Multiscale Modeling and Estimation of Poisson Processes with Application to Photon-limited Imaging

Klaus E. Timmermann, Student Member, IEEE*, and Robert D. Nowak,* Member, IEEE*,*

Abstract

Many important problems in engineering and science are well-modeled by Poisson processes. In many applications it is of great interest to accurately estimate the intensities underlying observed Poisson data. In particular, this work is motivated by photon-limited imaging problems. This paper studies a new Bayesian approach to Poisson intensity estimation based on the Haar wavelet transform. It is shown that the Haar transform provides a very natural and powerful framework for this problem. Using this framework, a novel multiscale Bayesian prior to model intensity functions is devised. The new prior leads to a simple, Bayesian intensity estimation procedure. Furthermore, we characterize the correlation behavior of the new prior and show that it has $1/f$ spectral characteristics. The new framework is applied to photon-limited image estimation and its potential to improve nuclear medicine imaging is examined.

Keywords: Poisson processes, multiscale analysis, wavelets, Bayesian inference, photon-limited imaging.

Department of Electrical Engineering, Michigan State University, East Lansing, MI 48824–1226. Fax: (517) 353–1980, Emails: timmerm4@egr.msu.edu, nowak@egr.msu.edu. Web: http://www.egr.msu.edu/spc/

This work is supported by the National Science Foundation, grant no. MIP–9701692.

A shorter version of this paper was presented at the 31st Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, Nov. 3, 1997.

I. Introduction

A great number of important phenomena in science and engineering are modeled as Poisson processes. In many instances, it is of interest to estimate the underlying intensity which gives rise to these phenomena. The intensity estimation problem is encountered in many fields including medicine [1], astronomy [2], communications [3], and networks [4]. This paper considers the problem of estimating the intensity of a Poisson process from a single observation of the process. We observe counts

$$\mathbf{c} \sim \text{Poisson}(\boldsymbol{\lambda}), \tag{1}$$

where the intensity $\boldsymbol{\lambda}$ may be a 1-d signal, 2-d image, or 3-d volume.

For example, the basic photon-limited imaging problem is described as follows. We observe photon emissions in a compact region of the plane. The photon emissions are the result of an underlying two-dimensional continuous intensity function. We are interested in estimating the intensity function from the counts of photon detections. Nuclear medicine imaging is one application that motivates our study of the photon-limited imaging problem. The major limitation of nuclear medicine imaging is the low-count levels acquired in typical studies, due in part to the limited level of radioactive dosage required to insure patient safety. Because of the variability of low-count images, it is very common to employ a post-filtering or estimation procedure to obtain a "better" estimate of the underlying intensity [5].

To simplify the presentation we work with 1-d intensity functions, but all the results are easily extended to multidimensional problems. Furthermore, we assume that the intensity function is discretized so that $\boldsymbol{\lambda}$ is represented as vector of length N with elements $(\lambda_k)_{k=0}^{N-1}$. The counts $c_k \sim \text{Poisson}(\lambda_k)$ are the elements of the vector $\mathbf{c}$, also of length N. The contribution of this paper is a novel *multiscale, Bayesian* approach to Poisson intensity estimation.

There are several reasons for adopting a multiscale estimation framework.

- Many signals, including intensity functions, have sparse multiscale representations, *i.e.,* a few large coefficients dominate the representation. Consequently useful Bayesian priors are easily specified in the multiscale analysis domain.
- The Poisson distribution is self-reproducing across scale (sum of independent Poisson variates is Poisson). This implies that the data have Poisson statistics at all resolutions.
- Coarse-scale estimators of intensities are very reliable (high signal-to-noise ratio). Reliable coarse-scale information can be leveraged to improve fine-scale estimators.

We discuss some of these motivations in more detail in sections II and III.

Many earlier approaches to Poisson intensity estimation were based on the idea of modeling the variability of the process by Gaussian fluctuations with non-stationary characteristics, *e.g.,* [6, 7]. Recently, simple wavelet-based approaches to this problem make use of the square-root of the counts (a variance stabilizing transformation that makes the data approximately Gaussian) and then apply standard wavelet thresholding techniques for Gaussian noise removal [8]. More sophisticated wavelet-based estimation procedures attempt to deal with the Poisson statistics directly. Kolaczyk has developed a wavelet-based thresholding scheme for the estimation of a special class of Poisson processes termed "burst-like" processes [9]. The burst-like Poisson process is characterized by a homogeneous, low intensity background with spatially isolated bursts of high intensity, and is motivated by problems in astronomical imaging. Nowak and Baraniuk propose a wavelet-based method for the estimation of more general Poisson intensities in [10] using the cross-validation estimator developed in [11]. This method is applied to nuclear medicine image estimation in [12]. Both methods [9, 10] can provide satisfactory results in certain situations. However, neither method adopts a Bayesian perspective, and hence they do not explicitly make use of prior information that may be available. We note, however, that several wavelet-based Bayesian estimation procedures have been proposed for Gaussian data, *e.g.,* [13, 14, 15, 16, 17], but such methods are not applicable to the Poisson problem considered here.

In this paper, we present and analyze a new, multiscale, Bayesian framework for the modeling and estimation of Poisson processes that was first proposed by the authors in the conference paper [18]. The new framework provides a very natural and powerful approach to studying a wide variety of Poisson processes, and we show how it can be applied in photon-limited imaging. The framework makes full use of the Poisson probability model and enables the incorporation of realistic prior information into the estimation process. There are four major contributions in the paper. First, we describe a new, multiscale, prior probability model for non-negative intensity functions. This model employs a multiplicative innovations structure in the scale-space domain. Second, based on this new prior we derive a simple and computationally efficient, Bayesian estimator of the intensity given an observation of counts, under squared error loss. It is shown through examples that the Bayesian estimation procedure significantly outperforms existing wavelet-based methods. Third, we extend the multiscale, intensity prior to a shift-invariant one, and develop a shift-invariant estimation procedure. Furthermore, we obtain closed-form expressions for the correlation functions of both priors, and show that the correlation behavior of the shift-invariant prior has $1/f$ spectral characteristics and is more regular than that of the shift-variant prior. Fourth, we apply the framework to photon-limited imaging and examine its potential to improve nuclear medicine imaging.

The paper is organized as follows. Section II sets some notation and briefly reviews the (unnormalized) Haar wavelet representation of signals. Section III introduces the multiscale multiplicative innovations (MMI) model as a new probability model for intensity functions. The MMI model is the cornerstone for the proposed estimator, which is developed in Section IV. The mathematical details of this development are left for the Appendices. Section V discusses a shift-invariant extension to the basic MMI model and compares the correlation behaviors of the two models. In Section VI, we compare the performance of the new estimation procedure to existing methods via simulated benchmark problems. In Section VII, we discuss applications of the new framework to photon-limited imaging. We finish with comments and conclusions in Section VIII.

II. The Haar Wavelet Transform

This paper uses the following wavelet representation and notation throughout. Let $\mathbf{c}_0$ be the data sequence of counts $\mathbf{c}$ of length $N = 2^J$, and let $c_{0,k}$ be its k^{th} element. The subscript 0 denotes the finest scale (resolution) of analysis. Similarly, let $\boldsymbol{\lambda}_0$ be the finest resolution representation of the intensity sequence $\boldsymbol{\lambda}$, *i.e.*, $\boldsymbol{\lambda}_0 = \boldsymbol{\lambda}$, also of length N. Then,

$$\mathbf{c}_0 \;\sim\; \text{Poisson}(\boldsymbol{\lambda}_0)\,, \tag{2}$$

and the objective is to estimate $\boldsymbol{\lambda}_0$ from the observation $\mathbf{c}_0$.

A multiscale analysis of $\mathbf{c}_0$ can be obtained by iterating

$$c_{j,k} \;=\; c_{j-1,2k} + c_{j-1,2k+1}, \tag{3}$$

$$d_{j,k} \;=\; c_{j-1,2k} - c_{j-1,2k+1}, \tag{4}$$

for $j = 1, \ldots, J$ and $k = 0, \ldots, N/2^j - 1$, and $J = \log_2(N)$. Here, J denotes the coarsest scale of analysis.[1] $c_{j,k}$ and $d_{j,k}$ are termed the scaling and wavelet coefficients of the data, respectively, at scale j and position (translation) k. These coefficients are simply the unnormalized Haar transform coefficients. The scaling coefficients $\mathbf{c}_j = (c_{j,k})_{k=0}^{N/2^j-1}$ represent a lower resolution representation of the data $\mathbf{c}_{j-1}$. The "detail" information in $\mathbf{c}_{j-1}$, which is absent in $\mathbf{c}_j$, is conveyed by the sequence of wavelet coefficients $\mathbf{d}_j = (d_{j,k})_{k=0}^{N/2^j-1}$. Note that $\mathbf{c}_{j-1}$ can be perfectly reconstructed from $\mathbf{c}_j$ and $\mathbf{d}_j$. Fig. 1 depicts the functional dependencies among the various scaling coefficients of a sequence $\mathbf{c}$ of length $N = 8$, using a tree-structured representation.

Similarly, as for $\mathbf{c}_0$, we define the scaling coefficients $\lambda_{j,k}$ and the wavelet coefficients $\theta_{j,k}$ of the intensity function $\boldsymbol{\lambda}_0$:

$$\lambda_{j,k} \;=\; \lambda_{j-1,2k} + \lambda_{j-1,2k+1}, \tag{5}$$

$$\theta_{j,k} \;=\; \lambda_{j-1,2k} - \lambda_{j-1,2k+1}. \tag{6}$$

For the case when the intensity of interest is of discrete nature, $\boldsymbol{\lambda}_0$ is simply the sequence $\boldsymbol{\lambda}$. If, on the contrary, the intensity signal is a function of a continuous variable $t \in [0,1]$, say $\lambda(t)$, then $\boldsymbol{\lambda}_0$ can be viewed as the sequence of integrated values of that function. That is, in accordance with the definition of the (unnormalized) Haar wavelet transform on the interval [19], $\lambda_{0,k} = \langle \lambda, \phi_{0,k} \rangle = \int_{k/N}^{(k+1)/N} \lambda(t)\, dt$, and more

[1] In this paper, scales corresponding to higher values of the index j represent coarser resolution levels. This convention is advantageous in keeping notation simple, however, it is in contrast to the convention followed in some literature.

generally, $\lambda_{j,k} = \langle \lambda, 2^{j/2}\phi_{j,k}\rangle = \int_{2^j k/N}^{2^j (k+1)/N} \lambda(t)\, dt$. Here, $\langle \cdot, \cdot \rangle$ denotes the inner product, and $\phi_{j,k}$, the Haar scaling function at scale j and shift k. Similarly, the wavelet coefficients have also a congruent meaning with (6) in the continuous variable case: $\theta_{j,k} = \langle \lambda, 2^{j/2}\psi_{j,k}\rangle = \lambda_{j-1,2k} - \lambda_{j-1,2k+1}$, where $\psi_{j,k}$ is the Haar wavelet at scale j and shift k.

A. Why the Unnormalized Haar Transform?

Multiscale analysis based on the unnormalized version of the Haar transform has the unique property that every scaling coefficient is the sum of two finer-scale scaling coefficients, and consequently, due to the reproducing property of the Poisson distribution,[2] every scaling coefficient is Poisson distributed. Furthermore, it is well known that given two Poisson variates, $C_1 \sim \text{Poisson}(\lambda_1)$ and $C_2 \sim \text{Poisson}(\lambda_2)$, the conditional distribution of C_1 given the sum $C_1 + C_2$ is binomial [20]. This reveals a very simple "parent-child" relationship between the scaling coefficients across scales. In Section IV, these facts are crucial in the development of the proposed intensity estimator. Similar attributes (reproducibility and simple parent-child relationship) do not hold for more general multiscale analyses of Poisson processes, based on other wavelet systems for example. Hence, the unnormalized Haar transform can be viewed as the *canonical* multiscale analysis tool for Poisson processes. This is in marked contrast to the Gaussian case, in which such attributes hold for a wide variety of wavelet analyses (including all orthogonal wavelet systems). In short, multiscale transforms of Poisson processes other than the unnormalized Haar transform are much more difficult to analyze and process. The natural match between the unnormalized Haar transform and the Poisson process is our primary motivation for choosing it.

The use of the unnormalized Haar wavelet transform to carry out the multiscale analysis of the data has additional (secondary) benefits springing from the following points. Poisson processes result from counting independent events occurring in disjoint regions of time or space of equal size. In such cases, the unnormalized Haar scaling coefficients correspond exactly to these counts occurring at intervals of sizes varying according to scale. Thus, the scaling and wavelet coefficient have a very natural interpretation according to (3) and (4). The Haar basis functions also have the property of being completely localized in space. By this we mean that at each scale, scaling functions, as well as wavelets, do not overlap. Therefore, at each scale, scaling coefficients are conditionally independent, that is,

$$\mathrm{P}(\mathbf{c}_j|\boldsymbol{\lambda}_j) = \prod_k \mathrm{P}(c_{j,k}|\lambda_{j,k}).$$

Also, Poisson processes are inherently nonnegative; therefore, a good estimator should always produce intensity estimates that are either positive or zero. An estimator based on the Haar system may be designed with this quality.

[2] $C_i \sim \text{Poisson}(\lambda_i),\ C_i$ independent $\Rightarrow \sum C_i \sim \text{Poisson}(\sum \lambda_i)$.

III. A New Probability Model for Intensity Functions

A. Multiscale Signal Model Framework

To formulate a Bayesian estimator for this problem, we must first propose a prior probability model for the unknown intensity. The observed data $\mathbf{c}$ is the realization of a random sequence $\boldsymbol{C}$ ($\sim$ Poisson($\boldsymbol{\lambda}$)), and $\boldsymbol{\lambda}$ is regarded as an unknown realization of a random sequence $\boldsymbol{\Lambda}$ with prior density $f_{\boldsymbol{\Lambda}}$. Given this prior, we seek the optimal estimate of $\boldsymbol{\lambda}$ with respect to squared error loss. The optimal estimate is the posterior mean

$$
\begin{aligned}
\widehat{\boldsymbol{\lambda}} &= \mathrm{E}\left[\boldsymbol{\Lambda} \,|\, \boldsymbol{C} = \mathbf{c}\right] \\
&= \int \boldsymbol{\lambda}\, f(\boldsymbol{\lambda}|\mathbf{c})\, d\boldsymbol{\lambda},
\end{aligned} \tag{7}
$$

where $f(\boldsymbol{\lambda}|\mathbf{c})$ is the posterior density function $f_{\boldsymbol{\Lambda}|\mathbf{C}}(\boldsymbol{\lambda}|\mathbf{c})$, and $\mathrm{E}[\cdot]$ denotes the expectation operator. We often follow this simplifying convention of defining pdf's and probability mass functions solely by their arguments.

A Bayesian approach facilitates the solution of (7) by expressing it in terms of the prior density $f_{\boldsymbol{\Lambda}}$. Applying Bayes' theorem to (7)

$$
\widehat{\boldsymbol{\lambda}} = \frac{\int \boldsymbol{\lambda}\, \mathrm{P}(\mathbf{c}|\boldsymbol{\lambda}) f(\boldsymbol{\lambda})\, d\boldsymbol{\lambda}}{\int \mathrm{P}(\mathbf{c}|\boldsymbol{\lambda}) f(\boldsymbol{\lambda})\, d\boldsymbol{\lambda}}, \tag{8}
$$

where $\mathrm{P}(\mathbf{c}|\boldsymbol{\lambda})$ is the likelihood that $\boldsymbol{C} = \mathbf{c}$ given that $\boldsymbol{\Lambda} = \boldsymbol{\lambda}$. The Bayes' estimator (8) poses two interrelated problems. First, the specification of a meaningful and useful prior $f_{\boldsymbol{\Lambda}}$. Second, the numerical computation of the estimator. The role the above quantities play in the estimation process is illustrated in Fig. 2. The caret symbol over a variate denotes its estimate. The remainder of this section describes a new prior probability model for the Haar scaling and wavelet coefficients of a non-negative intensity that leads to a very simple specification of $f_{\boldsymbol{\Lambda}}$. In Section IV, we derive an efficient algorithm for computing the optimal estimator (8).

There are two important reasons for adopting a multiscale approach to this problem:

- Prior models that are mathematically tractable, computationally practical, and empirically supported can be specified very naturally.
- Poisson data is much more reliable at coarse scales than at fine resolutions (higher counts $\Rightarrow$ higher signal-to-noise ratio). Therefore, more reliable coarse-scale estimates can be leveraged to improve high resolution estimators.

The first point is due partly to the fact that multiscale decompositions of real-world intensities are often statistically *self-similar*. By this we mean the property that the various scale representations preserve the major features and characteristics of the original object, except for the usual gradual loss of resolution. In particular, it has been widely recognized that the distribution of the wavelet coefficients of real-world signals tend to be similar at all scales of analysis, and are usually concentrated around the origin and unimodal [14].

The self-similarity captured by the Haar multiscale analysis is illustrated by considering the distributions of the wavelet coefficients at various scales. Fig. 3 illustrates this phenomena. The histograms in this figure correspond to the wavelet coefficients[3] at scales 1, 2, 3, and 4 for the cameraman image of Fig. 8(a). The similarity between these distributions facilitates the specification of Bayesian prior for the intensity in a very natural way.

The second point above motivates an estimation process that evolves from coarse to fine scales. It is easily verified that the signal-to-noise ratio (SNR) in a Poisson process increases linearly with the underlying intensity (signal). Thus, according to (3), $c_{J,0} = \sum_{k=0}^{N-1} c_{0,k}$, and so the signal-to-noise ratio at scale J is 2^J times as large as that for the average data point $c_{0,k}$. For example, for a 128 by 128 pixel image, this represents a SNR improvement of 42 dB.

B. Multiscale Multiplicative Innovations Model

We are now in a position to describe a new, Haar-based probability model for the intensity. Let $\Lambda_{j,k}$ and $\Theta_{j,k}$ denote the random variables corresponding to the j,k-th scaling and wavelet coefficient of the intensity, respectively. At the coarsest scale $j = J$, the single scaling coefficient $\Lambda_{J,0}$ has a density with support on $\mathbb{R}^+$. In this work, we choose the gamma density, since it is especially easy to use in conjunction with the Poisson mass function, and because it provides a reasonable mechanism for incorporating prior knowledge of the intensity range. However, as noted earlier, the SNR in the count $c_{J,0}$ is typically very high, and therefore any reasonable prior with support on $\mathbb{R}^+$ will not significantly influence our estimation of $\lambda_{J,0}$.[4]

Next, introduce statistically independent perturbation variables $\{\Delta_{j,k}\}$ and model the wavelet coefficients by

$$\Theta_{j,k} \;=\; \Lambda_{j,k}\,\Delta_{j,k}. \tag{9}$$

Each wavelet coefficient is modeled as independent perturbation of its corresponding scaling coefficient. Furthermore, the perturbations at all scales and positions are assumed to be mutually independent. Applying recursions (5) and (6) to these coefficients, we find that $\Lambda_{j-1,k} = \frac{1}{2}(\Lambda_{j,[k/2]} + (-1)^k\,\Theta_{j,[k/2]})$, where $[\cdot]$ stands for the largest integer no greater than the argument.

To gain some insight into this model, consider the random variable $Y_{j,k}$ defined by

$$Y_{j,k} \;=\; \frac{1}{2}\,(1 + \Delta_{j,k})\,. \tag{10}$$

[3]More precisely, here we are plotting the histogram of the ratio of the wavelet coefficient relative to the corresponding scaling coefficient. That is, each wavelet coefficient is divided by the corresponding scaling coefficient at the same scale and position. If the scaling coefficient is zero, the operation maps to zero. The motivation for this ratio will become apparent in the following sections.

[4]In fact, in practice we often use the estimate $\widehat{\lambda}_{J,0} = c_{J,0}$.

The variable $Y_{j,k}$ can be viewed as the *canonical* multiscale parameter for Poisson processes because of the following parent-child relationship. It is well known that given two Poisson variates, $C_1 \sim \text{Poisson}(\lambda_1)$ and $C_2 \sim \text{Poisson}(\lambda_2)$, the conditional distribution of C_1 given the sum $C_1 + C_2 = c$ is binomial with parameter $y = \frac{\lambda_1}{\lambda_1+\lambda_2}$ [20]. In the context of our multiscale analysis, this special property implies a very simple parent-child relationship. Specifically, the conditional distribution of child $C_{j-1,2k}$ given the parent $C_{j,k} = C_{j-1,2k} + C_{j-1,2k+1} = c$ is binomial with parameter $Y_{j,k}$. This relationship demonstrates the fundamental role of $Y_{j,k}$ in the multiscale analysis of Poisson processes.

Using $Y_{j,k}$ in conjunction with (9) we have

$$\Lambda_{j-1,2k} = \Lambda_{j,k}\, Y_{j,k}, \tag{11}$$

$$\Lambda_{j-1,2k+1} = \Lambda_{j,k}\, (1 - Y_{j,k}). \tag{12}$$

We can interpret these refinements as a multiscale innovations structure, with the innovations $Y_{j,k}$ and $(1-Y_{j,k})$ entering in a multiplicative fashion, in contrast to the more standard additive innovations structure encountered in Gaussian estimation problems [21]. We call this model a *multiscale multiplicative innovations* (MMI) model. The model is graphically depicted in Fig. 4. The following are some key properties of the MMI model.

So long as the distributions for the perturbations $\{\Delta_{j,k}\}$ are chosen to be similar across scales, the MMI model gives rise to self-similar intensity representations, which as discussed in section III-A, are typical of real-world intensities. Also, here we only consider temporally homogeneous processes, it would be undesirable if the prior depended on the observation time interval in a complicated manner. Due to the multiplicative innovations structure, only the coarsest scale of the prior is dependent on the observation time interval. Thus, the model is essentially invariant to the length of the observation time. Moreover, in Section IV the MMI model is shown to provide a mathematically tractable match to the Poisson nature of the data which leads to a very simple estimator formulation.

The MMI model is closely related to other models studied in physics and statistics. The MMI model belongs to the class of *cascade models*, which are used in statistical physics for modeling a variety of natural phenomena including turbulence modeling [22] and rainfall distributions [23]. In fact, because MMI model is a type of cascade model, it can be shown that the MMI model is a random multifractal [24]. It is also interesting to note that the MMI model is a special case of a *Polya tree* [25, 26]. In the statistics community, Polya trees are used to model probability distributions, a problem analogous to modeling a non-negative intensity function.

C. Prior Distribution for Innovations

In the MMI model, the prior density f_Δ for $\Delta_{j,k}$ can be chosen as identical for all j, k, or may be defined to be distributed differently at each scale for added flexibility. Location dependence can also be introduced, but we have not pursued this at present. Desired properties for f_Δ include support on the $[-1, 1]$ interval, symmetry about the origin, unimodality, and concentration around zero. The first property is due to the fact that the range of $\Theta_{j,k}$ is $[-\Lambda_{j,k}, \Lambda_{j,k}]$. The second arises from the assumption that there is no reason *a priori* to favor $\Lambda_{j-1,2k}$ over $\Lambda_{j-1,2k+1}$, or vice-versa. The third and fourth properties are based on the characteristics of observed wavelet coefficient distributions resulting from natural signals [14], and which have been exploited in other areas including wavelet-based compression [27, 28]. These properties are also illustrated in the histograms of Fig. 3.

One very general class of probability density functions that possesses the desired characteristics, and which we use throughout the remainder of the paper, are beta-mixture densities of the form

$$f(\delta) = \sum_{i=1}^{M} p_i \frac{(1-\delta^2)^{s_i-1}}{B(s_i, s_i)\, 2^{2s_i-1}}, \tag{13}$$

for $-1 \leq \delta \leq 1$, where B is the Euler beta function, $0 \leq p_i \leq 1$ is the weight of the i-th beta density $\frac{(1-\delta^2)^{s_i-1}}{B(s_i,s_i)\, 2^{2s_i-1}}$ with parameter $s_i \geq 1$, and $\sum_{i=1}^{M} p_i = 1$. Fig. 5 depicts a mixture of three beta densities. A similar method was also recently proposed using a prior based on a mixture of a Dirac impulse and a single beta distribution [29].

Other classes of density functions may also provide the desired characteristics, but the beta family has a significant computational advantage. As pointed out above, $Y_{j,k}$ parameterizes the conditional distribution of the child $C_{j-1,2k}$ given the parent $C_{j,k}$. This conditional distribution is binomial. Hence, from a practical perspective, the use of a prior that is conjugate to the binomial will greatly facilitate computations [30]. It is well known that the beta family is conjugate to the binomial. For this reason, the beta mixture prior described above leads to a very simple, closed-form estimator which is discussed in the next section. However, for the sake of brevity, in our derivation of the optimal estimator, given in the Appendix, we directly compute the posterior means based on a beta mixture prior, without explicitly noting the use of conjugacy.

IV. Estimation

A. Bayesian Multiscale Intensity Estimator

In this paper, we focus on the posterior mean estimator, although other estimators (*e.g.*, MAP) may also be considered within our framework. The posterior mean is the optimal Bayes estimate under a quadratic loss function. The posterior mean estimate of an intensity $\lambda_{j,k}$, given all the information available in the data $\mathbf{c}_0$ and the MMI prior model $f_{\mathbf{\Lambda}}$, is the conditional mean $\widehat{\lambda}_{j,k} = \mathrm{E}[\Lambda_{j,k}|\mathbf{c}_0]$. In this section we derive simple closed-form expressions for the posterior mean.

First, based on the analysis in Appendix A,

$$\widehat{\lambda}_{j,k} = \mathrm{E}[\Lambda_{j,k}|\mathbf{c}_0] = \mathrm{E}[\Lambda_{j,k}|\mathbf{c}_j].$$

This implies that a simple coarse-to-fine procedure can be employed in the estimation process.

At the coarsest scale $j = J$, the intensity is represented by a single scaling coefficient $\lambda_{J,0}$. Let us begin by considering the estimation of $\lambda_{J,0}$. We have $\widehat{\lambda}_{J,0} = E\left[\Lambda_{J,0}|\mathbf{c}_0\right] = E\left[\Lambda_{J,0}|c_{J,0}\right]$. As argued in Section III, the corresponding count $c_{J,0}$ is itself usually a very good estimate for $\lambda_{J,0}$ provided the total number of counts is sufficient large. This choice has the added advantage of insuring the preservation of total number of counts, *i.e.*, $\sum_k \widehat{\lambda}_{0,k} = \sum_k c_{0,k}$.

The posterior mean estimate for the wavelet coefficient $\theta_{j,k}$ is given by

$$\begin{aligned} \widehat{\theta}_{j,k} &= \mathrm{E}\left[\Theta_{j,k}|\mathbf{c}_0\right] \\ &= \mathrm{E}\left[\Lambda_{j,k}|\mathbf{c}_0\right] \mathrm{E}\left[\Delta_{j,k}|\mathbf{c}_0\right], \end{aligned}$$

where we have used (9), and have exploited the independence between $\Lambda_{j,k}$ and $\Delta_{j,k}$. Now, we may simply write

$$\widehat{\theta}_{j,k} = \widehat{\lambda}_{j,k}\, \widehat{\delta}_{j,k}, \tag{14}$$

with the obvious definition $\widehat{\delta}_{j,k} = \mathrm{E}\left[\Delta_{j,k}|\mathbf{c}_0\right]$.

The desired form for $\widehat{\lambda}_{j,k}$ may be obtained using (5) and (6), and the linearity property of the expectation operator:

$$\begin{aligned} \widehat{\lambda}_{j,k} &= \mathrm{E}\left[\frac{1}{2}\left(\Lambda_{j+1,[k/2]} + (-1)^k\, \Theta_{j+1,[k/2]}\right) \middle| \mathbf{c}_0\right] \\ &= \frac{1}{2}\left(\widehat{\lambda}_{j+1,[k/2]} + (-1)^k\, \widehat{\theta}_{j+1,[k/2]}\right). \end{aligned} \tag{15}$$

Here is where we exploit the multiscale framework for estimating the desired intensity. The estimate for $\widehat{\lambda}_{j,k}$ is leveraged by the more robust coarser-scale scaling coefficient's estimate $\widehat{\lambda}_{j+1,[k/2]}$.

In Appendix B we show that

$$\widehat{\delta}_{j,k} = d_{j,k}\, \frac{\sum_i p_i \frac{B(s_i+c_{j-1,2k}\,,\, s_i+c_{j-1,2k+1})}{B(s_i,s_i)\;(2s_i+c_{j,k})}}{\sum_i p_i \frac{B(s_i+c_{j-1,2k}\,,\, s_i+c_{j-1,2k+1})}{B(s_i,s_i)}}. \tag{16}$$

The parameters $\{p_i, s_i\}$ are the defining parameters for the beta mixture model in (13). Note that $\widehat{\delta}_{j,k}$ guarantees nonnegativity of the resulting intensity estimates. That is, $\widehat{\lambda}_{j,k} \geq 0$ for $j = 0, \ldots, J$ and all k. To verify this simply rewrite (16) with the factor $d_{j,k}$ inside the upper summand and consider the fact that $\left|\frac{d_{j,k}}{2s_i+c_{j,k}}\right| \leq 1$ for all i.

The overall algorithm is described below.

Bayesian Multiscale Intensity Estimation

1. Estimate coarsest scale coefficient

$$\widehat{\lambda}_{J,0} = c_{J,0}$$

2. For $j = J$ down to 1 and $k = 0$ to $N/2^j - 1$

Compute:

$\widehat{\delta}_{j,k}$ according to (16)

$$\widehat{\theta}_{j,k} = \widehat{\lambda}_{j,k}\,\widehat{\delta}_{j,k}$$

Refine:

$$\widehat{\lambda}_{j-1,2k} = \tfrac{1}{2}\left(\widehat{\lambda}_{j,k} + \widehat{\theta}_{j,k}\right)$$

$$\widehat{\lambda}_{j-1,2k+1} = \tfrac{1}{2}\left(\widehat{\lambda}_{j,k} - \widehat{\theta}_{j,k}\right)$$

These simple procedural steps produce posterior mean estimates for finer and finer representations of the underlying intensity, and terminate with the desired, finest-scale estimate $\widehat{\boldsymbol{\lambda}}_0$. The complexity of the proposed estimator is $O(N)$, the same order as the fast wavelet transform itself.

For large data sets it is possible that the full $J = \log_2(N)$ iterations over the scales in Step 2 above are not necessary. For instance, for long data sequences the estimator may be initiated at a scale $J^* < \log_2(N)$, for which the estimate $\widehat{\boldsymbol{\lambda}}_{J^*} = \mathbf{c}_{J^*}$ is already very reliable. In practice, even for low-count data, we have found that $J^* = 5$ provides excellent results. Using J^* other than J is equivalent to dividing the original data sequence into equal subsequences, estimating their underlying intensities separately, and concatenating the individual results to obtain the overall intensity estimate. However, in Section V we introduce a shift-invariant version of the estimator above, for which the equivalence just described does not hold.

B. Selection and Analysis of the Beta-Mixture Prior

Our experiments and analysis with real-world intensity functions have led to several conclusions.

1. The perturbation densities of many real-world intensities are very well characterized by a weighted combination of three beta densities with shape parameters $s_1 = 1$, $s_2 = 100$, and $s_3 = 10000$.

2. The $s_1 = 1$ component is the uniform density, and we have found that fixing the corresponding weight to a small positive constant (*e.g.*, $p_1 = 0.001$), to insure that the prior density f_Δ is bounded from below over the entire $[-1, 1]$ interval, is appropriate in most situations we have studied.

3. The key parameter that distinguishes the characteristics of different intensity functions is the trade-off between the $s_2 = 100$ component and the lower variance $s_3 = 10000$ component. The trade-off is parameterized by the probability $p_2 \in [0, 1 - p_1]$. (Note $p_3 = 1 - p_1 - p_2$.)

To gain some insight into the functioning of the MMI model estimator, consider Fig. 6 which plots $\widehat{\delta}_{j,k}$ versus the ratio $d_{j,k}/c_{j,k}$ for three cases: low ($c = 10$), medium ($c = 30$), and high counts ($c = 1000$). These are, respectively, the dashed, solid, and dot-dashed curves. The ratio $d_{j,k}/c_{j,k}$ may be regarded as an empirical counterpart to the perturbation variate $\delta_{j,k}$. Fig. 6(a) and (b) correspond to two δ-priors with $p_2 = 0.1$, and $p_2 = 0.9$, respectively. The rest of the parameters take the values given in points 1–3 above.

From these figures we can observe the following. At high counts, when the SNR is high, the estimator regards $d_{j,k}/c_{j,k}$ to be a good estimate of $\delta_{j,k}$ for almost every value of $d_{j,k}/c_{j,k}$; thus, the resemblance to the unit-slope linear function. In contrast, for low counts, the SNR is much lower, and consequently the estimator severely attenuates $d_{j,k}/c_{j,k}$. This phenomenon is reminiscent of the behavior of wavelet-domain threshold estimators designed for additive Gaussian white noise (AGWN) [8], except that the threshold is adaptive to the local intensity.

The MMI model estimator minimizes the expected squared error with respect to the MMI model prior. Of course, the error is minimized by balancing the trade-off between fidelity to the data and fitting to the prior model. If the data are very 'reliable,' then the estimator favors the data. If the data are unreliable, then the estimator favors the prior. As noted, at low counts the ratios $d_{j,k}/c_{j,k}$ are not accurate estimates for δ, and so, the Bayesian estimator attenuates them to minimize the expected squared error in accordance with the prior. The nonlinearities in Fig. 6(a) correspond to a lower-variance prior (smaller p_2) than that giving rise to the nonlinearities of Fig. 6(b). As a result, the nonlinearities in Fig. 6(a) display a 'dead-band' characteristic, since the prior requires that a greater number of $d_{j,k}/c_{j,k}$ samples be mapped towards zero, in contrast to the higher-variance prior which displays a less harsh attenuation of small $d_{j,k}/c_{j,k}$ in Fig. 6(b). This behavior is similar to that observed in other wavelet-based Bayesian estimators designed for AGWN [13, 15, 16].

However, the functioning of the MMI model estimator is in contrast to that of estimators for AGWN processes. In the latter cases, all wavelet coefficients are typically attenuated independently and in disregard to the values of the corresponding scaling coefficients [13, 15, 16]. The MMI model estimator, on the other hand, adapts not just to the statistics of a particular scale, but also naturally incorporates information from coarser scales. This is crucial in the Poisson problem since the coarse-scale intensities (scaling coefficients) are indicative of the statistical reliability of the wavelet coefficient.

C. Estimation of prior parameters

The Haar wavelet coefficient distributions of real-world intensities often fit a profile which resembles that of Fig. 5 as previously discussed. However, while many distributions follow this general characteristic, one expects that subtle variations will exist from application to application. Therefore, it is of interest to adapt the prior to the problem at hand. Here we give a very simple approach to fitting the prior based on a moment-matching method. This adaptation can be viewed as an empirical Bayesian extension of framework described above.

Recall, we assume that for each scale j, the set $\{\Delta_{j,k}\}$ is independent identically distributed (i.i.d.) with an M-component beta-mixture density distribution with parameters $\{p_{j,i}, s_i\}_{i=1}^{M}$. We let the mixing probabilities $\{p_{j,i}\}$ depend on scale to enable variations of the density across scale. Also note that here the index i does not refer to shift (as does k in $Y_{j,k}$ for example), but rather it refers to the i-th component of the mixture density. Then, by (10), at each scale j, $\{Y_{j,k}\}_{k=0}^{N/2^j-1}$ is also i.i.d. with an M-component standard-beta-mixture density distribution with parameters $\{p_{j,i}, s_i\}_{i=1}^{M}$:

$$f_j(y) = \sum_{i=1}^{M} p_{j,i} \frac{y^{s_i-1}(1-y)^{s_i-1}}{B(s_i, s_i)}, \quad 0 \leq y \leq 1. \tag{17}$$

Since $Y_{j,k}$, is independent of $\Lambda_{j,k}$, $\mathrm{E}[\Lambda_{j-1,2k}^n] = \mathrm{E}[\Lambda_{j,k}^n]\,\mathrm{E}[Y_{j,k}^n]$ and

$$\mathrm{E}\left[Y_{j,k}^n\right] = \frac{\mathrm{E}[\Lambda_{j-1,2k}^n]}{\mathrm{E}[\Lambda_{j,k}^n]}. \tag{18}$$

The moments $\mathrm{E}[Y_{j,k}]$ need not be computed using this expression since the prior model (17) gives the mean $\frac{1}{2}$ for any choice of parameters. For $n \geq 2$, the moments $\mathrm{E}[\Lambda_{j-1,2k}^n]$ and $\mathrm{E}[\Lambda_{j,k}^n]$ are easily estimated from the data. Since $\mathrm{E}[C_{j,k}|\Lambda_{j,k}] = \Lambda_{j,k}$,

$$\mathrm{E}[\Lambda_{j,k}] = \mathrm{E}[\mathrm{E}[C_{j,k}|\Lambda_{j,k}]] = \mathrm{E}[C_{j,k}].$$

And in general, it can be shown that for all n there exist a degree n polynomial $p_n(\cdot)$ such that

$$\mathrm{E}[\Lambda_{j,k}^n] = \mathrm{E}[p_n(C_{j,k})].$$

For example,

$$\mathrm{E}[\Lambda_{j,k}^2] = \mathrm{E}[C_{j,k}^2 - C_{j,k}] \approx \frac{1}{N/2^j} \sum_k (c_{j,k}^2 - c_{j,k}).$$

Substituting these estimates for various n into (18) we can obtain empirical estimates for the various moment of $Y_{j,k}$. Equating these to the moments of the beta-mixture model (17) produces a set of equations that can be solved for the parameters $\{p_{j,i}, s_i\}_{i=1}^{M}$.

As mentioned above, we have found that the choice of a three component prior beta-mixture model suffices for many real-world intensities. At all scales j, we suggest the shape parameters $s_1 = 1$, $s_2 = 100$, and $s_3 = 10000$, with weights $p_{j,1} = .001$, $p_{j,3} =$ 1-$p_{j,1}$-$p_{j,2}$, and $p_{j,2}$ adapted at each scale using the second moment estimate for $Y_{j,k}$:

$$\mathrm{E}[Y_{j,k}^2] = \int_0^1 y^2 f_j(y)\,dy = \sum_{i=1}^{3} p_{j,i} \frac{s_i+1}{4s_i+2} \approx \frac{\sum_m (c_{j-1,2m}^2 - c_{j-1,2m})}{\sum_m (c_{j,m}^2 - c_{j,m})}. \tag{19}$$

V. A Stationary Intensity Model and Estimator

One potential limitation of MMI model is that it is not stationary due to the fact that the Haar wavelet transform is shift-dependent. That is, the analysis and estimation depends on the alignment between the Haar basis functions and the data. Moreover, the coarse scale approximation by Haar wavelets is piece-wise constant, an unrealistic intensity model. To circumvent such problems, *shift-invariant wavelet transforms* have been proposed in literature [31, 32, 33, 34, 35, 36]. In this section, we provide a unified Bayesian framework for shift-invariant analysis and estimation based on the MMI model. We characterize the autocorrelation functions of the MMI model (non-stationary) and a shift-invariant MMI model (stationary), and show that the shift-invariant MMI model has a more regular correlation behavior which may be better suited for modeling real-world intensities.

A. Shift-Invariant MMI Models

Shift-invariant MMI intensity models can be easily constructed within our Bayesian framework. Specifically, the shift of the intensity function with respect to the Haar wavelet system can be viewed as an additional degree of freedom in the model, and a probability measure on the shift parameter can be introduced. It is assumed that all shifts are circular. If we regard the original model as "unshifted" (shift $= 0$), then the standard MMI model introduced in Section III is denoted $f(\boldsymbol{\lambda}|\text{shift} = 0)$. Now let $\mathrm{P}(\text{shift} = m)$ denote a probability mass function for the shift, and consider the averaged MMI model

$$f(\boldsymbol{\lambda}) = \sum_{m} f(\boldsymbol{\lambda}|\text{shift} = m)\mathrm{P}(\text{shift} = m). \tag{20}$$

If $\mathrm{P}(\text{shift} = m)$ is the uniform distribution, a non-informative prior expressing no preference for any particular shift, then the averaged MMI model provides a shift-invariant (stationary) intensity prior and we call it the shift-invariant (SI) MMI model. Moreover, it is shown in the next section that the SI-MMI model is more regular than the basic MMI model.

Estimation using the SI-MMI model is easily carried out by computing the optimal Bayes shift-dependent estimator given in Section IV for each shift, and then computing an average of the results. The complexity of the shift-invariant estimator is $O(N^2)$ operations. However, note that if we employ a J^*-scale SI-MMI model, with $J^* < \log_2(N)$, then the estimator is invariant to shifts modulo 2^{J^*}. This is due to the fact that the scaling functions at the coarsest scale have support over 2^{J^*} samples. In general the J^*-scale SI-MMI model only requires a uniform shift prior over a 2^{J^*} sample region of support, rather than over the entire range of circular shifts. Thus, if $J^* < \log_2(N)$, then the complexity of shift-invariant estimation is only $O(N2^{J^*})$. As pointed out earlier (see Section IV), in many applications it suffices to take J^* smaller than $\log_2(N)$. Also note that the fast shift-invariant methods described in [32, 33] are not applicable in this case due to the dependence between wavelet coefficients across scale. This dependence stems from the multiplicative relationship between scaling and wavelet coefficients.[5]

B. Autocorrelation Functions of MMI and SI-MMI Models

The underlying beta mixture densities capture the key heavy-tailed, non-Gaussian nature of wavelet coefficient distributions. However, the shift-dependent nature of the Haar wavelet transform generates a non-stationary correlation structure as illustrated next. For the sake of illustration, we focus on the 1-d case. Extensions to higher dimensions are straightforward. Also, to keep things simple, we assume that the maximum number of scales $J = \log_2(N)$ are computed in the analysis, where N is the length of the intensity vector. The results easily extend to other choices for $J^* \neq J$.

First, consider that basic MMI model (shift$=0$) and let us introduce the following notation. Let $\mu_J^2 = \mathrm{E}[\Lambda_{J,0}^2]$, the second moment of the scaling coefficient at the coarsest scale, and let $\rho_j^2 = \mathrm{E}\left[Y_{j,k}^2\right] = \mathrm{E}\left[(1-Y_{j,k})^2\right]$, the second moment of the innovations variates. Recall that we assume the distribution of $Y_{j,k}$ does not depend on position k. The variables $Y_{j,k}$ and $1 - Y_{j,k}$ have a common second moment due to the symmetry of the distribution of $Y_{j,k}$ about $\frac{1}{2}$.

For illustration consider the correlation between the intensities $\Lambda_{0,0}$ and $\Lambda_{0,2}$ in the MMI model depicted in Fig. 4. Note that the finest scale for which $\Lambda_{0,0}$ and $\Lambda_{0,2}$ have a common predecessor above in the tree is $j = 2$ and the predecessor is $\Lambda_{2,0}$. Therefore we can write

$$\begin{aligned}\Lambda_{0,0} &= Y_{1,0}\ Y_{2,0}\ \Lambda_{2,0},\\ \Lambda_{0,2} &= Y_{1,1}\ (1-Y_{2,0})\ \Lambda_{2,0}.\end{aligned}$$

The correlation between the two intensities is computed

$$\mathrm{E}\left[\Lambda_{0,0}\Lambda_{0,2}\right] = \mathrm{E}\left[Y_{1,0}\ Y_{1,1}\ Y_{2,0}\ (1-Y_{2,0})\ \Lambda_{2,0}^2\right].$$

Exploiting the independence of the innovations variates, we have

$$\mathrm{E}\left[\Lambda_{0,0}\Lambda_{0,2}\right] = \mathrm{E}\left[Y_{1,0}\right]\ \mathrm{E}\left[Y_{1,1}\right]\ \mathrm{E}\left[Y_{2,0}\ (1-Y_{2,0})\right]\ \mathrm{E}\left[\Lambda_{2,0}^2\right].$$

Examining the individual product terms and making use of the moments defined above,

$$\begin{aligned}\mathrm{E}\left[Y_{1,0}\right] = \mathrm{E}\left[Y_{1,1}\right] &= 2^{-1},\\ \mathrm{E}\left[Y_{2,0}\ (1-Y_{2,0})\right] &= 1/2 - \rho_2^2,\\ \mathrm{E}\left[\Lambda_{2,0}^2\right] &= \mu_3^2\rho_3^2.\end{aligned}$$

Hence,

$$\mathrm{E}\left[\Lambda_{0,0}\Lambda_{0,2}\right] = 2^{-2}\ (1/2-\rho_2^2)\ \mu_3^2\rho_3^2.$$

[5]The fast shift-invariant methods treat all wavelet coefficients independently.

Now consider a general case in which we are interested in the correlation between two intensities, say Λ_{0,k_1} and Λ_{0,k_2}. Let j^* be the finest scale for which Λ_{0,k_1} and Λ_{0,k_2} have a common predecessor (above) $\Lambda_{j^*,k}$ in the MMI model tree. The scale j^* can be explicitly calculated using the binary representations of k_1 and k_2, and depends on the exact positions of k_1 and k_2 with respect to the alignment of the Haar basis functions. Assuming that $k_1 < k_2$, we have

$$\Lambda_{0,k_1} = \prod_{i=1}^{j^*-1} Y_{i,k}\ Y_{j^*,k}\ \Lambda_{j^*,k}, \tag{21}$$

$$\Lambda_{0,k_2} = \prod_{i=1}^{j^*-1} Y'_{i,k}\ (1 - Y_{j^*,k})\ \Lambda_{j^*,k}. \tag{22}$$

In the expressions for Λ_{0,k_1} and Λ_{0,k_2} above, we use a generic spatial index k, since the distributions of independent innovations variates do not depend on position. We do, however, use $Y_{j,k}$ and $Y'_{j,k}$ to distinguish the independent innovations variates corresponding to Λ_{0,k_1} and Λ_{0,k_2}, respectively. The correlation is given by

$$\mathrm{E}\left[\Lambda_{0,k_1}\Lambda_{0,k_2}\right] = \mathrm{E}\left[\prod_{i=1}^{j^*-1} Y_{i,k} \prod_{i=1}^{j^*-1} Y'_{i,k}\ Y_{j^*,k}\ (1 - Y_{j^*,k})\ \Lambda^2_{j^*,k}\right]. \tag{23}$$

Again exploiting the independence of the innovation variates and making use of the moments defined above, we have $\mathrm{E}\left[\Lambda^2_{j^*,k}\right] = \mu_J^2 \prod_{i=j^*+1}^{J} \rho_i^2$ and

$$\begin{aligned} r(k_1, k_2) &\triangleq \mathrm{E}\left[\Lambda_{0,k_1}\Lambda_{0,k_2}\right] \\ &= 2^{-2(j^*-1)}\ (1/2 - \rho^2_{j^*})\ \mu_J^2 \prod_{i=j^*+1}^{J} \rho_i^2. \end{aligned} \tag{24}$$

Note that because j^* depends on the alignment between the intensity function and the Haar basis, $r(k_1, k_2)$ is not a function of the difference $k_1 - k_2$ alone. This shows that the intensity distribution represented by the MMI model is non-stationary. Two columns (fixed k_1) of the autocorrelation function of a 256 length MMI model intensity prior are shown in Fig. 7. Note also that the autocorrelation function is highly irregular (piecewise-constant), which is an undesirable model for real-world intensities.

Now consider the autocorrelation function of the SI-MMI model. Given a displacement of n between two intensities, say $\Lambda_{0,0}$ and $\Lambda_{0,n}$, we can compute the probability of the finest scale j^* of a common predecessor, with respect to the uniform distribution over the shift parameter, as follows. We need to count the number of shifts that give rise to each possible value of j^*, or more precisely the probability of the set of shifts that give rise to each possible j^*. For example, suppose $n = 1$ and consider the tree in Fig. 4. In this case, four shifts (out of eight) result in $j^* = 1$, another two shifts result in $j^* = 2$, and two shifts (one wrapping around due to circularity) result in $j^* = 3$. Hence, we compute $\mathrm{P}(j^* = 1|n = 1) = 1/2$, $\mathrm{P}(j^* = 2|n = 1) = 1/4$, and $\mathrm{P}(j^* = 3|n = 1) = 1/4$. Similarly, if $n = 2$, then $\mathrm{P}(j^* = 1|n = 2) = 0$, $\mathrm{P}(j^* = 2|n = 2) = 1/2$, and

$\mathrm{P}(j^* = 3|n = 2) = 1/2$, where again we must be careful to account for the wrap-around effect of the circular shifting.

Given a displacement of $n = k$ between two scaling coefficients in a J-scale MMI model, the probability of the scale j^* is determined by inspecting the associated binary tree. We need only consider $|k| \leq 2^{j-1}$ due to the periodicity of the MMI. First note that we have $\mathrm{P}(j^* \leq J \,|\, n = k) = 1$. Next, notice that $(2^m - k)_+$ is the number of shifts of a length-k sequence that fit within a length-2^m sequence, where $(x)_+ = \max(x, 0)$. In other words, $(2^m - k)_+$ is the total number of scaling coefficient pairs spaced k apart at the bottom of a m-level binary tree. Also note that the number of m-level subtrees, $m < J$, at the bottom of a larger J-level binary tree is precisely 2^{J-m}. Hence, for $m < J$

$$\begin{aligned} \mathrm{P}(j^* \leq m \,|\, n = k) &= (2^m - |k|)_+ \, 2^{J-m} 2^{-J} \\ &= (2^m - |k|)_+ \, 2^{-m}. \end{aligned} \tag{25}$$

It follows that

$$\begin{aligned} p(m|k) &:= \mathrm{P}(j^* = m|n = k) \\ &= \begin{cases} 0, & m < \lceil \log_2(|k| + 1) \rceil \\ 1 - 2^{-m}|k|, & m = \lceil \log_2(|k| + 1) \rceil \\ 2^{-m}|k|, & \lceil \log_2(|k| + 1) \rceil < m < J \\ |k| 2^{-J+1}, & m = J, \end{cases} \end{aligned}$$

where $\lceil \log_2(|k| + 1) \rceil$ denotes the smallest integer greater than or equal to $\log_2(|k| + 1)$.

With these probabilities defined, the autocorrelation function is given by

$$r(k) = \mathrm{E}\left[\Lambda_{0,l} \Lambda_{0,l+k}\right] = \sum_{m=1}^{J} \nu_m \, p(m|k), \tag{26}$$

where

$$\nu_m = 2^{-2(m-1)} \left(1/2 - \rho_m^2\right) \mu_J^2 \prod_{i=m+1}^{J} \rho_i^2 \tag{27}$$

is simply the autocorrelation between two intensities at the bottom of an MMI binary tree model for which $j^* = m$, as in (24). Note that the SI-MMI autocorrelation is stationary. The autocorrelation function for a 256 length SI-MMI intensity prior is shown in Fig. 7. Unlike the autocorrelation of the MMI model, the SI-MMI model's autocorrelation is piece-wise linear (compared to piece-wise constant) and hence is more regular and potentially better suited for the analysis of natural intensities. These results are similar in spirit to the the analysis in [32], where it is observed that the shift-invariant Haar wavelet transform is line-preserving.[6] Moreover, the results here suggest possible schemes for choosing the parameters of the

[6]These conclusions extend to the SI-MMI model as well.

SI-MMI model. For example, the decay of the autocorrelation function may be tailored by appropriate choices of the $\rho_i^2,\ i = 1, \ldots, J$. Larger values for the ρ_i^2 cause $r(k)$ to decay faster. We also note that similar correlation analysis may be carried out for related wavelet-domain signal models developed for Gaussian data [13, 14, 15, 16, 17]. We also note that the issue of combining or mixing trees (which is essentially what is being done in the SI-MMI model) has been studied in the more general setting of the Polya tree. Some very interesting theoretical results concerning the continuity of densities generated by mixtures of Polya trees (SI-MMI models are a special case) are given in [26]. These results may provide further insights into the SI-MMI model and may suggest other extensions of our framework, but these issues are beyond the scope of this paper.

C. SI-MMI Model and $1/f$ Processes

Remarkably, the SI-MMI correlation function has a fractal $1/f$-like character. Fractal $1/f$ random process models are commonly used in image modeling, and it has been observed that natural signals and images often display a correlation structure similar to that of the MMI model [37, 38, 39]. To see that the SI-MMI correlation is $1/f$, consider the simple case in which the second moments of the innovations are constant independent of scale, *i.e.*, $\rho_j = \rho,\ j = 1, \ldots, J$. Then

$$\begin{aligned} \nu_m &= 2^{-2(m-1)}\ (1/2 - \rho^2)\ \mu_J^2 \rho^{2(J-m)} && (28) \\ &= C\,(2\rho)^{-2m}, && (29) \end{aligned}$$

where C is a constant independent of m. Next equate $(2\rho)^{-2m}$ with $2^{(\gamma-1)m}$ and solve for γ. This gives

$$\nu_m = C\, 2^{(\gamma-1)m}, \tag{30}$$

where $\gamma = -1 - \log_2 \rho^2$. Note that $1/4 < \rho^2 < 1/2$, where the lower and upper bounds corresponds to the two extreme limits of the beta density: a point mass at 1/2 or two point masses at 0 and 1, respectively. This implies $0 < \gamma < 1$. Combining (30) with (26) and after some algebra, we have

$$\begin{aligned} r(k) &= C(2^{(\gamma-1)\lceil \log_2(|k|+1) \rceil} - \beta|k|2^{(\gamma-2)J} \\ &\quad + \beta|k|2^{(\gamma-2)\lceil \log_2(|k|+1)\rceil}) && (31) \end{aligned}$$

for $|k| > 0$, where $\beta = \frac{1-2^{(\gamma-1)}}{1-2^{(\gamma-2)}}$. In the case $k = 0$, $r(0) = \mu_J^2 \rho^{2J}$. Finally, making use of the approximation $2^{\lceil \log_2(|k|+1) \rceil} \approx |k|$, for $|k| > 0$ we have

$$r(k) \approx C\left[(1-\beta)|k|^{(\gamma-1)} + \beta|k|2^{(\gamma-2)J}\right]. \tag{32}$$

For large J and $\gamma < 1$, the term $\beta|k|2^{(\gamma-2)J}$ is negligible, and hence the correlation function behaves like $|k|^{(\gamma-1)}$ and the power spectrum decays like $\frac{1}{|f|^\gamma}$ (see [40] for the relationship between autocorrelation functions and power spectrums of $1/f$ processes). That is, the SI-MMI model produces a non-negative, stationary processes with $1/f$ characteristics. For more details on SI wavelet models and $1/f$ processes see [36].

VI. Numerical Comparison of Wavelet-Based Intensity Estimators

Here we compare the performance of the new Bayesian estimation algorithm with several existing methods. To assess the performance of each method four test intensity functions were used. These functions were the "Doppler," "Blocks," "HeaviSine," and "Bumps" test signals proposed in [8]. Each test function is 1024 samples long (*i.e.,* $N = 1024$, $J = 10$). These functions serve as benchmark tests for signal estimators, and they were designed to be representative of a variety natural signal structures. We refer the reader to [8] for more information about the test functions. Since the intensity functions must be non-negative, each test function was shifted and scaled to obtain an intensity with a desired peak value and a minimum value of $\frac{1}{\text{peak value}}$. Realizations of counts are generated from each intensity using a standard Poisson random number generator [30]. We compare the performance of the simple estimator based on the raw counts (COUNT), a shift-invariant version of the the cross-validation estimator[7] (CV) proposed in [10], the SI-MMI model estimator described in this paper with a three component beta-mixture model for the innovations with parameters[8] $s_1 = 1$, $s_2 = 100$, and $s_3 = 10000$, the square-root estimation methods using a shift-invariant version of the Haar wavelet transform (D2), and the square-root estimation method using a shift-invariant version of the Daubechies-8 (D8) wavelet. The method proposed in [9] is not compared since it is derived under a "burst-like" process model which is not appropriate for these test functions with the exception of the Bumps function.

The square-root method first computes the square-root of the counts, then treats the square-root data as though it were Gaussian, takes the shift-invariant discrete wavelet transform [32, 33], applies a soft-threshold nonlinearity to wavelet coefficients, and computes the inverse transform of the thresholded coefficients. After this processing, the result is squared to obtain an intensity estimate. For both square-root methods the universal threshold proposed in [8] was used. We consider both the D2 and D8 wavelet (which can be applied in the case of Gaussian data) to demonstrate that our Haar-based method can outperform the square-root method even when more regular wavelets like the D8 are used. All methods employ a 5-scale wavelet transform. In practice, we could use full J-scale transforms, but their performances were roughly the same as that of the 5-scale transforms in our experiments, and the 5-scale transforms are more computationally efficient. Table 1 gives the average mean-square errors (AMSE) of the various methods for

[7]This estimator is a wavelet threshold type operation that is derived using the statistical method of cross-validation [11].

[8]The mixing parameter $p_1 = 0.001$, and parameter p_2 (and hence p_3) is determined using the data-adaptive moment-matching method given in Section IV-C.

a peak intensity of 8. Table 2 gives the AMSE of each method for a peak intensity of 128. The AMSE is estimated using 25 independent trials in each case, and each AMSE is normalized by the squared Euclidean norm of the underlying intensity function. Inspection of the tables shows that all methods offer significant improvements over the simple, count estimator. Moreover, the SI-MMI based estimator outperforms all others in every case. We also note that similar tests and comparisons were made with the *shift-variant* versions of each estimator. As expected, the shift-variant estimators did not perform as well as their shift-invariant counterparts. However, the MMI model estimator outperformed the other shift-variant methods in all cases as well.

Table I. *AMSE results for various test intensities and estimation algorithms. Peak intensity* = 8.

Intensity	*COUNT*	*CV*	*BAYES*	*D2*	*D8*
Doppler	0.1786	0.0588	0.0154	0.0548	0.0443
Blocks	0.1935	0.0617	0.0178	0.0700	0.0800
HeaviSine	0.1833	0.0552	0.0052	0.0294	0.0300
Bumps	0.5207	0.1877	0.1475	0.4570	0.4317

Table II. *AMSE results for various test intensities and estimation algorithms. Peak intensity* = 128.

Intensity	*COUNT*	*CV*	*BAYES*	*D2*	*D8*
Doppler	0.0111	0.0047	0.0026	0.0095	0.0059
Blocks	0.0120	0.0040	0.0027	0.0077	0.0126
HeaviSine	0.0115	0.0039	0.0007	0.0036	0.0028
Bumps	0.0324	0.0171	0.0143	0.1046	0.0908

VII. Application to Photon-Limited Imaging

In this section we apply the MMI models and estimation procedure to the problem of photon-limited imaging. Photon-limited imaging arises in many fields including medicine and astronomy. The fundamental problem in photon-limited imaging is the variability due to quantum effects in the emission and detection of photons. In many problems, the photon counts collected during image acquisition are well-modeled by a temporally homogeneous and spatially inhomogeneous Poisson process.

Assume that we detect photon emissions in a compact region of the plane. The photon emissions are the result of an underlying two-dimensional continuous intensity function. We are interested in estimating the intensity function from the photon detections. For practical reasons (computing and display), we seek an estimate of the intensity at a finite scale (resolution) represented by a "pixelized" intensity. A crude estimate of the pixelized intensity is obtained by simply counting the number of photons detected in each square pixel region of the plane. This "count" image is highly variable due to the random nature of the photon emission process. However, lower resolution images, obtained by counting the number of photons detected in larger square pixel regions of the plane, provide better (less variable) estimates of the low-resolution intensities. This illustrates the advantage of multiscale analysis in photon-limited imaging. Relatively reliable coarse-scale estimators of the intensity can be leveraged to obtain finer details using our multiscale Bayesian framework.

To illustrate the effectiveness of the framework in photon-limited imaging applications, we consider a simulated experiment and a real-world application to nuclear medicine imaging. Note that the MMI models and estimator are easily generalized to two dimensions. Specifically, we take the 2-d multiscale parameters to be the factors corresponding to the *multiplicative* refinement of a coarse scaling coefficient (intensity) into four finer scaling coefficients by first splitting it vertically (horizontally) into two halves, then next horizontally (vertically) splitting each half into two quarters. That is, if $\Lambda_{k,l}$ is a 2-d intensity function, then we define $\Lambda_{0,k,l} \equiv \Lambda_{k,l}$ at the finest scale $j = 0$ and for coarser scales take

$$\begin{aligned}
\Lambda_{j+1,k,l} &= \Lambda_{j,2k,2l} + \Lambda_{j,2k,2l+1} + \Lambda_{j,2k+1,2l} + \Lambda_{j,2k+1,2l+1}, \\
Y^1_{j+1,k,l} &= \frac{\Lambda_{j,2k,2l} + \Lambda_{j,2k,2l+1}}{\Lambda_{j,2k,2l} + \Lambda_{j,2k,2l+1} + \Lambda_{j,2k+1,2l} + \Lambda_{j,2k+1,2l+1}}, \\
Y^2_{j+1,k,l} &= \frac{\Lambda_{j,2k,2l}}{\Lambda_{j,2k,2l} + \Lambda_{j,2k,2l+1}}, \\
Y^3_{j+1,k,l} &= \frac{\Lambda_{j,2k+1,2l}}{\Lambda_{j,2k+1,2l} + \Lambda_{j,2k+1,2l+1}}.
\end{aligned} \tag{33}$$

Note that in the 2-d case we have three sets of multiplicative innovations, one vertical set Y^1 and two horizontal sets Y^2 and Y^3. In the analysis of count images, each scaling coefficient is the sum of four counts, and each wavelet coefficient is simply the difference of two counts. Hence, all the machinery developed for the one-dimensional case, based on sums and differences of pairs of counts, is immediately applicable to two (or even higher) dimensional data. Note that this 2-d multiscale analysis defined above differs from the standard 2-d Haar wavelet analysis [19], which involves a vertical, horizontal and diagonal differences. We use the alternative 2-d analysis because, unlike the standard 2-d Haar analysis, it allows us to decouple the Poisson problem, just as in the 1-d case. In both experiments, the 5-scale Haar transform is employed and a three component beta-mixture density is used for the SI-MMI model with fixed shape parameters $s_1 = 1$, $s_2 = 100$, and $s_3 = 10000$. The mixing probability p_1 is fixed at 0.001, and p_2 (and p_3) is adapted to the data at each scale using the moment-matching method described in Section IV-C. We have found these choices of s_1, s_2, and s_3, combined with the flexibility of the data-adaptive p_2, to provide very good results for a wide-variety of imagery.

A. Photon-Limited Imaging Simulation

Fig. 8 depicts a typical realization of a simulated photon-limited imaging application and the resulting estimates provided by the MMI and SI-MMI models. The maximum intensity in the image in Fig. 8(a) is 60.00, and the average intensity is 25.40. Hence, this simulation models a fairly low intensity (low SNR) imaging problem. A realization of counts is generated from this intensity using a standard Poisson random number generator [30]. Note the visual improvement provided by the MMI and SI-MMI model estimates in Fig. 8(c) and (d), respectively, in comparison to the count image Fig. 8(b). Furthermore, the estimate based on the SI-MMI model appears to be better than that of the MMI model. In fact, in 25 independent trials of this experiment we estimated the average mean squared pixel error to be 25.28 for the count image, 7.36 for the MMI model estimator, and 4.01 for the SI-MMI model estimator.

B. Application to Nuclear Medicine Imaging

Nuclear medicine imaging is a widely used commercial imaging modality [41]. Unlike many other medical imaging techniques, nuclear medicine imaging can provide both anatomical *and* functional information. However, nuclear medicine imaging has a much lower signal-to-noise ratio relative to other imaging techniques. Hence, improvements in image quality via optimized signal processing represent a significant opportunity to advance the state-of-the-art in nuclear medicine.

Nuclear medicine images are acquired by the following procedure [41]. Radioactive pharmaceuticals that are targeted for uptake in specific regions of the body are injected into the patient's bloodstream. As the radioactive pharmaceuticals decay, gamma rays are emitted from within the patient. Imaging the gamma ray emissions provides a mapping of the distribution of the pharmaceutical, and hence a mapping of the anatomy or physiologic function of the patient. Gamma rays are detected and spatially located using a gamma camera, which converts gamma rays into light. Photomultiplier tubes then detect and locate the emissions. The raw nuclear medicine data is an image of photon detections (counts). The raw data may be viewed directly or used for tomographic reconstruction. The major limitation of nuclear medicine imaging is the low-count levels acquired in typical studies, due in part to the limited level of radioactive dosage required to insure patient safety. Because of the variability of low-count images, it is very common to employ a post-filtering or estimation procedure to obtain a "better" estimate of the underlying intensity. An excellent discussion of the potential diagnostic benefits of various frequency domain processing methods is given in [5]. Advantages of multiscale methods over frequency domain methods for photon-limited imaging problems are discussed in [42].

To illustrate the potential of our multiscale Bayesian framework in nuclear medicine imaging, consider the spine and heart studies depicted in Fig. 9. Fig. 9(a) depicts the count image from a nuclear medicine spine study. The radiopharmaceutical used here is Technetium-99m labeled diphosphonate. In bone studies such as this, brighter areas indicate increased uptake of blood in areas where bone growth is occurring. This may reflect areas where bone damage has occurred. Functional changes in bone can be detected using nuclear medicine images before they will show up in x-ray images. The maximum count in this image is 178 in the "hot-spot" at the bottom of the spine. The maximum count in the upper portion of the spine is 75. Fig. 9(b) shows the SI-MMI model estimate of the underlying intensity. Fig. 9(c) depicts an image of a heart obtained from a nuclear medicine study. The image was obtained using the radiopharmaceutical Thallium-201, and the maximum count is 33. In this type of study, the radiopharmaceutical is injected into the bloodstream of the patient and moves into the heart wall in proportion to the local degree of blood perfusion. The purpose of this procedure is to determine if there is decreased blood flow to the heart muscle. Fig. 9(d) shows the SI-MMI model estimate. In both studies, we see that the SI-MMI model estimator preserves the important image structure and has significantly lower variance compared to the raw count image. The intensity estimates provided by the SI-MMI model may enable better diagnosis in clinical nuclear medicine. The feedback we have received from our collaborators in several radiology departments has been very encouraging, and we are beginning more exhaustive studies to assess the potential of our new framework to improve the diagnostic capability of nuclear medicine imaging.

VIII. Conclusions

We have introduced a Bayesian approach for Poisson intensity estimation. We argued that the multiscale analysis is the right framework for carrying the Bayesian estimation. We introduced a novel MMI prior model for intensity functions based on a *multiplicative* innovations structure. The MMI captures many of the key features of real-world intensity functions and provides an excellent match to the Poisson distribution. The MMI model facilitates a multiscale Bayesian estimation procedure that proceeds in a natural fashion from coarse-to-fine resolutions. The estimator has a simple closed expression and can be implemented in $O(N)$ operations, where N is the dimension of the finest resolution of the discretized intensity. The issue of choosing the parameters of the prior model was addressed, and a simple moment-matching method was proposed for fitting the parameters to a given set of data. The MMI model was extended to a shift-invariant (stationary) model called the SI-MMI model, and it was shown that the SI-MMI model has a $1/f$ correlation structure that is more regular than that of the MMI model.

We have illustrated the performance of the multiscale Bayesian estimator by comparing its performance to other wavelet-based approaches on several benchmark problems. We have also studied the application of the framework to photon-limited imaging problems, and examined its potential to improve the quality of nuclear medicine images. Our initial investigations are promising, and the feedback we have received from our collaborators in various radiology departments has been very encouraging. Ongoing work is investigating the use of the multiscale Bayesian framework in other imaging applications, including computed tomography.

The framework also appears to have potential in other applications such as network tomography [4] and network traffic modeling and synthesis. We are also investigating relationships between the MMI model and the theories of cascade models of physics [24, 23] and Polya trees in statistics [25, 26]. Finally, the framework for multiscale Bayesian analysis of Poisson processes presented here can be extended to more general priors based on hidden Markov models, which capture the key inter-scale dependencies present in many natural intensities. The interested reader is referred to [43] for more information.

Acknowledgments

Special thanks go to Dr. Robert Hellman of the Medical College of Wisconsin for supplying the nuclear medicine data. The authors also thank David Nowak for his expert advice on nuclear medicine imaging and stimulating discussions on this subject. The second author thanks Mike West for pointing out the connection between the MMI model and Polya trees. Finally, both authors thank the anonymous reviewers for the helpful comments and suggestions and especially for pointing out the connection between this work and cascade models.

Appendix A
Posterior Distributions

In this appendix we show that

$$f\left(\boldsymbol{\lambda}_j|\mathbf{c}_0\right) \;=\; f\left(\boldsymbol{\lambda}_j|\mathbf{c}_j\right) \tag{34}$$

and

$$f\left(\delta_{j,k}|\mathbf{c}_0\right) \;=\; f\left(\delta_{j,k}|c_{j-1,2k}, c_{j-1,2k+1}\right). \tag{35}$$

These are proved in the following two propositions.

Proposition 1 $\quad f(\boldsymbol{\lambda}_j|\mathbf{c}_0) = f(\boldsymbol{\lambda}_j|\mathbf{c}_j) \quad$ for $j = 0, \cdots, J$

Proof: Since the same information contained in the set $\{\mathbf{c}_0\}$ is contained in $\{\mathbf{c}_0, \mathbf{c}_j\}$, applying Bayes' theorem $f\left(\boldsymbol{\lambda}_j|\mathbf{c}_0\right)$ may be written as

$$f\left(\boldsymbol{\lambda}_j|\mathbf{c}_0\right) \;=\; f\left(\boldsymbol{\lambda}_j|\mathbf{c}_0, \mathbf{c}_j\right) \;=\; \mathrm{P}\left(\mathbf{c}_0|\mathbf{c}_j, \boldsymbol{\lambda}_j\right) \frac{f(\boldsymbol{\lambda}_j|\mathbf{c}_j)}{\mathrm{P}\left(\mathbf{c}_0|\mathbf{c}_j\right)}.$$

Thus, it sufficies to show that $\mathrm{P}(\mathbf{c}_0|\mathbf{c}_j, \boldsymbol{\lambda}_j) = \mathrm{P}(\mathbf{c}_0|\mathbf{c}_j)$ for $j = 1, \cdots, J$.

Define sequences of even and odd elements of $\mathbf{c}_{j-1}$ and $\boldsymbol{\lambda}_{j-1}$: $\mathbf{c}^e_{j-1} = (c_{j-1,0}, c_{j-1,2}, \cdots, c_{j-1,N/2^{j-1}-2})$ and $\mathbf{c}^o_{j-1} = (c_{j-1,1}, c_{j-1,3}, \cdots, c_{j-1,N/2^{j-1}-1})$, and similarly for $\boldsymbol{\lambda}^e_{j-1}$ and $\boldsymbol{\lambda}^o_{j-1}$. Clearly,

$$\mathrm{P}(\mathbf{c}_{j-1}|\mathbf{c}_j, \boldsymbol{\lambda}_j) \;=\; \begin{cases} \frac{\mathrm{P}(\boldsymbol{C}^e_{j-1}=\mathbf{c}^e_{j-1}, \boldsymbol{C}^o_{j-1}=\mathbf{c}^o_{j-1}|\boldsymbol{\lambda}_j)}{\mathrm{P}(\boldsymbol{C}_j=\mathbf{c}_j|\boldsymbol{\lambda}_j)} & \text{if } \mathbf{c}^o_{j-1} = \mathbf{c}_j - \mathbf{c}^e_{j-1} \geq \mathbf{0} \\ 0 & \text{otherwise.} \end{cases} \tag{36}$$

Then, with the aid of the total probability theorem we may write for the non-trivial case

$$\begin{aligned}
\mathrm{P}(\mathbf{c}_{j-1}|\mathbf{c}_j,\boldsymbol{\lambda}_j) &= \frac{\int \mathrm{P}(\mathbf{c}^e_{j-1},\mathbf{c}_j-\mathbf{c}^e_{j-1}|\boldsymbol{\lambda}_j,\boldsymbol{\lambda}^e_{j-1})\, f(\boldsymbol{\lambda}^e_{j-1}|\boldsymbol{\lambda}_j)\, d\boldsymbol{\lambda}^e_{j-1}}{\mathrm{P}(\mathbf{c}_j|\boldsymbol{\lambda}_j)} \\
&= \frac{\int \prod_k \mathrm{P}(c_{j-1,2k}|\lambda_{j-1,2k}) \prod_k \mathrm{P}(c_{j,k}-c_{j-1,2k}|\lambda_{j,k}-\lambda_{j-1,2k})\, f(\boldsymbol{\lambda}^e_{j-1}|\boldsymbol{\lambda}_j)\, d\boldsymbol{\lambda}^e_{j-1}}{\prod_k \mathrm{P}(c_{j,k}|\lambda_{j,k})} \\
&= \frac{\int \prod_k e^{-\lambda_{j-1,2k}} \frac{(\lambda_{j-1,2k})^{c_{j-1,2k}}}{(c_{j-1,2k})!} \prod_k e^{-\lambda_{j,k}+\lambda_{j-1,2k}} \frac{(\lambda_{j,k}-\lambda_{j-1,2k})^{c_{j,k}-c_{j-1,2k}}}{(c_{j,k}-c_{j-1,2k})!}\, f(\boldsymbol{\lambda}^e_{j-1}|\boldsymbol{\lambda}_j)\, d\boldsymbol{\lambda}^e_{j-1}}{\prod_k e^{-\lambda_{j,k}} \frac{(\lambda_{j,k})^{c_{j,k}}}{(c_{j,k})!}} \\
&= \int \prod_k \binom{c_{j,k}}{c_{j-1,2k}} \left(\frac{\lambda_{j-1,2k}}{\lambda_{j,k}}\right)^{c_{j-1,2k}} \left(1-\frac{\lambda_{j-1,2k}}{\lambda_{j,k}}\right)^{c_{j,k}-c_{j-1,2k}} f(\boldsymbol{\lambda}^e_{j-1}|\boldsymbol{\lambda}_j)\, d\boldsymbol{\lambda}^e_{j-1},
\end{aligned}$$

where all the indicated products are over $k=0$ through $N/2^j-1$. In these expressions we have exploited the conditional independence of the data scaling coefficients at any specific scale given their corresponding intensity scaling coefficients.

Since the innovation variates are given by $Y_{j,k}=\frac{\Lambda_{j-1,2k}}{\Lambda_{j,k}}$ (see (11)), the conditional densities within the integral signs may also be written as $f_{\boldsymbol{\Lambda}^e_{j-1}|\boldsymbol{\Lambda}_j}(\boldsymbol{\lambda}^e_{j-1}|\boldsymbol{\lambda}_j)=\frac{1}{\prod_k \lambda_{j,k}} f_{\boldsymbol{Y}_j}\left(\left(\frac{\lambda_{j-1,2k}}{\lambda_{j,k}}\right)_k | (\lambda_{j,k})_k\right)$, where $\boldsymbol{Y}_j=(Y_{j,0},Y_{j,1},\cdots,Y_{j,N/2^j-1})$. Then, with a change of variables and recalling the mutual independence of $\boldsymbol{Y}_j$ and $\boldsymbol{\Lambda}_j$, we obtain the equivalent expression

$$\mathrm{P}(\mathbf{c}_{j-1}|\mathbf{c}_j,\boldsymbol{\lambda}_j) = \int \prod_k \binom{c_{j,k}}{c_{j-1,2k}} (y_{j,k})^{c_{j-1,2k}} (1-y_{j,k})^{c_{j,k}-c_{j-1,2k}} f(\boldsymbol{y}_j)\, d\boldsymbol{y}_j. \tag{37}$$

The absence of $\boldsymbol{\lambda}_j$ in this expression indicates that $\mathbf{c}_{j-1}|\mathbf{c}_j$ is, at least explicitly, independent of $\boldsymbol{\lambda}_j$. However, one may argue that $\boldsymbol{\lambda}_j$ intervenes only functionally to define the probability mass $\mathrm{P}(\mathbf{c}_{j-1}|\mathbf{c}_j,\boldsymbol{\lambda}_j)$. That is, if we denote the right side of (37) by $g(\mathbf{c}_{j-1},\mathbf{c}_j)$, we must still verify whether $g(\mathbf{c}_{j-1},\mathbf{c}_j)=\mathrm{P}(\mathbf{c}_{j-1}|\mathbf{c}_j)$. We may achieve this as follows.

$$\begin{aligned}
\mathrm{P}(\mathbf{c}_{j-1}|\mathbf{c}_j) &= \int \mathrm{P}(\mathbf{c}_{j-1}|\mathbf{c}_j,\boldsymbol{\lambda}_j)\, f(\boldsymbol{\lambda}_j|\mathbf{c}_j)\, d\boldsymbol{\lambda}_j \\
&= \int g(\mathbf{c}_{j-1},\mathbf{c}_j)\, f(\boldsymbol{\lambda}_j|\mathbf{c}_j)\, d\boldsymbol{\lambda}_j \;=\; g(\mathbf{c}_{j-1},\mathbf{c}_j).
\end{aligned}$$

Therefore, $\mathrm{P}(\mathbf{c}_{j-1}|\mathbf{c}_j,\boldsymbol{\lambda}_j)=\mathrm{P}(\mathbf{c}_{j-1}|\mathbf{c}_j)$. Using this result one can arrive at the desired final conclusion $\mathrm{P}(\mathbf{c}_0|\mathbf{c}_j,\boldsymbol{\lambda}_j)=\mathrm{P}(\mathbf{c}_0|\mathbf{c}_j)$ using induction.

Proposition 2 $f(\delta_{j,k}|\mathbf{c}_0)=f(\delta_{j,k}|c_{j-1,2k},c_{j-1,2k+1})$ for $j=1,\cdots,J$

Proof: In order to keep the mathematical notation as clean as possible in these and subsequent derivations, we introduce the following simplifying notations: $c_{\mathbf{1}}=c_{j-1,2k}$ and $c_{\mathbf{2}}=c_{j-1,2k+1}$, $\mathbf{c}^*_{j-1}=\mathbf{c}_{j-1}\backslash(c_{\mathbf{1}},c_{\mathbf{2}})$, and $\boldsymbol{y}^*_j=\boldsymbol{y}_j\backslash y_{j,k}$. Now, since $y_{j,k}=\frac{1}{2}(1+\delta_{j,k})$ (see (10)), it suffices to show that $f(y_{j,k}|\mathbf{c}_0)=f(y_{j,k}|c_{\mathbf{1}},c_{\mathbf{2}})$ for $j=1,\cdots,J$. This may be done as follows.

By the total probability and Bayes' theorems,

$$\begin{aligned} f(y_{j,k}|\mathbf{c}_0) &= \int f(\boldsymbol{y}_j, \boldsymbol{\lambda}_j|\mathbf{c}_0)\, d\boldsymbol{y}_j^*\, d\boldsymbol{\lambda}_j \\ &= \int \frac{\mathrm{P}(\mathbf{c}_0|\boldsymbol{y}_j, \boldsymbol{\lambda}_j)}{\mathrm{P}(\mathbf{c}_0)} f(\boldsymbol{y}_j, \boldsymbol{\lambda}_j)\, d\boldsymbol{y}_j^*\, d\boldsymbol{\lambda}_j. \end{aligned}$$

Since the information conveyed by $\{\boldsymbol{y}_j, \boldsymbol{\lambda}_j\}$ is the same as that conveyed by $\{\boldsymbol{\lambda}_{j-1}\}$, $\mathrm{P}(\mathbf{c}_0|\boldsymbol{y}_j, \boldsymbol{\lambda}_j) = \mathrm{P}(\mathbf{c}_0|\boldsymbol{\lambda}_{j-1}) = f(\boldsymbol{\lambda}_{j-1}|\mathbf{c}_0)\mathrm{P}(\mathbf{c}_0)/f(\boldsymbol{\lambda}_{j-1})$. Using Proposition 1, this becomes $\mathrm{P}(\mathbf{c}_0|\boldsymbol{y}_j, \boldsymbol{\lambda}_j) = f(\boldsymbol{\lambda}_{j-1}|\mathbf{c}_{j-1})\mathrm{P}(\mathbf{c}_0)/f(\boldsymbol{\lambda}_{j-1}) = \mathrm{P}(\mathbf{c}_{j-1}|\boldsymbol{\lambda}_{j-1})\mathrm{P}(\mathbf{c}_0)/\mathrm{P}(\mathbf{c}_{j-1})$. Upon substitution in the above integral

$$\begin{aligned} f(y_{j,k}|\mathbf{c}_0) &= \int \frac{\mathrm{P}(\mathbf{c}_{j-1}|\boldsymbol{\lambda}_{j-1})}{\mathrm{P}(\mathbf{c}_{j-1})} f(\boldsymbol{y}_j, \boldsymbol{\lambda}_j)\, d\boldsymbol{y}_j^*\, d\boldsymbol{\lambda}_j \\ &= \int \frac{\mathrm{P}(\mathbf{c}_{j-1}|\boldsymbol{y}_j, \boldsymbol{\lambda}_j)}{\mathrm{P}(\mathbf{c}_{j-1})} f(\boldsymbol{y}_j, \boldsymbol{\lambda}_j)\, d\boldsymbol{y}_j^*\, d\boldsymbol{\lambda}_j \\ &= \int \mathrm{P}(c_{\mathbf{1}}, c_{\mathbf{2}}|y_{j,k}, \lambda_{j,k}) \frac{\mathrm{P}(\mathbf{c}_{j-1}^*|\boldsymbol{y}_j^*, \boldsymbol{\lambda}_j)}{\mathrm{P}(\mathbf{c}_{j-1})} f(\boldsymbol{y}_j, \boldsymbol{\lambda}_j)\, d\boldsymbol{y}_j^*\, d\boldsymbol{\lambda}_j \\ &= f(y_{j,k}) \int \mathrm{P}(c_{\mathbf{1}}, c_{\mathbf{2}}|y_{j,k}, \lambda_{j,k}) \frac{\mathrm{P}(\mathbf{c}_{j-1}^*|\boldsymbol{\lambda}_j)}{\mathrm{P}(\mathbf{c}_{j-1})} f(\boldsymbol{\lambda}_j)\, d\boldsymbol{\lambda}_j \\ &= f(y_{j,k})\, y_{j,k}^{c_{\mathbf{1}}} (1 - y_{j,k})^{c_{\mathbf{2}}} \int \frac{\lambda^{c_{\mathbf{1}}+c_{\mathbf{2}}} e^{-\lambda}}{c_{\mathbf{1}}!\, c_{\mathbf{2}}!} \frac{\mathrm{P}(\mathbf{c}_{j-1}^*|\boldsymbol{\lambda}_j)}{\mathrm{P}(\mathbf{c}_{j-1})} f(\boldsymbol{\lambda}_j)\, d\boldsymbol{\lambda}_j. \end{aligned}$$

Thus, $f(y_{j,k}|\mathbf{c}_0) = f(y_{j,k}|c_{\mathbf{1}}, c_{\mathbf{2}})$.

Appendix B

Multiplicative Innovation's Optimal Estimation

In this section we find a closed form for the optimal estimate $\widehat{\delta}_{j,k}$ of the innovation coefficient $\delta_{j,k}$. For the sake of simplicity, in the few steps that follow we will disregard the indices j and k, and simply write δ for $\delta_{j,k}$ and similary for other quantities.

The minimum mean square error (mmse) optimal estimate of the innovation coefficient δ, given all the information available in $\mathbf{c}_0$ is given by

$$\widehat{\delta} \;=\; \mathrm{E}\,[\Delta|\mathbf{c}_0] \;=\; \int_{-1}^{1} \delta\, f(\delta|\mathbf{c}_0)\, d\delta \;=\; \int_{-1}^{1} \delta\, f(\delta|c_{\mathbf{1}}, c_{\mathbf{2}})\, d\delta$$

in accordance with (35). Applying Bayes' theorem to this expression we obtain

$$\widehat{\delta} \;=\; \frac{\int_{-1}^{1} \delta\, \mathrm{P}(c_{\mathbf{1}}, c_{\mathbf{2}}|\delta) f(\delta)\, d\delta}{\int_{-1}^{1} \mathrm{P}(c_{\mathbf{1}}, c_{\mathbf{2}}|\delta) f(\delta)\, d\delta} \;=\; \frac{\int_{-1}^{1} \int_0^\infty \delta\, \mathrm{P}(c_{\mathbf{1}}, c_{\mathbf{2}}|\delta, \lambda) f(\lambda|\delta)\, d\lambda\, f(\delta)\, d\delta}{\int_{-1}^{1} \int_0^\infty \mathrm{P}(c_{\mathbf{1}}, c_{\mathbf{2}}|\delta, \lambda) f(\lambda|\delta)\, d\lambda\, f(\delta)\, d\delta}.$$

Here, $f(\lambda|\delta) = f(\lambda)$ due to the independence of $\Lambda_{j,k}$ and $\Delta_{j,k}$. Also, notice that $c_{\mathbf{1}}$ and $c_{\mathbf{2}}$ only depend on λ and δ by way of their respective statistical means $\lambda_{\mathbf{1}} = \frac{1}{2}\lambda\,(1+\delta)$ and $\lambda_{\mathbf{2}} = \frac{1}{2}\lambda\,(1-\delta)$ (See (11) and (12)). Given $\lambda_{\mathbf{1}}$ and $\lambda_{\mathbf{2}}$, λ and δ do not convey any further information on the behavior of $c_{\mathbf{1}}$ and $c_{\mathbf{2}}$. Moreover, given $\lambda_{\mathbf{1}}$ and $\lambda_{\mathbf{2}}$, $c_{\mathbf{1}}$ and $c_{\mathbf{2}}$ are independent. Therefore,

$$\begin{aligned}
\widehat{\delta} &= \frac{\int_{-1}^{1}\int_{0}^{\infty} \delta \, \mathrm{P}\left(c_{\mathbf{1}}, c_{\mathbf{2}} \,\middle|\, \lambda\frac{1+\delta}{2}, \lambda\frac{1-\delta}{2}\right) f(\lambda)\, d\lambda\, f(\delta)\, d\delta}{\int_{-1}^{1}\int_{0}^{\infty} \mathrm{P}\left(c_{\mathbf{1}}, c_{\mathbf{2}} \,\middle|\, \lambda\frac{1+\delta}{2}, \lambda\frac{1-\delta}{2}\right) f(\lambda)\, d\lambda\, f(\delta)\, d\delta} \\
&= \frac{\int_{-1}^{1}\int_{0}^{\infty} \delta \, \frac{e^{-\lambda(1+\delta)/2}}{c_{\mathbf{1}}!}\left(\lambda\frac{1+\delta}{2}\right)^{c_{\mathbf{1}}} \frac{e^{-\lambda(1-\delta)/2}}{c_{\mathbf{2}}!}\left(\lambda\frac{1-\delta}{2}\right)^{c_{\mathbf{2}}} f(\lambda)\, d\lambda\, f(\delta)\, d\delta}{\int_{-1}^{1}\int_{0}^{\infty} \frac{e^{-\lambda(1+\delta)/2}}{c_{\mathbf{1}}!}\left(\lambda\frac{1+\delta}{2}\right)^{c_{\mathbf{1}}} \frac{e^{-\lambda(1-\delta)/2}}{c_{\mathbf{2}}!}\left(\lambda\frac{1-\delta}{2}\right)^{c_{\mathbf{2}}} f(\lambda)\, d\lambda\, f(\delta)\, d\delta}.
\end{aligned}$$

It is now possible to factor out every λ-dependent term from the integrals in δ. Doing this, the resulting integrals in λ in the numerator and denominator cancel out leading to

$$\widehat{\delta} = \frac{\int_{-1}^{1} \delta \,(1+\delta)^{c_{\mathbf{1}}}\,(1-\delta)^{c_{\mathbf{2}}}\, f(\delta)\, d\delta}{\int_{-1}^{1} (1+\delta)^{c_{\mathbf{1}}}\,(1-\delta)^{c_{\mathbf{2}}}\, f(\delta)\, d\delta}. \tag{38}$$

Substituting the beta mixture model (13) into this expression and carrying out the integration we obtained the desired result:

$$\widehat{\delta}_{j,k} = d_{j,k}\, \frac{\sum_i p_i \frac{B(s_i+c_{\mathbf{1}}, s_i+c_{\mathbf{2}})}{B(s_i,s_i)\,(2s_i+c)}}{\sum_i p_i \frac{B(s_i+c_{\mathbf{1}}, s_i+c_{\mathbf{2}})}{B(s_i,s_i)}}. \tag{39}$$

References

[1] D. L. Snyder and M. I. Miller, *Random Point Processes in Time and Space.* New York: Springer-Verlag, 1991.

[2] D. L. Snyder, A. M. Hammoud, and R. L. White, "Image recovery from data acquired with a charge-coupled-device camera," *J. Opt. Soc. Am. A*, vol. 10, no. 5, pp. 1014–1023, 1993.

[3] H. Heffes and D. M. Lucantoni, "A Markov modulated characterization of packetized voice and data traffic and related statistical multiplexer performance," *IEEE J. Selected Areas Comm.*, vol. 4, pp. 856–868, 1986.

[4] Y. Vardi, "Network tomography: Estimating source-destination traffic intensities from link data," *J. Amer. Statist. Assoc.*, vol. 91, pp. 365–377, 1996.

[5] M. A. King, R. B. Schwinger, P. W. Doherty, and B. C. Penney, "Two-dimensional filtering of SPECT images using the Metz and Wiener filters," *J. Nuc. Med.*, vol. 25, pp. 1234–1240, 1984.

[6] H. C. Andrews and B. R. Hunt, *Digital Image Restoration.* Englewood Cliffs, New Jersey: Prentice Hall, 1977.

[7] D. T. Kuan, A. A. Sawchuk, T. C. Strand, and P. Chavel, "Adaptive noise smoothing filter for images with signal-dependent noise," *IEEE Trans. Pattern Anal. Machine Intelligence*, vol. 7, pp. 165–177, 1985.

[8] D. Donoho and I. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *J. Amer. Statist. Assoc.*, vol. 90, pp. 1200–1224, Dec. 1995.

[9] E. D. Kolaczyk, "Wavelet shrinkage estimation of certain Poisson intensity signals using corrected thresholds," *Statistica Sinica,* under revision, 1997.

[10] R. D. Nowak and R. G. Baraniuk, "Wavelet-domain filtering for photon imaging systems," *Proc. SPIE, Wavelet Applications in Signal and Image Processing V*, vol. 3169, pp. pp. 55–66, August 1997.

[11] R. D. Nowak, "Optimal signal estimation using cross-validation," *IEEE Signal Processing Letters*, vol. 4, no. 1, pp. 23–25, 1997.

[12] R. Nowak, R. Hellman, D. Nowak, and R. Baraniuk, "Wavelet domain filtering for nuclear medicine imaging," in *Proc. IEEE Med. Imaging Conf.*, pp. 279–290, 1996.

[13] J.-C. Pesquet, H. Krim, and E. Hamman, "Bayesian approach to best basis selection," in *IEEE Int. Conf. on Acoust., Speech, Signal Proc. — ICASSP '96*, (Atlanta), pp. 2634–2637, 1996.

[14] M. Crouse, R. Nowak, and R. Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models," *IEEE Trans. Signal Processing*, vol. 46, pp. 886–902, 1998.

[15] H. Chipman, E. Kolaczyk, and R. McCulloch, "Adaptive Bayesian wavelet shrinkage," *J. Amer. Statist. Assoc.*, vol. 92, pp. 1413–1421, 1997.

[16] B. Vidakovic, "Nonlinear wavelet shrinkage with Bayes rules and Bayes factors," *J. Amer. Statist. Assoc.*, vol. 93, pp. 173–179, 1998.

[17] E. P. Simoncelli and E. H. Adelson, "Noise removal via Bayesian wavelet coring," in *IEEE Int. Conf. on Image Proc. — ICIP 1996*, (Switzerland), September 1996.

[18] K. Timmermann and R. Nowak, "Multiscale Bayesian estimation of Poisson intensities," in *Proc. Thirty-First Asilomar Conf. Signals, Systems, and Comp.,* Pacific Grove, CA, pp. 85–90, IEEE Computer Society Press, 1997.

[19] I. Daubechies, *Ten Lectures on Wavelets.* Philidelphia: SIAM CBMS-NSF Series in Applied Mathematics, no. 61, 1992.

[20] N. L. Johnson, S. Kotz, and A. W. Kemp, *Univariate Discrete Distributions.* New York: John Wiley and Sons, 1992.

[21] L. L. Scharf, *Statistical Signal Processing. Detection, Estimation, an Time Series Analysis.* Reading, MA: Addison-Wesley, 1991.

[22] B. Mandelbrot, "Intermittant turbulence in self-similar cascades: Divergence of high moments and dimension of the carrier," *J. Fluid Mech.*, 1974.

[23] S. Lovejoy and D. Schertzer, "Multifractals and rain" in *New Uncertainty concepts in Hydrology and Hydrological modelling.* Cambridge Press, 1995.

[24] C. J. G. Evertsz and B. B. Mandelbrot., "Multifractal measures" in *Chaos and Fractals: New Frontiers in Science.* Springer-Verlag, 1992.

[25] R. D. Mauldin, W. D. Sudderth, and S. C. Williams, "Polya trees and random distributions," *Ann. Stat.*, vol. 20, pp. 1203–1221, 1992.

[26] M. Lavine, "Some aspects of Polya tree distributions for statistical modelling," *Ann. Stat.*, vol. 20,

pp. 1222–1235, 1992.

[27] K. Birney and T. Fischer, "On the modeling of DCT and subband image data for compression," *IEEE Transactions on Image Processing*, vol. 4, pp. 186–193, February 1995.

[28] S. LoPresto, K. Ramchandran, and M. T. Orchard, "Image coding based on mixture modeling of wavelet coefficients and a fast estimation-quantization framework," in *Data Compression Conference '97*, (Snowbird, Utah), pp. 221–230, 1997.

[29] E. Kolaczyk, "Bayesian multi-scale models for Poisson processes," *Technical Report 468, Dept. of Statistics, University of Chicago*, 1998.

[30] C. Robert, *The Bayesian Choice: A Decision Theoretic Motivation.* New York: Springer-Verlag, 1994.

[31] E. Simoncelli, W. Freeman, E. Adelson, and D. Heeger, "Shiftable multiscale transforms," *IEEE Trans. Inform. Theory*, vol. 38, no. 2, pp. 587–607, 1992.

[32] R. Coifman and D. Donoho, "Translation invariant de-noising," in *Lecture Notes in Statistics: Wavelets and Statistics*, vol. New York: Springer-Verlag, pp. 125–150, 1995.

[33] M. Lang, H. Guo, J. E. Odegard, C. S. Burrus, and R. O. Wells, "Noise reduction using an undecimated discrete wavelet transform," *IEEE Signal Processing Letters*, vol. 3, no. 1, pp. 10–12, 1996.

[34] J.-C. Pesquet, H. Krim, and H. Carfantan, "Time-invariant orthonormal wavelet representations," *IEEE Trans. Signal Processing*, vol. 44, no. 8, pp. 1964–1970, 1996.

[35] G. P. Nason and B. W. Silverman, "The stationary wavelet transform and some statistical applications," in *Lecture Notes in Statistics: Wavelets and Statistics*, vol. New York: Springer-Verlag, pp. 281–299, 1995.

[36] R. Nowak, "Shift invariant wavelet-based statistical models and $1/f$ processes," *Proc. IEEE Digital Signal Processing Workshop*, paper no. 83, Bryce Canyon, UT, 1998.

[37] A. P. Pentland, "Fractal-based description of natural scenes," *IEEE Trans. Patt. Anal. Mach. Intell.*, pp. 661–674, July 1984.

[38] A. van der Schaaf amd J. van Hateren, "Modelling the power spectra of natural images," *Vision Research*, vol. 36, no. 17, pp. 2759–2770, 1996.

[39] B. J. West and M. F. Shlesinger, "On the ubiquity of 1/f noise," *Intl. J. of Modern Physics (B)*, vol. 3, no. 6, pp. 795–819, 1989.

[40] G. Wornell, *Signal Processing with Fractals. A Wavelet-Based Approach.* Englewood Cliffs, New Jersey: Prentice Hall, 1996.

[41] J. A. Sorenson and M. E. Phelps, *Physics in Nuclear Medicine.* New York: Grune & Stratton, 1980.

[42] R. D. Nowak and R. G. Baraniuk, "Wavelet-based filtering for photon imaging systems," *IEEE Trans. Image Processing*, submitted April 1997.

[43] R. Nowak, "Multiscale hidden Markov models for Bayesian image analysis," in to appear in *Bayesian*

Inference in Wavelet Based Models, Springer-Verlag, 1999. Editors B. Vidakovic and P. Müller.

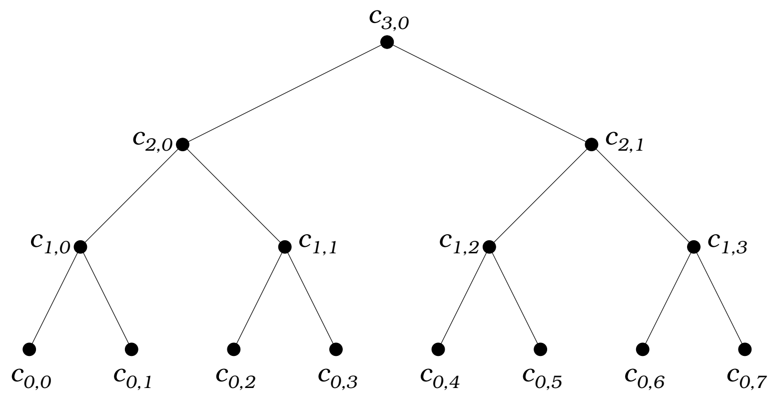


Fig. 1. *Multiscale scaling coefficients* $\{c_{j,k}\}$. *At the top, we have scaling coefficients at the coarsest resolution. At the bottom, we have the finest resolution, expressed by the data themselves. The connecting segments illustrate the functional dependencies among the various scaling coefficients* $c_{j,k}$ *according to expression* (3).

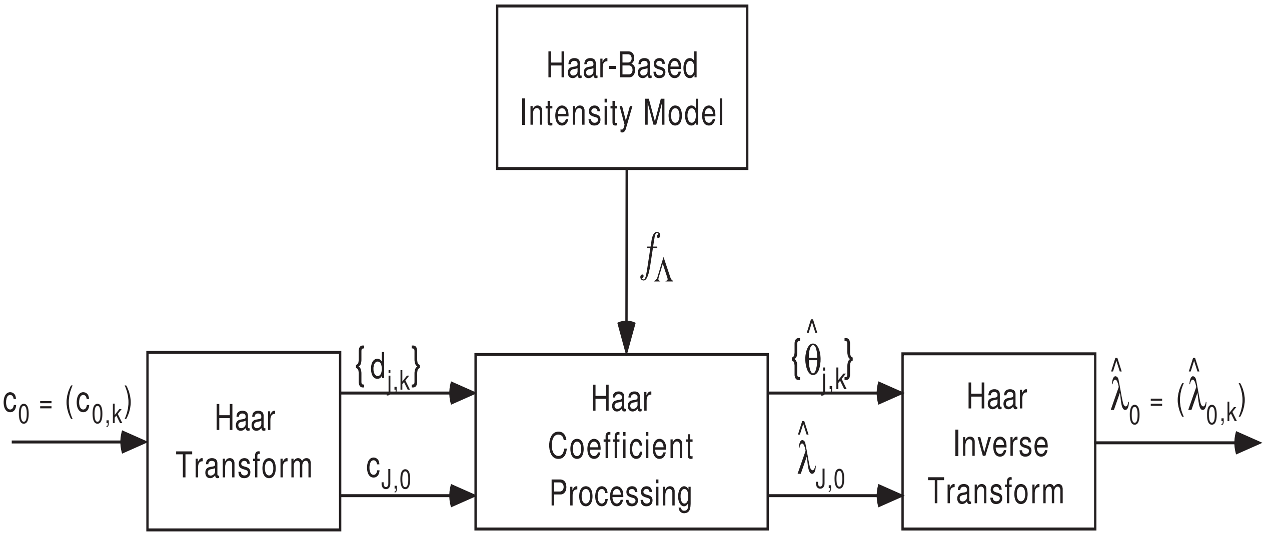


Fig. 2. *Structure of a Haar-based intensity estimator.*

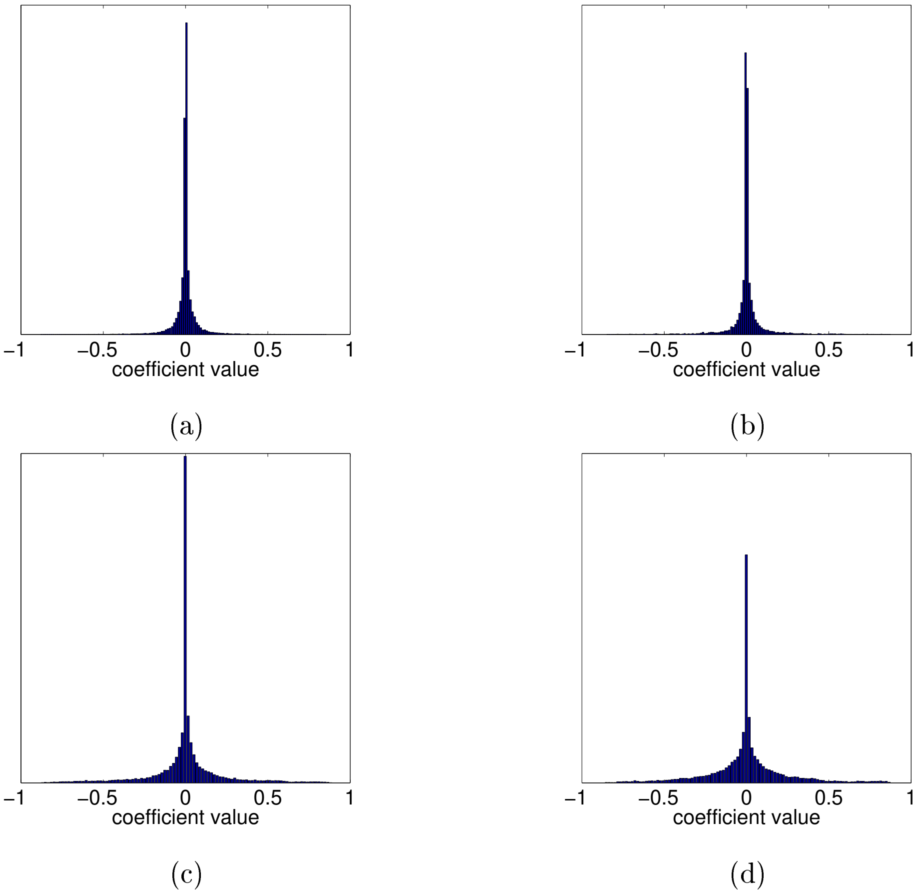


Fig. 3. *Histogram of perturbation variates ($\delta = \theta/\lambda$) for scale (a) $j = 1$, (b) $j = 2$, (c) $j = 3$, and (d) $j = 4$ of the cameraman image of Figure 8(a). The general invariance of the distributions' structure across scales illustrates a self-similar property of real-world image statistics.*

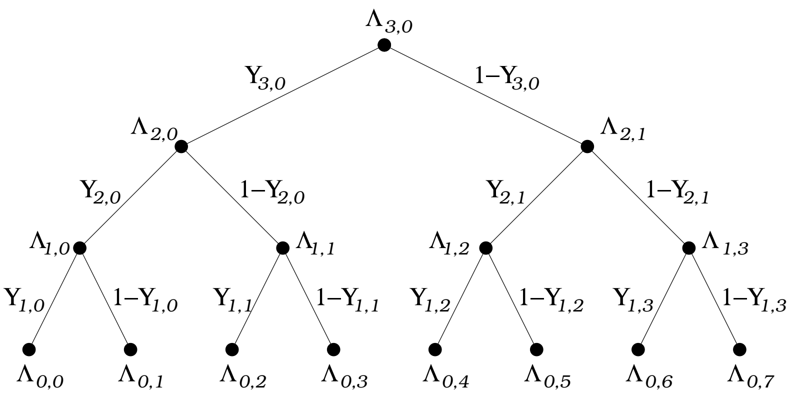


Fig. 4. *MMI model interpreted as a probabilistic tree. The MMI model can be viewed as a tree-structured probability model in which the intensity $\Lambda_{j,k}$ at coarse scale j is refined (split) via the multiplicative innovation $Y_{j,k}$ to obtain two new intensities $\Lambda_{j-1,2k}$ and $\Lambda_{j-1,2k+1}$ at the next finer scale of analysis $j-1$. The innovations variates $\{Y_{j,k}\}$ are mutually independent at all scales j and positions k.*

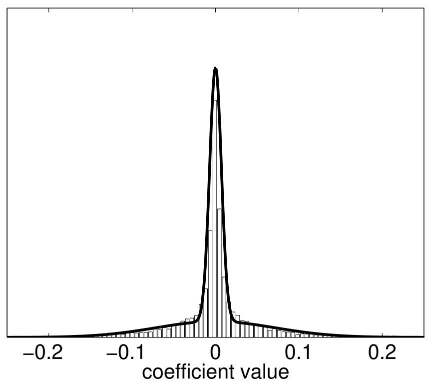


Fig. 5. *Three component Beta-mixture distribution (solid line) superimposed on the histogram of the perturbation variates ($\delta = \theta/\lambda$) of Figure 8(a). The beta mixture parameters here are $s_1 = 1$, $s_2 = 100$, $s_3 = 10000$, $p_1 = 0.001$, $p_2 = 0.400$, and $p_3 = 0.599$.*

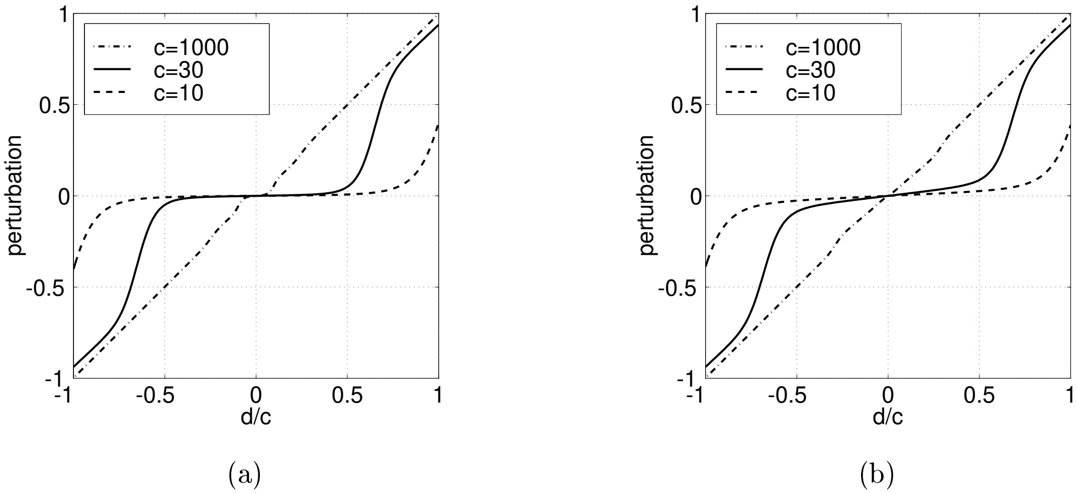


Fig. 6. *Perturbation estimate $\widehat{\delta}$ as a function of d/c for $c = 10$ (dash), $c = 30$ (solid), and $c = 1000$ (dot-dash). The estimator's defining parameters are $s_1 = 1$, $s_2 = 100$, $s_3 = 10000$, $p_1 = 0.01$, and (a) $p_2 = 1 - p_1 - p_3 = 0.100$, and (b) $p_2 = 1 - p_1 - p_3 = 0.900$.*

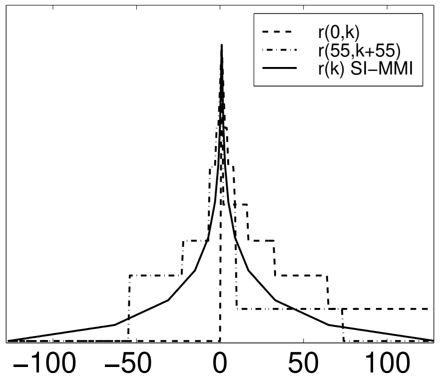
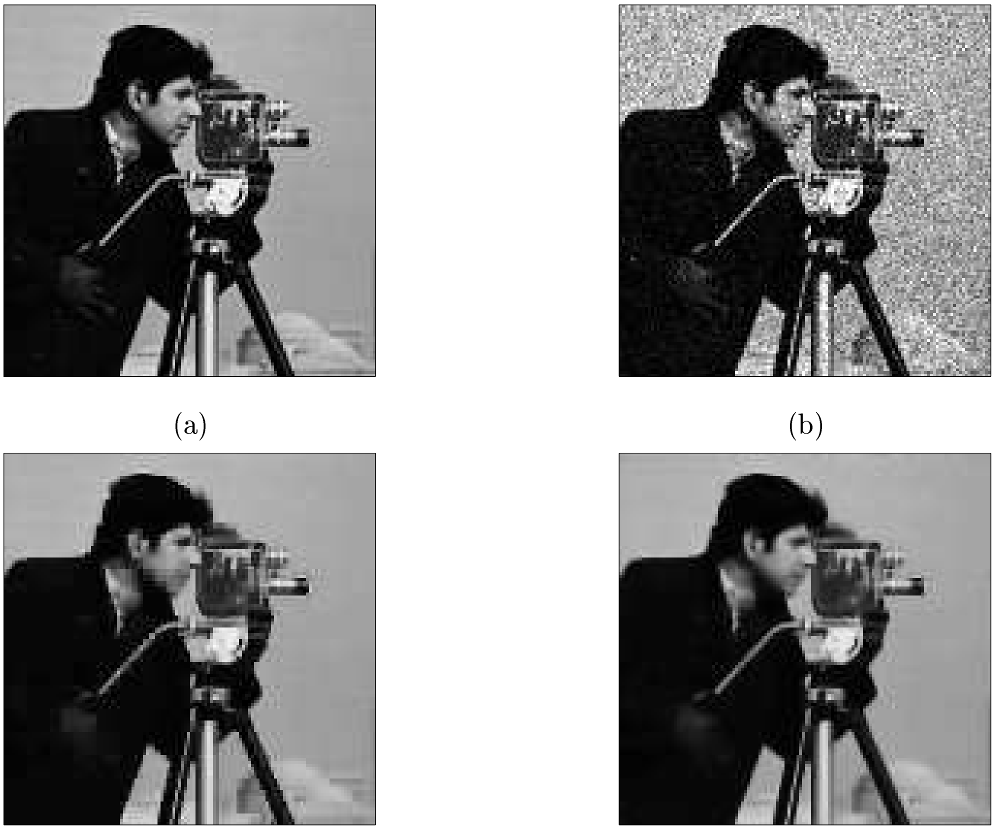


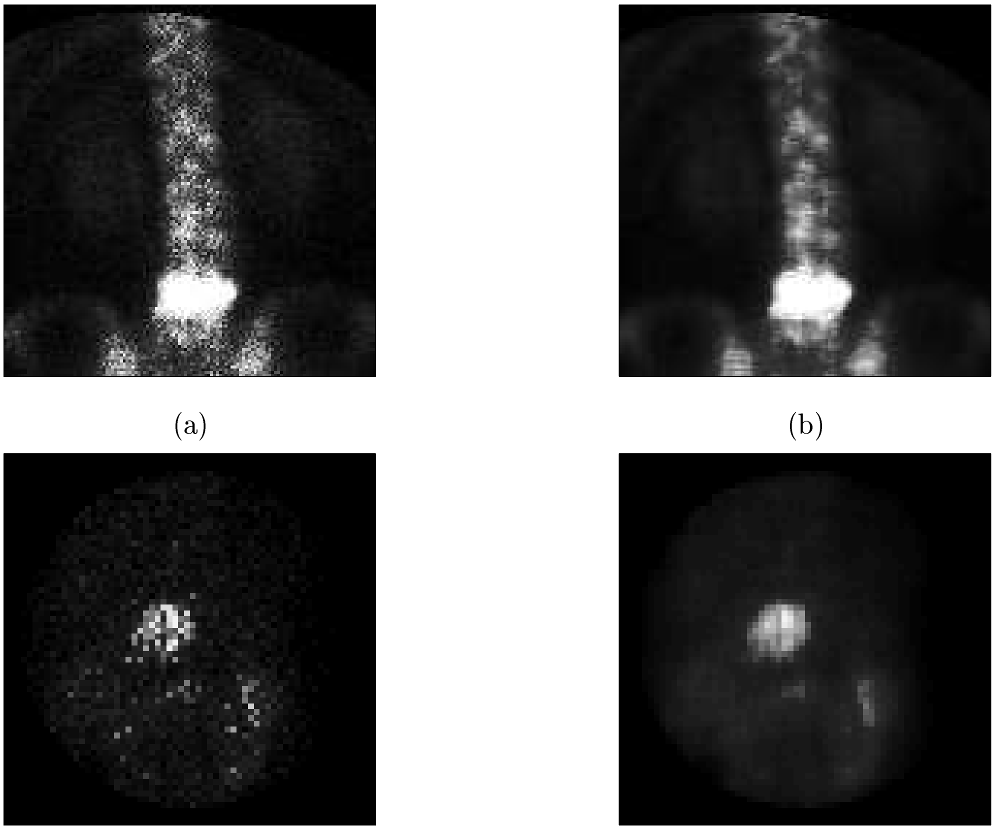
Fig. 7. *Correlations functions for MMI and SI-MMI 256-point ($J = 8$) intensity priors. MMI model autocorrelation $r(0,k)$ (dash-dash) and $r(55,k+55)$ (dash-dot) plotted as a function of k. Stationary SI-MMI model autocorrelation $r(k)$ (solid). For both the MMI and SI-MMI models in this example, $\mu_J^2 = 1$ and $\rho_j^2 = 0.26,\ j = 1,\ldots,J$.*



(a) (b)

(c) (d)

Fig. 8. *Photon-limited image estimation using MMI models. (a) intensity function, (b) realization of Poisson counts (average squared pixel error = 24.80), (c) intensity estimate using MMI model (average squared pixel error = 7.36), (d) intensity estimate using SI-MMI model (average squared pixel error = 3.98).*



(a) (b) (c) (d)

Fig. 9. *Nuclear medicine image estimation using SI-MMI models. (a) Spine count image. (b) SI-MMI model estimate of underlying intensity. (c) Heart count image. (d) SI-MMI model estimate.*