A DATA-DRIVEN HAAR-FISZ TRANSFORM FOR MULTISCALE VARIANCE STABILIZATION

P. Fryzlewicz[1] and V. Delouille[2]

[1] Department of Mathematics, Imperial College London, UK.
[2] Royal Observatory of Belgium, Brussels, Belgium.

ABSTRACT

We propose a *Data-Driven Haar Fisz Transform* (DDHFT): a fast, fully automatic, multiscale technique for approximately Gaussianising and stabilizing the variance of sequences of non-negative independent random variables whose variance is a non-decreasing (but otherwise unknown) function of the mean. We demonstrate the excellent performance of the DDHFT on Poisson data. We then use the DDHFT to denoise a solar irradiance time series recorded by the X-ray radiometer on board the GOES satellite: as the noise distribution is unknown, we first take the DDHFT, then use a standard wavelet technique for homogeneous Gaussian data, and then take the inverse DDHFT. The procedure is shown to significantly outperform its competitors.

1. INTRODUCTION

In non-parametric regression, we are often faced with the problem of estimating a one-dimensional function $f : [0,1] \mapsto \mathbb{R}$ from noisy observations X_i taken on an equispaced grid:

$$X_i = f(i/n) + \varepsilon_i, \quad i = 1, \ldots, n,$$

where ε_i's are random variables with zero mean, $\mathbb{E}(\varepsilon_i) = 0$. In cases where f possesses irregular features, several authors advocate the use of nonlinear estimators based on wavelet shrinkage [1–3]. Many of these estimators combine excellent finite-sample performance and (near-)optimal Mean Integrated Squared Error (MISE) behaviour over a variety of smoothness classes of f.

Several of those wavelet-based methods rely on the assumption that $(\varepsilon_i)_{i=1}^n$ is a sequence of independent $N(0, \sigma^2)$ variables. In practice, however, the distribution of ε_i often depends on the level of the underlying signal $f(i/n)$. For example, consider a possibly inhomogeneous Poisson process observed on the interval $[0, 1)$ and discretised into a vector $\mathbf{X} = (X_1, \ldots, X_n)$, where X_i is the number of events falling into the interval $[i/n, (i+1)/n)$. Each X_i can be thought of as coming from a Poisson distribution with an unknown parameter λ_i, which needs to be estimated. Here the "noise" $\varepsilon_i = X_i - \mathbb{E}(X_i) = X_i - \lambda_i$ is independent but not identically distributed.

One way to construct successful wavelet-based denoising algorithms is to tailor the method to one particular type of noise. Some recent techniques for estimating Poisson intensities are reviewed in [4]. Antoniadis et al. [5, 6] consider the case where the noise variance is (at most) a quadratic or cubic function of the mean. Wavelet thresholding of chi-square data is used in various time series contexts: see [7], Section 2.3.2, for a list of references. The major drawback of these techniques is that they essentially require a pre-estimate of the signal to be supplied to the estimation procedure.

Another approach is to transform the noisy signal into a signal contaminated with approximately Gaussian noise with constant variance, apply a denoising procedure designed for Gaussian noise, and then apply the inverse transform to the result. For Poisson data, Anscombe [8] proposed a square-root transform which induces exact asymptotic normality and stabilizes the variance. This is a particular case of the Box-Cox transformation [9]. Some recent variance-stabilizing transforms are reviewed in [10].

When the underlying noise distribution is unknown, the appropriate variance-stabilizing transformation needs to be estimated from the data. Examples of such data-driven transformations include the AVAS technique of [11], the ACE method of [12], as well as the procedure described in [13].

Fryzlewicz et al. [14] proposed a Haar-Fisz Transform (HFT) for Gaussianising and stabilizing the variance of sequences of Poisson counts. The HFT is performed in linear computational time as a straightforward modification of the discrete Haar Transform (HT) [15]. Unlike the Anscombe transform, the HFT is not diagonal and has a multiscale structure. The same authors proposed a similar HFT for chi-square data [16].

Our contribution in this paper is to combine the above ideas and propose a fully automatic Haar-Fisz method for (approximately) Gaussianising and stabilizing the variance of sequences of nonnegative independent variables whose variance is a non-decreasing (but otherwise unknown) function of the mean. To avoid a possible notational confusion with [14] and [16], we call our automatic multiscale transform the *Data-Driven Haar-Fisz Transform* (DDHFT). We use the DDHFT to denoise an X-ray irradiance time series recorded by GOES [17]. This time series can be subject to photon noise, readout noise, as well as aliasing. Moreover, the larger the fluctuations, the more the aliasing is expected to impact the usual Poisson noise distribution. A flexible method like the DDHFT is thus necessary to analyse the noise structure.

The paper is organised as follows. In Section 2, we specify our assumptions and recall the HT. In Section 3 (resp. 4), we introduce the DDHFT in the case when the link function between the mean and the variance of the noise is known (resp. unknown). In Section 5, we show the performance of our method on a data set arising in solar physics (namely, irradiance data). In Section 6, we conclude and discuss the use of the DDHFT in the analysis of data sets collected by radiometers such as LYRA on PROBA-2 [18].

2. ASSUMPTIONS ON DATA AND DISCRETE HAAR TRANSFORM

Let $\mathbf{X} = (X_i)_{i=1}^n$ denote an input vector to the DDHFT. The following list specifies the generic distributional properties of $\mathbf{X}$.

1. The length n of $\mathbf{X}$ must be a power of two. We denote $J = \log_2(n)$.

2. $(X_i)_{i=1}^n$ must be a sequence of independent, nonnegative random variables with finite positive means $\mu_i = \mathbb{E}(X_i) >$
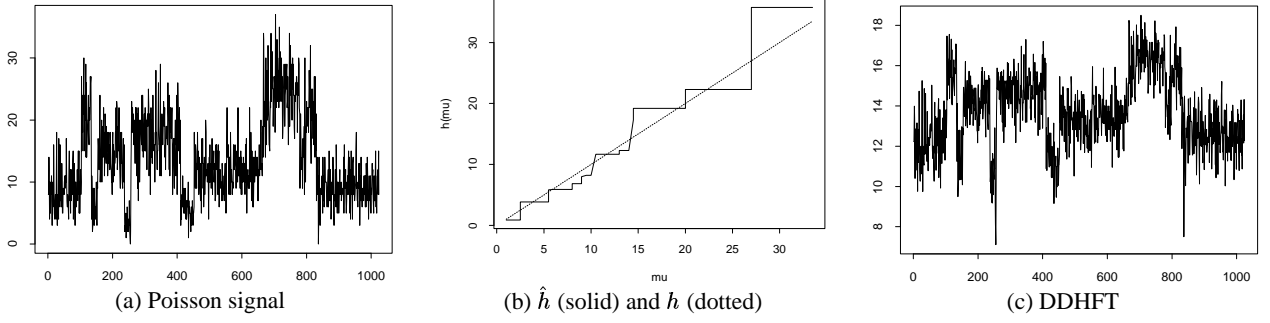
Fig. 1. Variance stabilization for a Poisson contaminated signal.

(a) Poisson signal   (b) $\hat{h}$ (solid) and h (dotted)   (c) DDHFT

0 and finite positive variances $\sigma_i^2 = \text{Var}(X_i) > 0$.

3. The variance σ_i^2 must be a non-decreasing function of the mean μ_i: we must have $\sigma_i^2 = h(\mu_i)$, where the function h is independent of i.

For example, let $X_i \sim \text{Pois}(\lambda_i)$. In this case, $\mu_i = \lambda_i$ and $\sigma_i^2 = \lambda_i$, which yields $h(x) = x$. Naturally, in many practical situations the exact form of h is unknown and needs to be estimated from the data. Sections 3 and 4 below describe the HFT in the cases where h is known and unknown, respectively.

Before we move on to describe the HFT and its data-driven version, we briefly recall the formula for the HT. The HT is a linear orthogonal transform $\mathbb{R}^n \to \mathbb{R}^n$ where $n = 2^J$. Given an input vector $\mathbf{X} = (X_i)_{i=1}^n$, the HT is performed as follows:

1. Let $s_i^J = X_i$.

2. For each $j = J-1, J-2, \ldots, 0$, recursively form vectors $\mathbf{s}^j$ and $\mathbf{d}^j$:
$$s_k^j = \frac{s_{2k-1}^{j+1} + s_{2k}^{j+1}}{2} \; ; \; d_k^j = \frac{s_{2k-1}^{j+1} - s_{2k}^{j+1}}{2}, \; k = 1, \ldots, 2^j.$$

The operator $\mathcal{H}$, where $\mathcal{H}\mathbf{X} = (\mathbf{s}^0, \mathbf{d}^0, \ldots, \mathbf{d}^{J-1})$, defines the HT. The inverse HT is performed as follows:

1. For each $j = 0, 1, \ldots, J-1$, recursively form $\mathbf{s}^{j+1}$:
$$s_{2k-1}^{j+1} = s_k^j + d_k^j \; ; \; s_{2k}^{j+1} = s_k^j - d_k^j, \; k = 1, \ldots, 2^j.$$

2. Set $X_i = s_i^J$.

The elements of $\mathbf{s}^j$ and $\mathbf{d}^j$ have a simple interpretation: they can be thought of as "smooth" and "detail" (respectively) of the original vector $\mathbf{X}$ at scale 2^j.

3. THE HAAR-FISZ TRANSFORM WITH H KNOWN

In this section, we introduce the HFT: a multiscale algorithm for (approximately) stabilizing the variance of $\mathbf{X}$ and bringing its distribution closer to normality.

The main idea of the HFT is to decompose $\mathbf{X}$ using the HT, then "Gaussianise" the coefficients d_k^j and stabilize their variance, and then apply the inverse HT to obtain a vector which is closer to Gaussianity and has its variance approximately stabilized. We now describe the middle step: the variance stabilization and "Gaussianisation" of d_k^j.

Consider first $d_1^{J-1} = (X_1 - X_2)/2$. Suppose for now that X_1, X_2 are identically distributed (i.d.): indeed, this is likely to occur if the underlying mean $\{\mu_i\}_i$ is e.g. piecewise constant. This implies that d_1^{J-1} is symmetric around zero. We want to stabilize the variance of d_1^{J-1} around $2^{(J-1)-J} = 1/2$. To do

so, we divide d_1^{J-1} by $2^{1/2}$ times its own standard deviation. We have
$$\text{Var}(d_1^{J-1}) = 1/4 \left(\text{Var}(X_1) + \text{Var}(X_2) \right) = \sigma_1^2/2,$$
which gives $2^{1/2} \left(\text{Var}(d_1^{J-1}) \right)^{1/2} = \sigma_1 = h^{1/2}(\mu_1)$. In practice μ_1 is unknown and we estimate it locally by $\hat{\mu}_1 = (X_1 + X_2)/2 = s_1^{J-1}$. The (approximately) variance-stabilized coefficient f_1^{J-1} is given by $f_1^{J-1} = d_1^{J-1}/h^{1/2}\left(s_1^{J-1} \right)$ (throughout the paper, we use the convention $0/0 = 0$).

Turning now to $d_1^{J-2} = (X_1 + X_2 - X_3 - X_4)/4$, we also first assume that the X_1, X_2, X_3, X_4 are i.d. In order to stabilize the variance of d_1^{J-2} around $2^{j-J} = 2^{J-2-J} = 1/4$, we divide d_1^{J-2} by 2 times its standard deviation. We have $2 \left(\text{Var}(d_1^{J-2}) \right)^{1/2} = \sigma_1 = h^{1/2}(\mu_1)$ as before, and we estimate μ_1 locally by s_1^{J-2}, which yields an approximately variance-stabilized coefficient $f_1^{J-2} = d_1^{J-2}/h^{1/2}\left(s_1^{J-2} \right)$. Asymptotic Gaussianity and variance stabilization of random variables of a form similar to f_k^j were studied by Fisz [19]: hence we label f_k^j the *Fisz coefficients* of $\mathbf{X}$, and the whole procedure — the *Haar-Fisz transform* of $\mathbf{X}$.

We now give the general algorithm for the Haar-Fisz transform when the function h is known.

1. Let $s_i^J = X_i$.

2. For each $j = J-1, J-2, \ldots, 0$, recursively form vectors $\mathbf{s}^j$ and $\mathbf{f}^j$:
$$s_k^j = \frac{s_{2k-1}^{j+1} + s_{2k}^{j+1}}{2} \; ; \; f_k^j = \frac{s_{2k-1}^{j+1} - s_{2k}^{j+1}}{2h^{1/2}\left(s_k^j \right)}, \; k = 1, \ldots, 2^j.$$

3. For each $j = 0, 1, \ldots, J-1$, recursively modify $\mathbf{s}^{j+1}$:
$$s_{2k-1}^{j+1} = s_k^j + f_k^j \; ; \; s_{2k}^{j+1} = s_k^j - f_k^j, \; k = 1, \ldots, 2^j.$$

4. Set $\mathbf{Y} = \mathbf{s}^J$.

The relation $\mathbf{Y} = \mathcal{F}_h \mathbf{X}$ defines a nonlinear, invertible operator $\mathcal{F}_h$ which we call *the Haar-Fisz transform (of $\mathbf{X}$) with link function h*.

4. THE HAAR-FISZ TRANSFORM WITH H UNKNOWN

In practice h is often unknown and needs to be estimated from the data. Since $\sigma_i^2 = h(\mu_i)$, ideally we would wish to estimate h by computing the empirical variances of $X_1, X_2, \ldots$ at points $\mu_1, \mu_2, \ldots$, respectively, and then smoothing the observations to obtain an estimate of h. Suppose for the time being that the μ_i's are known and, as an illustrative example, consider $\mu_i = \mu_{i+1}$. The empirical variance of X_i can be pre-estimated, for example, as $\hat{\sigma}_i^2 = (X_i - X_{i+1})^2/2$. Note that on any piecewise constant

stretch, our pre-estimate is exactly unbiased. The above discussion motivates the following regression setup:

$$\hat{\sigma}_i^2 = h(\mu_i) + \varepsilon_i,$$

where $\varepsilon_i = \hat{\sigma}_i^2 - \sigma_i^2 = (X_i - X_{i+1})^2/2 - \sigma_i^2$ and "in most cases" $\mathbb{E}(\varepsilon_i) = 0$. Of course, in practice, the μ_i's are not known and, since we pre-estimate the variance of X_i using X_i and X_{i+1}, it also makes sense to pre-estimate μ_i by $\hat{\mu}_i = (X_i + X_{i+1})/2$. Note that for each $k = 1, \ldots, 2^{J-1}$, we have $\hat{\mu}_{2k-1} = s_k^{J-1}$ and $\hat{\sigma}_{2k-1}^2 = 2(d_k^{J-1})^2$, which leads to our final regression setup

$$2(d_k^{J-1})^2 = h(s_k^{J-1}) + \varepsilon_k. \tag{1}$$

In other words, we estimate h from the finest-scale Haar smooth and detail coefficients of $(X_i)_{i=1}^n$, where the smooth coefficients serve as pre-estimates of μ_i and the squared detail coefficients serve as pre-estimates of σ_i^2.

As we restrict h to be a non-decreasing function of μ, we choose to estimate it from the regression problem (1) via least-squares isotone regression, using the "pool-adjacent-violators" algorithm described in detail in [20], Section 6.3. The resulting estimate, denoted here by $\hat{h}$, is a non-decreasing, piecewise constant function of μ.

The DDHFT is performed as in Section 3 except that h is replaced by $\hat{h}$. As an example, we exhibit the performance of our DDHFT for Poisson noise.

Figure 1 shows results for a Poisson-contaminated signal. Figure 1(a) shows a simulated Poisson vector X_t whose underlying intensity is Donoho's "blocks" function sampled at 1024 equispaced points and scaled to have a minimum (maximum) of 3 (25). Figure 1(b) shows the estimate $\hat{h}(\mu)$ (solid line), estimated for X_t from the regression problem (1) via least-squares isotone regression. The dotted line is the true $h(\mu)$ function: recall that for Poisson data, we have $h(\mu) = \mu$. The resulting DDHFT of X_t is displayed in Figure 1(c): the variance of the noise is now clearly well stabilized. The Anscombe square-root transform of X_t produces a result similar to the one in Figure 1(c). Note however that the Anscombe transform is specifically designed for Poisson noise whereas the DDHFT "does not know" the nature of the noise and needs to estimate some of its characteristics (namely, the link function $h(\mu)$) from the data.

5. APPLICATION TO SOLAR PHYSICS DATA

In this section, we analyse a time series provided by the X-ray sensor instrument on board the GOES satellite [17]. The series was recorded on 9th February 2001, and represents measurements for the whole Sun X-ray fluxes in the $0.1 - 0.8$ nm wavelength band, at the sampling rate of one measurement every 3 seconds. We only consider the last 2^{14} observations of the time series, which corresponds to the measurements taken between around 10.00 and 24.00 hours. The time series (denoted by $\{X_t\}$), is shown in Figure 2(a). It is visibly noisy (although the variance of the noise is not large), and it is of interest to solar physicists to (a) detect departure of the noise from the white noise model with constant variance, (b) remove the noise before proceeding with the analysis. We now denoise $\{X_t\}$ using a classical wavelet-based technique and our DDHFT-based approach, and compare the results.

5.1. Wavelet smoothing of the raw data

As the underlying signal appears "spiky", we denoise it using a wavelet-based method: namely, the translation-invariant version of a hard thresholding estimator proposed by [1] (we use Daubechies' Extremal Phase wavelet with 2 vanishing moments [21]). This procedure will be referred to as UNIV below. Figure 2(b) shows the

resulting estimate (displayed on part of the domain only, for clarity).

In the region shown, the UNIV estimate is extremely noisy. The reason is that UNIV assumes that the variance of the noise is constant over time, whereas this does not hold for X_t, see Figure 2(a). Some extensions of classical wavelet denoising to inhomogeneous variance have been developed. However, most of them use the universal threshold, and require a pre-estimate of the time-varying noise level. Hence the question of the choice of the pre-estimation method, and its parameters, arises [22]. In contrast, our approach based on the DDHFT is (a) completely automatic, i.e. no smoothing parameters need to be supplied to the user, and (b) modular, i.e. it can make use of any denoising procedure intended for homogeneous Gaussian data.

5.2. Approach via DDHFT

In this section, we apply a variant of our DDHFT-based denoising algorithm to X_t. The first step is to estimate the link function h from the data, see Figure 3(a). The sharp step around $\mu \approx 1.2 \times 10^{-6}$ seems to indicate (at least) two noise regimes: one with a lower variance for $\mu < 1.2 \times 10^{-6}$, and the other with a higher variance for $\mu > 1.2 \times 10^{-6}$. Figure 3(b) shows $\mathcal{F}_{\hat{h}} X_t$: the DDHFT of X_t computed using the estimated link function $\hat{h}$. The variance of the noise in $\mathcal{F}_{\hat{h}} X_t$ appears to be constant over time. We also tried the AVAS and ACE techniques mentioned in the Introduction, but none of them produced a good result: the noise was still strongly heterogeneous.

Next, we denoise $\mathcal{F}_{\hat{h}} X_t$ by means of the UNIV procedure described in Section 5.1. The variance of the corresponding empirical residuals appears to be constant over time. The denoised version of $\mathcal{F}_{\hat{h}} X_t$ is denoted by $\tilde{\mathcal{F}}_{\hat{h}} X_t$. We estimate $\mathbb{E}X_t$ by $\mathcal{F}_{\hat{h}}^{-1} \tilde{\mathcal{F}}_{\hat{h}} X_t$: the inverse DDHFT of $\tilde{\mathcal{F}}_{\hat{h}} X_t$. The estimate $\mathcal{F}_{\hat{h}}^{-1} \tilde{\mathcal{F}}_{\hat{h}} X_t$ is shown in Figure 2(c) (for comparison with the UNIV estimate from Figure 2(b), only part of the time domain is displayed). The noise-free character of $\mathcal{F}_{\hat{h}}^{-1} \tilde{\mathcal{F}}_{\hat{h}} X_t$, compared to UNIV, is remarkable.

Finally, Figure 3(c) shows the residuals $X_t - \mathcal{F}_{\hat{h}}^{-1} \tilde{\mathcal{F}}_{\hat{h}} X_t$ (solid line) and the plot of the function $10^{-7} \mathbb{I}(X_t < 1.2 \times 10^{-6})$, where $\mathbb{I}(\cdot)$ is the indicator function (dotted line). The fact that the residuals oscillate around zero, shows the apparent lack of bias in $\mathcal{F}_{\hat{h}}^{-1} \tilde{\mathcal{F}}_{\hat{h}} X_t$.

6. CONCLUSION AND FUTURE WORK

Through a simulated and a real data example, we have demonstrated that the DDHFT is a robust and flexible way of stabilizing the variance in heteroscedastic data sets. The method can handle various noise distributions and can be combined with any nonparametric denoising technology for homoscedastic data.

The DDHFT promises to be a useful tool for analysing data obtained from LYRA [18], a new radiometer to be launched in the year 2006. LYRA will measure the irradiance of the Sun in four wavebands at a sampling rate higher than 10 Hz. The data will be corrupted by different types of noise (Poisson and readout). In order to analyse the variance of the data, a flexible and fully automatic method such as the one presented here is needed. Moreover, variance stabilization and denoising are a necessary prelude to tackling solar physics problems, e.g. the problem of the extraction of peaks (which correspond to solar eruptions) from a smooth background.

7. REFERENCES

[1] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, 1994.
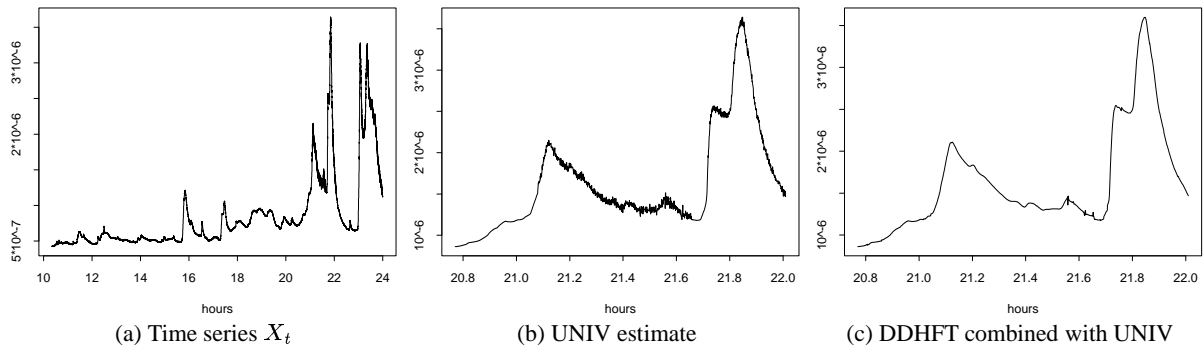
| (a) Time series X_t | (b) UNIV estimate | (c) DDHFT combined with UNIV |

Fig. 2. Denoising of the time series recorded by the X-flux radiometer on board GOES.

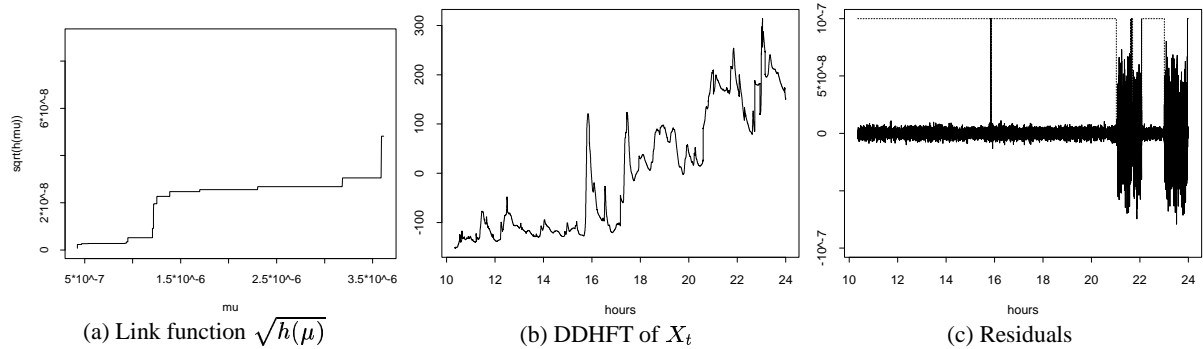| (a) Link function $\sqrt{h(\mu)}$ | (b) DDHFT of X_t | (c) Residuals |

Fig. 3. Analysis of GOES data X_t: (a) estimation of link function h, (b) DDHFT of X_t, (c) residuals of the final estimate.

[2] G. P. Nason, "Wavelet shrinkage using cross-validation," *J. Roy. Statist. Soc. Ser. B*, vol. 58, pp. 463–479, 1996.

[3] I. M. Johnstone and B. W. Silverman, "Empirical Bayes selection of wavelet thresholds," *Ann. Stat.*, vol. 33, 2005.

[4] P. Besbeas, I. De Feis, and T. Sapatinas, "A comparative simulation study of wavelet shrinkage estimators for Poisson counts," *Int. Stat. Rev.*, vol. 72, 2004.

[5] A. Antoniadis and T. Sapatinas, "Wavelet shrinkage for natural exponential families with quadratic variance functions," *Biometrika*, vol. 88, pp. 805–820, 2001.

[6] A. Antoniadis, P. Besbeas, and T. Sapatinas, "Wavelet shrinkage for natural exponential families with cubic variance functions," *Sankhya*, vol. 63, pp. 1–19, 2001.

[7] P. Fryzlewicz, *Wavelet techniques for time series and Poisson data*, Ph.D. thesis, University of Bristol, 2003.

[8] F. J. Anscombe, "The transformation of Poisson, binomial and negative-binomial data," *Biometrika*, vol. 35, pp. 246–254, 1948.

[9] G. E. P. Box and D. R. Cox, "An analysis of transformations (with discussion)," *J. Roy. Statist. Soc., Ser. B*, vol. 26, pp. 211–246, 1964.

[10] D. Ruppert, "Transformations of data," in *The International Encyclopedia of the Social & Behavioral Sciences*. Elsevier, 2001.

[11] R. Tibshirani, "Estimating transformations for regression via additivity and variance stabilization," *J. Am. Stat. Assoc.*, vol. 83, pp. 394–405, 1988.

[12] L. Breiman and J.H. Friedman, "Estimating optimal transformations for multiple regression and correlation," *J. Am. Stat. Assoc.*, vol. 80, pp. 580–619, 1985.

[13] O.B. Linton, R. Chen, N.S. Wang, and W. Härdle, "An analysis of transformations for additive nonparametric regression," *J. Am. Stat. Assoc.*, vol. 92, pp. 1512–1521, 1997.

[14] P. Fryzlewicz and G. P. Nason, "A Haar-Fisz algorithm for Poisson intensity estimation," *J. Comp. Graph. Stat.*, vol. 13, pp. 621–638, 2004.

[15] S. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Trans. Pattn Anal. Mach. Intell.*, vol. 11, pp. 674–693, 1989.

[16] P. Fryzlewicz and G. P. Nason, "Smoothing the wavelet periodogram using the Haar-Fisz transform," *Submitted*, 2004.

[17] P. L. Bornmann, D. Speich, J. Hirman, L. Matheson, R. Grubb, H. Garcia, and R. Viereck, "GOES x-ray sensor and its use in predicting solar-terrestrial disturbances," in *Proc. SPIE Vol. 2812, p. 291-298, GOES-8 and Beyond, Edward R. Washwell; Ed.*, Oct. 1996, pp. 291–298.

[18] J.-F. Hochedez, W. Schmutz, M.Nesladek, Y.Stockman, U.Schühle, A.BenMoussa, S.Koller, K.Haenen, J.-P Halain, D.Berghmans, and *other authors.*, "LYRA: the solar UV radiometer aboard the ESA PROBA-2," in *COSPAR*, 2005, To appear.

[19] M. Fisz, "The limiting distribution of a function of two independent random variables and its statistical application," *Colloquium Mathematicum*, vol. 3, pp. 138–146, 1955.

[20] I. M. Johnstone and B. W. Silverman, "EbayesThresh: R and S-Plus programs for Empirical Bayes thresholding," *J. Stat. Software,* to appear, 2005.

[21] I. Daubechies, *Ten Lectures on Wavelets*, SIAM, Philadelphia, Pa., 1992.

[22] R. von Sachs and B. MacGibbon, "Non-parametric curve estimation by wavelet thresholding with locally stationary errors," *Scand. J. Statist.*, vol. 27, pp. 475–499, 2000.