Stochastic Atomic Decompositions in a Wavelet Dictionary

F. Abramovich[1], T. Sapatinas[2] and B. W. Silverman[3]

[1]Department of Statistics & Operations Research,
Raymond & Beverly Sackler Faculty of Exact Sciences,
University of Tel Aviv, Ramat Aviv 69978, Israel.

[2]Institute of Mathematics and Statistics,
University of Kent, Canterbury CT2 7NF, United Kingdom.

[3]Department of Mathematics,
University of Bristol, Bristol BS8 1TW, United Kingdom.

April 9, 1998

Abstract

We consider random functions defined by atomic decompositions in a wavelet dictionary. The scale and dilation of the wavelet atoms are not dyadic constraints, but the function is modelled as a sum of wavelet functions at arbitrary positions and scales. The locations of the wavelet atoms and the magnitudes of their coefficients are chosen with respect to a certain marked Poisson process model, allowing intuitive notions about the functions genuinely to be modelled. We explore the properties of realizations from the proposed models by investigating the relationships between the parameters of the models and the Besov spaces within which realizations fall. Such relations may be seen as giving insight into the meaning of the Besov space parameters themselves. Moreover, although we do not elaborate on this, the established relations make it possible in principle to incorporate prior knowledge about the function's regularity into statistical applications of the proposed models.

Keywords: Atomic Decompositions; Besov Spaces; Poisson Processes; Wavelet Dictionaries.

1 Introduction

In recent years there has been growing interest in the atomic decomposition of functions (e.g. Mallat & Zhang, 1993; Davis, Mallat & Zhang, 1994; Chen, Donoho & Saunders, 1996). In mathematical terms, an atomic decomposition of a function (signal) g is an expression of g as a superposition of a parametric collection of waveforms $(\xi_\lambda)_{\lambda\in\Lambda}$

$$g(t) = \sum_{\lambda\in\Lambda} \nu_\lambda \xi_\lambda(t). \tag{1}$$

The collection of waveforms $(\xi_\lambda)_{\lambda\in\Lambda}$ is called a *dictionary*, and the waveforms ξ_λ are called *atoms*.

A traditional and widely-used dictionary is the Fourier dictionary, which is a collection of orthonormal sinusoidal waveforms of appropriate angular frequencies. However, it is well known that the Fourier dictionary provides a poor representation of spatially inhomogeneous functions whose frequency characteristics vary in time. An efficient alternative dictionary for such functions is the orthonormal wavelet dictionary generated by translations and dilations of a single basic function, called the *mother wavelet*. Unlike Fourier atoms, the corresponding orthonormal wavelet atoms are local in both frequency/scale (via dilations) and in time (via translations). This localization allows parsimonious wavelet expansions for a wide set of functions with various degrees and kinds of smoothness.

However, being a basis, orthonormal wavelet dictionaries are essentially *minimal* wavelet dictionaries, and this causes poor adaptivity (Mallat & Zhang, 1993). *Overcomplete* wavelet dictionaries are obtained by sampling indices more finely. This increases the adaptivity of the dictionary, because one can choose the most suitable representation among many available. An important example of overcomplete wavelet dictionaries is the stationary wavelet dictionary (e.g. Coifman & Donoho, 1995; Nason & Silverman, 1995).

In this paper, we consider random functions defined by atomic decompositions in an overcomplete wavelet dictionary which has both orthonormal and stationary wavelet dictionaries as particular subsets. The scale and dilation of the wavelet atoms are no longer subject to dyadic constraints, but the function is modelled as a sum of wavelet functions at arbitrary positions and scales. The locations of the wavelet atoms and the magnitudes of their coefficients are generated by a certain marked Poisson process model, allowing intuitive notions about the functions genuinely to be modelled.

Specifically, we consider the atomic decomposition of a function g on [0,1] in a wavelet dictionary $\mathcal{D} = \{\psi_\lambda : \lambda \in \Lambda\}$

$$g(t) = \sum_{\lambda\in\Lambda} \omega_\lambda \psi_\lambda(t), \tag{2}$$

where the wavelet atoms ψ_λ are given by

$$\psi_\lambda(t) = a^{1/2}\psi(a(t-b)), \quad a \geq 1, \quad 0 \leq b \leq 1.$$

The set Λ of indices $\lambda = (a, b)$ is sampled from a Poisson process S on $[1, \infty) \times [0, 1]$ with intensity $\mu(\lambda)$. Conditional on S, the coefficients ω_λ are assumed to be independent normal random variables

$$\omega_\lambda \mid S \sim N(0, \tau^2(\lambda)). \tag{3}$$

In addition, it is assumed that both the variance $\tau^2(\lambda)$ and the intensity $\mu(\lambda)$ depend on the scale a only, and are of the form

$$\tau_a^2 \propto a^{-\delta} \quad \text{and} \quad \mu_a \propto a^{-\zeta}, \quad a \geq 1, \tag{4}$$

where $\delta,\ \zeta \geq 0$, with $\delta + \zeta > 0$. (To avoid difficulties raised by boundary conditions, however, in Section 3, we will work with functions periodic on $[0, 1]$, slightly modifying the above definitions.)

The intuitive basis of the proposed model is an extension of the notion that the discrete wavelet transform of an unknown function is sparse. It is because of this notion that wavelets are useful for compression and statistical estimation of functions and images. The parameter δ controls the relative rarity of higher frequency wavelet atoms in the function, while the parameter ζ controls the size of the contribution of these atoms when they appear. For example, if δ is small and ζ is large, there will be a considerable number of high frequency atoms but these will each have fairly low contribution, so one might expect the functions to be reasonably smooth and homogeneous. On the other hand, if δ is large and ζ is small, there will be occasional large high frequency effects in the functions.

In Section 3, we investigate the properties of the functions generated by this model by revealing relations between the parameters δ and ζ and the Besov spaces within which the random functions will fall. For some readers, we hope that these relations give insight into the meaning of the Besov space parameters, because the atomic decomposition give an appealing constructive model for functions falling into the various spaces. Noting that the literal Russian meaning of the word "besov" is "devilish", this is an attempt to "exorcise" the "devilish" Besov spaces for statisticians!

Moreover, although we do not elaborate on this, the established relations make it possible in principle to incorporate prior knowledge about the function's regularity properties if the models are used as priors in a Bayesian approach to function estimation, along the lines set out in Abramovich, Sapatinas & Silverman (1998) for orthonormal wavelet dictionaries. Further comments on potential statistical applications are given in Section 4.

The rest of the paper is organized as follows: in Section 2 we briefly review relevant aspects of the Besov spaces on the interval and orthonormal wavelet expansions that we need throughout paper. Concluding remarks are made in Section 4, and some technical results on Poisson processes that are used to prove our main results are given in the Appendix; these may be of some interest in their own right.

2 Wavelets and Besov Spaces

2.1 Wavelets and scaling functions

Wavelet series are generated by dilations and translations of a function ψ, called the *mother wavelet*,

$$\psi_{jk}(t) = 2^{j/2}\psi(2^j t - k), \quad j, k \in \mathbb{Z}.$$

For suitable choices of ψ, the corresponding set of ψ_{jk} forms an orthonormal basis in $L^2(\mathbb{R})$. The wavelet series representation of a function $g \in L^2(\mathbb{R})$ is then

$$g(t) = \sum_{j\in\mathbb{Z}} \sum_{k\in\mathbb{Z}} w_{jk}\psi_{jk}(t), \tag{5}$$

where the wavelet coefficients w_{jk} are given by

$$w_{jk} = \int_{\mathbb{R}} g(t)\psi_{jk}(t)\, dt.$$

Intuitively, the ψ_{jk} represent "smooth wiggly functions" localized to spatial positions near $k2^{-j}$ and frequencies near 2^j. The sum in (5) can be restricted to $j \geq j_0$ by also including terms depending on suitable rescalings of a *scaling function* ϕ; see any standard text on wavelets for details. In contrast to standard Fourier sine and cosine series, wavelets are local in both frequency/scale (via dilations) and in time (via translations). This localization allows parsimonious representation for a wide range of different functions in wavelet series.

We shall assume that the the scaling function ϕ and the mother wavelet ψ have compact supports and correspond to an r-regular multiresolution analysis, for some integer $r > 0$ (e.g. Daubechies, 1992). This will imply that ϕ and ψ are members of the Hölder space C^r, and that ψ has vanishing moments up to order r. For examples of mother wavelets with compact support, and with various regularity properties, see Daubechies (1992).

For detailed comprehensive expositions of the mathematical aspects of wavelets we refer to Meyer (1992), Daubechies (1992) and Wojtaszczyk (1997). Jawerth & Sweldens (1994) provide an excellent overview of wavelet-based multiresolution analyses. For a general statistical introduction to wavelets see Nason & Silverman (1994), Bruce & Gao (1996) and Ogden (1997).

2.2 Besov spaces on the interval

In this section we briefly introduce some relevant aspects of the (inhomogeneous) Besov spaces on the interval that we need further. For more details we refer, for example, to DeVore & Popov (1988), DeVore, Jawerth & Popov (1992), Meyer (1992), Donoho *et al.* (1995) and Wojtaszczyk (1997).

The Besov space $B^s_{p,q}$ on $[0,1]$ is the class of functions

$$g : [0,1] \to \mathbb{R}$$

for which $g \in L^p[0,1]$ and the seminorm $|g|_{B^s_{p,q}}$ (as defined, for example, in Donoho *et al.*, 1995) is finite. The Besov function norm is defined as

$$||g||_{B^s_{p,q}} = ||g||_{L^p[0,1]} + |g|_{B^s_{p,q}}, \quad 1 \leq q < \infty, \tag{6}$$

$$||g||_{B^s_{p,\infty}} = ||g||_{L^p[0,1]} + |g|_{B^s_{p,\infty}}. \tag{7}$$

The parameter s measures the number of derivatives, where the existence of derivatives is required in an L^p-sense, while the parameter q provides a further finer gradation.

The Besov spaces include, in particular, the well-known Sobolev and Hölder spaces of smooth functions H^m and C^s ($B^m_{2,2}$ and $B^s_{\infty,\infty}$ respectively), but in addition less traditional spaces, like the space of functions of bounded variation, sandwiched between $B^1_{1,1}$ and $B^1_{1,\infty}$. The latter functions are of statistical interest in providing particular models of spatial inhomogeneity (e.g. Donoho *et al.*, 1995).

Choose now a collection of scaling functions $\{\phi_{j_0k}\}$ and wavelets $\{\psi_{jk}, j \geq j_0\}$ that correspond to an r-regular multiresolution analysis as noted in Section 2.1. To adapt wavelets and multiresolution analyses in such a way that they result in an orthonormal basis of $L^2[0,1]$ then requires some modifications. Anderson *et al.* (1993) and Cohen, Daubechies & Vial (1993), amongst others, obtained necessary boundary corrections to retain orthonormality. Another possibility is to confine attention to periodic functions on $\mathbb{R}$ with unit period and work in effect with periodic wavelets (e.g. Daubechies, 1992, Section 9.3). The equivalences we set out below are valid whichever boundary correction is used, and as Johnstone (1994) points out, any boundary modifications affect only a fixed number of wavelet coefficients at each resolution level j.

An interesting property of orthonormal wavelet bases is that, by choosing the mother wavelet ψ to be of regularity r, one can generate unconditional wavelet bases for the Besov spaces $B^s_{p,q}$, where $\max(0, 1/p - 1/2) < s < r$, $1 \leq p, q \leq \infty$. An important tool in our argument will be an equivalence between the Besov norm of a function g and a norm based on its wavelet expansion. For $j \geq j_0$, define w_j to be the vector of coefficients

$$w_{jk} = \langle g, \psi_{jk} \rangle, \quad 0 \leq k \leq 2^{j_0} - 1.$$

Define the vector u_{j_0} to have elements

$$u_{j_0 k} = \langle g, \phi_{j_0 k} \rangle, \quad 0 \le k \le 2^{j_0} - 1.$$

Let $s' = s + 1/2 - 1/p$. The sequence space norm of the array w is defined as

$$||w||_{b^s_{p,q}} = \left\{ \sum_{j=j_0}^{\infty} 2^{js'q} ||w_j||_p^q \right\}^{1/q}, \quad 1 \le q < \infty, \tag{8}$$

$$||w||_{b^s_{p,\infty}} = \sup_{j \ge j_0} \left\{ 2^{js'} ||w_j||_p \right\}. \tag{9}$$

Then the Besov norm (6) or (7) is equivalent to the sum of the ℓ_p norm of u_{j_0} and the sequence space norm (8) or (9) respectively, that is

$$K_1 ||g||_{B^s_{p,q}} \le ||u_{j_0}||_p + ||w||_{b^s_{p,q}} \le K_2 ||g||_{B^s_{p,q}}, \tag{10}$$

where K_1 and K_2 are constants, not depending on g (e.g. Donoho *et al.*, 1995).

3 Properties of functions from the model

3.1 Periodic atomic decompositions

We now explore the connections between the parameters δ and ζ of the model (3) with hyperparameters specified by (4), and the Besov space parameters s and p. To avoid difficulties raised by boundary conditions, we work with functions periodic on $[0, 1]$, slightly modifying the original definitions as we now explain. It turns out that our results are true for all values of the parameter q; the work of Abramovich *et al.* (1998) for orthonormal wavelet dictionaries suggests that addition of a third model parameter might produce results sensitive to the Besov parameter q, but we do not investigate this possibility here.

Suppose that the scaling function ϕ and the mother wavelet ψ have compact supports and correspond to an r-regular multiresolution analysis. Let $[L, U]$ be the support of the mother wavelet ψ. We model our function g as the sum of two functions f_0 and f. The function f_0 represents the 'low frequency' part of the function, while f is modelled by an atomic decomposition at scales above some fixed level. Since the function space membership of g is essentially determined by the behaviour of its 'high-frequency' components, the important step is the detailed study of the function f.

For each λ, let $\phi_\lambda(t) = a^{1/2}\phi\{a(t-b)\}$, wrapped onto the unit interval. Choose an integer $j_0 \ge \log_2\{2(U - L)\}$ and let $a_0 = 2^{j_0}$. The dictionary within which we work consists of all scaling functions ϕ_λ for $a \le a_0$, wrapped onto the unit interval, and all wavelets ψ_λ for $a \geq a_0$. Note in passing that, by the properties of the scaling functions, it is sufficient to restrict a to $\frac{1}{2}a_0 < a < a_0$ without loss of generality, since functions on lower scales can all be expressed in terms of these. The dictionary contains the basis for $L^2[0,1]$ consisting of all functions of the form ϕ_λ with $\lambda = (2^{j_0}, 2^{-j_0}k)$, $k = 0, 1, \ldots, 2^{j_0} - 1$, and all those of the form ψ_λ with $\lambda = (2^j, 2^{-j}k)$ for $j \geq j_0$ and $k = 0, 1, \ldots, 2^j - 1$, all functions being wrapped onto the unit interval. Therefore, any $L^2[0,1]$ function can, in principle, be expressed in terms of functions in the dictionary.

To construct f_0, let $\lambda_i = (a_i, b_i)$, $i = 1, \ldots, M$ be a finite set of indices, with $a_i \leq a_0$, and set

$$f_0(t) = \sum_{i=1}^{M} \eta_i \phi_{\lambda_i}(t) \tag{11}$$

for some real numbers η_i. Because f_0 is a finite linear combination of functions ϕ_λ, it will belong to the same Besov spaces as the scaling functions. From the equivalence of norms, it can be seen that the scaling functions will be members of all Besov spaces with $s < r$, where r is the regularity of the wavelet, and these will be the only Besov spaces we consider below.

Turn now to the 'high-frequency' function f. Let S_0 be a Poisson process of events $\lambda = (a, b)$ on $[a_0, \infty) \times [0, 1]$ with intensity proportional to $a^{-\zeta}$. With each event λ associate a mark ω_λ which is normally distributed with mean 0 and variance proportional to $a^{-\delta}$. Extend the process S_0 to a process S on $[a_0, \infty) \times (-\infty, \infty)$ by replicating the process S_0 (and the associated marks) periodically in the argument b. The resulting process S will be a Poisson process on any set $[a_0, \infty) \times (x, x+1)$. We now define

$$f(t) = \sum_{\lambda \in S} \omega_\lambda \psi_\lambda(t). \tag{12}$$

3.2 Regularity properties of the random functions

We now establish sufficient conditions for f given by (12) to fall (with probability one) in any particular Besov space. To prove our results, we consider the expansion of the function f in terms of an orthonormal wavelet basis

$$f(t) = \sum_{k=0}^{2^{j_0}-1} u_{j_0 k} \phi_{j_0 k}(t) + \sum_{j=j_0}^{\infty} \sum_{k=0}^{2^j - 1} w_{jk} \psi_{jk}(t),$$

where $\psi_{jk} = \psi_\lambda$ with $\lambda = \lambda_{jk} = (2^j, 2^{-j}k)$ and $\phi_{j_0 k}$ is the corresponding transform of the scaling function. After dealing with the scaling function coefficients, we find the relation between the atomic decomposition coefficients ω_λ and the orthonormal wavelet coefficients w_{jk}. Then, we exploit the equivalence between the Besov norms and the sequence space norms of the orthonormal wavelet coefficients w_{jk} given by (8) and (9). The last stage is along the lines of the arguments used by Abramovich *et al.* (1998) for random functions defined by atomic decompositions in orthonormal wavelet dictionaries, but there are many differences of detail.

We now prove results for the case $1 \leq p < \infty$. The case $p = \infty$ will be considered in Theorem 2.

Theorem 1 *Let ψ be a compactly supported mother wavelet of regularity r, and let $\rho \in (0,1)$ be the exponent of Hölder continuity of $\psi^{(r)}$. Consider constants s, p and q such that* $\max(0, 1/p - 1/2) < s < r$, $1 \leq p < \infty$, $1 \leq q \leq \infty$. *Consider a function f as defined in (12), with the conditional variances $\tau_a^2 \propto a^{-\delta}$ and the intensity of the Poisson process $\mu_a \propto a^{-\zeta}$. Assume that $\delta \geq 0$, $0 \leq \zeta \leq 1$, and that $\delta + \zeta > 0$. Assume also that the wavelets are sufficiently regular that* $\delta < 2r + 2\rho - 1$. *Then*

$$f \in B^s_{p,q} \quad \textit{almost surely}, \;\; 1 \leq p < \infty, \; 1 \leq q \leq \infty,$$

if

$$s + 1/2 - \zeta/p - \delta/2 < 0. \tag{13}$$

Proof. First of all, consider the vector u_{j_0} of scaling function coefficients $u_{j_0k} = \langle f, \phi_{j_0,k}\rangle$ for $k = 0, 1, \dots, 2^{j_0} - 1$. In this argument, and subsequently, we use C to denote a generic positive constant, not necessarily the same each time it is used. For fixed k, let $\mathcal{T}'$ be the range of indices $\lambda = (a, b)$ with $a > a_0$ for which the support of ψ_λ overlaps that of ϕ_{j_0k} for some fixed k. Then

$$\begin{aligned} E|u_{j_0k}| &\leq E \sum_{S \cap \mathcal{T}'} |\omega_\lambda| \\ &\leq C \int_{\mathcal{T}'} a^{-\delta/2} a^{-\zeta} d\lambda \\ &\leq C \int_{a_0}^{\infty} a^{-\delta/2} a^{-\zeta} da < \infty, \end{aligned}$$

since both the scaling function and the mother wavelet have bounded support, so that the range of b for which $(a, b) \in \mathcal{T}'$ is uniformly bounded over $a > a_0$. It follows that u_{j_0k} is finite almost surely and, therefore, $||u_{j_0}||_p$ is finite almost surely for all $1 \leq p \leq \infty$.

Now consider the orthonormal wavelet coefficients $w_{jk} = \langle f, \psi_{jk}\rangle$ and set $\lambda_{jk} = (2^j, 2^{-j}k)$. For resolution and spatial indices j and k with $j \geq j_0$ and $k = 0, 1, \dots, 2^j - 1$ respectively, we then have

$$w_{jk} = \sum_{\lambda \in S} K(\lambda, \lambda_{jk}) \omega_\lambda, \tag{14}$$

where

$$K(\lambda, \lambda') = \langle \psi_\lambda, \psi_{\lambda'} \rangle.$$

We now explore some properties of the reproducing kernel K that we use in subsequent calculations. Firstly, since $\int \psi_\lambda^2 = 1$ for all λ, we always have

$$K(\lambda, \lambda')^2 \leq 1. \tag{15}$$

Now define

$$K_0(u, v) = \langle \psi, \psi_{(u,v)} \rangle.$$

Let $\lambda = (a, b)$ and $\lambda' = (a', b')$. Simple calculus shows that

$$K(\lambda, \lambda') = K_0(a/a', a'(b - b')), \tag{16}$$

In the particular case where $\lambda' = \lambda_{jk}$,

$$K(\lambda, \lambda_{jk}) = K_0(2^{-j}a, 2^j b - k). \tag{17}$$

Let $[L, U]$ be the support of the mother wavelet ψ. Then

$$K_0(u, v) = \langle \psi, \psi_{u,v} \rangle \neq 0$$

only if

$$L - U/u \leq v \leq U - L/u. \tag{18}$$

We have, from Daubechies (1992, p. 48),

$$|K_0(u, v)| \leq C u^{-(r+\rho+1/2)}, \quad \text{uniformly in } u \geq 1. \tag{19}$$

For $u < 1$, apply the symmetry of K to show that

$$K_0(u, v) = K_0(1/u, -uv)$$

and hence

$$|K_0(u, v)| \leq C u^{(r+\rho+1/2)}, \quad \text{uniformly in } u \leq 1. \tag{20}$$

Suppose that condition (13) holds. It follows from (14) that, conditionally on S, the distribution of w_{jk} is normal with mean zero and variance

$$\sigma_{jk}^2(S) = \sum_{\lambda \in S} K(\lambda, \lambda_{jk})^2 a^{-\delta}, \tag{21}$$

where, as usual, $\lambda = (a, b)$. The unconditional distribution of w_{jk} will have finite variance if the expectation of $\sigma_{jk}^2(S)$ over S is finite. If the sum in (21) is infinite for a particular S, then, conditionally on S, the sum defining w_{jk} cannot converge, because it will not converge in distribution.

More generally, for $p > 0$, the pth absolute moment of w_{jk} will be given by

$$E|w_{jk}|^p = \nu_p E_S \left\{ \sum_{\lambda \in S} K^2(\lambda, \lambda_{jk}) a^{-\delta} \right\}^{p/2}, \tag{22}$$

where ν_p is the pth absolute moment of the standard normal distribution.

Let $\mathcal{T}_{jk}^0$ be the set $[a_0, \infty) \times (2^{-j}k - \frac{1}{2}, 2^{-j}k + \frac{1}{2})$. By the definition of a_0 and j_0, we may restrict attention in the sum (21) to $S \cap \mathcal{T}_{jk}^0$ since the support of any ψ_λ with $a > a_0$ will be of length at most $\frac{1}{2}$, and so the terms excluded by restricting the sum will all be zero.

Now let S_j' be a Poisson process on the half-plane $\{(u, v) : u > 0, -\infty < v < \infty\}$ of intensity $2^{-j\zeta} u^{-\zeta}$. Define the set

$$\mathcal{T}_j = [2^{-j}, \infty) \times [-2^{j-1}, 2^{j-1}].$$

Consider the transformation of $\lambda = (a, b)$ given by $(u, v) = (2^{-j}a, 2^j b - k)$. Applied to the process $S \cap \mathcal{T}_{jk}^0$ this gives a process with the same distribution as $S_j' \cap \mathcal{T}_j$. In addition, for each λ, we have from (16) that

$$K(\lambda, \lambda_{jk}) = K_0(u, v).$$

It follows that

$$\begin{aligned} E|w_{jk}|^p &= \nu_p 2^{-j\delta p/2} E \left\{ \sum_{(u,v) \in S_j' \cap \mathcal{T}_j} K_0^2(u, v) u^{-\delta} \right\}^{p/2} && (23) \\ &\leq \nu_p 2^{-j\delta p/2} E \left\{ \sum_{(u,v) \in S_j'} K_0^2(u, v) u^{-\delta} \right\}^{p/2}. && (24) \end{aligned}$$

To obtain a bound on the expectation in (24), define the random sum

$$Z_j = \sum_{(u,v) \in S_j'} K_0^2(u, v) u^{-\delta}. \tag{25}$$

The bounds (19) and (20) imply that $K_0^2(u, v) u^{-\delta}$ is bounded by $C u^{2r+2\rho+1-\delta}$ for $0 < u \leq 1$, and by $C u^{-2r-2\rho-1-\delta}$ for $u \geq 1$; hence it is uniformly bounded for all u and v.

We now apply Corollary 1 in the Appendix to investigate the behaviour of $EZ_j^{p/2}$. To check the first condition, it will be sufficient to have finiteness of the integral

$$\int_0^\infty \int_{-\infty}^\infty K_0^2(u, v) u^{-\delta-\zeta} dv\, du. \tag{26}$$

The bounds (18) on the support of the integrand, and those stated above on its order of magnitude, allow (26) to be dominated by

$$C\int_0^1 u^{2r+2\rho+1-\delta-\zeta}(1+1/u)du + C\int_1^\infty u^{-2r-2\rho-1-\delta-\zeta}(1+1/u)du.$$

The first of these integrals is finite since $2(r+\rho)-\delta>0$ by the assumptions of the theorem, while the second is clearly finite. Now consider the second integral in the corollary. By similar arguments to those just used,

$$\begin{aligned}&\int_0^\infty \int_{-\infty}^\infty \{K_0^2(u,v)u^{-\delta}\}^{p/2}u^{-\zeta}dv\,du\\ \leq\ & C\int_0^1 u^{(2r+2\rho+1-\delta)p/2-\zeta-1}du + C\int_1^\infty u^{-(2r+2\rho+1+\delta)p/2-\zeta}du.\end{aligned}$$

The assumptions of the theorem about the regularity of the wavelets imply easily that these integrals are both finite. It now follows from Corollary 1 in the Appendix that, for each fixed p,

$$EZ_j^{p/2} = C2^{-j\zeta} + o(2^{-j\zeta}) \text{ as } j\to\infty. \tag{27}$$

Now define w_j to be the vector with elements w_{jk} for $k=0,...,2^j-1$. Substituting (27) into (22) gives the result

$$E||w_j||_p^p \leq C2^{-j(\delta p/2+\zeta-1)} \text{ for all } j. \tag{28}$$

Set $\gamma = s+1/2-\zeta/p-\delta/2 = s'+1/p-\zeta/p-\delta/2$. By Jensen's inequality and (28), we have

$$E||w||_{b^s_{p,1}} = \sum_{j=j_0}^\infty 2^{js'}E||w_j||_p \leq \sum_{j=j_0}^\infty 2^{js'}(E||w_j||_p^p)^{1/p} \leq C\sum_{j=j_0}^\infty 2^{j\gamma},$$

which is finite if $\gamma<0$. Thus, if $\gamma<0$, $||w||_{b^s_{p,1}}$ is finite almost surely. We have already shown that the term $||u_{j_0}||_p$ in (10) is finite almost surely. By the equivalence of norms given by (8) and (10), we conclude that, if $\gamma<0$, $||f||_{B^s_{p,1}}$ is finite almost surely, and therefore, by the embedding properties of Besov spaces (e.g. Peetre, 1975), $f\in B^s_{p,q}$ almost surely for all $1\leq q\leq\infty$, completing the proof.

3.3 Remarks and extensions

Remark 1. Theorem 1 places an upper bound restriction on the value of ζ. In the case $\zeta>1$, the intensity $\mu_a \propto a^{-\zeta}$ is integrable over the range of λ for which ψ_λ has support intersecting $[0,1]$. Therefore, the number of relevant terms in the atomic decomposition of f is finite almost surely. With probability one, f will belong to the same Besov spaces as the mother wavelet ψ, namely those for which $\max(0,1/p-1/2)<s<r$, $1\leq p\leq\infty$, $1\leq q\leq\infty$.

Remark 2. To investigate to what extent the condition (13) is necessary for the Besov space membership of f, note that all the $\mathcal{T}_j$ contain the set $\mathcal{T}_0 = [1,\infty)\times(-\frac{1}{2},\frac{1}{2})$. Therefore an argument along the lines following (23) gives the *lower* bound

$$E|w_{jk}|^p \geq \nu_p 2^{-j\delta p/2} E(Z_j^*)^{p/2},$$

where

$$Z_j^* = \sum_{(u,v)\in S_j'\cap\mathcal{T}_0} K_0^2(u,v)u^{-\delta}.$$

Corollary 1 in the Appendix can be applied to show that, with different constants C, the expression (27) also applies to Z_j^*. Hence, for sufficiently large j, $E||w_j||_p^p$ is bounded below by a multiple of $2^{-j(\delta p/2+\zeta-1)}$. If condition (13) fails to hold, we can conclude that

$$\sum_{j=j_0}^{\infty} 2^{js'p} E||w_j||_p^p = \infty,$$

and hence that the expected value of the pth moment of the $B^s_{p,p}$ norm of f is infinite. While this does not imply that the norm itself is infinite, it provides support for the conjecture that condition (13) is necessary as well as sufficient for the finiteness of the $B^s_{p,q}$ norms of f.

Theorem 1 restricted attention to finite p; the following theorem extends the basic result to the case $p=\infty$.

Theorem 2 *Under the same assumptions as Theorem 1 with $p=\infty$,*

$$f \in B^s_{\infty,q} \text{ almost surely}, \quad 1\leq q\leq\infty$$

if

$$s + 1/2 - \delta/2 < 0.$$

Proof. For any positive θ and c, Markov's inequality implies that

$$P(|w_{jk}|\geq c) \leq 2e^{-\theta c}E(e^{\theta w_{jk}}). \tag{29}$$

To evaluate the expectation, we use the standard expression for the moment generating function of a normal distribution and Campbell's theorem (see, for example, Kingman, 1993, p. 28) applied to the random sum (21) to obtain

$$\begin{aligned}
\log E(e^{\theta w_{jk}}) &= \log E\{E(e^{\theta w_{jk}}|S)\} = \log E\ \exp\{\tfrac{1}{2}\theta^2\sigma^2(S)\} \\
&= \int_{\mathcal{T}^0_{jk}} [\exp\{\tfrac{1}{2}\theta^2 K(\lambda,\lambda_{jk})^2 a^{-\delta}\}-1]a^{-\zeta}d\lambda \\
&= 2^{-\zeta j}\int_{\mathcal{T}_j}[\exp\{\theta^2 K_0(u,v)^2 2^{-1-\delta j}u^{-\delta}\}-1]u^{-\zeta}du\,dv
\end{aligned}$$

by the usual change of variable. Now let $M = \sup\{u^{-\delta}K_0(u,v)^2\}$, which was shown to be finite in the proof of Theorem 1. Extending the integral to the whole of the half-plane $u > 0$, we have, by the convexity of the exponential function,

$$\begin{aligned}\log E(e^{\theta w_{jk}}) &\leq C2^{-\zeta j}\exp(2^{-1-\delta j}\theta^2 M)\int_0^\infty\int_{-\infty}^\infty K_0(u,v)^2 u^{-\delta-\zeta}dv\,du \\ &\leq C2^{-\zeta j}\exp(M\theta^2 2^{-1-\delta j}).\end{aligned}$$

Suppose $2^{j\delta/2}c > 1$. To obtain a bound for $P(|w_{jk}| > c)$, choose $\theta^2 = M^{-1}2^{1+\delta j}\log(2^{j\delta/2}c)$, and substitute into (29) to obtain, for positive constants C_1 and C_2,

$$\log P(|w_{jk}| > c) \leq \log 2 + C_1 2^{(\delta/2-\zeta)j}c - C_2 2^{\delta j/2}c\sqrt{\log\{2^{j\delta/2}c\}}.$$

Finally, set $\gamma = \delta/2 - s'$, where $s' = s + \frac{1}{2}$. Both s' and γ are positive by the hypotheses of the theorem. Choose ϵ such that $0 < \epsilon < \gamma$ and set $c = 2^{-(s'+\epsilon)j}$. Then $2^{j\delta/2}c = 2^{(\gamma-\epsilon)j} > 1$. We now have

$$\begin{aligned}\log P(2^{(s'+\epsilon)j}|w_{jk}| > 1) &\leq \log 2 + C_1 2^{(\gamma-\epsilon-\zeta)j} - C_2 2^{(\gamma-\epsilon)j}\sqrt{\log 2^{(\gamma-\epsilon)j}} \\ &\leq -2^{(\gamma-\epsilon)j},\end{aligned}$$

for sufficiently large j. Since w_j is of length 2^j, it follows that, for sufficiently large j,

$$P(2^{s'j}||w_j||_\infty > 2^{-\epsilon j}) < 2^j\exp(-2^{(\gamma-\epsilon)j}).$$

This very rapidly decreasing bound on the tail probabilities implies that, with probability one, the sequence $2^{s'j}||w_j||_\infty$ is bounded by a multiple of $2^{-\epsilon j}$ and hence its sum is convergent almost surely. The arguments at the beginning of the proof of Theorem 1 show that $||u_{j_0}||_\infty$ is finite almost surely. By the equivalence of norms given by (9) and (10), we conclude that $f \in B^s_{\infty,1}$ almost surely and, therefore, by the embedding properties of Besov spaces, $f \in B^s_{\infty,q}$ almost surely for all $1 \leq q \leq \infty$, completing the proof.

4 Concluding remarks

The key conclusion of the theorems we have proved is that, under suitable conditions, the function f falls in $B^s_{p,q}$ if $\delta + (2/p)\zeta$ exceeds $2s+1$. Since the high frequency content of model functions depends both on the intensity of high frequency atoms and on their size, it is not surprising that the smoothness as measured by the parameter s should depend on both parameters. The parameter p can be seen as discouraging inhomogeneity, in that the larger the value of p the more emphasis is placed on the parameter δ. For large δ, no matter how many high frequency atoms there are, they each make a relatively low contribution. On the other hand, if p is small, then there is a trade-off where large weights on high frequency atoms (small δ) can be tolerated if the corresponding atoms are relatively rare (large ζ).

The statistical applications of the models and results of this paper are an interesting subject for future work. Various methods have been proposed for the estimation of functions by means of dictionaries of wavelets and related functions. For example, Daubechies (1988) proposed the so-called Method of Frames that picks up the representation with the minimal l_2-norm of its coefficients. The result is essentially an average of all possible representations. Chen *et al.* (1996) argue for mimimizing the l_1-norm and develop the corresponding Basis Pursuit algorithm. The Matching Pursuit of Mallat & Zhang (1993) yields a stepwise dictionary atom selection when on each step an atom that best correlates with the residual is added. For certain dictionaries Coifman & Wickerhauser (1992) suggested the procedure that selects the "best orthogonal basis" with the minimal entropy.

The stochastic atomic decompositions described in this paper, can, in principle, be used to set up suitable priors that allow intuitive notions about the function genuinely to be modelled. Bayesian wavelet approaches based on standard orthonormal dictionaries have been considered in several papers recently; see Abramovich *et al.* (1998) and other references cited there. The results reported have been very encouraging. The models we have considered in this paper allow Bayesian ideas to be extended to a broader range of wavelet models, freed from the dyadic positions and scales considered in the classical case, but the algorithmic details, probably involving modern Bayesian computational methods, have yet to be worked out in detail. The improvement to standard wavelet thresholding methods obtained by moving from the classical to the stationary wavelet transform (e.g. Coifman & Donoho, 1995; Nason & Silverman, 1995; Lang *et al.*, 1996; Johnstone & Silverman, 1997) suggest that a Bayesian approach based on a general atomic decomposition may result in yet better performing wavelet shrinkage estimators. This is an interesting topic for further research.

Acknowledgments

We are delighted to thank Boris Tsirelson for fruitful discussions. The financial support of the Engineering and Physical Sciences Research Council, the Israel Academy of Science, and the Royal Society are gratefully acknowledged. This work was completed while TS was a Research Associate in the Department of Mathematics, University of Bristol, supported by EPSRC grant GR/K70236, and while BWS was a Fellow at the Center for Advanced Study in the Behavioral Sciences, Stanford, partly supported by NSF grant SBR-9601236.

Appendix: Moments of sums of thinned Poisson processes

In this appendix we prove a lemma and corollary used in the proof of Theorem 1. They are of interest in their own right.

Lemma 1 *Let μ be a measure on $\mathbb{R}$ and let S_ε be a Poisson Process on $\mathbb{R}$ with intensity measure $\varepsilon\mu$, where $\varepsilon > 0$. Assume that μ satisfies the condition:*

$$\int_{-\infty}^{\infty} \min(1, |x|)\ \mu(dx) < \infty. \tag{30}$$

Define:

$$Y_\varepsilon = \sum_{X \in S_\varepsilon} X \tag{31}$$

and assume that, for some $l > 0$:

$$c_l = \int_{-\infty}^{\infty} |x|^l\ \mu(dx) < \infty \tag{32}$$

(but that μ is not necessarily itself finite). Then:

$$E|Y_\varepsilon|^l = \varepsilon\, c_l + o(\varepsilon) \quad \text{as} \quad \varepsilon \to 0. \tag{33}$$

Proof. Appealing to Campbell's Theorem (e.g. Kingman, 1993, p. 28), condition (30) shows that the random variable defined by (31) is well defined with probability one, in the sense that the right hand side of (31) is absolutely convergent with probability one.

For any $\delta > 0$, define sets $A_\delta = [-\delta, \delta]$ and $B_\delta = \mathbb{R} \setminus A_\delta$. It follows from (30) that $\mu(B_\delta) < \infty$; define $F(\delta) = \mu(B_\delta)$. Now choose δ to depend on ε in such a way that $\delta \leq 1$, and $\delta \to 0$ and $\varepsilon F(\delta) \to 0$ as $\varepsilon \to 0$. Since $F(\delta)$ increases as $\delta \to 0$ this puts a constraint on the rate at which $\delta(\varepsilon) \to 0$ as $\varepsilon \to 0$. Henceforth, the dependence of δ on ε will be expressed explicitly.

Write:

$$Y_\varepsilon^{(1)} = \sum_{X \in S_\varepsilon \bigcap B(\delta)} X \tag{34}$$

$$Y_\varepsilon^{(2)} = \sum_{X \in S_\varepsilon \bigcap A(\delta)} X. \tag{35}$$

Consider, first, the asymptotic behaviour of $Y_\varepsilon^{(1)}$ defined in (34). The number of X in $S_\varepsilon \bigcap B(\delta)$ is a Poisson $(\varepsilon F(\delta))$ random variable and so:

$$E|Y_\varepsilon^{(1)}|^l = \sum_{j=1}^{\infty} \exp(-\varepsilon F(\delta)) \frac{\varepsilon^j F(\delta)^j}{j!} E\left(|\sum_{i=1}^{j} X_i|^l \right), \tag{36}$$

where $X_1, X_2, \ldots$ are independent and identically distributed random variables on B_δ with distribution $\mu/F(\delta)$. Let:

$$c_l^{(1)} = \int_{B_\delta} |x|^l \, \mu(dx). \tag{37}$$

For $j \geq 2$, we consider a bound for the expectation in (36). For $l \leq 1$, we immediately see that:

$$E\left(|\sum_{i=1}^{j} X_i|^l \right) \leq E\left(\sum_{i=1}^{j} |X_i|^l \right) = jE\left(|X_i|^l \right) = \frac{j c_l^{(1)}}{F(\delta)}. \tag{38}$$

For $l > 1$, using Jensen's inequality, we have that:

$$E\left(|\sum_{i=1}^{j} X_i|^l \right) \leq E\left(j^{l-1} \sum_{i=1}^{j} |X_i|^l \right) = \frac{j^l c_l^{(1)}}{F(\delta)}. \tag{39}$$

Hence, in either (38) or (39), we have from (36), separating the terms for $j = 1$ and $j > 1$, that:

$$E|Y_\varepsilon^{(1)}|^l = \varepsilon \, c_l^{(1)} \exp(-\varepsilon F(\delta)) + R_\varepsilon^{(1)}, \tag{40}$$

where:

$$\begin{aligned} R_\varepsilon^{(1)} &\leq \sum_{j=1}^{\infty} \exp(-\varepsilon F(\delta)) \frac{\varepsilon^j F(\delta)^j}{j!} \frac{j^{(l \wedge 1)} c_l^{(1)}}{F(\delta)} \\ &= \varepsilon^2 c_l^{(1)} F(\delta) \sum_{k=0}^{\infty} \exp(-\varepsilon F(\delta)) \frac{\varepsilon^k F(\delta)^k}{k!} \frac{(k+2)^{(l-1)_+}}{(k+1)}. \end{aligned} \tag{41}$$

As $\varepsilon \to 0$, the sum in (41) is a Poisson expectation that converges to $2^{(l-1)_+}$, since $\varepsilon F(\delta) \to 0$. It follows that $R_\varepsilon^{(1)} = o(\varepsilon)$ and hence, from (40), that, as $\varepsilon \to 0$:

$$\varepsilon^{-1} E|Y_\varepsilon^{(1)}|^l \to c_l, \tag{42}$$

using the facts that $c_l^{(1)} \to c_l$ and $\varepsilon F(\delta) \to 0$.

Consider now the asymptotic behaviour of $Y_\varepsilon^{(2)}$ defined in (35). For $l \leq 1$, by Campbell's theorem applied to the Poisson process $S_\varepsilon \bigcap A(\delta)$, we have that:

$$\begin{aligned} E|Y_\varepsilon^{(2)}|^l &\leq E\left(\sum_{X \in S_\varepsilon \bigcap A(\delta)} |X|\right)^l \\ &\leq E\left(\sum_{X \in S_\varepsilon \bigcap A(\delta)} |X|^l\right) \\ &= \varepsilon \int_{A_\delta} |x|^l \, \mu(dx). \end{aligned} \tag{43}$$

Therefore, we have from (43), as $\delta \to 0$, that $E|Y_\varepsilon^{(2)}|^l = o(\varepsilon)$. Since:

$$\left| |Y_\varepsilon|^l - |Y_\varepsilon^{(1)}|^l \right| \leq |Y_\varepsilon^{(2)}|^l,$$

it follows from (42) that:

$$E|Y_\varepsilon|^l = \varepsilon \, c_l + o(\varepsilon) \quad \text{as} \quad \varepsilon \to 0. \tag{44}$$

Finally, consider the case $l > 1$. Define:

$$Z_\varepsilon = \sum_{X \in S_\varepsilon \bigcap A(\delta)} |X|. \tag{45}$$

For $-\delta \leq x \leq \delta$, $0 \leq e^{|x|} - 1 \leq \delta^{-1}(e^\delta - 1)|x|$, and so, using (30), we have:

$$I_\delta = \int_{-\delta}^{\delta} (e^{|x|} - 1) \, \mu(dx) < \infty. \tag{46}$$

It then follows (using equation (3.17) of Kingman, 1993), that:

$$E \exp(Z_\varepsilon) = \exp(\varepsilon I_\delta). \tag{47}$$

Since $l > 1$, there exists a constant k_l such that $z^l \leq k_l(e^z - 1)$ for all $z \geq 0$. We then have:

$$\begin{aligned} EZ_\varepsilon^l &\leq k_l(E \exp(Z_\varepsilon) - 1) \\ &= k_l(\exp(\varepsilon I_\delta) - 1) \\ &= o(\varepsilon), \quad \text{as} \quad \varepsilon \to 0, \end{aligned} \tag{48}$$

since $I_\delta \to 0$ as $\delta \to 0$. Therefore, from (48), it follows at once that:

$$E|Y_\varepsilon^{(2)}|^l = o(\varepsilon), \quad \text{as} \quad \varepsilon \to 0.$$

Using Minkowski's inequality applied to the norm $||X||_l = \left(E|X|^l\right)^{1/l}$ we have:

$$\left(E|Y_\varepsilon^{(1)}|^l\right)^{1/l} - \left(E|Y_\varepsilon^{(2)}|^l\right)^{1/l} \leq \left(E|Y_\varepsilon|^l\right)^{1/l} \leq \left(E|Y_\varepsilon^{(1)}|^l\right)^{1/l} + \left(E|Y_\varepsilon^{(2)}|^l\right)^{1/l}$$

and hence, since $\varepsilon^{-1}E|Y_\varepsilon^{(2)}|^l \to 0$, the limiting values of $\varepsilon^{-1}E|Y_\varepsilon^{(1)}|^l$ and $\varepsilon^{-1}E|Y_\varepsilon|^l$ are the same. It follows from (42) that $\varepsilon^{-1}E|Y_\varepsilon|^l \to c_l$, completing the proof of the lemma.

Corollary 1 *Let μ be a measure on a set Ω, and let g be a real-valued function on Ω. Let S_ε be a Poisson Process on Ω with intensity measure $\varepsilon\mu$, where $\varepsilon > 0$. Assume that the induced measure $\mu(g^{-1}(A))$ is non-atomic for every measurable set $A \subseteq \mathbb{R}$, and assume that:*

$$\int_\Omega \min(1, |g(x)|)\ \mu(dx) < \infty.$$

Define:

$$Y_\varepsilon = \sum_{X \in S_\varepsilon} g(X)$$

and assume that, for some $l > 0$:

$$u_l = \int_\Omega |g(x)|^l\ \mu(dx) < \infty$$

(but that μ is not necessarily itself finite). Then:

$$E|Y_\varepsilon|^l = \varepsilon\ u_l + o(\varepsilon) \quad \text{as} \quad \varepsilon \to 0.$$

Proof. Define a measure on $\mathbb{R}$ by $\mu_g(A) = \mu(g^{-1}(A))$ and let $Z = g(X)$. Then, appealing to the Mapping theorem for Poisson processes (e.g. Kingman, 1993, p. 18), Z is a Poisson process on $\mathbb{R}$ with intensity measure $\varepsilon\mu_g$ and $Y_\varepsilon = \sum_{X \in S_\varepsilon} Z$. The proof of the corollary is completed by applying Lemma 1 to Z and μ_g.

References

Abramovich, F., Sapatinas, T. & Silverman, B.W. (1998). Wavelet thresholding via a Bayesian approach. *Journal of the Royal Statistical Society*, Ser. B, **60** (in the press).

Anderson, L., Hall, N., Jawerth, B. & Peters, G. (1993). Wavelets on closed subsets of the real line. In *Recent Advances in Wavelet Analysis*, Schumaker, L.L. and Webb, G. (Eds.), pp. 1-61, New York: Academic Press.

Bruce, A. & Gao, H.-Y. (1996). *Applied Wavelet Analysis with S-Plus.* New York: Springer-Verlag.

Chen, S.S., Donoho, D.L. & Saunders, M.A. (1996). Atomic decompositions by basis pursuit. *Technical Report*, February 1996, Department of Statistics, Stanford University, USA.

Cohen, A., Daubechies, I. & Vial, P. (1993). Wavelets on the interval and fast wavelet transforms. *Applied and Computational Harmonic Analysis* **1**, 54-81.

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

Coifman, R.R. & Wickerhauser, M.V. (1992). Entropy-based algorithms for best-basis selection. *IEEE Transactions on Information Theory* **38**, 713-718.

Daubechies, I. (1988). Time-frequency localization operators: a geometric phase space approach. *IEEE Transactions on Information Theory* **34**, 605-612.

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

Davis, G., Mallat, S.G. & Zhang, Z. (1994). Adaptive time-frequency approximations with matching pursuit. In *Wavelets: Theory, Algorithms, and Applications*, Chui, C.K., Montefusco, L. & Puccio, L. (Eds.), pp. 271-293, San Diego: Academic Press.

DeVore, R.A., Jawerth, B. & Popov, V. (1992). Compression of wavelet decompositions. *American Journal of Mathematics* **114**, 737-785.

DeVore, R.A. & Popov, V. (1988). Interpolation of Besov Spaces. *Transactions of the American Mathematical Society* **305**, 397-414.

Donoho, D.L. & Johnstone, I.M., Kerkyacharian, G. & Picard, D. (1995). Wavelet Shrinkage: Asymptopia? (with discussion). *Journal of the Royal Statistical Society*, Ser. B, **57**, 301-369.

Jawerth, B & Sweldens, W. (1994). An overview of wavelet based multiresolution analyses. *SIAM Review* **36**, 377-412.

Johnstone, I.M. (1994). Minimax Bayes, asymptotic minimax and sparse wavelet priors. In *Statistical Decision Theory and Related Topics, V*, Gupta, S.S and Berger, J.O. (Eds.), pp. 303-326, New-York: Springer-Verlag.

Johnstone, I.M. & Silverman, B.W. (1997). Wavelet threshold estimators for data with correlated noise. *Journal of the Royal Statistical Society*, Ser. B, **59**, 319-351.

Kingman, J.F.C. (1993). *Poisson Processes.* Oxford: Oxford University Press.

Lang, M., Guo, H., Odegard, J.E., Burrus, C.S. & Wells, R.O. (1996). Noise reduction using an undecimated discrete wavelet transform. *IEEE Signal Processing Letters* **3**, 10-12.

Mallat, S.G. & Zhang, Z. (1993). Matching pursuit in a time-frequency dictionary. *IEEE Transactions on Signal Processing* **41**, 3397-3415.

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

Nason, G.P. & Silverman, B.W. (1994). The discrete wavelet transform in S. *Journal of Computational* & *Graphical Statistics* **3**, 163-191.

Nason, G.P. & Silverman, B.W. (1995). The stationary wavelet transform and some statistical applications. In *Wavelets and Statistics*, Lecture Notes in Statistics **103**, Antoniadis, A. and Oppenheim, G. (Eds.), pp. 281-300, New York: Springer-Verlag.

Ogden, R.T. (1997). *Essential Wavelets for Statistical Applications and Data Analysis.* Boston: Birkhäuser.

Peetre, J. (1975). *New Thoughts on Besov Spaces.* Durham: Duke University Press.

Wojtaszczyk, P. (1997). *A Mathematical Introduction to Wavelets.* Cambridge: Cambridge University Press.