GOES-8 X-ray sensor variance stabilization using the multiscale data-driven Haar-Fisz transform.

Piotr Fryzlewicz

Department of Mathematics, University of Bristol, Bristol, UK

Véronique Delouille

Royal Observatory of Belgium, Brussels, Belgium

Guy P. Nason

Department of Mathematics, University of Bristol, Bristol, UK

Summary. We consider the stochastic mechanisms behind the data collected by the solar X-ray sensor (XRS) on board the the GOES-8 satellite. We discover and justify a non-trivial mean-variance relationship within the XRS data. Transforming such data so that its variance is stable and its distribution is taken closer to the Gaussian is the aim of many techniques (e.g. Anscombe, Box-Cox). Recently, new techniques based on the Haar-Fisz transform have been introduced that use a multiscale method to transform and stabilize data with a known mean-variance relationship. In many practical cases, such as the XRS data, the variance of the data can be assumed to increase with the mean, but other characteristics of the distribution are unknown. We introduce a method, the data-driven Haar-Fisz transform (DDHFT), which uses Haar-Fisz but also estimates the mean-variance relationship. For known noise distributions, the DDHFT is shown to be competitive with the fixed Haar-Fisz methods. We show how our DDHFT method denoises the XRS series where other existing methods fail.

Keywords: GOES-8, XRS, X-ray, variance stabilization, Gaussianisation, Haar-Fisz.

1. Introduction

1.1. GOES-8 XRS data

Central to this article is a set of solar flux time series obtained from the X-ray sensor (XRS) on the GOES-8 satellite. As well as background X-ray information the XRS provides warning that a solar-flare has occurred. This information can then be used to predict a solar-terrestrial disturbance at the Earth. Such a prediction is essential as the effects of large flares can be dramatic: they can disrupt communications, navigation systems and even knock out power grids. For example, in 1989 a severe solar storm disabled the electrical grid in Quebec for 9 hours, see Taubes (1999) for further information.

An example XRS data set is shown in Figure 1 of Section 2.2 and can be seen to be noisy. Scientists wish to denoise such data and before many further analyses, such as peak extraction, can be performed, it is essential that noise should be removed. Since the data clearly shows periods of rapid change methods such as wavelet shrinkage might be suitable, see Abramovich et al. (2000) for a review of wavelet methods. However, Section 2.2 shows that straightforward application of wavelet shrinkage is not appropriate and the reason is that the variance of the noise increases with the mean of the signal.

Another approach consists of transforming the problem to one where the variance of the noise is constant with respect to the mean of the signal, i.e. apply a variance stabilization transform to restore

homoscedasticity. Although our main focus is on the GOES-8 data, the technique we develop is applicable to many situations so we next make a few comments about the general situation.

1.2. Mean-variance relationships

A great deal of research in non-parametric regression is concerned with models where the noise is additive and Gaussian. For many real applications this is simply inappropriate. For example, for many types of count data, such as geophysical events like earthquakes, a Poisson-like model is appropriate. Another example is how to smooth the periodogram in time series analysis. In this situation the 'noise' is multiplicative and χ^2 in character. In many of these situations the variance of the noise is often some non-decreasing function, h, of the mean and hence there exists a *mean-variance* relationship. Many classical methods fail or are non-optimal in this situation.

Certain methods have been proposed to overcome the problems caused by a non-additive and/or non-constant variance functions with respect to the mean. One important class of these methods are the *variance stabilizing transforms* which attempt to transform the data to have approximately constant variance and additive noise. We will briefly review some of these in Section 1.3. Many variance stabilizing transforms require knowledge of the mean-variance relationship. One innovation in this paper consists of extending one of them, the Haar-Fisz transform, to enable it to simultaneously estimate the mean-variance function and then use that to stabilize variance. We call this new technique the Data-Driven Haar-Fisz transform (DDHFT). We use this DDHFT, along with a simple wavelet shrinkage technique, to denoise a set of GOES-8 XRS data on which other methods are shown to fail.

1.3. Brief review of existing methods

For Poisson and χ^2 data. Recent examples of Poisson intensity estimation techniques include the multiscale Bayesian methods of Kolaczyk (1999) and Timmermann and Nowak (1999), other methods are reviewed in Besbeas et al. (2004). Antoniadis and Sapatinas (2001) and Antoniadis et al. (2001) consider wavelet shrinkage for noise whose variance is (at most) a quadratic or cubic function of the mean, respectively. Several authors consider wavelet thresholding of χ^2 data arising in various time series contexts: see Fryzlewicz (2003), Section 2.3.2, for a list of references. The major drawback of these techniques is that they effectively require a pre-estimate of the signal to be supplied to the estimation procedure.

Variance stabilization. A more "modular" approach consists in transforming the noisy signal into a signal contaminated with approximately Gaussian noise with constant variance, then applying one of the many denoising procedures designed for Gaussian noise, and then transforming back to obtain an estimate of the original signal. For Poisson data, Anscombe (1948) proposed a square-root transform which induces exact asymptotic normality and stabilizes the variance. No such transform exists for scaled χ^2_2 data, although the log transform stabilizes the variance of scaled χ^2 exactly, and the cube root transform (Wilson and Hilferty (1931)) brings the data close to Gaussian but has no variance-stabilizing effect. All these transforms are, up to constants, special cases of the Box-Cox transformation (Box and Cox (1964)). Ruppert (2001) is a nice review article on variance-stabilizing transforms.

In many cases, the underlying noise distribution is not known and the appropriate variance-stabilizing transformation has to be estimated from the data. Examples of such data-driven transformations include the AVAS technique of Tibshirani (1988), the ACE method of Breiman and Friedman (1985), as well as the procedure of Linton et al. (1997).

Fryzlewicz and Nason (2004) proposed a Haar-Fisz (HF) transform for Gaussianising and stabilizing the variance of sequences of Poisson counts. The HF transform is performed in linear computational time as a computationally straightforward modification of the Discrete Haar Transform (Mallat (1989)). Unlike the Anscombe transform, the HF is not diagonal and has a multiscale structure. The authors also proposed a Poisson intensity estimation algorithm based on the HF transform and showed that it outperformed other state-of-the-art methods. Jansen (2006) extended that idea to other families of wavelets. Fryzlewicz and Nason (2006) proposed a similar HF transform for χ^2 data.

In this paper, we 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, h, of the mean (which we call the *variance* function). To avoid a possible notational clash with Fryzlewicz and Nason (2004) and Fryzlewicz and Nason (2006), we call our new automatic multiscale transform the *Data-Driven Haar-Fisz transform* (DDHFT).

1.4. Organization of the paper

Section 2 considers the GOES-8 XRS and justifies why data from it will have a non-trivial mean-variance relationship. This section also applies some standard methods to the XRS data, demonstrates their failure and previews the successful application of the DDHFT. Section 3 describes the general model setup and identifies the kinds of data that our method is likely to be suitable for. Section 4 provides some essential background information on the Haar wavelet transform which forms the basis for our new methodology, the DDHFT, which is described in Sections 5 and 6. Section 7 uses the DDHFT and a simple wavelet shrinkage technique to denoise the GOES-8 XRS data with excellent results. As mentioned previously the Haar-Fisz transform has already been proposed for known noise structures such as Poisson and χ^2. Section 8 investigates how well the *data driven* version does in those situations and concludes that there is little performance loss. Section 9 concludes.

2. The GOES-8 XRS and its mean-variance relationship

2.1. The sensor and its physical characteristics

The abbreviation GOES stands for 'Geostationary Operational Environmental Satellite'. GOESs are actually a series of geostationary satellites that form an integral part of US weather monitoring and forecasting. GOES-8 was launched on 13th April 1994 and carried a wide range of instrumentation including an imager (to take pictures of the earth at five frequencies), a sounder (measuring emitted radiation in 18 thermal infrared bands for inferring, e.g. atmospheric moisture and ozone), a space environment monitor subsystem (which includes the XRS, but also sensors for detecting energetic particles (EPS), high energy protons and alpha particles (HEPAD), and magnetometers for monitoring the Earth's geomagnetic field strength). GOES-8 ended operational service on 1st April 2003. A wide range of literature is available on GOES-8. For example, the NASA 'GOES Project Science' web page (`rsd.gsfc.nasa.gov/goes`). More traditionally, *Proceedings of SPIE*, Volume 2812, 1996 is a comprehensive treatise on the design and background to a particular GOES series of satellites of which GOES-8 was the first.

Our article concentrates on the X-ray sensor (XRS) instrument the design of which is described in detail by Hanser and Sellers (1996). Broadly speaking this detects solar X-rays and this detection is converted into an electrical signal which is quantized and then transmitted to the ground.

The XRS consists of two ion chambers (A and B) that detect X-rays in two frequency bands. Chamber A is filled with Xenon gas and responds to X-rays with wavelengths in the range 0.05nm to 0.3nm and chamber B, filled with Argon gas, to wavelengths in the range of 0.1nm to 0.8nm. As X-rays impinge on the ion chamber they generate ions in the gas which get swept into the chamber electrodes generating a tiny electric current. This current gets collected by an electronic circuit which is then further processed by the XRS electronics. Earlier GOES spacecraft possessed a similar instrument but often large solar flares caused readings that exceeded the maximum flux capacity of the device. As a consequence the GOES-8 electronics has a kind of 'auto-ranging' facility that effectively decreases the sensitivity during flare events.

The ion chamber and associated electronics is worthy of study from a statistical point of view. X-ray incidents are discrete events and number of events in a given time can be modelled using a counting stochastic process (e.g., a Poisson process). So, straightaway, *we will obtain a series of counts with a potentially non-trivial mean-variance relationship* (e.g., if the stochastic process was Poisson then clearly the variance equals to the mean). However, it should be noted that even if the X-ray arrivals were due to a conveniently nice Poisson process the actual output signal is unlikely to be so as the ion chambers have a non-linear response to X-rays of different energies, these get integrated by the first stage electronics and then the autoranging will put these quantities on different scales (and this is before we have considered other effects such as various sources or noise). So, although it is not entirely clear *a priori* what the mean-variance relationship would be it *is* clear that it is likely to be non-trivial.

As we will begin to see in the next section the GOES-8 XRS data exhibits an interesting mean-variance relationship. It should be stressed that similar relationships turn up in many places where the intensity of some kind of physical phenomena is being measured. For example, photon sensing in CCD imaging (Janesick, 2001) or in microarrays (Sebastiani et al., 2003).

2.2. An example GOES-8 XRS data set and some standard analyses

Figure 1 shows the series, X_t, which consists of $2^{14} = 16384$ observations recorded every 3 seconds from about 10am to midnight on 9th February 2001. They represent measurements for the whole Sun X-ray fluxes in the 0.1–0.8nm wavelength band. The series is visibly noisy and although the variance of the noise is not large it does appear to depend on the mean intensity of the signal. It is of interest to solar physicists both to remove the noise before proceeding with their analyses, and to understand the noise structure itself. In the remaining part of this section, we explain why some traditional approaches fail in denoising the signal, and then advocate the use of our DDHFT for this purpose.

Wavelet smoothing of the raw data. As the underlying signal appears "spiky", it is natural to attempt to denoise it by means of a wavelet-based method. Here, we make a preliminary attempt at denoising the signal using the translation-invariant version of the standard hard thresholding estimator with the universal threshold, as described in Donoho and Johnstone (1994) and Coifman and Donoho (1995). We use Daubechies' (1992) Extremal Phase wavelet with two vanishing moments (DEP2): the signal roughly resembles the Bumps signal (Donoho and Johnstone (1994)) and the DEP2 wavelet is known to perform well on that signal, see e.g. Fryzlewicz (2006). The above denoising procedure will be referred to as UNIV below. The resulting estimate (displayed on part of the domain only, for clarity) is shown as the bottom plot of Figure 1.

In the region shown, the UNIV estimate is extremely noisy. The reason for this poor performance is that UNIV (like many denoising techniques) assumes that the variance of the noise is constant over the support of the signal. This certainly does not hold for X_t: from visual inspection, it is apparent that the size of the noise between the 21st and 22nd hour is higher than that between,

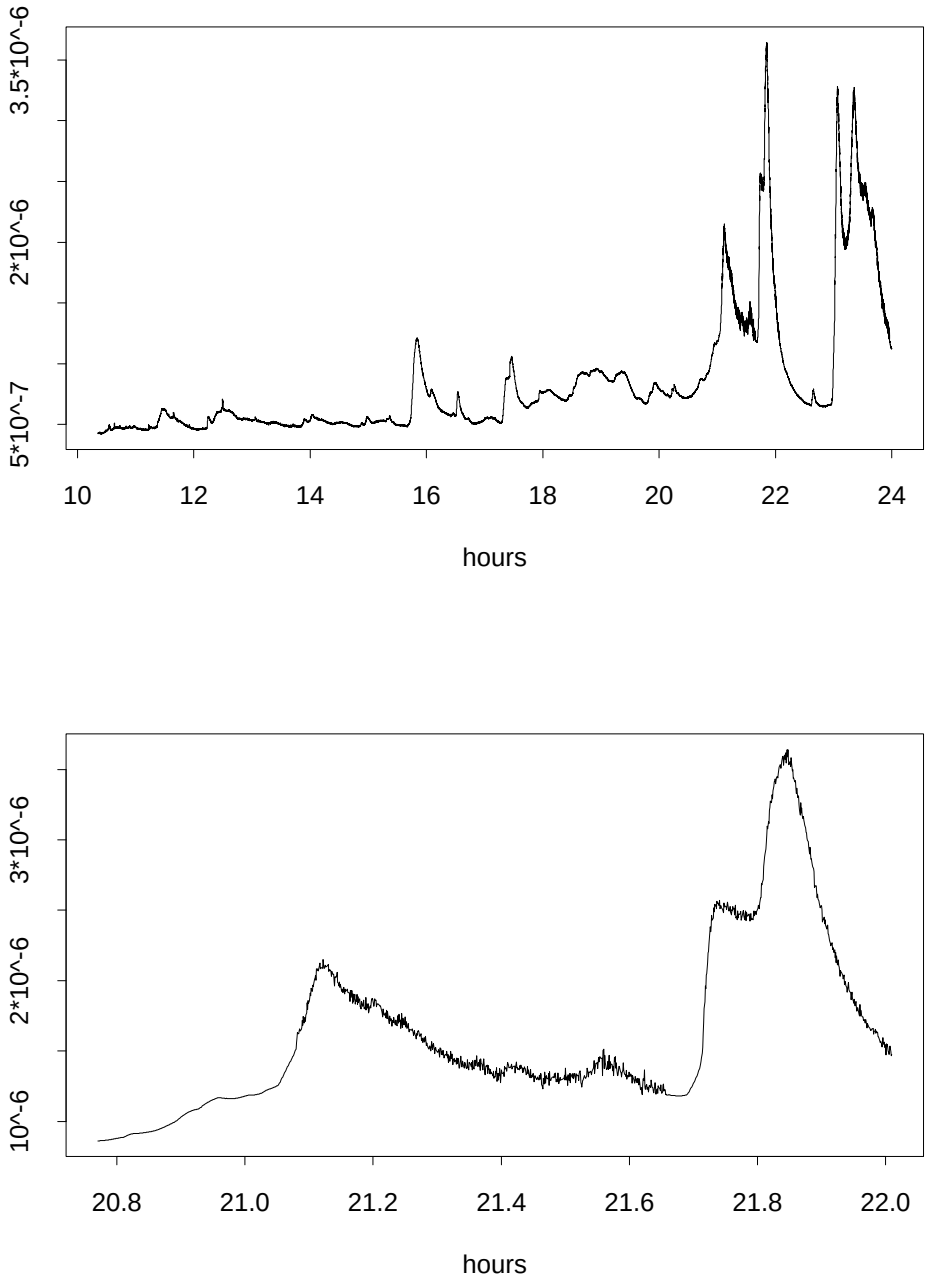

Fig. 1. Top: Solar X-ray flux X_t recorded by GOES-8 XRS on February 9, 2001. Units are Wm^{-2}. Bottom: fragment of X_t denoised by means of the UNIV procedure, see text for discussion.

say, the 18th and 19th hour. Some extensions of classical wavelet denoising to inhomogeneous variance have been developed, but they are still very preliminary: most of them only use the universal threshold, and require a pre-estimate of the noise magnitude, where the question of the choice of the pre-estimation method, and its parameters, arises, see e.g. von Sachs and MacGibbon (2000). In contrast to this methodology, our approach based on the DDHFT is (a) completely automatic, i.e. no smoothing parameters need to be supplied by the user, and (b) modular, i.e., it can make use of any denoising procedure intended for homogeneous Gaussian data.

Classical variance stabilization techniques. To avoid having to pre-estimate the time-varying noise level (which is an inherently difficult task as it requires knowledge of the noise which the analyst obviously does not have), an alternative is to apply a "data-driven" (automatic) variance stabilization technique, such as the AVAS or ACE techniques mentioned in the Introduction. Both are implemented in the popular statistical package *S-Plus* as `avas` and `ace`, respectively. Figure 2 exhibits the performance of AVAS and ACE on X_t. It is clear that neither procedure does a good job here: the variance of the noise in the transformed signals still varies from one region to another.

Preview of DDHFT performance. Before we introduce our methodology we refer forward to its results on the GOES-8 XRS data in the top-left and top-right figures of Figure 4. The left of these shows data used to estimate an empirical mean-variance relationship and the right plot shows the estimate itself. It is clear that the variance of the XRS data varies with the mean and the estimate provides a quantitative description of a possible two-regime system. After appropriate denoising using this information the final estimate of the underlying signal appears in the bottom-left plot of Figure 4 and is superior to the ones obtained above using classical methods.

Although our article concentrates on the analysis of the XRS data it is important to realize that our new DDHFT method is very general and is designed to variance stabilize (and take distributions closer to Gaussian) many kinds of data that may possess a wide range of mean-variance relationships. Hence, we need to formulate a widely applicable model and establish a general methodology. This we do next.

3. General Model Setup

In this section, we describe the probabilistic structure of input vectors to the Data-Driven Haar-Fisz transform (DDHFT), and give examples of distributions which satisfy these requirements. Let $\mathbf{X} = (X_i)_{i=1}^n$ denote an input vector to the DDHFT. The following list specifies the generic distributional properties of $\mathbf{X}$.

(a) The length n of $\mathbf{X}$ must be a power of two. We denote $J = \log_2(n)$.
(b) $(X_i)_{i=1}^n$ must be a sequence of independent, nonnegative random variables with finite positive means $\mu_i = \mathbb{E}(X_i) > 0$ and finite positive variances $\sigma_i^2 = \mathrm{Var}(X_i) > 0$.
(c) The variance σ_i^2 must be a non-decreasing function of the mean μ_i:

$$\sigma_i^2 = h(\mu_i), \tag{1}$$

where the function h is independent of i.

Note that the condition (a) above can be relaxed: even if n is not a power of two, we can still implement the DDHFT using the lifting scheme proposed by Sweldens (1996, 1997). We now give two examples of well-known probability distributions which satisfy the above requirements.

Poisson. Let $X_i \sim \mathrm{Pois}(\lambda_i)$. We have $\mu_i = \lambda_i$ and $\sigma_i^2 = \lambda_i$, which gives $h(x) = x$.

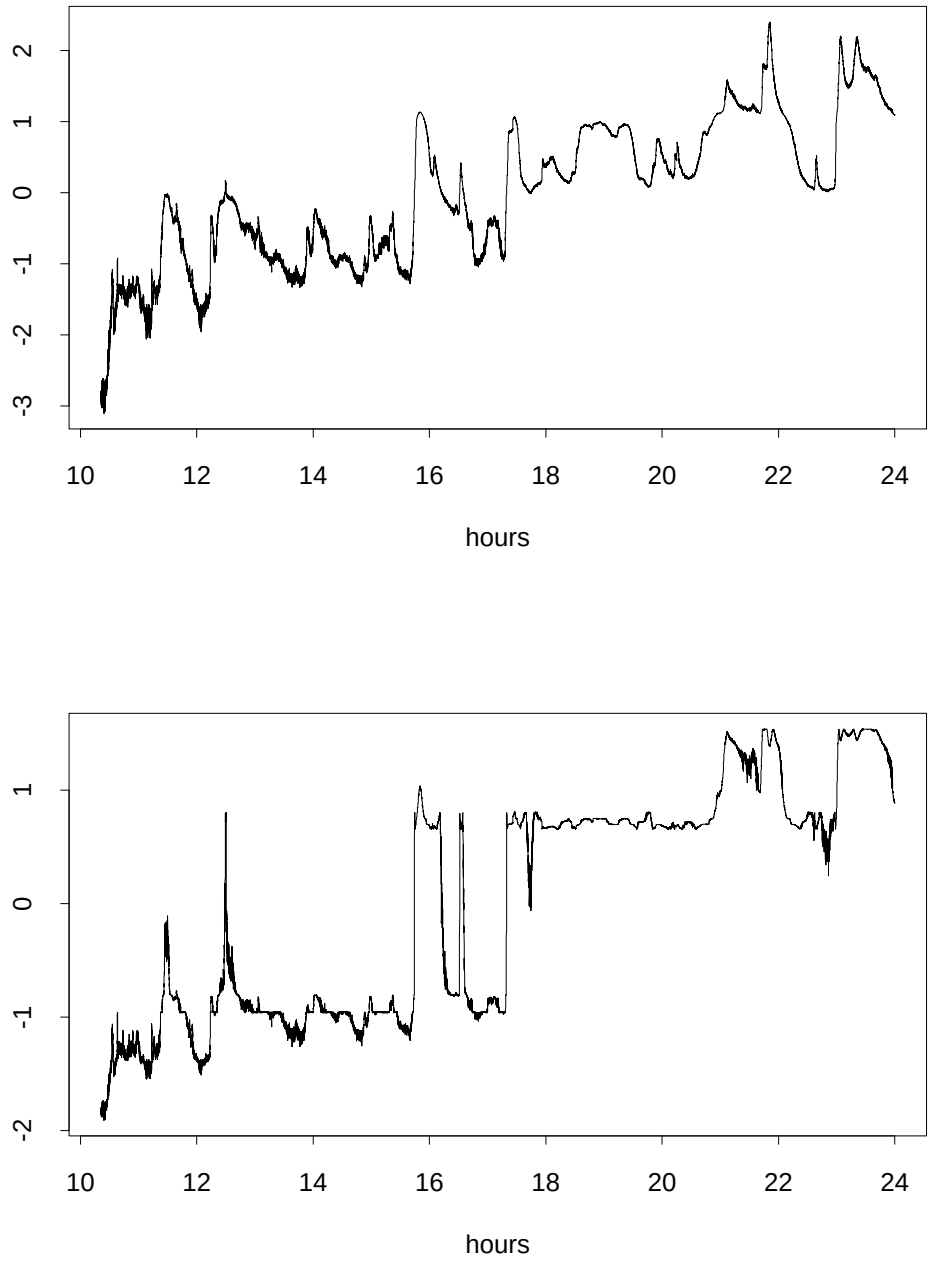

Fig. 2. Top: X_t transformed via the AVAS and (Bottom) ACE variance-stabilizing transforms..

Gamma. Let $X_i \sim \text{Gamma}(\lambda_i, \beta_i)$ where λ_i is the scale parameter and β_i is the shape parameter. We have $\mu_i = \beta_i/\lambda_i$ and $\sigma_i^2 = \beta_i/\lambda_i^2$. This yields two natural possibilities:

(a) $h(x) = x^2/\beta_i$, hence β_i must be independent of i. As an example, consider $X_i = \alpha_i \left(Z_1^2 + \ldots + Z_m^2\right)$ where $\alpha_i > 0$ and Z_j i.i.d. $N(0,1)$. Then

$$X_i \sim \text{Gamma}(1/(2\alpha_i), m/2)$$

and, for a fixed m, β_i is independent of i whereas λ_i is allowed to vary along the signal. Readers might recognize this parameterization which is the one conventionally used in generalized linear models.

(b) $h(x) = x/\lambda_i$, hence λ_i must be independent of i. Consider now

$$X_i = \alpha \left(Z_1^2 + \ldots + Z_{m_i}^2\right)$$

where $\alpha > 0$ and Z_j i.i.d. $N(0,1)$. Then $X_i \sim \text{Gamma}(1/(2\alpha), m_i/2)$ and, for a fixed α, λ_i is independent of i while β_i is allowed to vary along the signal.

Our motivation for restricting the variance function h to be non-decreasing can be summarised as follows:

- As described at the end of Section 2 and in Section 7 the GOES XRS data has a function h that can be modelled as non-decreasing.
- As demonstrated above, for the two canonical examples of Poisson and χ^2 data, the true h is increasing.
- For a number of measuring devices (for example Charge Coupled Devices, which form the basis of modern image sensors and can be found in both domestic video cameras and advanced astronomical instruments, see e.g. Janesick (2001)), we can expect a larger noise variance for observations X_i which are larger in value (i.e. have a larger mean).
- With a shape restriction of this form, it is natural to estimate h via isotone regression, which is a fully automatic procedure, i.e. no smoothing parameters (such as bandwidth) need to be supplied by the user. See Section 5 for details of the estimation procedure.

In the Gamma and Poisson examples the exact form of h was known. However, in many practical situations it is unknown and needs to be estimated from the data. Sections 5 and 6 below describe the Haar-Fisz transform in the cases where h is known and unknown, respectively.

4. The Haar transform

Before we move on to describe the Haar-Fisz transform and its data-driven version, we briefly explain the Discrete Haar Transform (DHT). Our Haar-Fisz transform will arise as a computationally straightforward modification of the DHT. The DHT 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 DHT is performed as follows:

(a) Let $s_i^J = X_i$, for $i = 1, \ldots, n$.

(b) For each $j = J-1, J-2, \ldots, 0$, recursively form vectors $\mathbf{s}^j$ and $\mathbf{d}^j$:

$$\begin{aligned} s_k^j &= \frac{s_{2k-1}^{j+1} + s_{2k}^{j+1}}{2} \\ d_k^j &= \frac{s_{2k-1}^{j+1} - s_{2k}^{j+1}}{2} \end{aligned}$$

for $k = 1, \ldots, 2^j$.

(c) The operator $\mathcal{H}$, where $\mathcal{H}\mathbf{X} = (\mathbf{s}^0, \mathbf{d}^0, \ldots, \mathbf{d}^{J-1})$, defines the DHT.

The Inverse DHT is performed by mechanically reversing the DHT steps as follows:

(a) For each $j = 0, 1, \ldots, J-1$, recursively form $\mathbf{s}^{j+1}$:

$$\begin{aligned} s_{2k-1}^{j+1} &= s_k^j + d_k^j \\ s_{2k}^{j+1} &= s_k^j - d_k^j \end{aligned}$$

for $k = 1, \ldots, 2^j$.

(b) Set $X_i = s_i^J$, for $i = 1, \ldots, n$.

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. As an example, assume $n = 8$ and consider the reconstruction of X_1:

$$\begin{aligned} X_1 &= s_1^3 \\ &= s_1^2 + d_1^2 \\ &= s_1^1 + d_1^1 + d_1^2 \\ &= s_1^0 + d_1^0 + d_1^1 + d_1^2 \\ &= \frac{\sum_{i=1}^8 X_i}{8} + \frac{\sum_{i=1}^4 X_i - \sum_{i=5}^8 X_i}{8} + \frac{X_1 + X_2 - X_3 - X_4}{4} + \frac{X_1 - X_2}{2}. \end{aligned} \quad (2)$$

Noting the pattern of operations in the above formula will make it easier to understand the origin of the more complicated formulae for the (Data-Driven) Haar-Fisz transform which we shall describe later.

5. The Haar-Fisz transform with h known

In this section, we introduce the Haar-Fisz (HF) transform: a multiscale algorithm for (approximately) stabilizing the variance of $\mathbf{X}$ and bringing its distribution closer to normality. For the time being, we assume that the function h (see formula (1)) is known. In Section 6, we propose a method for estimating it from the data.

The main idea of the HF transform is to decompose $\mathbf{X}$ using the Discrete Haar Transform (DHT), then "Gaussianise" the coefficients d_k^j and stabilize their variance, and then apply the Inverse DHT to obtain a vector which is closer to Gaussianity and has its variance approximately stabilized. We first describe the middle step: the "Gaussianisation" of d_k^j. As a motivating example, consider d_1^{J-1} and d_1^{J-2}. Later, we will describe the complete algorithm in its generality.

We first turn to d_1^{J-1}. Recall that $d_1^{J-1} = (X_1 - X_2)/2$. To facilitate understanding, the reader is invited to think of a situation where $(\mu_i)_{i=1}^n$ is a piecewise constant sequence with a small

number of jumps, and the entire distribution of X_i depends only on μ_i (as is the case, for example, with Poisson-distributed signals). We now assume that the distributions of X_1 and X_2 are identical (which is indeed likely in the piecewise-constant setup described above). This implies that the distribution of d_1^{J-1} is symmetric around zero. Our aim is to stabilize the variance of d_1^{J-1} around $2^{j-J} = 2^{J-1-J} = 1/2$. A natural way to achieve this is to divide d_1^{J-1} by $2^{1/2}$ times its own standard deviation. We have

$$\mathrm{Var}(d_1^{J-1}) = 1/4\,(\mathrm{Var}(X_1) + \mathrm{Var}(X_2)) = \sigma_1^2/2,$$

which gives $2^{1/2}\left(\mathrm{Var}(d_1^{J-1})\right)^{1/2} = \sigma_1 = h^{1/2}(\mu_1)$. Of course in practice μ_1 is unknown and we propose to estimate it locally by $\hat{\mu}_1 = (X_1 + X_2)/2 = s_1^{J-1}$. Thus, we obtain an approximately variance-stabilized coefficient f_1^{J-1}:

$$f_1^{J-1} = \frac{d_1^{J-1}}{h^{1/2}\left(s_1^{J-1}\right)},$$

with the convention $0/0 = 0$. In a different context, Fisz (1955) showed that for some well-known distributions of X_1, X_2 (such as Poisson or Gamma), the random variable f_1^{J-1} converges in distribution to $N(0, 1/2)$ in a certain asymptotic regime (for more details on the asymptotic regime in the *fixed* distribution Haar-Fisz case see Fisz (1955), Fryzlewicz and Nason (2004) or Fryzlewicz et al. (2006)). On the other hand, if the distributions of X_1 and X_2 are distinct, the distribution of f_1^{J-1} deviates in the mean from $N(0, 1/2)$ and the coefficient f_1^{J-1} carries significant information, and not only pure "noise". In this case we expect the coefficient f_1^{J-1} to survive any subsequent thresholding.

Turning now to $d_1^{J-2} = (X_1 + X_2 - X_3 - X_4)/4$, we also first look at the case where the distributions of X_1, X_2, X_3, X_4 are identical. As with d_1^{J-1}, our aim is to stabilize the variance of d_1^{J-2} around $2^{j-J} = 2^{J-2-J} = 1/4$ and to achieve this we divide d_1^{J-2} by 2 times its standard deviation. We have $2\left(\mathrm{Var}(d_1^{J-2})\right)^{1/2} = \sigma_1 = h^{1/2}(\mu_1)$ as before, and we estimate μ_1 locally by s_1^{J-2}, which yields an approximately variance-stabilized coefficient f_1^{J-2}:

$$f_1^{J-2} = \frac{d_1^{J-2}}{h^{1/2}\left(s_1^{J-2}\right)},$$

again with the convention $0/0 = 0$. Again, convergence of f_1^{J-2} to $N(0, 1/4)$ can be proved in a certain asymptotic regime. Note that if the distributions of X_1, X_2, X_3, X_4 are all identical, then using symmetry arguments it can readily be shown that f_1^{J-1} and f_1^{J-2} are exactly uncorrelated. Once more, if the distributions are not the same then we expect the coefficient f_1^{J-2} to survive any subsequent thresholding.

The asymptotic distribution of random variables of a form similar to f_k^j was studied by Fisz (1955): hence we label f_k^j the *Fisz coefficients* of $\mathbf{X}$. Motivated by the above discussion, we now introduce the algorithm for the Haar-Fisz transform where the function h is known. Given the input vector $\mathbf{X}$, the algorithm proceeds as follows:

(a) Let $s_i^J = X_i$, for $i = 1, \ldots, n$.
(b) For each $j = J-1, J-2, \ldots, 0$, recursively form vectors $\mathbf{s}^j$ and $\mathbf{f}^j$:

$$s_k^j \quad = \quad \frac{s_{2k-1}^{j+1} + s_{2k}^{j+1}}{2}$$

$$f_k^j = \frac{s_{2k-1}^{j+1} - s_{2k}^{j+1}}{2h^{1/2}\left(s_k^j\right)},$$

for $k = 1, \ldots, 2^j$, with the convention $0/0 = 0$.

(c) For each $j = 0, 1, \ldots, J-1$, recursively modify $\mathbf{s}^{j+1}$:

$$\begin{aligned} s_{2k-1}^{j+1} &= s_k^j + f_k^j \\ s_{2k}^{j+1} &= s_k^j - f_k^j \end{aligned}$$

for $k = 1, \ldots, 2^j$.

(d) 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 variance function* h.

As an example, we look again at the case $n = 8$ and consider Y_1:

$$Y_1 = \frac{\sum_{i=1}^8 X_i}{8} + \frac{\sum_{i=1}^4 X_i - \sum_{i=5}^8 X_i}{8h^{1/2}\left(\frac{\sum_{i=1}^8 X_i}{8}\right)} + \frac{X_1 + X_2 - X_3 - X_4}{4h^{1/2}\left(\frac{\sum_{i=1}^4 X_i}{4}\right)} + \frac{X_1 - X_2}{2h^{1/2}\left(\frac{X_1+X_2}{2}\right)} \qquad (3)$$

To gain a better understanding of the above pattern the reader might find it useful to compare (3) to the "resolution of identity" in formula (2). Each term in (3) applies the suitable division which Gaussianises and variance-stabilizes each individual Haar coefficient, and then the final coefficient summation induces further Gaussianisation.

Of course, after the Fisz transformation the transformed coefficients are not, in general, exactly distributed according to a Gaussian distribution. However, due to the joint action of the Fisz transform, and the repeated summation action of the wavelet transform, the transformed coefficients tend to be closer to Gaussianity than previously and coarse scale coefficients tend to be 'more Gaussian'.

6. The Haar-Fisz transform with h unknown

As we already mentioned before, 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 be able 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 μ_i's are known and, as an illustrative example, consider $\mu_i = \mu_{i+1}$ (recall the piecewise constant setup evoked before). The empirical variance of X_i can be pre-estimated, for example, as

$$\hat{\sigma}_i^2 = \frac{(X_i - X_{i+1})^2}{2}.$$

It is easily seen that, given that $\sigma_i^2 = \sigma_{i+1}^2$, we have $\mathbb{E}(\hat{\sigma}_i^2) = \sigma_i^2$, so 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" $\mathbb{E}(\varepsilon_i) \approx 0$ ("in most cases" means where the intensity function is reasonably smooth. If it is piecewise constant then $\mathbb{E}(\epsilon_i) = 0$

over the constant stretches). Of course, in practice μ_i's are not known and, since we pre-estimate the variance of X_i using X_i and X_{i+1}, we analogously pre-estimate μ_i by

$$\hat{\mu}_i = \frac{X_i + X_{i+1}}{2}.$$

Note that for each $k = 1, \ldots, 2^{J-1}$, we have $\hat{\mu}_{2k-1} = s_k^{J-1}$ and $\hat{\sigma}^2_{2k-1} = 2(d_k^{J-1})^2$, which leads us to our final regression setup

$$2(d_k^{J-1})^2 = h(s_k^{J-1}) + \varepsilon_k. \tag{4}$$

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.

Because we restrict h to be a non-decreasing function of μ, we choose to estimate it from the regression problem (4) via least-squares isotone regression, using the automatic "pool-adjacent-violators" algorithm described in detail in Johnstone and Silverman (2005), Section 6.3. The resulting estimate, denoted here by $\hat{h}$, is a non-decreasing, piecewise constant function of μ. (Although we typically obtain good results in practice it is important to note that the isotonic regression we use here is only intended as an approximate estimation method. In particular, the 'X' variable ($h(s_k^{J-1})$) is not deterministic and although the mean of ϵ_k is approximately zero and independently distributed the variance of ϵ_k is not likely to be constant. It should also be noted that for other kinds of mean-variance functions (e.g. like the Binomial) which are not increasing, isotonic regression would not be appropriate, but a more general smoother, e.g. loess, kernel or wavelet, might be useful. A full technical treatment of these issues is beyond the scope of the present paper).

The DDHFT is performed as in Section 5 except that h is replaced by $\hat{h}$.

Simulation exhibiting our method when h considered unknown. We now briefly exhibit the performance of our DDHFT for Poisson and χ^2_2 noise. We report these simulations here so that the reader can understand the behaviour of our methodology in a controlled situation so as to better prepare for the analysis of the real data in the next section.

- *Poisson noise.* The left column in Figure 3 shows results for a Poisson-contaminated signal. The top plot 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). The plot underneath shows the estimate $\hat{h}(\mu)$ (solid line), estimated for X_t from the regression problem (4) 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 next plot down the column shows the DDHFT of X_t: the variance of the noise is now clearly well-stabilized. For comparison, the bottom plot shows the Anscombe square-root transform of X_t, scaled to have the same vertical range as the DDHFT of X_t. Visually, there is little to choose between the two, however one must remember 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 characeristics (namely, the variance function $h(\mu)$) from the data. The analysis of the noise (not shown) of the DDHF-transformed vector reveals that no spurious autocorrelation is induced.

- *χ^2_2 noise.* The right column in Figure 3 shows results for independent scaled χ^2 noise with two degrees of freedom (which is similar to the periodogram setting). The series is denoted by Y_t. The underlying mean of the data is again Donoho's Blocks function sampled at 1024 equispaced points and scaled to have a minimum (maximum) of 6 (50). The plot underneath shows the estimate $\hat{h}(\mu)$ (solid line) and the true $h(\mu)$: in this setting, $h(\mu) = \mu^2$. The next

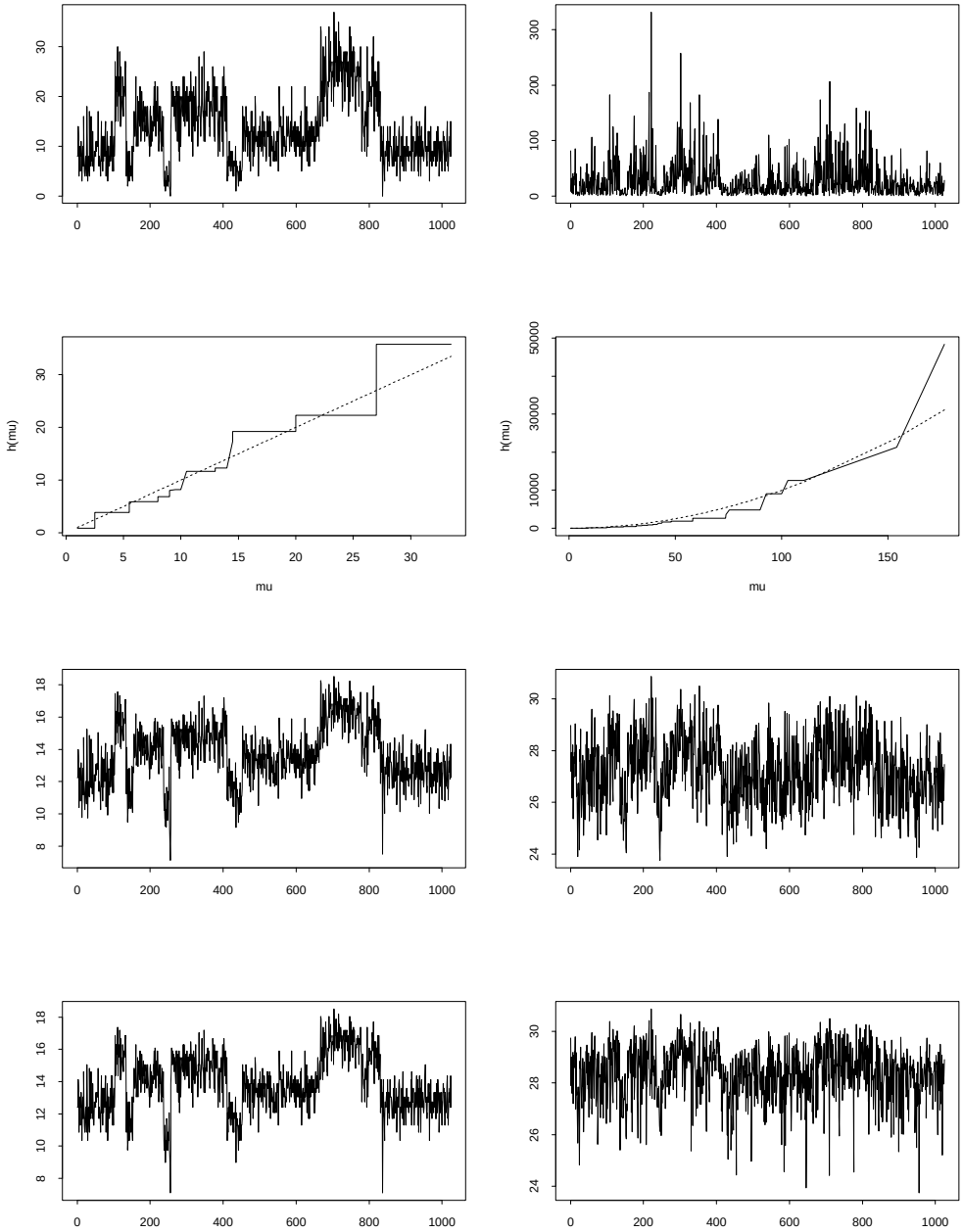

Fig. 3. Left-hand column corresponds to the Poisson case and right-hand column to the χ^2 case. Top row: Blocks function contaminated with selected noise. Second row: the true variance function h (dotted line) and its DDHFT estimate (continuous line). Third row shows the DDHFT of both signals from the top row. Bottom row: Poisson data stabilized via the Anscombe transform and the χ^2_2 data stabilized via the log-transform.

two plots show the DDHFT of Y_t and the log transform of Y_t, scaled to have the same range. The noise in the DDHFT of Y_t appears to be more symmetric, and the shape of the underlying signal is clearer. Again, the analysis of the noise (not shown) of the DDHF-transformed vector reveals that no spurious autocorrelation is present.

7. DDHFT analysis of GOES-8 XRS X-ray data

In this section, we combine our DDHFT algorithm with a wavelet denoising method. First, we estimate the variance function from the data set $\{X_t\}$. The top left plot of Figure 4 shows the pairs $(s_k^{J-1}, 2(d_k^{J-1})^2)$ plotted on a square-root scale for clarity, where s_k^j (d_k^j) are the empirical Haar smooth (detail) coefficients of X_t (see formula (4)). The top right plot shows the estimated function $\hat{h}(\mu)$, again plotted on a square-root scale. 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 $\mu > 1.2 \times 10^{-6}$. We suspect that this change in variance is due, in part, to the autoranging function of the sensor electronics which can suddenly change ranges in response to changes of solar X-ray flux.

The middle left plot shows $\mathcal{F}_{\hat{h}} X_t$: the DDHFT of X_t computed using the estimated variance function $\hat{h}$. Note that unlike in AVAS or ACE, the variance of the noise in $\mathcal{F}_{\hat{h}} X_t$ appears to be constant over time.

This observation is confirmed when $\mathcal{F}_{\hat{h}} X_t$ is denoised: the middle right plot shows the empirical residuals obtained from denoising $\mathcal{F}_{\hat{h}} X_t$ by means of the UNIV procedure described in Section 2.2 (except that, for speed, the DWT-based, rather than the translation-invariant, version of UNIV was used). The variance of the empirical residuals appears to be constant over time. We denote the denoised version of $\mathcal{F}_{\hat{h}} X_t$ by $\tilde{\mathcal{F}}_{\hat{h}} X_t$.

The last step of the denoising procedure is to take the inverse DDHFT transform of $\tilde{\mathcal{F}}_{\hat{h}} X_t$ to obtain the final estimate of $\mathbb{E}X_t$, denoted here by $\mathcal{F}_{\hat{h}}^{-1} \tilde{\mathcal{F}}_{\hat{h}} X_t$. The estimate $\mathcal{F}_{\hat{h}}^{-1} \tilde{\mathcal{F}}_{\hat{h}} X_t$ is shown in the bottom left plot of Figure 4: for comparison with the UNIV estimate from the bottom plot of Figure 1, 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, the bottom right plot on Figure 4 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). When the dotted line is 10^{-7} then X_t is in the low-variation mode. The plot clearly confirms our earlier observation that (at least) two noise regimes are present in X_t: one with a lower variance for $\mu < 1.2 \times 10^{-6}$, and the other with a higher variance for $\mu > 1.2 \times 10^{-6}$. Also, the fact that the residuals oscillate around zero, shows the apparent lack of bias in $\mathcal{F}_{\hat{h}}^{-1} \tilde{\mathcal{F}}_{\hat{h}} X_t$.

8. Comparison of DDHFT to regular Haar-Fisz with known noise

As mentioned in Section 1.3 above Fryzlewicz and Nason (2004) demonstrated how to use the Haar-Fisz transform for estimating the intensity of an inhomogeneous Poisson sequence. The general idea is to Haar-Fisz transform the "noisy" sequence, which produces a "signal+noise" representation with approximately constant variance and a distribution which is closer to Gaussian. A suitable denoiser can then be applied to estimate the signal and remove the noise. Generally speaking, since the (transformed) noise has reasonably constant variance and is "nearly Gaussian", many simple smoothers perform very well. The final step is to take the smooth and perform the inverse Haar-Fisz transform to obtain the desired estimate. Such a "HF-denoise-inverse HF" strategy can also make use of the DDHFT.

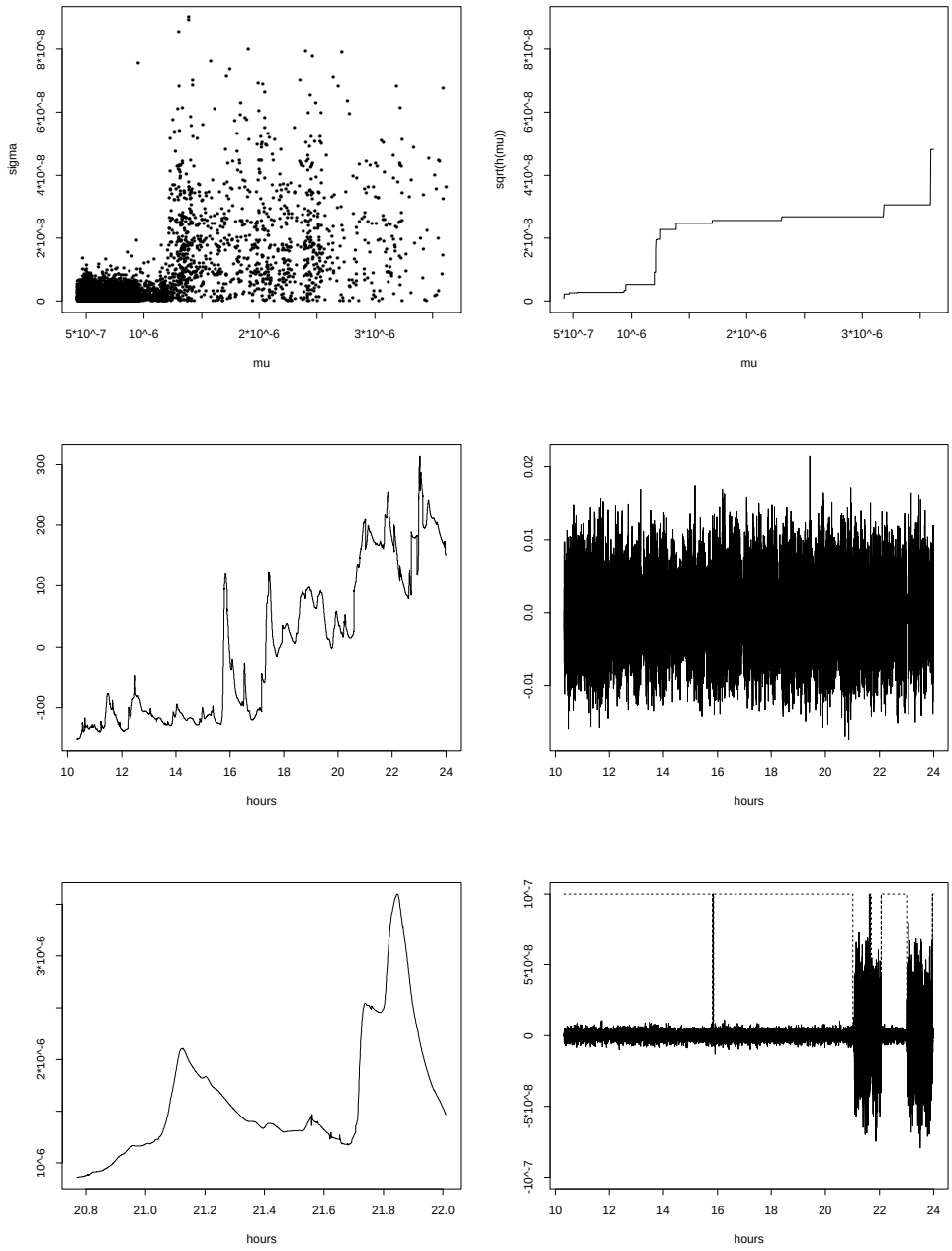

Fig. 4. DDHFT of GOES data X_t. Top row shows the scaled moduli of finest-scale Haar detail coefficients against the finest-scale Haar smooth coefficients (left) and the estimated variance function h (right). Middle row: DDHFT of X_t (left) and empirical residuals when the DDHFT of X_t is denoised (right). Bottom row: final estimate of $E(X_t)$ (left) and (right) its residuals (solid line) and indicator $10^{-7}\mathbb{I}(X_t < 1.2 \times 10^{-6})$ (dotted line). Units for top plots (x and y axes), and bottom plots (y axes) are Wm^{-2}.

Table 1. Scaled MISE for Poisson intensity estimation (Poisson) and local variance estimation for Gaussian data (χ^2_1): comparing Haar-Fisz with Data-Driven Haar-Fisz. Smoothing used Daubechies' least asymmetric wavelets with 10 vanishing moments, hard thresholding, universal threshold with variances computed by MAD and applied to scale level 3 and finer. All estimates are averaged over 50 circular shifts of the data.

	Poisson				χ^2_1	
	Peak intensity=8		Peak intensity=128		Peak intensity=8	
Intensity	HFT	DDHFT	HFT	DDHFT	HFT	DDHFT
Doppler	94	97	12	13	502	535
Blocks	287	290	31	32	803	835
Heavisine	39	40	6	6	196	203
Bumps	1243	1423	144	157	3874	3529

In this section, we compare intensity estimation results for Poisson data using (a) the Haar-Fisz transform suitable for Poisson data (with $h(\mu) = \mu$), and (b) the DDHFT. We use exactly the same denoising methodology and error measure as Fryzlewicz and Nason (2004). In summary, Fryzlewicz and Nason (2004) repeated the simulation study of Timmermann and Nowak (1999), which created "noisy" data by taking one of the Doppler, Blocks, Bumps or Heavisine signals from Donoho and Johnstone (1994), scaling them to have (minimum, maximum) intensity of $(1/8, 8)$ or $(1/128, 128)$, and then sampling 1024 independent Poisson replicates with mean equal to the intensity values at $i/1024$ for $i = 0, \ldots, 1023$. Fryzlewicz and Nason (2004) demonstrated the superiority of the Haar-Fisz algorithm over existing methods.

Table 1 shows the (scaled) Mean Integrated Square Errors averaged over 100 simulated noisy signals. The DDHFT-based algorithm is hardly worse than the Haar-Fisz algorithm, even though the variance function is being estimated by DDHFT. Coupled with the results for the Haar-Fisz transform given in Fryzlewicz and Nason (2004), this demonstrates the superiority of DDHFT over other existing methods. Moreover, the appearance of DDHFT and Haar-Fisz reconstructions is very similar; examples of the latter are given in Fryzlewicz and Nason (2004).

A similar experiment was carried out for χ^2 variation. Here the Timmermann and Nowak (1999) intensity functions were multiplied pointwise by independent χ^2_1 random variables and denoised by both a fixed Haar-Fisz transform designed for χ^2 data (with $h(\mu) = \mu^2$) and the DDHFT. Note that as the setup is now multiplicative (i.e. two noisy vectors corresponding to the two scalings are exactly multiples of each other), it is sufficient to investigate the performance for one intensity scaling. The results are also shown in Table 1. Overall, apart from the Bumps intensity, the DDHFT produces results which are up to 6% worse than the specific Haar-Fisz transform. However, for Bumps, the DDHFT does about 9% better.

9. Conclusions and future work

This article considered the GOES-8 XRS data. The prime motivation was to discover a way in which to successfully denoise the XRS data. Since the XRS data was found to possess a non-trivial mean-variance relationship standard wavelet shrinkage methods or other variance stabilization methods were shown to fail or are non-optimal in this situation. Our approach used a data-driven transform to first estimate the mean-variance relationship and then use a Haar-Fisz transform to variance stabilize the XRS data in a multiscale way. The transformed XRS data could then be denoised and the inverse DDHFT applied to obtain a successfully denoised estimate.

The DDHFT enables us to discover such mean-variance relationships and then, also, to stabilize the variance in these cases.

A great deal has been written in the statistics literature about the new paradigm of vast quantities of data collected by automatic systems. However, for many of these sensors (some of which work by collecting photons) the mean-variance relationship is usually not of the 'signal+noise' kind, nor is it likely to be Gaussian nor have constant variance. Some important examples are to be found in astronomy (sensing over wide ranges of the electromagnetic spectrum), bioinformatics (in microarrays, for example), low-light vision in security and defence applications. In many of these cases it is likely that the use of the DDHFT to Gaussianise and stabilize variance will help improve and ease the analysis of these kinds of data.

Further, simulations showed that DDHFT performs nearly as well as the fixed Haar-Fisz transforms for Poisson and χ^2 noise (even though it does not know the mean-variance relationship).

For the future it would be interesting to think about how one might use smoother wavelets to replace Haar in the DDHFT, possibly using the ideas of Jansen (2006). Another interesting question is what is the best way of performing isotonic regression when the 'X' variable is stochastic? Here we apply 'standard' isotonic algorithm but future work could investigate a formal framework and alternative methods that take account of the stochastic nature of 'X'. Or when the variance of the error, ϵ_k, in model (4), is not constant.

We mentioned in the introduction that our technique does not require pre-estimation. However, use of the quantity s_k^j in the formation of the transformed coefficients is a fixed pre-estimate of the mean. Future work might investigate whether performance can be improved by using alternatives to s_k^j but it should be remembered that more complex alternatives might require additional computations whereas the s_k^j is efficiently and automatically produced by the discrete wavelet transform.

Software implementing the DDHFT can be obtained upon request from the first author.

10. Acknowledgments

PF is grateful to the Department of Mathematics, Imperial College London, where a large part of this work was done, for an excellent research environment. VD acknowledges the support from the Belgian Federal Science Policy Office through the ESA-PRODEX programme. GPN gratefully acknowledges support from EPSRC Advanced Research Fellowship Grant GR/A01664/01 and Responsive Mode Grant EP/D005221/1.

References

Abramovich, F., T. C. Bailey, and T. Sapatinas (2000). Wavelet analysis and its statistical applications. *J. Roy. Statist. Soc. Ser. D 49*, 1–29.

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

Antoniadis, A., P. Besbeas, and T. Sapatinas (2001). Wavelet shrinkage for natural exponential families with cubic variance functions. *Sankhya 63*, 1–19.

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

Besbeas, P., I. De Feis, and T. Sapatinas (2004). A comparative simulation study of wavelet shrinkage estimators for Poisson counts. *Int. Statist. Rev. 72*, 209–238.

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

Breiman, L. and J. Friedman (1985). Estimating optimal transformations for multiple regression and correlation. *J. Am. Statist. Ass. 80*, 580–619.

Coifman, R. R. and D. L. Donoho (1995). Translation-invariant de-noising. In A. Antoniadis and G. Oppenheim (Eds.), *Lect. Notes Stat., 103*, pp. 125–150. Springer.

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

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

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. (2003). *Wavelet techniques for time series and Poisson data*. Ph. D. thesis, University of Bristol.

Fryzlewicz, P. (2006). Bivariate hard thresholding in wavelet function estimation. *Statistica Sinica 16*, (to appear).

Fryzlewicz, P. and G. P. Nason (2004). A Haar-Fisz algorithm for Poisson intensity estimation. *J. Comput. Graph. Stat. 13*, 621–638.

Fryzlewicz, P. and G. P. Nason (2006). Haar-Fisz estimation of evolutionary wavelet spectra. *J. Roy. Statist. Soc. Ser. B 68*, 611–634.

Fryzlewicz, P., T. Sapatinas, and S. Subba Rao (2006). A Haar-Fisz technique for locally stationary volatility estimation. *Biometrika 93*, (to appear).

Hanser, F. and F. B. Sellers (1996). Design and calibration of the goes-8 solar x-ray sensor: the xrs. In E. Washwell (Ed.), *GOES-8 and Beyond*, Volume 2812 of *Proc. SPIE*, pp. 344–352. SPIE.

Janesick, J. (2001). *Scientific Charge Coupled Devices*, Volume PM83. SPIE Press.

Jansen, M. (2006). Multiscale Poisson data smoothing. *J. Roy. Stat. Soc. B 68*, 27–48.

Johnstone, I. M. and B. W. Silverman (2005). EbayesThresh: R and S-Plus programs for Empirical Bayes thresholding. *J. Statist. Soft. 12.8*, 1–38.

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

Linton, O., R. Chen, N. Wang, and W. Härdle (1997). An analysis of transformations for additive nonparametric regression. *J. Am. Statist. Ass. 92*, 1512–1521.

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

Ruppert, D. (2001). Transformations of data. In *The International Encyclopedia of the Social & Behavioral Sciences*. Elsevier.

Sebastiani, P., E. Gussoni, I. Kohane, and M. Ramoni (2003). Statistical challenges in functional genomics. *Statist. Sci. 18*, 33–70.

Sweldens, W. (1996). The lifting scheme: a custom-design construction of biorthogonal wavelets. *Appl. Comput. Harmon. Anal. 3*, 186–200.

Sweldens, W. (1997). The lifting scheme: A construction of second generation wavelets. *SIAM J. Math. Anal. 29*, 511–546.

Taubes, G. (1999). Forecasting the storms and showers of space. *Science 286*, 2438–2440.

Tibshirani, R. (1988). Estimating transformations for regression via additivity and variance stabilization. *J. Am. Stat. Ass. 83*, 394–405.

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

von Sachs, R. and B. MacGibbon (2000). Non-parametric curve estimation by wavelet thresholding with locally stationary errors. *Scand. J. Statist. 27*, 475–499.

Wilson, E. B. and M. M. Hilferty (1931). The distribution of chi square. *Proc. Natn. Acad. Sci. USA 17*, 684–688.