give the values of $\sum \hat{\alpha}_0^{(r)}/100$, etc. The columns headed mse (α_0), etc., give the mean-squared errors $\sum (\hat{\alpha}_0^{(r)} - \alpha_0)^2/100$, etc., and that headed mse (l) the average per-pixel mean-squared error

$$\sum_{r=1}^{100} \sum_{s=1}^{128} (\hat{x}_s^{(r)} - x_s)^2/12800,$$

where $\hat{x}_s^{(r)}$ is the estimated ramp-plus-hole surface based on $\hat{\alpha}_0^{(r)}$, $\hat{\alpha}_1^{(r)}$ and $\hat{\gamma}^{(r)}$. For the two local priors, the mean-squared errors for the posterior means before post-processing were almost exactly four times the mse (l) figures. From these means and mean-squared errors, bias and variance can readily be estimated. Since the replicates are independent, approximate t-statistics can be calculated to assess the significance of any apparent biases. While there is no significant evidence that the estimates

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

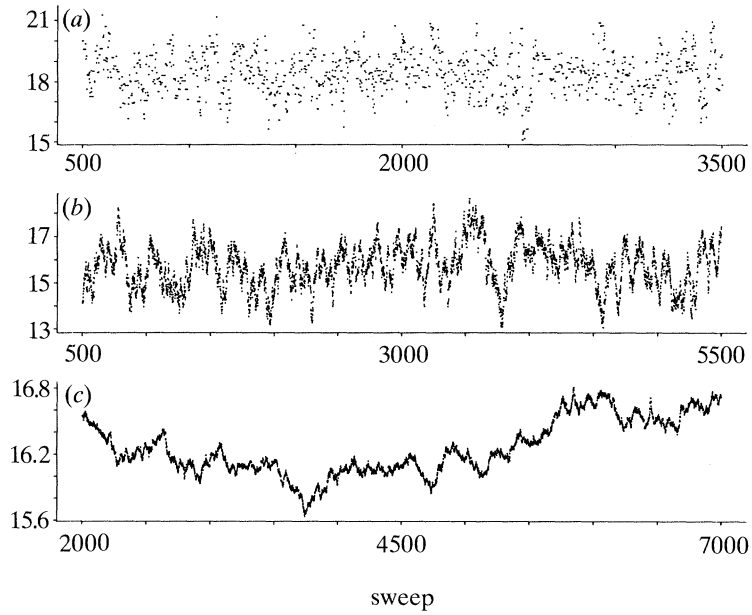


Figure 5. Monitoring statistics. (*a*) Global prior: intercept parameter. (*b*) Local difference prior: intercept of LS line. (*c*) Local laplacian prior: intercept of LS line.

based on the global and laplacian priors are biased, the method based on the pixel difference prior on average estimates the ramp as too shallow, with the hole position too far to the right.

In terms of overall mean-squared errors, however, all three methods are performing comparably well. Note the role of the blur parameter σ. As this increases, the errors in estimating γ increase while those of α_0 and α_1 reduce slightly.

A benchmark for the estimated sampling variances in this experiment is provided by those obtained by maximum likelihood estimation of (α_0, α_1) in fitting the ramp alone to the first of the stimulated data-sets, for example by Fisher scoring. The resulting variances were (0.968, 0.00272), slightly less than those obtained here.

In making firm comparisons between the methods, some caution is necessary. First, we need to take account of the computing times involved. For the numbers of sweeps indicated in table 1, the times for each of the three methods were between 3 min and 4 min for a single data-set in the case $\sigma = 0$ (all times quoted were obtained on a Sun Sparcstation 1). Thus the effort saved in only resampling three values each sweep instead of 128 is offset by the additional effort in repeatedly recomputing the ramp-plus-hole surface and all likelihood terms. However, as the blur increases to the rather excessive maximum value considered, $\sigma = 5$, while the computing times for the global method increase by a factor of about 2.5, those for the local methods multiply 14 times. Thus, from this perspective, the local methods become increasingly uncompetitive as blur increases.

More seriously, there is the question of whether convergence is really achieved. In this one-dimensional case, convergence of the pixel-based methods, especially the laplacian one, is very slow indeed. The parameter settings used here were selected after exhaustive experimentation as giving the fastest convergence available, yet plots of our monitoring statistics, such as those in figure 5*c*, suggest that autocorrelation times remain long. This difficulty does not apply to the global

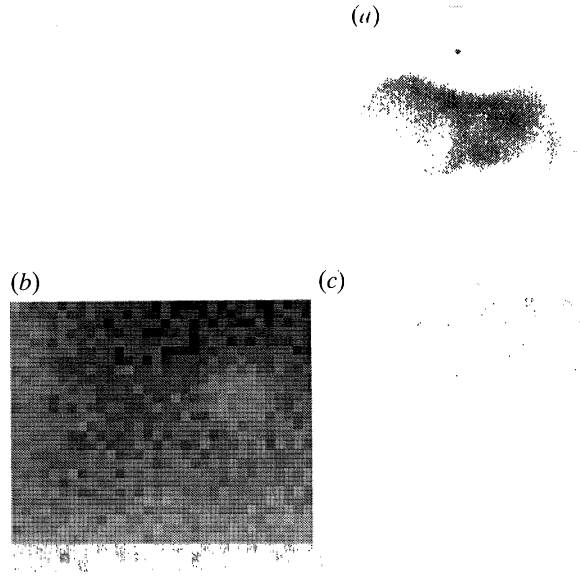


Figure 6. Gamma-camera data. (*a*) Image of liver. (*b*) Extract of phantom data. (*c*) Thinned phantom data.

Table 2. *Variances for global prior*

	estimate variances			Monte Carlo variances		
σ	$\bar{\alpha}_0$	$\bar{\alpha}_1$	$\bar{\gamma}$	$\bar{\alpha}_0$	$\bar{\alpha}_1$	$\bar{\gamma}$
0	1.442	3.004×10^{-3}	0.4222	0.010976	3.027×10^{-6}	0.56251×10^{-3}
1	1.186	3.081×10^{-3}	0.5021	0.005000	1.470×10^{-6}	0.88771×10^{-3}
3	1.128	3.035×10^{-3}	0.7750	0.006648	1.765×10^{-6}	1.47200×10^{-3}
5	1.113	2.859×10^{-3}	1.3955	0.011473	3.417×10^{-6}	2.04096×10^{-3}

method (see figure 5*a*), for which in table 2 we give Monte Carlo variances calculated by the blocking method. Comparing these with the sampling variances in the same table confirms that the Monte Carlo runs we used were more than adequate in length.

Our conjecture is that the slow convergence in the local methods is due to the one-dimensional nature of this example, a view which can be supported by heuristic arguments about connectivity. Notwithstanding this difficulty, we have seen that the methods do perform well; but we cannot confidently claim we are in fact evaluating the posterior means.

(*b*) *Gamma-camera phantom data*

To investigate the performance of gamma-camera imaging, and to calibrate equipment, medical physicists occasionally conduct experiments in which the patient is replaced by a physical 'phantom', usually a perspex box containing radioactive liquid, emitting photons. The data we use here were kindly provided by Dr C. J. Gibson, of the Northern Regional Medical Physics Department at the Dryburn Hospital, Durham.

Figure 6 displays three gamma-camera data-sets as grey-scale images. Panel (*a*) is an image of a patient's liver: the existence of a very large lesion is evident from visual

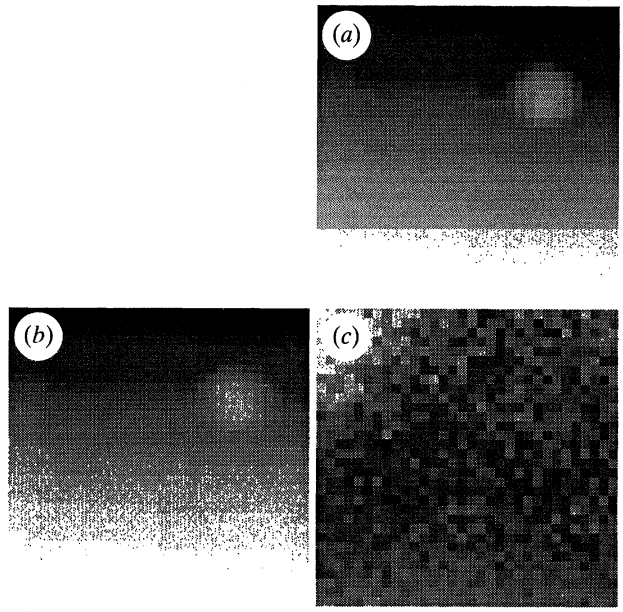


Figure 7. Reconstructions using global prior. (*a*) Posterior mean. (*b*) Fitted values. (*c*) Raw residuals.

Table 3. *Parameter estimates from two-dimensional examples*

(Phantom data. ($\delta = 30$. Number of sweeps = 2000 (after discarding 500). Global model, $\tau^2 = \{0.025, 0.00025, 0.01, 0.025, 0.025\}$; local difference, $\tau^2 = 500$; local laplacian, $\tau^2 = 180$.))

prior	β	α_0	α_1	α_2	γ_1	γ_2
global		48.88	0.4366	5.9674	24.67	22.93
local difference	1	48.51	0.6409	5.4295	24.18	22.85
local laplacian	10	48.08	0.6680	5.4402	24.33	22.78

(Thinned phantom data. ($\delta = 10$. Number of sweeps = 2000 (after discarding 500). Global model, $\tau^2 = \{0.025, 0.00025, 0.01, 0.025, 0.025\}$; local difference, $\tau^2 = 125$; local laplacian, $\tau^2 = 45$.))

prior	β	α_0	α_1	α_2	γ_1	γ_2
global		15.01	0.1211	1.7879	24.79	23.57
local difference	0.3	15.44	0.1802	1.6036	24.61	23.60
local laplacian	5.0	15.32	0.1800	1.6096	23.73	23.29

inspection. Panel (*b*) shows part of a phantom image. The source of the hole evident in this display is a steel ball 2 cm in diameter, suspended within the perspex box (pixels in this image are approximately 0.3 cm across). In both of these panels there is a high level of noise due to the Poisson variation: the maximum counts per pixel are only 248 and 288 respectively. There is interest in the medical physics community in the development of image analysis procedures for such data that perform well even at considerably smaller photon levels than this. This naturally means a worse signal-to-noise ratio, but is attractive from the point of view of patient safety (as isotope dosage can be reduced), and to reduce motion blur due to patient movement or isotope flow. To simulate lower photon levels, we have used a thinned image,

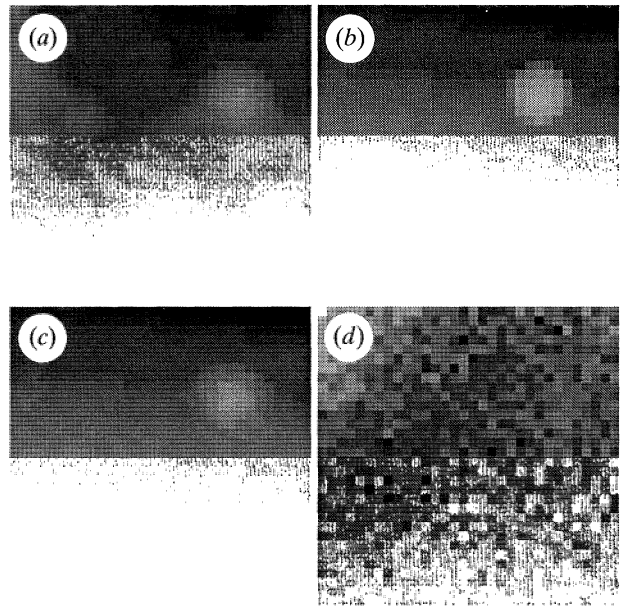


Figure 8. Reconstructions using local difference prior. (*a*) Posterior mean. (*b*) Post processed mean. (*c*) Fitted values. (*d*) Raw residuals.

displayed in figure 6*c*. This is found by binomial resampling from the image in figure 6*b*, retaining each photon, independently, with probability $\frac{1}{3}$. This clearly retains the Poisson character of the random variation, while producing a noisier image in which the maximum count per pixel is 95.

Our analysis of these data has assumed a bivariate gaussian point-spread function with $\sigma = 1.5$ pixels, truncated at $\pm 2\sigma$, a hole radius of 3.5 and depths 80 and 27 for the unthinned and thinned data-sets respectively. These values are based on the dimensions of the phantom, and information about the point-spread function provided by Dr Gibson. A similar modification to that used in our one-dimensional experiments was used to deal with edge effects, and again spatial discretization was imposed by integration, not sampling. The hyperparameter values and numbers of sweeps used are given in table 3.

Figures 7, 8 and 9 display results of our analysis of the phantom data, using each of the three methods, and figures 10, 11 and 12, the corresponding results for the binomially thinned data. In each case the estimated posterior mean is displayed, together with the post-processed output in the case of the pixel-based priors. The remaining panels display the 'fitted values' resulting from applying the point-spread function to the estimates of the truth, and the raw residuals, that is, pixelwise differences between data and fitted values. These latter images will be affected by various departures from the model: if all is well, we should expect to see random noise superimposed on a smooth trend representing deviation of the ramp from planarity. That seems to be the case in these figures. On the other hand, if the lesion had been incorrectly located, its depth or shape mis-specified, or if there had been in fact no lesions present, or more than one, then corresponding patterns would have been apparent in the residual image.

All of the methods produce similar estimates of the lesion location (γ_1, γ_2), and the ramp intercept (α_0) and gradient (α_1, α_2), on both the original and the thinned data-

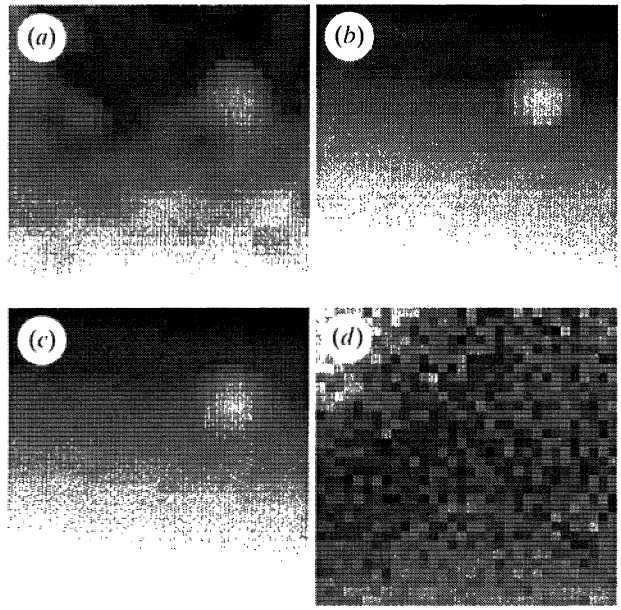


Figure 9. Reconstructions using local laplacian prior. (*a*) Posterior mean. (*b*) Post processed mean. (*c*) Fitted values. (*d*) Raw residuals.

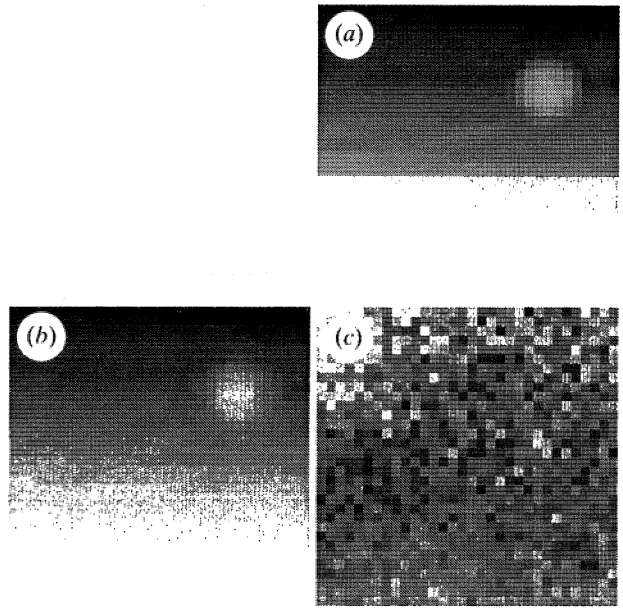


Figure 10. Reconstructions using global prior (thinned data). (*a*) Posterior mean. (*b*) Fitted values. (*c*) Raw residuals.

sets. These estimates are presented in table 3. It is particularly pleasing that even when the signal-to-noise ratio is as low as it is in the thinned data of figure 6*c*, these procedures do not break down.

In figure 13, we display traces of two monitoring statistics for each of the three

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

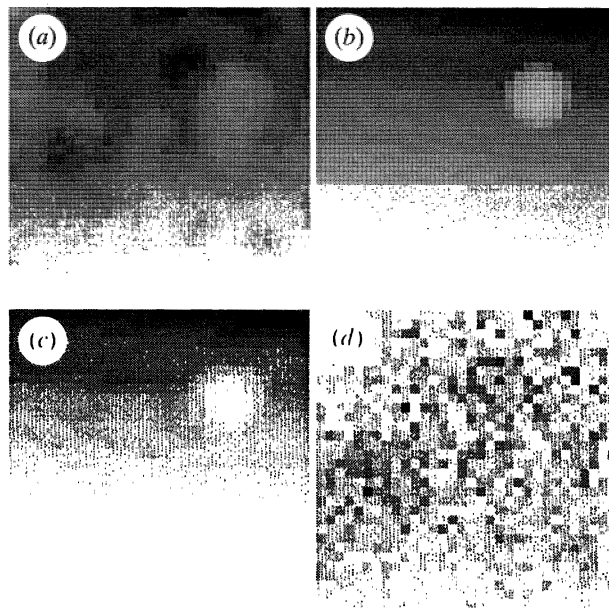


Figure 11. Reconstructions using local difference prior (thinned data). (*a*) Posterior mean. (*b*) Post processed mean. (*c*) Fitted values. (*d*) Raw residuals.

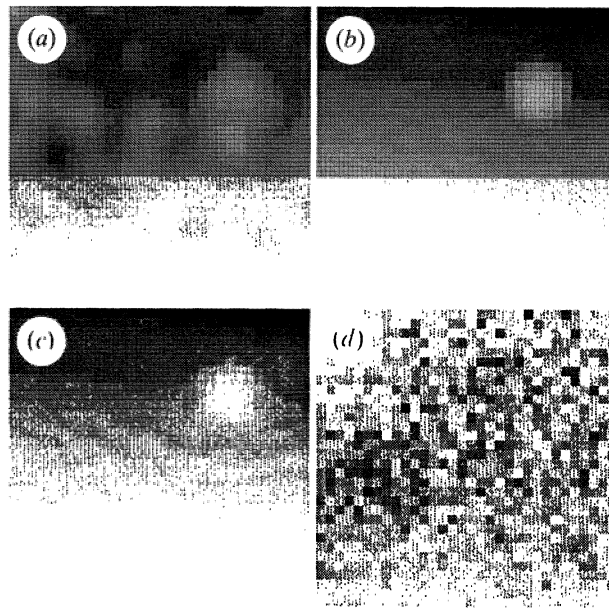


Figure 12. Reconstructions using local laplacian prior (thinned data). (*a*) Posterior mean. (*b*) Post processed mean. (*c*) Fitted values. (*d*) Raw residuals.

methods, and this time no difficulty is apparent. Monte Carlo variances for parameter estimates are given in table 4 for each of the methods.

We do not have estimated sampling variances for the estimates given in table 3, although we do intend to explore in future work a possible approach to obtaining such variances in bootstrap fashion, by exploiting the binomial resampling idea that

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

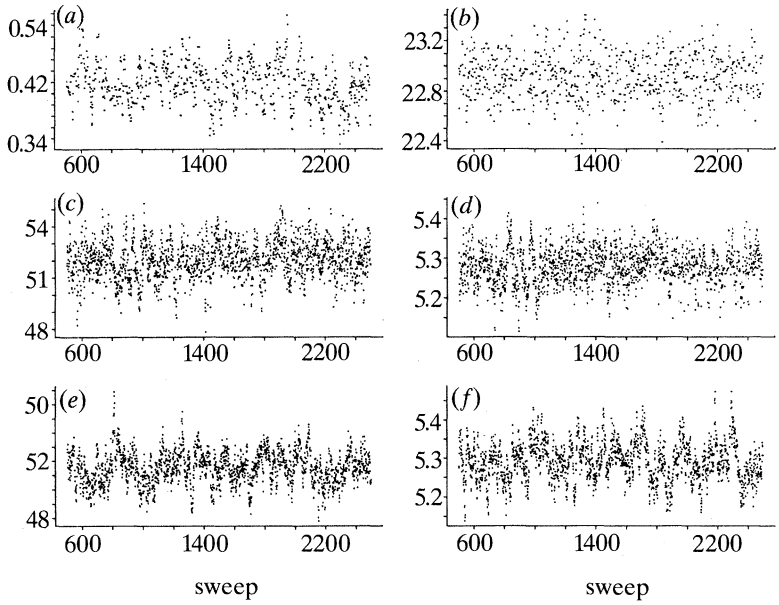


Figure 13. Monitoring statistics. Global prior: (*a*) horizontal gradient parameter; (*b*) vertical location parameter. Local difference prior: (*c*) intercept of LS plane; (*d*) vertical gradient of LS plane. Local laplacian prior: (*e*) intercept of LS plane; (*f*) vertical gradient of LS plane.

Table 4. *Monte Carlo variances*

Phantom data

prior	k	b	α_0	α_1	α_2	γ_1	γ_2
global	50	40	0.00903	1.229×10^{-5}	1.671×10^{-5}	6.316×10^{-5}	4.634×10^{-5}
local difference	100	20	0.00542	0.781×10^{-5}	0.816×10^{-5}	0.004432	0.0040727
local laplacian	50	40	0.01217	2.563×10^{-5}	2.330×10^{-5}	0.045221	0.0194146

Thinned phantom data

prior	k	b	α_0	α_1	α_2	γ_1	γ_2
global	50	40	1.513×10^{-3}	2.436×10^{-6}	4.012×10^{-6}	9.557×10^{-5}	2.797×10^{-4}
local difference	50	40	0.592×10^{-3}	1.521×10^{-6}	1.521×10^{-6}	0.19339	0.098933
local laplacian	100	20	4.146×10^{-3}	6.439×10^{-6}	5.681×10^{-6}	0.37729	0.119511

was used to produce the thinned data-set. Estimated sampling variances for maximum likelihood estimation of $(\alpha_0, \alpha_1, \alpha_2)$ in fitting the planar ramp, without the hole, to the unthinned data of figure 6*b* are approximately (0.724, 0.00142, 0.00147) respectively.

We are indebted to Chris Gibson for providing the data, and to the referees for their help in improving the paper. We acknowledge financial support for R.G.A. and provision of computing equipment by the Science and Engineering Research Council, through its Complex Stochastic Systems initiative.

References

Amit, Y., Grenander, U. & Piccioni, M. 1991 Structured image restoration through deformable templates. *J. Am. statist. Ass.* (In the press.)

Besag, J. 1974 Spatial interaction and the statistical analysis of lattice systems (with discussion). *Jl R. statist. Soc.* B **36**, 192–236.

Besag, J. 1983 Discussion of paper by P. Switzer. *Bull. Int. statist. Inst.* **50** (3), 422–425.

Besag, J. 1986 On the statistical analysis of dirty pictures (with discussion). *Jl R. statist. Soc.* B **48**, 259–302.

Besag, J., York, J. & Mollié, A. 1991 Bayesian image restoration, with two applications in spatial statistics (with discussion). *Ann. Inst. statist. Math.* **43**, 1–59.

Chow, Y., Grenander, U. & Keenan, D. M. 1988 Hands: a pattern theoretic study of biological shape. *Research notes in neural computing*, vol. 2. Berlin: Springer.

Clifford, P. 1986 Contribution to discussion of paper by J. Besag. *Jl R. statist. Soc.* B **48**, 284.

Dupuis, P., Geman, D., Horowitz, J. & Reynolds, G. 1991 Statistical inference on the shape of circumstellar disks from HST observations. Working paper, University of Massachusetts, Amherst, U.S.A.

Gelfand, A. E. & Smith, A. F. M. 1990 Sampling-based approaches to calculating marginal densities. *J. Am. statist. Ass.* **85**, 398–409.

Geman, D. 1991 Random fields and inverse problems in imaging. *Lecture Notes in Math.* Berlin: Springer. (In the press.)

Geman, S. & Geman, D. 1984 Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. Pattern Analys. Mach. Intell.* **6**, 721–741.

Geman, S. & McClure, D. 1987 Statistical methods for tomographic image reconstruction. *Bull. Int. statist. Inst.* **52** (4), 5–21.

Green, P. J. 1986 Contribution to discussion of paper by J. Besag. *Jl R. statist. Soc.* B **48**, 284–285.

Green, P. J. 1987 Penalized likelihood for general semi-parametric regression models. *Int. Statist. Rev.* **55**, 245–259.

Green, P. J. 1990 Bayesian reconstructions from emission tomography data using a modified EM algorithm. *IEEE Trans. Medical Imaging* **9**, 84–93.

Green, P. J. & Han, X.-L. 1991 Metropolis methods, gaussian proposals and antithetic variables. Stochastic models, statistical methods and algorithms in image analysis (ed. P. Barone, A. Frigessi & M. Piccioni). *Lecture Notes in Statist.* Berlin: Springer. (In the press.)

Hammersley, J. M. & Handscomb, D. C. 1964 *Monte Carlo methods.* London: Methuen.

Hastings, W. K. 1970 Monte Carlo simulation methods using Markov chains, and their applications. *Biometrika* **57**, 97–109.

Jennison, C. & Jubb, M. 1991 Aggregation and refinement in binary image restoration. In *Spatial statistics and imaging* (ed. A. Possolo), IMS Lecture Notes. (In the press.)

Jones, M. C. & Silverman, B. W. 1989 An orthogonal series density estimation approach to reconstructing position emission tomography images. *J. appl. Statist.* **16**, 177–191.

Larsson, S. A. 1980 Gamma camera emission tomography. *Acta radiologica Suppl.* 363.

Qian, W. & Titterington, D. M. 1991*a* Stochastic relaxations and E-M algorithms for Markov random fields. *J. statist. Comput. Simulation.* (In the press.)

Qian, W. & Titterington, D. M. 1991*b* Estimation of parameters in hidden Markov models. *Phil. Trans. R. Soc. Lond.* A **337**, 407–428. (This volume.)

Ripley, B. D. & Sutherland, A. I. 1990 Finding spiral structures in images of galaxies. *Phil. Trans. R. Soc. Lond.* A **332**, 477–485.

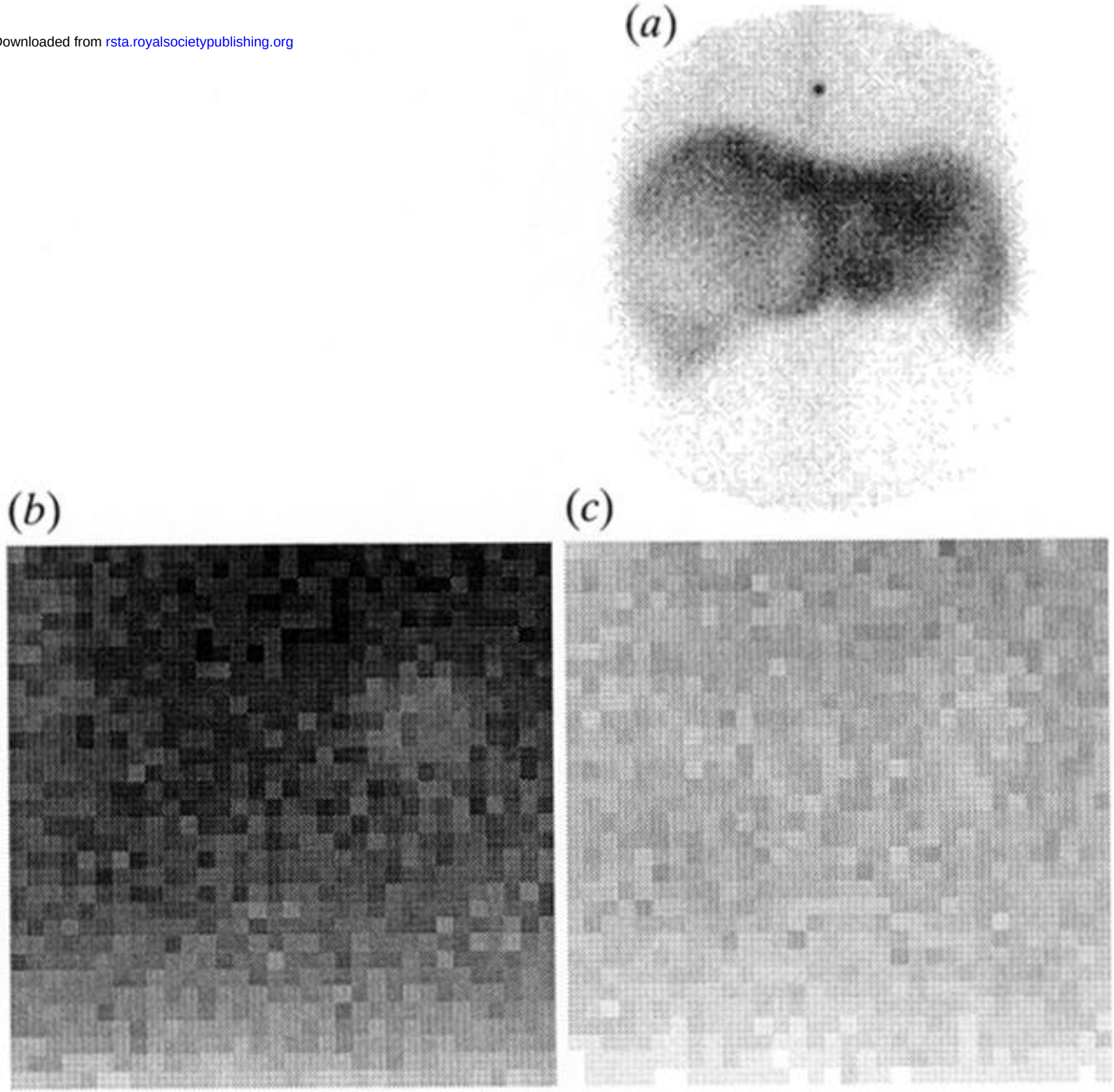
Sheehan, N. A. & Thomas, A. W. 1991 On the irreducibility of a Markov chain defined on a space of genotype configurations by a sampling scheme. (Submitted.)

Shepp, L. A. & Vardi, Y. 1982 Maximum likelihood reconstruction for emission tomography. *IEEE Trans. Medical Imaging* **1**, 113–122.

Smith, A. F. M. 1991 Bayesian computational methods. *Phil. Trans. R. Soc. Lond.* A **337**, 369–386. (This volume.)

Sokal, A. D. 1989 Monte Carlo methods in statistical mechanics: foundations and new algorithms. Cours de Troisième Cycle de la Physique en Suisse Romande, Lausanne.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES



igure 6. Gamma-camera data. (*a*) Image of liver. (*b*) Extract of phantom data. (*c*) Thinned phantom data.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

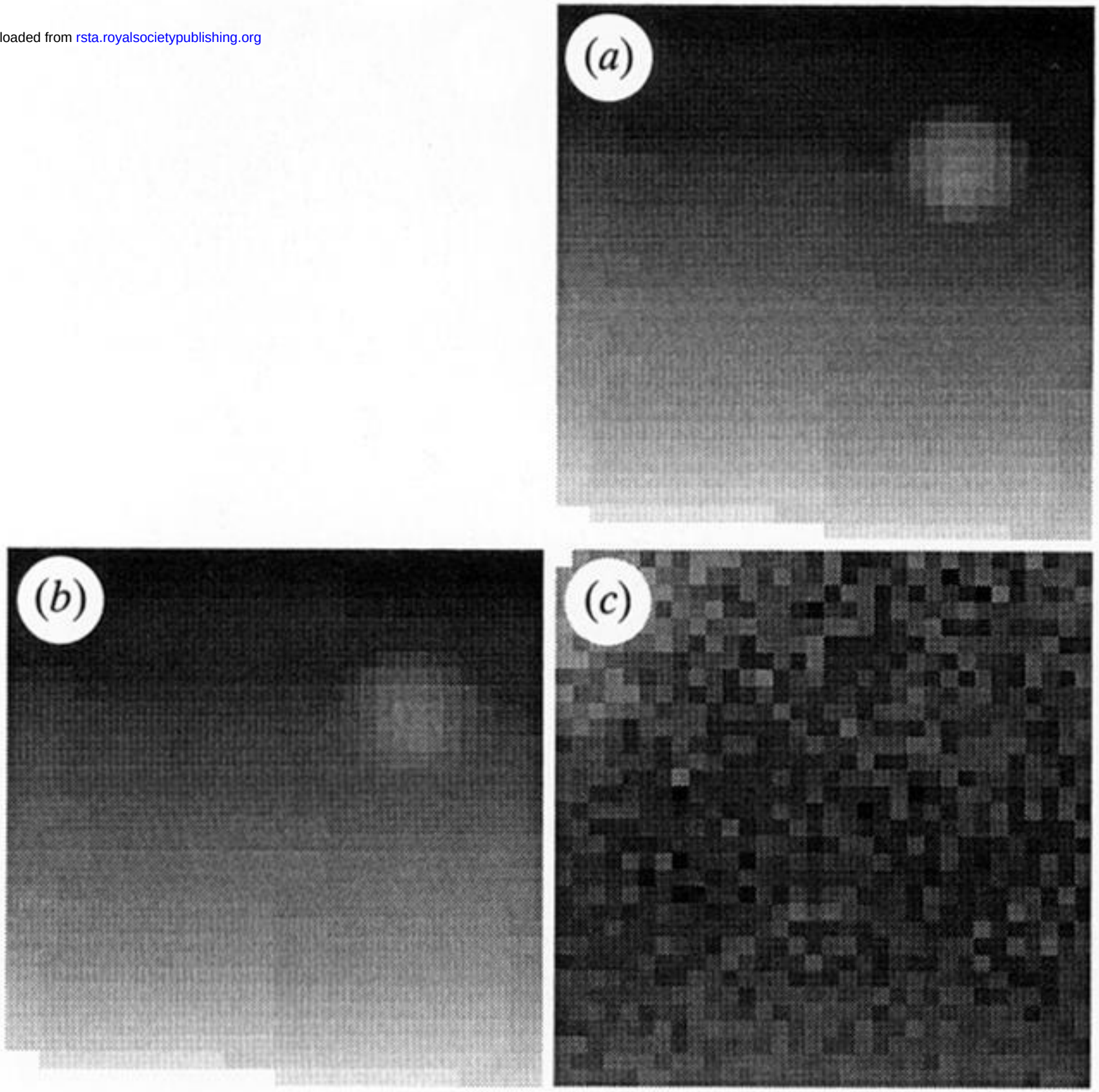


Figure 7. Reconstructions using global prior. (*a*) Posterior mean. (*b*) Fitted values. (*c*) Raw residuals.

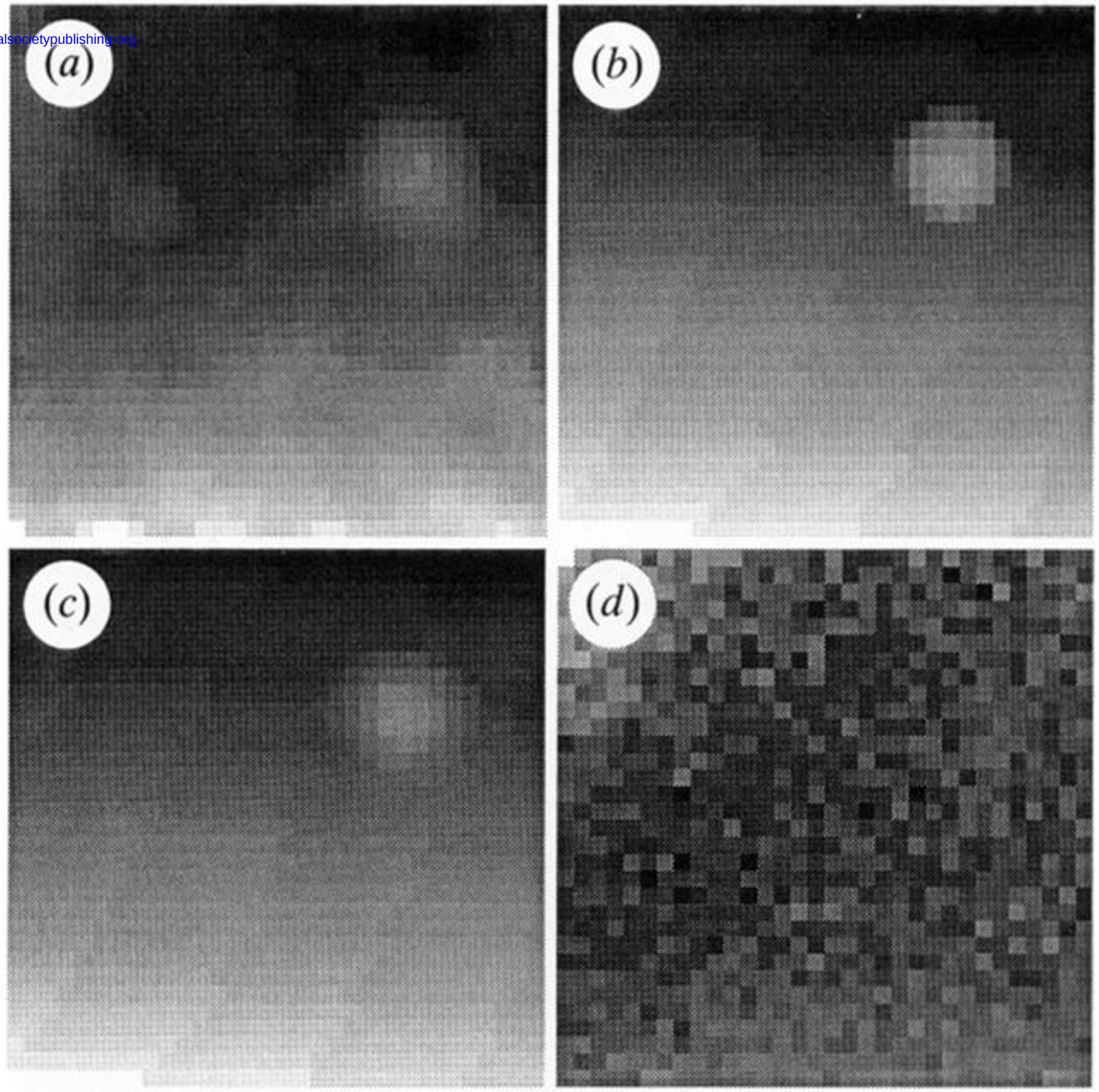


Figure 8. Reconstructions using local difference prior. (*a*) Posterior mean. (*b*) Post processed mean. (*c*) Fitted values. (*d*) Raw residuals.

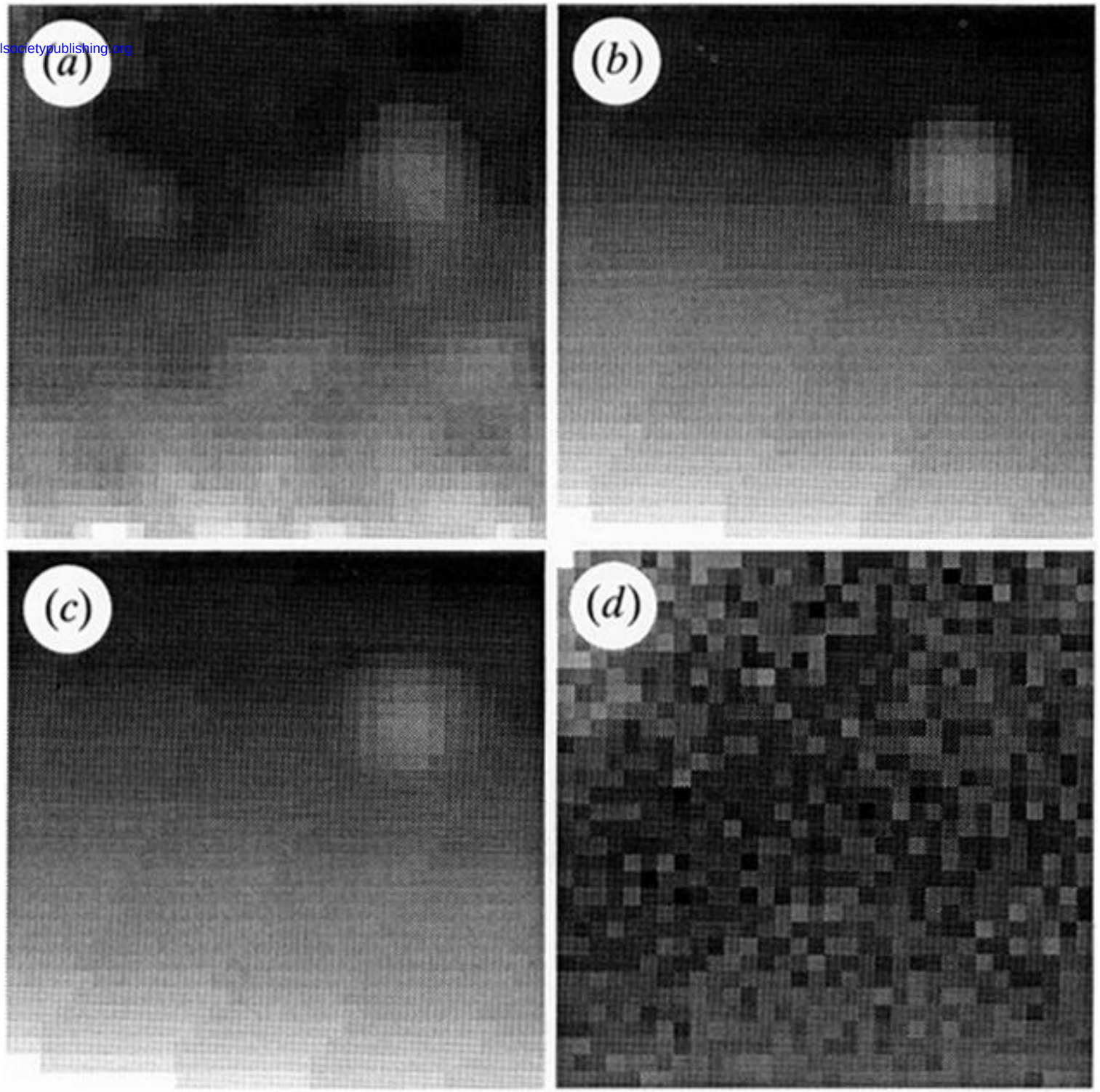
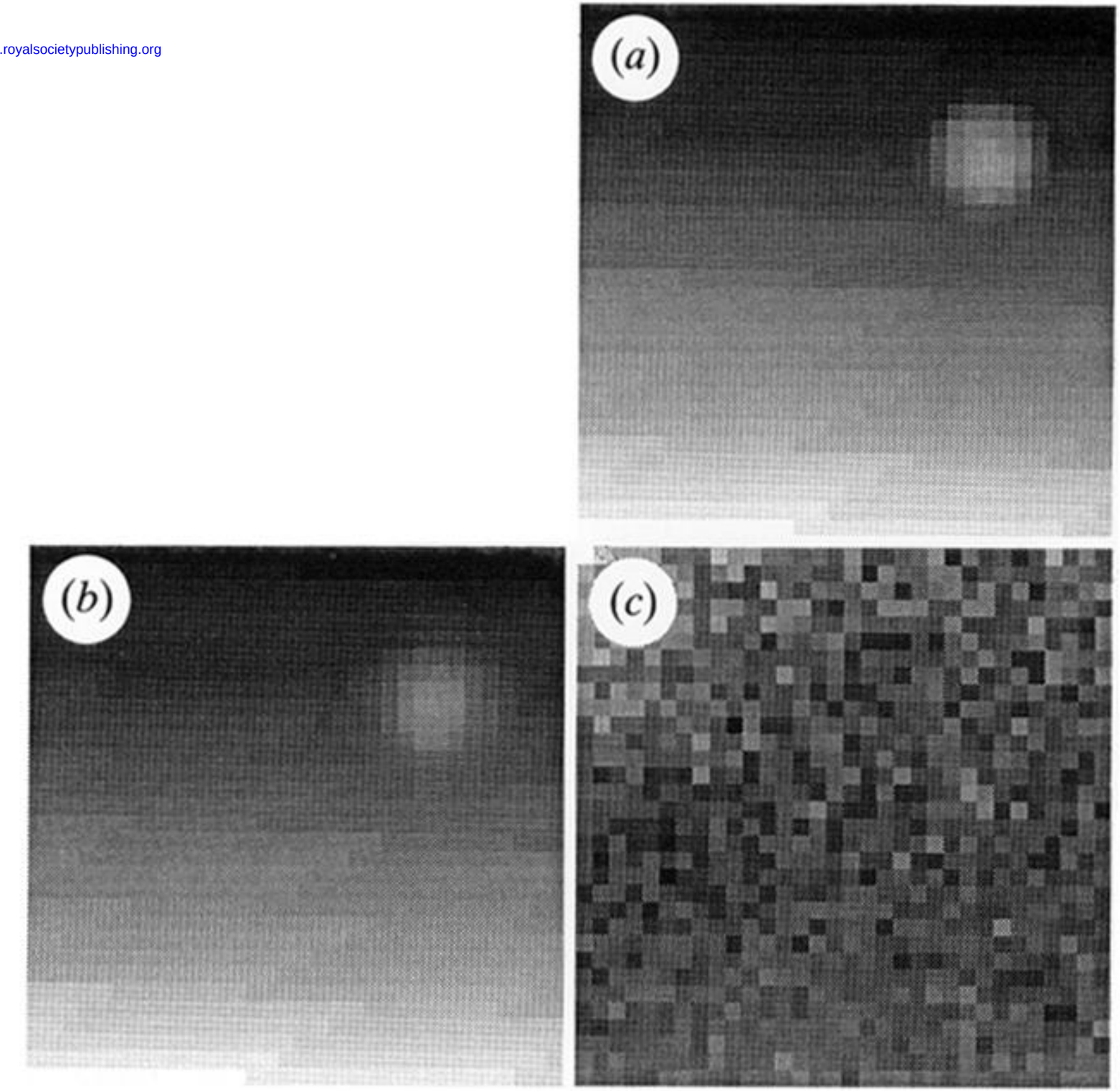


Figure 9. Reconstructions using local laplacian prior. (*a*) Posterior mean. (*b*) Post processed mean. (*c*) Fitted values. (*d*) Raw residuals.



igure 10. Reconstructions using global prior (thinned data). (*a*) Posterior mean. (*b*) Fitted values. (*c*) Raw residuals.

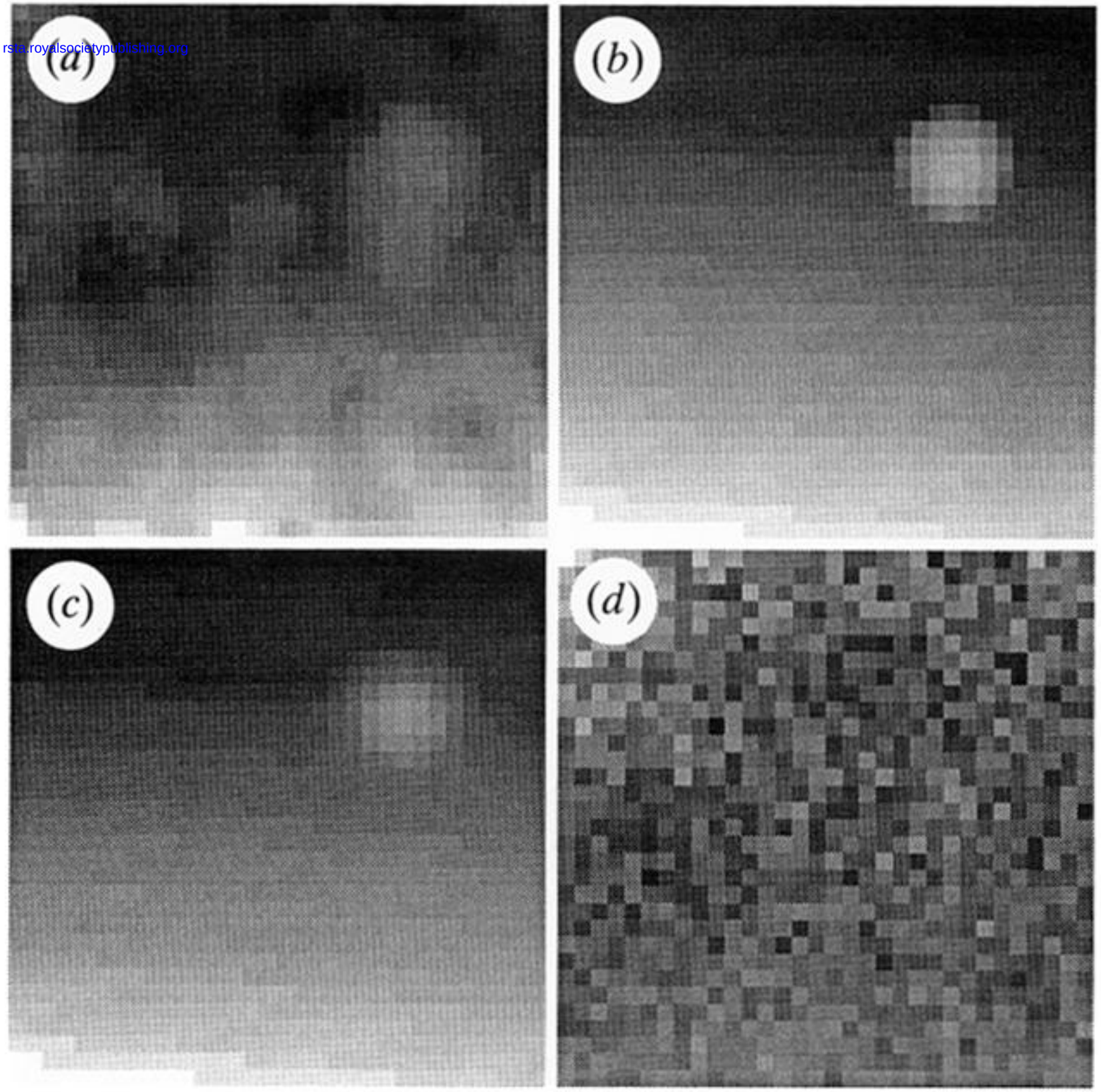
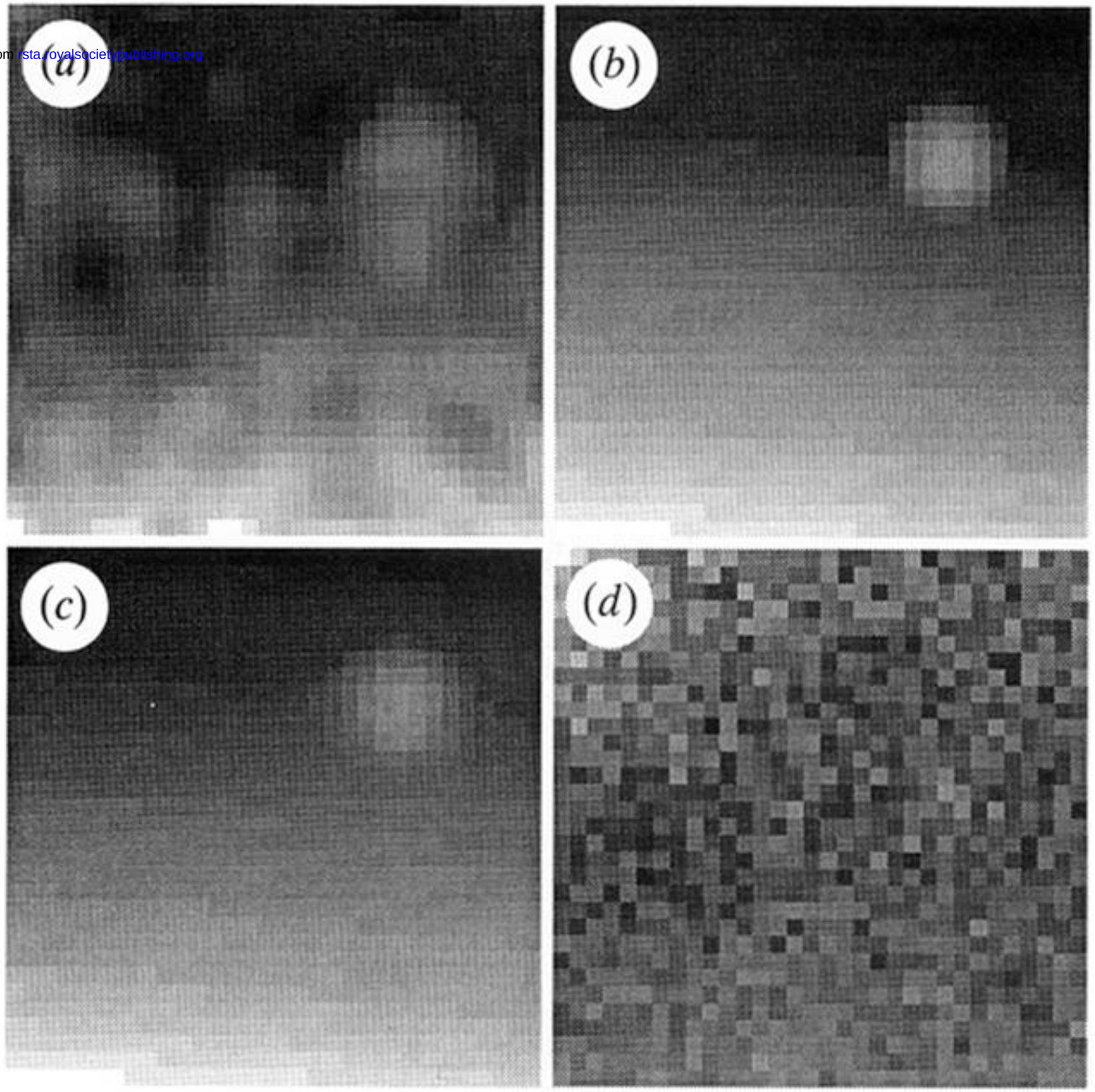


Figure 11. Reconstructions using local difference prior (thinned data). (*a*) Posterior mean. (*b*) Post processed mean. (*c*) Fitted values. (*d*) Raw residuals.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES



igure 12. Reconstructions using local laplacian prior (thinned data). (*a*) Posterior mean. (*b*) Post processed mean. (*c*) Fitted values. (*d*) Raw residuals.