International Statistical Review (2004), **72,** 2, 209–237, Printed in The Netherlands

A Comparative Simulation Study of Wavelet Shrinkage Estimators for Poisson Counts

Panagiotis Besbeas[1], Italia De Feis[2] and Theofanis Sapatinas[3]

[1]*University of Kent at Canterbury, Canterbury, Kent CT2 7NF, UK. E-mail: P.T.Besbeas@ukc.ac.uk*
[2]*Consiglio Nazionale delle Ricerche, 80131 Napoli, Italy. E-mail: defeis@iam.na.cnr.it*
[3]*University of Cyprus, CY 1678 Nicosia, Cyprus. E-mail: T.Sapatinas@ucy.ac.cy*

Summary

Using computer simulations, the finite sample performance of a number of classical and Bayesian wavelet shrinkage estimators for Poisson counts is examined. For the purpose of comparison, a variety of intensity functions, background intensity levels, sample sizes, primary resolution levels, wavelet filters and performance criteria are employed. A demonstration is given of the use of some of the estimators to analyse a data set arising in high-energy astrophysics. Following the philosophy of reproducible research, the MATLAB **programs and real-life data example used in this study are made freely available.**

Key words: Bayesian inference; Gamma-ray bursts; Monte Carlo experiments; Multiscale analysis; Nonparametric regression; Poisson processes; Wavelets.

1 Introduction

Let $N(0, t]$ be an inhomogeneous Poisson process on the interval $[0, 1]$ with intensity function $\mu(t) > 0$ and suppose that we actually observe a discrete-time version of the process, a Poisson counting process. That is, the observation interval $[0, 1]$ is divided into n equal subintervals and we observe y_i, $i = 0, 1, \ldots, n-1$, the number of events occurring in each subinterval. The observed data $\mathbf{y} = (y_0, y_1, \ldots, y_{n-1})'$ can thus be expressed as n independent Poisson random variables, i.e.,

$$y_i \sim \text{Poisson}(\mu_i), \quad i = 0, 1, \ldots, n-1, \tag{1}$$

where the mean vector $\boldsymbol{\mu} = (\mu_0, \mu_1, \ldots, \mu_{n-1})'$ is unknown. This model is called the Poisson regression model in the literature and the aim here is to recover the underlying intensity function $\mu(t)$ from the data $\mathbf{y}$ without assuming any particular parametric structure for it.

Amongst the available nonparametric approaches in the literature to tackle the aforementioned problem, the kernel smoothing method proposed by Diggle (1985) and Diggle & Marron (1988), the penalized likelihood method proposed by O'Sullivan, Yandell & Raynor (1986), and the P-splines method proposed by Eilers & Marx (1996) have several theoretical merits and have been successfully applied to real-life data sets. The first method constructs a boundary adjustment kernel smoother for $\mu(t)$ using a quartic kernel; the second method penalizes the negative logarithm of the likelihood of the logarithm of $\mu(t)$ with a smoothness penalty based on the integral of the square of the second derivative of the logarithm of $\mu(t)$; and the third method penalizes the regression of the logarithm of $\mu(t)$ using a set of B-splines with a smoothness penalty based on a higher-order difference of the coefficients of adjacent B-splines. Although the smoothing kernel-based method has been obtained directly using the Poisson regression model (1), both spline-based methods have been proposed as

general methodologies for nonparametric regression estimation in generalised linear models, Poisson regression being a particular statistical model within this family of models. Moreover, asymptotic results and finite-sample properties of the above methods have been studied under the assumption that the underlying intensity function $\mu(t)$ obeys a smooth behaviour.

On the other hand, wavelet methods have a long history in the nonparametric estimation of spatially-variable objects with applications to diverse disciplines including astronomy, biology, electronics, engineering, medicine and physics. The potential of wavelets in the classical nonparametric regression context, where the data are modelled as observations of a response function contaminated by additive Gaussian noise, was convincingly demonstrated by Donoho, Johnstone, Kerkyacharian & Picard (1995). Their work showed that a simple nonlinear estimator obtained by thresholding the empirical wavelet coefficients is essentially optimal (in the minimax sense) over a broad range of function classes. Furthermore the estimation procedure is computationally efficient, making it very appealing in practice.

However the potential scope of wavelet methods is broader than just the classical nonparametric regression setting. In the case where the data are Poisson counts, various techniques have been recently proposed, including the use of a variance-stabilising transformation with a normal approximation to facilitate wavelet shrinkage. Donoho (1993) and Fryźlewicz & Nason (2001) are examples of this approach, employing the Anscombe (1948) and the Fisz (1955) transformation respectively. Alternatively, Kolaczyk (1997, 1999a) and Nowak & Baraniuk (1999) focused on adapting wavelet shrinkage to the original Poisson counts, whilst Kolaczyk (1999b) and Timmermann & Nowak (1999) developed another approach based on Bayesian inference.

Wavelet shrinkage techniques have also been proposed for using wavelets when the data are coming from some general families of distributions, Poisson counts being a particular type of data obtained within these families of distributions. Using ideas by Beran & Dümbgen (1998), Antoniadis & Sapatinas (2001) and Antoniadis, Besbeas & Sapatinas (2001) presented a wavelet shrinkage methodology for natural exponential families with quadratic and cubic variance functions respectively. These families encompass some very famous distributions including the Gaussian, Poisson, gamma, binomial, negative binomial and inverse Gaussian. On the other hand, Sardy, Antoniadis & Tseng (2004) considered an l_1-penalised likelihood method to develop a wavelet shrinkage methodology for Gaussian, exponential, Poisson and Bernoulli distributions.

While there are merits to each of the above approaches, the most common types of data in the nonparametric regression context are Gaussian and Poissonian. Surprisingly, and in contrast to the Gaussian case (see Marron, Adak, Johnstone, Neumann & Patil, 1998; Antoniadis, Bigot & Sapatinas, 2001), there is little coherence in the literature about the relative performance of the different wavelet schemes for Poisson data. Furthermore this leaves a number of open questions regarding their small sample properties, particularly relative to each other. The aim of this article is, therefore, to compare most of the currently available wavelet shrinkage methods for estimating the underlying intensity function $\mu(t)$ based on observations from the Poisson regression model (1). In our comparison we have included the classical methods of Donoho (1993), Kolaczyk (1997, 1999a), Antoniadis & Sapatinas (2001), Fryźlewicz & Nason (2001) and Sardy, Antoniadis & Tseng (2004), and the Bayesian methods of Kolaczyk (1999b) and Timmermann & Nowak (1999). We proceed by simulation and employ a variety of intensity functions, background intensity levels, sample sizes, primary resolution levels, wavelet filters and performance criteria; insight about the performance of these estimators is obtained from graphical outputs and numerical tables.

The rest of this article is organised as follows. In Section 2 we briefly review some general results on wavelets, including the discrete wavelet transform. Sections 3 and 4 review respectively the classical and Bayesian wavelet shrinkage methods for Poisson data that we employ in the simulation study, which is described in Section 5. Section 6 presents simulation results and some discussions, and Section 7 contains the main conclusions. Finally, in Section 8 and the Appendix we demonstrate,

by application to an astronomical gamma-ray burst data set, how the (MATLAB) programs used for the calculations of this article may be used in practice. Following the philosophy of reproducible research (see Buckheit & Donoho, 1995), the programs and the gamma-ray data set can be found on the World Wide Web at `http://www.ucy.ac.cy/~fanis/links/software.html`.

2 Wavelet Background

In this section we provide some general material which is at the heart of the wavelet paradigm. Further details on wavelet theory can be found in, for example, Daubechies (1992), Meyer (1992) and Mallat (1999).

A wavelet is a function that, roughly speaking, looks like a localised wiggle. We define a collection of wavelets—called a wavelet basis—by dilating and translating two basic functions, a 'father' wavelet ϕ and a 'mother' wavelet ψ. The wavelets ϕ and ψ are assumed to be compactly supported and are called r-regular if they have r vanishing moments and r continuous derivatives.

A wavelet basis has an associated r-regular multiresolution analysis on $\mathbb{R}$. For simplicity in exposition, we shall focus exclusively on periodised wavelet bases on $[0, 1]$, letting

$$\phi^{\mathrm{p}}_{jk}(t) = \sum_{l\in\mathbb{Z}} \phi_{jk}(t-l) \quad \text{and} \quad \psi^{\mathrm{p}}_{jk}(t) = \sum_{l\in\mathbb{Z}} \psi_{jk}(t-l),$$

where

$$\phi_{jk}(t) = 2^{j/2}\phi(2^j t - k) \quad \text{and} \quad \psi_{jk}(t) = 2^{j/2}\psi(2^j t - k)$$

for $j, k \in \mathbb{Z}$, the set of integers. For any primary resolution level $j_0 \geq 0$, the collection $\{\phi^{\mathrm{p}}_{j_0k},\ k = 0, 1, \ldots, 2^{j_0}-1;\ \psi^{\mathrm{p}}_{jk},\ j \geq j_0,\ k = 0, 1, \ldots, 2^j-1\}$ constitutes an orthonormal basis of $L^2([0, 1])$. The superscript "p" will be suppressed from the notation, for typographical convenience.

The idea underlying the wavelet approach is that a broad class of functions can be arbitrarily well approximated by a wavelet series; i.e., for any function $g(t) \in L^2([0, 1])$

$$g(t) = \sum_{k=0}^{2^{j_0}-1} \alpha_{j_0k}\phi_{j_0k}(t) + \sum_{j=j_0}^{\infty}\sum_{k=0}^{2^j-1} \beta_{jk}\psi_{jk}(t),$$

where

$$\alpha_{j_0k} = \int_0^1 g(t)\phi_{j_0k}(t)\,dt, \quad j_0 \geq 0, \quad k = 0, 1, \ldots, 2^{j_0}-1,$$
$$\beta_{jk} = \int_0^1 g(t)\psi_{jk}(t)\,dt, \quad j \geq j_0 \geq 0, \quad k = 0, 1, \ldots, 2^j-1.$$

The coefficients α_{j_0k} and β_{jk} are called respectively the scaling and wavelet coefficients of $g(t)$.

The continuous formulation of the wavelet transform is however not very useful for the discretely sampled functions that are common in practice. For these functions it is the wavelet analog of the discrete Fourier transform that is applicable and this is referred to as the discrete wavelet transform (DWT). In particular, given a vector of function values $\mathbf{g} = (g(t_1), \ldots, g(t_n))'$ at equally spaced points t_i, the discrete wavelet transform takes the form of an orthogonal matrix W that carries $\mathbf{g}$ to its discrete wavelet coefficients

$$\mathbf{d} = W\mathbf{g}.$$

Here, the coefficients are stored as an $n \times 1$ vector consisting of both the discrete scaling coefficients, c_{j_0k}, and the discrete wavelet coefficients, d_{jk}. These coefficients are only approximately their continuous counterparts, α_{j_0k} and β_{jk}. The precise relationship between them is discussed in any standard reference; see, for example, Abramovich, Bailey & Sapatinas (2000). By the orthogonality of W, the transpose matrix W' inverts the transform and returns the original input. This process is called

reconstruction and the reconstruction algorithm is called the inverse discrete wavelet transform (IDWT). A crucial point is that both the DWT and the IDWT are not implemented by matrix multiplication, but by a sequence of filtering operations which produce an order $O(n)$ algorithm. This algorithm requires $n = 2^J$, for some positive integer J, and we will implicitly assume this in the sequel. See Mallat (1989) for further details.

3 Classical Wavelet Approaches for Poisson Data

Consider now the Poisson regression model (1) and assume that the underlying intensity function $\mu(t)$ admits an orthonormal wavelet expansion with scaling and wavelet coefficients c_{j_0k} and d_{jk} respectively. Let also $\hat{c}_{j_0k}$ and $\hat{d}_{jk}$ be their empirical counterparts from the data vector $\mathbf{y} = (y_0, y_1, \ldots, y_{n-1})'$, where $n = 2^J$, for some positive integer J. The sparseness of the wavelet expansion implies that most of the structure in $\mu(t)$ is concentrated in relatively few 'large' d_{jk}. We would thus obtain a reasonable estimate of $\mu(t)$ if we could extract the wavelet coefficients of largest magnitude accurately (by modifying appropriately the $\hat{d}_{jk}$), even when we set the rest to zero. Note that, in general, the scaling coefficients c_{j_0k} are not altered (they are estimated by $\hat{c}_{j_0k}$) because the primary resolution level j_0 represents 'low-frequency' terms that usually contain important components of the intensity function.

In parallel to the classical nonparametric regression setting, a simple nonlinear wavelet estimator of $\mu(t)$ can be constructed by thresholding the empirical wavelet coefficients $\hat{d}_{jk}$. We can allow for a broad class of thresholding schemes but common choices include hard thresholding $[\delta_\tau(x) = x1_{\{|x|>\tau\}}]$ and soft thresholding $[\delta_\tau(x) = sgn(x)(|x| - \tau)_+]$ for a fixed threshold $\tau > 0$, where 1_A is the indicator function of A, $sgn(x)$ is the signum function of x and $(x)_+ = \max(0, x)$. Hard thresholding is thus a 'keep' or 'kill' rule, while soft thresholding is a 'shrink' or 'kill' rule. In either case, application of the appropriate inverse transform to the empirical coefficients $\{\hat{c}_{j_0k},\ \delta_\tau(\hat{d}_{jk})\}$ yields a denoised estimate of the underlying intensity function $\mu(t)$.

From both practical and theoretical standpoints, the success of this approach is closely related to the properties of the noise, assumed to be Gaussian, additive and stationary. The noise in (1) shares none of these characteristics and, as a result, direct application of the methodology is clearly inappropriate. The remainder of this section is devoted to the classical wavelet-based approaches that have been proposed in the literature for estimating the underlying intensity function $\mu(t)$ based on observations from the Poisson regression model (1). The methods we describe are due to Donoho (1993), Kolaczyk (1997, 1999a), Antoniadis & Sapatinas (2001), Fryźlewicz & Nason (2001) and Sardy, Antoniadis & Tseng (2004), and they are used in the simulation study in Sections 5 and 6. We refer to these articles for more details on these methods.

3.1 Wavelet Shrinkage of General Intensity Functions Using Transformations

A straightforward approach that was initially proposed by Donoho (1993) involves preprocessing the data using a normalising and variance-stabilising transformation. Donoho (1993) employed the Anscombe (1948) transformation, whilst recently Fryźlewicz & Nason (2001) advocated the use of the Fisz (1955) transformation. Following this preprocessing the usual wavelet methodology with a global threshold τ can be applied as if the noise was actually Gaussian. Application of the inverse transformation leads to an estimate of the underlying intensity function $\mu(t)$. The effectiveness of the approach using either the Anscombe or the Fisz transformation, followed by one of the minimax (Donoho & Johnstone, 1994; Bruce & Gao, 1996), universal (Donoho & Johnstone, 1994) or 'leave-out-half' cross-validation (Nason, 1996) threshold, is evaluated in the simulation study in Sections 5 and 6. The computational complexity of this approach for the Anscombe or the Fisz transformation, with either the minimax or the universal threshold, is $O(n)$. On the other hand, its computational

complexity with the 'leave-out-half' cross-validation threshold is $O(n^2)$, although an $O(n \log n)$ algorithm is possible if hard thresholding is used (see Hurvich & Tsai, 1998).

Finally, making use of the translation invariant methodology of Coifman & Donoho (1995), we have also implemented the translation invariant versions of these combinations. The result of this process is to suppress many of the artifacts frequently found in standard wavelet shrinkage estimates that are the result of pseudo-Gibbs phenomena, and to correct unfortunate misalignments between features (of interest) in the intensity function $\mu(t)$ and features in the wavelet basis. The translation invariant algorithm itself requires $O(n \log n)$ and $O(n^2)$ operations for the Anscombe and the Fisz transformation respectively, and the effectiveness of this approach is also evaluated in the simulation study in Sections 5 and 6.

To make the exposition fairly well self-contained, however, we briefly discuss below the idea behind the minimax, universal and 'leave-out-half' cross-validation thresholds. Hereafter, we assume that the original Poisson count data $\mathbf{y} = (y_0, \ldots, y_{n-1})'$ from model (1) have been preprocessed, using either the Anscombe or the Fisz transformation, implying that the resulting transformed vector of observations $\mathbf{y}^\star = (y_0^\star, \ldots, y_{n-1}^\star)'$ follows a multivariate Gaussian distribution with mean $\mathbf{g}$ (the vector of underlying response function values) and variance equal to I_n. By a slight abuse of notation, we again use the generic d and $\hat{d}$ to denote the wavelet and empirical wavelet coefficients of the underlying response function $g(t)$ and the transformed data $\mathbf{y}^\star$ respectively.

The Minimax Threshold

An optimal threshold, derived to minimize the constant term in an upper bound of the risk involved in estimating the underlying response function $g(t)$, was obtained by Donoho & Johnstone (1994). The proposed minimax threshold, that depends on the sample size n, is defined as

$$\tau^{\mathrm{M}} = \tau_n^\star,$$

where $\tau_n^\star$ is defined as the value of τ which achieves

$$\Lambda_n^\star = \inf_\tau \sup_d \left\{ \frac{R_\tau(d)}{n^{-1} + R_{\mathrm{oracle}}(d)} \right\}, \tag{2}$$

where $R_\tau(d) = E(\delta_\tau(\hat{d}) - d)^2$ and $R_{\mathrm{oracle}}(d)$ is the ideal risk achieved with the help of an oracle.

Two oracles were considered by Donoho & Johnstone (1994): diagonal linear projection (DLP), an oracle which tells us when to 'keep' or 'kill' each empirical wavelet coefficient, and diagonal linear shrinker (DLS), an oracle which tells you how much to shrink each wavelet coefficient. The ideal risks for these oracles are given by

$$R_{\mathrm{oracle}}^{\mathrm{DLP}}(d) = \min(d^2, 1) \quad \text{and} \quad R_{\mathrm{oracle}}^{\mathrm{DLS}}(d) = \frac{d^2}{d^2 + 1}.$$

Donoho & Johnstone (1994) computed the DLP minimax thresholds for the soft thresholding rule, while the DLP minimax thresholds for the hard thresholding rule and the DLS minimax thresholds for both soft and hard thresholding rules were obtained by Bruce & Gao (1996).

Table 1

Diagonal linear projection minimax thresholds for hard and soft thresholding rules for various sample sizes.

n	128	256	512	1024
Hard	2.913	3.117	3.312	3.497
Soft	1.669	1.859	2.045	2.226

Since the type of the oracle has little impact on the minimax thresholds, Table 1 only reports the numerical values of the DLP minimax thresholds (for both hard and soft thresholding rules) for the sample sizes that we will be using in the simulative study in Sections 5 and 6, and can be used as a look-up table in any software. These values were computed using a grid search over τ with increments $\Delta_\tau = 0.0001$. At each point, the supremum over d in (2) was computed using a quasi-Newton optimisation with numerical derivatives (see, for example, Dennis & Mei, 1979).

The Universal Threshold

As an alternative to the use of minimax thresholds, Donoho & Johnstone (1994) suggested thresholding each of empirical wavelet coefficients $\hat{d}$ by using the universal threshold

$$\tau^{\mathrm{U}} = \sqrt{2 \log n}. \tag{3}$$

This threshold is easy to remember and its implementation in software requires no costly development of look-up tables. The universal threshold ensures, with high probability, that every sample in the wavelet transform in which the underlying function is exactly zero will be estimated as zero. This is so, because if $X_1, \ldots, X_n$ are independent and identically distributed standard Gaussian random variables, then

$$P\left\{\max_{1 \le i \le n} |X_i| \le \sqrt{2 \log n}\right\} \to 1, \quad \text{as} \quad n \to \infty.$$

The rate at which the probability above tends to one is actually quite slow. Faster rates, achieved by larger thresholds, are possible; however, they lead to oversmoothing with wavelet shrinkage.

The 'Leave-out-half' Cross-Validation Threshold

One way to choose the threshold level τ is by minimising the mean integrated squared error between a wavelet threshold estimator $\hat{g}_\tau(t)$ and the true function $g(t)$. In symbols, the threshold τ should minimise

$$M(\tau) = \mathbb{E} \int \left(\hat{g}_\tau(t) - g(t)\right)^2 \, dt. \tag{4}$$

In practice, the function $g(t)$ is unknown and so an estimate of M is required. The approach to cross-validation in wavelet regression was adopted by Nason (1996) who, in order to directly apply the DWT, suggested breaking the original data set into two subsets of equal size: one containing only the even-indexed data, and the other, the odd-indexed data. The odd-indexed data will be used to 'predict' the even-indexed data, and vice-versa, leading to a 'leave-out-half' strategy.

To be more specific, given the transformed Gaussian data $\mathbf{y}^\star = (y_0^\star, \ldots, y_{n-1}^\star)'$ obtained from model (1) with $n = 2^J$, remove all the odd-indexed $y_i^\star$ from the set. This leaves 2^{J-1} evenly indexed $y_i^\star$ which are re-indexed from $j = 1, \ldots, 2^{J-1}$. These re-indexed data are then used to construct a function estimate $\hat{g}_\tau^{\mathrm{E}}$ by using a particular threshold parameter τ with either hard thresholding or soft thresholding. To compare the function estimator with the left-out noisy data an interpolated version of $\hat{g}_\tau^{\mathrm{E}}$ is formed

$$\bar{g}_{\tau,j}^{\mathrm{E}} = \frac{1}{2}\left(\hat{g}_{\tau,j+1}^{\mathrm{E}} + \hat{g}_{\tau,j}^{\mathrm{E}}\right), \quad j = 1, \ldots, n/2,$$

setting $\hat{g}_{\tau,n/2+1}^{\mathrm{E}} = \hat{g}_{\tau,1}^{\mathrm{E}}$ because g is assumed to be periodic. The $\bar{g}_\tau^{\mathrm{O}}$ is computed for the odd-indexed points and the interpolant is, similarly, formed as

$$\bar{g}_{\tau,j}^{\mathrm{O}} = \frac{1}{2}\left(\hat{g}_{\tau,j+1}^{\mathrm{O}} + \hat{g}_{\tau,j}^{\mathrm{O}}\right), \quad j = 1, \ldots, n/2.$$

The full estimate of M given in (4) compares the interpolated wavelet estimators and the left out

points

$$\hat{M}(\tau) = \sum_{j=1}^{n/2} \left[\left(\bar{g}^{\mathrm{E}}_{\tau,j} - y_{2j+1} \right)^2 + \left(\bar{g}^{\mathrm{O}}_{\tau,j} - y_{2j} \right)^2 \right]. \tag{5}$$

It can be shown that one can almost always find a unique minimum of (5)

$$\tau_{\min} = \arg\min_{\tau \geq 0} \hat{M}(\tau).$$

This minimum value depends on $n/2$ data points (since both estimates of g, $\hat{g}^{\mathrm{E}}_{\tau}$ and $\bar{g}^{\mathrm{O}}_{\tau}$ are based on $n/2$ data points) and, therefore, a correction for the sample size is needed. Nason (1996) considered the universal threshold τ^{U} given in (3) to supply a heuristic method for obtaining a cross-validated threshold for n data points. By using this adjustment, the 'leave-out-half' cross-validation threshold is defined as

$$\tau^{\mathrm{CV}} = \left(1 - \frac{\log 2}{\log n} \right)^{-1/2} \tau_{\min}.$$

3.2 Wavelet Shrinkage of General Intensity Functions Using Modulation Estimators

A different approach with a wider scope of application was considered by Antoniadis & Sapatinas (2001). In this approach, one can construct estimators of the underlying intensity function $\mu(t)$ by diagonal linear shrinkages that are asymptotically minimax for a class of submodels for $\mu(t)$, namely the class of functions belonging to an ellipsoid of the Sobolev class W^s_2 of smoothness index $s > 1/2$. These estimators take the form $\hat{H}W\mathbf{y}$, where $\hat{H} = \mathrm{diag}(\hat{\mathbf{h}})$ is the diagonal matrix of order n and $\hat{\mathbf{h}} : T = \{0, 1, \ldots, n-1\} \to [0, 1]$ (depends on $\hat{\mathbf{d}} = W\mathbf{y}$) is chosen to minimize the estimated risk of the linear estimator $\hat{\mathbf{d}}_{\mathbf{h}} = HW\mathbf{y}$ over all functions $\mathbf{h}$ in a class $\mathcal{H} \subset [0, 1]^T$. Each function $\mathbf{h}$ in a class $\mathcal{H} \subset [0, 1]^T$ is called a modulator and the resulting estimator $\hat{\mathbf{d}}_{\hat{\mathbf{h}}} = \hat{H}W\mathbf{Y}$ is referred to as the modulation estimator.

Let $\mathcal{B}_n$ be a partition of the set T and define

$$\mathcal{H} = \left\{ \sum_{B \in \mathcal{B}_n} 1_B c(B) : \; c \in [0, 1]^{\mathcal{B}_n} \right\}, \tag{6}$$

where 1_B is the indicator function of B. Extending the arguments in Beran & Dümbgen (1998), Antoniadis & Sapatinas (2001) put forth the modulation estimator

$$\hat{\mathbf{d}}_{\hat{\mathbf{h}}} = \sum_{B \in \mathcal{B}_n} \frac{\mathrm{ave}(1_B(\hat{\mathbf{d}}^2 - \tilde{\boldsymbol{\sigma}}^2))_+}{\mathrm{ave}(1_B \hat{\mathbf{d}}^2)} \, 1_B \hat{\mathbf{d}}$$

of $\mathbf{d} = W\boldsymbol{\mu}$, where $\mathrm{ave}(\mathbf{g}) = n^{-1} \sum_{t \in T} g(t)$ for any function $\mathbf{g} \in \mathbb{R}^n$, $\tilde{\boldsymbol{\sigma}}^2(t) = \sum_{l=0}^{n-1} w_{tl}^2 y_l$ ($t \in T$), and the wavelet-filter coefficients w_{tl} in $\tilde{\boldsymbol{\sigma}}^2(t)$ are associated with the DWT matrix W. By performing the IDWT of $\hat{\mathbf{d}}_{\hat{\mathbf{h}}}$, the resulting modulation estimator can be used to estimate the underlying intensity function $\mu(t)$. This approach requires $O(n)$ operations, whilst its translation invariant version requires $O(n^2)$ operations.

Various examples of modulator classes $\mathcal{H}$ in (6) can now be constructed. The optimal modulation estimator is a multiple Stein estimator and results when the partition $\mathcal{B}_n$ has cardinality $|\mathcal{B}_n| = o(n)$ (see, Beran & Dümbgen, 1998). One such choice is, for example, when $\mathcal{B}_n$ is a partition of T in intervals of length $[\log_e(\log_e n)]$, where $[x]$ denotes the integer part of x. However, for practical purposes, in the simulation study in Sections 5 and 6 we have taken $\mathcal{B}_n$ to be a partition of T in intervals of length one, resulting in a pointwise modulation estimator (see Antoniadis & Sapatinas,

2001) that is related to the estimator considered by Nowak & Baraniuk (1999).

3.3 *Wavelet Shrinkage of General Intensity Functions Using an l_1-Penalised Likelihood Method*

Another approach with a wider scope of application was considered by Sardy, Antoniadis & Tseng (2004). In this approach, one can construct estimators of the underlying intensity function $\mu(t)$ by considering an l_1-penalised likelihood method. In other words, the resulting estimator is obtained as the solution of the following convex programming problem

$$\min_{\boldsymbol{\mu},\mathbf{d}} \left\{ -l(\boldsymbol{\mu}; \mathbf{y}) + \lambda \sum_{j=j_0}^{J-1} \sum_{k=0}^{2^j-1} |d_{jk}| \quad \text{with} \quad \boldsymbol{\mu} = W'\mathbf{d} \quad \text{and} \quad \boldsymbol{\mu} \in (0,\infty)^n \right\}, \tag{7}$$

where $-l(\boldsymbol{\mu}; \mathbf{y})$ is the log-likelihood function of $\boldsymbol{\mu}$ based on the observations $\mathbf{y}$ from model (1) and $\lambda > 0$ is the smoothing parameter.

The above convex programming problem is a reminiscent of the Basis Pursuit estimator of Chen, Donoho & Saunders (1999) obtained by an interior point method (based on non-orthonormal wavelet bases) for Gaussian data. However, (7) is more complex in that it has inequality constraints, the l_1-penalty only applies to the wavelet coefficients d_{jk} ($j = j_0, \ldots, J-1$), and the log-likelihood is not quadratic. Moreover, special care must also be taken to handle the case of zero Poisson counts. Sardy, Antoniadis & Tseng (2004) solved the above convex programming problem (7) by first deriving its dual problem and then developing a primal-dual log-barrier internal point method. Furthermore, using results from inhomogeneous continuous-time Poisson processes, they proved that the appropriate smoothing parameter λ in (7) is actually level-dependent and equal to

$$\lambda_j = M(\mu_\phi, \psi) 2^{j/2} \sqrt{2 \log n} / \sqrt{n}, \quad j = j_0, \ldots, J-1,$$

where

$$M(\mu_\phi, \psi) := \max_{u \in [0,1]} \{\psi^2(u)\} \int_0^1 \mu_\phi^{-1}(t)\, dt.$$

The above universal-type thresholds are then used to estimate the underlying intensity function $\mu(t)$. The constant $M(\mu_\phi, \psi)$ depends both on the mother wavelet ψ and the knowledge of μ_ϕ which represents the situation where the underlying intensity function is in the range of the scaling functions (i.e., all fine-scale wavelet coefficients are zero). In practice, Sardy, Antoniadis & Tseng (2004) suggested to estimate μ_ϕ in the range of the scaling functions using the estimator proposed by Donoho, Johnstone, Kerkyacharian & Picard (1995) for Poisson data, based on the variance-stabilising transformation of Anscombe (1948).

As it was pointed out in Sardy, Antoniadis & Tseng (2004), the computational complexity of this approach is difficult to access. The resulting estimator is based on an iterative algorithm solving a set of non-linear equations by a Newton-type method, and for the Newton direction a system of linear equations is solved by a variant of the conjugate gradient algorithm. Therefore, the method is computationally intensive, but some hints of how to speed up the conjugate gradient algorithm are given in Sardy, Antoniadis & Tseng (2004). The relative performance of this approach is evaluated in the simulation study in Sections 5 and 6.

3.4 *Wavelet Shrinkage of Burst-like Intensity Functions Using Haar Thresholds*

An approach for burst-like Poisson processes has been described in Kolaczyk (1997). A burst-like Poisson process is defined as the sum of a homogeneous 'background' Poisson process (with constant intensity function $\mu_0 > 0$) and a second, inhomogeneous Poisson process (with intensity function $\tilde{\mu}(t) \geq 0$) that tends to generate observations in bursts.

The approach considered by Kolaczyk (1997) constructs estimators of the underlying intensity function $\mu(t)$ using Haar wavelets and applying suitable level-dependent thresholds τ_j to the resulting empirical wavelet coefficients $\hat{d}_{jk}$ of the original Poisson counts, that are similarly mild in nature to the universal threshold of Donoho & Johnstone (1994). More specifically, under the null hypothesis $H_0: \mu(t) \equiv \mu_0$ (i.e, no burst being present), the resulting empirical wavelet coefficients $\hat{d}_{jk}$, at each resolution level, are independent and identically distributed according to a symmetric distribution about zero with variance μ_0, and we have that

$$P\left(\max_{0\le k\le 2^j-1} |\hat{d}_{jk}| \le \tau_j\right) = \left[1 - 2\ P\left(\chi^2_{(2m_j)}(\nu_j) < \nu_j\right)\right]^{2^j},$$

where $\nu_j = 2^{(J-j)}\mu_0$ and $m_j = 2^{(J-j)/2}\tau_j$. By using an approximation for non-central chi-square random variables due to Patnaik (1949) and appealing to the central limit theorem, Kolaczyk (1997) was able to derive the following level-dependent thresholds

$$\tau_j = 2^{-(J-j)/2}\left\{\log(2^j) + \sqrt{\log^2(2^j) + 2\mu_0\log(2^j)2^{J-j}}\ \right\}, \quad j = j_0, \ldots, J-1,$$

which are then used to estimate the underlying intensity function $\mu(t)$. The dependence of the thresholds on the unknown true value of μ_0 has practical implications and we come back to this issue in Section 6. For fixed μ_0, the method is implemented within the translation invariant framework of Coifman & Donoho (1995) to eliminate the 'staircase-like' structure of non-translation invariant Haar estimates. As a result, the computational complexity of the resulting approach is $O(n \log n)$. The relative performance of this approach is evaluated in the simulation study in Sections 5 and 6.

3.5 Wavelet Shrinkage of Burst-like Intensity Functions Using Corrected Thresholds

Kolaczyk (1999a) attempted the generalisation of the Haar-based thresholds discussed in Section 3.4 for burst-like Poisson processes to arbitrary wavelet bases. In particular, a new set of level-dependent thresholds τ_j is sought calibrated so that

$$P\left(\max_{0\le k\le 2^j-1} \hat{d}_{jk} \le \tau_j\right) \to 1, \quad \text{as} \quad j \to \infty,$$

at a rate similar to that obtained for the universal threshold of Donoho & Johnstone (1994) in the Gaussian case. Using the Aldous' Poisson Clumping Heuristic (see, Aldous, 1989), the following approximation is obtained

$$P\left(\max_{0\le k\le 2^j-1} \hat{d}_{jk} \le \tau_j\right) \approx \exp\left\{-2^j \frac{P(\hat{d}_{jk} > \tau_j)}{\mathbb{E}(\mathcal{C}_{\tau_j})}\right\}, \quad j = j_0, \ldots, J-1, \tag{8}$$

where $\mathbb{E}(\mathcal{C}_{\tau_j})$ is the expected 'clump size', roughly the expected number of local exceedances of τ_j conditional on there being at least a single exceedance, which implies $\mathcal{C}_{\tau_j} \ge 1$.

Although a closed form expression for $P(\hat{d}_{jk} > \tau_j)$ in (8) exists when using Haar wavelets, this is not generally true for other choices of wavelet bases. However, Kolaczyk (1999a) derived implicit level-dependent thresholds τ_j that also depend on the background intensity μ_0. The idea behind this method is a large deviation approximation, yielding an expression that serves to account for effects of the first few cumulants of the Poisson distribution on the tails of the empirical wavelet coefficient distributions. The resulting pair of thresholds (which accommodate the asymmetric nature of the distribution of $\hat{d}_{jk}$) were called the 'corrected thresholds' due to the fact that they are, essentially, corrected versions of the usual Gaussian-based thresholds for arbitrary wavelet bases. They are given

by

$$\tau_j^{\max} = b_j^{\max}\kappa_2^{1/2} \quad \text{and} \quad \tau_j^{\min} = b_j^{\min}\kappa_2^{1/2}, \quad j = j_0, \ldots, J-1, \tag{9}$$

where $\kappa_2 = 2^J \mu_0 \int_0^1 [\psi_{jk}(x)]^2 \, dx$, and the proportionality constants $b_j^{\max}$, $b_j^{\min}$ are defined as the solution of particular equations. These thresholds are then used to estimate the underlying intensity function $\mu(t)$. However, the asymptotic approximations used by Kolaczyk (1999a) in deriving $\tau_j^{\max}$ and $\tau_j^{\min}$ in (9) result in the fact that they may not exist for small values of μ_0 (since reasonable solutions of $b_j^{\max}$ and $b_j^{\min}$ may not exist). When they exist, the approach can be computed in $O(n)$ time.

The relative performance of this approach is evaluated in the simulation study in Sections 5 and 6, along with its translation invariant version which requires $O(n \log n)$ operations.

4 Bayesian Wavelet Approaches for Poisson Data

The wavelet shrinkage problem of Section 2 has also been approached from a Bayesian point of view. In this section we present two Bayesian estimators of the mean vector $\boldsymbol{\mu}$ (and thus of the intensity function $\mu(t)$ at the design points) that use a multiscale data analysis. In general, multiscale analysis refers to the study of structure in signals or data at various spatial and/or temporal resolutions. Perhaps the simplest technique is the Haar multiscale analysis defined according to

$$\begin{aligned} &\hat{c}_{Jk} \equiv y_k, \quad k = 0, 1, \ldots, 2^J - 1 \\ &\hat{c}_{jk} = \hat{c}_{j+1,2k} + \hat{c}_{j+1,2k+1}, \quad j = 0, 1, \ldots, J-1; \quad k = 0, 1, \ldots, 2^j - 1. \end{aligned}$$

The parameters $\hat{c}_{jk}$ are the (unnormalised) Haar empirical scaling coefficients. The relationship between a 'parent' (e.g., $\hat{c}_{jk}$) and a 'child' (e.g., $\hat{c}_{j+1,2k}$) is of fundamental interest in multiscale analysis. This relationship is expressed by the conditional likelihood $p(\hat{c}_{j+1,2k} \mid \hat{c}_{jk}, \boldsymbol{\mu})$ which happens to have a very simple form for the Poisson regression model (1). Defining $c_{Jk} \equiv \mu_k$ and $c_{jk} = c_{j+1,2k} + c_{j+1,2k+1}$, for $j = 0, 1, \ldots, J-1$ and $k = 0, 1, \ldots 2^j - 1$, we find that

$$p(\hat{c}_{j+1,2k} \mid \hat{c}_{jk}, \boldsymbol{\mu}) = \text{Bin}\left(\hat{c}_{j+1,2k} \mid \hat{c}_{jk}, \frac{c_{j+1,2k}}{c_{jk}}\right),$$

where $\text{Bin}(x \mid n, p)$ denotes the binomial distribution with parameters n and p. The canonical multiscale parameters $\theta_{jk} = c_{j+1,2k}/c_{jk}$ can thus be viewed as splitting factors that govern the multiscale refinement of the mean vector $\boldsymbol{\mu}$. The simplicity of the parent-child relationship is quite exceptional here and, as a result, leads to the following factorisation of the likelihood function

$$p(\mathbf{y} \mid \boldsymbol{\mu}) = p(\hat{c}_{00} \mid c_{00}) \prod_{j=0}^{J-1} \prod_{k=0}^{2^j-1} p(\hat{c}_{j+1,2k} \mid \hat{c}_{jk}, \theta_{jk}).$$

This factorisation also greatly facilitates multiscale analysis and modelling. For example, estimates of the multiscale parameters θ_{jk} can be used to reconstruct an estimate of the underlying intensity function $\mu(t)$. From a Bayesian perspective, multiscale modelling requires (i) specification of a suitable prior model for the multiscale parameters θ_{jk}, and (ii) determination of the posterior distribution resulting from the likelihood and prior. The remainder of this section is devoted to two types of prior models that have been proposed in the literature for estimating the underlying intensity function $\mu(t)$ based on observations from the Poisson regression model (1). The methods we describe are due to Kolaczyk (1999b) and Timmermann & Nowak (1999), and they are used in the simulation study in Sections 5 and 6. We refer to these articles for more details on these methods.

4.1 Wavelet Shrinkage of General Intensity Functions Using Multiscale Model Estimators

Kolaczyk (1999b) introduced a class of Bayesian multiscale models (BMSMs) in which the prior distribution of the θ_{jk}'s is a mixture of a point mass at 1/2 and a symmetric beta distribution, i.e.,

$$\begin{aligned}\theta_{jk} \mid \gamma_{jk}, B_{jk} &\sim \gamma_{jk}\,\frac{1}{2} + (1-\gamma_{jk})\,B_{jk},\\ \gamma_{jk} \mid p_j &\sim \text{Bernoulli}(p_j),\\ B_{jk} \mid a_j &\sim \text{beta}(a_j, a_j).\end{aligned}$$

Note that, at each scale $j = 0, 1, \ldots, J-1$, the hyperparameters $0 \le p_j \le 1$ and $a_j > 0$ are constant across all locations $k = 0, 1, \ldots, 2^j - 1$.

If the parameters θ_{jk} are modelled independently of the $\hat{c}_{jk}$, then a multiscale factorisation of the posterior distribution is obtained. This factorisation allows inferences to be made on each multiscale parameter individually, leading to a computationally efficient, recursive expression (across scales) for the posterior mean. The final posterior mean estimate of the mean vector $\boldsymbol{\mu}$ is calculated from a translation invariant framework that eliminates the "staircase-like" structure of each individual estimate. The entire estimation process may be calculated in $O(n \log n)$ time by implementing it in a computationally efficient manner.

For the functioning of the BMSM estimator, Kolaczyk (1999b) recommended the choices $c_{00} = \hat{c}_{00}$ and $a_j = 1$ for all $j = 0, 1, \ldots, J-1$. In addition, he presented an expectation-maximisation (EM) algorithm for computing an empirical Bayes estimate of the scale-dependent mixing parameters p_j. We have also found these choices to work well but, for computational reasons, the simulation results we present in Section 6 have been based on $p_j = 0.90$ (for all $j = 0, 1, \ldots, J-1$). This choice was found to work reasonably well in preliminary experiments.

4.2 Wavelet Shrinkage of General Intensity Functions Using Multiscale Multiplicative Innovations Model Estimators

The Bayesian multiscale multiplicative innovations model (BMMIM) of Timmermann & Nowak (1999) moves beyond the BMSM of Kolaczyk (1999b) by specifying more general beta mixtures for the multiscale parameters. Formally the mixture priors adopted by Timmermann & Nowak (1999) have the form

$$p(\theta_{jk}) = \sum_{i=1}^{M} p_j \frac{\{\theta_{jk}(1-\theta_{jk})\}^{s_i - 1}}{B(s_i, s_i)}, \quad j = 0, 1, \ldots, J-1; \quad k = 0, 1, \ldots, 2^j - 1,$$

where M is the number of components, $B(\alpha, \beta)$ is the standard beta function, $0 \le p_i \le 1$ denotes the *a priori* probability of the ith component, and $\sum_{i=1}^{M} p_i = 1$.

The mixture priors above are conjugate for the likelihood function, factorise the posterior distribution of the θ_{jk}'s and lead to a simple closed-form expression for the posterior mean of the mean vector $\boldsymbol{\mu}$. The computational complexity of its translation invariant version, which eliminates the "staircase-like" structure of each individual estimate, is $O(n^2)$. However, for large n, the complexity of the translation invariant estimator may be reduced to $O(n2^{J-j_0})$ by initiating it at some primary resolution level $j_0 > 0$.

Timmermann & Nowak (1999) found that three components suffice for many applications and suggested the choices $s_1 = 1$, $s_2 = 100$ and $s_3 = 10000$ with weights $p_1 = 0.001$ and a moments-based estimate for p_2 (and hence p_3). However, we have found that these moment-based estimates do not work particularly well and, therefore, the simulation results we present in Section 6 have been based on $p_2 = 0.07$ (and hence $p_3 = 0.929$) for all $j = j_0, \ldots, J-1$. This choice was found to work reasonably well in preliminary experiments.

5 Description of the Simulation

To compare the performance of the classical and Bayesian wavelet shrinkage methods described in Sections 3 and 4 we have designed a simulation study. Three factors of interest were identified and included in the study: the intensity function $\mu(t)$, the background intensity level μ_0, and the sample size n. Motivated by various applications, six different intensity functions were used. These were the 'Smooth' (Beran & Dümbgen, 1998), 'Angles' (Marron, Adak, Johnstone, Neumann & Patil, 1998), 'Clipped Blocks' (Fryźlewicz & Nason, 2001), 'Bumps' (Donoho & Johnstone, 1994), 'Spikes' (Cai, 1999) and 'Bursts' (Kolaczyk, 1997) functions, and they are plotted in Figure 1. We refer to these papers for more information about these functions. The background intensity level varied over 5, 20 and 50 counts per time point so as to gain some insight about the behaviour of the procedures at 'low', 'medium' and 'high' background intensity levels. Finally, the sample size varied over 256, 512 and 1024 observations in order to examine the behaviour in 'small', 'medium' and 'large' samples respectively.

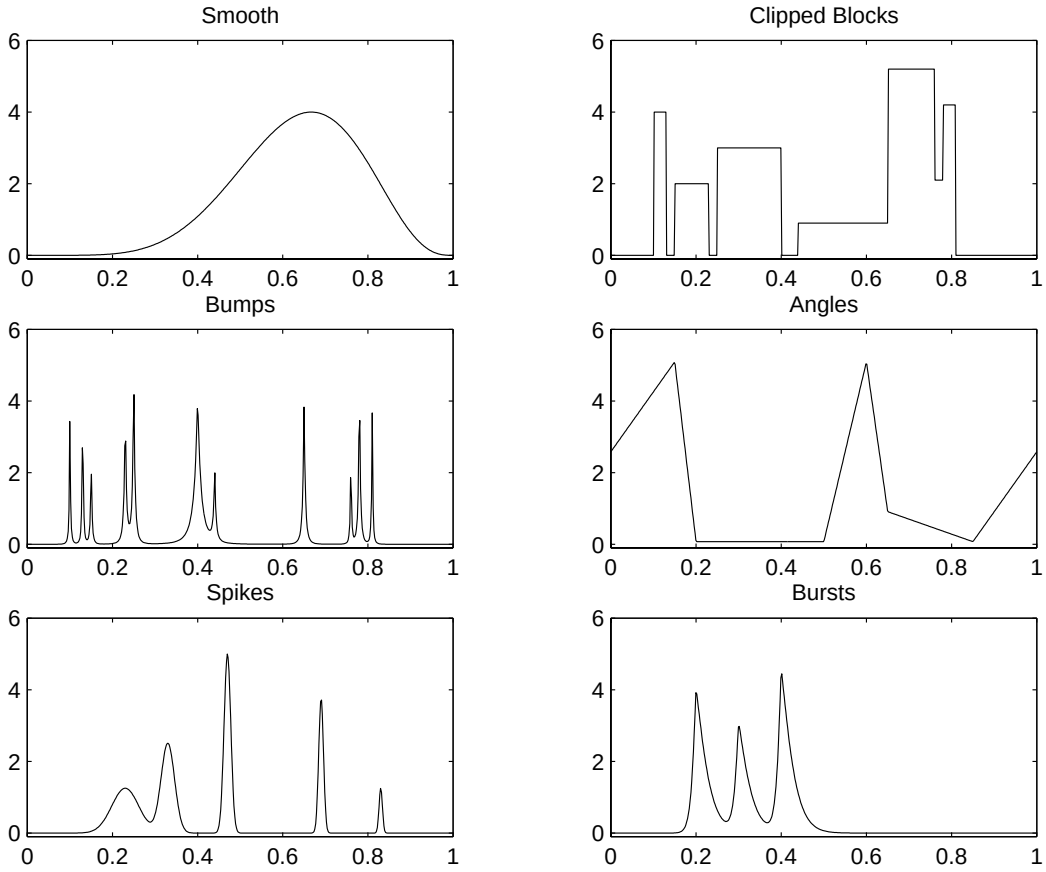

Figure 1. *The six intensity functions used in the simulation study sampled at our 'medium' sample size* $n = 512$.

For each combination of the 3 factors, 100 realisations of Poisson counts were generated from model (1) using the Poisson random number generator `poissrnd` in MATLAB. The noisy versions of the intensity functions were generated according to the formula: $\texttt{noisysignal} = \mu_0 \times (1 + \texttt{signal})$.

Figure 2 illustrates one realisation from each intensity function with $n = 512$ and $\mu_0 = 20$. For each realisation we compared the performance of the estimators based on the Anscombe and Fisz transformations, modulation models, l_1-penalised likelihood method, Haar wavelets, corrected thresholds, Bayesian multiscale model, and Bayesian multiscale multiplicative innovations model. Where applicable, we considered both soft and hard thresholding and one of the minimax, universal or 'leave-out-half' cross-validation thresholds, and we examined three primary resolution levels ($j_0 = 4$, 5 and 6). Based on their nice properties and wide applicability, we employed both the Symmlet 8 (Daubechies, 1992, p. 198) and the Coiflet 3 (Daubechies, 1992, p. 258) wavelet filter.

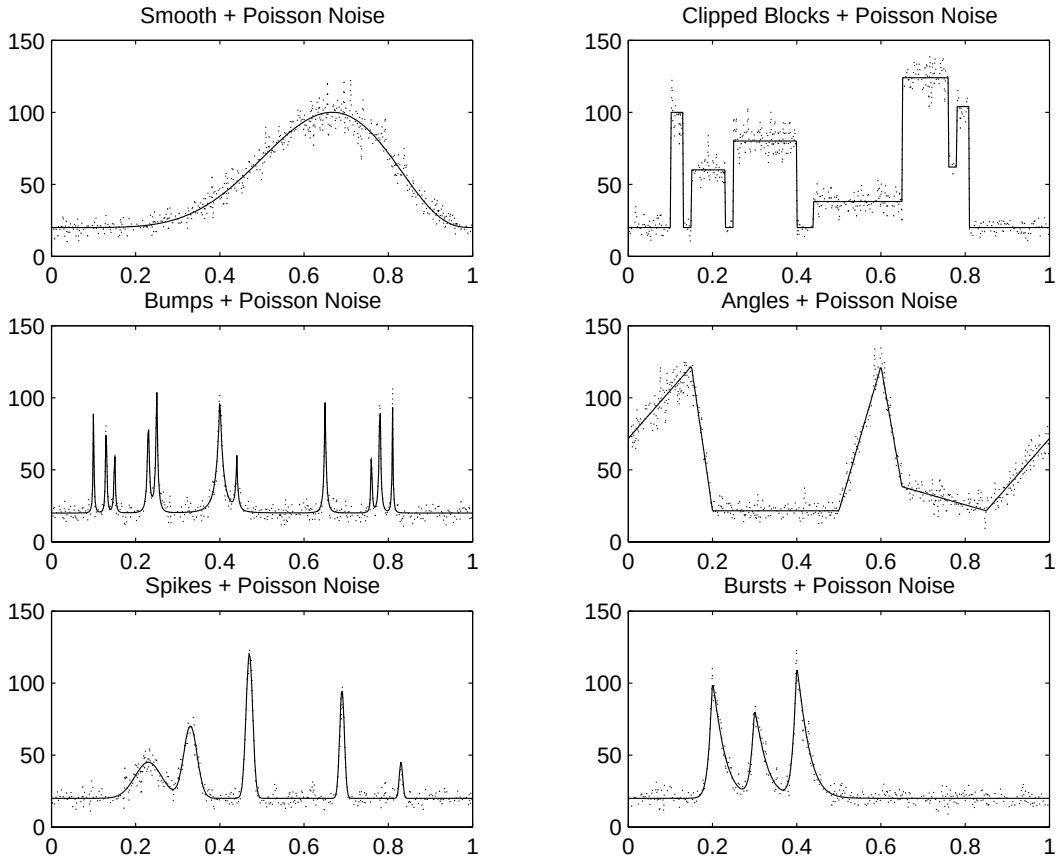

Figure 2. *Noisy versions of the six intensity functions shown in Figure 1, giving a visual impression of our 'medium background intensity' level* $\mu_0 = 20$.

Although we have implemented them, we have not compared in the main simulation study the translation invariant versions of the estimator based on the Fisz transformation and the modulation approach since they require $O(n^2)$ operations. Note however that the translation invariant versions of these estimators perform better than their non-translation invariant counterparts and may be preferred in individual applications. Table 2 lists the estimation procedures used in the simulation study and provides an easy-to-understand acronym for each procedure as a shorthand means of referring to it. As shown in the table, 22 different estimators were actually compared in total. We have not included the estimators based on the Fisz transformation noting that the results were as good as, or almost as good as, with the results obtained using the estimators based on the Anscombe transformation.

Table 2

Acronyms and details of the wavelet shrinkage estimation procedures used in the simulation study. The last column, "Section", refers to the sections of the text where these procedures have been described.

	Name	Thresholding	Threshold Value	Section
1	ANSCOMBE-MIN-H	Hard	Minimax	3.1
2	ANSCOMBE-MIN-S	Soft	Minimax	3.1
3	ANSCOMBE-UNI-H	Hard	Universal	3.1
4	ANSCOMBE-UNI-S	Soft	Universal	3.1
5	ANSCOMBE-CV-H	Hard	'Leave-Out-Half' Cross-Validation	3.1
6	ANSCOMBE-CV-S	Soft	'Leave-Out-Half' Cross-Validation	3.1
7	ANSCOMBE-MIN-TI-H	Hard	Minimax	3.1
8	ANSCOMBE-MIN-TI-S	Soft	Minimax	3.1
9	ANSCOMBE-UNI-TI-H	Hard	Universal	3.1
10	ANSCOMBE-UNI-TI-S	Soft	Universal	3.1
11	ANSCOMBE-CV-TI-H	Hard	'Leave-Out-Half' Cross-Validation	3.1
12	ANSCOMBE-CV-TI-S	Soft	'Leave-Out-Half' Cross-Validation	3.1
13	MODULATION		Multiple Stein Alike Shrinkages	3.2
14	l_1-PENALISED		Penalised Likelihood-Based Shrinkages	3.3
15	HAAR-TI-H	Hard	Level-Dependent Thresholds	3.4
16	HAAR-TI-S	Soft	Level-Dependent Thresholds	3.4
17	CORRECTED-H	Hard	Pair of Level-Dependent Thresholds	3.5
18	CORRECTED-S	Soft	Pair of Level-Dependent Thresholds	3.5
19	CORRECTED-TI-H	Hard	Pair of Level-Dependent Thresholds	3.5
20	CORRECTED-TI-S	Soft	Pair of Level-Dependent Thresholds	3.5
21	BMSM-TI		Multiscale Model	4.1
22	BMMIM-TI		Multiscale Multiplicative Innovation Model	4.2

In the simulations, the performance of an estimator $\hat{\boldsymbol{\mu}}$ was summarised by the following criteria

1. mean squared error (MSE), defined as the average over the 100 replicates of
$$\frac{1}{n}\sum_{i=1}^{n}(\mu_i - \hat{\mu}_i)^2;$$
2. l_1-norm (L1), defined as the average over the 100 replicates of
$$\sum_{i=1}^{n}|\mu_i - \hat{\mu}_i|;$$
3. root mean squared error (RMSE), defined as the square root of the MSE;
4. SB/MSE, defined as the ratio of squared bias to the MSE;
5. maximum deviation (MXDV), defined as the average over the 100 replicates of
$$\max_{1\le i\le n}|\mu_i - \hat{\mu}_i|;$$
6. CPU time, defined as the average CPU over the 100 replicates.

The relative performance of the 22 estimators of Table 2 can be deduced from graphical outputs and numerical tables. The whole set of results across all intensity functions, background intensity levels, sample sizes, primary resolution levels, wavelet filters and performance criteria is very extensive.

Here, for reasons of space, we report summary results for all intensity functions pertaining to $\mu_0 = 20$, $n = 512$, $j_0 = 5$ and Symmlet 8. Different combinations of wavelet filters and primary resolution levels yielded basically the same results, although some methods seem to be quite sensitive to the choice of the primary resolution level; we come back to this issue in Sections 6 and 7. For $n = 256$ or $n = 1024$ we found respectively poorer/better individual performance of the estimators but their relative performance was roughly the same. This behaviour was also observed for $\mu_0 = 5$ or $\mu_0 = 50$, where the signal-to-noise ratio is low/high respectively, although some methods tend to oversmooth or attenuate the fine detail structure in the underlying intensity function especially in situations involving very low level of counts. Moreover, some methods seem to be very sensitive to the knowledge of the background intensity level; we come back to these issues in Sections 6 and 7. In Section 8, we provide further details and a demonstration session in which various estimators are considered for an astronomical gamma-ray burst data set.

6 Summary of Results

The results of the simulation study will now be presented, with the remainder of the section devoted to the discussion of these results. Note that, for brevity, the estimation procedures will be referred to by the acronyms introduced in Table 2.

6.1 Smooth and Angles Intensity Functions

The results for the 'Smooth' and 'Angles' intensity functions are reported first. For these functions, where there is relatively little underlying structure, all estimators produced very good estimates with the exception of HAAR-TI-H, CORRECTED-H and CORRECTED-TI-H estimators, despite their using the true parameter value of μ_0. The poor performance of these estimators may be due to the fact that they are best suited for burst-like intensity functions. It is interesting to note however that this group of estimators performed significantly better when using soft thresholding, in which case their performance became comparable with that of several other estimators. Figures 3 and 4 illustrate the relative performance of the estimators in terms of the six performance measures for the 'Smooth' and 'Angles' intensity functions respectively.

There is little to choose amongst the various Anscombe-based, MODULATION, l_1-PENALISED and BMMIM-TI estimators (for 'Smooth'), all of which outperformed the BMSM-TI estimator. Similarly, there is little to choose amongst the various Anscombe-based and MODULATION estimators (for 'Angles'), all of which outperformed the BMSM-TI and BMMIM-TI estimators. The performance of the BMSM-TI estimator (for both 'Smooth' and 'Angles') was found to improve by selecting their hyperparameters adaptively, as recommended by their authors, at the expense of increasing significantly the considerable computational cost. On the other hand, the performance of the BMMIM-TI estimator (for 'Angles') was found to improve by decreasing the primary resolution level, as found in preliminary experiments. It can be shown that in these cases the BMSM-TI and BMMIM-TI estimators perform as well as their non-Bayesian counterparts. Of these two estimators, BMMIM-TI performed best in terms of the measures, despite its tendency to produce estimates with slightly reduced amplitudes (for 'Angles').

6.2 Clipped Blocks Intensity Function

In contrast to the 'Smooth' and 'Angles' intensity functions, a range of performance was observed for the 'Clipped Blocks' intensity function. Figure 5 illustrates the relative performance of the estimators in terms of the six performance measures. As observed in the figure, the BMSM-TI estimator dominated its non-Bayesian counterparts in every aspect apart from CPU time. Although

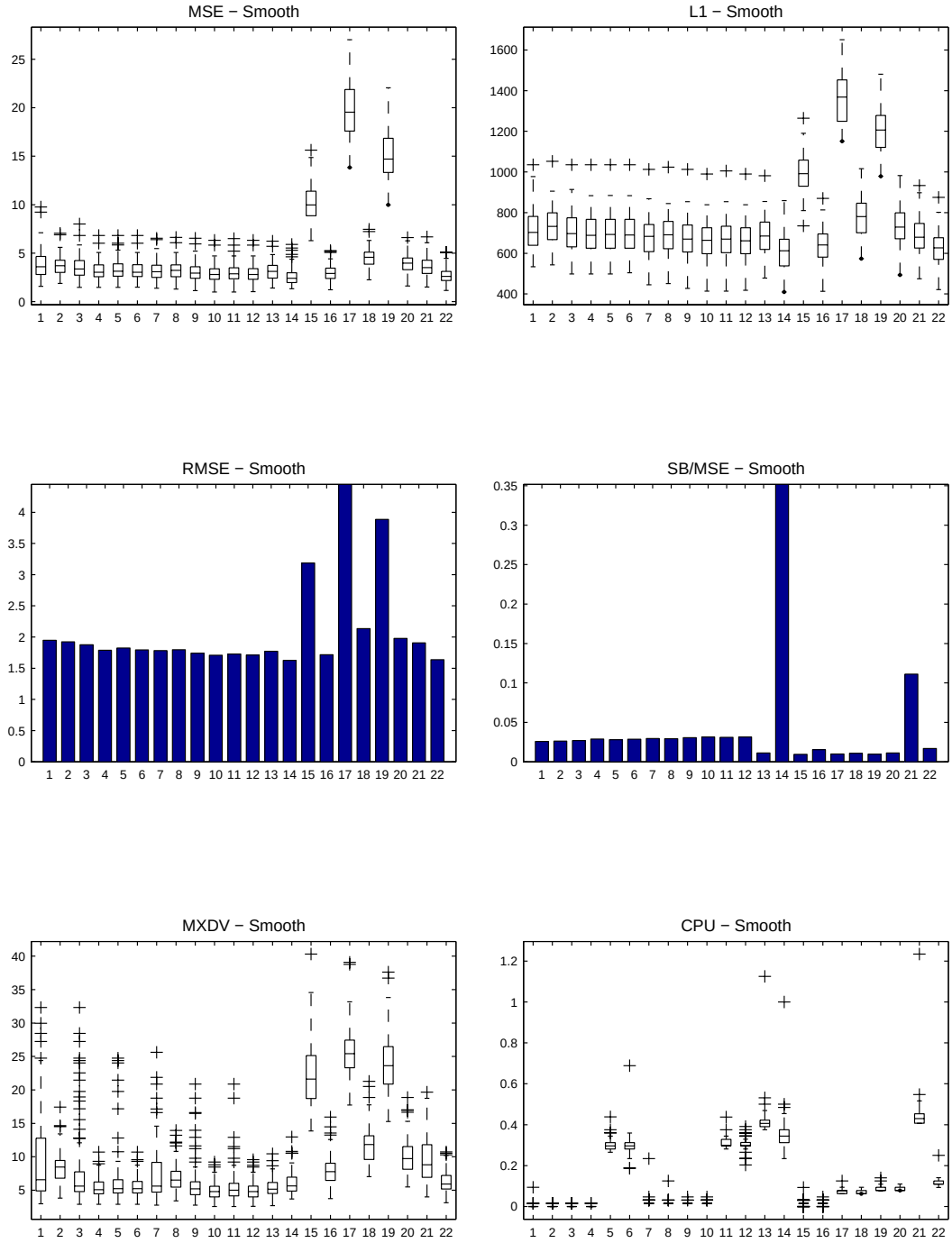

Figure 3. *Relative performance of the 22 methods of Table 2 for the estimation of the 'Smooth' intensity function when* $\mu_0 = 20$ *and* $n = 512$*. The performance of each method is given in terms of the six measures of Section 5. The methods are presented in the same order as in Table 2 and the results are based on 100 replications, Symmlet 8 and* $j_1 = 5$*.*

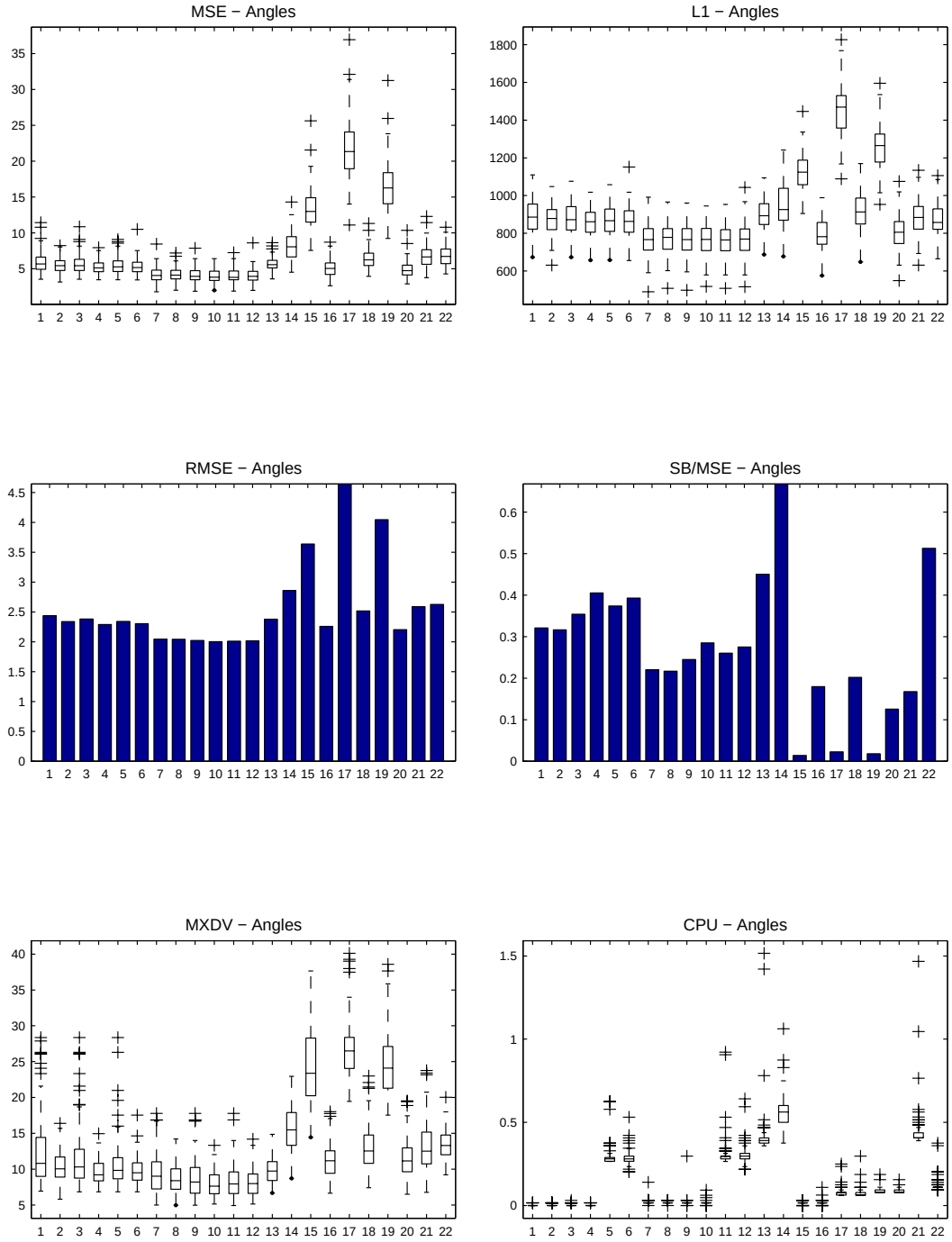

Figure 4. *Relative performance of the 22 methods of Table 2 for the estimation of the 'Angles' intensity function when* $\mu_0 = 20$ *and* $n = 512$*. The performance of each method is given in terms of the six measures of Section 5. The methods are presented in the same order as in Table 2 and the results are based on 100 replications, Symmlet 8 and* $j_1 = 5$*.*

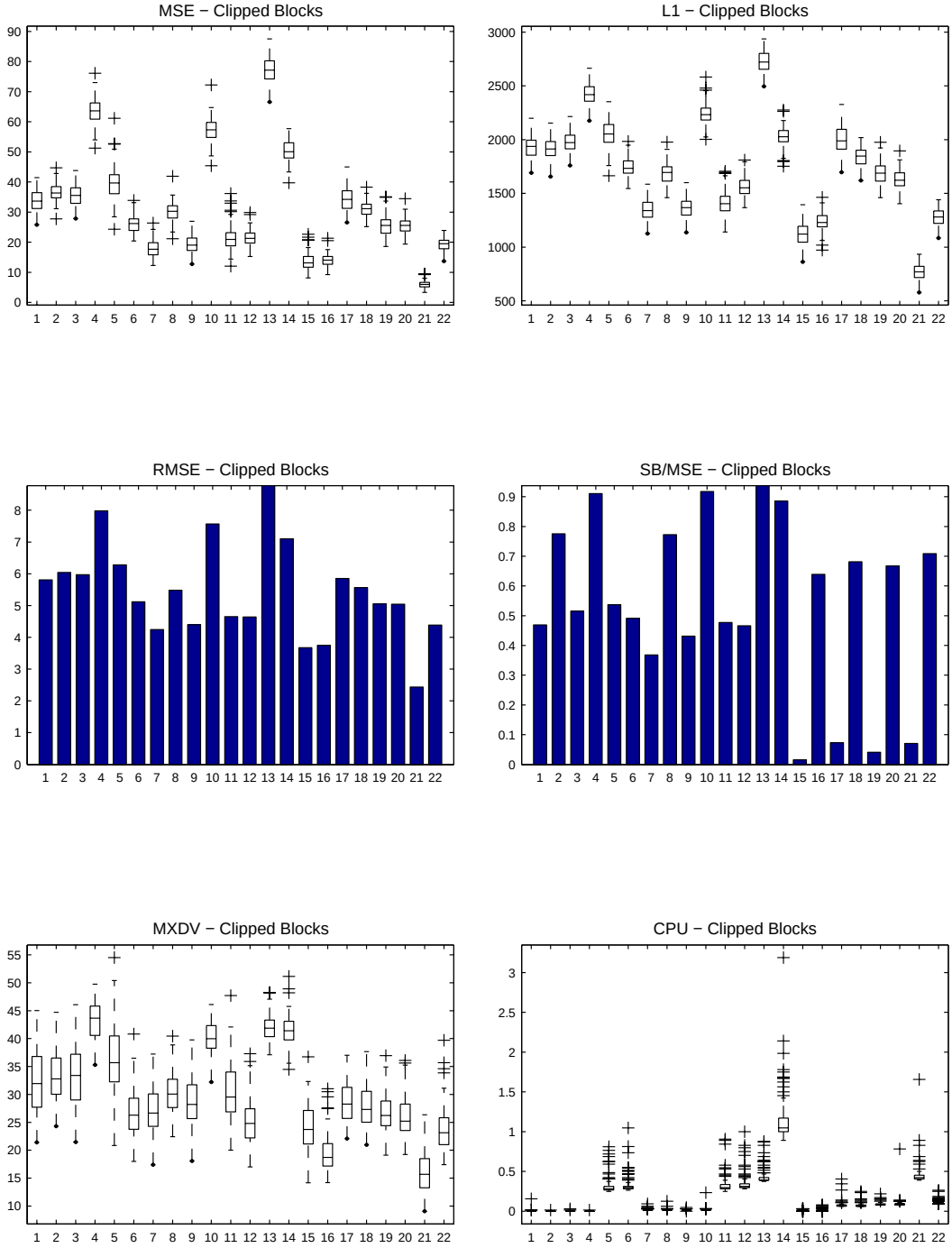

Figure 5. *Relative performance of the 22 methods of Table 2 for the estimation of the 'Clipped Blocks' intensity function when* $\mu_0 = 20$ *and* $n = 512$. *The performance of each method is given in terms of the six measures of Section 5. The methods are presented in the same order as in Table 2 and the results are based on 100 replications, Symmlet 8 and* $j_1 = 5$.

the BMMIM-TI estimator does not perform as well as the BMSM-TI estimator, its performance improves significantly by increasing the primary resolution level; it can be shown that in this case the BMMIM-TI estimator also performs better than its non-Bayesian counterparts. For the hyperparameter values we used, the BMSM-TI estimator fared better than the BMMIM-TI estimator but was not as computationally efficient. The computational advantage of the BMMIM-TI estimator is due to the five-scale wavelet transform we employed, although in practice we could use full J-scale transforms for improved performance (but see the discussion in Timmermann & Nowak, 1999).

Of the classical estimators, HAAR-TI-H and HAAR-TI-S performed best whilst MODULATION performed worst. The non-translation invariant Anscombe-based estimators outperformed, in almost all cases, the l_1-PENALISED estimator but, as expected, did not perform as well as their translation invariant counterparts. It is interesting to note however that the l_1-PENALISED estimator performs better than the ANSCOMBE-UNI-S and ANSCOMBE-UNI-TI-S estimators, but the gain in performance is not enough to justify its cost.

6.3 Bumps Intensity Function

A range of performance was also observed for the 'Bumps' intensity function. Figure 6 illustrates the relative performance of the estimators in terms of the six performance measures. As observed in the figure, the HAAR-TI-H and CORRECTED-TI-H estimators performed best, followed closely by the BMSM-TI estimator. The generally good performance of the various translation invariant Haar and Corrected threshold estimators is not entirely surprising since the 'Bumps' intensity function displays the burst-like behaviour for which these estimators were specifically designed. Note however that the performance of this group of estimators was based on the true background intensity level μ_0 and would therefore deteriorate in practice, where μ_0 is unknown. By way of contrast, the performance of the BMSM-TI estimator has the potential of improving in practice by using data-adaptive hyperparameters. The same is also true for the BMMIM-TI estimator by increasing the primary resolution level.

Amongst the Anscombe-based estimators, the type of thresholding scheme (hard, soft) and the choice of threshold value (minimax, universal, cross-validation) play an important role. Of the six pairs of thresholding scheme and threshold value, the pairs (soft, cross-validation) and (hard, minimax) performed best whilst the pair (soft, universal) performed worst. As expected, the translation invariant estimators offered improvements in performance and practically eliminated most of these differences. As for the 'Clipped Blocks' intensity function, the l_1-PENALISED estimator performs better than ANSCOMBE-UNI-S and ANSCOMBE-UNI-TI-S estimators, but the gain in performance is not enough to justify its cost. Finally, the MODULATION estimator performed worst and trailed all of its competitors in terms of all the measures apart from CPU time.

6.4 Spikes and Bursts Intensity Functions

For the 'Spikes' and 'Bursts' intensity functions, the estimators can be broadly categorised according to their performance as 'good', 'not so good' or 'bad'. Thus the 'bad' estimators are the MODULATION and l_1-PENALISED, the 'not so good' estimators are ANSCOMBE-UNI-S and ANSCOMBE-UNI-TI-S, and all the remaining estimators are 'good' estimators. It is interesting to note that the ANSCOMBE-UNI-TI-S estimator performed worse than the ANSCOMBE-UNI-S estimator for the 'Spikes' intensity function and the worst estimators was l_1-PENALISED, while the worst estimator for 'Bursts' was MODULATION. However, the MODULATION estimator performs better than the l_1-PENALISED estimator in terms of CPU time, for both intensity functions. Figures 7 and 8 illustrate this and the relative performance of the remaining estimators in terms of the six performance measures for the 'Spikes' and 'Burst' intensity functions respectively.

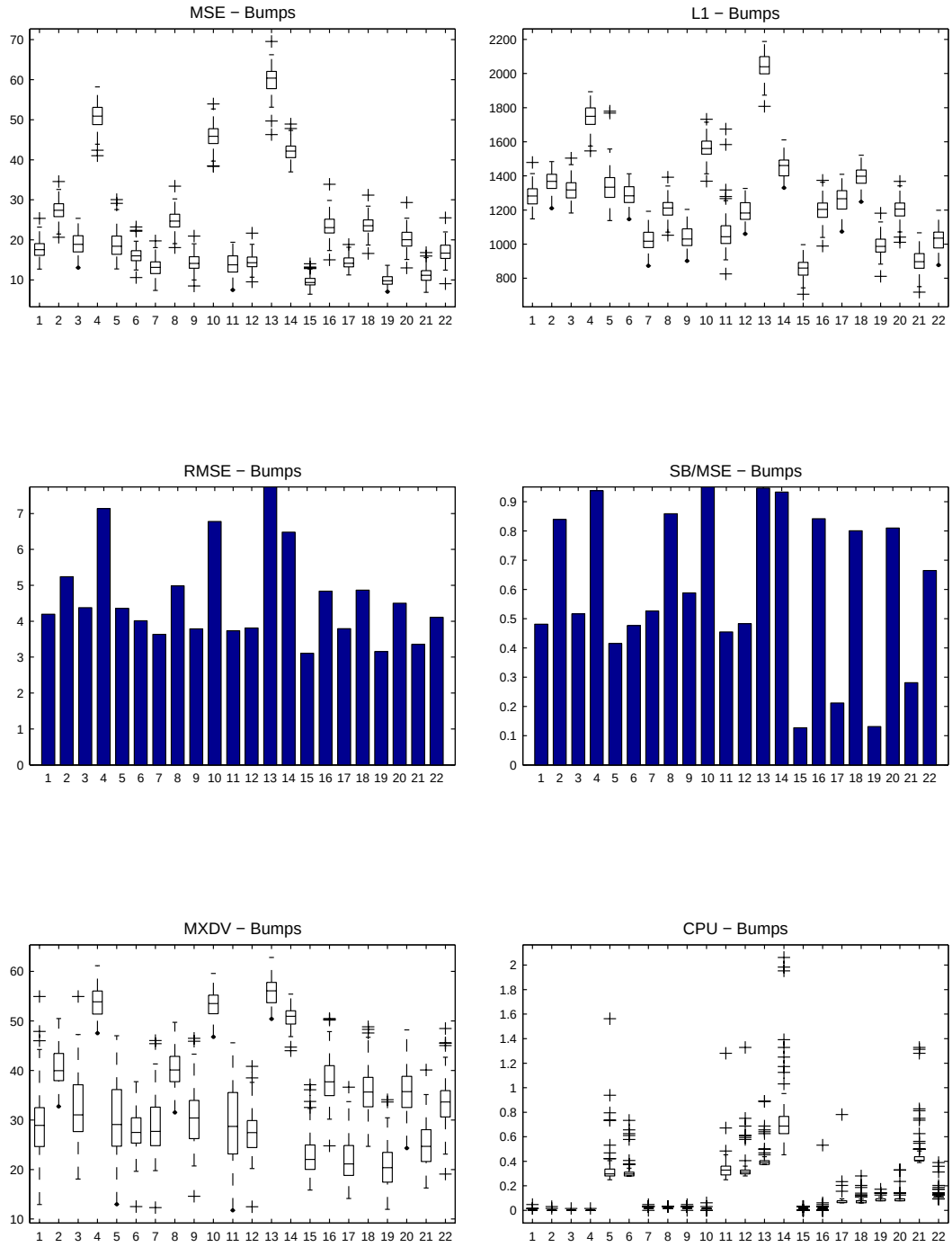

Figure 6. *Relative performance of the 22 methods of Table 2 for the estimation of the 'Bumps' intensity function when* $\mu_0 = 20$ *and* $n = 512$*. The performance of each method is given in terms of the six measures of Section 5. The methods are presented in the same order as in Table 2 and the results are based on 100 replications, Symmlet 8 and* $j_1 = 5$*.*

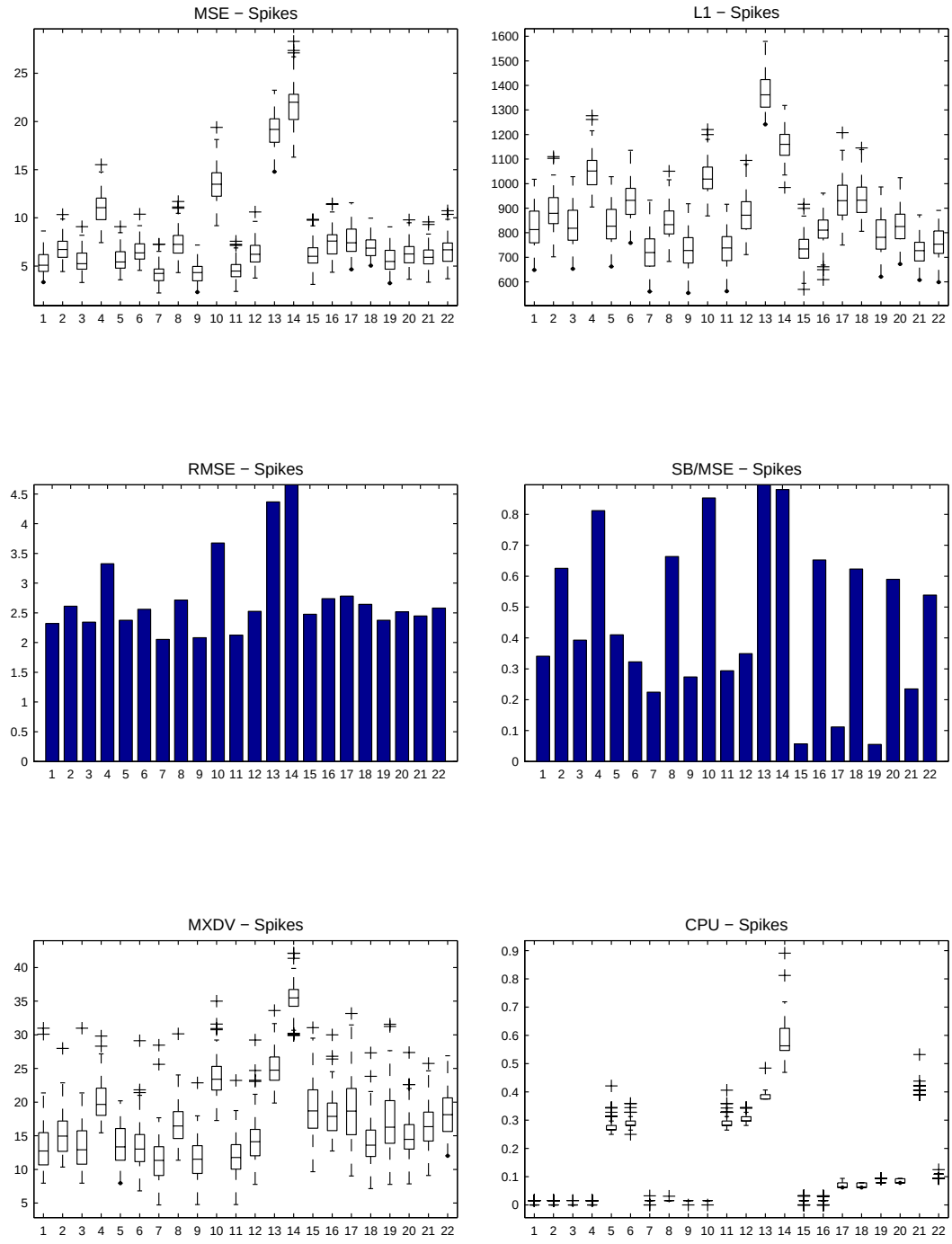

Figure 7. *Relative performance of the 22 methods of Table 2 for the estimation of the 'Spikes' intensity function when* $\mu_0 = 20$ *and* $n = 512$*. The performance of each method is given in terms of the six measures of Section 5. The methods are presented in the same order as in Table 2 and the results are based on 100 replications, Symmlet 8 and* $j_0 = 5$*.*

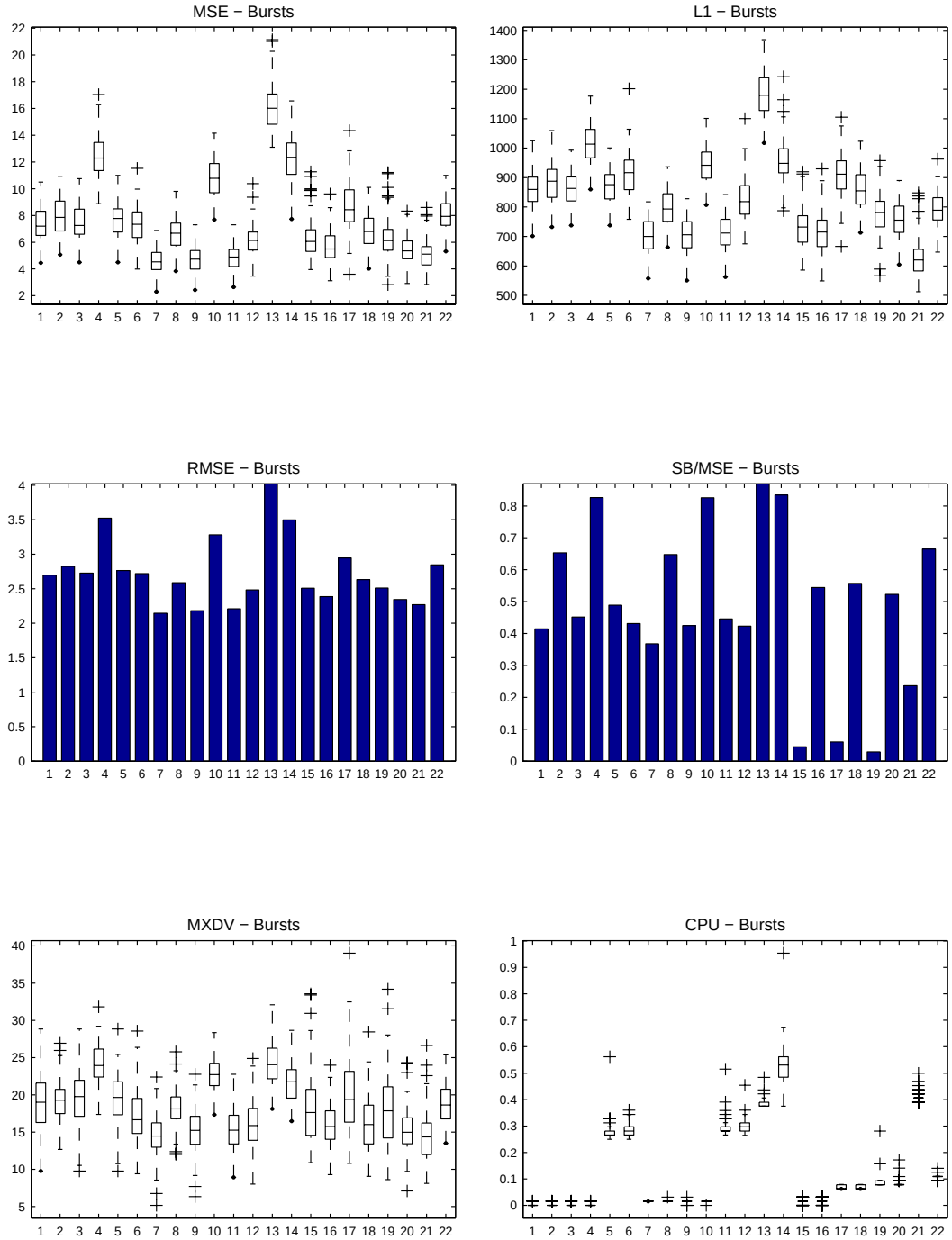

Figure 8. *Relative performance of the 22 methods in Table 2 for the estimation of the 'Bursts' intensity function when* $\mu_0 = 20$ *and* $n = 512$*. The performance of each method is given in terms of the six measures of Section 5. The methods are presented in the same order as in Table 2 and the results are based on 100 replications, Symmlet 8 and* $j_1 = 5$*.*

Of the better estimators, the ANSCOMBE-MIN-TI-H, ANSCOMBE-UNI-TI-H and ANSCOMBE-CV-TI-H estimators were almost identical and slightly outperformed their competitors. The translation invariant Haar and the various corrected threshold estimators performed more or less equally well in terms of the performance measures but the HAAR-TI-H and CORRECTED-TI-H estimators yielded noticeably better estimates visually. As before, the performance of these estimators was based on the true value of the background intensity level μ_0, which would be usually unknown in practice. Since estimation of μ_0 can be difficult in these intensity functions (and in some of the previous intensity functions), we have conducted a smaller study to examine the effect of μ_0 on their performance. We have found that deviations from the true value of μ_0 can substantially deteriorate the performance of the estimators, depending on the sign and magnitude of the deviation, the smoothness of the underlying intensity function, and the type of thresholding. The performance sensitivity of the translation invariant Haar and the various corrected threshold estimators on the true parameter value of μ_0 constitute a serious practical limitation on the use of these estimators.

The BMSM-TI estimator, on the other hand, can be readily applied in practice. Furthermore, apart from its favourable MSE performance, it produced estimates which did not contain the spurious fine-scale effects that were often present in the other estimates; its estimates, however, tended to have slightly lower peaks and higher troughs. The same behaviour is also observed to some extent for the BMMIM-TI estimator by increasing the primary resolution level, especially for the 'Bursts' function.

7 Overall Conclusions

We compared the finite sample performance of a number of classical and Bayesian wavelet shrinkage estimators for Poisson counts. Amongst the classical approaches, the methods of Donoho (1993), Kolaczyk (1997, 1999a), Antoniadis & Sapatinas (2001), Fryźlewicz & Nason (2001) and Sardy, Antoniadis & Tseng (2004) were considered. We also considered recent Bayesian approaches developed by Kolaczyk (1999b) and Timmermann & Nowak (1999). The output of our simulation study is an illustration of the no-free lunch principle; the results show that there is no single estimator that is superior to others in all cases.

When the underlying intensity function is relatively homogeneous, most estimators perform comparably and there is little to choose amongst them. As expected, the translation invariant version of the Anscombe-based estimators usually performs better than their non-translation versions. Moreover, since the universal threshold is larger than the minimax threshold obtained from a soft thresholding rule and since soft thresholding produces more biased estimates than hard thresholding, the resulting estimates obtained from soft thresholding with universal threshold are more biased (more coefficients are thresholded) than the ones obtained from soft thresholding with minimax threshold. On the other hand, since the universal threshold is smaller than the minimax threshold with a hard thresholding rule, the resulting estimates obtained from hard thresholding with minimax threshold are less biased (fewer coefficients are thresholded) than the ones obtained from hard thresholding with universal threshold. The same behaviour is also observed to a lesser degree for the Fisz-based estimators. The very good performance of the modulation estimator is explained by its linearity; linear estimators are appropriate for homogeneous intensity functions. The performance of the l_1-penalised likelihood estimator is very good even though it has been proposed for inhomogeneous intensity functions. On the other hand, despite their using the true parameter value of the background intensity level, the poor performance of the Haar-based and the corrected-based estimators is justified by the fact that these estimators have been proposed for burst-like intensity functions. Also, as expected, the translation invariant version of the corrected-based estimators usually performs better than their non-translation versions. However, the knowledge of the background intensity level for both Haar-based and corrected-based estimators constitutes a serious practical limitation on the use of these esti-

mators. Moreover, the asymptotic approximations used in obtaining the corrected-based estimators result in the fact that these estimators may not exist for small values of the background intensity level, resulting again in a serious limitation on the use of these estimators in practical applications. Finally the Bayesian estimators consistently perform well, although the Bayesian multiscale multiplicative innovations model estimator is sensitive to the choice of the primary resolution level.

When the underlying intensity function is inhomogeneous, a range of performance is observed. Again, as expected, the translation invariant version of Anscombe-based estimators usually performs better than their non-translation versions. Although these latter estimators show relatively good performance, they tend to oversmooth or attenuate the fine detail structure in the underlying intensity function especially in situations involving very low level of counts, as a result of the variance-stabilizing transformation. The same behaviour is also observed to a lesser degree for the Fisz-based estimators. The bad performance of the modulation estimator is explained by its linearity; linear estimators are usually inappropriate for inhomogeneous intensity functions. In spite of the fact that it has been proposed for inhomogeneous intensity functions, the l_1-penalised likelihood estimator does not perform as well as in the case of homogeneous functions. However, its performance for such functions is improved by considering a cross-validation criterion instead of the optimal universal threshold considered in the simulation study, at the expense of increasing significantly the considerable computational cost. The better performance of the Haar-based estimators over the corrected-based estimates can be explained by the fact that in the former case the resulting empirical wavelet coefficients are independent at each resolution level, which is not the case for the corrected-based estimators. Also, as expected, the translation invariant version of corrected-based estimators usually performs better than their non-translation versions. However, the very good performance of both Haar-based and corrected-based estimators depends heavily on the knowledge of the background intensity level; this is a serious practical limitation on the use of these estimators, especially in situations where a good knowledge of the background intensity level is not available. Again, the asymptotic approximations used in obtaining the corrected-based estimators result in the fact that these estimators may not exist for small values of the background intensity level, resulting again in a serious limitation on the use of these estimators in practical applications. Finally, as in the homogeneous case, the Bayesian multiscale model estimator consistently performs well, while the Bayesian multiscale multiplicative innovations model estimator is again sensitive to the choice of the primary resolution level, with better performance obtained using higher resolution levels.

In conclusion, the Anscombe-based, the Fisz-based, the modulation and the l_1-penalised likelihood estimators can be safely used for homogeneous intensity functions, but the transformation-based estimators are usually inappropriate for very low level of counts. The modulation estimator is usually inappropriate for inhomogeneous intensity functions while the l_1-penalised likelihood estimator can be mostly applied with the use of cross-validation threshold, especially when the computational cost is not a serious problem. On the other hand, the Haar-based thresholds should be preferred over the corrected-based thresholds for inhomogeneous intensity functions and are mostly appropriate for burst-like intensity functions; however, both should be used with caution, especially in cases where a good knowledge of the background intensity level is not available. Furthermore, the corrected-based estimators do not usually exist for low level of counts. On the other hand, the Bayesian estimators seem to perform well for both homogeneous and inhomogeneous intensity functions, although the Bayesian multiscale multiplicative innovations model is sensitive to the choice of the primary resolution level; better performances were obtained using low and high resolution levels for homogeneous and inhomogeneous intensity functions respectively.

Overall, although there is no single estimator that is superior to others in all cases, the Bayesian estimators consistently perform well and it is therefore safe to recommend their application in estimating various intensity functions obtained in diverse scientific fields.

8 An illustrative example

The calculations for this article were carried out using an implementation in MATLAB, a high-level programming language written by The MathWorks, Inc. In accordance with the principle of reproducible research (see Buckheit & Donoho, 1995), the (MATLAB) programs used in this study, including those we did not include in the simulations (i.e., the Fisz-based estimators and the translation invariant version of the modulation estimator), can be found on the World Wide Web at `http://www.ucy.ac.cy/~fanis/links/software.html`. All these programs make extensive use of Version 8 of the WaveLab package developed by Buckheit, Chen, Donoho, Johnstone & Scargle (1995). WaveLab is a library of MATLAB routines for wavelet analysis and is available free of charge over the Internet at `http://www-stat.stanford.edu/~wavelab`. In addition, the l_1-penalised likelihood estimator makes also use of the Wavelet toolbox of MATLAB.

Also available at our web site are some data sets related to the astronomical gamma-ray burst intensities. These data were detected by the Burst and Transient Source Experiment (BATSE) instruments on board NASA's Compton Gamma Ray Observatory (see Meegan *et al.*, 1992, for details). These instruments record the arrival times of the high-energy photons in which the burst is composed according to four energy channels (at 25-58 keV, 58-115 keV, 115-320 keV and $\geq$ 320 keV). In practice it is common to aggregate over all energies and to collect the arrival times into bins, yielding a sequence of counts. However, of additional interest to astronomers is sometimes an analysis of the data by energy channel. In the Appendix, we provide a MATLAB session which analyses the BATSE trigger 551 data at the third channel (115-320 keV). As this session is intended as a demonstration of the way our programs may be used in practice, we do not provide a serious analysis of these data. For serious analyses of BATSE data, see Megan *et al.* (1992) and Norris *et al.* (1996).

The BATSE trigger 551 data we analyse were recorded over approximately 0.94 seconds and collected into 1024 adjacent, equally-spaced bins. The resulting sequence of counts is plotted in Figure 9. As pulses occurred exclusively in the first half of the data (0.47 s), only this half is plotted so as to enlarge the relevant area. Also plotted in Figure 9 are estimates of the underlying intensity profile provided by the ANSCOMBE-MIN-TI-H, HAAR-TI-H and BMSM-TI estimators. The HAAR-TI-H estimate uses the mean of the last 60% observations as an estimate of the background intensity level, while the BMSM-TI estimate is based on data-adaptive hyperparameters.

As observed in the figure, the ANSCOMBE-MIN-TI-H estimator has an unacceptably rounded appearance. The estimate produced by the BMSM-TI estimator appears more plausible, but the HAAR-TI-H estimator has effectively removed the noise from the data while retaining the perceived sharp structure of the pulses.

Acknowledgements

The authors would like to thank Piotr Fryźlewicz and Guy Nason for providing their SPLUS routines, Eric Kolaczyk for providing his MATLAB routines and the BATSE data, and Sylvain Sardy for providing his MATLAB routines. Panagiotis Besbeas would like to thank Theofanis Sapatinas for financial support and excellent hospitality while visiting Nicosia to carry out part of this work. The work of Italia de Feis was supported by the Agenzia Spaziale Italiana (Italian Space Agency). The work of Theofanis Sapatinas was partly supported by 'University of Cyprus – 2001, 2002 Research Grants'. We are grateful to the Joint Editor (Elja Arjas) and the two anonymous referees whose valuable comments and suggestions led to a significant improvement of this article.

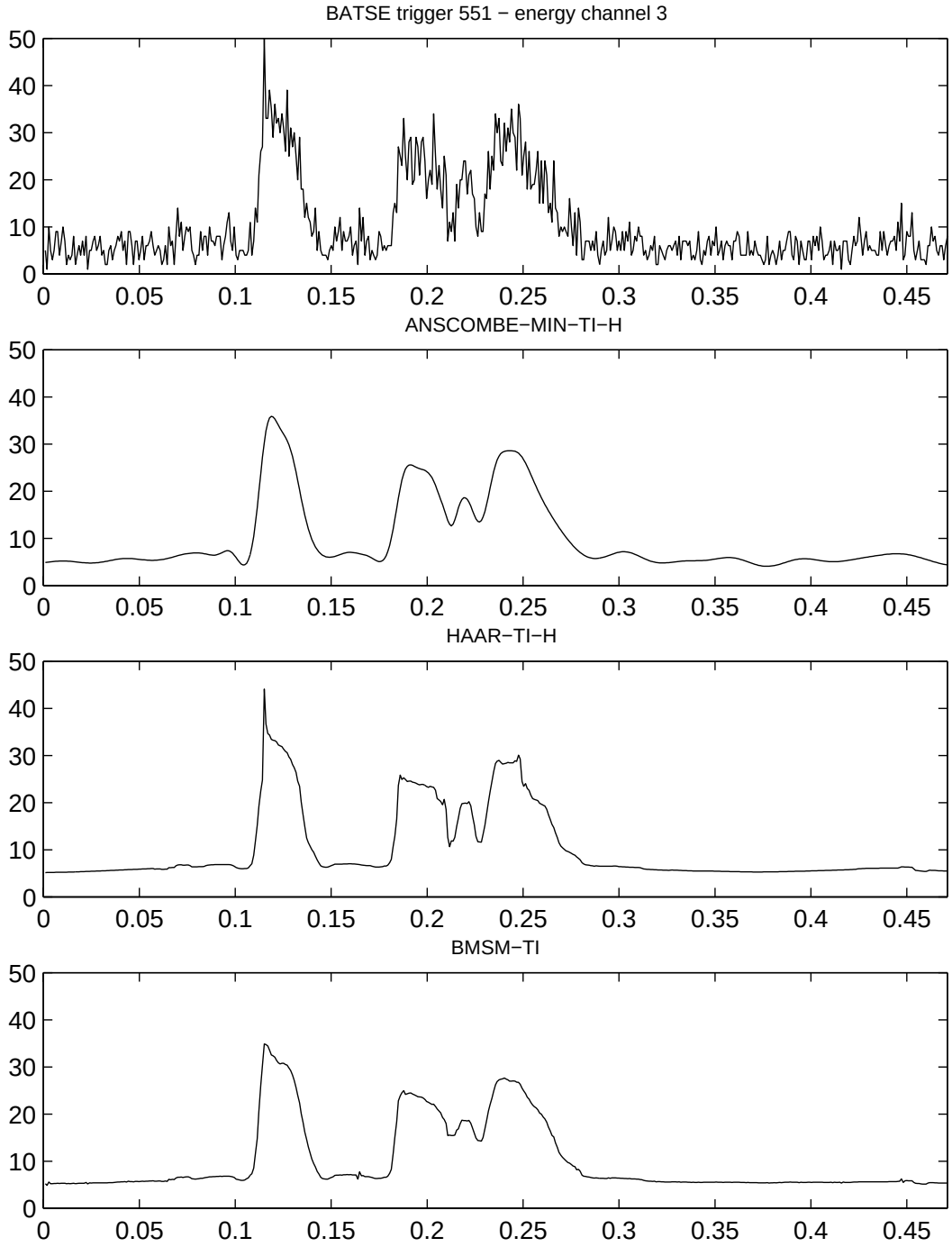

Figure 9. *Gamma-ray burst (BATSE trigger 551, energy channel 3, $n = 1024$) and estimates of its intensity function using three wavelet shrinkage estimators: ANSCOMBE-MIN-TI-H, HAAR-TI-H (with the mean of the last 60% observations as an estimate of μ_0) and BMSM-TI (with data-adaptive hyperparameters). Only the first half of the data (0.47 s), which contains the pulses, is shown so as to enlarge the relevant area.*

References

Abramovich, F., Bailey, T.C. & Sapatinas, T. (2000). Wavelet analysis and its statistical applications. *The Statistician*, **49**, 1–29.

Aldous, D. (1989). *Probability Approximations via the Poisson Clumping Heuristic*. New York: Springer-Verlag.

Anscombe, F.J. (1948). The transformation of Poisson, binomial and negative binomial data. *Biometrika*, **35**, 246–254.

Antoniadis, A. & Sapatinas, T. (2001). Wavelet shrinkage for natural exponential families with quadratic variance functions. *Biometrika*, **88**, 805–820.

Antoniadis, A., Besbeas, P. & Sapatinas, T. (2001). Wavelet shrinkage for natural exponential families with cubic variance functions. *Sankhyā*, Series A, **63**, 309–327.

Antoniadis, A., Bigot, J. & Sapatinas, T. (2001). Wavelet estimators in nonparametric regression: a comparative simulation study. *J. Statist. Software*, **6**(6), 1–83.

Beran, R. & Dümbgen, L. (1998). Modulation of estimators and confidence sets. *Ann. Statist.*, **26**, 1826–1856.

Bruce, A.G. & Gao, H.-Y. (1996). Understanding WaveShrink: variance and bias estimation. *Biometrika*, **83**, 727–745.

Buckheit, J.B. & Donoho, D.L. (1995). WaveLab and Reproducible Research. In *Wavelets and Statistics*, Eds. A. Antoniadis and G. Oppenheim. Lect. Notes Statist., **103**, 55–81, New York: Springer-Verlag.

Buckheit, J.B., Chen, S., Donoho, D.L., Johnstone, I.M. & Scargle, J. (1995). About WaveLab. Technical Report, Department of Statistics, Stanford University, USA.

Cai, T.T. (1999). Adaptive wavelet estimation: a block thresholding and oracle inequality approach. *Ann. Statist.*, **27**, 898–924.

Chen, S.S., Donoho, D.L. & Saunders, M.A. (1999). Atomic decompositions by basis pursuit. *SIAM J. Sci. Comput.*, **20**, 33–61.

Coifman, R.R. & Donoho, D.L. (1995). Translation-invariant de-noising. In *Wavelets and Statistics*, Eds. A. Antoniadis and G. Oppenheim. Lect. Notes Statist., **103**, 125–150, New York: Springer-Verlag.

Daubechies, I. (1992). *Ten Lectures on Wavelets*. Philadelphia: SIAM.

Dennis, J.E. & Mei, H.H.W. (1979). Two new unconstrained optimization algorithms which use function and gradient values. *J. Optimiz. Theory Appl.*, **28**, 453–483.

Diggle, P. (1985). A kernel method for smoothing point process data. *Appl. Statist.*, **34**, 138–147.

Diggle, P. & Marron, J.S. (1988). Equivalence of smoothing parameter selectors in density and intensity estimation. *J. Amer. Statist. Ass.*, **83**, 793–800.

Donoho, D.L. (1993). Non-linear wavelet methods for recovery of signals, densities and spectra from indirect and noisy data. In *Proceedings of Symposia in Applied Mathematics: Different Perspectives on Wavelets*, Vol. **47**, 173–205. Ed. I. Daubechies. San Antonio: American Mathematical Society.

Donoho, D.L. & Johnstone, I.M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, **81**, 425–455.

Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. & Picard, D. (1995). Wavelet shrinkage: asymptopia? (with discussion). *J. R. Statist. Soc., Series B*, **57**, 301–337.

Eilers, P.H.C. & Marx, B.D. (1996). Flexible smoothing with B-splines and penalties (with discussion). *Statist. Sci.*, **11**, 89–121.

Fisz, M. (1955). The limiting distribution of a function of two independent random variables and its statistical application. *Colloquium Mathematicum*, **3**, 138–146.

Fryźlewicz, P. & Nason, G.P. (2001). Poisson intensity estimation using wavelets and the Fisz transformation. Technical Report, **01/10**. Department of Mathematics, University of Bristol, United Kingdom.

Hurvich, C.M. & Tsai, C.-L. (1998). A crossvalidatory AIC for hard wavelet thresholding in spatially adaptive function estimation. *Biometrika*, **85**, 701–710.

Kolaczyk, E.D. (1997). Non-parametric estimation of Gamma-Ray burst intensities using Haar wavelets. *Astrophys. J.*, **483**, 340–349.

Kolaczyk, E.D. (1999a). Wavelet shrinkage estimation of certain Poisson intensity signals using corrected thresholds. *Statist. Sinica*, **9**, 119–135.

Kolaczyk, E.D. (1999b). Bayesian multiscale models for Poisson processes. *J. Amer. Statist. Ass.*, **94**, 920–933.

Mallat, S.G. (1989). A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. Pattn. Anal. Mach. Intell.*, **11**, 674–693.

Mallat, S.G. (1999). *A Wavelet Tour of Signal Processing*. 2nd Edition, San Diego: Academic Press.

Marron, J.S., Adak, S., Johnstone, I.M., Neumann, M.H. & Patil, P. (1998). Exact risk analysis of wavelet regression. *J. Comp. Graph. Statist.*, **7**, 278–309.

Meegan, C.A., Fishman, G.J., Wilson, R.B., Paciesas, W.S., Pendleton, G.N., Horack, J.M., Brock, M.N. & Kouveliotou, C. (1992). The Spatial Distribution of Gamma Ray Bursts Observed by BATSE. *Nature*, **355**, 143–145.

Meyer, Y. (1992). *Wavelets and Operators*. Cambridge: Cambridge University Press.

Nason, G.P. (1996). Wavelet shrinkage using cross-validation. *J. R. Statist. Soc., Series B*, **58**, 463–479.

Norris, J.P., Nemiroff, R.J., Bonnell, J.T., Scargle, J.D., Kouveliotou, C., Paciesas, W.S., Meegan, C.A. & Fishman, G.J. (1996). Attributes of Pulses in Long Bright Gamma-Ray Bursts. *Astrophys. J.*, **459**, 393–412.

Nowak, R.D. & Baraniuk, R.G. (1999). Wavelet domain filtering for photon imaging systems. *IEEE Trans. Image Proc.*, **8**, 666–678.

O'Sullivan, F., Yandell, B.S. & Raynor, W.J. (1986). Automatic smoothing of regression functions in generalized linear models. *J. Amer. Statist. Ass.*, **81**, 96–103.

Patnaik, P.B. (1949). The non-central χ^2-and F-distributions and their applications. *Biometrika*, **36**, 202-232.

Sardy, S., Antoniadis, A. & Tseng, P. (2004). Automatic smoothing with wavelets for a wide class of distributions. *J. Comp. Graph. Statist.*, **13**, (to appear).

Timmermann, K.E. & Nowak, R.D. (1999). Multiscale modeling and estimation of Poisson processes with applications to photon-limited imaging. *IEEE Trans. Inf. Theor.*, **45**, 846–862.

Résumé

Nous étudions l'aide de simulations numériques, le comportement à taille d'échantillon fini d'un certain nombre d'estimateurs de l'intensité de comptages Poissonniens, fondés sur des procédures de seuillage, classiques et bayesiennes, des coefficients d'ondelettes. A cet effet nous employons une large gamme de fonctions d'intensité, de bruit de fond, de tailles d'échantillons, de résolution primaire de décomposition et de famille d'ondelettes et des critères de performance variés. Les méthodes sont illustrées sur un exemple réel issus d'astrophysique. Selon les principes d'une recherche reproductible les procédures MATLAB et les exemples traités sont gracieusement disponibles.

Appendix

This appendix provides a MATLAB session which analyses the BATSE trigger 551 data, as described in Section 8. MATLAB is an easy-to-understand language but a few comments about the code below are in order.

1. Comments begin with the percent symbol (%).
2. The command `loadfile.mat` retrieves the variables from the MAT-file 'file.mat'.
3. The command `linspace(a, b, n)` generates `n` equally spaced points between `a` and `b`.
4. The command `hist(y, x)` returns the distribution of `y` among bins with centres specified by `x`.

```
%retrieve BATSE trigger 551 data
load burst551.mat
photon_times=burst3;     %arrivals at the third channel

%collect arrival times into 1024 bins
num_bins=1024;
t_min=min(photon_times);
t_max=max(photon_times);
t_range=t_max-t_min;
dt_bin=t_range/num_bins;
mid_first=t_min+dt_bin/2;
mid_last=t_max-dt_bin/2;
data_bins=linspace(mid_first,mid_last,num_bins);
poissig=hist(photon_times,data_bins);
x=data_bins';

%intensity function estimates

%translation invariant version of Anscombe-based estimate
%with hard minimax threshold
j0=6;                %primary resolution level
h=MakeONFilter('Symmlet',8);        %wavelet filter choice
anscombeMin_TI_H=recanscombeTI(poissig,'MinMax','Hard',j0,h);

%translation invariant version of Haar estimate with hard threshold
j0=4;                %primary resolution level
%Estimate of background intensity level
mu0=ceil(mean(poissig(floor(2/5*num_bins)+1:num_bins)));
haar_TI_H=rechaarTI(poissig,mu0,'Hard',j0);

%translation invariant version of Bayesian Multiscale model estimate
pqind=2;              %data-adaptive hyperparameters
```

```
BMSMShrink_TI=BMSMShrink(poissig,pqind);

%plot data and reconstructions
xsec=x/10^6;            %change time scale to seconds
n2=num_bins/2;          %concentrate on first half

subplot(4,1,1)
plot(xsec,poissig)
axis([0 xsec(n2) 0 50])
title('BATSE trigger 551 - energy channel 3','FontSize',8)

subplot(4,1,2) plot(xsec,anscombeMin_TI_H)
axis([0 xsec(n2) 0 50])
title('ANSCOMBE-MIN-TI-H','FontSize',8)


subplot(4,1,3) plot(xsec,haar_TI_H)
axis([0 xsec(n2) 0 50])
title('HAAR-TI-H','FontSize',8)

subplot(4,1,4)
plot(xsec,BMSMShrink_TI)
axis([0 xsec(n2) 0 50])
title('BMSM-TI','FontSize',8)
```

[*Received May 2002, accepted July 2003*]