A Haar-Fisz Algorithm for Poisson Intensity Estimation

Piotr FRYZLEWICZ and Guy P. NASON

This article introduces a new method for the estimation of the intensity of an inhomogeneous one-dimensional Poisson process. The Haar-Fisz transformation transforms a vector of binned Poisson counts to approximate normality with variance one. Hence we can use any suitable Gaussian wavelet shrinkage method to estimate the Poisson intensity. Since the Haar-Fisz operator does not commute with the shift operator we can dramatically improve accuracy by always cycle spinning before the Haar-Fisz transform as well as optionally after. Extensive simulations show that our approach usually significantly outperformed state-of-the-art competitors but was occasionally comparable. Our method is fast, simple, automatic, and easy to code. Our technique is applied to the estimation of the intensity of earthquakes in northern California. We show that our technique gives visually similar results to the current state-of-the-art.

Key Words: Cycle spinning; Denoising; Poisson process; Transform to Gaussian, Variance stabilizing transform.

1. INTRODUCTION

Wavelet methods have now become a useful tool in the area of curve estimation including, in particular, regression problems where noise is Gaussian, as well as density estimation. A general overview of wavelet methods in statistics can be found, for example, in Vidakovic (1999); see Daubechies (1992) for a mathematical introduction to wavelets. Some authors have also considered the problem of estimating the intensity of a Poisson process using a wavelet-based technique. The usual (regression) setting is as follows: the possibly inhomogeneous one-dimensional Poisson process is observed on the interval $[0, T)$, and discretized into a vector $\mathbf{v} = (v_0, v_1, \ldots, v_{N-1})$, where v_n is the number of events falling into the interval $[nT/N, (n+1)T/N)$, and $N = 2^J$ is an integer power of two. Each v_n can be thought of as coming from a Poisson distribution with an unknown parameter λ_n, which needs to be

Piotr Fryzlewicz is Graduate Student, Department of Mathematics, University of Bristol, University Walk, Bristol BS8 1TW, UK (E-mail: p.z.fryzlewicz@bristol.ac.uk). Guy P. Nason is Professor of Statistics, Department of Mathematics, University of Bristol, University Walk, Bristol BS8 1TW, UK (E-mail: g.p.nason@bristol.ac.uk).

Journal of Computational and Graphical Statistics, Volume 13, Number 3, Pages 621–638
DOI: 10.1198/106186004X2697

estimated. The approach proposed by Donoho (1993) consists in first preprocessing the data using Anscombe's (1948) square-root transformation, $\mathcal{A}\mathbf{v} = 2\sqrt{\mathbf{v} + 3/8}$, so that the noise becomes approximately Gaussian. Then the analysis proceeds as if the noise were indeed Gaussian, yielding (after applying the inverse square-root transformation) an estimate of the intensity of the process.

Besbeas, De Feis, and Sapatinas (2002) reported that the current state-of-the-art methods are those based on translation-invariant multiscale Bayesian techniques as described by Kolaczyk (1999a) and Timmermann and Nowak (1997, 1999). Due to frequent citation we abbreviate Timmermann and Nowak by "TN". Kolaczyk (1999a) introduced a Bayesian multiscale algorithm to estimate the discretized intensity. However, rather than transforming the data using a wavelet transform, he considered recursive dyadic partitions, and places prior distributions at the nodes of the binary trees associated with these partitions. The Bayesian methods outperform earlier techniques in Kolaczyk (1997, 1999b), Nowak and Baraniuk (1999), and also the recent technique of Antoniadis and Sapatinas (2001) [since the latter is equivalent to Nowak and Baraniuk (1999) for Poisson data]. Sardy, Antoniadis, and Tseng (2004) described a computationally intensive l_1-penalized likelihood method that can be used for estimating Poisson intensities.

Other recent contributions to the field of wavelet-based intensity estimation include Patil and Wood (2004), who concentrated on the theoretical MISE properties of wavelet intensity estimators, where the intensity is a random process rather than a deterministic function (or, after discretization, a deterministic vector). Brillinger (1998) gave a brief overview of wavelet-based methodology in the analysis of point process data, and obtains an estimate of the autointensity function of the well-known California earthquake data.

This article proposes an alternative wavelet-based algorithm for estimating the deterministic discretized intensity function of an inhomogeneous one-dimensional Poisson process. Our method is based on the asymptotic normality of a certain function of the Haar wavelet and scaling coefficients of the vector $\mathbf{v}$, the property first observed and proved by Fisz (1955), but (obviously) not set in the wavelet context at that time. In his article, Fisz used this property to test the hypothesis that two Poisson variables have equal means, and the hypothesis that their means are both equal to a given number.

The idea behind our algorithm is the following: we first preprocess the (Poisson) vector $\mathbf{v}$ using a nonlinear wavelet-based transformation, which we call the Haar-Fisz transformation, and then treat the preprocessed vector as if it were signal plus iid Gaussian noise of unit variance. In other words, we provide a new Gaussianizing and variance stabilizing transform, which operates in the wavelet domain, and not in the time domain, like the standard square-root transformation.

The main advantages of our method are the following.

1. Its performance is extremely good, see Section 3.
2. It is of computational order N (or NM if M cycle-spins are used); in practice the software is itself very fast.
3. It is simple and easy to code.
4. It is fully automatic (up to any parameters that the Gaussian denoiser requires).

5. It can make use of *any* signal+Gaussian noise denoising technology, an area where a vast amount of research effort has been and is being expended. Hence our method can only get better as we take advantage of newer Gaussian denoisers.

The articles by Kolaczyk (1999a), Timmermann and Nowak (1999), and the review article Besbeas, De Feis, and Sapatinas (2002) all conclude that the Bayesian methods proposed by Kolaczyk (1999a) and Timmermann and Nowak (1999) are the best currently available. Section 3.2 demonstrates that our algorithm usually significantly outperforms the above Bayesian methods but occasionally its performance was comparable.

Section 2 introduces the Haar-Fisz transform and describes its theoretical and empirical properties. Section 3 introduces the Haar-Fisz algorithm for estimating Poisson intensities and performs a thorough simulation study on its performance. Section 4 exhibits our algorithm and that of Kolaczyk (1999a) on a dataset derived from the well-known Northern Californian Earthquake database studied by Brillinger (1998). Finally, Section 5 provides some conclusions and ideas for future exploration.

2. THE HAAR-FISZ TRANSFORMATION

2.1 THE HAAR DISCRETE WAVELET TRANSFORM

The discrete wavelet transform (DWT) is a linear orthogonal transform $\mathbb{R}^N \to \mathbb{R}^N$, where $N = 2^J$. In this section we are only concerned with the Haar DWT and we shall describe the fast $O(N)$ version as devised by Mallat (1989). The Haar DWT works as follows: given an input vector $\mathbf{v} = (v_0, v_1, \ldots, v_{N-1})$ we set $\mathbf{s}^0 = \mathbf{v}$ and recursively perform the following steps:

$$s_i^j = (s_{2i}^{j-1} + s_{2i+1}^{j-1})/2, \tag{2.1}$$

and

$$d_i^j = (s_{2i}^{j-1} - s_{2i+1}^{j-1})/2, \tag{2.2}$$

and define $\mathbf{s}^j$ and $\mathbf{d}^j$ be the vectors of s_i^j and d_i^j all for $i = 0, \ldots, 2^{J-j} - 1$, where $j = 1, \ldots, J$. The elements of $\mathbf{s}^j$ (and $\mathbf{d}^j$) can be thought of as smooth (and detail) of the original vector $\mathbf{v}$ at scale 2^j. The result of the full Haar DWT retains the detail vectors along with the coarsest scale smooth element to give $(\mathbf{s}^J, \mathbf{d}^J, \mathbf{d}^{J-1}, \ldots, \mathbf{d}^1)$.

The Haar DWT coefficients have a simple interpretation. For example, for $N = 8$ the coarsest scale smooth coefficient is simply the mean of the original input vector: $s^3 = \frac{1}{8}\sum_{i=0}^{7} v_i$, and, for example, the coarse scale detail coefficient

$$d_1^2 = \frac{1}{4}(v_4 + v_5 - v_6 - v_7),$$

is a localized difference at scale 2. Both coefficients can be easily obtained from the recursive transform formulas (2.1) and (2.2).

The output from the Haar DWT may be inverted to recover the original vector $\mathbf{v}$. The inverse algorithm simply reverses formulas (2.1) and (2.2) to give:

$$s_{2i}^{j-1} = s_i^j + d_i^j, \tag{2.3}$$

and

$$s_{2i+1}^{j-1} = s_i^j - d_i^j \tag{2.4}$$

for $i = 0, \ldots, 2^{J-j} - 1$ where $j = J, \ldots, 1$. Remember that $\mathbf{v} = \mathbf{s}^0$.

As an aid to understanding the modifications to the Haar DWT we are about to propose consider the following example. Using steps (2.3) and (2.4) of the inverse Haar DWT with $N = 8$ we can represent v_2 as

$$v_2 = s_2^0 = s_0^3 + d_0^3 - d_0^2 + d_1^1. \tag{2.5}$$

Expanding these summands using the forward formulas (2.1) and (2.2) gives

$$v_2 = \frac{1}{8}\sum_{i=0}^{7} v_i + \frac{1}{8}\left(\sum_{i=0}^{3} v_i - \sum_{i=4}^{7} v_i\right) - \frac{1}{4}(v_0 + v_1 - v_2 - v_3) + \frac{1}{2}(v_2 - v_3), \tag{2.6}$$

Formula (2.6) demonstrates the action of the forward and then the inverse Haar DWT. Noting the pattern of operations in formula (2.6) makes it easier to understand the origin of the more complicated formulas (2.9)–(2.16) in Section 2.2.

Note that formulas (2.1) and (2.2) can also be viewed as the convolution of an input sequence with finite impulse response filters $\frac{1}{2}(1, 1)$ and $\frac{1}{2}(1, -1)$. Most presentations of the Haar DWT use alternative filters where $\frac{1}{2}$ is replaced by $\frac{1}{\sqrt{2}}$ which makes the Haar DWT ortho*normal*.

2.2 THE HAAR-FISZ TRANSFORM

The Haar-Fisz transform of the vector $\mathbf{v} = (v_0, v_1, \ldots, v_{N-1})$ for $N = 2^J$, where $v_i \geq 0$ for all i is defined as follows.

1. Take the Haar DWT of $\mathbf{v}$ but with the modification that as each set of $\mathbf{s}^j$ and $\mathbf{d}^j$ is produced immediately define $\mathbf{f}^j$ by

$$f_i^j = \begin{cases} 0 & \text{if} \quad s_i^j = 0, \\ d_i^j / \sqrt{s_i^j} & \text{otherwise.} \end{cases} \tag{2.7}$$

2. Apply the inverse Haar DWT to the modified transform $(\mathbf{s}^J, \mathbf{f}^J, \mathbf{f}^{J-1}, \ldots, \mathbf{f}^1)$ to produce the vector $\mathbf{u}$.

For fixed i, j formula (2.7) is an example of the Fisz (1955) transform which we shall define and explain in Section 2.3. There is a one-to-one correspondence between vectors $\mathbf{v}$ and $\mathbf{u}$, and we will denote

$$\mathbf{u} = \mathcal{F}\mathbf{v}. \tag{2.8}$$

The nonlinear operator $\mathcal{F}$ defines the *Haar-Fisz transform*. A general formula for the Haar-Fisz transform appears in Fryzlewicz and Nason (2003, sec. 8). Because the only difference between Haar-Fisz and the Haar DWT transform are the $O(1)$ in-place coefficient modifications in (2.7), the computational speed and memory requirement of the Haar-Fisz transform is $O(N)$. The inverse Haar-Fisz transform simply reverses the above Steps 2 and 1: Apply the Haar DWT to $\mathbf{u}$ to produce $(\mathbf{s}^J, \mathbf{f}^J, \mathbf{f}^{J-1}, \ldots, \mathbf{f}^1)$ and then apply the inverse Haar DWT while undoing the effect of (2.7) as each scale j is produced. The inverse Haar-Fisz transform has the same computational complexity $O(N)$.

As an example we demonstrate the Haar-Fisz transform applied to the input vector $\mathbf{v}$ of length 8 now with all positive entries. The Haar-Fisz transform $\mathbf{u} = \mathcal{F}\mathbf{v}$ is given by

$$u_0 = \frac{\sum_{i=0}^7 v_i}{8} + \frac{\sum_{i=0}^3 v_i - \sum_{i=4}^7 v_i}{2\sqrt{2}\sqrt{\sum_{i=0}^7 v_i}} + \frac{v_0 + v_1 - (v_2 + v_3)}{2\sqrt{\sum_{i=0}^3 v_i}} + \frac{v_0 - v_1}{\sqrt{2}\sqrt{v_0 + v_1}}, \quad (2.9)$$

$$u_1 = \frac{\sum_{i=0}^7 v_i}{8} + \frac{\sum_{i=0}^3 v_i - \sum_{i=4}^7 v_i}{2\sqrt{2}\sqrt{\sum_{i=0}^7 v_i}} + \frac{v_0 + v_1 - (v_2 + v_3)}{2\sqrt{\sum_{i=0}^3 v_i}} - \frac{v_0 - v_1}{\sqrt{2}\sqrt{v_0 + v_1}}, \quad (2.10)$$

$$u_2 = \frac{\sum_{i=0}^7 v_i}{8} + \frac{\sum_{i=0}^3 v_i - \sum_{i=4}^7 v_i}{2\sqrt{2}\sqrt{\sum_{i=0}^7 v_i}} - \frac{v_0 + v_1 - (v_2 + v_3)}{2\sqrt{\sum_{i=0}^3 v_i}} + \frac{v_2 - v_3}{\sqrt{2}\sqrt{v_2 + v_3}}, \quad (2.11)$$

$$u_3 = \frac{\sum_{i=0}^7 v_i}{8} + \frac{\sum_{i=0}^3 v_i - \sum_{i=4}^7 v_i}{2\sqrt{2}\sqrt{\sum_{i=0}^7 v_i}} - \frac{v_0 + v_1 - (v_2 + v_3)}{2\sqrt{\sum_{i=0}^3 v_i}} - \frac{v_2 - v_3}{\sqrt{2}\sqrt{v_2 + v_3}}, \quad (2.12)$$

$$u_4 = \frac{\sum_{i=0}^7 v_i}{8} - \frac{\sum_{i=0}^3 v_i - \sum_{i=4}^7 v_i}{2\sqrt{2}\sqrt{\sum_{i=0}^7 v_i}} + \frac{v_4 + v_5 - (v_6 + v_7)}{2\sqrt{\sum_{i=4}^7 v_i}} + \frac{v_4 - v_5}{\sqrt{2}\sqrt{v_4 + v_5}}, \quad (2.13)$$

$$u_5 = \frac{\sum_{i=0}^7 v_i}{8} - \frac{\sum_{i=0}^3 v_i - \sum_{i=4}^7 v_i}{2\sqrt{2}\sqrt{\sum_{i=0}^7 v_i}} + \frac{v_4 + v_5 - (v_6 + v_7)}{2\sqrt{\sum_{i=4}^7 v_i}} - \frac{v_4 - v_5}{\sqrt{2}\sqrt{v_4 + v_5}}, \quad (2.14)$$

$$u_6 = \frac{\sum_{i=0}^7 v_i}{8} - \frac{\sum_{i=0}^3 v_i - \sum_{i=4}^7 v_i}{2\sqrt{2}\sqrt{\sum_{i=0}^7 v_i}} - \frac{v_4 + v_5 - (v_6 + v_7)}{2\sqrt{\sum_{i=4}^7 v_i}} + \frac{v_6 - v_7}{\sqrt{2}\sqrt{v_6 + v_7}}, \quad (2.15)$$

$$u_7 = \frac{\sum_{i=0}^7 v_i}{8} - \frac{\sum_{i=0}^3 v_i - \sum_{i=4}^7 v_i}{2\sqrt{2}\sqrt{\sum_{i=0}^7 v_i}} - \frac{v_4 + v_5 - (v_6 + v_7)}{2\sqrt{\sum_{i=4}^7 v_i}} - \frac{v_6 - v_7}{\sqrt{2}\sqrt{v_6 + v_7}}. \quad (2.16)$$

The structure of the underlying Haar inverse DWT can clearly be seen in the formula. In particular, compare the formula for u_2 above with formula (2.6). Formulas (2.9) to (2.16) are a special case of the general formula for $\mathcal{F}$ given in Fryzlewicz and Nason (2003, sec. 8).

2.3 RATIONALE BEHIND THE HAAR-FISZ TRANSFORM

The Haar-Fisz transform possesses Gaussianizing and variance stabilization properties stemming from the following result which is a special case of the theorem in Fisz (1955).

Theorem 1. *(**Fisz**) Let $X_i \sim Pois(\lambda_i)$ for $i = 1,2$ and X_1, X_2 independent. Define the function $\zeta : \mathbb{R}^2 \to \mathbb{R}$ by*

$$\zeta(X_1, X_2) = \begin{cases} 0 & \text{if} \quad X_1 = X_2 = 0, \\ (X_1 - X_2)/(X_1 + X_2)^{\frac{1}{2}} & \text{otherwise.} \end{cases} \tag{2.17}$$

If $(\lambda_1, \lambda_2) \to (\infty, \infty)$ and $\lambda_1/\lambda_2 \to 1$ then $\zeta(X_1, X_2) - \zeta(\lambda_1, \lambda_2) \xrightarrow{d} N(0, 1)$.

We call ζ the *Fisz transform*. We illustrate the use of Fisz's theorem in our algorithm using the specific example in (2.9)–(2.16). Each summand, apart from the first (the mean of all data), in formulas (2.9)–(2.16) is of the form $a\zeta(X_1, X_2)$ where X_i are independent Poisson with mean λ_i, $i = 1, 2$ and $a > 0$. If λ_1 and λ_2 are large and close then, by Fisz's theorem, each summand will be close to normal, with variance of a^2 and mean of $a\zeta(\lambda_1, \lambda_2)$.

In general, since we are dealing with inhomogeneous intensities with means λ_1, λ_2 which are not always large and close, it would be desirable to gain some insight into

- how well the Fisz transform can Gaussianize and stabilize variance, and
- how well we can determine the mean, that is, how close $E\zeta(X_1, X_2)$ is to $\zeta(\lambda_1, \lambda_2)$,

for a whole range of λ_i. These issues would be challenging to investigate theoretically. However, to cast some light we performed the following simulation experiment. We chose values of λ_i to range from 1 to 40 in steps of 1. For each pair (λ_1, λ_2) we draw 10^5 values of $\zeta(X_1, X_2)$ as defined by (2.17) and denote the sample by $\mathbf{z}(\lambda_1, \lambda_2)$. For a comparison of Gaussianization we also compute Anscombe's transform, as mentioned in Section 1, to the X_i which arises from the larger λ_i (this comparison is charitable to Anscombe: either X_1 or X_2 could be used but Anscombe works better for larger intensities).

Figure 1 gives some idea of how well the Fisz transform Gaussianizes, stabilizes variance and how close $\bar{\mathbf{z}}(\lambda_1, \lambda_2)$ is to $\zeta(\lambda_1, \lambda_2)$. The top left figure shows that Fisz is always "more Gaussian" than Anscombe. The top right figure merely shows that $\bar{\mathbf{z}}(\lambda_1, \lambda_2)$ is very close to $\zeta(\lambda_1, \lambda_2)$. The bottom row of Figure 1 shows that the variance of $\mathbf{z}(\lambda_1, \lambda_2)$ is stable and close to one for a wide range of (λ_1, λ_2). To summarize, the above experiment shows that $\zeta(X_1, X_2)$, the Fisz transform of X_1 and X_2, can be thought of as an approximately Gaussian variable with mean $\zeta(\lambda_1, \lambda_2)$ and variance bounded above by (and close to) one.

The above discussion concentrates on the distributional properties of individual Fisz transformed variates. However, it is also of interest to ascertain how well the fully Haar-Fisz transformed variables $\mathbf{u}$ from (2.8) are Gaussianized and variance-stabilized. Another property we desire of the Haar-Fisz transform is that it should not unduly introduce correlation in $\mathbf{u}$: especially when none existed in $\mathbf{v}$. In an attempt to empirically investigate the degree of (de)correlation, Gaussianization and variance stabilization ("D-G-S") of the Haar-Fisz transform we carried out an extensive simulation study which is fully documented in Fryzlewicz and Nason (2003) but we briefly summarize the conclusions here.

1. The degree of D-G-S was strikingly similar for sample sizes of $N = 128$ and $N = 1{,}024$: we suspect that the degree of D-G-S is not strongly dependent on N. We consider $N = 128$ to be a short vector in this situation.
2. The greater the minimum of the intensity vector, $\boldsymbol{\lambda}$, the higher the degree of D-G-S.

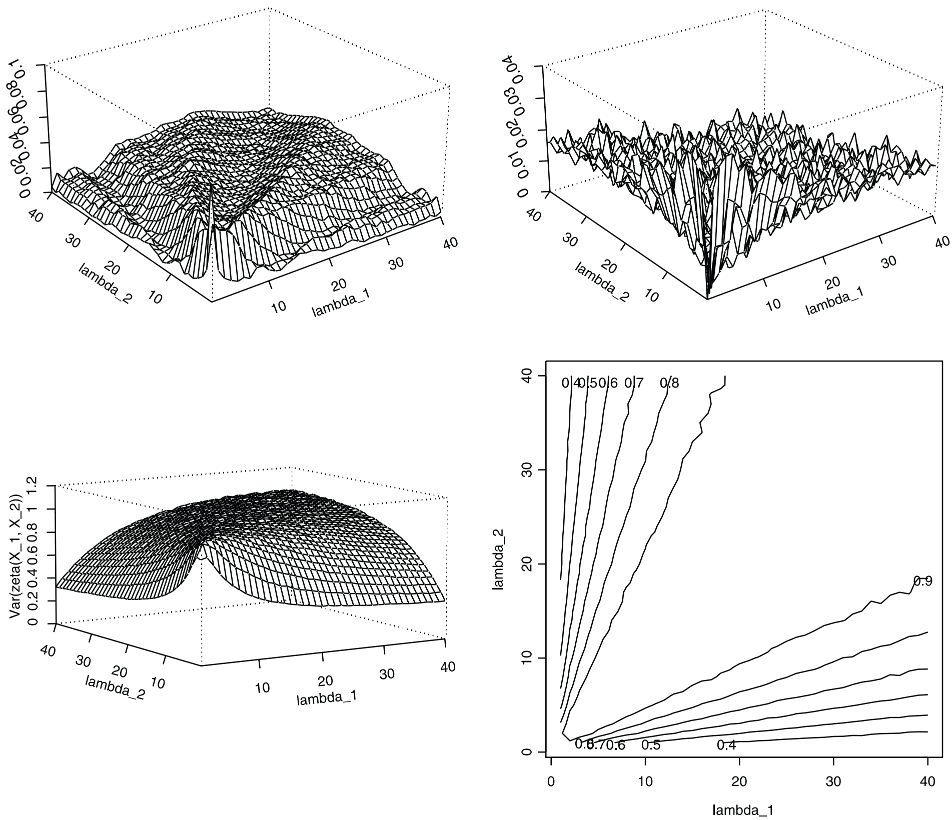

Figure 1. Top left: Difference between Kolmogorov-Smirnov test statistics computed on Anscombe-transformed Poisson variables with intensity max(λ_1, λ_2)*, and* $\mathbf{z}(\lambda_1, \lambda_2)$*. Positive difference means that Haar-Fisz is closer to Gaussian. Top right:* $|\bar{\mathbf{z}}(\lambda_1, \lambda_2) - \zeta(\lambda_1, \lambda_2)|$*. Bottom left (and right): perspective (and contour) plot of var*$(\mathbf{z}(\lambda_1, \lambda_2))$*.*

For constant intensities D-G-S is *extremely* effective from about $\boldsymbol{\lambda} = 4$.

3. For nonconstant intensities, the degree of D-G-S depends not only on $\min \boldsymbol{\lambda}$ but also on the length of the stretch where the intensity is equal, or close, to $\min \boldsymbol{\lambda}$. The shorter the stretch, the lower the "acceptable" value of $\min \boldsymbol{\lambda}$ for which D-G-S is still very effective. For example, if the intensity is at its constant minimum, 2, for 25% of the time and the remaining intensity is constant at 10, then the D-G-S is extremely effective.

The following example compares the D-G-S properties of the Haar-Fisz transform, $\mathcal{F}$, Anscombe's transform, $\mathcal{A}$, and the identity transform. Let us consider the intensity as in the top plot of Figure 2 (a rescaled and shifted version of the Donoho and Johnstone (1994) **bumps** function). This intensity vector will be denoted by $\boldsymbol{\lambda}$, and $\mathbf{v}$ will denote a sample path generated from it.

Figure 2 compares the Q-Q plots of $\mathbf{v} - \boldsymbol{\lambda}$, $\mathcal{A}\mathbf{v} - \mathcal{A}\boldsymbol{\lambda}$, and $\mathcal{F}\mathbf{v} - \mathcal{F}\boldsymbol{\lambda}$ averaged over 100 samples of $\mathbf{v}$. Clearly, the Q-Q plot shows that the Haar-Fisz transformation does a better job

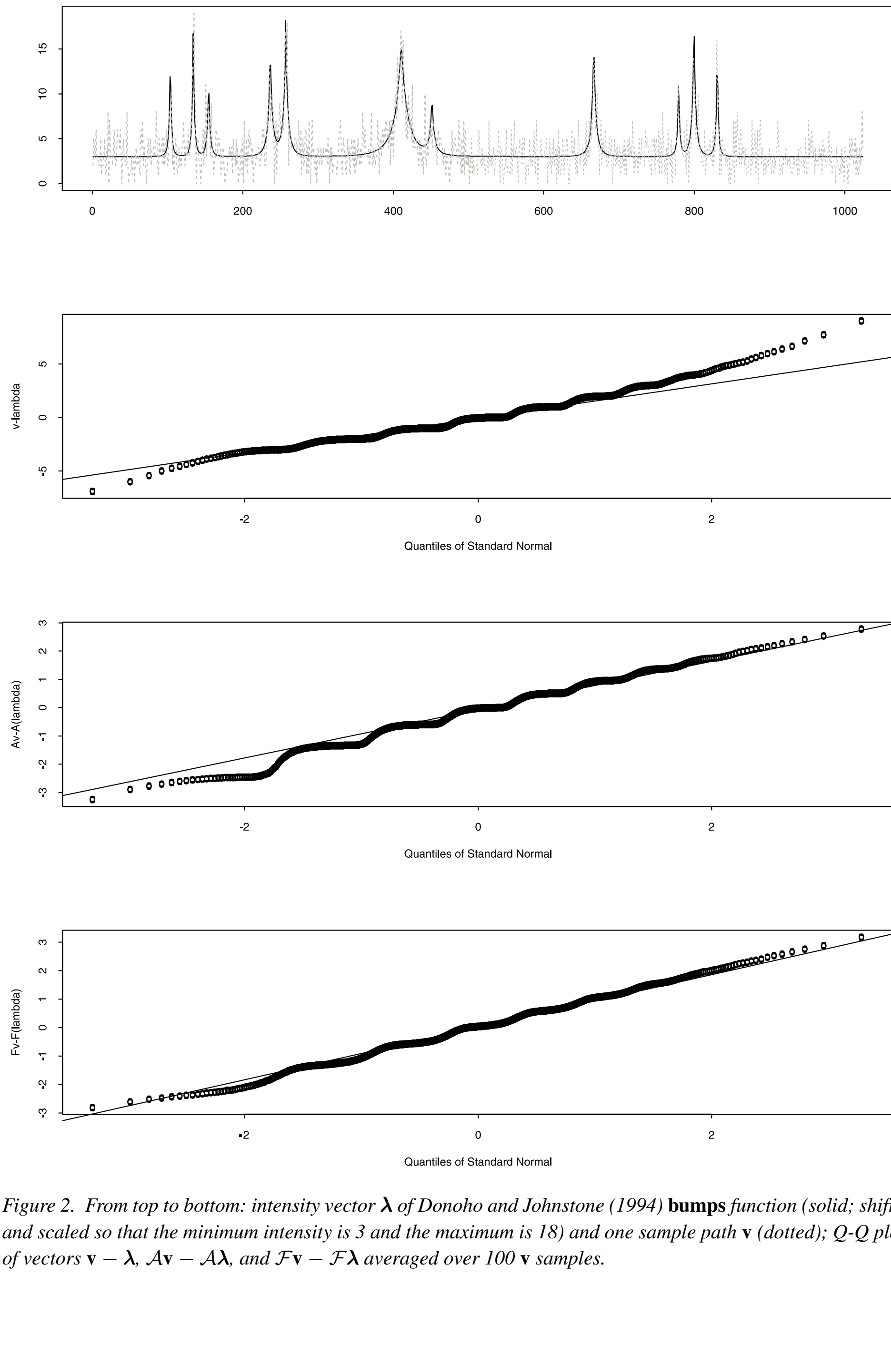

Figure 2. From top to bottom: intensity vector $\boldsymbol{\lambda}$ of Donoho and Johnstone (1994) **bumps** *function (solid; shifted and scaled so that the minimum intensity is 3 and the maximum is 18) and one sample path $\mathbf{v}$ (dotted); Q-Q plots of vectors $\mathbf{v} - \boldsymbol{\lambda}$, $\mathcal{A}\mathbf{v} - \mathcal{A}\boldsymbol{\lambda}$, and $\mathcal{F}\mathbf{v} - \mathcal{F}\boldsymbol{\lambda}$ averaged over 100 $\mathbf{v}$ samples.*

in "Gaussianization." In particular, the Haar-Fisz transformed data are less "stepped" and look more like variates from a continuous distribution than a discrete one. The Anscombe-transformed data appear more "stepped" for lower quantiles than for higher ones. Further, the tails for Haar-Fisz are more normal than for Anscombe which in turn is more normal than the raw count data.

The top two plots in Figure 3 show the squared residual $\mathcal{A}\mathbf{v} - \mathcal{A}\boldsymbol{\lambda}$ for Anscombe and $\mathcal{F}\mathbf{v} - \mathcal{F}\boldsymbol{\lambda}$ for Haar-Fisz averaged over 1,000 samples. It can be seen that both transforms stabilize variance around 1 very well. Further extensive simulations, reported by Fryzlewicz and Nason (2003), show that Haar-Fisz stabilizes variance extremely well, in many cases much better than Anscombe. In particular, when the intensity is larger than three both Haar-Fisz and Anscombe perform well. Haar-Fisz is also good for intensities 2–3. However, for low intensities, less than 2, neither Haar-Fisz nor Anscombe stabilize very well, but Haar-Fisz always does better.

The bottom plot in Figure 3 shows the sample autocorrelation of $\mathcal{F}\mathbf{v} - \mathcal{F}\boldsymbol{\lambda}$ averaged over 100 samples and demonstrates that there is little autocorrelation in the Haar-Fisz transformed variates. This is a common feature provided the vector is well Gaussianized: see conclusions 1–3 above and see also Fryzlewicz and Nason (2003) for further simulation results.

Finally, in this section, we quote two propositions for the operator $\mathcal{F}$ for constant intensities [the proofs appear in Fryzlewicz and Nason (2003, sec. 8)]. In actuality we are interested in nonconstant intensities but as mentioned earlier such theory is beyond the scope of the present article. Our propositions provide some reassurance that using denoisers for signals contaminated with uncorrelated Gaussian is not an unreasonable thing to do (even for nonconstant intensities, particularly those that are slowly changing or piecewise constant). This will be confirmed later by our simulation results in Section 3.

Proposition 1 says that the coefficients of the Haar-Fisz transformed vector of Poisson counts are asymptotically uncorrelated.

Proposition 1. *Let* $\mathbf{v} = (v_0, v_1, \ldots, v_{N-1})$ *be a vector of iid Poisson variables with mean* λ, *and let* N *be an integer power of two. Let* $\mathbf{u} = \mathcal{F}\mathbf{v}$. *For* $m \neq n$, *we have*

$$cor(u_m, u_n) \to 0 \quad as \quad \lambda \to \infty \quad and \quad \lambda/N \to 0. \tag{2.18}$$

Proposition 2 says that these coefficients are also asymptotically normal with variance one.

Proposition 2. *Let* $\mathbf{v} = (v_0, v_1, \ldots, v_{N-1})$ *be a vector of iid Poisson variables with mean* λ, *and let* N *be an integer power of two. Let* $\mathbf{u} = \mathcal{F}\mathbf{v}$. *For all* $n = 0, 1, \ldots, N-1$, *we have*

$$u_n - \lambda = \nu + Y_n, \tag{2.19}$$

where

$$\begin{aligned} \nu &\xrightarrow{d} 0 \quad as \quad \lambda/N \to 0 \\ Y_n &\xrightarrow{d} N(0,1) \quad as \quad (\lambda, N) \to (\infty, \infty). \end{aligned} \tag{2.20}$$

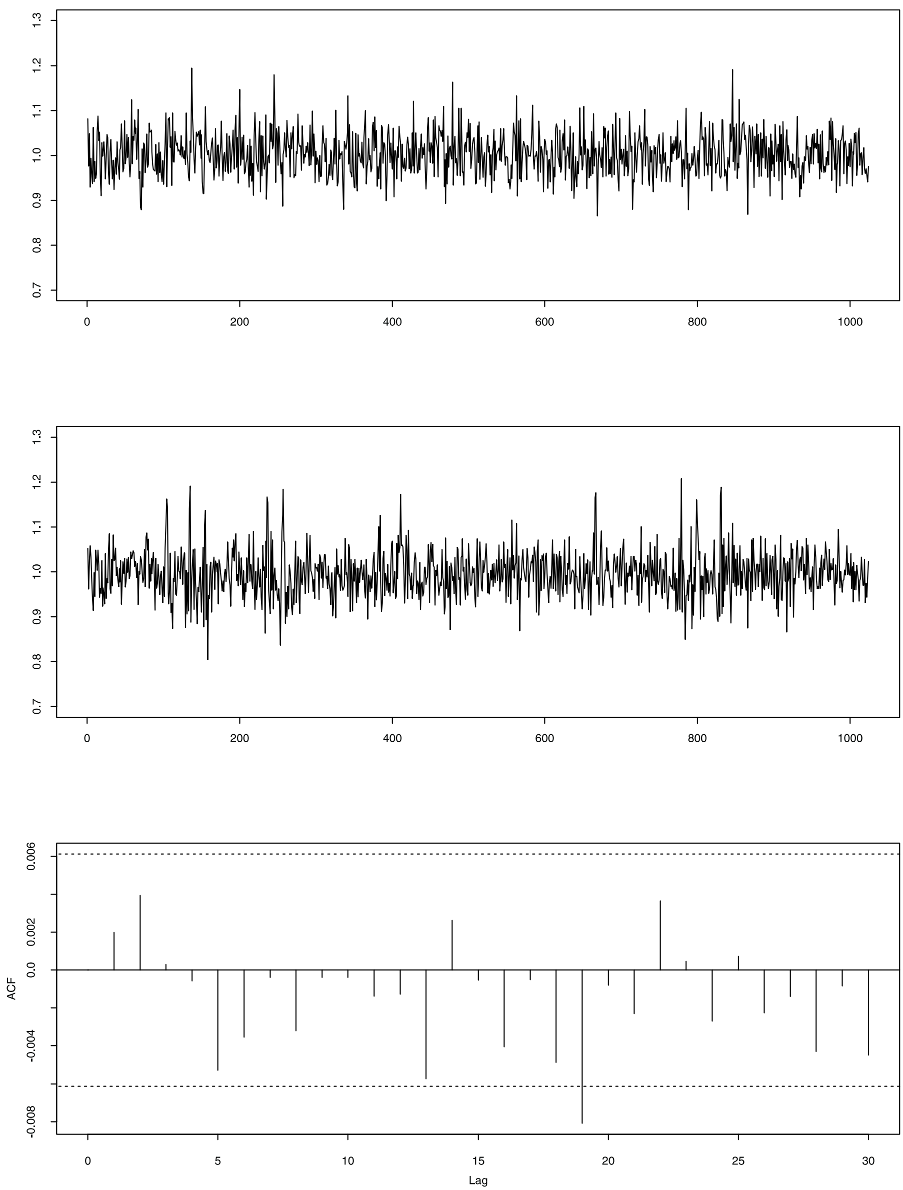

Figure 3. Top and middle: $|\mathcal{A}\mathbf{v} - \mathcal{A}\boldsymbol{\lambda}|^2$, $|\mathcal{F}\mathbf{v} - \mathcal{F}\boldsymbol{\lambda}|^2$ *both averaged over 1,000 samples of* $\mathbf{v}$. *Bottom: acf of* $\mathcal{F}\mathbf{v} - \mathcal{F}\boldsymbol{\lambda}$ *averaged over 100 sample paths.*

3. POISSON INTENSITY ESTIMATION

We propose the following core algorithm for estimating the intensity $\boldsymbol{\lambda}$ of a Poisson process:

[A1] Given the vector $\mathbf{v}$ of Poisson observations, preprocess it using the Haar-Fisz transformation to obtain $\mathcal{F}\mathbf{v}$.

[A2] Denoise $\mathcal{F}\mathbf{v}$ using *any* suitable ordinary wavelet denoising technique, appropriate for Gaussian noise (i.e., DWT—thresholding—inverse DWT). Denote the smoothed version of $\mathcal{F}\mathbf{v}$ by $\widehat{\mathcal{F}\boldsymbol{\lambda}}$. We can optionally exploit the fact that the asymptotic variance of the noise is equal to one.

[A3] Perform the inverse Haar-Fisz transform to obtain $\mathcal{F}^{-1}(\widehat{\mathcal{F}\boldsymbol{\lambda}})$ and take it to be the estimate of the intensity.

Section 3.1 describes a range of methods for Poisson intensity estimation including our new ones. Section 3.2 evaluates the performance of our Poisson intensity estimation algorithm and compares it to existing techniques on a variety of test intensities.

3.1 METHODS FOR POISSON INTENSITY ESTIMATION

Existing methods. As mentioned in the introduction, the Bayesian methods due to Kolaczyk (1999a) and Timmermann and Nowak (1997, 1999) are currently state-of-the-art; see Besbeas, De Feis, and Sapatinas (2002). Our simulation study compared our technique with these Bayesian methods, as well as with the computationally intensive l_1-penalized likelihood technique of Sardy, Antoniadis, and Tseng (2004) and with a choice of methods based on the Anscombe transformation [to save space, for the latter two see Fryzlewicz and Nason (2003)]. To compare our technique with Kolaczyk (1999a) we used Kolaczyk's *BMSMShrink* MATLAB software. As we did not have access to TN's software we exactly reproduced the simulation setup as in Timmermann and Nowak (1999) and compared our results to their Tables I and II. [Incidentally, the methods in Kolaczyk (1999a) and Timmermann and Nowak (1999) are very similar: the underlying Bayesian model is exactly the same, although the hyperparameter estimation is slightly different (E. D. Kolaczyk, personal communication, 2001)].

Our method. The following describes the common features for our Poisson intensity estimation.

1. All our techniques always involve the Haar-Fisz transform, **[A1]**, of the data, and the inverse Haar-Fisz transform, **[A3]**.
2. In step **[A2]** of our algorithm the wavelet denoising technique may be of a translation invariant (TI) transform type; see Coifman and Donoho (1995). We refer to TI-denoising at this stage as "internal" cycle spinning (CS).
3. In step **[A2]** we could use any one of a number of wavelet families [e.g., multiwavelet, see Downie and Silverman (1998), complex-valued, see Lina (1997) etc.] for the denoising. In our simulations below we use Haar wavelets and Daubechies least-asymmetric wavelets of order 10; see Daubechies (1992).

4. Let $\mathcal{S}$ be the shift-by-one operator from Nason and Silverman (1995). The Haar-Fisz transform is not translation-equivariant because $\mathcal{FS} \neq \mathcal{SF}$. This noncommutativity implies that it is beneficial to apply CS to the whole algorithm **[A1]**–**[A3]** even if **[A2]** uses a TI technique. We call this "external" CS.
   Due to the particular type of nonlinearity of the Haar-Fisz transform there is no fast $\mathcal{O}(N \log N)$ algorithm for the external CS. Therefore, we implement external CS by actually shifting the data before **[A1]**, shifting back the estimate after **[A3]**, and averaging over the estimates obtained through several different shifts.
   For a dataset of length N there are N possible shifts. However, through empirical investigation detailed by Fryzlewicz and Nason (2003), we have found that 50 shifts are enough for data of length $\leq 1{,}024$. We postulate that using more shifts for longer datasets is likely to be beneficial.

Note that there is no point in doing external CS with the Anscombe transformation, $\mathcal{A}$, provided one has carried out internal CS, since Anscombe's transformation commutes with the shift operator: $\mathcal{AS} = \mathcal{SA}$.

The following list labels and describes the wavelet denoising methods that we choose to use in **[A2]**. In each case **F**$\bowtie$ denotes the use of the Haar-Fisz transform and its inverse.

F$\bowtie$**U:** Universal hard thresholding from Donoho and Johnstone (1994) as implemented in *WaveThresh* (Nason 1998) with default parameters (e.g., uses MAD variance estimation on all coefficients). No CS is performed.

F$\bowtie$**CV:** Cross-validation method from Nason (1996) as implemented in *WaveThresh* using default parameters but hard thresholding. No CS is performed.

F$\bowtie$**BT:** A variant of the greedy tree algorithm from Baraniuk (1999). No CS is performed.

F$\bowtie$**U CS, F**$\bowtie$**CV CS, F**$\bowtie$**BT CS:** As above, but supplemented with external CS (50 shifts).

Hybrids. We also looked at the performance of certain hybrid methods. These estimate the intensity by averaging the results of two of the above Haar-Fisz methods. Our main hybrid, **H:CV+BT CS**, combines **F**$\bowtie$**CV CS** and **F**$\bowtie$**BT CS**. Note that hybrids can be easily formulated due to the large number of methods available for denoising Gaussian contaminated signals.

During our investigations we made use of several other denoisers including the eBayes procedure as described by Johnstone and Silverman (2001) using software kindly supplied by Bernard Silverman; universal hard threshold with internal cycle-spinning; and hybrids of these with **F**$\bowtie$**CV CS**.

3.2 SIMULATION RESULTS FOR STANDARD TEST FUNCTIONS

Simulated data, $\mathbf{v}$, for the model described in the first paragraph of the Section 1, can be easily obtained: fix the known intensities λ_n and then draw a sequence of Poisson variables v_n each with known intensity λ_n.

The simulation setup in this section is the same as that described by Timmermann and Nowak (1999) and the results here can be directly compared. The results cited here as

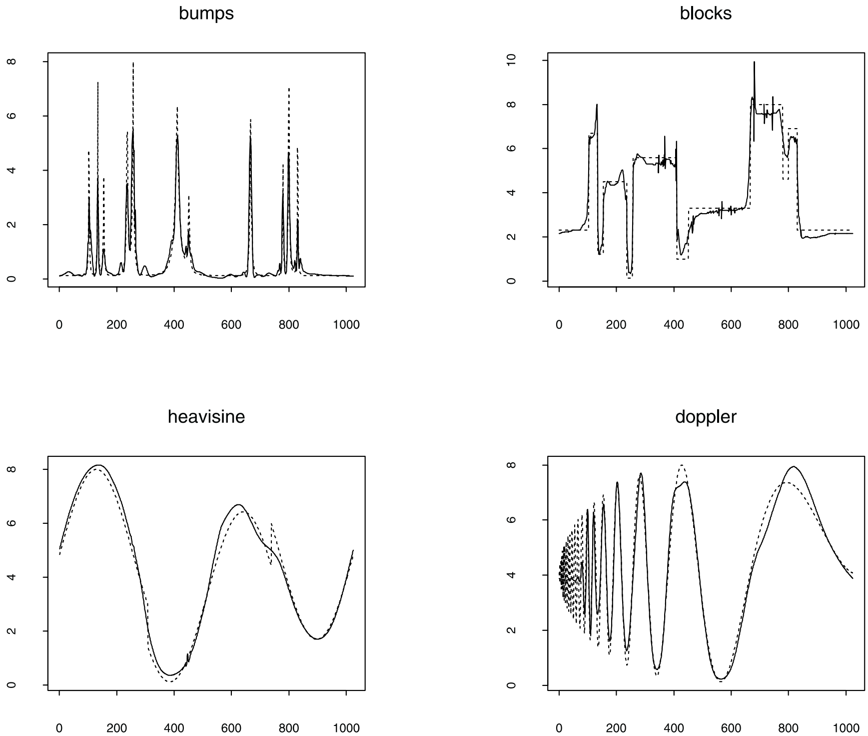

Figure 4. Selected estimates for the Donoho and Johnstone intensity functions (dashed, described in text). Each estimate gives an idea of "average" performance in that in each case its MISE is the closest to the median MISE obtained over 50 sample paths. The estimation method in each case was **F ⋈ U CS** *with Daubechies least-asymmetric wavelets with 10 vanishing moments except for* **blocks** *which used* **H:CV+BT CS** *with Haar wavelets.*

BAYES are taken from and refer to the Bayesian method developed in Timmermann and Nowak (1999) obtained over 25 independent simulations. The study by Timmermann and Nowak (1999) obtains two sets of intensity functions of length $N = 1{,}024$ from the test functions from Donoho and Johnstone (1994). Each set is obtained by shifting and scaling to achieve (min,max) intensities of $(1/8, 8)$ and $(1/128, 128)$. The true intensity functions for the $(1/8, 8)$ case are shown as dashed lines in Figure 4. All the results in this section on our methods and BMSMShrink are based on 100 independent simulations.

The results reported in Table 1 are the MISE normalized by the squared l_2 norm of the true intensity vector, multiplied by 10,000 and then rounded for clarity of presentation [this is exactly the same performance measure as in Timmermann and Nowak (1999) which is useful for comparability].

The results show that our **F ⋈ U CS** method with the LA10 wavelet outperforms the existing state-of-the-art methods especially for the lower intensity, except for the **blocks** function. The main reason why its performance for **blocks** is less impressive is that a smooth wavelet is used in the Gaussian denoising step **[A2]**. As expected, the performance of **F ⋈ U CS** with the Haar wavelet is much better in this context, but still not as good as that of *BMSMShrink*, which is the better of the two Bayesian competitors for **blocks**. However,

Table 1. Normalized MISE Values ($\times$ 10,000) for Existing Bayesian Techniques and our **F $\bowtie$ U**, **F $\bowtie$ U CS** and **H:CV + BT CS** Methods Using Haar Wavelets and Daubechies' Least Asymmetric Wavelets with 10 Vanishing Moments (LA10), on the Test Functions with Peak Intensities 8 and 128. The best results are indicated by bold.

			F $\bowtie$ U		**F $\bowtie$ U CS**		**H:CV+BT CS**
Intensity	*BAYES*	*BMSMShrink*	*Haar*	*LA10*	*Haar*	*LA10*	*Haar*
			Peak intensity = 8				
Doppler	154	146	380	181	201	**99**	159
Blocks	178	**129**	374	450	191	302	135
HeaviSine	52	46	214	79	68	**40**	64
Bumps	1475	1871	3245	1892	2826	**1268**	2266
			Peak intensity = 128				
Doppler	26	20	92	24	29	**12**	23
Blocks	27	8	22	64	8	37	**7**
HeaviSine	**7**	**7**	35	13	9	**7**	9
Bumps	143	174	357	211	185	**133**	163

the hybrid method **H:CV+BT CS** with the Haar wavelet achieves performance comparable to *BMSMShrink*. We should emphasize here that our **F $\bowtie$ U CS** method is far simpler to implement than the current state-of-the-art techniques. The non-CS results for **F $\bowtie$ U** are reported for information only as both *BAYES* and *BMSMShrink* use cycle-spinning.

In Figure 4 the small spike in the **heavisine** function is not picked up well at intensity 8 but is almost always clearly estimated at intensity 128 (not shown). However, it should be said that the spike is almost completely obscured by noise in all realizations at intensity 8 so it would be extremely difficult for any method to detect it. We are impressed with the quality of the estimates using the new Haar-Fisz method, particularly with **bumps** and **doppler**. Also, the reconstruction of **blocks**, using the hybrid method **H:CV+BT CS**, is very accurate. Overall, it must be remembered that the reconstructions are usually going to be less impressive than the classical wavelet shrinkage problem where the test functions are contaminated with *Gaussian* noise with variance one.

We also performed an indirect comparison with the computationally intensive l_1-penalized likelihood method proposed by Sardy, Antoniadis, and Tseng (2004) using results presented therein, as well as a direct comparison with methods using the Anscombe transform but otherwise constructed in exactly the same way as **F $\bowtie$ U** and **F $\bowtie$ U CS**. Although our technique significantly outperformed these competitors for peak intensity 8, its performance was almost identical to the Anscombe-based method for peak intensity 128. Also, the method of Sardy, Antoniadis, and Tseng (2004) based on Haar wavelets turned out to be slightly superior to our technique for the **bumps** function for peak intensity 128: the average MISE achieved by their method was about 8% lower. This investigation was detailed by Fryzlewicz and Nason (2003).

4. APPLICATION TO EARTHQUAKE DATA

In this section, we analyze Northern Californian earthquake data, available from:

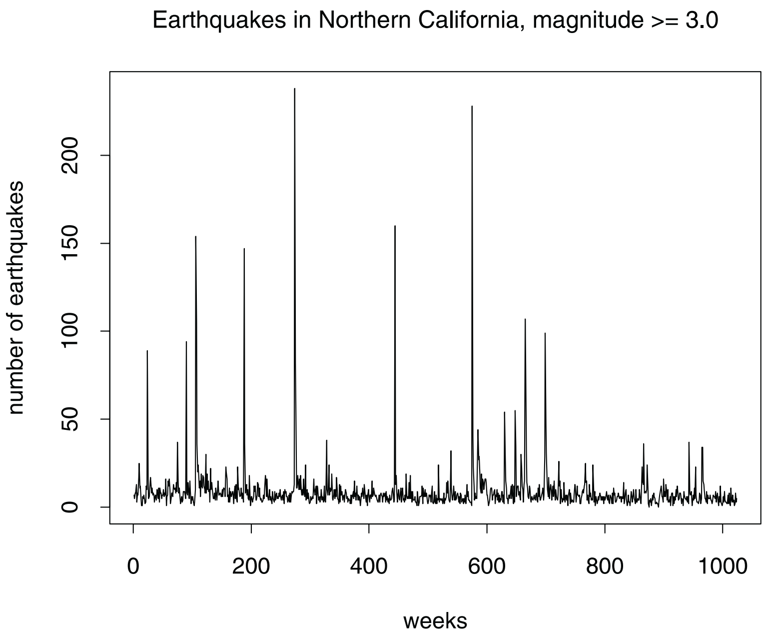

Figure 5. The number of earthquakes of magnitude $\geq$ 3.0 which occurred in Northern California in 1,024 consecutive weeks, the last week being 29 November–5 December 2000.

http://quake.geo.berkeley.edu. We analyze the time series N_k, $k = 1, \ldots, 1{,}024$, where N_k is the number of earthquakes of magnitude 3.0 or more which occurred in the kth week, the last week under consideration being 29 November–5 December 2000. The time series, imported into S-Plus, is plotted in Figure 5.

Our aim is to extract the intensity which underlies the realization of this process. For the purposes of this example we shall use the *BMSMShrink* methodology of Kolaczyk (1999a) and our hybrid **H:CV+BT CS** method with Haar wavelets. The rationale for using **H:CV+BT CS** is that:

- it appears that the true earthquake intensity is highly nonregular and **H:CV+BT CS** with Haar wavelets worked the best on the **blocks** simulation example from the previous section;
- the earthquake data exhibits medium to high intensities and **H:CV+BT CS** was better than the other hybrids that we tried in this situation.

Figure 6 shows the intensity estimates obtained using *BMSMShrink* and **H:CV+BT CS** plotted on a log scale. (Due to the large peak at 274 weeks the original scale of 0–250 is not suitable for analyzing the subtle differences between the estimates.) Visually the estimates are very similar however the **H:CV+BT CS** estimate is a little less variable. Although with this real data there is clearly no right or wrong answer it is reassuring that they do give such similar visual results even though *BMSMShrink* and **H:CV+BT CS** are based on completely different philosophies.

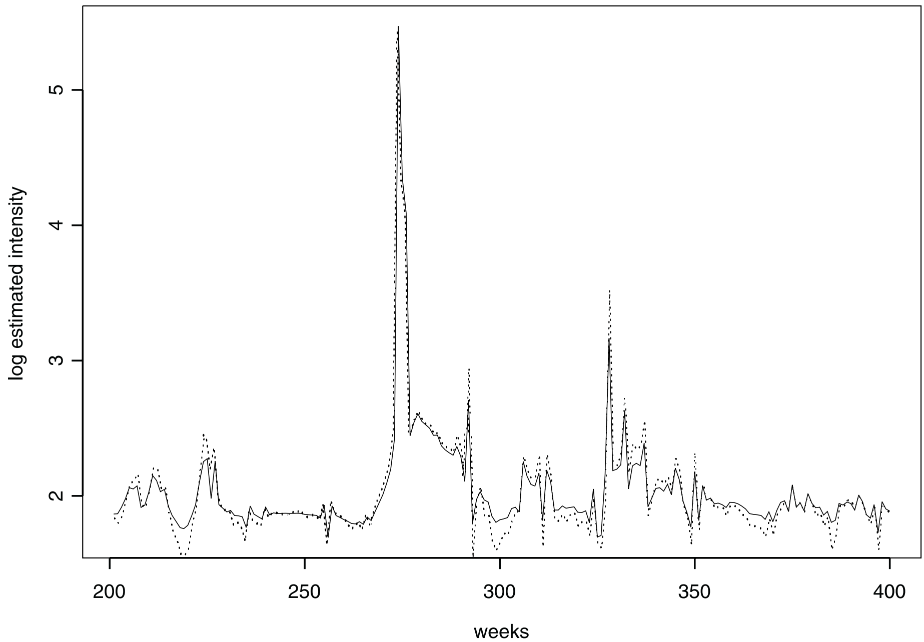

Figure 6. Intensity estimates for earthquake data for weeks 201 to 400. Dotted line is BMSMShrink estimate and solid is **H:CV+BT CS** *estimate.*

5. CONCLUSIONS

This article describes a new wavelet-based technique for bringing vectors of Poisson counts to normality with variance one. The technique, named the Haar-Fisz transformation, was applied to estimating the intensity of an inhomogeneous Poisson process, yielding a method whose performance was nearly always better than that of the current state-of-the-art.

For Poisson intensity estimation our methodology requires two components. The first, the Haar-Fisz transform, is very simple and easy to code. The second component can be any suitable Gaussian denoising procedure: we have used and compared a variety of wavelet methods ranging from the fast universal thresholding to more complicated techniques such as cross-validation, Baraniuk trees, and empirical Bayes. Because *any* Gaussian denoiser can be used, the Haar-Fisz algorithm can only improve as the field develops.

If computational speed is not an issue, and little is known about the smoothness of the true intensity, we recommend that several denoisers be used and a hybrid averaging all of their results, with optional full cycle-spinning, be considered. However, if speed is important, then there is an issue over which one denoiser should be chosen: not all denoisers are appropriate for all types of intensity as our earlier simulations confirmed. Our recommendation is that if one suspects the intensity is piecewise constant then one should use Haar wavelets and a hybrid method such as **H:CV+BT CS**; otherwise, we strongly recommend the use of **F** $\bowtie$ **U CS** with a smooth wavelet.

We believe that one of the reasons why the performance of the Haar-Fisz algorithm is so good is due to the non-commutativity of the Fisz and shift operators, which enables meaningful cycle spinning. Also, the Fisz transform itself is a more effective normalizer than Anscombe.

Future ideas. The Haar-Fisz transform has the potential to "Gaussianize" a wide range of other noise distributions such as binomial, Gamma, and negative binomial. As χ^2 is a special case of the Gamma distribution we believe that the Haar-Fisz algorithm could be a potentially useful variance-stabilizer for a putative (wavelet) periodogram smoothing technique for (locally) stationary time series analysis. See Priestley (1981) for details on periodogram smoothing for stationary time series and Nason, von Sachs, and Kroisandt (2000) for smoothing of wavelet periodograms of locally stationary wavelet processes.

Also, we note that our algorithm (1) can be easily extended to more than one dimension to handle, for example, images; (2) can be used in conjunction with denoisers other than wavelets; and (3) could profit from a theoretical study of the properties of the Haar-Fisz transform for inhomogeneous intensities. These extensions are avenues for further research.

Software. The S-Plus routines written and used by us can be downloaded from the associated Web page: http://www.stats.bris.ac.uk/~mapzf/Poisson/Poisson.html.

ACKNOWLEDGMENTS

We are grateful to the associate editor and three referees for providing insightful comments and helpful suggestions. We would like to thank Eric Kolaczyk for kindly suppying his *BMSMShrink* software. Thanks also to Anestis Antoniadis, Eric, Rob Nowak, Theofanis Sapatinas, and Kamila Żychaluk for helpful discussions. Piotr Fryzlewicz was supported by a University of Bristol Research Scholarship, an ORS award and Unilever Research. Guy Nason gratefully acknowledges support from EPSRC grants GR/M10229/01 and GR/A01664/01.

[Received June 2002. Revised June 2003.]

REFERENCES

Anscombe, F. J. (1948), "The Transformation of Poisson, Binomial and Negative-Binomial Data," *Biometrika*, 35, 246–254.

Antoniadis, A., and Sapatinas, T. (2001), "Wavelet Shrinkage for Natural Exponential Families with Quadratic Variance Functions," *Biometrika*, 88, 805–820.

Baraniuk, R.G. (1999), "Optimal Tree Approximation with Wavelets," in *Wavelet Applications in Signal and Image Processing VII*, Proceedings of SPIE, eds. M. A. Unser, A. Aldroubi, and A. F. Laine, 3813, pp. 206–214.

Besbeas, P., De Feis, I., and Sapatinas, T. (2002), "A Comparative Simulation Study of Wavelet Shrinkage Estimators for Poisson Counts," *Rapporto Tecnico*, 240/02, Istituto per le Applicazioni del Calcolo "Mauro Picone"—Sezione di Napoli, Consiglio Nazionale delle Ricerche, Italy.

Brillinger, D. R. (1998), "Some Wavelet Analyses of Point Process Data," in *Conference Record of the Thirty-First Asilomar Conference on Signals, Systems and Computers, 1997*, 2, pp. 1087–1091.

Coifman, R. R., and Donoho, D. L. (1995), "Translation-Invariant De-noising," in *Wavelets and Statistics*, eds. A. Antoniadis and G. Oppenheim, New York: Springer-Verlag, pp. 125–150.

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

Donoho, D. L. (1993), "Nonlinear Wavelet Methods for Recovery of Signals, Densities, and Spectra from Indirect and Noisy Data," in *Proceedings of Symposia in Applied Mathematics*, 47, Providence, RI: American Mathematical Society.

Donoho, D. L., and Johnstone, I. M. (1994), "Ideal Spatial Adaptation by Wavelet Shrinkage," *Biometrika*, 81, 425–455.

Downie, T., and Silverman, B. (1998), "The Discrete Multiple Wavelet Transform and Thresholding Methods," *IEEE Transactions on Signal Processing*, 46, 2558–2561.

Fisz, M. (1955), "The Limiting Distribution of a Function of Two Independent Random Variables and its Statistical Application," *Colloquium Mathematicum*, 3, 138–146.

Fryzlewicz, P., and Nason, G. P. (2003), "Some Comments on the Haar-Fisz Algorithm for Poisson Intensity Estimation," Technical Report 03:04, Statistics Group, Department of Mathematics, University of Bristol. Available at http://www.amstat.org/publications/jcgs/ in the data archive section.

Johnstone, I. M., and Silverman, B. W. (2001), "Risk Bounds for Empirical Bayes Estimates of Sparse Sequences With Applications to Wavelet Smoothing," unpublished manuscript.

Kolaczyk, E. D. (1997), "Non-parametric Estimation of Gamma-ray Burst Intensities using Haar Wavelets," *The Astrophysical Journal*, 483, 340–349.

Kolaczyk, E. D. (1999a), "Bayesian Multiscale Models for Poisson Processes," *Journal of the American Statistical Association*, 94, 920–933.

——— (1999b), "Wavelet Shrinkage Estimation of Certain Poisson Intensity Signals using Corrected Thresholds," *Statistica Sinica*, 9, 119–135.

Lina, J-M. (1997), "Image Processing with Complex Daubechies Wavelets," *Journal of Mathematical Imaging and Vision*, 7, 211–223.

Mallat, S. G. (1989), "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11, 674–693.

Nason, G. P. (1996), "Wavelet Shrinkage Using Cross-Validation," *Journal of the Royal Statistical Society*, Series B, 58, 463–479.

——— (1998), *WaveThresh3 Software*, Department of Mathematics, University of Bristol, Bristol, UK.

Nason, G. P., and Silverman, B. W. (1995), "The Stationary Wavelet Transform and Some Statistical Applications," in *Lecture Notes in Statistics*, 103, New York: Springer-Verlag, 281–299.

Nason, G. P., von Sachs, R., and Kroisandt, G. (2000), "Wavelet Processes and Adaptive Estimation of the Evolutionary Wavelet Spectrum," *Journal of the Royal Statistical Society*, Series B, 62, 271–292.

Nowak, R. D., and Baraniuk, R. G. (1999), "Wavelet Domain Filtering for Photon Imaging Systems," *IEEE Transactions on Image Processing*, 8, 666–678.

Patil, P. N., and Wood, A. T. A. (2004), "Counting Process Intensity Estimation by Orthogonal Wavelet Methods," *Bernoulli*, 10, 1–24.

Priestley, M. B. (1981), *Spectral Analysis and Time Series*, London: Academic Press.

Sardy, S., Antoniadis, A., and Tseng, P. (2004), "Automatic Smoothing With Wavelets for a Wide Class of Distributions," *Journal of Computational and Graphical Statistics*, 13, 399–421.

Timmermann, K. E., and Nowak, R. D. (1997), "Multiscale Bayesian Estimation of Poisson Intensities," *Proceedings of the Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA: IEEE Computer Press, pp. 85–90.

——— (1999), "Multiscale Modeling and Estimation of Poisson Processes With Application to Photon-Limited Imaging," *IEEE Transactions on Information Theory*, 45, 846–862.

Vidakovic, B. (1999), *Statistical Modeling by Wavelets*, New York: Wiley.