Submitted to IEEE Transactions on Image Processing

Hierarchical Posterior Sampling for Continuous-State Random Fields

Paul Fieguth
Systems Design Engineering
University of Waterloo
Waterloo, Ontario, Canada

***Abstract*—The posterior sampling problem computes a random sample from a posterior distribution. Typically this problem is solved through Markov-Chain Monte-Carlo / Simulated Annealing, however these can be computationally challenging and slow to converge.**

In this paper we use a little-known property of multiscale statistical models to formulate a posterior sampler, exact in the case of Markov random fields, and approximate for other distributions. The proposed approach benefits from and builds upon past work on multiscale model inference and approximation, yielding a fast approach to sampling continuous-state images.

I. Introduction

The estimation of images and random fields from sparse and/or noisy data is a highly-developed discipline, to the point where methods such as least-squares estimation, simulated annealing, and wavelet shrinkage appear in standard textbooks[3], [20]. Indeed, beyond the standardization of such methods, there also exists a variety of mature, efficient estimators.

For the majority of current image analysis problems, characterized by densely-measured images from a camera, typically showing macroscopic scenes (e.g., people, faces, houses, trees, cities, etc.), available estimation algorithms are more than adequate, and the challenge lies, instead, in the high-level modeling and interpretation of such images, not the subject of this paper.

However there exists a comparatively small, but important, class of problems, mostly scientific, in which the "image" is governed by some mathematical behaviour (the prior model), and the measurements are sparse (constrained by physics, time, and/or cost), of which important examples include satellite remote sensing (sparsity by orbital constraints) and imaging-MRI (sparsity by time constraints). In particular, the research of this paper is motivated by a current interdisciplinary research project at the University of Waterloo involving the imaging and modeling of porous media, such as limestone, concrete, cartilage, or sand. Although many studies of porous media are discrete-state (e.g., rock versus pore), we focus on continuous-state models, appropriate in our MRI context (in which we measure pore *density*, since pores may be too small to resolve) and in many remotely-sensed contexts.

Throughout this paper we assume that the prior model is known. That is, we focus exclusively on the computational aspects of posterior sampling. The system identification problem to learn such models is of great importance, but beyond the present scope.

Suppose we have a Gaussian random field $\underline{z}$ (equivalently, a grid of image pixels) on a lattice $\mathcal{L}$, with a prior probability density $p(\underline{z})$ and linear measurements

$$\underline{m} = C\underline{z} + \underline{v} \tag{1}$$

with measurement noise $\underline{v}$, where all quantities are column vectors. A *random* sample from the posterior distribution is typically solved via Markov-Chain Monte-Carlo / Simulated Annealing [11], [12], [13], [21], [22], which can be computationally challenging and slow to converge. In this paper we use multiscale models, where the domain-decomposition of the multiscale approach leads to an efficient posterior sampler which benefits from a variety of previously-developed multiscale tools and models.

Section 2 briefly describes the mathematical background behind sampling, followed by a development of the multiscale approach in Section 3 and experimental results in Section 4.

II. Background

We briefly move through the sequence of Prior Sampling, Estimation, Posterior Sampling, and Dynamic Models. Suppose an n-element random field $\underline{z} \in R^n$ has a known Normal distribution

$$\underline{z} \sim \mathcal{N}(\underline{\mu}, P). \tag{2}$$

Finding the eigendecomposition of P, we write

$$P = V\Lambda V^T,\ L = V\Lambda^{1/2}, \tag{3}$$

where L is referred to as a matrix square-root of P. In practice there are a variety of ways to compute a matrix square root, most notably the $\mathcal{O}(n^3)$ Cholesky decomposition. From L, a random sample may be found as

$$\underline{z} = \underline{\mu} + L\underline{w},\ \underline{w} \sim \mathcal{N}(0, I). \tag{4}$$

Next, given linear measurements

$$\underline{m} = C\underline{z} + \underline{v},\ \ \underline{v} \sim \mathcal{N}(0, R) \tag{5}$$

the linear least-squares estimates may be computed, as usual, as

$$\hat{\underline{z}} = E[\underline{z}|\underline{m}] = \underline{\mu} + \left(P^{-1} + C^T R^{-1} C\right)^{-1} C^T R^{-1}(\underline{m} - C\underline{\mu}) \tag{6}$$

$$\tilde{P} = \mathrm{cov}(\tilde{\underline{z}}) = \left(\mathrm{P}^{-1} + \mathrm{C}^{\mathrm{T}}\mathrm{R}^{-1}\mathrm{C}\right)^{-1} \tag{7}$$

The former computation of $\hat{\underline{z}}$ is relatively straightforward, typically $\mathcal{O}(n^{1.5})$ to $\mathcal{O}(n^2)$, and is easily cast into a linear-systems framework for iterative solution; the latter $\mathcal{O}(n^3)$ computation of $\tilde{P}$ is much more difficult, and is omitted in many contexts.

A *posterior sample* of $\underline{z}$ is one constrained by the measurements $\underline{m}$, that is, a random sample of

$$\underline{z}|\underline{m} \sim \mathcal{N}(\hat{\underline{z}}, \tilde{P}). \tag{8}$$

If the $\mathcal{O}(n^3)$ computation of $\tilde{P}$ in (7) can be undertaken, then the parameters of this distribution are known. If we let $\tilde{L}$ be the matrix square root of $\tilde{P}$, then the posterior samples are given by

$$(\underline{z}|\underline{m}) = \hat{\underline{z}} + \tilde{L}\tilde{\underline{w}},\ \ \tilde{\underline{w}} \sim \mathcal{N}(0, I). \tag{9}$$

This requires finding $\hat{\underline{z}}$ $\mathcal{O}(n^{1.5})$, the posterior covariance $\tilde{P}$ $\mathcal{O}(n^3)$ and its square root $\mathcal{O}(n^3)$.

Finally, *if* pieces of the random field obey a dynamic relationship then sampling may be simplified. For example, if the columns $\underline{z}_i$ of an image $Z = [\underline{z}_o\ \underline{z}_1\ \ldots]$ obey a standard Gauss-Markov model

$$\underline{z}_{i+1} = A_i\underline{z}_i + B_i\underline{w}_i,\ \ \underline{z}_o \sim \mathcal{N}(\underline{\mu}_o, P_o), \tag{10}$$

where $\underline{w}_i$ is a white Gaussian noise process, then only matrix $P_o = L_o L_o^T$ need be decomposed and the sampling process proceeds iteratively, per (10), initialized with

$$\underline{z}_o = L_o\underline{w}_o,\ \ \underline{w}_o \sim \mathcal{N}(0, I). \tag{11}$$

[0]The support of the Natural Science & Engineering Research Council of Canada is acknowledged.
EDICS Category FLT-LFLT, MRP-OTHR

This appears to be a convenient solution to the sampling problem, however two issues remain:

1) In many cases the process parameters B_i are not, in fact, known; rather (10) takes the form

$$\underline{z}_{i+1} = A_i \underline{z}_i + \underline{q}_i, \;\; \underline{z}_o \sim \mathcal{N}(\underline{\mu}_o, P_o), \;\; \underline{q}_i \sim \mathcal{N}(0, Q_i) \qquad (12)$$

where $\underline{q}_i$ is a noise process with covariance Q_i. To generate a posterior sample requires finding the square root of each Q_i. The key to efficiency is to keep the state dimension of $\underline{z}_i$, and thus the size of matrix Q_i, as small as possible.

2) In order to do posterior sampling, it is really a dynamic relationship for the estimation *errors* $\underline{\tilde{z}} = \underline{z} - \underline{\hat{z}}$ which we require:

$$\underline{\tilde{z}}_{i+1} = \tilde{A}_i \underline{\tilde{z}}_i + \underline{\tilde{q}}_i, \;\; \underline{\tilde{z}}_o \sim \mathcal{N}(\underline{\hat{z}}_o, \tilde{P}_o), \;\; \underline{\tilde{q}}_i \sim \mathcal{N}(0, \tilde{Q}_i). \qquad (13)$$

The error process of the Kalman smoother *does* obey such a dynamic process [1], [2].

III. Multiscale Sampling

Motivated by the long history of success of the Kalman filter, there have been efforts to similarly model random processes over scale rather than over time. One particular hierarchical framework, the multiscale one[4], [8], [14], [17], has seen considerable effort and model development. A scale-to-scale Gauss-Markov process is asserted,

$$\underline{z}(s) = A(s)\underline{z}(ps) + B(s)\underline{w}(s), \;\; \underline{w}(s) \sim \mathcal{N}(\underline{0}, I), \;\; \underline{z}(0) \sim \mathcal{N}(\underline{\mu}_0, P_0), \qquad (14)$$

where s is an index on the nodes of a tree, 0 is the node corresponding to the tree root, and p is an operator, returning the parent ps of node s. This model has a clear relationship with (10), except that in (14) a given node s may have more than one descendant.

The usual arrangement, for two-dimensional random fields, is a quad-tree representation, in which the leaf-nodes of the tree represent the pixels of the random field of interest, whereas the higher-level nodes serve to produce (to the extent possible) the desired statistical structure at the finest scale.

The immediate question is to what extent (14) can model random fields of interest. Two-dimensional Markov random fields can be represented exactly [17], and there have been successes in modeling fractals [5], $1/f$-like processes [8], [9], and results for more general stochastic realization [6], [15].

We define states $\underline{z}(s) = F(s)\underline{z}$, where $F(s)$ is highly sparse (such as the sparse sampling of boundary pixels in Figure 1). The node-to-node cross-statistics can be computed as

$$P(s_1, s_2) = F(s_1) P F^T(s_2) \qquad (15)$$

where P is the covariance of $\underline{z}$. The multiscale model parameters follow as

$$\begin{aligned} A(s) &= P(s, ps) P^{-1}(ps, ps) && (16) \\ B(s)B^T(s) &= P(s,s) - A(s) P(ps, ps) A^T(s). && (17) \end{aligned}$$

The estimator [4], [8], [17] which follows from (14) greatly resembles the Kalman smoother, producing estimates $\underline{\hat{z}}(s)$ and error covariances $\tilde{P}(s)$ at each node, based on all of the measurements on the entire tree. The computed estimates are exact, and any approximation lies in the choice of model itself.

The algorithm is normally highly efficient, although this depends greatly on the dimension of $\underline{z}(s)$ at coarse scales [14]. A typical example is shown in Figure 1; although the sparse pixellation in the state is exaggerated in the figure, the key point is that the size of the state will, in any case, be only a tiny fraction of the size of the entire image.

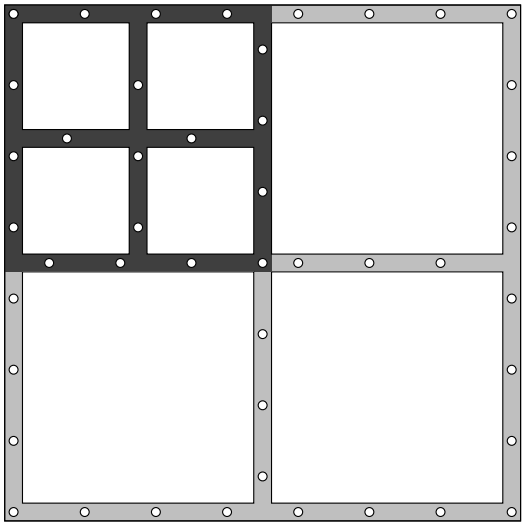


Fig. 1. A typical state arrangement at the coarsest level of a 2D multiscale quad-tree (light grey) and the top-left state at the next scale (dark grey). The state $F(s)\underline{z}$ is a subsampled set (circles) of boundary pixels. The degree of subsampling is a function of the prior statistics, accuracy requirements, and computational limitations.

The *key* benefit of the multiscale framework, for the context of this paper, is a mostly-overlooked result [16] which shows that the multiscale error process $\underline{\tilde{z}}(s)$ *also* obeys a multiscale model:

$$\underline{\tilde{z}}(s) = \tilde{A}(s)\underline{\tilde{z}}(ps) + \underline{\tilde{w}}(s), \;\; \underline{\tilde{w}}(s) \sim \mathcal{N}(0, \tilde{Q}(s)), \;\; \underline{\tilde{z}}(0) \sim \mathcal{N}(\underline{\hat{z}}(0), \tilde{P}_0), \qquad (18)$$

where $\tilde{A}(s), \tilde{Q}(s)$ are expressed [16] in terms of the model parameters $A(s), B(s)$, prior covariances $P(s)$, and estimation error covariances $\tilde{P}_u(s)$:

$$\begin{aligned} \tilde{A}(s) &= \tilde{P}_u(s) G^T(s) \tilde{P}_u^{-1}(ps) && (19) \\ \tilde{Q}(s) &= \tilde{P}_u(s) - \tilde{P}_u(s) G^T(s) \tilde{P}_u^{-1}(ps) G(s) \tilde{P}_u(s) && (20) \\ G(s) &= P(ps) A^T(s) P^{-1}(s). && (21) \end{aligned}$$

The original intent of the smoothing-error model was predominantly for other purposes — for data fusion, incorporating new measurements into existing estimates by running the estimator on the new model $\tilde{A}(s), \tilde{Q}(s)$, and for the computation of inter-node posterior estimation error covariances. However in the context of this paper the availability of a multiscale smoothing error model leads to an approach to hierarchical posterior sampling, following the dynamic approach discussed in the previous section:

The multiscale estimator computes $\underline{\hat{z}}(s), \tilde{P}(s)$ at each node on the tree.

We then compute the smoothing error model $\tilde{A}(s), \tilde{Q}(s)$.

Use a Cholesky decomposition compute the matrix square-roots

$$\tilde{L}(0) = \left(\tilde{P}(0)\right)^{1/2}, \;\; \tilde{L}(s) = \left(\tilde{Q}(s)\right)^{1/2}, s \neq 0 \qquad (22)$$

Sample the smoothing error process, starting from the root node and proceeding down the tree:

$$\begin{aligned} (\underline{\tilde{z}}|\underline{m})(0) &= \tilde{L}(0)\underline{\tilde{w}}(0), \;\; \underline{\tilde{w}}(0) \sim \mathcal{N}(0, I) && (23) \\ (\underline{\tilde{z}}|\underline{m})(s) &= \tilde{A}(\underline{\tilde{z}}|\underline{m})(ps) + \tilde{L}(s)\underline{\tilde{w}}(s), \;\; \underline{\tilde{w}}(s) \sim \mathcal{N}(0, I) && (24) \end{aligned}$$

Finally, add the mean (the estimates) to produce the posterior sample:

$$(\underline{z}|\underline{m})(s) = \underline{\hat{z}}(s) + (\underline{\tilde{z}}|\underline{m})(s) \qquad (25)$$

If the state $\underline{z}(s)$ has dimension $d(s)$, then each of the estimation, smoothing, and square-root steps have complexity $\mathcal{O}(d(s)^3)$ per node. For a two-dimensional n-pixel Markov process, modeled as in Figure 1, $d(0) \propto \sqrt{n}$ and thus total complexity is $\mathcal{O}(n^{1.5})$.

With the principle of the method in place, the following section demonstrates the approach, its versatility, and computational issues.

IV. Results

Two prior models were used to illustrate and assess the proposed approach: a thin-plate[1] Markov model and a Markov wood-grain texture, the coefficients shown in Table I.

Figure 2 illustrates the posterior sampling process. Because of stationarity and periodicity, the prior samples can be generated very efficiently using standard FFT methods, however the aperiodic structure of the measurements implies that the FFT cannot be used for posterior sampling[2]. The sampled estimation error $\underline{\tilde{z}}|\underline{m}$ shows that part of the stochastics not constrained by the measurements; indeed, the constrained domain is clearly visible down the middle of the image.

The multiscale model makes no assumptions regarding symmetry or stationarity, so it applies to non-isotropic models, such as in Figure 3, or a spatially nonstationary case such as in Figure 4. This latter figure shows a process consisting of two independent parts (inside and outside of a disk). The nonstationary structure of the decoupling is preserved, clearly visible in the sampled images. Note that the decoupled model is not in any way learned from the measurements; it is specified *a priori*.

Figure 5 shows sampled results for the types of scenarios motivating this work. The imaging-MRI device being applied to 3D samples of porous media can take measurements pointwise, in planes, or volumetrically. Limiting our studies to the two-dimensional domain, we can approximate these scenarios by looking at a single 2D cross section, such that measured planes become lines (here rows or columns), and low-resolution volumetric measurements become low-resolution planar ones.

It should be noted that cases (a-c) and (d) are fundamentally different: in (a-c) we measure a subset of the finest-scale pixels, in (d) we measure only low-resolution averages of 4×4 blocks of pixels, with no measurements at all at the finest scale. Because the multiscale framework does model the process at coarser scales, accommodating such low-resolution measurements is straightforward and natural.

In each case, two posterior samples are shown to emphasize the random sampling process. It is quite clear that the degree to which the two posterior samples are similar is a function of the degree to which the random field is specified by the measurements, with significant variations visible in (a), and much less so in (c) and (d). Although (d) may appear to take on aspects of super-resolution methods[7], [18], in posterior sampling we are adding details from the error process $\underline{\tilde{z}}$ which is, by definition, orthogonal or *not* inferable from the measurements, whereas super-resolution methods seek to infer fine-scale details which *are* inferable from the measurements.

Finally, Figure 6 looks at issues of computational complexity related to multiscale posterior sampling. Given a prior model, there are two questions to ask in undertaking a multiscale implementation:

1) What state sampling $F(s)$, as in Figure 1, is needed / required to produce adequate results?
2) Are there other multiscale ideas, such as averaging multiple models [8], overlapping trees [14], model realization [6], [15] or possibly multiply-rooted trees [10] which may be used?

Panel (a) of Figure 6 has a dense state definition, in which *every* boundary pixel is included in the state, such that the coarsest state contains 759 elements. If we do not require the model to be top-bottom and left-right periodic, then by eliminating boundary pixels in the coarse scales (b), a huge reduction in effort is realized. Next (c), because of the substantial horizontal correlation in this particular model, it is not necessary to keep *every* pixel in the horizontal direction, so further reductions are possible. Of course, if the state reduction is continued too far (d) decorrelative artifacts will appear along tree boundaries. Finally (e), an overlapping approach [14] may allow a reduced-state model to produce good results, as is the case here.

Although the results are shown here for only one example, extensive work in model inference and realization implies the extensibility of these results to other contexts. In any event, very reasonable results can be obtained for 128×128 images in less than 30 seconds. By comparison, a direct approach to sampling via Cholesky decomposition would take hours on a 2GHz workstation, in addition to problems with matrix storage, whereas FFT and steady-state Kalman filters may be used only in special circumstances with stationarity.

V. Conclusions

We have developed a multiscale approach for posterior sampling - the random sampling from the posterior distribution of a prior model plus constraints due to measurements, a method particularly useful in contexts of sparse measurements.

The method of this paper does not, to be sure, address all of the challenges raised by the 3D modeling and imaging of porous media. In particular, the most pressing extension to this work is the development of more meaningful prior models, in particular, the inference of prior models from data.

VI. Acknowledgements

We thank the anonymous reviewers for helpful comments in earlier revisions of this manuscript.

References

[1] F. Badawi, A. Lindquist, M. Pavon, "A Stochastic Realization Approach to the Smoothing Problem," *IEEE Trans. Automatic Control* (24), pp.878-887, 1979

[2] M. Bello, A. Willsky, B. Levy, "Construction and applications of discrete-time smoothing error models," *Int. J. of Control* (50), pp.203-223, 1989

[3] K. Castleman, *Digital image processing*, Prentice Hall, 1996

[4] K. Chou, A. Willsky, A. Benveniste, "Multiscale Recursive Estimation, Data Fusion, and Regularization", *IEEE Trans. on Automatic Control* (39) #3, pp.464–478, 1994

[5] M. Daniel, A. Willsky, "A multiresolution methodology for signal-level fusion and data assimilation with applications to remote sensing," *Proceedings of the IEEE*, (85), pp.164–180

[6] M. Daniel, A. Willsky, "The modeling and estimation of statistically self-similar processes in a multiresolution framework," *IEEE Trans. Information Theory* (45) #3 , pp.955–970, 1999

[7] M. Elad, A. Feuer, "Superresolution Restoration of an Image Sequence: Adaptive Filtering Approach," *IEEE Trans. on Image Processing* (8) #3, 1999

[8] P. Fieguth, W. Karl, A. Willsky, C. Wunsch, "Multiresolution Optimal Interpolation and Statistical Analysis of TOPEX/POSEIDON Satellite Altimetry", *IEEE Trans. Geoscience and Remote Sensing* (33) #2, pp.280–292, 1995

[9] P. W. Fieguth, A. S. Willsky, "Fractal Estimation Using Models on Multiscale Trees", *IEEE Transactions on Signal Processing* (44) #5, pp.1297–1300, 1996

[10] P. Fieguth, "Multiply-Rooted Multiscale Models for Large-Scale Estimation," *IEEE Image Processing* (10) #11, pp.1676–1686, 2001

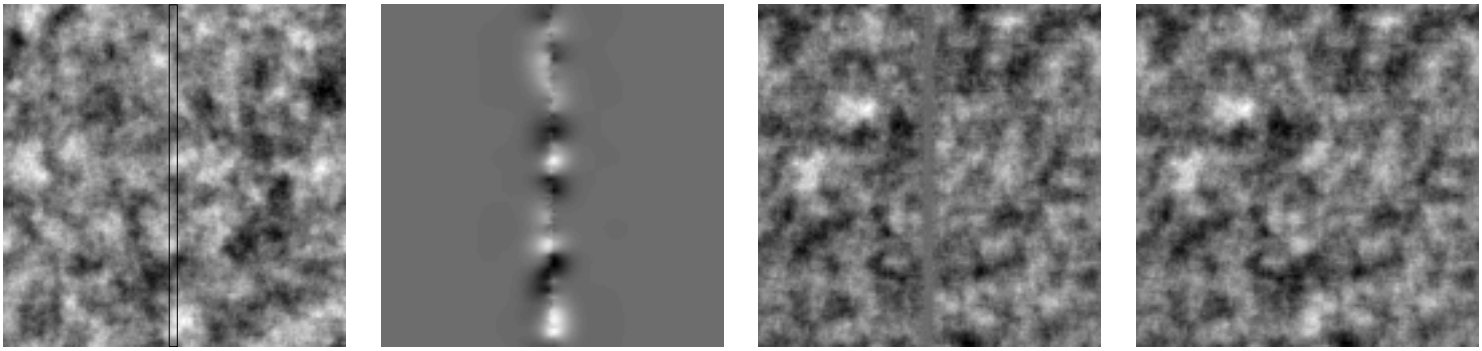
[11] A. Gelfand, S. Hills, A. Racine-Poon, A. Smith, "Illustration of Bayesian Inference in Normal Data Models Using Gibbs Sampling," *J. American Statistical Association* (85), pp.972–985, 1990

[1] The discretization of a second-order model, penalizing first and second order spatial derivatives, like a steel plate.

[2] In the artificial, special case that both prior and measurements are stationary / periodic, then indeed the FFT can be used to compute estimates, error variances, and posterior samples.

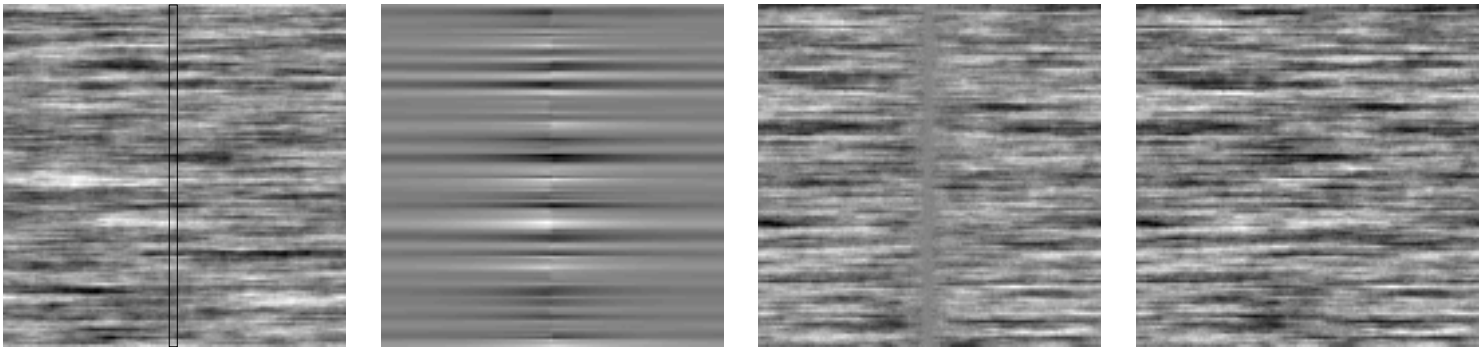
0	0	1	0	0	0	-0.0091	0.0517	0.0008	0
0	2	-8	2	0	0.0058	0.1405	-0.5508	0.1164	0.0085
1	-8	20.01	-8	1	-0.0139	-0.2498	1.0000	-0.2498	-0.0139
0	2	-8	2	0	0.0085	0.1164	-0.5508	0.1405	0.0058
0	0	1	0	0	0	0.0008	0.0517	-0.0091	0
		Thin Plate					Wood Grain		

TABLE I
MARKOV COEFFICIENTS FOR THE "THIN-PLATE" AND "WOOD-GRAIN" MODELS.



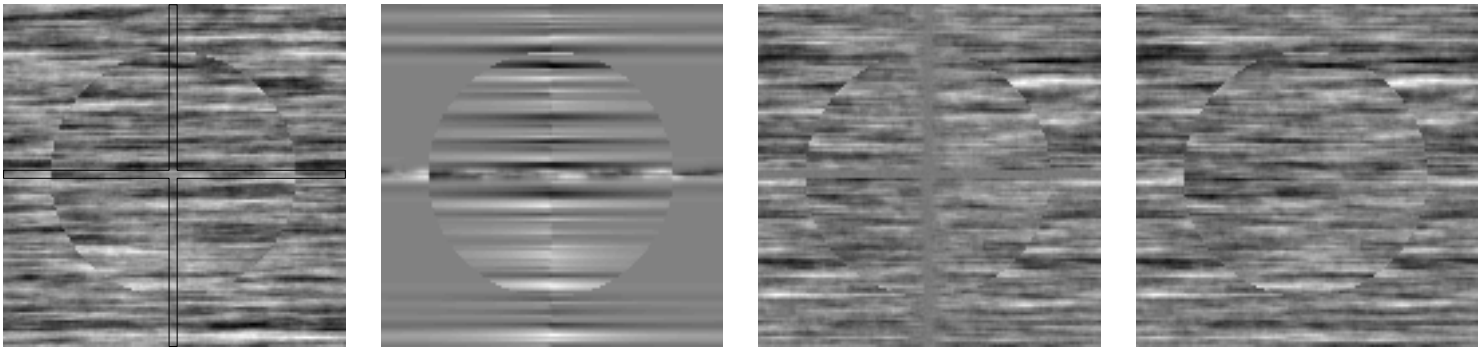
$\underline{z}$ $\quad$ $\underline{\hat{z}}|\underline{m}$ $\quad$ $\underline{\tilde{z}}|\underline{m}$ $\quad$ $\underline{z}|\underline{m}$

Fig. 2. The process of posterior sampling. The leftmost two panels show a sample from the prior model and estimates based on the central measured column. The third panel shows the sampled estimation error, where a low-variance zero-mean band can be seen where the estimation uncertainties are small, around the measurements. The final panel shows the sampled posterior.



$\underline{z}$ $\quad$ $\underline{\hat{z}}|\underline{m}$ $\quad$ $\underline{\tilde{z}}|\underline{m}$ $\quad$ $\underline{z}|\underline{m}$

Fig. 3. As in Figure 2, repeated for the case of an anisotropic prior model, a wood-grain like texture.



$\underline{z}$ $\quad$ $\underline{\hat{z}}|\underline{m}$ $\quad$ $\underline{\tilde{z}}|\underline{m}$ $\quad$ $\underline{z}|\underline{m}$

Fig. 4. As in Figure 2, but for a nonstationary model. We have two decoupled domains - one inside the circle, and the other outside. Measurements are taken in the central column and row.

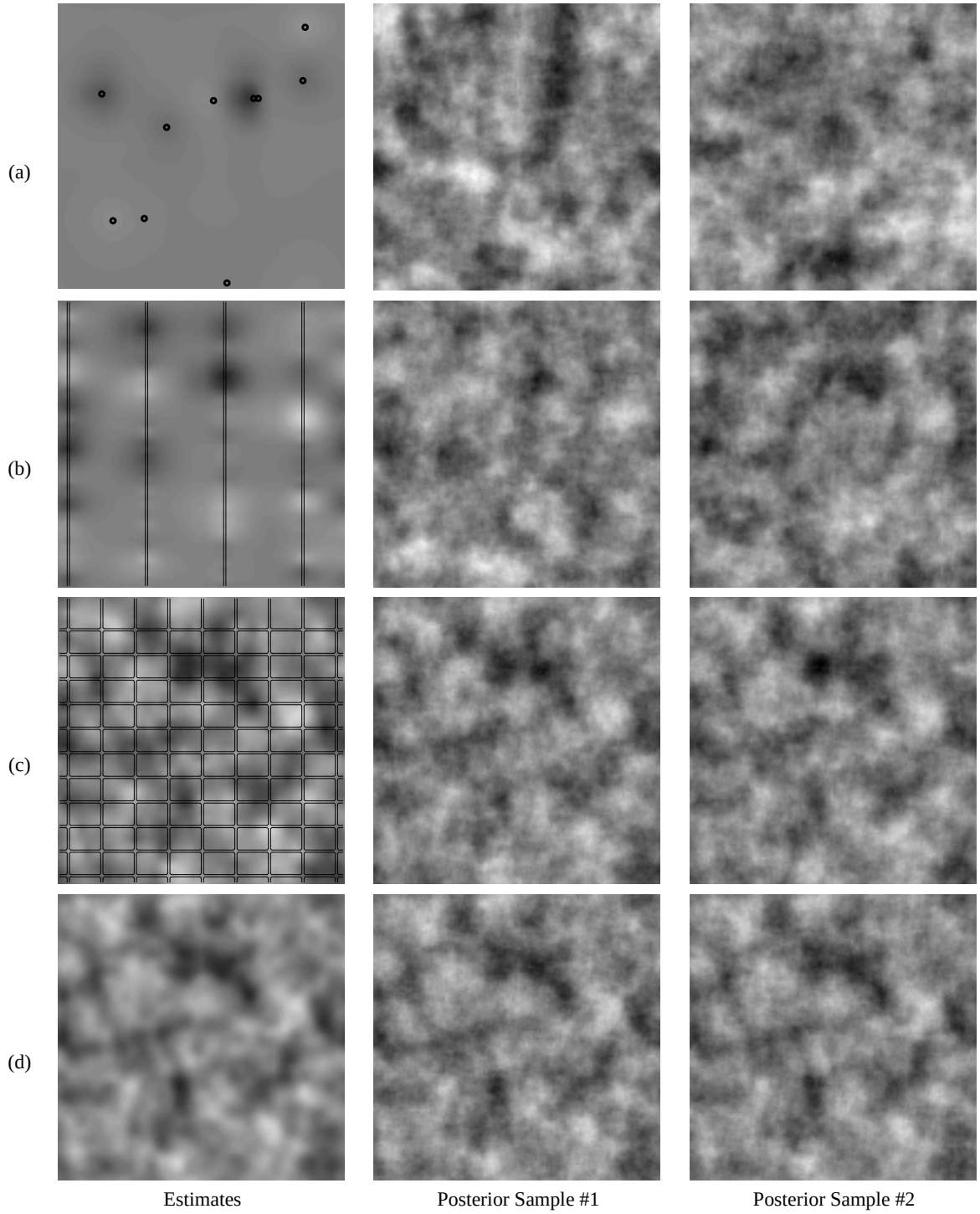
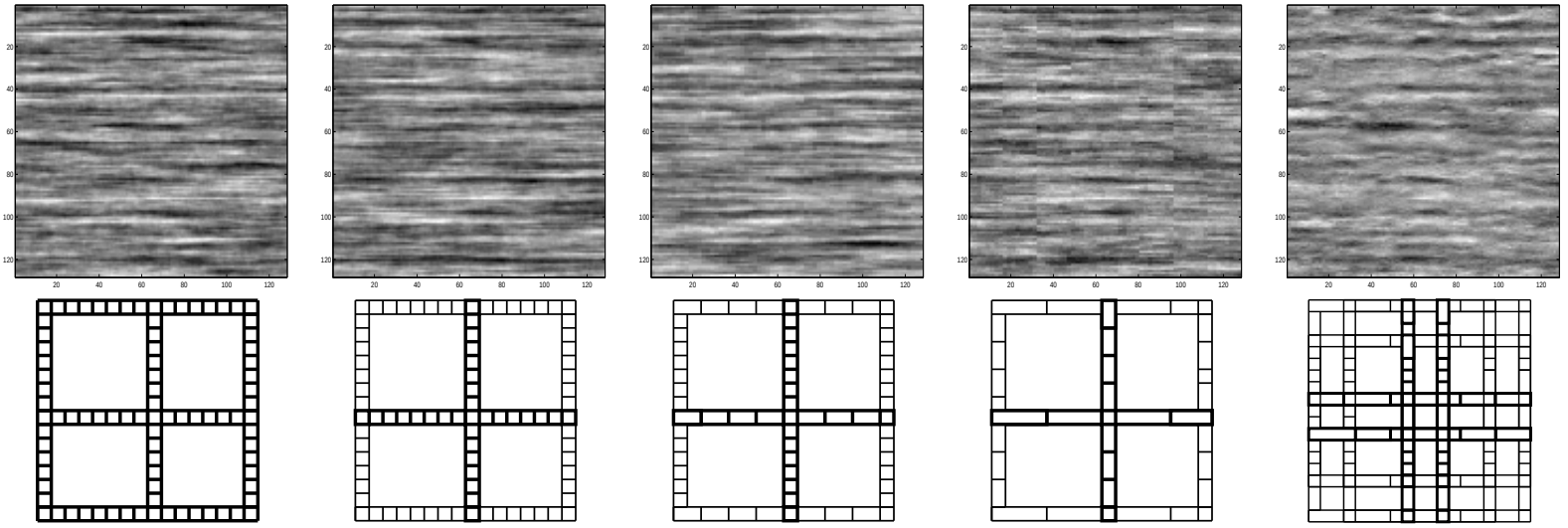


Fig. 5. Sampling for various measurement contexts: (a) sparse pointwise, (b) sparse columns, (c) dense rows/columns, and (d) densely at lower resolution (one measurement of each 4×4 pixel block average). Two posterior samples are shown in each case; their similarity depends on the degree to which the estimation errors are constrained by the available measurements, which increase in density from top to bottom.



a) Dense State 303 sec. b) Aperiodic 74 sec. c) Sparse Horz. 21 sec. d) Very Sparse 10 sec. e) Sparse, Overlapped 20 sec.

Fig. 6. An illustration of computational complexity and tradeoffs. The examples differ in the nature of the multiscale state of Figure 1, corresponding to (a) dense pixel-wise sampling, (b) a reduction in the coarse states to relax boundary periodicity, (c) a reduction in horizontal sampling density, (d) a further reduction in density (so that artifacts appear), and (e) a reduced state on an overlapped tree[14]. Bold state elements appear at all scales, unbolded ones only at finer scales. The running times, in seconds, are for the entire process — multiscale model computation, estimation, estimation error variances, and posterior sampling. Direct posterior sampling using Cholesky decomposition would require several hours.

[12] S. Geman, D. Geman, "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images", *IEEE Trans. PAMI* (6) #6, pp. 721–741, 1984

[13] K. Hanson, G. Cunningham, "Posterior Sampling with Improved Efficiency," *Proceedings of the SPIE* (3338), pp.371-382, 1998

[14] W. Irving, P. Fieguth, A. Willsky, "An Overlapping Tree Approach to Multiscale Stochastic Modeling and Estimation", *IEEE Trans. on Image Processing* (6) #11, pp.1517–1529, 1997

[15] W. Irving, A. Willsky, "A Canonical Correlations Approach to Multiscale Stochastic Realization," IEEE Transactions on Automatic Control (46) #10, pp.1514-1528, 2001

[16] M. Luettgen, A. Willsky, "Multiscale Smoothing Error Models", *IEEE Trans. on Automatic Control* (40) #1, 1995

[17] M. Luettgen, W. Karl, A. Willsky, R. Tenney, "Multiscale Representations of Markov Random Fields", *IEEE Trans. Signal Processing* (41) #12, pp.3377–3396, 1993

[18] N. Nguyen, P. Milanfar, G. Golub, "A Computationally Efficient Superresolution Image Reconstruction Algorithm," *IEEE Trans. on Image Processing* (10) #4, pp.573-583, 2001

[19] H. Rauch, F. Tung, C. Striebel, "Maximum Likelihood Estimates of Linear Dynamic Systems",*AIAA Journal*, (3) #8, 1965

[20] J. Starck, F. Murtagh, A. Bijaoui *Image Processing and Data Analysis, The Multiscale Approach*, Cambridge University Press, 1998

[21] M. Tanner, W. Wong, "The Calculation of Posterior Distributions by Data Augmentation," *J. American Statistical Association* (82), pp.548–550, 1987

[22] G. Winkler, *Image analysis, random fields, and dynamic Monte Carlo methods : a mathematical introduction*, Springer-Verlag, 1995.